#### REVIEW



## Clinical use of polygenic scores in type 2 diabetes: challenges and possibilities

Rashmi B. Prasad<sup>1,2</sup> · Liisa Hakaste<sup>2,3</sup> · Tiinamaija Tuomi<sup>1,2,3,4</sup>

Received: 25 October 2024 / Accepted: 30 January 2025 / Published online: 5 April 2025 © The Author(s) 2025

#### Abstract

Resulting from a combination of genetic and environmental factors, type 2 diabetes is highly heterogeneous in manifestation and disease progression, with the only common feature being chronic hyperglycaemia. In spite of vigorous efforts to elucidate the pathogenetic origins and natural course of the disease, there is still a lack of biomarkers and tools for prevention, disease stratification and treatment. Genome-wide association studies have reported over 1200 variants associated with type 2 diabetes, and the decreased cost of generating genetic data has facilitated the development of polygenic scores for estimating an individual's genetic disease risk based on combining effects from most—or all—genetic variants. In this review, we summarise the current knowledge on type 2 diabetes-related polygenic scores in different ancestries and outline their possible clinical role. We explore the potential applicability of type 2 diabetes polygenic scores to quantify genetic liability for prediction, screening and risk stratification. Given that most genetic risk loci are determined from populations of European origin while other ancestries are under-represented, we also discuss the challenges around their global applicability. To date, the potential for clinical utility of polygenic scores for type 2 diabetes is limited, with such scores outperformed by clinical measures. In the future, rather than predicting risk of type 2 diabetes, the value of polygenic scores may be in stratification of the severity of disease (risk for comorbidities) and treatment response, in addition to aiding in dissecting the pathophysiological mechanisms involved.

**Keywords** Ancestries · Comorbidities · Genetic risk · Mechanisms · Polygenic scores · Prediction · Review · Screening · Subtypes · Type 2 diabetes

#### Abbreviations

Artificial intelligence
Coronary artery disease
Genome-wide association study
Mild age-related diabetes

Rashmi B Prasad and Liisa Hakaste are joint first authors.

Rashmi B. Prasad rashmi.prasad@med.lu.se

- <sup>1</sup> Lund University Diabetes Centre, Department of Clinical Sciences, Genetics and Diabetes, CRC, Lund University, Malmö, Sweden
- <sup>2</sup> Institute for Molecular Medicine Finland (FIMM), Helsinki University, Helsinki, Finland
- <sup>3</sup> Folkhalsan Research Centre, Helsinki, Finland
- <sup>4</sup> Helsinki University Hospital, Abdominal Centre/Endocrinology, Helsinki, Finland

PGS	Polygenic score
pPS	Partitioned polygenic score
PXS	Polyexposure score
SIDD	Severe insulin-deficient diabetes
SIRD	Severe insulin-resistant diabetes
T2D-PGS	Type 2 diabetes polygenic score

#### Introduction

Type 2 diabetes is one of the most common diseases in the world, with diagnosis involving only chronic hyperglycaemia and exclusion of other specific causes [1]. It manifests as heterogeneous combinations of varying degrees and mechanisms of insulin deficiency and insulin resistance together with varying rates and types of comorbidities. Lifestyle factors have a significant impact on diabetes risk and progression but there is also a major genetic component [2, 3]. While recent breakthroughs



**Fig. 1** Clinical utility of polygenic scores: possibilities and challenges. (**a**) Creation of optimal polygenic scores for type 2 diabetes requires generation of large, accessible data sources covering diverse ancestries to identify transancestry as well as population-specific signals. Simultaneous method development to improve PGS performance in diverse ancestries will also be vital to facilitate opportunities for clinical application. (**b**) Integration of PGSs with clinical risk factors such as age, body composition and exposures (e.g. smoking) and evaluating the calibration of risk models will be important steps towards clinical implementation. (c) PGSs have potential for use in early prediction of type 2 diabetes and its more severe forms, subgrouping and evaluating response to treatment modalities. (d) Several challenges such as missing heritability (pertaining to both rare and common variants), gene–environment interactions, economics and phenotypic heterogeneity currently challenge the clinical implementation of T2D-PGSs. GWAS, genome-wide association study; T2D, type 2 diabetes. Created in BioRender. Prasad, R. (2025) https://BioRender.com/g39f279. This figure is available as part of a downloadable slideset

have identified over 1200 common and rare associated genetic variants [4, 5], several of which are associated with related metabolic traits [5-9], the low effect sizes of these variants mean that they have little clinical utility as predictive or diagnostic markers at individual level. However, combining the effects of several variants as polygenic scores (PGS) (also known as genetic risk scores or polygenic risk scores) may be useful. In the review, we discuss the clinical translation potential of type 2 diabetes PGSs (T2D-PGSs) and related PGSs for (1) risk prediction and screening of type 2 diabetes and its comorbidities and (2) risk stratification in clinical care, paying special attention to usability in populations from different ancestries. An outline of the review is presented in Fig. 1. An overview of type 2 diabetes risk stratification and classification is presented in Fig. 2.

🖉 Springer

## PGSs for prediction and screening of type 2 diabetes

Numerous risk calculators have been developed for the early detection of type 2 diabetes [10–12], as it can remain asymptomatic for years despite the development of comorbidities [10, 13]. Typically, risk calculators are based on traditional risk factors such as age, sex, ethnicity, history of hypertension, BMI, waist circumference and family history of diabetes (Fig. 2). Some risk calculators also consider fasting plasma glucose and HbA<sub>1c</sub>. Usually, screening is considered in adults of any age with overweight or obesity and in all individuals aged  $\geq$ 45 years, as BMI is the strongest risk factor for type 2 diabetes [14]. Performance metrics used for prediction,

screening and diagnostic models are summarised in the text box [15, 16].

# Performance metrics for prediction, screening and diagnostic models

A receiver operator characteristic (ROC) curve is generated by analysing the trade-off between the sensitivity and the specificity of a test, typically to study the combined effect of included variants and their predictive value.

The **area under the curve (AUC)** (also known as C-statistic or C-index) provides a measure of the proportion of times such a test will correctly assign type 2 diabetes status between a pair of individuals, one who has the disease and another who does not. The AUC, while providing an overall performance measure, may not be clinically useful for screening, as it does not focus on specific thresholds relevant to decision-making.

Reporting the **detection rate (sensitivity)** at a sensible false-positive rate is more practical. Additionally, the incidence or prevalence of the condition impacts the interpretation of positive results, particularly affecting the positive predictive value (PPV) [15, 16].

The high degree of type 2 diabetes heritability (69%) among European populations aged 35-60 years [17] has led to expectations that adding T2D-PGSs to models would considerably improve type 2 diabetes risk discrimination. However, the results of such studies have been modest. The largest T2D-PGS to date is a metaPGS based on summary statistics from 44 genome-wide association studies (GWAS) performed in both European and non-European genetic ancestries for type 2 diabetes and its risk factors. Compared with the QDiabetes risk prediction score [12] comprising traditional risk factors, the metaPGS had a larger C-index for 10 year type 2 diabetes risk prediction (0.716; 95% CI 0.708, 0.723) than all individual traditional risk factors, including family history (C-index: 0.687; 95% CI 0.679, 0.695), except for BMI (C-index: 0.780; 95% CI 0.773, 0.787) and HbA<sub>1c</sub> (C-index: 0.826; 95% CI 0.819, 0.833) [18]. In addition, when incorporating the metaPGS into absolute risk predictions made by QDiabetes risk scores, the metaPGS significantly improved the risk stratification of QDiabetes, increasing the number of correctly classified future incident type 2 diabetes cases.

In some situations, such as in younger individuals and individuals without obesity, T2D-PGSs have clear advantages. While the performance of the clinical models varies with age and adiposity [10], the genetic variants remain constant [19], allowing for risk prediction either at an early age before the typical risk factors manifest, or in normal-weight individuals. It is especially important to identify individuals with younger-onset type 2 diabetes, as they are prone to more severe forms of disease and higher mortality rates [20, 21].

Among many populations, especially in East and South-East Asia, a considerable proportion (up to 60%) of people with type 2 diabetes are of normal weight or lean. Even in Western countries, up to 25% of individuals with type 2 diabetes have a normal BMI [22]. Overall, T2D-PGSs have revealed differences in both the relative and the absolute risk of type 2 diabetes among individuals in all categories of BMI, with the risk of incident type 2 diabetes being considerably higher in individuals with the highest T2D-PGS in all BMI categories in both men and women [23, 24]. The genetically determined risk appears to be higher for individuals at the lower end of each BMI range than for those at the higher end (Tables 1, 2, 3 and 4) [23–25], pointing to a stronger genetic risk or specific pathophysiology in those without the classical type 2 diabetes phenotype. In a Chinese population, individuals with normal weight showed higher partitioned polygenic scores (pPSs; see text box: Types of polygenic scores) for beta cell dysfunction and lipodystrophy than those with overweight, while the pPSs for obesity were associated with faster progression to clinical requirement of insulin treatment [26]. Furthermore, in a separate study of a US population, each SD increase in a T2D-PGS was associated with a decrease in age at diagnosis of type 2 diabetes by 1.3 years. In addition, in individuals without type 2 diabetes, an elevated T2D-PGS increased the odds of reported high blood sugar by 23% and type 2 diabetes by 43% within a year [27].

It has been suggested that family history of diabetes should be used for prediction of type 2 diabetes instead of T2D-PGSs. However, family history and PGSs are independent and not interchangeable measures. A PGS provides complementary information on inherited disease susceptibility [28], especially for those without a known first-degree family history of diabetes (Tables 1, 2, 3 and 4) [23]. Moreover, not all people know their family history, and lack of a family history of diabetes may also be a consequence of the healthier lifestyle of parents



**Fig. 2** Type 2 diabetes: risk stratification and classification. Current prediction and screening methods for type 2 diabetes, diabetes sub-types/medication response and comorbidities are based on traditional clinical risk factors. Adding PGSs to current practices as one risk factor among others may aid in the detection of individuals at risk of type 2 diabetes and associated complications at early stages of disease development and allow for better subclassification of disease. Examples of PGS models assessed for prediction of susceptibility to disease and comorbidites include those for risk factors such as obesity and coronary artery disease. PGSs also provide clues to the underlying pathophysiology of diabetes subtypes defined by both clinical

variable-derived and genetic clustering. Ahlqvist et al [92] defined five clinical variable-derived clusters. Suzuki et al [4] reported eight genetic clusters using a hard clustering approach, allowing for one SNP to belong to only one cluster. Smith et al [90] identified 12 clusters using a soft clustering approach whereby a SNP can associate with more than one cluster. ALP, alkaline phosphatase; HFpEF, heart failure with preserved ejection fraction; LpA, lipoprotein A; MARD, mild age-related diabetes; MOD, mild obesity-related diabetes; SAID, severe autoimmune diabetes; SIBG, sex hormone-binding globulin; SIDD, severe insulin-deficient diabetes; SIRD, severe insulin-resistant diabetes. This figure is available as part of a downloadable slideset

Table 1	Heritability potential of	various strata of risk factor	s assessed using heritab	oility h2 (LDSC regression) [24]
---------	---------------------------	-------------------------------	--------------------------	----------------------------------

Study population	Groups/model	Heritability h2 <sup>a</sup> (LDSC regression)
BBJ <i>n</i> (T2D)=27,642, <i>n</i> (control)=70,242	Unstratified	0.145
UKBB n(T2D)=27,642, n(control)=70,242		0.131
BBJ <i>n</i> (T2D)=13,821, <i>n</i> (control)=35,121	BMI stratified, two groups	+84.5% in lowest vs highest BMI group (0.198 vs 0.107)
UKBB n(T2D)=13,821, n(control)=35,121	BMI stratified, two groups	+64.7% in lowest vs highest BMI group (0.152 vs 0.092)
BBJ <i>n</i> (T2D)=9214, <i>n</i> (control)=23,414	BMI stratified, three groups	Non-significant difference between lowest and highest BMI groups, but a similar trend as in the BMI-strati- fied datasets separated into two groups
UKBB <i>n</i> (T2D)=9214, <i>n</i> (control)=23,414	BMI stratified, three groups	Non-significant difference between lowest and highest BMI groups, but a similar trend as in the BMI-strati- fied datasets separated into two groups

<sup>a</sup>h2 is broad-sense heritability that describes the contribution of genetic variation to the variation in a given trait

BBJ, BioBank Japan; LDSC, linkage disequilibrium score; T2D, type 2 diabetes; UKBB, UK Biobank

Table 2	Dradiation r	otontial of	1100000000	atrata of	might footone	accord	maina	liobility	$D^2$	F <b>^</b> / 1
I a pie z	Prediction L	Jolenniai oi	various	suata or	TISK TACIOIS	assessed	using	парши	Λ	1241

Groups/model	Liability $R^2$ (pseudo- $R^2$ value)				
Unstratified	0.072				
Unstratified	0.051				
BMI stratified, two groups	+22% in lowest BMI group vs unstratified (0.088 vs 0.072)				
BMI stratified, two groups	+23.5% in lowest BMI group vs unstratified (0.063 vs 0.051)				
BMI stratified, three groups	+30.6% in lowest BMI group vs unstratified (0.094 vs 0.072)				
BMI stratified, three groups	+21.6% in lowest BMI group vs unstratified (0.062 vs 0.051)				
BMI stratified, two groups	+60% in lowest vs highest BMI group (0.088 vs 0.055)				
BMI stratified, two groups	+57.5% in lowest vs highest BMI group (0.063 vs 0.040)				
BMI stratified, three groups	+100% in lowest vs highest BMI group (0.094 vs 0.047)				
BMI stratified, three groups	+72.2% in lowest vs highest BMI group (0.062 vs 0.036)				
	Groups/model Unstratified Unstratified BMI stratified, two groups BMI stratified, two groups BMI stratified, three groups BMI stratified, three groups BMI stratified, two groups BMI stratified, two groups BMI stratified, two groups BMI stratified, three groups BMI stratified, three groups				

BBJ, BioBank Japan; T2D, type 2 diabetes; UKBB, UK Biobank

Table 3 Prediction potential of various strata of risk factors assessed using AUC [25]

Study population	Groups/model	AUC (95% CI)
UKBB <i>n</i> (total)=121,113 (2544 incident T2D)	GRIT-T2D: age, sex, T2D-PGS, antihypertensives, statin, CVD, FH, smoking status, gestational diabetes	0.810 (0.803, 0.818)
	GRIT-T2D+: GRIT-T2D + TG, HDL, systolic BP	0.827 (0.820, 0.834)
Stratified by age: $<55$ vs $\geq 55$ years		
UKBB <i>n</i> (total)=121,113 (2544 incident T2D)	GRIT-T2D age <55 years	0.846 (0.830, 0.861)
	GRIT-T2D age ≥55 years	0.780 (0.770, 0.789)
	GRIT-T2D+ age <55 years	0.862 (0.847, 0.877)
	GRIT-T2D+ age ≥55 years	0.798 (0.789, 0.807)
Stratified by BMI: $<30 \text{ vs} \ge 30 \text{ kg/m}^2$		
UKBB n(total)=121,113 (2544 incident T2D)	GRIT-T2D BMI <30 kg/m <sup>2</sup>	0.786 (0.773, 0.799)
	GRIT-T2D BMI ≥30 kg/m <sup>2</sup>	0.708 (0.695, 0.721)
	GRIT-T2D+ BMI <30 kg/m <sup>2</sup>	0.808 (0.795, 0.820)
	GRIT-T2D+ BMI $\geq$ 30 kg/m <sup>2</sup>	0.734 (0.721, 0.746)

FH, family history of diabetes; GRIT-T2D, Genomics-enhanced Risk Tool; TG, triacylglycerol; T2D, type 2 diabetes; UKBB, UK Biobank

Table 4	Prediction potential	of various strata of risk factors assess	ed using T2D-PGS HR	for incident type 2 diabetes 23
	1		0	<b>1</b>

Study population	Groups/model	T2D-PGS HR (95% CI) <sup>a</sup>
Stratified by BMI		
UKBB <i>n</i> (total)=431,658 (17,259 incident T2D)	Normal BMI <sup>b</sup> . Model: age + sex + BMI + FH + genetic array + first four genetic principal components + PGS × BMI	2.21 (1.92, 2.56)
	Overweight <sup>c</sup> . Model: age + sex + BMI + FH + genetic array + first four genetic principal components + PGS × BMI	2.19 (2.04, 2.35)
	Obese <sup>d</sup> . Model: age + sex + BMI + FH + genetic array + first four genetic principal components + PGS × BMI	1.80 (1.70, 1.91)
Stratified by first-degree FH		
UKBB <i>n</i> (total)=431,658 (17,259 incident T2D)	First-degree FH. Model: age + sex + BMI + FH + genetic array + first four genetic principal components + PGS × FH	1.81 (1.68, 1.95)
	No first-degree FH <sup>b</sup> . Model: age + sex + BMI + FH + genetic array + first four genetic principal components + PGS × FH	2.06 (1.96, 2.17)

<sup>a</sup>Cox proportional hazards model; highest quintile vs middle PGS quintiles

<sup>b</sup>BMI  $\geq$ 18.5 to <25 kg/m<sup>2</sup>

<sup>c</sup>BMI  $\geq$ 25 to <30 kg/m<sup>2</sup>

p<0.001 for both comparisons (BMI strata and first-degree FH strata)

FH, family history of diabetes; T2D, type 2 diabetes; UKBB, UK Biobank

 $<sup>^{</sup>d}BMI \ge 30 \text{ kg/m}^2$ 

#### Types of polygenic scores

#### Restricted-to-significant polygenic score (rsPS)

This score is composed of variants showing a statistically significant association (generally Bonferroni corrected  $p < 10^{-8}$ ) in a genome-wide association study (GWAS) wherein each variant's contribution is weighted by its effect size (also called weighted PGS). Hence, individuals who carry several risk alleles have an increased risk even when the individual effect sizes are minute. This is the standard approach to quantify liability in disease risk prediction [5, 6, 10].

#### Global extended polygenic score (gePS)

These scores are calculated using either variants at specific *p* value thresholds or all SNPs from a GWAS, weighted by their effect and sample sizes. As GWAS often show only marginal associations for most causal variants, including the maximum number of potential risk variants may improve the performance of a PGS. This approach assumes that (1) several true risk variants that might normally escape detection due to limitations in the sample size will be included in the PGS and (2) non-associated variants will cancel each other out by random fluctuations across the null distribution [51].

#### Partitioned polygenic score (pPS)

A clustering approach incorporating orthogonal lines of evidence, grouping type 2 diabetes variants into specific clusters that reflect their underlying biology [83, 87–89].

protecting against diabetes in their offspring, despite them having a high genetic risk that may have been passed to their offspring. Of note, individuals with a high risk of diabetes often also have a high risk of CVD, which may lead to premature death before the diagnosis of diabetes. Another point has been that knowledge of one's own genetic risk can cause anxiety. This can be counteracted by information. Obviously, individuals with a high genetic risk are not destined to develop type 2 diabetes, and protective lifestyle changes have been shown to attenuate the risk of diabetes in those with a high T2D-PGS [29, 30]. On the other hand, as a high T2D-PGS itself has been associated with unhealthy dietary and physical activity habits [31], genetic information may encourage the adoption of a healthier lifestyle. Indeed, receiving personal risk data for CVD (including polygenic risk) motivated positive health behaviour changes and healthcare contacts among a cohort of over 7000 individuals with a mean age of 56 years, which supports the integration of genomic information into clinical risk calculations [32].

## PGSs for prediction and screening of comorbidities

As the individual and financial burdens of type 2 diabetes largely derive from symptoms, screening and treatment of vascular comorbidities (nephropathy, retinopathy, neuropathy, coronary and peripheral artery disease, cerebrovascular disease), it is of utmost importance to identify individuals at high risk of these comorbidities [33]. Risk calculators based on clinical factors have suboptimal predictive power [34–36], and similar screening practices are usually applied for all individuals. The predictions rely heavily on the presence of hyperglycaemia, which is a strong predisposing factor [37], but optimal blood glucose levels do not guarantee protection against comorbidities. Also, screening for diabetic nephropathy relies on existing signs of kidney damage: albuminuria or a decline in GFR [38]. PGS for nephropathy may help stratify risk at earlier stages as suggested in a recent systematic review [39].

T2D-PGSs as such are associated with increased risks of retinopathy, kidney disease, peripheral artery disease, neuropathy and coronary artery disease (CAD), pointing to genomic pathways that link type 2 diabetes to vascular outcomes [7], but there are few studies on PGSs designed for comorbidities of diabetes (Fig. 2). A study among 6079 individuals with type 2 diabetes of European, Hispanic, African and other ancestries showed that individuals in the top PGS decile for retinopathy had retinopathy 1.8 times more often and earlier than those in the bottom decile [40]. PGSs for type 2 diabetes and CAD independently predict future cardiovascular mortality risk [41], and a large biobank study indicated that PGSs for CAD had potential for clinical utility, at least for those of European ancestry [42]. However, the use of PGSs as an addition to clinical risk models has also been challenged with regard to CAD, stroke and heart failure [43]. Given the differences in clinical presentation of CAD and stroke in people with and without diabetes (the former having more extensive disease also affecting the small arteries), it can be speculated that a general CAD-PGS might perform less well than a T2D-specific CAD-PGS in individuals with type 2 diabetes. However, at least among individuals of European and South Asian descent, no evidence of different genetic architecture of CAD was noted between those with type 2 diabetes and those without [44].

## PGSs for predicting response to glucose-lowering medications

It would be clinically useful to employ PGSs to predict individual responses to glucose-lowering medications; however, data in this area are limited [45]. To date, instead of using partitioned scores, studies have used only known loci for type 2 diabetes and related traits. Of note, in addition to glycaemic response, it would also be important to look for any associations with side effects.

Metformin is the most common initial treatment for type 2 diabetes; however, many individuals do not achieve adequate glycaemic control on metformin, resulting in delays in commencement of other therapies. It has been suggested that the metformin response in type 2 diabetes is associated with variants in, for example, SLC22A1, ATM and SLC2A2; however, the replicability of the SLC22A1 locus associations has been inconsistent, potentially owing to differences between studies in study design, population characteristics or disease stage [46-50]. Interestingly, individuals with a high global extended polygenic score [51] (gePS; see text box: Types of polygenic scores) for fasting glucose had a reduced glucose response to metformin in the Study to Understand the Genetics of the Acute Response to Metformin and Glipizide in Humans (SUGAR-MGH) [52], suggesting that PGSs may aid in choosing the primary treatment for type 2 diabetes.

Response to sulfonylureas may also be heritable. Individuals with a high restricted-to-significant polygenic score (rsPS; see text box: Types of polygenic scores) for type 2 diabetes had greater acute and sustained responses to sulfonylureas in the SUGAR-MGH study, and many variants (e.g. in *CYP2C9*, *KCNJ11*, *TCF7L2*, *GXYLT1* and *SLCO1B1*) were also independently associated with this trait [53]. Other studies have reported the association of *GLP1R* and *ARBB1* variants with response to glucagon-like peptide-1 receptor agonists [54], *GLP1R* and *DPP4* variants with response to dipeptidyl peptidase-4 inhibitors [55, 56], and *SLC5A2* variants with response to sodium–glucose cotransporter 2 inhibitors [57].

T2D-PGSs have also been associated with insulin treatment (as a proxy for advanced disease) in diverse populations [27, 58]. Among a South Asian population, populationspecific pharmacogenetic variant profiles were marked by an excess of alleles associated with poor treatment response to various non-insulin glucose-lowering drug classes. This calls for further pharmacogenetic studies in multiple ancestries and reconsideration of dosage recommendations for glucose-lowering medications to ensure optimal efficacy and safety [59].

#### PGSs for type 2 diabetes in different ancestries

As GWAS findings have largely been based on populations of European ancestry, it is unsurprising that PGSs often fail to predict disease risk in other populations. The genetic architecture of populations varies markedly (e.g. effect sizes, allele frequencies and patterns of linkage disequilibrium), as do the genetic associations with diseases [60]. In recent years, there has been a rapid increase in the inclusion of participants from diverse ancestral backgrounds, and several type 2 diabetes GWAS have been performed on non-European populations [9, 61, 62]. Indeed, the latest GWAS of ~2.5 million participants (17% with type 2 diabetes) included 40% of individuals with non-European ancestry and identified 1289 independent signals, of which 46% were attributed to inclusion of previously under-represented ancestry groups [4]. While the majority of the discovered loci have similar effects across ancestries, ancestry-specific loci have also been identified [5], such as loci in SLC16A11 and SLC16A13 in Latin American individuals [63], TMEM163 in Asian Indian individuals [64], DNER in American Indian individuals [65], SCTR, GP2 and ZNF257 in Japanese individuals [66], ZRANB3, AGMO, ANKH, INS-IGF2-KCNQ1, TGFB1 and AGMO in African individuals [4, 5, 67–69], and UBE2E2, PAX4, KLF14, ANK1 and INS in East Asian individuals [4]. Adding ancestry-specific loci to European T2D-PGSs outperformed the European T2D-PGSs alone in different populations [62, 70, 71]. For instance, adding ancestry-specific loci to European T2D-PGSs boosted type 2 diabetes prediction in a population from continental Africa [71-73]. and ancestry-specific PGSs outperformed European-based PGSs in South Asian populations, with a more than 20-fold higher type 2 diabetes risk predictability among individuals in the top (ninth) compared with the middle (fifth) decile, showing the sensitivity and effectiveness of the PGS models even at the lower extremes of the distribution [74]. In another study, T2D-PGS based on South Asian individuals showed an approximately fourfold higher risk between the top and the bottom quartiles [75]. Polfus et al showed that, in large sample sizes and using multiancestry weights, both of which are more likely to accurately reflect the true causal effect of a variant, a multiancestry PGS outperformed a population-specific PGS [67].

Approaches to increase the predictive potential of PGSs could include adding rare ancestry-specific variants to the common variant-based PGSs and considering that associations may indicate mechanistic pleiotropy. For example, in a GWAS on HbA<sub>1c</sub> levels in European, East Asian and South Asian populations, some variants predicted type 2 diabetes risk whereas others influencing HbA<sub>1c</sub> through erythrocytic pathways did not [76]. A dominant effect of the erythrocytic *G6PD* variant on type 2 diabetes was seen only in individuals with African American ancestry [76]. These approaches were combined to construct PGSs for HbA<sub>1c</sub> based on a selection of rare and common variants, and including variants in genes with known erythrocytic roles. Including 22 common and 21,293 rare (across 154)

genes; 73% observed in fewer than three people) variants augmented the diagnostic potential compared with PGSs including only common variants [77]; adding other dimensions such as metabolite data further enhanced type 2 diabetes risk prediction [78].

A type 2 diabetes predictive model including a PGS, physical measurements and clinical risk factors increased the prediction performance compared with models without the PGS, and predicted conversion from normal glucose tolerance to prediabetes to type 2 diabetes, in 5490 Korean individuals [58]. A polyexposure score (PXS) combining multiple lifestyle and exposure factors showed modest improvement in predicting risk over clinical factors and PGSs in White British individuals from the UK Biobank [73]. Applying a similar PXS in addition to clinical factors and PGSs may also improve prediction in non-European populations; however, the variability in exposures between populations and the gene–environment interactions may add to the complexity in transferring European-derived GWAS scores across ancestries.

#### Potential challenges around clinical implementation of PGSs in diverse ancestries

The clinical implementation of PGSs in diverse ancestries faces several challenges (Fig. 3). While the inclusion of participants from different ancestral backgrounds has increased, and European-derived scores may be applicable to other ancestries, the genetic background of type 2 diabetes in non-European populations is still underexplored. There is some promise in the improving, albeit more complex, methodology for calculating multiple-ancestry PGSs, which leverages machine learning/artificial intelligence (AI) and reported summary statistics [79, 80]. The contribution of known type 2 diabetes genetic signals to heritability in non-European populations remains unclear, and ancestryspecific variants are yet to be identified, complicated by accessibility and economic barriers. Phenotypic differences across ethnicities also affect risk variations. For example, BMI cut-offs for obesity ( $\geq 30 \text{ kg/m}^2$ ) were developed in European populations but may not be appropriate for other groups; in this regard, Asian populations exhibit higher type 2 diabetes prevalence at lower BMI and younger age, highlighting the need to consider population-specific cut-offs [81]. Moreover, the genetic predisposition captured by PGSs may render an individual's risk assessment incomplete, as it may not fully account for the influence of early life factors and diverse environments, as well as their interactions with type 2 diabetes genetic loci, complicating PGS development and utility [4, 82].



Fig. 3 Challenges around clinical implementation of PGSs in understudied/diverse populations. Insufficient knowledge of ancestry-specific signals for common and rare variants can decrease their prediction potential. Inadequate information on early life factors, such as birthweight, maternal nutrition, early growth patterns and childhood obesity, can interact with genetic susceptibility to modulate T2D risk. This can enhance risk stratification and provide information on gene-environment interactions. Vast variations in environmental conditions, including cultural diversity, and economic challenges such as lack of access to infrastructure can also present significant challenges for the clinical implantation of T2D-PGSs. Moreover, the wide heterogeneity in disease manifestation (e.g. lean type 2 diabetes in South Asian populations) and underlying genetic architecture, and differences in phenotypic measures compared with European populations, renders the application of European-derived PGSs especially challenging. This figure is available as part of a downloadable slideset

#### Using PGSs to unravel the heterogeneity of type 2 diabetes

Hyperglycaemia may be driven by various mechanisms that directly or indirectly affect insulin secretion and action [83]. Most individuals are likely to have mixed phenotypes arising from the contribution of several pathways, which can be difficult to deduce from clinical measurements [72]. Variants associated with type 2 diabetes may reflect these processes independently or through pleiotropic effects. Different approaches have been used to partition diabetes subtypes using clinical and/or genetic data, but the tentative subtypes are not yet ready to be used in the clinic. More advanced methods using AI on detailed phenotypes together with genetic information may allow for better subclassification and identification of individuals with high risk for the disease and development of complications [62, 84, 85].

The PGS approach Using machine learning, known variants have been grouped into 'clusters' or pPSs (see text box: Types of polygenic scores), representing different biological pathways (Figs 1 and 2). Each individual receives a score for each cluster, giving a more versatile view of their genetic risk. Applying pPSs to previously reported type 2 diabetes loci, six clusters were initially defined for impaired lipid metabolism, insulin secretion 1 and 2, adiposity, insulin action and insulin secretion/action [86]. Also using 'hard clustering', Suzuki et al defined eight clusters in the multiancestry GWAS comprising 40% non-European ancestry participants and allowed individuals to belong to only one cluster (Fig. 2). These clusters were characterised by distinct profiles of cardiometabolic trait associations; the obesity and lipodystrophy clusters associated with CAD, with the former also associating with peripheral artery disease and end-stage diabetic nephropathy (Fig. 2) [4]. Ancestry-related differences were observed: allelic effects were strongest in East Asian individuals for beta cell dysfunction clusters, and in European individuals for insulin resistance clusters. After accounting for BMI variation across ancestry groups, significant differences in association signals persisted between African and European cohorts, but not between African and East Asians cohorts [4]. In addition, the heterogeneity in effects for beta cell dysfunction clusters seemed to be accounted for by BMI differences between East Asian, European and African cohorts [4].

To allow for a particular variant to be associated with several clusters, Udler et al applied 'soft clustering' to 95 type 2 diabetes loci (associated with 47 diabetes-related traits) and identified five robust clusters, two related to insulin secretion ('beta cell' and 'proinsulin') and three to insulin response or action ('obesity', 'lipodystrophy' and 'liver/lipid') [87]. The obesity and lipodystrophy clusters were associated with increased blood pressure and hypertension, and the latter was also associated with CAD. The liver/lipid cluster was associated with decreased CAD and reduced renal function, suggesting a shared pathway [88, 89], perhaps via insulin resistance. Further clustering of 650 type 2 diabetes variants with 110 trait associations identified 12 clusters (recapturing the previously defined five clusters) [90] (Fig. 2). Applying this multiancestry pPS to British South Asian individuals revealed that genetic predisposition to insulin deficiency and lipodystrophy was linked to earlier onset of type 2 diabetes, more rapid progression to complications, insulin dependence and diminished response to medication [91].

An advantage of this approach is the consistent categorisation of genetically driven propensity to malfunction of certain molecular pathways and propensity for related metabolic disease outcomes. However, limitations exist. The assignment of the variants to certain pathways and clusters is based on limited numbers of individuals with deep phenotyping data. As the clusters are defined by allele distributions and scores are continuous, the boundaries can be unclear. Also, while heritability of type 2 diabetes is at most 69% and the known risk variants explain only about 50% of the heritability, these genetic clusters are not comprehensive. Adding more variants may facilitate the identification of more clusters with several possible permutations and combinations, while also allowing for refining for more precise definitions. Population-specific scores and environmental contributions will need to be taken into consideration for clinical applicability.

The clinical variable-based approach This approach involves using clinical variables to categorise individuals. Datadriven machine learning performed on commonly available variables (age at diagnosis, sex, BMI, HbA<sub>1c</sub>) and HOMA2-B and HOMA2-IR in newly diagnosed individuals with diabetes from southern Sweden identified five reproducible clusters: severe autoimmune (SAID), insulin-deficient (SIDD) and insulin-resistant (SIRD) diabetes and mild obesity-related (MOD) and age-related (MARD) diabetes (Fig. 2) [92]. The groups (replicated in several populations [81, 93]) differ regarding clinical features, complication risk and disease progression [92]. Despite partially different genetic associations, for example SIRD uniquely associated with a fasting insulin PGS, while most tested PGSs associated with more than one cluster [94], no subgroup-specific PGSs are available. When these subgroups were applied to Indian cohorts, the PGS associations were partially replicated, but some were unique to this population. For instance, association with a liver-related PGS was seen in India but not Europe and was related to poor liver growth in utero and other early life factors that may contribute to the thin-fat phenotype [70, 95]. However, in another study from India, inclusion of other variables resulted in two similar (SIDD and MARD) and two new (combined insulin-resistant and -deficient diabetes [CIRDD] and insulin-resistant obese diabetes [IROD]) clusters [96]. Integrating a PGS for insulin secretion or resistance (without actual measurements) with an existing risk tool in over 22,000 British Pakistani and Bangladeshi individuals still identified a probable severe insulin-deficient diabetes (pSIDD) subgroup, underscoring the robustness of these associations [97]. While the clinical variable-based approach seems less robust in populations of diverse ancestry, this could be attributed to differences in distribution (e.g. body composition) and genetic factors limiting the utility of PGSs in this context. Overall, this highlights the need for further refinement of clusters and inclusion of other important clinical variables, and the need to take into consideration parental and early life effects and population-specific definitions [70].

**Combined polygenic and phenotype approach** Combining genetics with phenotypic information derived from OGTTs, body fat distribution and liver fat content [98] resulted in six clusters of prediabetes characterised by low or very low type 2 diabetes risk, beta cell failure, low risk + obesity, high risk + insulin resistance + fatty liver, and high risk + visceral fat + nephropathy. This subgrouping targets underlying pathogenic defects to refine the stratification, but extension to non-European populations is still needed.

#### **Future perspectives**

The share of heritability of type 2 diabetes that can be explained by identified variants has increased from 25% in the first GWAS decade (2007-2016) to about 50% currently. However, this means that at least half of the heritability is unexplained. In addition, type 2 diabetes risk loci have been shown to have sex-specific and parent-of-origin effects, which need to be taken into consideration to improve prediction. Moreover, the effects of gene-gene and gene-environment interactions are largely unknown. Future research and more affordable whole-genome sequencing will likely expand the numbers of common and rare genetic loci as well as ancestry-specific signals associated with diabetes, with particular efforts needed to identify ancestry-specific signals before clinical implementation can be tested. Examining the distribution of PGSs in different populations may then, perhaps in combination with clinical risk factors, allow for identification of groups who are at high risk of developing different subtypes of 'type 2 diabetes' and, especially, comorbidities. This approach could be piloted in a clinical setting to enable targeted interventions and lifestyle advice to be provided, to reduce the risk of developing the disease. Most importantly, it is hoped that PGSs could be used to identify individuals at risk of more aggressive disease forms to receive stringent monitoring and treatment, including targeted initiation of medication to prevent the development and progression of comorbidities. However, before that is feasible, further understanding is needed of the molecular mechanisms, biological functions and clinical implications of these loci.

The rapidly expanding healthcare-associated biobanks and direct-to-consumer genetic testing will facilitate not only gathering of genotype-associated follow-up data for risk evaluation, but also population-based implementation of PGSs. As technological advances render low-coverage whole-genome sequencing (which can be used to calculate PGSs and identify monogenic variants) less expensive, PGS construction will become increasingly cost-effective for preventive health approaches, particularly when performed early in life. The potential for applicability of PGSs in clinics worldwide will increase with more complete profiles of genetic risk variants across diverse populations. At the same time, pipelines for constructing PGSs (and updating them based on new knowledge), as well as interpreting them to aid clinicians and patients, are needed, together with frameworks connecting them to medical records.

To date, the accuracy of PGSs for type 2 diabetes is limited by our current understanding of the genetic causes of the disease. At best, they may improve our understanding of the underlying biology of type 2 diabetes and lead to the development of more effective preventive and therapeutic strategies. The variant-based clustering approaches represent a forerunner in this field. However, even if the soft clustering approach allows same variants to be included in different clusters, we need ways of analysing each individual's profiles for multiple clusters and how they interact in providing the individual's phenotype and affect risk estimates for comorbidities. This is likely to need solutions based on AI.

Although current knowledge cannot support the use of PGSs for diagnostic purposes, that is, to prove that an individual has type 2 diabetes, they could be used for elimination purposes, that is, to raise awareness of a high likelihood that an individual has a different type of diabetes and provide an incentive for screening for monogenic diabetes.

#### Conclusion

Although high-throughput genomic studies have revealed significant breakthroughs in type 2 diabetes genetics and much of the heritability can be explained, we still lack insight into the molecular mechanisms, biological functions and clinical implications of most loci. Future research is also needed regarding population-specific genetics and missing heritability. Of note, despite the huge advances on the research side, translational pilot studies are scarce. The first use cases for PGSs in the clinic are likely to involve integrating them with current practice to predict high risk of diabetes or its comorbidities, which may lead to the development of pilot interventions for those at risk. The prediction of low risk of diabetes, without subsequent intervention, might be more cost-effective than targeting high-risk groups (given the presumably larger numbers of individuals affected), but this will be hard to accomplish given the current gaps in our knowledge.

Supplementary Information The online version contains a slideset of the figures for download available at https://doi.org/10.1007/ s00125-025-06419-1.

Acknowledgements We gratefully acknowledge J. Postma and U. Blom-Nilsson (Department of Clinical Sciences, Diabetes and Endocrinology, CRC, Lund University, Malmö, Sweden) for excellent administrative support. **Funding** Open access funding provided by Lund University. Work in the authors' laboratories is supported by grants from the Swedish research council (2021–02623), Vinnova (2023–04234), Diabetesfonden (DIA2023–847) and EFSD and Lilly European Diabetes Research Programme to RBP, and the Folkhalsan Research Foundation, Academy of Finland (grants no. 263401, 267882, 312063, 336822, 312072 and 336826) and University of Helsinki to LH and TT.

Authors' relationships and activities The authors declare that there are no relationships or activities that might bias, or be perceived to bias, their work.

**Contribution statement** All authors were responsible for drafting the article and reviewing it critically for important intellectual content. All authors approved the version to be published.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

#### References

- Holt RIG, Cockram CS, Flyvbjerg A, Goldstein BJ (2017) Textbook of diabetes. Wiley-Blackwell, Chichester, UK
- Neel JV (1976) Diabetes mellitus a geneticist's nightmare. In: Creutzfeldt W, Köbberling J, Neel JV (eds) The genetics of diabetes mellitus. Springer, Berlin, Heidelberg, pp 1–11. https://doi. org/10.1007/978-3-642-66332-1\_1
- Rich SS (2016) Diabetes: Still a geneticist's nightmare. Nature 536:37–38. https://doi.org/10.1038/nature18906
- Suzuki K, Hatzikotoulas K, Southam L et al (2024) Genetic drivers of heterogeneity in type 2 diabetes pathophysiology. Nature 627:347–357. https://doi.org/10.1038/s41586-024-07019-6
- Mahajan A, Spracklen CN, Zhang W et al (2022) Multi-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation. Nat Genet 54:560–572. https://doi.org/10.1038/s41588-022-01058-3
- Mahajan A, Taliun D, Thurner M et al (2018) Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. Nat Genet 50(11):1505– 1513. https://doi.org/10.1038/s41588-018-0241-6
- Vujkovic M, Keaton JM, Lynch JA et al (2020) Discovery of 318 new risk loci for type 2 diabetes and related vascular outcomes among 1.4 million participants in a multi-ancestry meta-analysis. Nat Genet 52:680–691. https://doi.org/10.1038/s41588-020-0637-y
- Prasad RB, Groop L (2015) Genetics of type 2 diabetes-pitfalls and possibilities. Genes (Basel) 6:87–123. https://doi.org/10.3390/ genes6010087
- Spracklen CN, Horikoshi M, Kim YJ et al (2020) Identification of type 2 diabetes loci in 433,540 East Asian individuals. Nature 582:240–245. https://doi.org/10.1038/s41586-020-2263-3
- Kengne AP, Beulens JW, Peelen LM et al (2014) Non-invasive risk scores for prediction of type 2 diabetes (EPIC-InterAct): a

validation of existing models. Lancet Diabetes Endocrinol 2:19-29. https://doi.org/10.1016/S2213-8587(13)70103-7

- Di Camillo B, Hakaste L, Sambo F et al (2018) HAPT2D: high accuracy of prediction of T2D with a model combining basic and advanced data depending on availability. Eur J Endocrinol 178:331–341. https://doi.org/10.1530/EJE-17-0921
- Hippisley-Cox J, Coupland C (2017) Development and validation of QDiabetes-2018 risk prediction algorithm to estimate future risk of type 2 diabetes: cohort study. BMJ 359:j5019. https://doi.org/10.1136/bmj.j5019
- Bonora E, Trombetta M, Dauriz M et al (2020) Chronic complications in patients with newly diagnosed type 2 diabetes: prevalence and related metabolic and clinical features: the Verona Newly Diagnosed Type 2 Diabetes Study (VNDS) 9. BMJ Open Diabetes Res Care 8(1):e001549. https://doi.org/10. 1136/bmjdrc-2020-001549
- Jayedi A, Soltani S, Motlagh SZT et al (2022) Anthropometric and adiposity indicators and risk of type 2 diabetes: systematic review and dose-response meta-analysis of cohort studies. BMJ 376:e067516. https://doi.org/10.1136/bmj-2021-067516
- Pepe MS, Janes H, Longton G, Leisenring W, Newcomb P (2004) Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. Am J Epidemiol 159:882–890. https://doi.org/10.1093/aje/kwh101
- Lendrem BC, Lendrem DW, Pratt AG et al (2019) Between a ROC and a hard place: Teaching prevalence plots to understand real world biomarker performance in the clinic. Pharm Stat 18:632–635. https://doi.org/10.1002/pst.1963
- Almgren P, Lehtovirta M, Isomaa B et al (2011) Heritability and familiality of type 2 diabetes and related quantitative traits in the Botnia Study. Diabetologia 54:2811–2819. https://doi.org/ 10.1007/s00125-011-2267-5
- Ritchie SC, Taylor HJ, Liang Y et al (2024) Integrated clinical risk prediction of type 2 diabetes with a multifactorial polygenic risk score. MedRxiv (Preprint). 23 Sep 2024. Available from: https://doi.org/10.1101/2024.08.22.24312440
- Manolio TA, Collins FS, Cox NJ et al (2009) Finding the missing heritability of complex diseases. Nature 461:747–753. https://doi.org/10.1038/nature08494
- Bjornstad P, Drews K, Zeitler PS (2021) Long-term complications in youth-onset type 2 diabetes. Reply. N Engl J Med 385:2016
- 21. Lin B, Coleman RL, Bragg F, Maddaloni E, Holman RR, Adler AI (2024) Younger-onset compared with later-onset type 2 diabetes: an analysis of the UK Prospective Diabetes Study (UKPDS) with up to 30 years of follow-up (UKPDS 92). Lancet Diabetes Endocrinol 12:904–914. https://doi.org/10.1016/S2213-8587(24) 00242-0
- 22. Brunetti P (2007) The lean patient with type 2 diabetes: characteristics and therapy challenge. Int J Clin Pract Suppl 153:3–9. https://doi.org/10.1111/j.1742-1241.2007.01359.x
- Liu X, Collister JA, Clifton L, Hunter DJ, Littlejohns TJ (2023) Polygenic risk of prediabetes, undiagnosed diabetes, and incident type 2 diabetes stratified by diabetes risk factors. J Endocr Soc 7:bvad020. https://doi.org/10.1210/jendso/bvad020
- Ojima T, Namba S, Suzuki K et al (2024) Body mass index stratification optimizes polygenic prediction of type 2 diabetes in crossbiobank analyses. Nat Genet 56:1100–1109. https://doi.org/10. 1038/s41588-024-01782-y
- Tamlander M, Mars N, Pirinen M, FinnGen Widen E, Ripatti S (2022) Integration of questionnaire-based risk factors improves polygenic risk scores for human coronary heart disease and type 2 diabetes. Commun Biol 5:158. https://doi.org/10.1038/ s42003-021-02996-0
- 26. Yu G, Tam CHT, Lim CKP et al (2025) Type 2 diabetes pathwayspecific polygenic risk scores elucidate heterogeneity in clinical presentation, disease progression and diabetic complications in

18,217 Chinese individuals with type 2 diabetes. Diabetologia 68(3):602–614. https://doi.org/10.1007/s00125-024-06309-y

- Ashenhurst JR, Sazonova OV, Svrchek O et al (2022) A polygenic score for type 2 diabetes improves risk stratification beyond current clinical screening factors in an ancestrally diverse sample. Front Genet 13:871260. https://doi.org/10.3389/fgene.2022.871260
- Mars N, Lindbohm JV, Della BriottaParolo P et al (2022) Systematic comparison of family history and polygenic risk across 24 common diseases. Am J Hum Genet 109:2152–2162. https:// doi.org/10.1016/j.ajhg.2022.10.009
- 29. Hivert MF, Jablonski KA, Perreault L et al (2011) Updated genetic score based on 34 confirmed type 2 diabetes Loci is associated with diabetes incidence and regression to normoglycemia in the diabetes prevention program. Diabetes 60:1340–1348. https://doi.org/10.2337/db10-1119
- Merino J, Guasch-Ferre M, Li J et al (2022) Polygenic scores, diet quality, and type 2 diabetes risk: an observational study among 35,759 adults from 3 US cohorts. PLoS Med 19:e1003972. https:// doi.org/10.1371/journal.pmed.1003972
- Yun JS, Jung SH, Shivakumar M et al (2022) Polygenic risk for type 2 diabetes, lifestyle, metabolic health, and cardiovascular disease: a prospective UK Biobank study. Cardiovasc Diabetol 21:131. https://doi.org/10.1186/s12933-022-01560-2
- 32. Widen E, Junna N, Ruotsalainen S et al (2022) How communicating polygenic and clinical risk for atherosclerotic cardiovascular disease impacts health behavior: an observational follow-up study. Circ Genom Precis Med 15:e003459. https://doi.org/10. 1161/CIRCGEN.121.003459
- Chatterjee S, Khunti K, Davies MJ (2017) Type 2 diabetes. Lancet 389:2239–2251. https://doi.org/10.1016/S0140-6736(17)30058-2
- 34. Razaghizad A, Oulousian E, Randhawa VK et al (2022) Clinical prediction models for heart failure hospitalization in type 2 diabetes: a systematic review and meta-analysis. J Am Heart Assoc 11:e024833. https://doi.org/10.1161/JAHA.121.024833
- Chamnan P, Simmons RK, Sharp SJ, Griffin SJ, Wareham NJ (2009) Cardiovascular risk assessment scores for people with diabetes: a systematic review. Diabetologia 52:2001–2014. https:// doi.org/10.1007/s00125-009-1454-0
- Dunkler D, Gao P, Lee SF et al (2015) Risk prediction for early CKD in type 2 diabetes. Clin J Am Soc Nephrol 10:1371–1379. https://doi.org/10.2215/CJN.10321014
- Boussageon R, Pouchain D, Renard V (2017) Prevention of complications in type 2 diabetes: is drug glucose control evidence based? Br J Gen Pract 67:85–87. https://doi.org/10.3399/bjgp17X689317
- Garg D, Naugler C, Bhella V, Yeasmin F (2018) Chronic kidney disease in type 2 diabetes: does an abnormal urine albumin-to-creatinine ratio need to be retested? Can Fam Physician 64:e446–e452
- Ali AS, Pham C, Morahan G, Ekinci EI (2024) Genetic risk scores identify people at high risk of developing diabetic kidney disease: a systematic review. J Clin Endocr Metab 109:1189–1197. https:// doi.org/10.1210/clinem/dgad704
- Forrest IS, Chaudhary K, Paranjpe I et al (2021) Genome-wide polygenic risk score for retinopathy of type 2 diabetes. Hum Mol Genet 30:952–960. https://doi.org/10.1093/hmg/ddab067
- 41. Yun JS, Jung SH, Shivakumar M et al (2022) Associations between polygenic risk of coronary artery disease and type 2 diabetes, lifestyle, and cardiovascular mortality: a prospective UK Biobank study. Front Cardiovasc Med 9:919374. https://doi.org/ 10.3389/fcvm.2022.919374
- 42. Mars N, Kerminen S, Feng YA et al (2022) Genome-wide risk prediction of common diseases across ancestries in one million people. Cell Genom 2(4):100118. https://doi.org/10.1016/j.xgen. 2022.100118
- 43. Tsao NL, Judy R, Levin MG et al (2023) Evaluation of the performance of the RECODe equation with the addition of polygenic

risk scores for adverse cardiovascular outcomes in individuals with type II diabetes. MedRxiv (Preprint). 5 May 2023. Available from: https://doi.org/10.1101/2023.05.03.23289457

- 44. van Zuydam NR, Ladenvall C, Voight BF et al (2020) Genetic predisposition to coronary artery disease in type 2 diabetes mellitus. Circ Genom Precis Med 13:e002769. https://doi.org/10.1161/ CIRCGEN.119.002769
- Li JH, Florez JC (2022) On the verge of precision medicine in diabetes. Drugs 82:1389–1401. https://doi.org/10.1007/ s40265-022-01774-4
- 46. Kerb R, Brinkmann U, Chatskaia N et al (2002) Identification of genetic variations of the human organic cation transporter hOCT1 and their functional consequences. Pharmacogenetics 12:591–595. https://doi.org/10.1097/00008571-200211000-00002
- Zhou K, Yee SW, Seiser EL et al (2016) Variation in the glucose transporter gene SLC2A2 is associated with glycemic response to metformin. Nat Genet 48:1055–1059. https://doi.org/10.1038/ ng.3632
- 48. van Leeuwen N, Nijpels G, Becker ML et al (2012) A gene variant near ATM is significantly associated with metformin treatment response in type 2 diabetes: a replication and meta-analysis of five cohorts. Diabetologia 55:1971–1977. https://doi.org/10.1007/ s00125-012-2537-x
- Christensen MM, Brasch-Andersen C, Green H et al (2011) The pharmacogenetics of metformin and its impact on plasma metformin steady-state levels and glycosylated hemoglobin A1c. Pharmacogenet Genomics 21:837–850. https://doi.org/10.1097/ FPC.0b013e32834c0010
- Shu Y, Sheardown SA, Brown C et al (2007) Effect of genetic variation in the organic cation transporter 1 (OCT1) on metformin action. J Clin Invest 117:1422–1431. https://doi.org/10.1172/JCI30558
- Euesden J, Lewis CM, O'Reilly PF (2015) PRSice: polygenic risk score software. Bioinformatics 31:1466–1468. https://doi.org/10. 1093/bioinformatics/btu848
- Li JH, Brenner LN, Kaur V et al (2023) Genome-wide association analysis identifies ancestry-specific genetic variation associated with acute response to metformin and glipizide in SUGAR-MGH. Diabetologia 66:1260–1272. https://doi.org/10.1007/ s00125-023-05922-7
- Li JH, Szczerbinski L, Dawed AY et al (2021) A polygenic score for type 2 diabetes risk is associated with both the acute and sustained response to sulfonylureas. Diabetes 70:293–300. https:// doi.org/10.2337/db20-0530
- Dawed AY, Mari A, Brown A et al (2023) Pharmacogenomics of GLP-1 receptor agonists: a genome-wide analysis of observational data and large randomised controlled trials. Lancet Diabetes Endo 11:33–41. https://doi.org/10.1016/S2213-8587(22)00340-0
- 55. Wilson JR, Shuey MM, Brown NJ, Devin JK (2017) Hypertension and type 2 diabetes are associated with decreased inhibition of dipeptidyl peptidase-4 by sitagliptin. J Endocr Soc 1:1168–1178. https://doi.org/10.1210/js.2017-00312
- 56. Javorsky M, Gotthardova I, Klimcakova L et al (2016) A missense variant in GLP1R gene is associated with the glycaemic response to treatment with gliptins. Diabetes Obes Metab 18:941–944. https://doi.org/10.1111/dom.12682
- 57. Zimdahl H, Haupt A, Brendel M et al (2017) Influence of common polymorphisms in the SLC5A2 gene on metabolic traits in subjects at increased risk of diabetes and on response to empagliflozin treatment in patients with diabetes. Pharmacogenet Genomics 27:135–142. https://doi.org/10.1097/FPC.000000000000268
- Kim NY, Lee H, Kim S et al (2024) The clinical relevance of a polygenic risk score for type 2 diabetes mellitus in the Korean population. Sci Rep 14:5749. https://doi.org/10.1038/ s41598-024-55313-0
- 59. Sivadas A, Sahana S, Jolly B et al (2024) Landscape of pharmacogenetic variants associated with non-insulin antidiabetic

drugs in the Indian population. BMJ Open Diabetes Res Care 12(2):e003769. https://doi.org/10.1136/bmjdrc-2023-003769

- Martin AR, Gignoux CR, Walters RK et al (2020) Human demographic history impacts genetic risk prediction across diverse populations. Am J Hum Genet 107:788–789. https://doi.org/10. 1016/j.ajhg.2020.08.020
- Flannick J, Florez JC (2016) Type 2 diabetes: genetic data sharing to advance complex disease research. Nat Rev Genet 17:535–549. https://doi.org/10.1038/nrg.2016.56
- Prasad RB, Ahlqvist E, Groop L (2018) Heterogeneity of diabetes - An Indian perspective. Diabetes Metab Syndr 13(5):3065– 3067. https://doi.org/10.1016/j.dsx.2018.07.001
- Consortium STD, Williams AL, Jacobs SB et al (2014) Sequence variants in SLC16A11 are a common risk factor for type 2 diabetes in Mexico. Nature 506:97–101. https://doi.org/ 10.1038/nature12828
- 64. Tabassum R, Chauhan G, Dwivedi OP et al (2013) Genomewide association study for type 2 diabetes in Indians identifies a new susceptibility locus at 2q21. Diabetes 62:977–986. https:// doi.org/10.2337/db12-0406
- 65. Hanson RL, Muller YL, Kobes S et al (2014) A genome-wide association study in American Indians implicates DNER as a susceptibility locus for type 2 diabetes. Diabetes 63:369–376. https://doi.org/10.2337/db13-0416
- 66. Suzuki K, Akiyama M, Ishigaki K et al (2019) Identification of 28 new susceptibility loci for type 2 diabetes in the Japanese population. Nat Genet 51:379–386. https://doi.org/10.1038/ s41588-018-0332-4
- Polfus LM, Darst BF, Highland H et al (2021) Genetic discovery and risk characterization in type 2 diabetes across diverse populations. HGG Adv 2(2):100029. https://doi.org/10.1016/j. xhgg.2021.100029
- Ge T, Irvin MR, Patki A et al (2022) Development and validation of a trans-ancestry polygenic risk score for type 2 diabetes in diverse populations. Genome Med 14:70. https://doi.org/10. 1186/s13073-022-01074-2
- Chen J, Sun M, Adeyemo A et al (2019) Genome-wide association study of type 2 diabetes in Africa. Diabetologia 62:1204– 1211. https://doi.org/10.1007/s00125-019-4880-7
- 70. Yajnik CS, Wagh R, Kunte P et al (2023) Polygenic scores of diabetes related traits in subgroups of type 2 diabetes in India: a cohort study. Lancet Reg Health Southeast Asia 14:100182. https://doi.org/10.1016/j.lansea.2023.100182
- Chikowore T, Ekoru K, Vujkovi M et al (2022) Polygenic prediction of type 2 diabetes in Africa. Diabetes Care 45:717–723. https://doi.org/10.2337/dc21-0365
- McCarthy MI (2017) Painting a new picture of personalised medicine for diabetes. Diabetologia 60:793–799. https://doi.org/ 10.1007/s00125-017-4210-x
- He Y, Lakhani CM, Rasooly D, Manrai AK, Tzoulaki I, Patel CJ (2021) Comparisons of polyexposure, polygenic, and clinical risk scores in risk prediction of type 2 diabetes. Diabetes Care 44:935–943. https://doi.org/10.2337/dc20-2049
- 74. Rout M, Wander GS, Ralhan S et al (2023) Assessing the prediction of type 2 diabetes risk using polygenic and clinical risk scores in South Asian study populations. Ther Adv Endocrinol Metab 14:20420188231220120. https://doi.org/10.1177/20420 188231220120
- Loh M, Zhang W, Ng HK et al (2022) Identification of genetic effects underlying type 2 diabetes in South Asian and European populations. Commun Biol 5:329. https://doi.org/10.1038/ s42003-022-03248-5
- 76. Wheeler E, Leong A, Liu CT et al (2017) Impact of common genetic determinants of Hemoglobin A1c on type 2 diabetes risk and diagnosis in ancestrally diverse populations: A transethnic

genome-wide meta-analysis. PLoS Med 14:e1002383. https://doi.org/10.1371/journal.pmed.1002383

- Dornbos P, Koesterer R, Ruttenburg A et al (2022) A combined polygenic score of 21,293 rare and 22 common variants improves diabetes diagnosis based on hemoglobin A1C levels. Nat Genet 54:1609–1614. https://doi.org/10.1038/s41588-022-01200-1
- Hahn SJ, Kim S, Choi YS, Lee J, Kang J (2022) Prediction of type 2 diabetes using genome-wide polygenic risk score and metabolic profiles: A machine learning analysis of population-based 10-year prospective cohort study. EBioMedicine 86:104383. https://doi. org/10.1016/j.ebiom.2022.104383
- Zhang H, Zhan J, Jin J et al (2023) A new method for multiancestry polygenic prediction improves performance across diverse populations. Nat Genet 55:1757–1768. https://doi.org/10.1038/ s41588-023-01501-z
- Lehmann B, Mackintosh M, McVean G, Holmes C (2023) Optimal strategies for learning multi-ancestry polygenic scores vary across traits. Nat Commun 14:4023. https://doi.org/10.1038/ s41467-023-38930-7
- Prasad RB, Asplund O, Shukla SR et al (2022) Subgroups of patients with young-onset type 2 diabetes in India reveal insulin deficiency as a major driver. Diabetologia 65:65–78. https://doi. org/10.1007/s00125-021-05543-y
- Yajnik CS, Bandopadhyay S, Bhalerao A et al (2021) Poor in utero growth, and reduced beta-cell compensation and high fasting glucose from childhood, are harbingers of glucose intolerance in young Indians. Diabetes Care 44:2747–2757. https://doi.org/10. 2337/dc20-3026
- Udler MS, McCarthy MI, Florez JC, Mahajan A (2019) Genetic risk scores for diabetes diagnosis and precision medicine. Endocr Rev 40:1500–1520. https://doi.org/10.1210/er.2019-00088
- Prasad RB, Groop L (2019) Precision medicine in type 2 diabetes. J Intern Med 285:40–48. https://doi.org/10.1111/joim.12859
- Ahlqvist E, Prasad RB, Groop L (2021) 100 years of insulin: towards improved precision and a new classification of diabetes mellitus. J Endocrinol 252:R59–R70. https://doi.org/10.1530/ JOE-20-0596
- Mahajan A, Wessel J, Willems SM et al (2018) Refining the accuracy of validated target identification through coding variant finemapping in type 2 diabetes. Nat Genet 50:559–571. https://doi. org/10.1038/s41588-018-0084-1
- Udler MS, Kim J, von Grotthuss M et al (2018) Type 2 diabetes genetic loci informed by multi-trait associations point to disease mechanisms and subtypes: a soft clustering analysis. PLoS Med 15:e1002654. https://doi.org/10.1371/journal.pmed.1002654
- DiCorpo D, LeClair J, Cole JB et al (2022) Type 2 diabetes partitioned polygenic scores associate with disease outcomes in 454,193 individuals across 13 cohorts. Diabetes Care 45:674–683. https://doi.org/10.2337/dc21-1395
- Kim H, Westerman KE, Smith K et al (2023) High-throughput genetic clustering of type 2 diabetes loci reveals heterogeneous mechanistic pathways of metabolic disease. Diabetologia 66:495– 507. https://doi.org/10.1007/s00125-022-05848-6
- 90. Smith K, Deutsch AJ, McGrail C et al (2024) Multi-ancestry polygenic mechanisms of type 2 diabetes. Nat Med 30:1065–1074. https://doi.org/10.1038/s41591-024-02865-3
- Hodgson S, Williamson A, Bigossi M et al (2024) Genetic basis of early onset and progression of type 2 diabetes in South Asians. Nat Med 31:323–331. https://doi.org/10.1038/s41591-024-03317-8
- Ahlqvist E, Storm P, Karajamaki A et al (2018) Novel subgroups of adult-onset diabetes and their association with outcomes: a datadriven cluster analysis of six variables. Lancet Diabetes Endocrinol 6:361–369. https://doi.org/10.1016/S2213-8587(18)30051-2
- Zaharia OP, Strassburger K, Strom A et al (2019) Risk of diabetes-associated diseases in subgroups of patients with recent-onset

diabetes: a 5-year follow-up study. Lancet Diabetes Endocrinol 7:684–694. https://doi.org/10.1016/S2213-8587(19)30187-1

- 94. Mansour Aly D, Dwivedi OP, Prasad RB et al (2021) Genomewide association analyses highlight etiological differences underlying newly defined subtypes of diabetes. Nat Genet 53:1534– 1542. https://doi.org/10.1038/s41588-021-00948-2
- 95. Yajnik CS, Kinare A, Chinchwadkar M et al (2024) Slower liver growth in-utero predicts the risk of prediabetes at 18 years of age, a DOHaD jump from birth size to organogenesis: novel human data in the Pune Maternal Nutrition Study (PMNS). Diabetologia 67(Suppl 1):1–593. https://doi.org/10.1007/s00125-024-06226-0
- 96. Anjana RM, Baskar V, Nair ATN et al (2020) Novel subgroups of type 2 diabetes and their association with microvascular outcomes in an Asian Indian population: a data-driven cluster analysis: the

INSPIRED study. BMJ Open Diabetes Res Care 8:e001506. https://doi.org/10.1136/bmjdrc-2020-001506

- 97. Hodgson S, Huang QQ, Sallah N et al (2022) Integrating polygenic risk scores in the prediction of type 2 diabetes risk and subtypes in British Pakistanis and Bangladeshis: a population-based cohort study. PLoS Med 19:e1003981. https://doi.org/10.1371/ journal.pmed.1003981
- Wagner R, Heni M, Tabak AG et al (2021) Pathophysiology-based subphenotyping of individuals at elevated risk for type 2 diabetes. Nat Med 27:49–57. https://doi.org/10.1038/s41591-020-1116-9

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.