Title: GENOMIC INSIGHTS INTO THE CAUSES OF TYPE 2 DIABETES

Article Type: Invited Review

Corresponding Author: XX Claudia Langenberg,
Corresponding Author's Institution:
First Author: Claudia Langenberg

Order of Authors: Claudia Langenberg; Luca A Lotta

Abstract: Genome-wide association studies of type 2 diabetes have implicated up to ~250 genomic regions in disease predisposition, with evidence for causal variants and genes emerging for several of these regions. Understanding of the underlying mechanisms, including the interplay between beta-cell failure, insulin sensitivity, appetite regulation, and adipose storage has been facilitated by the integration of multi-dimensional data on diabetes-related intermediate phenotypes, detailed genomic annotations, functional experiments and now multi-"omic" molecular features. Studies in diverse ethnicities and examples from population isolates have highlighted the value and need for a broader genomic approach to this global disease. Ongoing trans-ethnic discovery efforts and large-scale biobanks in diverse populations and ancestries may help to address some of the existing "Eurocentric" bias. Despite rapid progress in the discovery of the highly-polygenic architecture of type 2 diabetes, dominated by common alleles with small, cumulative effects on disease risk, current knowledge has shown little clinical utility for disease prediction or prevention, and only small contributions to subtype classification or stratified approaches to treatment. Successful development of academia-industry partnerships for exome or genome sequencing in large Biobanks can deliver economies of scale, with implications for the future of genomics-focused research.
GENOMIC INSIGHTS INTO THE CAUSES OF TYPE 2 DIABETES

Claudia Langenberg and Luca A. Lotta

MRC Epidemiology Unit, University of Cambridge, United Kingdom.

Abstract

Genome-wide association studies of type 2 diabetes have implicated up to ~250 genomic regions in disease predisposition, with evidence for causal variants and genes emerging for several of these regions. Understanding of the underlying mechanisms, including the interplay between beta-cell failure, insulin sensitivity, appetite regulation, and adipose storage has been facilitated by the integration of multi-dimensional data on diabetes-related intermediate phenotypes, detailed genomic annotations, functional experiments and now multi-“omic” molecular features. Studies in diverse ethnicities and examples from population isolates have highlighted the value and need for a broader genomic approach to this global disease. Ongoing trans-ethnic discovery efforts and large-scale biobanks in diverse populations and ancestries may help to address some of the existing “Eurocentric” bias. Despite rapid progress in the discovery of the highly-polygenic architecture of type 2 diabetes, dominated by common alleles with small, cumulative effects on disease risk, current knowledge has shown little clinical utility for disease prediction or prevention, and only small contributions to subtype classification or stratified approaches to treatment. Successful development of academia-industry partnerships for exome or genome sequencing in large Biobanks can deliver economies of scale, with implications for the future of genomics-focused research.
1. A global view of type 2 diabetes genomics

The genetic basis of type 2 diabetes

In the last decade, hypothesis-free genome-wide association studies (GWAS) have been the single most important contributor to identifying genetic determinants of type 2 diabetes (T2D), leading to the discovery of ~100 associated genomic regions or loci.\textsuperscript{1-4} The last year has seen a game-change, a leap forward from smaller, cumulative advances to the description of now up to ~250 genome-wide significant loci,\textsuperscript{5,8} including a large meta-analysis currently only available in pre-publication format.\textsuperscript{8} Several developments have enabled such rapid progress.

A tripling in effective sample size was achieved through the integration of large-scale, accessible resources such as the UK Biobank (http://www.ukbiobank.ac.uk), teamed with the openness of diabetes researchers worldwide to share results of previous studies (GWAS results: http://diagram-consortium.org; exome sequencing and genotyping results: http://www.type2diabetesgenetics.org). This has led to the inclusion of up to 74,000 T2D cases and 820,000 controls of European-descent,\textsuperscript{7,8} with an accompanying increase in power.

The breadth and depth of genetic variation ascertainment has also dramatically improved. Dense and accurate genotype imputation using the sequenced haplotypes of the 1000 Genomes\textsuperscript{5} and now the Haplotype Reference Consortium\textsuperscript{7,8} has enabled the interrogation of over 10 million genetic variants, with a ~100-fold increase compared with early GWAS efforts.\textsuperscript{1} Direct genotyping or sequencing of common to rare alleles of the exonic areas of the genome has enabled a better ascertainment of coding variation.\textsuperscript{7,10} As a consequence, T2D susceptibility variants now range from 0.02-50% in minor allele frequency (MAF), and from 1.04 to 8.05 in per-allele odds ratio.\textsuperscript{5,8} While most signals are led by common variants with ever smaller effects, new risk alleles include several that are low-frequency or rare.\textsuperscript{5-8} This is consistent with the model of heritability of T2D derived from whole-genome sequencing experiments, characterised by a prominent contribution to heritability of common variation, a small contribution of rare variation and evidence of low selective pressure on predisposition alleles.\textsuperscript{10} Future meta-analyses of multiple large biobanks are likely to expand the catalogue of susceptibility variants to some ultra-rare alleles with extreme effect sizes (odds ratios >10), in addition to finding even more common regulatory alleles with very small effects (odds ratios <1.01; Figure 1). Direct genotyping and sequencing will be critical to replicate and identify associations for risk variants that are in the rare allele frequency spectrum (MAF below 0.5%),\textsuperscript{11} specifically if discovered based on imputed genotypes given the difficulties in accurately imputing these variants even with expanded population reference panels.\textsuperscript{9}

Clinical translation of these genomic associations critically depends on our understanding of underlying mechanisms. Establishing causal variants and variant-gene links has been a challenge in genetic studies. Denser imputation,\textsuperscript{5,8} extended genotyping in coding or metabolic-trait associated regions,\textsuperscript{3,6,7} direct sequencing,\textsuperscript{10,12} larger sample size,\textsuperscript{6,8} integration of extensive genomic and regulatory annotations,\textsuperscript{13,14} and progress in analytical fine-mapping approaches\textsuperscript{15,16} have all made this fundamental task easier. For a given association signal, fine-mapping has been able to considerably narrow the size of the genomic region that likely contains the causal variant as well as the list of plausible causal variants in that genomic region.\textsuperscript{5,7,8,14} Identified missense or nonsense variants with evidence of causal association\textsuperscript{7,8} are now amenable to more direct in vitro and in vivo experimental follow-up. These represent critical advances in the translation of robust associations into biological understanding.
Trans-ethnic discovery efforts

The global burden of T2D with large differences in risk within and between populations warrants a global genomic approach to study its predisposition. Despite the progress in the discovery of its genetic basis, T2D is no exception when it comes to the continued underrepresentation of ethnic diversity in genetic research and discovery efforts. Genetic studies in multiple ethnicities are valuable for several reasons: (a) susceptibility variants may be present at appreciable allele frequency only in non-European populations, as elegantly shown by the discovery of associations at the HNF1A (via exome sequencing)\(^{18}\), SLC16A11 (via GWAS),\(^{19}\) and IGF2 (via exome genotyping)\(^{20}\) loci in populations of Latin American origin; (b) for shared susceptibility loci, trans-ethnic studies increase statistical power for new discoveries;\(^{4,6}\) (c) the diverse linkage disequilibrium patterns across ancestries increase the resolution of fine-mapping analyses for the identification of causal variants;\(^{5,4}\) (d) exposure to diverse environments may reveal the effect of susceptibility variants which are masked in other settings (i.e. gene-environment interactions).

Important evidence has been emerging from GWAS of South Asian, East Asian, Latin American and African American populations\(^{19,23}\) and trans-ethnic discovery efforts;\(^{4,6}\) however, studies for non-European ancestries remain small in comparison. Ongoing trans-ethnic discovery efforts under the umbrella of the DIAMANTE consortium include over 170,000 T2D cases with around 45% of the effective sample size accounted for by non-European ancestries, currently including African (7%), East Asian (23%), Hispanic or Latino (6%) and South Asian (9%) ancestry participants (personal communication Prof Andrew Morris, University of Liverpool). While this represents a substantial advance compared to previous studies, research in the field remains heavily “Eurocentric”.

Genetically isolated populations

Recent studies have clearly demonstrated the value of studying genetically isolated populations. In these settings deleterious variants with large phenotypic effect may raise by chance to higher allele frequencies, due to a phenomenon known as allelic drift.\(^{26}\) This makes it easier to identify disease associations for such variants in isolated compared to admixed populations, in which these variants may not be present or may be very rare. Such findings can provide insight into the aetiology of T2D that is generalizable outside of the context of the particular population in which they were discovered.

In the Inuit population from Greenland, homozygote carriers of a loss-of-function variant in TBC1D4 (p.Arg684*) were found to have a ~10 fold higher risk of diabetes, and ~1 standard deviation higher glucose and insulin at 2 hours following an oral glucose challenge.\(^{27}\) The variant segregates at high frequency in Inuits (MAF 17%), while being rare or monomorphic in other populations. Risk allele carriers have lower levels of GLUT4 in skeletal muscle\(^{27}\) (Figure 2). This finding highlights the causal role of insufficient GLUT4-mediated glucose uptake in muscle for postprandial hyperglycaemia and T2D risk (Figure 2). This corroborates evidence from the first report of a mutation significantly impairing GLUT4 translocation, identified in a child with acanthosis nigricans and extreme postprandial hyperinsulinemia carrying a heterozygous premature stop mutation (p.Arg363*) in TBC1D4.\(^{28}\) In the same population, exome sequencing has revealed the association of loss-of-function variant in ADCY3 with obesity and diabetes,\(^{29}\) a finding supported publicly available trans-ethnic T2D exome sequencing studies (http://www.type2diabetesgenetics.org),\(^{29}\) as well as studies of consanguineous families with severe obesity.\(^{30}\)

Other examples also highlight convergence of evidence from studies of monogenic disease and population isolates. An autosomal dominant missense mutation in AKT2 was originally identified as the cause of hyperinsulinemia and diabetes in a family with severe insulin resistance and partial lipodystrophy.\(^{31}\) A recent exome-array and sequencing study detected a low-frequency AKT2 coding variant (p.Pro50Thr, MAF 1.1% in Finnish participants, but not detected in other populations) to be associated with higher fasting insulin levels,\(^{32}\) lower uptake of glucose in insulin-sensitive tissues
(shown in a separate recall-by-genotype study) and higher T2D risk, further highlighting the role of AKT2 in insulin sensitivity.

In Samoans, a founder population with a high prevalence of obesity and T2D, a common (26% allele frequency in Samoans but extremely rare in other populations) “thrifty” missense variant in CREBRC (rs373863828, p.Arg457Gln) was associated with substantially higher body mass index (1.4 kg/m² per allele). Interestingly, the adiposity-raising allele was associated with lower fasting glucose and protection from T2D. In an adipocyte model, overexpression of Arg457Gln selectively decreased energy use and increased fat storage, in line with its effect on body fat, potentially highlighting the metabolic benefits of a greater capacity of fat storage.
2. Insights into pathways to diabetes through genomic discovery

Genetic studies of diabetes-related intermediate phenotypes in non-diabetic individuals have emerged as a way to gain mechanistic insights into T2D susceptibility that is complementary to disease-focused discovery GWAS.

**Glycaemic control and susceptibility to diabetes**

Genetic studies of glucose and insulin related measures have been defined by the struggle between sample size and degree of refinement of phenotype ascertainment. Large efforts have focused on widely-available, simple measures as fasting glucose or insulin, and glycated haemoglobin. In contrast, efforts based on “gold-standard” measures involving frequently-sampled oral glucose tolerance tests or continuous intravenous measurements that are difficult to obtain at scale have had limited sample sizes. The former approach has been successful in the discovery of loci influencing glycaemic traits in non-diabetic individuals, and helped to identify insulin secretory effects as a major driver of associations for several of the common diabetes susceptibility loci (Figure 3). These studies also revealed considerable aetiologic heterogeneity in pathways to T2D, highlighting the multifactorial nature of T2D predisposition.

Overlying diabetes susceptibility variants, glycaemic traits and pancreatic islet regulatory and functional data has provided the foundation for an improved understanding of mechanisms linking beta-cell glucose sensing and insulin secretion with T2D risk. High-throughput functional screens of gene silencing in human beta-cell lines are now available and can empower systematic characterisation of the functional impact of novel likely-causal genes on insulin secretion.

A future challenge in this field will be for intermediate trait studies to stay apace with the rapid increases in sample size of diabetes association analyses. This is complicated by the practical difficulties of obtaining fasting samples, let alone more invasive “gold standard” intravenous measurements, in large biobanks.

**Excess overall fat and “central” role of peripheral fat**

Excess fat is the hallmark of overeating and lack of physical exercise and has been a major focus of genetic research. Large-scale studies of body mass index and related measures have linked genes highly expressed in the central nervous system with general obesity in different ancestries. This has provided complementary evidence to original discoveries that have revealed the fundamental role of appetite regulation in monogenic obesity, in particular the leptin-melanocortin axis.

