





















Precision subclassification of type 2 diabetes: a systematic review

Shivani Misra ^{1,2,209}✉, Robert Wagner ^{3,4,5,209}, Bige Ozkan ^{6,7}, Martin Schön ^{4,5,8}, Magdalena Sevilla-Gonzalez ^{9,10,11}, Katsiaryna Prystupa ^{4,5}, Caroline C. Wang⁶, Raymond J. Kreienkamp ^{10,12,13,14}, Sara J. Cromer^{10,11,12,13}, Mary R. Rooney ^{6,15}, Daisy Duan ¹⁶, Anne Cathrine Baun Thuesen ¹⁷, Amelia S. Wallace ^{6,15}, Aaron Leong^{10,11,12,18}, Aaron J. Deutsch ^{10,11,12,13}, Mette K. Andersen ¹⁷, Liana K. Billings ^{19,20}, Robert H. Eckel²¹, Wayne Huey-Herng Sheu^{22,23,24}, Torben Hansen ¹⁷, Norbert Stefan ^{5,25,26}, Mark O. Goodarzi ²⁷, Debashree Ray ^{15,28}, Elizabeth Selvin ^{6,15}, Jose C. Florez^{10,11,12,13}, ADA/EASD PMDI*, James B. Meigs^{10,11,18,210} & Miriam S. Udler ^{10,11,12,13,210}

Abstract

Background Heterogeneity in type 2 diabetes presentation and progression suggests that precision medicine interventions could improve clinical outcomes. We undertook a systematic review to determine whether strategies to subclassify type 2 diabetes were associated with high quality evidence, reproducible results and improved outcomes for patients.

Methods We searched PubMed and Embase for publications that used ‘simple subclassification’ approaches using simple categorisation of clinical characteristics, or ‘complex subclassification’ approaches which used machine learning or ‘omics approaches in people with established type 2 diabetes. We excluded other diabetes subtypes and those predicting incident type 2 diabetes. We assessed quality, reproducibility and clinical relevance of extracted full-text articles and qualitatively synthesised a summary of subclassification approaches.

Results Here we show data from 51 studies that demonstrate many simple stratification approaches, but none have been replicated and many are not associated with meaningful clinical outcomes. Complex stratification was reviewed in 62 studies and produced reproducible subtypes of type 2 diabetes that are associated with outcomes. Both approaches require a higher grade of evidence but support the premise that type 2 diabetes can be subclassified into clinically meaningful subtypes.

Conclusion Critical next steps toward clinical implementation are to test whether subtypes exist in more diverse ancestries and whether tailoring interventions to subtypes will improve outcomes.

Plain language summary

In people with type 2 diabetes there may be differences in the way people present, including for example, their symptoms, body weight or how much insulin they make. We looked at recent publications describing research in this area to see whether it is possible to separate people with type 2 diabetes into different subgroups and, if so, whether these groupings were useful for patients. We found that it is possible to group people with type 2 diabetes into different subgroups and being in one subgroup can be more strongly linked to the likelihood of developing complications over others. This might mean that in the future we can treat people in different subgroups differently in ways that improves their treatment and their health but it requires further study.

Type 2 diabetes is a global health problem posing substantial burdens on human health¹. The diagnosis of type 2 diabetes is based on elevated blood glucose coupled with the absence of clinical features indicating alternative subtypes, such as type 1, monogenic, pancreatic or medication-induced diabetes². A diagnosis of type 2 diabetes is generally the default or can be arrived at through exclusion of other types. Traditionally, most type 2 diabetes care guidelines have advocated treatment choice based on cost-effectiveness and side effects of specific medications, which have no relationship to underlying pathophysiology in the individual. More recent guidelines have suggested differential glucose-lowering therapies on the basis of higher body mass index (BMI) (favouring use of glucagon-like peptide analogue, GLP-1) or presence or absence of cardiovascular and/or renal disease and/or heart failure (favouring GLP-1 and/or sodium-glucose co-transporter 2, SGLT-2 inhibitors)³.

There is considerable heterogeneity in the clinical characteristics of patients with type 2 diabetes. Clinicians recognise that differences in degree of obesity or body fat distribution, age, dyslipidaemia or presence of metabolic syndrome can influence prognosis in diabetes and can be important considerations in treatment and management^{4–6}. There is increasing awareness that type 2 diabetes heterogeneity may reflect differences in the underlying pathophysiology, environmental contributors, and the genetic risk of affected individuals. The mechanisms leading to the development of type 2 diabetes may differ from one individual to another and this could impact treatment and outcome.

Accurate characterisation of the heterogeneity in type 2 diabetes may help individualise care and improve outcomes. This goal has been realised in part for monogenic diabetes, where treatments can be tailored to genetic subtype to deliver precision care achieving better outcomes than standard care⁷. Given the complex pathophysiology and genetics of type 2 diabetes, applying precision medicine approaches is challenging. Critical to this endeavour is a better understanding of specific subtypes.

There are many studies of type 2 diabetes subtypes. The literature reflects diverse approaches based on the presence or absence of one or more simple clinical features or biomarkers and, more recently, sophisticated methods that deploy machine learning (ML) or use omics data. Classification approaches such as clustering methods to categorise this heterogeneity show inter-cluster differences in progression to complications or need for insulin treatment. These approaches consider clinical features at diagnosis⁸ or clinical information combined with genetic data to characterise disease heterogeneity^{9,10}. Simpler approaches are more easily implemented across all resource settings, while complex approaches may have greater precision in classifying heterogeneity. The breadth and scope of the evidence in favour of type 2 diabetes subclassification have not to date been thoroughly examined.

The *Precision Medicine in Diabetes Initiative* (PMDI) was established in 2018 by the American Diabetes Association (ADA) in partnership with the European Association for the Study of Diabetes (EASD). The ADA/EASD PMDI includes global thought leaders in precision diabetes medicine who are working to address the burgeoning need for better diabetes prevention and care through precision medicine¹¹. This Systematic Review is written with the ADA/EASD PMDI as part of a comprehensive evidence evaluation in support of the 2nd International Consensus Report on Precision Diabetes Medicine¹².

In this systematic review for the PMDI we aimed to provide a critical assessment of the evidence to date for type 2 diabetes subclassification using (i) simple approaches based on categorisation of clinical features, biomarkers, imaging, or other parameters, and (ii) complex subclassification approaches that use ML incorporating clinical data and/or genomic data. We aimed to

identify areas where further research is needed with the goal to improve patient and health system outcomes in type 2 diabetes care.

Our analysis shows that many simple approaches to subclassification have been tried but none have been replicated and most are not associated with meaningful clinical outcomes. However, a more complex stratification, using machine learning applied to clinical variables, yielded reproducible subtypes of type 2 diabetes that are associated with outcomes. Both approaches, however, require a higher grade of evidence but support the premise that type 2 diabetes can be subclassified into clinically meaningful subtypes.

Methods

This systematic review was written and conducted in accordance with our pre-established protocol (PROSPERO ID CRD42022310539) and reported using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Statement (PRISMA)¹³. We systematically reviewed papers to address two research questions devised by an expert working group: 1) What are the main subtypes of type 2 diabetes defined using simple clinical criteria and/or routinely available laboratory tests (simple approaches), and 2) What subphenotypes of type 2 diabetes can be reproducibly identified using ML and/or genomics approaches (complex approaches)? Subsequently, we refer to the first question as *simple approaches* and the second question as *complex approaches*. The quality of each paper was reported, and the aggregate of data evaluated using the Grading of Recommendations, Assessment, Development, and Evaluations (GRADE) system¹⁴.

Study eligibility criteria. We included English-language original research studies of all design types that analysed populations with prevalent or new-onset type 2 diabetes and attempted in some way to stratify or subgroup patients with type 2 diabetes. We used broad terms to identify stratification studies and all approaches to stratification (the exposure) were included (supplementary table 1). We excluded studies examining risk for the development of type 2 diabetes, use of glycaemic control (e.g. HbA1c strata) alone to stratify, studies of stratification in types of diabetes other than type 2 diabetes, and review articles or case reports.

For simple approaches the exposure was defined as any of the following; a routine blood or urine biomarker that was widely available in most clinic settings; a blood or urine biomarker that might not be routinely available now but could have the potential to become easily accessible; any routinely available imaging modality; any physiological assessment that could be undertaken in an outpatient setting or results from routinely available dynamic tests. The stratification approach was either a cut-off or categorisation based on one or more of the above or if an index, ratio, trend or other analysis was undertaken, it could be calculated without complex mathematics. Finally, all outcomes were accepted for example clinical characterisation of subgroups, association with specific biomarkers and association with complications or mortality.

For complex approaches, the exposure used was defined as any of the inputs for the simple approach outlined above and/or any form of genetic data. However, unlike the simple approach, the stratification approach either deployed ML approaches or used other complex statistical approaches for stratification. All outcomes were accepted, as above, for simple.

Literature search and selection strategy. PUBMED and EMBASE databases were searched from inception to May 2022 for relevant articles using a strategy devised by expert health sciences librarians (supplementary methods). We undertook

independent searches for each systematic review question. From both searches, each abstract and subsequently, full text paper, was screened by two independent team members for eligibility. In addition to the initial exclusion criteria, at the full-text review stage, we further excluded studies where exposures were not clearly defined and/or if the data on outcomes of the stratification were not available in results or supplementary material. We also excluded studies where the *only* stratification modality was a measure of glycaemic control, as this itself provides the diagnosis of type 2 diabetes. In cases of disagreement between two reviewers, a third reviewer made the final decision. The process involved group-based discussions to resolve disagreements to ensure all decisions were made on the same grounds.

Data extraction. Data were manually extracted from each full-text paper by individual team members and cross-checked by an independent team member at the data synthesis stage. We extracted relevant data on study design (observational or clinical trial), analysis design (cross-sectional or prospective), study population characteristics, stratification method and results (exposure), outcomes, and study quality assessment. For population characteristics, we extracted data on whether the type 2 diabetes population was new-onset or prevalent, the sample size, ethnicity and gender, the duration of diabetes (for cross-sectional analysis) and duration of follow-up (for longitudinal follow-up). For exposures, we extracted the approach to stratification and the number and nature of subgroups identified. For outcomes, we documented the type of outcome studied and the findings according to stratified subgroup.

Data synthesis. Following full-text data extraction, we undertook a qualitative analysis of exposures (measures used to stratify individuals) for each systematic review question. For simple subclassification approaches, we extracted the details of stratification criteria in each paper (supplementary methods), then categorised

the exposure as blood/urine test, imaging, age). After data extraction, these exposures were further refined into subcategories based on common emerging themes (e.g., use of pancreatic autoantibodies, BMI categories, measures of beta-cell function, use of lipid profiles). For complex approaches, the exposure included both the input clinical and/or genetic data used and the ML approach to analysis (e.g., k-means, hierarchical clustering, latent-class analysis), deployed. In both reviews, outcomes were heterogeneous, so we broadly categorised them where possible. Due to the variability in exposures and outcomes, it was not possible to undertake formal meta-analyses of any outcome. All coding, categorisation and thematic synthesis was undertaken and agreed upon by at least three members of the research team.

Quality assessment. The GRADE system was used to assess the quality of the studies extracted¹³. At least two members assessed whether study exposures and outcomes were clearly defined, valid and reliable, and whether confounders were appropriately accounted and adjusted for. Disagreements were resolved by discussion between the joint first and senior authors during group discussion. Assessors evaluated study limitations, consistency of results, imprecision, and reporting bias to assign study-specific and overall GRADE certainty ratings as very low, low, moderate and high¹⁵.

Reporting summary. Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Results

Search and screening for simple and complex systematic review questions. The first question examined simple stratification approaches using clinical variables that may reveal type 2 diabetes heterogeneity. A total of 6097 studies met the inclusion criteria and were screened (Fig. 1A). Of these, 183 studies were included for full text data review, of which 132 studies were subsequently

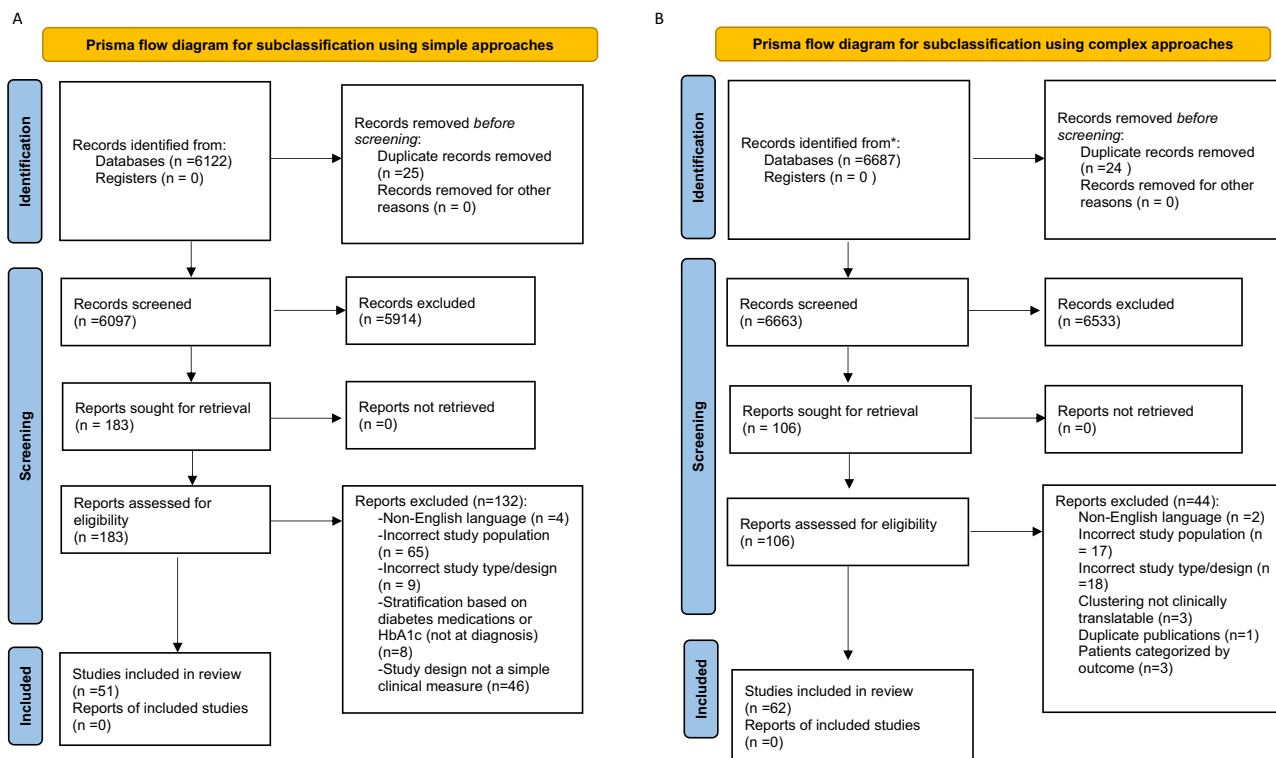


Fig. 1 PRISMA systematic review attrition diagram. A This shows the flow diagram for simple approaches to subclassification and **B** Complex approaches.

excluded. The most common reasons for exclusion at the full-text review stage were studies conducted in populations without prevalent or incident type 2 diabetes, study designs that used ML approaches or stratification approaches that used HbA1c or diabetes medications. In total, 51 “simple approach” studies underwent full-text data extraction.

The second question aimed to identify papers with complex approaches, mostly ML-based strategies, to identify subgroups of patients with type 2 diabetes (Fig. 1B). A total of 6639 studies were screened, of which 106 were found eligible for full-text review. The most common reasons for exclusion were study populations not comprising participants with type 2 diabetes or classification approaches not using ML. In total, 62 ‘complex’ studies underwent full-text data extraction.

Use of simple approaches to subclassify type 2 diabetes

Description of extracted studies: The 51 studies using simple type 2 diabetes subclassification approaches incorporated 1,751,350 participants with prevalent or new-onset type 2 diabetes. Among them, 39% (20/51) of studies included participants of white European ancestry, 43% (22/51) incorporated exclusively participants from non-white European ancestries and 17% (9/51) included mixed ancestry groups (Supplementary Data 1). The majority of the studies (78%, 40/51) were conducted in populations with prevalent type 2 diabetes, and 22% (11/51) in new-onset type 2 diabetes. Approximately half the studies had a prospective design (25/51), the remaining half had a cross-sectional (26/51) design. For longitudinal studies, study follow-up periods ranged from <1 year to 22 years.

Studies included a wide range of exposures (Fig. 2) based on routine clinical measurements with standard cut-offs or groupings. These included assessment of individual routine clinic-based measurements (e.g., levels of BMI, or biomarker variability over time) or composite stratification incorporating two or more tiers of criteria (e.g. groupings combining one or more biomarkers or anthropometric measurements) including both routine and non-routine but clinically available tests, including oral glucose tolerance tests (OGTT) which, while a glycaemic test, also

indirectly measures insulin resistance. The associations of stratified exposure characteristics were investigated with various outcomes: 1) measures of glycaemia, 2) clinical characteristics, 3) measures of diabetes progression such as time-to-insulin treatment or development of microvascular complications and 4) cardiovascular outcomes and/or mortality.

Description of categorised subgroups: Simple approaches to classification included use of lipid profiles (*n* = 8), BMI (*n* = 6), pancreatic beta-cell related measures (*n* = 6), pancreatic auto-antibodies (*n* = 6), age at diagnosis (*n* = 2), OGTT data (*n* = 4), cardiovascular measures (*n* = 3), other biomarkers in urine or blood and alternative approaches (*n* = 5) (Table 1).

Different categories of triglycerides, low-density lipoprotein (LDL) cholesterol, high-density lipoprotein (HDL) cholesterol, atherogenic small dense lipoproteins with and without features of metabolic syndrome were used to stratify type 2 diabetes in eight studies. Cardiovascular disease (CVD) outcomes were assessed in 3/8 of the studies^{16–18} which showed that a more atherogenic metric of the specific lipid exposure (e.g., higher LDL cholesterol) was associated with a greater frequency of CVD outcomes. Other outcomes included pulse wave velocity¹⁹ or clinical characteristics; age, BMI, presence of metabolic syndrome in specific subgroups.

The six studies assessing pancreatic autoantibodies focused on glutamic acid decarboxylase 65 (GAD-65) levels. Studies used positive versus negative status or high versus low titre, and one study sub-stratified by age. Outcomes included time-to-insulin treatment^{20,21}, associations with other clinical characteristics such as lipid profiles, BMI and blood pressure^{22–24} and measures of beta-cell function. There was no consistency in study design and most were observational with low to moderate evidence grade; two studies showed that GAD-65 positivity was associated with faster time-to-insulin treatment^{20,21}.

Patients with type 2 diabetes were stratified according to their BMI in six studies, either by BMI alone (*n* = 5) or BMI in combination with HbA1c. The number of BMI categories varied between two and six in the identified studies. The association between BMI and glycaemic outcomes (change in HbA1c from

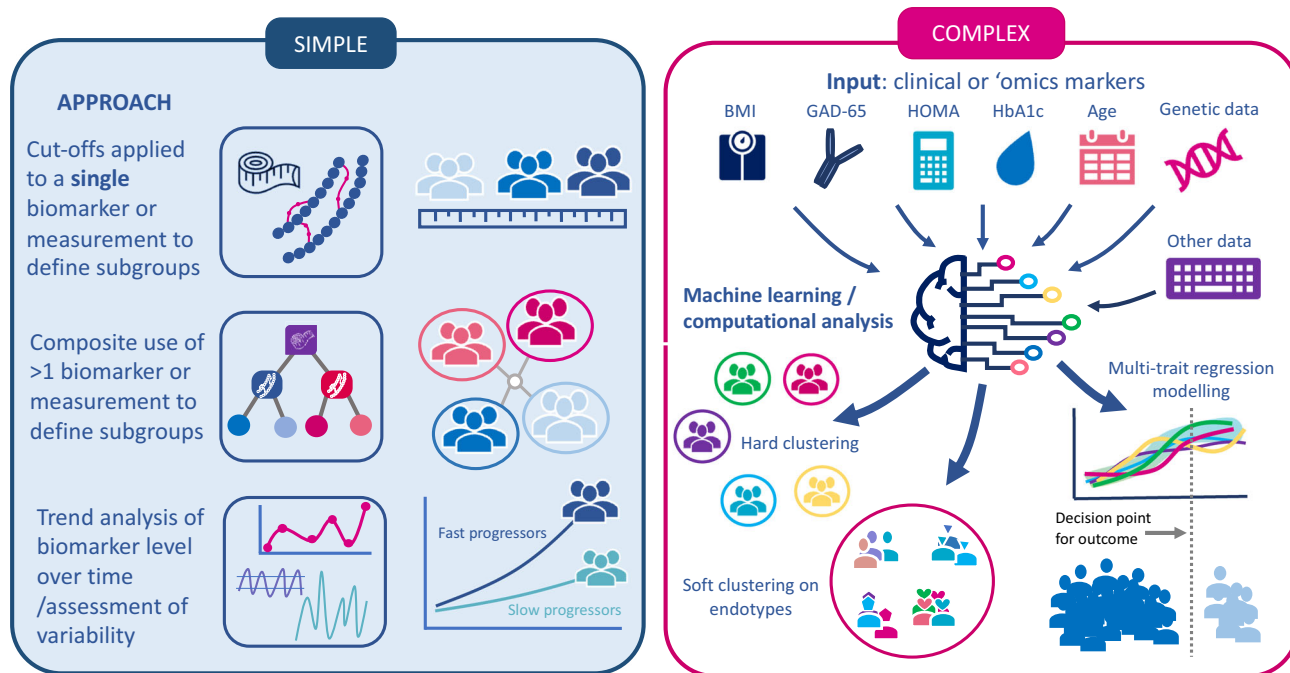


Fig. 2 Schematic overview of approaches used to subclassify type 2 diabetes. The figure summarises simple approaches that have been taken to subclassify type 2 diabetes and complex approaches. HbA1c glycated haemoglobin, BMI body mass index, GAD-65 glutamic acid decarboxylase-65 antibodies.

Table 1 Summary of published studies using simple approaches to type 2 diabetes classification.

Approach Total n participants % Female Mean age	Race/ethnicity breakdown, %	% Prevalent or new onset T2D, n (%) Prospective or Cross- sectional, n (%)	T2D subclassified Groups identified	Outcome category, n (%)	Overall quality, n (%)	References
Simple: Pancreatic autoantibodies (6 studies) n = 8350 42.2% female 54.6 years	87% European/ White 11% East Asian (Korean) 2% Black (Nigerian)	Prevalent (n = 5, 83%) New onset (n = 1, 17%) Longitudinal (n = 2, 33%) Cross-sectional (n = 4, 67%)	1) GAD antibody positive vs negative 2) GAD status sub-stratified by age 3) High and low titre GAD positivity	1) Earlier time to insulin treatment (n = 2, 33%) 2) Association with variable clinical characteristics (n = 3, 50%) 3) Beta-cell function physiological measure (n = 1, 17%)	Moderate (n = 4, 67%) Low (n = 1, 17%) Very low (n = 1, 17%)	20-24,55
Simple: Beta-cell related measure or index (6 studies) n = 17,881 38.1% female 56.9 years Simple: Age at diagnosis (2 studies) n = 1535 47.6% female Mean age not reported	75% East Asian (Chinese, Japanese) 24% European/ White, 1% South Asian (Indian) 57% European/ White 43% South Asian (Indian)	Prevalent (n = 4, 66%) New onset (n = 2, 33%) Longitudinal (n = 1, 17%) Cross-sectional (n = 5, 83%) Prevalent (n = 2, 100%) Longitudinal (n = 2, 100%)	1) High versus normal/low fasting insulin/C-peptide level/ UCPCR level 2) Insulin sensitive versus resistant 1) Age 60-75 years vs. >75 years. 2) Age at diagnosis <25 years v. >25 years	1) Biochemical measures (n = 2, 33%) 2) Micro/macrovacular disease (n = 3, 50%) 3) Response to medications (n = 2, 33%) 1) Total and cardiovascular mortality (n = 1, 50%) 2) Severity of retinal disease (n = 1, 50%)	Moderate (n = 2, 33%) Very low (n = 4, 66%) Low (n = 2, 100%)	56-61 4,6
Simple: BMI (6 studies) n = 74,104 41.6% female 62 years Simple: Lipid profile/metabolic syndrome (8 studies) n = 23,933 41.0% female 61.7 years	25% European/ White 23% Asian (Chinese, South Asian, Other) 2% non-White 50% not reported 52% European/ White, 32% East Asian (Chinese) 16% non-White	Prevalent (n = 6, 100%) Longitudinal (n = 4, 67%) Cross-sectional (n = 2, 33%) Prevalent (n = 7, 87.5%) New-onset (n = 1, 12.5%) Longitudinal (n = 3, 37.5%) Cross-sectional (n = 5, 62.5%)	1) BMI categories 2) BMI and HbA1c 3) BMI and TG 4) sex-stratified BMI 1) Atherogenic dyslipidaemia (y/n) 2) Tertiles of triglyceride- glucose index 3) Quartiles of small dense LDL-cholesterol 4) Triglyceride-glucose index and visceral adiposity 5) Atherogenic index 6) Hypertriglyceridaemic waist phenotype 7) Triglyceride levels 8) Metabolic syndrome (y/n)	1) Physiological measure (n = 1, 17%) 2) T2D complication other than CVD (n = 1, 17%) 3) