Common and rare forms of diabetes mellitus: towards a continuum of diabetes subtypes

Jason Flannick, Stefan Johansson and Pål R. Njølstad

Abstract | Insights into the genetic basis of type 2 diabetes mellitus (T2DM) have been difficult to discern, despite substantial research. More is known about rare forms of diabetes mellitus, several of which share clinical and genetic features with the common form of T2DM. In this Review, we discuss the extent to which the study of rare and low-frequency mutations in large populations has begun to bridge the gap between rare and common forms of diabetes mellitus. We hypothesize that the perceived division between these diseases might be due, in part, to the historical ascertainment bias of genetic studies, rather than a clear distinction between disease pathophysiology. We also discuss possible implications of a new model for the genetic basis of diabetes mellitus subtypes, where the boundary between subtypes becomes blurred.

Type 2 diabetes mellitus (T2DM) affects ~400 million people worldwide and imposes a substantial public health burden owing to a range of diabetic complications and the inadequacy of current therapies. As T2DM is a heritable disorder, the possibility of studying human genetics to provide insight into disease biology, treatment and prevention has received much attention. However, the large environmental component of T2DM susceptibility poses challenges for identifying informative mutations.

In contrast to T2DM, much more is known regarding the pathophysiology of other forms of diabetes mellitus. Type 1 diabetes mellitus (T1DM) is caused by autoimmune-mediated destruction of β cells, a pathogenesis (presumably) distinct from that of T2DM. However, many rarer forms of diabetes mellitus do share clinical similarities with T2DM, including maturity-onset diabetes of the young (MODY), neonatal diabetes mellitus (NDM), mitochondrial diabetes mellitus and certain multiorgan syndromes such as Wolcott–Rallison syndrome and Wolfram syndrome. For many of these conditions, the genetic causes are known and have led, in some cases, to diagnostic screening tools or effective treatments.

As diagnoses of diabetes mellitus are based on clinical presentation rather than specific molecular defect(s), a longstanding hypothesis has been that the genetic aetiologies of common and rare forms of diabetes mellitus might not be wholly distinct. An obvious question is, thus: to what extent can knowledge about rare forms of diabetes mellitus inform our understanding of T2DM and vice versa? Here, we discuss how genetic studies have begun to elucidate the links between these disorders. We first review the clinical presentations of these conditions and classic approaches to dissect their genetic bases. We then discuss studies that combine features of these traditional approaches: studies of rare variants (traditionally analysed for rare diseases) for association with common T2DM; and studies of large populations (traditionally analysed for common diseases) for mutations causing rare forms of diabetes mellitus. In our view, genetic evidence now supports a unified risk model for rare and common forms of diabetes mellitus; we explore the consequences for future research, diagnosis and treatment.

Clinical presentation

T2DM

T2DM presents as chronic hyperglycaemia, typically after 40 years of age, and is thought to result from the progressive failure of pancreatic β cells to secrete sufficient insulin to meet metabolic demands. Obesity and age are major risk factors for T2DM, as they result in insulin resistance that typically precedes diabetes mellitus. Most drugs for T2DM, such as insulin, metformin and sulfonylureas, are primarily intended to lower blood glucose levels, but have, thus far, been unable to change the progressive course of T2DM or adequately reduce the risk of late-stage complications associated with diabetes mellitus.

Rare forms of diabetes mellitus

A number of much rarer, monogenic conditions can also result in nonautoimmune diabetes mellitus [FIG. 1]. The most notable, MODY, is characterized by...
We propose that a unified diabetes mellitus risk model, spanning variants of all relevant to monogenic diseases suggests an overlapping disease aetiology, possibly including distinct T2DM subtypes. We further display mild fasting hyperglycaemia (5.4–7.6 mmol/L) from birth but with mild deterioration with ageing. More than 80 disease genes, most of which have been thoroughly reviewed elsewhere, have also successfully identified causal variants in known genes. Novel genes have been comparatively less frequently identified via such approaches, although successes have been observed for GATA6 and pancreatic agenesis33; KCNJ11 (REF. 32) and APPL1 (REF. 33) for MODY; and CDKN1C8 and PRPS1 (REF. 35) for syndromic forms of diabetes mellitus characterized by intrauterine growth restriction.

Many of our insights into the disease mechanisms of diabetes mellitus are due to knowledge of genes associated with monogenic forms. The KATP channel, which has an essential role in glucose-stimulated insulin secretion, was shown to have inactivating mutations, which cause hyperinsulinaemia and hypoglycaemia34, as well as activating mutations, which can variously cause NDM35 or MODY36. In the same pathway, the role of glucokinase (GCK) as a glucose sensor was confirmed when homozygous mutations, which cause complete GCK deficiency, were shown to cause NDM due to impaired glucose-dependent ATP production and inhibition of KATP channel closure37. In other pathways, the identification of MODY mutations in genes encoding liver-enriched transcription factors active during embryogenesis (HNF1A, HNF4A and HNF1B) confirmed observations that several MODY subtypes are in fact early-onset multiorgan diseases38–41, whereas the identification of MODY mutations in CEL (expressed in pancreatic acinar cells) showed that cell types other than the β-cell might be implicated in diabetes mellitus40.

This improved understanding of the aetiology and pathogenesis of monogenic diabetes mellitus has, in many cases, led to improved patient care. Many countries have introduced screening programs for all children diagnosed with diabetes mellitus before 6 months of age42, as a diagnosis of activating KATP channel mutations can usually enable patients to safely discontinue insulin and start high-dose sulfonylureas43,44. Patients diagnosed with a mutation in HNF1A or HNF4A are similarly highly sensitive to sulfonylurea treatment45. Patients who carry a mutation in GCK, who further display mild fasting hyperglycaemia (5.4–8.3 mmol/L, HbA1c range 5.8–7.6%) from birth but with mild deterioration with ageing46, do not require any treatment at all, as the increased fasting glucose...
Genome-wide association studies (GWAS). An approach for identifying disease genes by testing many common genetic variants in different individuals for associations with a trait.

Expression quantitative trait loci (eQTL). A genomic region that is associated with variation in expression levels of mRNA.

‘set-point’ for GCK–MODY is seldom sufficient to cause microvascular complications\(^2\). Insulin therapy is usually required for other conditions, however, such as HNF1B–MODY or the majority of mitochondrial diabetes mellitus conditions.

**Common variants for T2DM**

Unlike monogenic diabetes mellitus, T2DM lacks a clear inheritance pattern, being characterized instead by familial aggregation; for example, a twofold to threefold relative risk and a heritability of 30–70%\(^3\). A substantial amount of T2DM risk is environmental, which leads to much smaller genetic effect sizes than those for monogenic forms (BOX 1). Consequently, linkage scans have yielded few signals for T2DM\(^5\).

By contrast, genome-wide association studies (GWAS) of common variants have proven far more successful than linkage studies for mapping T2DM loci (the major findings are reviewed elsewhere\(^1\)). In brief, sample sizes now exceed 100,000 (REFS 53,54), spanning multiple ethnicities and identifying around 100 genomic loci (FIG. 2). All variants affect risk modestly (with relative risks usually around 1.1–1.2 (REF 55)) and explain a significant but small amount (10–15%) of T2DM heritability. These findings are increasingly interpreted as consistent with a polygenic model for the genetic basis of T2DM\(^5,56\), which is characterized by possibly thousands of distinct genetic risk factors\(^5\). Most associations map to noncoding regions of the genome, preferentially acting through regulatory elements such as pancreatic islet enhancers\(^57,59\).

As GWAS associations implicate genomic regions that span hundreds of kilobases, less is known about the causal genes or variants responsible for these associations. Efforts to localize signals to individual variants have had limited success, yielding ‘credible sets’ of candidate variants that are sometimes manageable but usually ambiguous\(^54,60\). Additional data and experiments have in some cases suggested promising genes or variants at these loci. For the FTO locus, for example, a series of studies used cross-species conservation\(^61\) or chromatin conformation capture\(^62,63\) to link the GWAS association to the distant IRX3 transcript; a later comprehensive investigation of the same locus used epigenomics, transcript expression, motif conservation and network analysis to suggest an IRX3-mediated regulatory mechanism underlying adipocyte thermogenesis that was subsequently validated by directed perturbations, mouse models and endogenous CRISPR–Cas9 genome editing\(^64\). Parent-of-origin effects and differential imprinting patterns have also been used to gain insight into effector transcripts (for the KLF14 locus\(^65\)), the stage of development that mediates disease risk (for the KCNQ1 locus\(^66\)) or tissue-specific effects (for the GRB10 locus\(^67\)). Additional insights into causal genes or variants have come from cross-species comparisons at the PPARG locus\(^68\), overlap with chromatin immunoprecipitation binding sites at the MTNR1B locus\(^69\), or expression quantitative trait loci (eQTL) mapping at the ZMIZ1 locus\(^70\). For most loci, however, the causal genes or biological mechanism remain to be elucidated.

For a subset of common variant signals, a gene implicated in monogenic diabetes mellitus lies nearby and, therefore, a strong candidate for causality. For example, an early candidate gene study of HNF1A identified a common variant private to the Canadian Ojib-Cree population with a moderate (twofold) effect on T2DM risk\(^71\); additional common variant associations for T2DM were identified from candidate gene studies of PPARG\(^72\), a gene implicated in a severe form of insulin-resistant diabetes mellitus and lipodystrophy, and KCNJ11, which is known for its role in PNDM\(^73\). Early T2DM GWAS identified further associations near HNF1B, HNF1A and WFS1 (REF 73). As studies have increased in size, the overlap between common variant signals and genes associated with monogenic diabetes mellitus has continued to increase; as of 2015, >10% of common variant signals for T2DM localized within 250 kb of a gene associated with monogenic diabetes mellitus, with additional overlap observed for glucose or insulin common variant associations (FIG. 2). Furthermore, a substantial number of signals localize within the same regulatory networks affected by mutations linked to monogenic diabetes mellitus, as demonstrated by the enrichment of T2DM or glucose GWAS associations within binding sites for transcription factors implicated in MODY\(^57,58,64\).

**Figure 1 | Monogenic forms of diabetes mellitus.** Shown are genes implicated in a restricted set of monogenic diabetes mellitus subtypes (discussed in the article). Each class of disorder is shown as a box in the central column: neonatal diabetes mellitus (NDM, red); maturity-onset diabetes of the young (MODY, blue); and mitochondrial diabetes mellitus (grey), while syndromic forms are divided based on closer similarity to NDM (pink) or MODY (light blue). The major clinical features of each disorder are described in each box. NDM and MODY with organ dysfunction can have one or more organs affected. Genes are shown as boxes in the outer two columns, with arrows connecting to the disorders for which the genes are implicated. Genes implicated in more than one disorder have arrows to each. CNS, central nervous system.
Box 1 | Distinction between rare and common forms of diabetes mellitus

Rare and common forms of diabetes mellitus are characterized by vastly different ‘genetic architectures’, defined as the number, frequencies and effect sizes of causal variants in a population. Rare forms are characterized by highly penetrant, often assumed deterministic, variants within a single gene, although incomplete penetrance can occur and lead to overestimation of risk. Although heterogeneity is often high, with numerous causal variants, the combined allele frequency is very low, with individual mutations very rare or even private to a familial lineage.

The genetic architecture of type 2 diabetes mellitus is far less certain than for rare forms of diabetes mellitus. In general, ‘rare variant’ models, in which very low frequency variants explain the bulk of genetic risk of the disease, ‘common variant’ models, in which common variants explain the bulk of risk, and intermediate models are all possible. Rare variant models have been justified on the basis of the modest effects of common variants identified thus far. Common variant models have, conversely, been justified based on the substantial amount of disease risk that common variants explain when analysed in aggregate.

Although perhaps academic, genetic architecture greatly determines the degree to which genetic studies can inform on disease risk and mechanisms. Rare coding variants with large effect clearly implicate disease mechanisms, as they suggest a specific gene for which a perturbation can have a substantial phenotypic effect; loss-of-function protective mutations can further suggest a directional perturbation with a clinical benefit. Furthermore, disease risk can be substantially elevated in individuals who carry large-effect mutations, which enables risk prediction or preventive measures. By contrast, common variant associations of weak effect are less valuable for risk prediction and can be challenging to localize to individual genes or disease mechanisms. Links between rare and common forms of diabetes mellitus, therefore, provide opportunities to translate insights from well-understood to less-understood disease forms.

Low frequency variants for T2DM

Findings from T2DM GWAS have, thus, shown a clear genetic link between common and monogenic forms of diabetes mellitus, although the properties of variants differ greatly between disease forms. Variants for monogenic diabetes mellitus are rare and result in large, possibly ‘deterministic’, increases in disease risk. GWAS variants are generally common and result in modest effects on disease risk. The increasing scale of NGS has, in the past 2–3 years, enabled studies to test whether a broader ‘allelic series’ — consisting of low frequency or rare variants with moderate-to-large effect on risk — contributes to T2DM.

The studies most successful at identifying lower frequency associations (that is, a frequency of ≤5%) have used the Illumina Human Exome Array, which includes ~70–80% of coding variants with a frequency >0.5% in the European population. The first study, in 2013, analysed ~8,000 individuals and identified five independent low-frequency variants, each explaining 0.4–0.9% of genetic variance for fasting proinsulin levels or insulinogenic index. Two 2015 studies of 33,000–61,000 individuals identified a series of low-frequency variants in G6PC2, as well as a variant in GLP1R, associated with fasting glucose levels. An additional low-frequency variant in AKT2 was associated with fasting insulin levels. A striking finding from a separate study, which used a similar genotyping array (the Metabochip) in the isolated Inuit population, was a protein-truncating variant in TBC1D4 for which homozygous carriers have a 10-fold increased risk of T2DM in addition to increased postprandial levels of glucose and insulin. Despite these positive examples, however, the largest exome array study conducted to date for T2DM identified only one statistically significant low-frequency association (a previously reported variant in PAM), even though numerous common variants far exceeded stringent significance thresholds.

Genome-wide sequencing studies have yielded a few additional associations. Although the first WES study identified no low-frequency or rare variants associated with T2DM in 2,000 Danish individuals, a similar study of ~4,000 Mexican and Latino individuals identified one low-frequency variant in HNF1A with a fivefold increased risk of T2DM. In a larger study that imputed variation from ~2,600 Icelandic whole genomes into ~278,000 individuals, four rare or low-frequency variants were associated with T2DM: two in PAM, which were reported in the first exome array study as associated with insulinogenic index, and one in each of CCND2 and PDX1, with twofold effect sizes on T2DM risk. Sequencing of more heterogeneous samples (13,000 multiethnic exome sequences and 2,600 Northern European genome sequences with imputation into 44,000 additional Europeans) did not identify any additional low-frequency or rare variants associated with T2DM, apart from the previously reported PAM and CCND2 signals.

Relative to genome-wide studies, targeted sequencing studies of selected genes have yielded more promising findings. In most cases, restriction of analysis to predicted functional variation has been necessary to demonstrate association. When 36 MTNR1B variants with frequencies below 0.1% were assessed for melatonin binding and signalling capacity, partial or total loss-of-function variants were associated with a 5.6-fold increased risk of T2DM. Similarly, nine rare PPARG variants that resulted in defects in adipocyte differentiation were associated with a 7.2-fold increase in T2DM risk, and 11 rare variants that impaired transcriptional activity, DNA binding or nuclear localization of HNF1A were associated with a sixfold increase in T2DM risk. In perhaps the largest targeted gene study to date, beginning with exon sequencing of Finnish individuals and culminating in analysis of 150,000 individuals of five ethnicities, a series of 12 predicted protein-truncating rare variants in SLC30A8 (which also harbours a common missense variant with a modest effect on T2DM risk) were associated with a threefold reduction in T2DM risk.

Many of the rare or low-frequency variants identified in these studies are located near GWAS signals or within genes linked to monogenic diabetes mellitus (FIG. 2), which supports the ‘allelic series’ hypothesis for T2DM risk. Two sequencing studies have examined this hypothesis further. The first study sequenced seven MODY genes in ~4,000 individuals from the general population and demonstrated an excess of predicted deleterious mutations in individuals with diabetes mellitus. The second study analysed the sequences of 82 genes associated with monogenic diabetes mellitus in 13,000 individuals and demonstrated a strong enrichment of predicted deleterious alleles in individuals with diabetes mellitus. Enrichment increased with
Figure 2 | Variants associated with monogenic diabetes mellitus, T2DM or levels of glucose or insulin. Shown is a list of reported associations between genes and diabetes mellitus or related traits. Associations are labelled based on the gene through which the associated variant is hypothesized to act (for variants identified in genome-wide association studies (GWAS), in many cases the gene is uncertain). The outer three circles show common variant associations for type 2 diabetes mellitus (T2DM) (blue), fasting or 2-h glucose levels (yellow), and fasting insulin or proinsulin levels (green). The middle circle (red) shows variants reported for monogenic disorders due to abnormal glucose or insulin levels, or adipocyte function. The inner three circles show higher-effect variants (in most cases of low frequency) associated with insulin (light green), glucose (light yellow), or T2DM (light blue). Pink lines are drawn through genes for which associations have been reported for a monogenic disorder as well as with a common or low-frequency variant. Grey lines are drawn through genes for which associations with both a higher-effect and common variant association have been reported.
predicted molecular severity of the variant or predicted relevance to diabetes mellitus of the affected gene. In each of these studies, analyses ruled out undiagnosed monogenic diabetes mellitus as an explanation for the associations.

**Variant effects in the population**

NGS studies of T2DM have, thus, suggested that rare or low-frequency alleles of moderate effect for T2DM coexist alongside common alleles of low effect for T2DM, as well as alongside rare alleles of high penetrance for monogenic diabetes mellitus. Plausibly, upon re-examination, variants previously reported to cause monogenic diabetes mellitus might, in fact, have more modest effects on diabetes mellitus risk. For example, linkage or familial studies of monogenic diabetes mellitus discover individuals with severe forms of the disease; these individuals are enriched not only for disease alleles but also for many other risk factors, which, when unaccounted for, can increase risk estimates for the general population. Other biases, such as unrecognized confounding or ‘winner’s curse’, can also lead to overestimated association strengths. Benign variants within disease genes can also be falsely classified as molecularly damaging and, indeed, large-scale sequencing efforts have shown that putatively healthy individuals carry a substantial number of such variants.

A study published in 2013 investigated the extent to which variants in genes implicated in MODY affected diabetes mellitus risk in the general population. Despite stringent bioinformatics filters, ~1% of individuals were found to carry variants potentially pathogenic for MODY; however, none developed MODY and there was, at most, a weak trend towards the development of diabetes mellitus later in life. Similarly, an exome sequencing study showed that variants reported in the literature to cause MODY were only modestly associated with diabetes mellitus risk, with effect sizes of the order of those for common variants identified from GWAS.

To illustrate specific examples of variants that might have lower penetrance than previously assumed, we examined a set of mutations initially reported as associated with monogenic diabetes mellitus but subsequently observed in large exome sequencing datasets. We intersected variants for the five most common MODY subtypes (GCK, HNF1A, HNF4A, HNF1B and INS), as reported in the ClinVar database, with those identified in a T2DM exome sequencing study of ~17,000 individuals (Type 2 Diabetes Genetics Portal). We then followed a protocol used in clinical diagnostic laboratories and guidelines provided by the American College of Medical Genetics, together with a manual examination of the original published evidence of pathogenicity, to assign each variant to one of five classes: class 1, benign; class 2, likely benign; class 3, variants of unknown significance (VUS); class 4, likely pathogenic; and class 5, pathogenic.

A total of 15 variants (cumulative allele frequency of 0.5%) in the T2DM exome sequencing study were labelled as either pathogenic or likely pathogenic in the ClinVar database. Our manual assessment based on current protocols, however, showed that only three were assessed as pathogenic, three as probably pathogenic, and the remaining nine as either VUS or likely benign. For many variants that we reclassified as VUS or likely benign, the original report of pathogenicity provided little information on whether the variant segregated with the disease in the families of the probands. Few of these variants were, or have been since, tested for functionality. Thus, in this small sample, almost two-thirds of variants initially reported as fully penetrant for MODY have, based on the best-practices used today, very limited evidence for being pathogenic.

Indeed, the variants reclassified as VUS or likely benign were, in aggregate, observed at equal frequencies in individuals with and without diabetes mellitus (Type 2 Diabetes Genetics Portal). Of the nine variants, more than half were observed in individuals without diabetes mellitus and three were observed in at least five individuals without diabetes mellitus. By contrast, five of the six variants reclassified as pathogenic or likely pathogenic were observed exclusively in individuals with diabetes mellitus. This example, while not statistically significant, suggests that modern criteria for variant classification might improve upon those originally used to identify MODY mutations. However, the presence of pathogenic or likely pathogenic variants in individuals with reported T2DM, but not MODY, suggests that even modern criteria can overstate the penetrance of mutations classified as pathogenic for monogenic diseases, and that even more stringent criteria (such as functional characterization) might be necessary to establish pathogenicity.
Table 1 | Analysis of variants originally reported as causal for MODY

<table>
<thead>
<tr>
<th>Variant/MODY subtype</th>
<th>cDNA change</th>
<th>Protein change</th>
<th>Assessment of pathogenicity</th>
<th>Frequencies in 17,000 exomes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>ClinVar</td>
<td>Manual</td>
</tr>
<tr>
<td><strong>Manual evaluation: All pathogenic and likely pathogenic</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GCK/MODY2</td>
<td>c.683C&gt;T</td>
<td>p.T229M</td>
<td>Pathogenic</td>
<td>5: Pathogenic</td>
</tr>
<tr>
<td>GCK/MODY2</td>
<td>c.676G&gt;A</td>
<td>p.V227M</td>
<td>Pathogenic</td>
<td>5: Pathogenic</td>
</tr>
<tr>
<td>GCK/MODY2</td>
<td>c.1240A&gt;G</td>
<td>p.K415E</td>
<td>Likely pathogenic</td>
<td>5: Pathogenic</td>
</tr>
<tr>
<td>GCK/MODY2</td>
<td>c.1160C&gt;T</td>
<td>p.A388V</td>
<td>Likely pathogenic</td>
<td>4: Likely pathogenic</td>
</tr>
<tr>
<td>HNF1A/MODY3</td>
<td>c.965A&gt;G</td>
<td>p.Y322C</td>
<td>Likely pathogenic</td>
<td>4: Likely pathogenic</td>
</tr>
<tr>
<td>HNF4A/MODY1</td>
<td>c.991C&gt;T</td>
<td>p.R331C</td>
<td>Likely pathogenic</td>
<td>4: Likely pathogenic</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td><strong>Manual Evaluation: All VUS or benign</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HNF1A/MODY3</td>
<td>c.92G&gt;A</td>
<td>p.G31D</td>
<td>Pathogenic</td>
<td>3: VUS</td>
</tr>
<tr>
<td>HNF4A/MODY1</td>
<td>c.406C&gt;T</td>
<td>p.R136W</td>
<td>Pathogenic</td>
<td>3: VUS</td>
</tr>
<tr>
<td>GCK/MODY2</td>
<td>c.440G&gt;A</td>
<td>p.G148D</td>
<td>Likely pathogenic</td>
<td>3: VUS</td>
</tr>
<tr>
<td>HNF1A/MODY3</td>
<td>c.1424C&gt;T</td>
<td>p.P475L</td>
<td>Likely pathogenic</td>
<td>3: VUS</td>
</tr>
<tr>
<td>HNF1A/MODY3</td>
<td>c.1745A&gt;G</td>
<td>p.H582R</td>
<td>Likely pathogenic</td>
<td>3: VUS</td>
</tr>
<tr>
<td>HNF1A/MODY3</td>
<td>c.1748G&gt;A</td>
<td>p.R583Q</td>
<td>Likely pathogenic</td>
<td>3: VUS</td>
</tr>
<tr>
<td>HNF1B/MODY5</td>
<td>c.140C&gt;T</td>
<td>p.P47L</td>
<td>Likely pathogenic</td>
<td>3: VUS</td>
</tr>
<tr>
<td>HNF4A/MODY1</td>
<td>c.505G&gt;A</td>
<td>p.V169I</td>
<td>Likely pathogenic; uncertain</td>
<td>2: Likely benign</td>
</tr>
<tr>
<td><strong>INS/Hyperpro-insulinaemia</strong></td>
<td>c.266G&gt;A</td>
<td>p.R89H</td>
<td>Pathogenic*</td>
<td>3: VUS*</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td><strong>ClinVar: All pathogenic and likely pathogenic</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

*As a representative example of variants initially reported as pathogenic for maturity-onset diabetes of the young (MODY), but which might in fact have reduced penetrance, we intersected variants reported in the ClinVar database as causal for the five most common MODY subtypes (GCK, HNF1A, HNF4A, HNF1B and INS) with those identified in a type 2 diabetes mellitus exome sequencing study of ~17,000 individuals and we then re-analysed the original reported evidence for pathogenicity according to a modern protocol used in clinical diagnostic laboratories. The table above shows the 15 analysed variants, together with their modern classification and frequency information from the exome sequencing study. Gene/MODY subtype: the reported MODY phenotype; cDNA change/protein change: the effect on the coding transcript or protein; ClinVar: the classification as reported in the ClinVar database; Manual: the classification using modern criteria; minor allele frequency (MAF): the frequency across 17,000 exomes; minor allele count (MAC) cases/controls: the number of case and control observations in the 17,000 exomes. VUS, variant of unknown significance.

A continuum of genetic effects

In light of these findings, a series of alleles ranging from fully-penetrant to weak effect, from low frequency to high frequency, and from coding to regulatory, might lie within or near many genes or pathways relevant to diabetes mellitus. Although T2DM and monogenic diabetes mellitus might be caused by variants at extreme ends of this spectrum, elucidation of the full allelic series might make it impossible to establish a clear division between these classes of disease.

Variants in HNF1A provide an excellent example of this hypothesis. HNF1A encodes a master transcription factor (HNF-1A) that regulates genes expressed in the liver, kidney and pancreas. The protein structure of HNF-1A consists of distinct domains for dimerization, DNA binding and transactivation. Most mutations in HNF1A that are associated with MODY result in an altered DNA binding domain of the encoded protein, causing a near-complete impairment of HNF-1A transactivation activity (Fig. 3). By contrast, common coding variants within HNF1A (for example, encoding p.I27L and p.A98V) modestly increase risk of T2DM and have only a minimal effect on transactivation activity. In between these two extremes, a low-frequency HNF1A variant (encoding p.E508K) identified in Mexican and Latino populations has intermediate phenotypic and molecular effects.

Modest effects on T2DM risk have also been observed for variants that alter regulation of the transcriptional network involving HNF-1A. Variants in HNF-1A binding sites are enriched for T2DM associations, which suggests that perturbation of genes regulated by HNF-1A can cause less extreme forms of diabetes mellitus. In addition, noncoding common variants upstream of HNF1A have also been associated with T2DM, which suggests that regulation of HNF1A itself might also predispose to T2DM. Notably, common variants near HNF1A associated with other traits such as lipid levels or coronary artery disease are distinct from those associated with T2DM, which suggests that
multiple disease-relevant regulatory pathways might involve HNF1A. A range of possible molecular perturbations of HNF1A, as well as genes upstream or downstream of it within multiple regulatory networks, can, thus, result in a range of phenotypes relevant to diabetes mellitus.

On the basis of overlap between genes associated with monogenic diabetes mellitus and those associated with T2DM, multiple pathways relevant to diabetes mellitus could have properties similar to those observed for HNF1A. Three such pathways are highlighted in Fig. 4. One pathway consists of genes responsible for glucose-stimulated insulin secretion. Whereas carriers of severe mutations might develop NDM or MODY, carriers of less severe mutations might exhibit a T2DM subtype characterized primarily by insufficient glucose secretion, with potential response to sulfonylureas or progression to minimal complications (similar to patients with GCK mutations). A second pathway controls insulin signalling in muscle and fat. Whereas carriers of loss-of-function variants might develop severe insulin resistance or lipodystrophies, carriers of less severe mutations might exhibit a T2DM subtype characterized primarily by insulin resistance, defects in insulin response or partially degenerative conditions of adipose tissue. A final pathway consists of the transcriptional network of β cells, which regulates multiple genes, influences β-cell growth and differentiation, and affects internal metabolic processes. Whereas carriers of loss-of-function variants might develop MODY, carriers of less severe mutations might exhibit a T2DM subtype characterized primarily by inadequate transcriptional response or decreased β-cell mass and renewal.

Under this model, T2DM and monogenic diabetes mellitus would be better viewed as extreme points on a landscape of phenotypes, rather than fundamentally distinct conditions. Many phenotypes considered T2DM might in fact be diabetes mellitus subtypes caused by mutations in specific pathways. Such phenotypes might more so resemble monogenic disorders caused by large-effect mutations in the same pathway, rather than subtypes of T2DM due predominantly to mutations in different pathways.
Figure 4 | A unified model of diabetes mellitus risk. Increasingly, diabetes mellitus–associated monogenic, low-frequency and common variants cluster within not only individual genes but within pathways as well. Shown are subsets of three such pathways for which multiple genes carry a series of variants: a | glucose–stimulated insulin secretion (GSIS), b | insulin signalling and c | β–cell transcriptional regulation. Variants reported as associated with monogenic diabetes mellitus (red), type 2 diabetes mellitus (T2DM; blue: moderate effect, light blue: weak effect), glucose (yellow: moderate effect, light yellow: weak effect) or insulin (green: moderate effect, light green: weak effect). The gradients underneath each pathway illustrate the hypothesis that, rather than being distinct conditions, monogenic disorders caused by mutations in one of these pathways might in reality be extreme points on a phenotypic continuum. The current broad classification of T2DM might disguise disease subtypes characterized by presentations similar to those of monogenic forms, such as conditions of moderately impaired GSIS (variants in a), moderately impeded insulin signalling (variants in b), and reduced β–cell mass or function (variants in c). Potential implications of this model are discussed in TABLE 2. GLUT-2, glucose transporter type 2; MODY, maturity-onset diabetes of the young; NDM, neonatal diabetes mellitus.

Future directions

Links have been hypothesized between T2DM and monogenic diabetes mellitus for some time\(^4^{–}\text{10}\), helping in some cases to understand diabetes mellitus pathophysiology or treatment responses\(^4^{–}\text{10}\). Today, however, the diseases continue to be studied somewhat independently, within partially overlapping research communities and with different experimental approaches. A unified model of diabetes mellitus risk, encompassing both common and rare forms of the disease, could have implications for the study and treatment of each (TABLE 2).

Toward the goal of diagnosis and study of monogenic forms of diabetes mellitus, we see several areas that a unified diabetes mellitus risk model could influence. First, a unified model predicts that many pathways will harbour highly penetrant mutations for monogenic diabetes mellitus, and skilled analysis of whole-exome or whole-genome sequence data will, thus, be necessary to diagnose many families. Multiple mutations, some expected to be regulatory, will need to be analysed simultaneously. Analysis of causal mutations could assign higher likelihood to those present in genes implicated in T2DM, but incorporating this idea into diagnostics will require statistical sophistication. In some cases, families thought to harbour Mendelian diseases might need reconsideration for more complex inheritance patterns, possibly approaching those of earlier-onset forms of T2DM.
Second, large genetic or functional datasets should be increasingly used as essential resources to calibrate the frequencies and effects of variants observed from diagnostic sequencing. Public datasets such as the Exome Variant Server or the ExAC Browser have already been recommended for use as frequency filters on potential causal variants. However, under a unified risk model, many variants with appreciable frequencies in these datasets will in fact contribute to risk of diabetes mellitus, some with significant effect. Public datasets established for diabetes mellitus, such as those in the Type 2 Diabetes Genetics Portal, might be better incorporated into diagnostic procedures for diabetes mellitus, as they will enable a more accurate estimate of the association with diabetes mellitus. Ultimately, however, molecular or cellular assays might become necessary in many diagnostic settings, as they could be the only means to obtain sufficient precision in the estimate of functional and, hence, phenotypic effect. If so, the degree to which such assays reflect human physiology, measurable in part by correlations between assay readouts and phenotypic associations, will need to be communicated, and a balance between accuracy and turnaround time, addressed potentially by new techniques to screen possible mutations in advance of diagnosis.

Finally, standards for reporting variants that cause monogenic diabetes mellitus should be revisited. Some considerations apply in all clinical genetics settings, as they could be the only means to obtain sufficient precision in the estimate of functional and, hence, phenotypic effect. If so, the degree to which such assays reflect human physiology, measurable in part by correlations between assay readouts and phenotypic associations, will need to be communicated, and a balance between accuracy and turnaround time, addressed potentially by new techniques to screen possible mutations in advance of diagnosis, will need to be struck.

In addition to these considerations for monogenic forms of diabetes mellitus, we anticipate several aspects in which a unified risk model could also affect efforts to understand or treat the common disease of T2DM. In the near future, analysis of genes linked to monogenic diabetes mellitus might be the most profitable means to gain genetic insights into T2DM. Relative to the few rare variant associations identified by exome-wide and genome-wide analyses, the associations within monogenic diabetes mellitus genes have been striking. A unified model of diabetes mellitus risk would predict that an allelic series exists within many monogenic genes, and a more narrow focus on these genes might prioritise variants for study that stringent genome-wide significance thresholds might disregard. Logical candidates for study beyond monogenic genes might be those within the same biological pathway; the association of a common

<table>
<thead>
<tr>
<th>Area</th>
<th>Current practice</th>
<th>Future approaches?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Monogenic forms of diabetes mellitus</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sequencing patients with disease</td>
<td>Targeted gene panels</td>
<td>Whole-exome or whole-genome sequencing</td>
</tr>
<tr>
<td></td>
<td>Mendelian inheritance pattern</td>
<td>More complex inheritance patterns</td>
</tr>
<tr>
<td></td>
<td>Simple filters on individual variants</td>
<td>Multi-gene analysis, prioritise T2DM regions</td>
</tr>
<tr>
<td>Variant interpretation</td>
<td>Phenotype-agnostic or in-house ‘control’ datasets</td>
<td>Databases of samples ascertained for disease</td>
</tr>
<tr>
<td></td>
<td>Bioinformatic predictors and small-scale functional assays</td>
<td>High-throughput functional assay ‘look-up’ tables</td>
</tr>
<tr>
<td>Disease risk reporting</td>
<td>Qualitative likelihood of pathogenicity with ‘deterministic’ pathogenic effects</td>
<td>Quantitative estimates of both pathogenicity and risk</td>
</tr>
<tr>
<td></td>
<td>Risk for monogenic disease</td>
<td>Risk for spectrum of diabetes mellitus subtypes</td>
</tr>
<tr>
<td><strong>T2DM</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identification of novel genes or disease mechanisms</td>
<td>Unbiased genome-wide or exome-wide association studies</td>
<td>Prior likelihood of association for genes in monogenic pathways</td>
</tr>
<tr>
<td></td>
<td>Enrichment or aggregation tests within monogenic genes</td>
<td>Dissect and characterize signals within monogenic genes</td>
</tr>
<tr>
<td>Ascertainment of samples for genetic studies</td>
<td>Ascertain individuals with diabetes mellitus and matched controls</td>
<td>Ascertain based on a spectrum of clinical features</td>
</tr>
<tr>
<td></td>
<td>Exclude individuals with potential monogenic diabetes mellitus</td>
<td>Group samples or genes on the basis of clinical or functional similarity</td>
</tr>
<tr>
<td>Targeted medications for patients with T2DM</td>
<td>Little genetic information available to target treatments</td>
<td>MODY therapeutic targets with moderate-effect T2DM variants might be precision medicine candidates</td>
</tr>
</tbody>
</table>

Genetic evidence now supports an ‘allelic series’ across numerous genes and pathways involved in both rare and common forms of diabetes mellitus. A model in which these diseases are viewed as different subtypes of a continuous spectrum of disorders might have implications as to how we study, diagnose and treat diabetes mellitus. MODY, maturity-onset diabetes in the young; T2DM, type 2 diabetes mellitus.
Greenlandic variant within TBC1D4 (REF 80), a phospho-
rylation target of a gene (AKT2) in which variants cause
 dominant disorders of severe insulin resistance109 and
lipodystrophy110, is an example of how an allelic series can
span multiple genes in the same pathway.

Association studies for T2DM might also improve
power to detect association by stratifying patients accord-
ing to phenotypic subtypes. Under a unified model, anal-
yzing T2DM as a single phenotype might introduce
substantial genetic heterogeneity114, whereas stratifying
patients based on phenotypic similarity with rarer disor-
ders might reduce heterogeneity. One means to achieve
this stratification might be to measure characteristics
related to monogenic disease symptoms, such as renal
function (relevant to HNF1B–MODY), C-reactive pro-
tein levels (HNF1A–MODY), or body mass distribution
(relevant to lipodystrophies). Genes could also be strat-
ified for analysis on the basis of functional similarity to
those implicated in a monogenic disorder. For example,
adipocyte differentiation has proven an effective assay
to characterize variants in PPARG16, a gene implicated
in lipodystrophies. With the advent of new genome-
editing technologies such as CRISPR/Cas9 (REF 111),
unbiased genome-wide knockout screens could be per-
formed to identify additional genes relevant to adipocyte
differentiation for subsequent genetic studies.

Finally, medications used to treat monogenic dis-
orders might be explored for efficacy in subtypes of T2DM. For example, the p.E508K variant in HNF1-A,
which has a fivefold effect on T2DM risk, is present in
~2% of Mexicans with T2DM. Carriers of this muta-
tion might have a subtype of T2DM closely related to
HNF1A–MODY, a monogenic diabetes mellitus subtype
for which patients can show sensitivity to sulfonyl-
ureas112. Future research could assess whether screening
for the p.E508K variant could guide personalized treat-
ment for diabetes mellitus in the Mexican population.
As other low-frequency mutations are discovered in
other fairly well-understood disease pathways, these
mutations might offer additional candidates to inform
personalized medicine.

Conclusions
On the basis of these considerations, we believe that rec-
oning of knowledge of monogenic diabetes mellitus with
findings from T2DM association studies should be a high
priority in the diabetes mellitus research community.
Building an accurate, unified model of diabetes mellitus
risk would provide a valuable framework to organize our
current knowledge of diabetes mellitus pathophysiology,
enable future findings to be placed in a richer context
than is available today, and provide valuable ideas to
catalyse future research, treatments and diagnostics.
At a minimum, we anticipate that researchers who today
study solely common or rare forms of diabetes mellitus
will be well-served to be familiar with advances across
both fields, as data sharing and unified analysis will be
a fertile research area for the foreseeable future.
et al. (2012). Agarwala, V., Flannick, J., Sunyaev, S., GoT2D DIAbetes Genetics Replication Meta-analysis. Hemminki, K., Li, X., Sundquist, K. & Sundquist, J. Chakera, A. J. Sagen, J. V. Irgens, H. U. Haldorsen, I. S. (2012). in 3 common diseases. type 2 diabetes susceptibility loci. Reference 53, a large meta-analysis, provides evidence supporting a major role for common variants in genes encoding the pancreatic β-cell transcription factor binding site patterns: from human islet cell samples links regulatory signals from genome-wide association studies for type 2 diabetes and glyemic traits to their downstream effectors. van de Bunt, M. et al. Transcript expression data from human islet cell lines and metabolites signals reveal a major role for common variants in genes encoding the pancreatic β-cell transcription factor binding site patterns: from human islet cell samples links regulatory signals from genome-wide association studies for type 2 diabetes and glyemic traits to their downstream effectors. van de Bunt, M. et al. Transcript expression data from human islet cell lines and metabolites signals reveal a major role for common variants in genes encoding the pancreatic β-cell transcription factor binding site patterns: from human islet cell samples links regulatory signals from genome-wide association studies for type 2 diabetes and glyemic traits to their downstream effectors. van de Bunt, M. et al. Transcript expression data from human islet cell lines and metabolites signals reveal a major role for common variants in genes encoding the pancreatic β-cell transcription factor binding site patterns: from human islet cell samples links regulatory signals from genome-wide association studies for type 2 diabetes and glyemic traits to their downstream effectors. van de Bunt, M. et al. Transcript expression data from human islet cell lines and metabolites signals reveal a major role for common variants in genes encoding the pancreatic β-cell transcription factor binding site patterns: from human islet cell samples links regulatory signals from genome-wide association studies for type 2 diabetes and glyemic traits to their downstream effectors. van de Bunt, M. et al. Transcript expression data from human islet cell lines and metabolites signals reveal a major role for common variants in genes encoding the pancreatic β-cell transcription factor binding site patterns: from human islet cell samples links regulatory signals from genome-wide association studies for type 2 diabetes and glyemic traits to their downstream effectors. van de Bunt, M. et al. Transcript expression data from human islet cell lines and metabolites signals reveal a major role for common variants in genes encoding the pancreatic β-cell transcription factor binding site patterns: from human islet cell samples links regulatory signals from genome-wide association studies for type 2 diabetes and glyemic traits to their downstream effectors. van de Bunt, M. et al. Transcript expression data from human islet cell lines and metabolites signals reveal a major role for common variants in genes encoding the pancreatic β-cell transcription factor binding site patterns: from human islet cell samples links regulatory signals from genome-wide association studies for type 2 diabetes and glyemic traits to their downstream effectors. van de Bunt, M. et al. Transcript expression data from human islet cell lines and metabolites signals reveal a major role for common variants in genes encoding the pancreatic β-cell transcription factor binding site patterns: from human islet cell samples links regulatory signals from genome-wide association studies for type 2 diabetes and glyemic traits to their downstream effectors. van de Bunt, M. et al. Transcript expression data from human islet cell lines and metabolites signals reveal a major role for common variants in genes encoding the pancreatic β-cell transcription factor binding site patterns: from human islet cell samples links regulatory signals from genome-wide association studies for type 2 diabetes and glyemic traits to their downstream effectors. van de Bunt, M. et al. Transcript expression data from human islet cell lines and metabolites signals reveal a major role for common variants in genes encoding the pancreatic β-cell transcription factor binding site patterns: from human islet cell samples links regulatory signals from genome-wide association studies for type 2 diabetes and glyemic traits to their downstream effectors. van de Bunt, M. et al. Transcript expression data from human islet cell lines and metabolites signals reveal a major role for common variants in genes encoding the pancreatic β-cell transcription factor binding site patterns: from human islet cell samples links regulatory signals from genome-wide association studies for type 2 diabetes and glyemic traits to their downstream effectors. van de Bunt, M. et al. Transcript expression data from human islet cell lines and metabolites signals reveal a major role for common variants in genes encoding the pancreatic β-cell transcription factor binding site patterns: from human islet cell samples links regulatory signals from genome-wide association studies for type 2 diabetes and glyemic traits to their downstream effectors. van de Bunt, M. et al. Transcript expression data from human islet cell lines and metabolites signals reveal a major role for common variants in genes encoding the pancreatic β-cell transcription factor binding site patterns: from human islet cell samples links regulatory signals from genome-wide association studies for type 2 diabetes and glyemic traits to their downstream effectors. van de Bunt, M. et al. Transcript expression data from human islet cell lines and metabolites signals reveal a major role for common variants in genes encoding the pancreatic β-cell transcription factor binding site patterns: from human islet cell samples links regulatory signals from genome-wide association studies for type 2 diabetes and glyemic traits to their downstream effectors. van de Bunt, M. et al. Transcript expression data from human islet cell lines and metabolites signals reveal a major role for common variants in genes encoding the pancreatic β-cell transcription factor binding site patterns: from human islet cell samples links regulatory signals from genome-wide association studies for type 2 diabetes and glyemic traits to their downstream effectors. van de Bunt, M. et al. Transcript expression data from human islet cell lines and metabolites signals reveal a major role for common variants in genes encoding the pancreatic β-cell transcription factor binding site patterns: from human islet cell samples links regulatory signals from genome-wide association studies for type 2 diabetes and glyemic traits to their downstream effectors. van de Bunt, M. et al. Transcript expression data from human islet cell lines and metabolites signals reveal a major role for common variants in genes encoding the pancreatic β-cell transcription factor binding site patterns: from human islet cell samples links regulatory signals from genome-wide association studies for type 2 diabetes and glyemic traits to their downstream effectors. van de Bunt, M. et al. Transcript expression data from human islet cell lines and metabolites signals reveal a major role for common variants in genes encoding the pancreatic β-cell transcription factor binding site patterns: from human islet cell samples links regulatory signals from genome-wide association studies for type 2 diabetes and glyemic traits to their downstream effectors. van de Bunt, M. et al. Transcript expression data from human islet cell lines and metabolites signals reveal a major role for common variants in genes encoding the pancreatic β-cell transcription factor binding site patterns: from human islet cell samples links regulatory signals from genome-wide association studies for type 2 diabetes and glyemic traits to their downstream effectors. van de Bunt, M. et al. Transcript expression data from human islet cell lines and metabolites signals reveal a major role for common variants in genes encoding the pancreatic β-cell transcription factor binding site patterns: from human islet cell samples links regulatory signals from genome-wide association studies for type 2 diabetes and glyemic traits to their downstream effectors. van de Bunt, M. et al. Transcript expression data from human islet cell lines and metabolites signals reveal a major role for common variants in genes encoding the pancreatic β-cell transcription factor binding site patterns: from human islet cell samples links regulatory signals from genome-wide association studies for type 2 diabetes and glyemic traits to their downstream effectors. van de Bunt, M. et al. Transcript expression data from human islet cell lines and metabolites signals reveal a major role for common variants in genes encoding the pancreatic β-cell transcription factor binding site patterns: from human islet cell samples links regulatory signals from genome-wide association studies for type 2 diabetes and glyemic traits to their downstream effectors. van de Bunt, M. et al. Transcript expression data from human islet cell lines and metabolites signals reveal a major role for common variants in genes encoding the pancreatic β-cell transcription factor binding site patterns: from human islet cell samples links regulatory signals from genome-wide association studies for type 2 diabetes and glyemic traits to their downstream effectors. van de Bunt, M. et al. Transcript expression data from human islet cell lines and metabolites signals reveal a major role for common variants in genes encoding the pancreatic β-cell transcription factor binding site patterns: from human islet cell samples links regulatory signals from genome-wide association studies for type 2 diabetes and glyemic traits to their downstream effectors. van de Bunt, M. et al. Transcript expression data from human islet cell lines and metabolites signals reveal a major role for common variants in genes encoding the pancreatic β-cell transcription factor binding site patterns: from human islet cell samples links regulatory signals from genome-wide association studies for type 2 diabetes and glyemic traits to their downstream effectors. van de Bunt, M. et al. Transcript expression data from human islet cell lines and metabolites signals reveal a major role for common variants in genes encoding the pancreatic β-cell transcription factor binding site patterns: from human islet cell samples links regulatory signals from genome-wide association studies for type 2 diabetes and glyemic traits to their downstream effectors. van de Bunt, M. et al. Transcript expression data from human islet cell lines and metabolites signals reveal a major role for common variants in genes encoding the pancreatic β-cell transcription factor binding site patterns: from human islet cell samples links regulatory signals from genome-wide association studies for type 2 diabetes and glyemic traits to their downstream effectors. van de Bunt, M. et al. Transcript expression data from human islet cell lines and metabolites signals reveal a major role for common variants in genes encoding the pancreatic β-cell transcription factor binding site patterns: from human islet cell samples links regulatory signals from genome-wide association studies for type 2 diabetes and glyemic traits to their downstream effectors. von Degen, J. et al. Pharmacogenetic analysis of primary human pancreatic islets provides insights into type 2 diabetes susceptibility loci. Cell Metab. 12, 445–455 (2010). The Welcome Trust Case Control Consortium et al. Bayesian refinement of association signals for 14 loci in 3 common diseases. Nat. Genet. 44, 1294–1301 (2012).


Reference 114 is a highly cited discussion of potential explanations for the ‘missing heritability’ that was often debated after the first wave of GWAS.


Reference 119 is one of the most influential papers to demonstrate the power of protective loss-of-function mutations to suggest therapeutic targets.


Reference 123 is a thorough review of the history of genetic mapping.