Observational epidemiology has clearly shown that, for a given level of overall adiposity, the distribution of fat in the body is associated with susceptibility to insulin resistance, diabetes and its complications. Genetic discovery approaches integrating multiple insulin-resistance related phenotypes have identified variants associated with insulin resistance, increased risk of diabetes and coronary disease, but lower fat mass in peripheral body compartments, in particular legs and subcutaneous regions. These insulin resistance loci are enriched with genes harbouring mutations in Mendelian forms of lipodystrophy and are associated with increased odds of severe lipodystrophic insulin resistance. In the context of other evidence about the protective role of fat deposition in peripheral compartments, these results suggest that reduced ability to safely store excess energy...
in the peripheral regions of the body leads to ectopic fat storage and higher cardio-metabolic risk in the general population, similar to clinical manifestations of severe forms of lipodystrophy (Figure 4).

Molecular mechanisms underlying this specific aetiology include the impaired ability to generate new adipocytes and the regulation of gene expression in these cells, which is supported by experimental evidence around PPAR, KLFL4, IRS1, CCDC92, DNAH10 and L3MBTL3. Impaired intravascular lipoprotein lipase (LPL)-mediated lipolysis, the mechanism that regulates lipid buffering from the circulation to peripheral tissues, is also implicated (Figure 4). Further studies are necessary to understand how fat deposition in specific body compartments influences metabolic disease risk in the general population and move beyond the very simple notion of “apple” and “pear” body shapes.

Prioritising causal pathways by integrating multi-omic data with clinical outcomes

As the sample size of genetic studies has rapidly grown, so has the ability to measure detailed molecular features in biological samples using high-throughput technology. Global patterns of methylation and other epigenetic features (epigenomics), gene expression (transcriptomics), proteins (proteomics) or metabolites (metabolomics) can now be measured at epidemiological scale enabling genetic mapping in genome-wide studies.

Genetic studies of “-omics” molecular features can help advance the understanding of the causes of diabetes (and other complex) diseases in multiple ways, by (a) characterising the phenotypic consequences of diabetes susceptibility variants identified by GWAS; (b) helping to identify causal variants and genes at known susceptibility loci; (c) enabling the estimation of causal associations between molecular traits and disease risk using the principles of Mendelian randomisation.

So far, no studies have systematically followed-up associations of T2D susceptibility variants with circulating metabolomic or proteomic profiles. Studies overlaying regulatory annotations and gene expression in pancreatic beta-cells with T2D GWAS results show the value of transcriptomic analyses for the identification of diabetes susceptibility genes and mechanistic understanding. The increasing availability of similar data on a variety of cell types (including other relevant metabolic tissues, such as skeletal muscle, adipose, liver) will make it possible to systematically assess the relevance of different tissues, cell types and tissue-specific mechanisms in T2D pathophysiology.

“Mendelian Randomization” studies using genetics to assess causal associations between molecular traits and T2D risk have traditionally focused on specific pathways and biomarkers rather than “-omics” profiles. This partly reflects challenges in applying causal inference frameworks in the context of correlated and co-regulated molecular exposures such as blood metabolites or proteins. For example, studies of the branched chain amino acid pathway have provided human genetic evidence of multi-directional causal relationships between their metabolism, insulin resistance and risk of diabetes, building upon observations dating back to the 1960s. This and other metabolomics genetic studies illustrate how variation at key metabolic regulators (e.g. enzymes) affects a large set of biologically and phenotypically related measures within a pathway. Hence, inference from associations with diabetes may be restricted to a specific locus of regulatory importance rather than levels of one or more metabolites. Studies of the epigenetic patterns associated with obesity and T2D have illustrated that these are mostly consequences rather than the cause of disease processes.

New methods integrating genomic and other “-omics” data have been developed to facilitate more sophisticated studies in this field. While individual multi-omic studies have already been conducted in sample sizes of several thousands, these deeply phenotyped epidemiological studies are still comparatively small in the context of genomic research and the relevance of identified loci to metabolic diseases remains largely unexplored. However, they can serve as models for what might be achieved when technology and cost developments enable high-throughput multi-omic phenotyping at
the scale of large national Biobanks with hundreds of thousands of participants systematically followed up for a broad range of diseases.
3. Genomics and therapeutics in T2D

The Chief Medical Officer’s Report “Generation Genome” published in 2017 highlighted key areas in which genomics can inform disease therapeutics, providing a helpful framework to describe recent developments of genomics-informed therapeutics in T2D (Table 1).

New targets

The identification of novel drug targets is a major stated objective of T2D genetic studies. The finding that drug targets supported by human genetics evidence are more likely to succeed in the selective drug development process and recent success in the development of new lipid-lowering drugs following genetic findings from different approaches and sources at the PCSK9, LPA, APOC3 and ANGPTL3 loci have exponentially increased interest in this approach.

Three diabetes-susceptibility genes identified in early GWAS studies encode drug targets for existing glucose lowering therapy (PPARG, thiazolidinediones; KCNJ11/ABCC8, sulfonylureas), suggesting that other genes identified through hypothesis-free approaches may become new drug targets. However, human genetics has played a peripheral role in the development of the most recently approved classes of glucose-lowering drugs, including dipeptidyl peptidase-4 inhibitors, GLP1R agonists and sodium-glucose co-transporter-2 inhibitors.

With over a hundred genetic loci now robustly associated with diabetes, why is there still not a clear “PCSK9-like” example? Generic and therapeutic-area specific obstacles may play a role. First, several of the causal genes implicated by GWAS have only recently emerged, but it takes several years for new drugs to enter clinical development. Specifically targeting pancreatic islets, adipocytes, skeletal myocytes or brain cells, as opposed to circulating proteins is challenging. Safety concerns have been a limitation in the development of appetite suppressants modifying genetically-validated targets, as illustrated by the failure of first generation agonists of the melanocortin 4 receptor due to on-target side effects. Also, many loci for T2D act via impaired insulin secretion. The existence of different classes of approved and widely-used insulin secretagogues may limit the interest of pharmaceutical companies in new drug development in this area, given the focus on developing commercially-differentiated products.

Protective loss-of-function variants are particularly interesting for drug development purposes because they provide insights into the likely consequences of inhibiting a gene product and, if carriers are healthy, provide initial evidence of the likely safety of pharmacological inhibition. Sequencing of the early T2D GWAS gene SLC30A8, encoding a pancreatic islet zinc transporter (ZnT8), has identified rare loss-of-function variants associated with protection against T2D (odds ratio for carriers, 0.34). More recently, an exome array genotyping study in populations of Latino descent identified a protective variant (odds ratio per allele, 0.80) in IGF2 associated with incorrect splicing of isoform 2 of the gene, suggesting that selectively inhibiting this isoform in relevant tissues may be therapeutically exploited. However, efforts to inhibit ZnT8 or insulin-like growth factor 2 have yet to reach clinical development. The potential therapeutic implications of recently reported protective associations of a loss-of-function variant in GPR151 against obesity, diabetes and coronary artery disease also deserve consideration.

Genetically-tailored treatment

Diabetes medicine holds some of the most elegant examples of tailoring treatment to the specific underlying genetic or molecular defect, but all of these relate to monogenic forms of the disease. No such examples exist for common susceptibility loci, but this area is certainly understudied.

Response to glucose-lowering treatment shows a measurable degree of heritability and evidence of polygenicity, indirectly suggesting that combinations of multiple alleles might be able to identify
patients who would be more or less responsive to certain drugs. Opportunities in this field arise from
the definition and characterization of polygenic scores combining common variants that capture a
particular aetiology. Variants with large effects, such as those from population isolates or rare variants
with large effects from GWAS in admixed populations could also provide suitable basis for
pharmacogenetics applications. By studying the functional consequences of all theoretically possible
missense variants in PPARG, Majithia and colleagues elegantly showed that diabetes-associated
mutations in the gene display heterogeneous in vitro response to thiazolidinediones, which could
provide the basis for tailored therapy or dosing in carriers of these specific alleles, as illustrated in
initial case reports.

**Drug dosing or response**

Efforts to identify interactions between genetic background and T2D treatment have been the
subject of a recent systematic review, reporting that research in the field is mostly based on
observational studies rather than randomized controlled trials and candidate gene rather than
hypothesis-free approaches, with a few notable exceptions discussed below. In a pharmacogenetic
clinical trial, Srinivasan et al. found that TCF7L2 variants associated with T2D influences the acute
response to both glipizide and metformin in people with risk factors for T2D or treatment-naïve T2D
patients. Two genome-wide association studies have identified common genetic variants at the
ATM and SLC2A2 loci associated with response to metformin. The difference in the effect of
metformin for these variants was estimated at around ~0.15-0.17% of HbA1c per allele, roughly
corresponding to a daily dose of ~250 mg of metformin. A genome-wide discovery embedded
into a clinical trial found novel associations for common and rare variants in \( \text{PPAR}\_\text{G} \), \( \text{CPA}\_\text{6} \), and
\( \text{STAT}\_\text{3} \) with metformin response. While these findings are important for the understanding of
genetic susceptibility to drug response, the low price of metformin and pragmatic focus on reaching
the HbA1c therapeutic target or the maximum tolerated dose of this drug are barriers to clinical use of
these genetic tests.

**Drug repurposing**

If “pharmacomimetic” genetic variants can be used to find new targets, they could theoretically
be used to find new indications for existing drugs. While there are not yet any established examples of
genetically-directed repurposing of approved drugs in diabetes, Imamura et al. used a systematic
bioinformatics approach to identify new T2D drug targets, revealing potential repurposing
opportunities for drugs targeting the gene products of \( \text{GSK}\_\text{3B} \) and \( \text{JUN} \). Recent findings around
lipoprotein lipase may offer an example of genetically-driven extension of the target population for
drugs that are in active development. In late 2016, we reported a gain-of-function variant (rs328,
p.Ser447*) in \( \text{LPL} \) associated with insulin sensitivity and protection from diabetes and an independent
loss-of-function variant (rs1801177, p.Asp36Asn) associated with higher diabetes risk. This
followed directionally consistent findings for triglyceride levels and heart disease, leading to the
hypothesis that the several agents targeting the LPL pathway that are in development for the treatment
of hypertriglyceridemia could also be valuable as insulin sensitizing agents. The association
with diabetes of rs328 has since been replicated and the variant or its proxies have emerged in
recent GWAS of diabetes, with consistent findings published for a loss-of-function in the natural
LPL-inhibitor \( \text{ANGPTL}\_\text{4} \) (rs116843064, p.Glu40Lys), lending powerful support to this hypothesis.

**Drug safety**

Genetic variants have been used to understand both desired and undesired secondary effects of
pharmacological modulation. Similar to statins and genetic variants at their target \( \text{HMGCR} \), also
cholesterol lowering alleles at \( \text{NPC1L1} \) (encoding the target of ezetimibe) and \( \text{PCSK9} \) (PCSK9
inhibitors) are associated with a lower risk of coronary heart disease, but higher diabetes risk.
While this suggested that also non-statin cholesterol lowering agents could be associated with higher diabetes risk, recent randomized controlled trials of ezetimibe\textsuperscript{125} and PCSK9-inhibitors\textsuperscript{126} did not identify large or statistically-significant diabetogenic effects.

Cardiovascular associations of genetic variants that mimic diabetes medications are particularly insightful, given the regulatory requirement that glucose-lowering drugs should not be associated with a higher risk of cardiovascular disease\textsuperscript{127} and the ongoing paradigm shift from glycemic control to prevention of complications in diabetes management.\textsuperscript{128} The concomitant publication of randomized controlled trials of a glucagon-like peptide receptor 1 (GLP1R) agonists\textsuperscript{129, 130} and of genetic studies of a putative gain-of-function variant of GLP1R (rs10305492, p.Ala316Thr),\textsuperscript{131} both showing cardio-protective associations for GLP1R activation, illustrates this concept. A similar approach has shown cardio-protective associations for a functional variant in ABCC8 (rs757110, p.Ala1369Ser),\textsuperscript{132} providing genetic insights into the cardiovascular effects of sulfonylureas for which clinical trial evidence is inconclusive.\textsuperscript{133}
4. Clinical relevance and future outlook

Genomics in the clinic

In addition to diabetes therapeutics, genomics has been proposed to help disease prediction and diagnosis of common subtypes. The polygenic architecture of T2D dominated by many common variants with small effects and tagging several different aetiologies has critical implications for both applications.

The level of prediction achieved by common alleles contrasts with patients’ often more deterministic understanding of what “genetic risk” constitutes, which is informed by examples of highly penetrant causes of monogenic diseases. Accurate prediction in combination with preventive lifestyle interventions can be used for targeted primary prevention and avoidance of complications arising from metabolic dysregulation that is present for years before diagnosis. However, existing diabetes prediction models perform well in the general population and can be used to non-invasively identify individuals at high risk. Earlier prospective studies using up to 65 variants have shown that polygenic T2D risk scores improve prediction performance only modestly when considered over and above risk factors that can be assessed non-invasively, such as age, sex, body mass index or family history. New approaches based on machine learning and complex statistical modelling have been proposed as improved methods for genetic prediction. Critically, the availability of good inexpensive predictors from a patient’s anamnesis or examination (e.g. family history, BMI) together with clinically established tests (i.e. blood glucose, HbA1c) that inform both future risk prediction and diagnosis greatly limits the scope for introduction of genetic information for T2D prediction in the clinic. Evidence exists that established genetic variants have the weakest relative effect and add the least to prediction in people with highest levels of traditional risk factors and hence at highest absolute risk, the exact subgroup of the population in which preventive interventions and the cost-effectiveness of screening would be greatest.

Diabetes is a multifactorial disease and it has been proposed that genetics could help classify common disease subtypes. Apart from gestational diabetes and rare, specific Mendelian forms for which genetics already helps to guide diagnosis and treatment, diabetes is currently crudely classified into two broad types (>90% T2D) based on clinical presentation and rapid requirement of insulin. Thomas et al. have provided evidence for the presentation of T1D up to the sixth decade of life and that a polygenic score specifically associated with type 1 but not type 2 diabetes can help to rule out T1D in late onset cases. But due to the overwhelming predominance of T2D at older ages, the score’s positive predictive value is too low to confidently identify late onset T1D patients, in whom initial management may not be optimal if misdiagnosed as T2D. Interesting studies aiming to identify T2D subgroups using data-driven agnostic approaches have recently emerged. For instance, using six diabetes-related parameters (glutamate decarboxylase antibodies, age at diagnosis, BMI, HbA1c, and indices of insulin secretion and resistance), Ahlqvist et al. have provided evidence of five subtypes of T2D that differ in disease trajectories and risk of complications. Using electronic medical records, Li et al. previously reported evidence of three subtypes, suggesting that data-driven aetiologic classifications are influenced by context and data availability. Investigations using “-omics” measurements may further extend these initial attempts at a more refined disease categorisation.

While the existence of different aetiological subtypes in T2D is widely accepted, a robust and definitive classification is missing. In contrast with approaches aimed at classifying T2D in subgroups and categories, it has been proposed that a more nuanced approach to aetiological classification would better suit the highly polygenic and multifactorial background of this disease.

Genomic medicine and academia-industry partnerships

In the UK, transformative sequencing projects are currently underway. The 100,000 Genome Project delivers the benefits of genomic medicine to NHS cancer and rare disease patients now and has so far sequenced over 55,000 whole genomes (https://www.genomicsengland.co.uk/the-100000-
This project has shown the need for economies of scale in the delivery of genomic medicine, and NHS England is recommissioning and modernising NHS Genomic Laboratory Services to develop a first-class genomic service. UK Biobank (http://www.ukbiobank.ac.uk/) has transformed opportunities for population research in this country and internationally and is entering a new phase with the announcement to exome-sequence all 500,000 participants by 2019. Such an accelerated timeline was only possible through major funding by a consortium of five pharmaceutical companies, brought together by Regeneron Pharmaceuticals, following from their first initiative to sequence 50,000 participants in collaboration with GSK. In the USA, a partnership between Regeneron and the Geisinger Health System for the exome sequencing of over 50,000 people has already shown the value of integrating genetic and electronic health record data at scale. These are only two examples of commercial partners having access to participant and patient data, including information collected as part of routine clinical care in electronic health records. The financial benefits of such an effort are something to be considered in the light of the scientific opportunities that this investment and the generated sequence information will provide to biomedical researchers worldwide when they gain access, as will be the case for UK Biobank.

At the same time, large-scale Biobanks collecting genetic, physiological, longitudinal electronic health records and other health data have been established in many countries around the world, focusing on clinical hospital populations (e.g. BioVU: https://www.vumc.org/dbmi/biovu), insurance or care provider populations (e.g. Million Veteran Program: https://www.research.va.gov/mvp/), or national population cohorts (e.g. China Kadoorie Biobank, German National Cohort etc). The NIH funded 1 million persons “All of Us” Research Program was originally launched as a “new Precision Medicine Initiative to bring us closer to curing diseases like cancer and diabetes”. It specifically aims to actively recruit ethnic minority populations to help address the existing bias by including sufficient numbers of some of the many currently underrepresented groups.
Genomic sequencing puts the patient at the centre of drug discovery and validation. Close collaboration between academia, the pharmaceutical and other industries can catalyse the development of novel therapies for T2D based on genomic insight. To make full use of these opportunities for patients in this country and elsewhere, the research community needs to engage in a dialogue with patients and the public about genomic medicine and research and their implications for uses and misuses of genomic data.

Notwithstanding the potential that such partnerships have to improve health research and outcomes, research commissioned by the Wellcome Trust (https://wellcome.ac.uk/sites/default/files/public-attitudes-to-commercial-access-to-health-data-wellcome-mar16.pdf) and work delivered by Genomics England as part of their “Genomics Conversation” (https://www.genomicsengland.co.uk/a-year-of-conversations-about-genomics) have highlighted that patients’ and the public have concerns about data safety, i.e. non-legitimate uses, and commercial access to health data, including pharmaceutical and insurance companies.

The Chief Medical Officer’s Report “Generation Genome” considered the ethical, social and legal implications of genomic medicine in this country and highlighted the need for highest levels of data security for storage of identifiable data. However, it is not possible to give an absolute guarantee of data security and the potential harms arising from criminal data breaches need to be put in proportion with harms arising from restricting legitimate research uses of health data. For people living in the UK, universal free access to the National Health Service means that there is less reason to fear discrimination with regards health care insurance on the grounds of genetic testing. While there is currently no explicit legislation, the existing voluntary agreement with insurance providers also means that an estimated 95% of insurance customers would not need to disclose genetic test results for example for life assurance, critical illness cover, or income protection, as disclosure is tied to the policy value. The House of Commons Select Committee on Science and Technology recently recommended to extend the existing voluntary agreement, but closely monitor patient’s views and the experiences in other countries with a legal prohibition (https://publications.parliament.uk/pa/cm201719/cmselect/cmsctech/349/34908.htm#_idTextAnchor041).

Conclusions

Advances in genomic research have facilitated rapid progress in clarifying the genetic basis of T2D and characterising causal variants and variant-gene links. Future opportunities lie in larger-scale sequencing, discovery across diverse ancestries, studies in genetically isolated populations and in massive-scale biobanks. Successful development of academia-industry partnerships can deliver economies of scale, with implications for the future of genomics-focused research.
**Review Methods**

We searched Pubmed from inception to March 1\textsuperscript{st} 2018 using the following search strategy: (Diabetes Mellitus, Type 2[MeSH] OR NIDDM OR Maturity-Onset Diabetes OR Diabetes Mellitus, Noninsulin-Dependent OR Diabetes Mellitus, Adult-Onset OR Adult-Onset Diabetes Mellitus OR Diabetes Mellitus, Adult Onset OR Diabetes Mellitus, Ketosis-Resistant OR Diabetes Mellitus, Ketosis Resistant OR Ketosis-Resistant Diabetes Mellitus OR Diabetes Mellitus, Maturity-Onset OR Diabetes Mellitus, Maturity Onset OR Diabetes Mellitus, Non Insulin Dependent OR Diabetes Mellitus, Non-Insulin-Dependent OR Non-Insulin-Dependent Diabetes Mellitus OR Diabetes Mellitus, Noninsulin Dependent OR Diabetes Mellitus, Slow-Onset OR Diabetes Mellitus, Slow Onset OR Slow-Onset Diabetes Mellitus OR Diabetes Mellitus, Stable OR Stable Diabetes Mellitus OR Diabetes Mellitus, Type II OR Maturity-Onset Diabetes Mellitus OR Maturity Onset Diabetes Mellitus OR MODY OR Type 2 Diabetes Mellitus OR Noninsulin-Dependent Diabetes Mellitus OR T2D OR T2DM OR Type 2 Diabetes[tiab] OR Type 2 diabetes mellitus OR diabetes[ti]) AND (Genome-Wide Association Study[MeSH] OR Association Studies, Genome-Wide OR Association Study, Genome-Wide OR Genome-Wide Association Studies OR Studies, Genome-Wide Association OR Study, Genome-Wide Association OR Genome Wide Association Scan OR Genome Wide Association Studies OR GWA Study OR GWA Studies OR Studies, GWA OR Study, GWA OR Whole Genome Association Analysis OR Whole Genome Association Study OR Genome Wide Association Analysis OR Genome Wide Association Study). To look for new studies published in pre-publication (non peer-reviewed) form, we searched BioRxiv using the advanced search function (https://www.biorxiv.org/search): articles posted in the “genetics” or “genomics” collections, with the key word “diabetes” in the title or abstract, posted between the 1\textsuperscript{st} of January 2017 and the 9\textsuperscript{th} of April 2018. These literature searches were integrated with reference files of the authors and their colleagues, reference lists of original articles, reviews, and meta-analyses. Given exhaustive reviews on early genetic association studies by McCarthy\textsuperscript{1} and Morris\textsuperscript{2}, amongst others, we focused on recent developments and articles providing insights into clinical translation of genetic findings.
### Table 1. Contribution of genetic findings to T2D therapeutics in key areas.

<table>
<thead>
<tr>
<th>Area of contribution</th>
<th>Rationale</th>
<th>Considerations and examples in T2D research</th>
</tr>
</thead>
<tbody>
<tr>
<td>New drug target identification</td>
<td>In retrospective analyses, drugs with human genetics support are more likely to successfully transition through the drug development pipeline. Rapid development of new lipid-lowering drugs with genetic validation illustrates potential. Loss-of-function variants provide insights into likely efficacy and safety of inhibition, while gain-of-function on stimulation of target.</td>
<td>Genes encoding the targets of glucose lowering agents have been found by early GWAS, but new classes of diabetes drugs have not been developed as a result of human genetics findings. Most loci identified by GWAS have not lead to new drug development. Protective loss-of-function variants in SLC30A8 and IGF2 provide interesting examples that still await new drug development.</td>
</tr>
<tr>
<td>Mutation specific treatment</td>
<td>Pharmacological interventions may be particularly effective in patients with particular underlying aetiology or genetic predisposition.</td>
<td>Prominent examples are from Mendelian genetics, lack of examples for common forms of diabetes with polygenic aetiologic contribution. Elegant exemplar from systematic study of all possible missense variants of PPARG. Opportunities in specific areas have not been fully exploited.</td>
</tr>
<tr>
<td>Drug dosing or response</td>
<td>Drugs may require dose-adjustment according to genetic background.</td>
<td>Common variants at the ATM and SLC2A2 loci have been robustly associated with response to metformin, but genetic testing is not used in the clinic. Several studies including a recent trial have proposed an effect of TCF7L2 variants on response to glucose-lowering drugs.</td>
</tr>
<tr>
<td>Drug repurposing</td>
<td>As with new target identification, genetic variants that “mimic” existing therapeutic agents may provide the basis for repurposing.</td>
<td>No established example of repurposing from other therapeutic areas to diabetes. Recent GWAS have explored repurposing opportunities using bioinformatics approaches. Genetic findings around the LPL pathway may provide basis for extension of future indications and target population for emerging drugs targeting this pathway.</td>
</tr>
<tr>
<td>Drug safety</td>
<td>Genetic variants can inform on desired and undesired secondary effects of pharmacological modulation of the encoded drug target. In diabetes, it is critical to study the cardiovascular safety of existing and new agents.</td>
<td>The example of low-density lipoprotein cholesterol lowering genetic variants in genes encoding targets of cholesterol lowering therapy (i.e. HMGCR, NPC1L1 and PCSK9) and diabetes risk illustrate power and challenges of genetic approaches, given the partial consistency between genetic and clinical trial results. GLP1R and ABCG8 variants have been used to gain insights into cardiovascular safety of existing glucose-lowering drugs.</td>
</tr>
</tbody>
</table>

Abbreviations: GWAS, genome-wide association studies.
Figure Legends

Figure 1. Illustrative representation of genome-wide studies in type 2 diabetes and their power to detect certain types of susceptibility alleles for a given sample size. Susceptibility alleles above the solid black lines are detectable with a given approach. The graph is informed by the results of actual historical and current GWAS studies as well as whole-genome and exome sequencing studies that provided an empirical model for the genetic architecture of type 2 diabetes. Exemplar genetic susceptibility loci are reported in the figure. Abbreviations: GWAS, genome-wide association studies; OR, odds ratio.

Figure 2. Aetiologic model for the role of TBC1D4 in GLUT4 translocation and insulin-mediated glucose uptake in the skeletal muscle.

Figure 3. Models for normal and impaired insulin secretion. Genetic variants affecting these processes result in impaired insulin secretion and higher diabetes risk (right panel), e.g. variants at KCNJ11 and ABCC8 identified in genome-wide association studies.

Figure 4. Aetiologic model for the contribution of peripheral adipose storage capacity to metabolic and cardiovascular disease and role of adipogenesis and intravascular lipolysis in this process. Some of the images have been samples and modified from SMART, Servier Medical Art, (URL: https://smart.servier.com/) under Creative Commons Attribution 3.0 Unported License (https://creativecommons.org/licenses/by/3.0/).
References


126. de Carvalho LSF, Campos AM, Sposito AC. Proprotein Convertase Subtilisin/Kexin Type 9 (PCSK9) Inhibitors and Incident Type 2 Diabetes: A Systematic Review and Meta-analysis With Over 96,000 Patient-Years. Diabetes care 2018; 41(2): 364-7.


GENOMIC INSIGHTS INTO THE CAUSES OF TYPE 2 DIABETES

Claudia Langenberg and Luca A. Lotta

MRC Epidemiology Unit, University of Cambridge, United Kingdom.

Abstract

Genome-wide association studies of type 2 diabetes have implicated up to ~250 genomic regions in disease predisposition, at several of which there is now strong evidence for the underlying causal alleles variants and genes emerging for several of these regions. Understanding of the underlying mechanisms, including the interplay between beta-cell failure, insulin sensitivity, appetite regulation, and adipose storage has been facilitated by the integration of multi-dimensional data on diabetes-related intermediate phenotypes, detailed genomic annotations, functional work results and now multi-"omic" molecular features has furthered the understanding of mechanisms underpinning these associations, including the interplay between beta-cell failure, insulin sensitivity, appetite regulation and adipose storage. Studies in diverse ethnicities and examples from population isolates have highlighted the value and need for a broader genomic approach to this global disease. Ongoing trans-ethnic discovery efforts and emerging large-scale biobanks in diverse populations and ancestries may help to address some of the existing "Eurocentric" bias. While studies in diverse ethnicities and selected examples from population isolates have stressed the value and need for a broader genomic approach to this global disease, research in the field is still heavily "Eurocentric". Despite rapid progress in the discovery of the highly-polygenic architecture of type 2 diabetes, dominated by common alleles with small, cumulative effects on disease risk, current knowledge has shown no little clinical utility for disease prediction or prevention, and only small contributions to subtype classification or stratified approaches to treatment. Successful development of academia-industry partnerships for exome or genome sequencing in large biobanks can deliver economies of scale, with implications for the future of genomics-focused research. Development of academia-industry partnerships can help to deliver economies of scale and provide opportunities for genomics-informed drug development and validation.
1. A global view of type 2 diabetes genomics

The genetic basis of type 2 diabetes

In the last decade, hypothesis-free genome-wide association studies (GWAS) have been the single most important contributor to identifying genetic determinants of type 2 diabetes (T2D), leading to the discovery of ~100 associated genomic regions or loci. The last year has seen a game-change, a leap forward from smaller, cumulative advances to the description of now up to ~250 genome-wide significant loci including a large meta-analysis currently only available in pre-publication format. Several developments have enabled such rapid progress.

A tripling in effective sample size was achieved through the integration of large-scale, accessible resources such as the UK Biobank (http://www.ukbiobank.ac.uk), teamed with the openness of diabetes researchers worldwide to share results of previous studies (GWAS results: http://diagram-consortium.org; exome sequencing and genotyping results: http://www.type2diabetesgenetics.org). This has led to the inclusion of up to more than 74,000 T2D cases and 820,000 controls of European descent with an accompanying increase in power.

The breadth and depth of genetic variation ascertainment has also dramatically improved. Dense and accurate genotype imputation using the sequenced haplotypes of the 1000 Genomes and now, the Haplotype Reference Consortium has enabled the interrogation of over 20 over 10 million genetic variants, with a ~100-fold increase compared with early GWAS efforts. Direct genotyping or sequencing of common to rare alleles of the exomic areas of the genome has enabled a better ascertainment of coding variation. As a consequence, T2D susceptibility variants now range from 0.02-50% in minor allele frequency (MAF), and from 1.04 to 8.05 in per-allele odds ratio. While most signals are led by common variants with ever smaller effects, new risk alleles include several that are low-frequency or rare, including two examples of very rare variants with odds ratio ~8.

This is consistent with the model of heritability of T2D derived from sequencing experiments, whole-genome sequencing experiments, characterised by a prominent contribution to heritability of common variation, a small contribution of rare variation and evidence of low selective pressure on predisposition alleles. Future meta-analyses of multiple large biobanks is likely to expand the catalogue of susceptibility variants to some ultra-rare alleles with extreme effect sizes (odds ratios >10), in addition to finding even more common regulatory alleles with very small effects (odds ratios <1.01; Figure 1). Direct genotyping and sequencing will be critical to replicate and identify associations for imputed risk variants that are in the rare allele frequency spectrum (MAF below 0.5%), specifically if discovered based on imputed genotypes, given the difficulties of achieving quality of in accurately imputing these variants even with expanded population reference panels.

Clinical translation of these genomic associations critically depends on our understanding of underlying mechanisms. Establishing causal variants and variant-gene links has been a challenge in genetic studies. Denser imputation, extended genotyping in coding or metabolic- trait associated regions, direct sequencing, larger sample size, integration of extensive genomic and regulatory annotations, and progress in analytical fine-mapping approaches have all made this fundamental task easier. For a given association signal, fine-mapping has been able to considerably narrow the size of the genomic region that likely contains the causal variant as well as the list of plausible causal variants in that genomic region. Of 380 signals amenable to fine mapping in the latest analysis, 18 could be attributed to a single causal variant with a posterior probability of association >99%, with over 50 signals at 41 loci exceeding 80% for the strongest variant. While most of these likely causal alleles map to regulatory regions in pancreatic islets, several overlap enhancers or promoters in other diabetes-related tissues as adipose, skeletal muscle and liver. Also, there are now a number of likely causal missense or nonsense variants with evidence of causal association, are now amenable to more direct in vitro
and in vivo experimental follow-up. These represent critical advances in the translation of robust associations into biological understanding.
Trans-ethnic discovery efforts

- The global health issue like burden of T2D with large differences in risk within and between populations warrants a global genomic approach to study its predisposition. Despite the progress in the discovery of its genetic basis, T2D is no exception when it comes to the continued underrepresentation of ethnic diversity in genetic research and discovery efforts. Genetic studies in multiple ethnicities are valuable for several reasons: (a) susceptibility variants may be present at appreciable allele frequency only in non-European populations, as elegantly shown by the discovery of associations at the HNF1A (via exome sequencing) and IGF2 (via exome genotyping) loci in populations of Latin American origin; (b) for shared susceptibility loci, trans-ethnic studies increase statistical power for new discoveries; (c) the diverse linkage disequilibrium patterns across ancestries increase the resolution of fine-mapping analyses for the identification of causal variants; (d) exposure to diverse environments may reveal the effect of susceptibility variants which are masked in other settings (i.e. gene-environment interactions).

Important evidence has been emerging from GWAS of South Asian, East Asian, Latin American and African American populations and trans-ethnic discovery efforts; however, studies for non-European ancestries remain small in comparison. Collaborative Ongoing trans-ethnic discovery efforts are now underway under the umbrella of the DIAMANTE consortium with over 170,000 T2D cases across 120 studies and with around 45% of the effective sample size accounted for by non-European ancestries, currently including African (7%), East Asian (23%), Hispanic or Latino (6%) and South Asian (9%) ancestry participants (personal communication Prof Andrew Morris, University of Liverpool). While this represents a substantial advance compared to previous studies, research in the field remains heavily “Eurocentric”. This will represent a substantial advance in employing genetic diversity across populations for T2D genetic discovery, and will improve our ability to generalise findings with important implications for drug discovery and clinical translation for non-European descent populations, many of which are at high risk of T2D.

Genetically isolated populations

Recent studies have clearly demonstrated the value of studying genetically isolated populations. In these settings, extended linkage disequilibrium and higher allelic frequencies of deleterious variants due to greater allelic drift allowed unique insights into the etiology of T2D, deleterious variants with large phenotypic effect may raise by chance to higher allele frequencies, due to a phenomenon known as allelic drift. This makes it easier to identify disease associations for such variants in isolated compared to admixed populations, in which these variants may not be present or may be very rare. Such findings can provide insight into the etiology of T2D that is generalizable outside of the context of the particular population in which they were discovered.

In the Inuit population from Greenland, homozygote carriers of a loss-of-function variant in TBC1D4 (p.Arg684*) were found to have a ~10 fold higher risk of diabetes, and ~1 standard deviation higher glucose and insulin at 2 hours following an oral glucose challenge. The variant segregates at high frequency in Inuits (MAF 17%), while being rare or monomorphic in other populations, and influences risk in a fashion consistent with recessive inheritance. The mechanism of action of this variant is also particularly insightful. Risk allele carriers had lower levels of GLUT4 in skeletal muscle, which is critical in the insulin-mediated control of post-prandial glycaemia (Figure 2). This finding highlights the causal role of insufficient GLUT4-mediated glucose uptake in muscle for postprandial hyperglycaemia and T2D risk (Figure 2). This corroborates evidence from the first report of a mutation significantly impairing GLUT4 translocation, identified in a child with acanthosis nigricans and extreme postprandial hyperinsulinemia carrying a heterozygous premature stop mutation (p.Arg363*) in TBC1D4.
In the same population, exome sequencing has revealed the association of More recently, an analysis of rare loss-of-function variants from exome sequencing in nine Greenlandic family trios identified a loss-of-function variant in ADCY3 with obesity and diabetes, a finding supported that decreases gene expression via a splicing defect. The variant was associated with greater BMI, abdominal fat, insulin resistance, dyslipidemia and risk of T2D, with evidence for a recessive inheritance. While this variant was not present in non-Greenlandic populations, the authors were able to identify heterozygous carriers of seven other predicted loss-of-function variants in ADCY3 in publicly available trans-ethnic T2D exome sequencing studies (URL: http://www.type2diabetesgenetics.org), as well as studies of consanguineous in families with severe obesity, and demonstrate their enrichment in T2D cases versus controls. A parallel study identified the gene by sequencing patients with severe obesity. This demonstrates the success of prioritizing genomic regions for systematic follow up in other populations based on strong evidence from population isolates or extreme phenotypes in family studies.

Other examples also highlight convergence of evidence from studies of monogenic disease and population isolates. An autosomal dominant heterozygous missense mutation in AKT2 was originally identified as the cause of fasting and postprandial hyperinsulinemia and diabetes in a family with severe insulin resistance and partial lipodystrophy. A recent exome-array and sequencing study detected a low-frequency AKT2 coding variant (p.Pro50Thr, MAF 1.1% in Finnish participants, but not detected in other populations) to be associated with higher fasting insulin levels, lower uptake of glucose in insulin-sensitive tissues (shown in a separate recall-genotype study), and higher T2D risk, further highlighting the role of AKT2 in insulin sensitivity.

Another interesting example comes from a study of Samoans, a founder population with a high prevalence of obesity and T2D, in whom a common population-specific (26% allele frequency in Samoans but extremely rare in other populations) “thrifty” missense variant in CREBF (rs373863828, p.Arg457Gln) showed an association with substantially higher body mass index (1.4 kg/m² per allele). Interestingly, the adiposity-raising allele was associated with lower fasting glucose and protection from T2D. In an adipocyte model, overexpression of Arg457Gln selectively decreased energy use and increased fat storage, in line with its effect on overall body fat, greater abdominal (waist) and glutero/femoral (hip) fat, potentially highlighting the metabolic benefits of a greater capacity of fat storage.
2. Insights into pathways to diabetes through genomic discovery

Genetic studies of diabetes-related intermediate phenotypes—continuous metabolic traits in nondiabetic individuals have emerged as a way to gain mechanistic insights into T2D susceptibility that is complementary to disease-focused discovery GWAS.

Glycaemic regulation and susceptibility to diabetes

In the absence of intermediate metabolic traits associations, the interpretation and mechanistic follow-up of diabetes susceptibility loci would be in some cases impossible. Genetic studies of glucose and insulin related glycaemic and insulinemic measures have been defined by the struggle between sample size and degree of refinement of phenotype ascertainment. Large efforts have focused on widely-available, and simple measures as fasting glucose or insulin, and glycated haemoglobin or insulin, while smaller ones in contrast, efforts based on “gold-standard” assays measures involving frequently-sampled oral glucose tolerance tests or continuous intravenous measurements that are difficult to obtain at scale. The former approach has been very successful in the discovery of loci influencing glycaemic traits and insulinemic levels in non-diabetic individuals, and helped to identify insulin secretory and has provided large-scale datasets for the initial characterisation effects as ation of diabetes-associated variants a major driver of associations for several of the common diabetes susceptibility loci (Figure 3). The latter has been used primarily for gold-standard based studies have provided in-depth physiologic characterisation of diabetes susceptibility variants (REFS) and has also shown promise for a discovery approach for the identification of novel by identifying loci (e.g. GRB10, BCL2, FAM19A2, NAT2) that have eluded discovery in much larger meta-analyses of more widely-available, but simpler measures or indices (REFS). Both approaches have been instrumental in understanding the underlying mechanisms of common T2D predisposition and the genetic influences on circulating glucose levels, insulin secretion and resistance.

Studies of intermediate traits highlighted the dominant role of impaired insulin secretion as a major driver of associations for several of the common diabetes susceptibility loci (Figure 3). Overlapping diabetes susceptibility variants, glycaemic traits and pancreatic islet regulatory and functional data has provided the foundation for an improved understanding of mechanisms linking beta-cell glucose sensing and insulin secretion with T2D risk. High-throughput functional screens of gene silencing in human beta-cell lines are now available that can empower systematic characterisation of the functional impact of novel likely-causal genes on insulin secretion. Studies of “gold-standard” measures have highlighted novel and specific genetic associations.

In the absence of intermediate metabolic traits associations, the interpretation and mechanistic follow-up of diabetes susceptibility loci would be in some cases impossible. A future challenge in this field will be for intermediate trait studies to stay apace with the rapid inflation-increases in sample size of diabetes association analyses. This is a task that is complicated by the practical difficulties of obtaining fasting samples, let alone more invasive “gold standard” intravenous measurements, in large biobanks.

Excess overall fat and “central” role of peripheral fat

Excess fat is the hallmark of overeating and lack of physical exercise and has been a major focus of genetic research. Large-scale studies of body mass index and related measures have linked genes highly expressed in the central nervous system with general obesity in different ancestries.
This has provided complementary evidence to original discoveries that have revealed the fundamental role of appetite regulation in monogenic obesity, in particular the leptin-melanocortin axis.

Observational epidemiology has clearly shown that, for a given level of overall adiposity, the distribution of fat in the body is associated with susceptibility to insulin resistance, diabetes and its complications. \(^{64-65}\) Genetic discovery approaches integrating multiple insulin-resistance related phenotypes \(^{52-53}\) has been instrumental in the study of Mendelian lipodystrophies. \(^{64-65}\) However, the complex relationship between regional body fat deposition and metabolic risk go beyond the simple notion of “apple” and “pear” shaped body. This concept is reinforced by the aetiologic and phenotypic heterogeneity of partial and generalised lipodystrophies. \(^{64-65}\) Using integrative genomic approaches, we and others have identified 53 genomic regions associated with integrative genomic approaches to the identification of insulin resistance genetic determinants. \(^{64-65}\) These insulin resistance loci were enriched with genes harbouring mutations in Mendelian forms of lipodystrophy and were associated with increased odds of severe lipodystrophic insulin resistance. \(^{64-65}\) In the context of other evidence about the protective role of fat deposition in peripheral compartments, \(^{65-67}\) these results suggest that reduced mean that being unable to safely store excess energy in the peripheral regions of the body may lead to ectopic fat storage and, higher circulating lipids and cardio-metabolic risk in the general population similar to clinical manifestations of severe forms of lipodystrophy (Figure 34).

Molecular mechanisms underlying this specific aetiology have only started to emerge. These include the impaired ability to generate new adipocytes and the regulation of gene expression in these cells, as suggested which is supported by initial experimental evidence around PPARG, \(^{73}\) KLF14, \(^{70}\) IRS1, CCDC92, DNAH10 and L3MBTL3 \(^{52-53}\) as well as the impaired intravascular lipoprotein lipase (LPL)-mediated lipolysis, \(^{64-65}\) the mechanism that regulates lipid buffering from the circulation to peripheral tissues, is also implicated. \(^{54-55}\) (Figure 21). Further studies are necessary to understand whether and how these mechanisms play into the well-established relationship between centripetal fat distribution and metabolic risk and are necessary to understand how fat deposition in specific body compartments influences metabolic disease risk of disease in the general population and move beyond the very simple notion of “apple” and “pear” body shaped body.

Prioritising causal pathways by integrating multi-omic data with clinical outcomes

As the sample size of genetic studies has rapidly grown, so has our ability to measure detailed molecular features in biological samples using high-throughput technology. Global patterns of methylation and other epigenetic features (epigenomics), \(^{75,76}\) gene expression (transcriptomics), \(^{75,76}\) proteins (proteomics) \(^{75,76}\) or metabolites (metabolomics) \(^{75,76}\) can now be measured at epidemiological scale and show enough heritability to enable a genetic mapping in genome-wide studies.

Genetic studies of “omics” molecular features can help advance the understanding of the causes of diabetes (and other complex) diseases in multiple ways, by (a) characterising the phenotypic consequences of diabetes susceptibility variants identified by GWAS; (b) helping to identify causal variants and genes at known susceptibility loci; (c) enabling the estimation of causal associations between molecular traits and disease risk using the principles of Mendelian randomisation.

So far, no studies have systematically followed-up associations of T2D susceptibility variants with circulating metabolomic or proteomic profiles. Studies overlaying regulatory annotations and gene expression in pancreatic beta-cells with T2D GWAS results show the value of omics\textsuperscript{13} transcriptomic technology analyses for the identification of diabetes susceptibility genes and mechanistic understanding. \(^{23-25}\) The increasing availability of similar data on a variety of cell types...
(including other relevant metabolic tissues, such as skeletal muscle, adipose, liver) will make it possible to systematically assess the relevance of different tissues, cell types and tissue-specific mechanisms in T2D pathophysiology.

“Mendelian Randomization” studies using genetics to assess causal associations between molecular traits and T2D risk have traditionally focused on specific pathways and biomarkers rather than “-omics” profiles. This is partly the reflection of challenges in applying causal inference frameworks in the context of correlated and co-regulated molecular exposures such as blood metabolites or proteins. 42 For example, studies of the branched chain amino acid pathway have provided human genetic evidence of multi-directional causal relationships between their metabolism, insulin resistance and risk of diabetes, 43 building upon observations dating back to the 1960s. 84-86 This and other metabolomics genetic studies illustrate how variation at key metabolic regulators (e.g. enzymes) affects a large set of biologically and phenotypically related measures within a pathway. Hence, inference from associations with diabetes may be restricted to a specific locus of regulatory importance rather than levels of one or more metabolites. Studies of the epigenetic patterns associated with obesity and secondarily with T2D have illustrated how often, that these complex molecular patterns are mostly the consequences rather than the cause of disease processes. 87

Several new methods that integrating genomic etic and other “-omics” data have been developed to facilitate these more sophisticated studies in this field. 88-91 In principle, these approaches can be used to systematically study the causal influences of molecular traits on diabetes risk. In practice, they present a number of challenges that complicate aetiologic inference. 79-83 While individual multi-omic studies have already been conducted in sample sizes of several thousands, these deeply phenotyped epidemiological studies are still comparatively small in the context of genomic research and the relevance of identified loci to metabolic diseases remains largely unexplored. However, they can serve as models for what might be achieved when technology and cost developments enable high-throughput multi-omic phenotyping at the scale of large national Biobanks with hundreds of thousands of participants systematically followed up for a broad range of diseases.

Building upon the first wave of discovery, large scale meta-analyses of GWAS of metabolite profiles, proteins and epigenetic markers are underway which will enable the systematic study of causal relationships between genetic differences in circulating metabolite/molecular patterns and risk of diabetes and other health outcomes. These investigations may also explain the association of disease at dozens of T2D associated genetic loci where the underlying molecular mechanisms remain elusive. Ultimately, the hope is that these multi-layered genetic investigations will reveal novel associations and make sense of known ones, but the distance between expectations and achievement in this field remains wide.
3. Genomics and therapeutics in T2D

The Chief Medical Officer’s Report “Generation Genome” published in 2017 highlighted key areas in which genomics can inform disease therapeutics, providing a helpful framework to describe recent developments of genomics-informed therapeutics in T2D (Table 1).

New targets

The identification of novel drug targets is a major stated objective of T2D genetic studies. The finding that drug targets supported by human genetics evidence are more likely to succeed in the selective drug development process, and recent success in the development of lipid-lowering new lipid-lowering drugs follows genetic findings from different approaches and sources at the PCSK9, LPA, APOC3 and ANGPTL3 loci have exponentially increased interest in this field approach.

Three diabetes-susceptibility candidate genes identified in early GWAS studies encode drug targets for existing glucose lowering therapy (PPARG, thiazolidinediones; KCNJ11/ABCC8, sulfonylureas). Mutations in these genes cause severe monogenic diseases characterized by loss of glycemic control and common variants at these loci were identified in the first wave of diabetes GWAS. This suggests that other genes identified through hypothesis-free approaches may provide useful new drug targets, even if common lead variants have small effects.

However, human genetics has played a peripheral role in the development of the most recently approved classes of glucose-lowering drugs, including dipeptidyl peptidase-4 inhibitors, GLP1R agonists, and sodium-glucose co-transporter-2 inhibitors.

With ~250 over a hundred genetic loci now robustly associated with diabetes, why is there still not a clear “PCSK9-like” example? Generic and therapeutic-area specific obstacles may play a role. First, several of the causal genes implicated by GWAS have only recently emerged, but it takes several years for new drugs to enter clinical development. Specifically targeting pancreatic islets, adipocytes, skeletal myocytes or brain cells, as opposed to circulating proteins is challenging. Safety concerns have been a limitation in the development of appetite suppressants modifying genetically-validated targets, as illustrated by the failure of first generation agonists of the melanocortin 4 receptor due to on-target side effects. Also, many loci for T2D act via impaired insulin secretion. The existence of different classes of approved and widely-used insulin secretagogues may limit the interest of pharmaceutical companies in new drug development in this area, given the focus on developing commercially-differentiated products.

Protective loss-of-function variants are particularly interesting for drug development purposes because they provide insights into the likely consequences of inhibiting a gene product and, if carriers are healthy, provide substantial implied evidence of the likely safety profile of pharmacological inhibition of the existence of a therapeutic window. Sequencing of the early T2D GWAS gene SLC30A8, encoding a pancreatic islet zinc transporter (ZnT8), has identified rare non-conservative amino acid changes in populations of Latino descent identified a protective variant (odds ratio for carriers, 0.34). More recently, an exome array genotyping study in populations of Latino descent identified a protective variant (odds ratio per allele, 0.80) in IGF2 associated with incorrect splicing of isoform 2 of the gene, suggesting that selectively inhibiting this isoform in relevant tissues may be therapeutically exploited. However, efforts to inhibit ZnT8 or insulin-like growth factor 2 have yet to reach clinical development. The potential therapeutic implications of newly discovered recently reported protective associations of SLC30A8 with obesity, diabetes and coronary artery disease (and of a loss-of-function variant in GPR151 against consideration.

Genetically-tailored treatment
Diabetes medicine holds one of the most elegant examples of tailoring treatment to the specific underlying genetic or molecular defect, but all of these relate to monogenic forms of the disease.\textsuperscript{103-106} No such examples exist for common susceptibility loci, but this area is certainly understudied.

Response to glucose-lowering treatment shows a measurable degree of heritability and evidence of polygenicity,\textsuperscript{108} indirectly suggesting that combinations of multiple alleles might be able to identify patients who would be more or less responsive to certain drugs. Opportunities in this field arise from the definition and characterization of polygenic scores combining common variants that capture a particular aetiology, Variants with large effects, such as those (b) variants with large effect size from population isolates (c) or rare variants with very large effects from GWAS in admixed populations could also provide suitable basis for pharmacogenetics applications. By studying the functional consequences of all theoretically possible missense variants in PPARG, Majithia and colleagues elegantly showed that diabetes-associated mutations in the gene display heterogeneous in vitro response to thiazolidinediones,\textsuperscript{109,110} which could provide the basis for tailored therapy or dosing in carriers of these specific alleles, as illustrated in initial case reports.\textsuperscript{110}

Drug dosing or response

The several efforts to identify interactions between treatment and genetic background and T2D\textsuperscript{111} treatment T2D have been the subject of a recent systematic review,\textsuperscript{112} which revealed the reliance on clinical studies and candidate gene rather than hypothesis-free approaches, with a few notable exceptions discussed below. In a pharmacogenetic clinical trial, Srinivasan et al. found that TCF7L2 variants associated with T2D influences the acute response to both glipizide and metformin in people with risk factors for T2D or treatment-naive T2D patients.\textsuperscript{113} Two genome-wide association studies have identified common genetic variants at the ATM\textsuperscript{114} and CPAP\textsuperscript{115} and SLC2A2\textsuperscript{116} loci associated with response to metformin. The difference in the effect of metformin effect size of for these variants on HbA1c was estimated at around \textasciitilde 0.15-0.17\% of HbA1c per allele, roughly corresponding to a daily dose of \textasciitilde 250 mg of metformin.\textsuperscript{117,118} A new genome-wide study\textsuperscript{119} discovered a new association with T2D treatment response, with T2D patients' response to metformin being dependent on the ATM locus. This follows directionally consistent findings for triglyceride levels and heart disease,\textsuperscript{120} which could provide the basis for tailored therapy or dosing in carriers of these specific alleles, as illustrated in initial case reports.\textsuperscript{110}

Gene testing in this setting -- entering the clinic

If “pharmacomimetic” genetic variants can be used to find new targets, they could theoretically be used to find new indications for existing drugs. While there, but there are not yet any established currently no clear examples of genetically-directed repurposing of approved drugs in diabetes.\textsuperscript{121} Imamura et al. used a systematic bioinformatics approach to identify new T2D drug targets, revealing potential repurposing opportunities for drugs targeting the gene products of GSK3B and JUN.\textsuperscript{122} Recent findings around lipoprotein lipase may offer an example of genetically-driven extension of the target population for drugs that are in active development. In late 2016, we reported a gain-of-function variant (rs328, p.Ser447*) in LPL associated with insulin sensitivity and protection from diabetes and an independent loss-of-function variant (rs1801177, p.Asp36Asn) associated with higher diabetes risk.\textsuperscript{123} This followed directionally consistent findings for triglyceride levels and heart disease,\textsuperscript{124} leading to the hypothesis that the several agents targeting the LPL pathway that are in development for the treatment of hypertriglyceridemia,\textsuperscript{125-127} could also be valuable as insulin sensitizing agents. The association with diabetes of rs328 has since been replicated.\textsuperscript{128}
the variant or its proxies have emerged in recent GWAS of diabetes, published for a loss-of-function in the natural LPL-inhibitor ANGPTL4 (rs116843064, p.Glu40Lys), lending powerful support to this hypothesis.

Drug safety

Genetic variants have also been used to understand both desired and undesired secondary effects of pharmacological modulation. For instance, statins are associated with cardiovascular protection, but with a modest increase in the risk of new-onset diabetes, a pattern observed also for cholesterol lowering variants in or near their target HMGCR, including NPCILI (encoding the target of ezetimibe) and PCSK9 (PCSK9 inhibitors). Also genetic studies of non-statin cholesterol lowering targets—cholesterol lowering alleles at ANGPTL4 (rs116843064, p.Glu40Lys)—have shown associations with a lower risk of coronary heart disease, but higher diabetes risk for cholesterol lowering alleles. While this has suggested modest diabetogenic effects for that also non-statin cholesterol lowering agents could be associated with higher diabetes risk, might have modest diabetogenic effects. It is important to note that recent randomized controlled trials of ezetimibe and PCSK9-inhibitors did not find associations with statistically significant higher diabetes risk, ruling out large or statistically significant diabetogenic effects. A meta-analysis of trials of PCSK9 inhibitors has revealed statistically significant albeit very modest increases in glycemic markers, but no significant association with diabetes risk.

Cardiovascular associations of genetic variants that mimic diabetes medications are particularly insightful, given the regulatory requirement that glucose-lowering drugs should not be associated with a higher risk of cardiovascular disease and the ongoing paradigm shift from glycemic control to prevention of complications in diabetes management. The concomitant publication of randomized controlled trials of a glucagon-like peptide receptor 1 (GLP1R) agonists and of genetic studies of a putative gain-of-function variant of GLP1R (rs10305492, p.Ala316Thr), both showing cardio-protective associations for GLP1R activation, illustrates this concept. A similar approach has shown cardio-protective associations for a functional variant in ABCG8 (rs757110, p.Ala1369Ser), providing genetic insights into the cardiovascular effects of sulfonylureas for which clinical trial evidence is inconclusive.
4. Clinical relevance and future outlook

Genomics in the clinic

In addition to diabetes therapeutics, genomics has been proposed to help disease prediction and diagnosis of common subtypes. The polygenic architecture of T2D is dominated by many common variants with small effects and tagging several different etiologies has critical implications for both applications.

The level of prediction achieved by common alleles contrasts with patients’ often more deterministic understanding of what “genetic risk” constitutes, which is informed by examples of highly penetrant causes of monogenic diseases. Accurate prediction in combination with preventive lifestyle interventions can be used for targeted primary prevention and avoidance of complications arising from metabolic dysregulation that is present for years before diagnosis. However, existing diabetes prediction models perform well in the general population and can be used to non-invasively identify individuals at high risk. Earlier prospective studies using up to 65 variants have shown that polygenic T2D risk scores improve prediction performance only marginally when considered over and above risk factors that can assessed non-invasively, such as age, sex, body mass index or family history. New approaches based on machine learning and complex statistical modeling such as those proposed by Shigemizu et al. could perhaps further improve on these results. In the context of a clinically established “minimally” invasive test, critically, the existence of good inexpensive predictors from a patient’s anamnesis or examination (e.g. family history, BMI, together with clinically established inexpensive tests (i.e. blood glucose, HbA1c) that inform both diagnosis and future risk prediction and diagnostic prediction, i.e. blood glucose levels or HbA1c, greatly limits the scope for introduction of genetic information forin T2D prediction in the clinic currently has no meaningful clinical contribution to make. Evidence exists that established genetic variants have the weakest relative effect and add the least to prediction in people with highest levels of traditional risk factors and hence at highest absolute risk, the exact subgroup of the population in which preventive interventions and the cost-effectiveness of screening would be greatest, i.e. blood glucose levels or HbA1c, genetically predicted risk currently has no meaningful clinical contribution to make. Evidence exists that established genetic variants have the weakest relative effect and add the least to prediction in people with highest levels of traditional risk factors and hence at highest absolute risk, the exact subgroup of the population in which preventive interventions and the cost-effectiveness of screening would be greatest.

Diabetes is a multifactorial disease and it has been proposed that genetics could help classify common disease subtypes. Apart from gestational diabetes and rare, specific Mendelian forms for which genetics already helps to guide diagnosis and treatment, diabetes is currently crudely classified into two broad types (>90% T2D) based on clinical presentation and rapid requirement of insulin. Thomas et al. have provided evidence for the presentation of T1D up to the sixth decade of life and that a polygenic score specifically associated with type 1 but not type 2 diabetes can help to rule out T1D in late onset cases But due to the overwhelming predominance of T2D at older ages, the score’s positive predictive value is too low to confidently identify late onset T1D patients, in whom initial management may not be optimal if misdiagnosed as T2D.

Recent interest in efforts to classify diabetes on the basis of electronic health records, physiological measures or genetic data using data-driven agnostic approaches are interesting, but their clinical relevance is yet to be demonstrated. For instance, using six diabetes-related variables (glutamate decarboxylase antibodies, age at diagnosis, BMI, HbA1c, and homeostatic model assessment 2 estimates of insulin secretion and β-cell function and insulin resistance), Ahlgqvist et al. have provided evidence of five subtypes of T2D that differ in disease,
participants by 2019. Such an investment and the generated sequence information will provide to biomedical researchers worldwide when they gain access to participant and patient data, including information collected as part of routine clinical care in electronic health records. The financial benefits of such an effort are something to be considered in the light of the scientific opportunities that this investment and the generated sequence information will provide to biomedical researchers worldwide when they gain access, as will be the case for UK Biobank researchers.

At the same time, large-scale Biobanks collecting genetic, physiological, longitudinal electronic health records and other health data have been established in many countries around the world, focusing on clinical hospital populations (e.g. BioVU: https://www.vumc.org/dbmi/biovu), insurance or care provider populations (e.g. Million Veteran Program: https://www.research.va.gov/mvp/), or national population cohorts (e.g. China Kadoorie Biobank, German National Cohort etc). The NIH funded 1 million persons “All of Us” Research Program was originally launched as a new Precision Medicine Initiative to bring us closer to curing diseases like cancer and diabetes144. It specifically aims to actively recruit ethnic minority populations to help address the existing bias by including. Whether or not it will be successful in accruing sufficient numbers required for genetic research of some of the many currently underrepresented groups remains to be established.

Genomic medicine and academia-industry partnerships

In the UK, transformative sequencing projects are currently underway. The 100,000 Genomes Project delivers the benefits of genomic medicine to NHS cancer and rare disease patients now and has so far sequenced over 545,000 whole genomes (https://www.genomicsengland.co.uk/the-100000-genomes-project/). This project has shown the need for economies of scale in the delivery of genomic medicine, and NHS England is recommissioning and modernising NHS Genomic Laboratory Services to develop a first-class genomic service. UK Biobank (http://www.ukbiobank.ac.uk/) has transformed opportunities for population research in this country and internationally and is entering a new phase with the announcement to exome-sequence all 500,000 participants by 2019. Such an accelerated timeline was only possible through major funding by a consortium of five pharmaceutical companies, brought together by Regeneron Pharmaceuticals, following from their first initiative to sequence 50,000 participants in collaboration with GSK. In the USA, a partnership between Regeneron and the Geisinger Health System for the exome sequencing of over 50,000 people has already shown the value of integrating genetic data and electronic health record data at a large scale145. This is just one example of commercial partners having access to participant and patient data, including information collected as part of routine clinical care in electronic health records. The financial benefits of such an effort are something to be considered in the light of the scientific opportunities that this investment and the generated sequence information will provide to biomedical researchers worldwide when they gain access, as will be the case for UK Biobank researchers.

trrajectories and risk of complications 141. Using electronic medical records, Li et al. had previously reported evidence of three subtypes, suggesting that data-driven aetiological classifications are influenced by context and data availability. Investigations using “-omes” measurements may further extend these initial attempts at a more refined disease categorisation. While the existence of different aetiological subtypes in T2D is widely accepted, a robust and definitive classification is missing. In contrast with approaches aimed at classifying T2D in subgroups and categories, it has been proposed that a more nuanced approach to aetiological classification would better suit the highly polygenic and multifactorial background of this disease.

Required at diagnosis

Required at diagnosis

In contrast with approaches aimed at classifying T2D in subgroups and categories, it has been proposed that a more nuanced approach to aetiological classification would better suit the highly polygenic and multifactorial background of this disease.
Engaging patients in research

Genomic sequencing puts the patient at the centre of drug discovery and validation. Close collaboration between academia, the pharmaceutical and other industries can catalyse the development of novel therapies for T2D based on genomic insight. To make full use of these opportunities for patients in this country and elsewhere, the research community needs to engage in a dialogue with patients and the public about genomic medicine and research and their implications for uses and misuses of genomic data.

Notwithstanding the potential that such partnerships have to improve health research and outcomes, research commissioned by the Wellcome Trust (https://wellcome.ac.uk/sites/default/files/public-attitudes-to-commercial-access-to-health-data-wellcome-mar16.pdf) and work delivered by Genomics England as part of their “Genomics Conversation” (https://www.genomicsengland.co.uk/a-year-of-conversations-about-genomics) have highlighted that patients’ and the public have concerns about data safety, i.e. non-legitimate uses, and commercial access to health data, including pharmaceutical and insurance companies.

The Chief Medical Officer’s Report “Generation Genome” considered the ethical, social and legal implications of genomic medicine in this country92 and highlighted the need for highest levels of data security for storage of identifiable data. However, it is not possible to give an absolute guarantee of data security and the potential harms arising from criminal data breaches need to be put in proportion with harms arising from restricting legitimate research uses of health data. For people living in the UK, universal free access to the National Health System means that there is less reason to fear discrimination with regards health care insurance on the grounds of genetic testing. While there is currently no explicit legislation, the existing voluntary agreement with insurance providers also means that an estimated 95% of insurance customers would not need to disclose genetic test results for example for life assurance, critical illness cover, or income protection, as disclosure is tied to the policy value. The House of Commons Select Committee on Science and Technology recently recommended to extend that the existing voluntary agreement, but is extended closely monitor but that patient’s views and the experiences in other countries with a legal prohibition are closely monitored (https://publications.parliament.uk/pa/cm201719/cmselect/cmsctech/349/34908.htm#_idTextAnchor041).

Conclusions

Advances in genomic research have facilitated rapid progress in clarifying the genetic basis of T2D and characterising causal variants and variant-gene links. Future opportunities lie in larger-scale sequencing, discovery across diverse ancestries, studies in genetically isolated populations and in massive-scale biobanks. Successful development of academia-industry partnerships can deliver economies of scale, with implications for the future of genomics-focused research, informed drug development and population research in this field.
Review Methods

We searched Pubmed from inception to March 1st 2018 using the following search strategy: (Diabetes Mellitus, Type 2[MeSH] OR NIDDM OR Maturity-Onset Diabetes Mellitus OR Diabetes Mellitus, Noninsulin-Dependent OR Diabetes Mellitus, Adult-Onset OR Diabetes Mellitus OR Diabetes Mellitus, Ketosis-Resistant OR Diabetes Mellitus, Ketosis Resistant OR Ketosis-Resistant Diabetes Mellitus OR Diabetes Mellitus, Maturity-Onset OR Diabetes Mellitus, Maturity Onset OR Diabetes Mellitus, Non Insulin Dependent OR Diabetes Mellitus, Non-Insulin-Dependent OR Non-Insulin-Dependent Diabetes Mellitus OR Diabetes Mellitus, Noninsulin Dependent OR Diabetes Mellitus, Slow-Onset OR Diabetes Mellitus, Slow Onset OR Slow-Onset Diabetes Mellitus OR Diabetes Mellitus, Stable OR Stable Diabetes Mellitus OR Diabetes Mellitus, Type II OR Maturity-Onset Diabetes Mellitus OR Maturity Onset Diabetes Mellitus OR MODY OR Type 2 Diabetes Mellitus OR Noninsulin-Dependent Diabetes Mellitus OR T2DM OR Type 2 Diabetes[tiab] OR Type 2 diabetes mellitus OR diabetes[ti]) AND (Genome-Wide Association Study[MeSH] OR Association Studies, Genome-Wide OR Association Study, Genome-Wide OR Genome-Wide Association Studies OR Studies, Genome-Wide Association OR Study, Genome-Wide Association OR Genome Wide Association Scan OR Genome Wide Association Studies OR GWA Study OR GWA Studies OR Studies, GWA OR Study, GWA OR Whole Genome Association Analysis OR Whole Genome Association Study OR Genome Wide Association Analysis OR Genome Wide Association Study). To look for new studies published in pre-publication (non peer-reviewed) form, we searched BioRxiv using the advanced search function (https://www.biorxiv.org/search): articles posted in the “genetics” or “genomics” collections, with the keyword “diabetes” in the title or abstract, posted between the 1st of January 2017 and the 9th of April 2018. These literature searches were integrated with reference files of the authors and their colleagues, reference lists of original articles, reviews, and meta-analyses. Given exhaustive reviews on early genetic association studies by McCarthy¹ and Morris², amongst others, we focused on recent developments and articles providing insights into clinical translation of genetic findings.
<table>
<thead>
<tr>
<th>Area of contribution</th>
<th>Rationale</th>
<th>Considerations and examples in T2D research</th>
</tr>
</thead>
</table>
| New drug target identification | In retrospective analyses, drugs with human genetics support are more likely to successfully transition through the drug development pipeline. Rapid development of new lipid-lowering drugs with genetic validation illustrates potential. Loss-of-function variants provide insights into likely efficacy and safety of inhibition, while gain-of-function on stimulation of target. | Genes encoding the targets of glucose lowering agents have been found by early GWAS, but new classes of diabetes drugs have not been developed as a result of human genetics findings. Most loci identified by GWAS have not lead to new drug development. Prominent examples are from Mendelian genetics, lack of examples for common forms of diabetes with polygenic aetiologic contribution. |}
| Mutation specific treatment  | Pharmacological interventions may be particularly effective in patients with particular underlying aetiology or genetic predisposition.                                                                                                                             | Prominent examples are from Mendelian genetics, lack of examples for common forms of diabetes with polygenic aetiologic contribution. Elegant exemplar from systematic study of all possible missense variants of PPARG. Opportunities in specific areas have not been fully exploited.                                                                                                                                 |
| Drug dosing or response      | Drugs may require dose-adjustment according to genetic background.                                                                                                                                                                                                   | Common variants at the ATM and SLC2A2 loci have been robustly associated with response to metformin, but genetic testing is not used in the clinic. Several studies including a recent trial have proposed an effect of TCF7L2 variants on response to glucose-lowering drugs. No established example of repurposing from other therapeutic areas to diabetes. Recent GWAS have explored repurposing opportunities using bioinformatics approaches. Genetic findings around the LPL pathway may provide basis for extension of future indications and target population for emerging lipid-lowering drugs targeting this pathway. |
| Drug repurposing             | As with new target identification, genetic variants that “mimic” existing therapeutic agents may provide the basis for repurposing.                                                                                                                                     | The example of low-density lipoprotein cholesterol lowering genetic variants in genes encoding targets of cholesterol lowering therapy (i.e. HMGCR, NPC1L1 and PCSK9) and diabetes risk illustrate power and challenges of genetic approaches, given the partial consistency between genetic and clinical trial results. |}
| Drug safety                  | Genetic variants can inform on desired and undesired secondary effects of pharmacological modulation of the encoded drug target. In diabetes, it is critical to study the cardiovascular safety of existing and new agents.                                         | GLP1R and ABCC8 variants have been used to gain insights into cardiovascular safety of existing glucose-lowering drugs.                                                                                                                                                                                                 |

Abbreviations: GWAS, genome-wide association studies.
Figure Legends

Figure 1. Illustrative representation of genome-wide studies in type 2 diabetes and their power to detect certain types of susceptibility alleles for a given sample size. Susceptibility alleles above the solid black lines are detectable with a given approach. The graph is informed by the results of actual historical and current GWAS studies as well as whole-genome and exome sequencing studies that provided an empirical model for the genetic architecture of type 2 diabetes. Exemplar genetic susceptibility loci are reported in the figure. Abbreviations: GWAS, genome-wide association studies; OR, odds ratio.

Figure 2. Aetiologic model for the role of TBC1D4 in GLUT4 translocation and insulin-mediated glucose uptake in the skeletal muscle.

Figure 3. Models for normal and impaired insulin secretion. Genetic variants affecting these processes result in impaired insulin secretion and higher diabetes risk (right panel), e.g., variants at KCNJ11 and ABCC8 identified in genome-wide association studies.

Figure 24. Aetiologic model for the contribution of peripheral adipose storage capacity to metabolic and cardiovascular disease and role of adipogenesis and intravascular lipolysis in this process. Some of the images have been samples and modified from SMART, Servier Medical Art, (URL: https://smart.servier.com/) under Creative Commons Attribution 3.0 Unported License (https://creativecommons.org/licenses/by/3.0/).
References


42. — Strawbridge RJ, Dupuis J, Prokopenko I, et al. Genome wide association identifies nine common variants associated with fasting proinsulin levels and provides new insights into the pathophysiology of type 2 diabetes. *Diabetes* 2011; 60(10): 2621-34.


Genetic mapping of diabetes susceptibility in the era of large biobanks

Early GWAS, N ~ 10,000s

Biobank-size GWAS, N ~ 100,000s

Meta-analysis of large biobanks, N ~ 1,000,000s

Effect size

- Extremely large (OR>20)
- Very large (5<OR<20)
- Large (1.5<OR<5)
- Modest (1.05<OR<1.5)
- Small (OR<1.05)

Minor allele frequency

- Ultra-rare <0.01%
- Rare 0.01%<MAF<5%
- Low-frequency 0.5%<MAF<5%
- Common >5%

Sample size:

- TCF7L2
- INSR
- IRS2
- ANKH
- DDX18

Fig 1
Normal skeletal muscle GLUT4 translocation

- Insulin
- Insulin receptor
- Blood glucose
- TBC1D4
- Independent pathways
- GLUT4

Cell membrane

Skeletal Myocyte

Impaired skeletal muscle GLUT4 translocation

- Insulin receptor
- Postprandial hyper-insulinaemia
- Higher blood glucose
- TBC1D4
- Independent pathways
- Reduced GLUT4 translocation and glucose uptake

Cell membrane

Skeletal Myocyte

Fig 2
Normal insulin secretion

- Blood glucose
- GLUT2
- K$_{ATP}$ channel
- ABCC8
- KCNJ11
- ATP
- Membrane depolarisation
- Ca$^{2+}$
- Ca$^{2+}$
- Ca$^{2+}$
- Ca$^{2+}$
- Insulin granule
- Metabolism

Impaired insulin secretion

- Blood glucose
- GLUT2
- K$_{ATP}$ channel
- ABCC8
- KCNJ11
- ATP
- Membrane depolarisation
- Ca$^{2+}$
- Ca$^{2+}$
- Ca$^{2+}$
- Insulin granule
- Metabolism

Pancreatic beta-cell

Impaired secretory response

Reduced Insulin granule secretion
**Functional peripheral adipose storage capacity**

- Ectopic fat susceptible tissues
- Lipoprotein Lipase mediated intravascular lipolysis
- Peripheral adipose storage
- Adipogenesis

**Impaired peripheral adipose storage capacity**

- Impaired storage capacity
- Obesogenic environment, impaired appetite regulation and positive energy balance
- Ectopic fat storage

Fig 4
Editor

Editor comment 1 (EC1): “We feel that addressing the reviewers' concerns will considerably strengthen the article, particularly those of reviewer 1. Some editors raised concerns about the heavy emphasis on findings that are not yet available in peer reviewed journals. While it is fine to use data from these sources, we would suggest that you use them as you would a conference abstract and clearly signpost when data are sourced from the bioRxiv and that these are not peer reviewed. If, in the meantime, the articles have been accepted for publication, of course these comments become irrelevant. With this in mind, please also add a sentence to your search strategy about how you found papers on BioRxiv, was this a systematic search?”

EC1, Authors’ reply: We agree and have now amended our review and references to pre-publication articles accordingly. Please also see response to reviewer 1, comment 1. Following the Editor’s and first reviewer’s concern, we have downplayed the relevance of pre-print publications and now state more clearly when we refer to them in the text. One of the articles has since been published in Nature Genetics. We now also provide the search strategy that was used to systematically identify relevant articles in BioRxiv.

EC2: “We wondered if you might also like to comment on recent contributions of genetic studies to classifications of disease beyond the monogenic forms of diabetes. Notably, Ahlqvist et al was recently published in TLDE and other clustering studies might be worth considering. This might fit with pathways rather than being a separate section.”

EC2, Authors’ reply: We agree that it is important to cite these approaches. The interesting article by Ahlqvist et al. has emerged after submission and we have now discussed and referenced this (P 11, L 420 of the new manuscript).

EC3: “Although we recognise that for many patients with type 2 diabetes their disease and obesity are intrinsically linked, we suggest that you shorten the discussion on obesity and lipids to make space for the additional suggestions by from reviewers.”

EC3, Authors’ reply: We have amended the manuscript to accommodate this suggestion (see detailed responses below).
Reviewer #1

Reviewer 1 comment 1 (R1C1): “this reviewer is surprised that this review is largely focused on a paper that has not been yet reviewed and published (only a preprint). In this regard the 250 loci for T2D should be cautiously presented especially for a large audience of the Lancet. The preprints are made to disseminate new ideas and to increase debate quality but this reviewer believes that only published peer reviewed papers should be presented in state of art review in particular in high IF journals.”

R1C1, Authors’ reply: We agree with this reviewer and editor that we needed to better highlight the non-peer reviewed nature of the BioRxiv articles and have amended the text of our article accordingly. One of the two manuscripts by Mahajan et al. has now been published in Nature Genetics and we reference it accordingly. The other has been reviewed and resubmitted and at this stage results are expected to not change dramatically. Because this large-scale meta-analysis a) substantially expands current knowledge, b) has been conducted by a team of internationally recognised experts with an outstanding track record in this area, and c) uses methodological standards that are accepted and go beyond what is expected for a discovery effort such as this one, we do think that reference to these results is appropriate and important and increases the timeliness and topicality of this review.

R1C2: “The abstract poorly summarizes the paper and should be rewritten. The last sentence is pure hope but not based on any robust evidence (that it may work) so far.”

R1C2, Authors’ reply: We have revised the abstract to better reflect the structure of the paper.

R1C3: “p 2: The 130 new loci should not be presented as granted. This introduction especially the last paragraph is too technical for geneticists only. This reviewer is not convinced that outside this small world people would understand what a posterior probability means. It is only a copy and paste of the preprint which is not the spirit of such a review for a large audience of readers.”

R1C3, Authors’ reply: We have revised the text to improve clarity and facilitate the understanding for a general and clinical audience, for instance by removing reference to the concept of posterior probability.

R1C4: “p3: the first paragraph is confusing as it does not differentiate GWAS analyses in various ethnicities and DNA sequencing of T2D genes (or exomes) in specific outbred populations. The DIAMANTE project is only a project and should not be presented as such in this review.”

R1C4, Authors’ reply: We have revised the text to reflect this suggestion and to more clearly distinguish findings arising from exome sequencing versus GWAS studies (P 3, L 83). We have trimmed the text referring to DIAMANTE, now mentioned only briefly but still highlighting the large scale and multi-ethnic nature of this important effort (P 3, L 92).
R1C5: “P 3 last paragraph: the ADCY3 story relates to food intake and obesity first and only secondarily to diabetes. The mechanism is likely to be very different from other T2D genes. In addition the ADY3 mutations, apart from the Inuit mutation were found in consanguineous populations. Of note these consanguinous populations are not isolated (such as Greenland Inuits who are 45K people) and constitute 20% of mankind. The study of consanguinous families was useful for monogenic diabetes but has not been correctly studied so far by GWAS and exome sequencing.”

R1C5, Authors’ reply: We agree with the reviewer that the ADCY3 example is relatively peripheral to the theme of this article and have shortened the corresponding text considerably (P 3, L 118), while still referring to the parallel publication based on the elegant study of consanguineous families.

R1C6: “p5: the pathway analysis part is very weak and too general in its findings. Many groups work (eg in Oxford, Lille, Malmo) on insulin secretion modulating genes found by GWAS and this quest should be presented. BTW the recent controversy on melatonin receptor opens the debate about genes directly or indirectly impairing insulin secretion (through brain effects) The last paragraph is very self centered on authors' own research. This reviewer has nothing against referring to fat deposition GWAS genes but it is rather far from T2D physiology (even if there is an overlap that should be presented with care). The lipodystrophy hypothesis is very interesting but deserves more explanation. Figure 2 is rather biased as seems to pretend that fat deposition is THE mechanism by which T2D develops in response to gene variants which is not proven. To this reviewer's knowledge recent papers from Gloyn/McCarthy, Bonnefond etc were strongly in favor of GWAS associated genes mostly expressed and functional in pancreatic beta cells. The fat deposition genes are more relevant to inflammation and cardiometabolic diseases. This part of the paper should be fixed.”

R1C6, Authors’ reply: It was not our intention to centre this part too strongly on fat deposition and following the reviewer’s suggestion we now present a more balanced view of the different pathways contributing to T2D pathophysiology, clarifying that fat deposition is a specific contributory mechanism but not the only one (P 5, L 150-208). We have also drawn additional figures depicting other relevant mechanisms in T2D aetiology, including impaired GLUT4 translocation and insulin secretion, aiming to provide a broader set of relevant mechanistic examples (Figures 2-3). Since these latter mechanisms have been discussed elsewhere, we defer to the Editor on the final decision of whether these additional figures add to the review and should be included in the final version.

R1C7: “p7: it is too early to compare the status of the 250 loci (in preprint for half!) and PCSK9 with drugs in the market. BTW this part (in particular the last paragraph) confuses GWAS analyses (the preprint paper and all others before) and targeted/exome sequencing for rare mutations (protective of LoF). Very confusing indeed.”

R1C7, Authors’ reply: Thank you for bringing this to our attention; we have revised the text to distinguish the more clearly between GWAS versus sequencing studies (e.g. P 8, L 293-299). It was not our intention to draw a direct comparison between newly emerging GWAS loci and PCSK9. We merely refer to this as a successful example of rapid translation (PCSK9 variants discovered by gene sequencing that prompted the development of new cholesterol
lowering drugs have been discovered around the time of first GWAS of T2D), because this supports the value of genomic discovery approaches and precision medicine more generally.

R1C8: “p8: drug response and safety: the paper forgot the most important papers in the field that are related to metformin complications (Pearson's papers). This reviewer is obliged to say that according to the title the paper is not on lipids, fat and PCSK9 but on diabetes genes!"

R1C8, Authors’ reply: We completely agree about the importance of Prof Pearson’s papers on metformin and these are now cited and discussed (P 9, L 326-337). We also trimmed the discussion of examples of lipid-lowering drugs to stay closer to the main focus of the article (P 9, L 364).

R1C9: “p10: genomics in the clinic: authors have a negative view of GWAS outcomes in clinic. It is their right although recent evidence from genomic Risk Scores show hope that soon absolute risk may be assessed by genetics. Authors totally ignore prospective studies of incident risk (means not in case control studies) that have illustrated the interest and limits of SNPs in the prediction of incident cases (see Swedish, French and US studies). In addition on top of genetics metabonomics offers strong hope for accurate prediction of incident cases (which BTW is THE question in clinic). See UK Lolipop, French Desir, Dutch studies...”

R1C9, Authors’ reply: This comment highlights the need to clarify our position and avoid any misunderstandings. Our comments on the relevance of GWAS findings for clinical prediction are meant to solely apply to the example of T2D, for which good, clinically accepted and used non-invasive and invasive predictive markers (i.e. HbA1c or glucose) already exist. These comments can and should not be applied to other outcomes of equally high clinical importance that are currently poorly predicted and/or for which genetic prediction may play an important role even in the near future.

In the light of the reviewer’s comment we have now revised this section (P 11, L 393) to also make it clearer that results are not based on case-control but prospective evidence (cohort and case-cohort).

Subgroups of the population where existing prediction algorithms work less well, and genetics adds relatively more (e.g. leaner and/or younger individuals) are at low absolute risk and hence from a population or universal clinical provider perspective, the possible benefits of testing for genotypes or other markers in these groups has to be weighed against their low absolute risk. We have modified the text to reflect these considerations (P 11, L 407).

R1C10: “The fact that Regeneron/GSK contribute to sequence patients genome does not mean that it will provide key insight on T2D genetics. Indeed their interest is more on LoF "protective " mutations in important genes found in general population modifying important phenotypes (lipids, glucose...). Thus wait and see...”

R1C10, Authors’ reply: We accept the views of the reviewer and have amended that section accordingly (P 11, L 436-453). As reviewer 3 highlights, “genomic studies for the identification of relevant genomic regions and validation of new therapeutic targets show tremendous promise for transformative clinical impact”. As academia-industry partnerships
delivering large-scale genomics are becoming more prevalent, the challenges and opportunities arising from such efforts for T2D deserve some mention in our view.
Reviewer #3:

Reviewer 3 comment 1 (R3C1): “The authors reviewed the literature and have discussed several points concerning the hereditary basis of type 2 diabetes, pathways involved in diabetes identified through genomic studies, the genomics and therapeutics of type 2 diabetes, and the clinical relevance and future outlook. Genome-wide association studies have implicated many genomic regions, and there has been considerable advancement in the understanding of the mechanisms underlying these associations. Genomic studies for the identification of relevant genomic regions and validation of new therapeutic targets show tremendous promise for transformative clinical impact. Therefore, the authors focused on recent developments and offer insights into clinical translation of genetic findings.”

R3C1, Authors’ reply: We are grateful to the reviewer for thoroughly reviewing our work and providing important suggestions for improvements.

R3C2: “Major comments: Page 5, lines 24-43.
Shungin et al. identified several loci associated with body fat distribution with stronger effects in one sex than the other (Nature. 2015;518(7538):187-96). This finding should be mentioned because variants with sex-specific effects are interesting and important for considering the genetic basis of diabetes.”

R3C2, Authors’ reply: We agree and now cite this relevant article in the new version of the manuscript (P 5, L 191). However, we had to considerably shorten this section and the discussion of body fat distribution following the requests of the editor and other reviewers, so were unable to discuss details of any sex specific effects.

R3C3: “Page 8, lines 9-18.
The limitations of pharmacogenetics study design should be addressed. Recently, Srinivasan et al. conducted a novel prospective pharmacogenetic clinical trial (SUGAR-MGH), which revealed that a TCF7L2 variant associated with T2D influences the acute response to both glipizide and metformin in non-diabetic participants (Diabetes Care. 2018;41(3):554-61). The study design of SUGAR-MGH is very important because it is free of the uncontrolled nature of retrospective clinical data sets. The authors should describe variants at not only ATM and SLC2A2 but also TCF7L2 loci that were associated with response to diabetes drugs in Table 1.”

R3C3, Authors’ reply: Thank you for raising this; we have cited this interesting work and also added TCF7L2 to the table (P 9, L 329).

R3C4: “Page 8, lines 19-32.
A systematic bioinformatics approach would be useful for repurposing of approved drugs in diabetes. For example, Imamura et al. identified two genes, GSK3<beta> and JUN, whose products directly interact with those of multiple biological T2D susceptibility genes, using a bioinformatics approach (Nature communications. 2016;7:10531). While therapeutic drugs for diseases other than diabetes targeting
GSK3β and JUN were under clinical trials, these compounds could also be potential treatments for T2D. The authors may want to cite this article.”

R3C4, Authors’ reply: We thank the reviewer for this suggestion and now cite the article as a possible way forward in genetic-evidence driven re-purposing (P 9, L 347).

R3C5: “Minor comments:
Figure 1.
The authors should clarify the meaning of the number next to each given approach (e.g. ~10,000s). Does it indicate the total number of variants genotyped?”

R3C5, Authors’ reply: We have modified the figure to clarify that numbers refer to the sample size (see Figure 1).

R3C6: “Page 10, lines 7-21.
Regarding diabetes prediction models, the authors should mention not only the clinical utility of a genetic risk score but also machine learning applications. For example, Shigemizu et al. developed a predictive model for T2D that consisted of nine SNPs selected using a Bayes Factor and lasso method with three clinical risk factors (age, gender, and BMI) and conducted a two-stage study (training and test sets) in a prospective cohort in Japan. The predictive model exhibited a 1.5% increase in the AUC over the clinical risk factors alone (PLoS One. 2014;9(3):e92549).”

R3C6, Authors’ reply: This is an interesting point and we have modified the text to reference advanced statistical modelling approaches for improved prediction (P 11, L 402).
Reviewer #4:

Reviewer 4 comment 1 (R4C1): “I really liked this review. I think the last two sections in particular focus on the translational/clinical side of T2D genetics in a way I haven't seen too often before. Below are hopefully constructive comments:”

R4C1, Authors’ reply: We would like to thank the reviewer for assessing our work so carefully and for the helpful and constructive comments. We have made changes to the manuscript in response to these suggestions and feel that the manuscript has greatly improved as a result.

R4C2: “Section 1 (global view)
In general I found the first two sections to need the most work. They seemed relatively unfocused and not necessarily making an argument leading up to the last two sections, which I felt to be quite strong. The shift in fact was quite abrupt. Maybe the two sections could state the overriding hypothesis of the review (which to me is how T2D genomics can impact translational and clinical research), frame the findings in the first two sections as such, and use their results as a foundation to support the major conclusions (e.g., to evaluate drug targets you need a lot of data and quantitative traits; for genetic risk prediction you need a lot of multi-ethnic data). I realize this is high level so below I'll call out specific areas that struck me.”

R4C2, Authors’ reply: Thank you for this suggestion; we have re-structured the first two sections in light of these suggestions and provide more detailed responses to related changes in our answers to the following comments below.

R4C3: “Very early on in Section 1 the authors state the reported (strong) ORs and lower frequencies from the recent T2D GWAS as evidence of much higher effect variants. It's important to mention (in a review like this) that these variants are of a different ilk than those previously reported: they haven't been validated in individual cohorts or even directly genotyped to make sure the imputation is accurate. Additionally, with the extreme case/control imbalance rare variant test statistics may not be well calibrated; a sentence suggesting this caution would be appropriate (particularly since future studies are going to only exacerbate this issue). Related to the first comment, I think at some point imputation quality should be addressed for these newer studies of much rarer variants.”

R4C3, Authors’ reply: We agree with these insights and have now added these points and downplayed our original comments about the newly identified rare variants (P 2, L 56).

R4C4: “The genetic isolate section seemed a little unfocused. It was framed as helping address the global basis of T2D, but then it goes into detail on three genes, the mechanism of which isn't really related to populations in which they are discovered. Mechanistic understanding of genes could come from variants discovered in any population; the fact that these variants were discovered in isolates is kind of orthogonal to the main content. The point of why isolates are useful is that large effect variants may (by chance) rise to high frequency in the isolate, which in some ways means that
individual high effect variants are better powered for detection in isolates. But this point doesn't come through with the micro focus on gene mechanisms. I personally think the gene mechanisms are well-covered in countless reviews and this one could focus more on why isolates can help the global epidemic of T2D. Restating this more broadly, the "global view" in the title doesn't really come through”

R4C4, Authors’ reply: We agree with this comment and try to now provide a higher level “global view” by giving more prominence to the discussion of the overall utility, advantages and insights provided by studying population isolates (P 3, L 102). We have trimmed the discussion of specific mechanisms accordingly.

R4C5: “Is it worth mentioning the T2D knowledge portal, which is funded by five pharmaceutical companies and the NIH/FNIH Accelerating Medicines Partnership? T2D has a unique commitment to data sharing.”

R4C5, Authors’ reply: We completely agree and we now reference the T2D Knowledge Portal as a very important resource for T2D researchers (P 2, L 38).

R4C6: “How were the lines drawn in Figure 1? Is there a reason the bottom curve has a sharp point at 0.01%? If not, I would make it smooth. At minimum, some justification for this figure beyond pure heuristics would be useful (e.g., show actual discovered MAF/OR combinations)?”

R4C6, Authors’ reply: We have drawn the figure using Microsoft pptx. Figures will be redrawn and improved by the Lancet’s team for the final version. Following this comment, we have now added exemplar genetic variants to the graph and made it clear that the graph follows the empirical results of previous GWAS and sequencing studies (see Figure 1 and its legend).

R4C7: “Section 2
The idea about the trade-off between sample size and degree of refinement of phenotype ascertainment was very interesting and something I had not seen clearly articulated before. It wasn't really developed though. This whole section was extremely vague and didn't tell me a whole lot. I would love to see it refocused on the first sentence, and to tell me how different studies had chosen one of these trade-offs rather than the other, and what each could tell us about T2D. When do you do one and not the other? What findings were from one but not the other?”

R4C7, Authors’ reply: We thank the reviewer for this helpful comment and have amended the new version accordingly (P 5, L 152-167).

R4C8: “Should PPARG be mentioned in the second subsection?”

R4C8, Authors’ reply: We agree and have now added PPARG to this section (P 6, L 203).
R4C9: “The multi-omic data section I didn't really follow at all. It sets up a bunch of technologies, but then there are two paragraphs on metabolomics based discovery. Where are the other technologies? How did these prioritize pathways? A nested association is not a pathway. I think there is a case to be made that multi-omics is a major future direction, but this section seems to only cover metabolomics. In general, I found this section to be really weak. I think the two ideas (trade-offs in how secondary traits are studied, multomics) are interesting to address, but they are not (at least upon my reading) actually addressed.”

R4C9, Authors’ reply: We have amended the section to now include a wider and more structured discussion of “-omics” approaches (P 6, L 211-254).

R4C10: “Should Mendelian randomization be covered, particularly given the clinical bent of the paper?”

R4C10, Authors’ reply: While we agree that genetic approaches to causality are an interesting topic but given that this has been reviewed and covered extensively elsewhere, we think that this is beyond the scope of this review. However, we have now included clear reference and a brief statement to bring this topic to an interested reader’s attention (P 6, L 229).

R4C11: “I would downplay the pathway/mechanism findings in general. They are overcovered by every other T2D review, and pruning them could focus this on the interesting ideas it broaches that have not been covered elsewhere.”

R4C11, Authors’ reply: We have shortened the sections about specific mechanisms to be able to expand the topics that have not been extensively covered by previous reviews as suggested (e.g. section on ADCY3, P3 L 118).

R4C12: “Sections 3 & 4

I really liked these sections, as mentioned above. My only suggestion is to refocus sections 1 and 2 so that they set up this section even more. I don't have any major comments here.”

R4C12, Authors’ reply: We thank the reviewer for these supportive comments.

R4C13: “My only minor comment (maybe a question): is this review supposed to only cover genomics of T2D in the UK? All of the future data mentioned is in UK, and similar US efforts like the MVP and All of Us are not mentioned. The explicit caveat that UK citizens have less to worry about with respect to discrimination made me think perhaps it is UK focused, but given that the title of the first section is "A global view of T2D genomics" the focus on the UK seems odd.”

R4C13, Authors’ reply: We agree and present a more balanced discussion of these topics in the new version, including specific reference to the MVP and All of Us efforts (P 12, L 455-463).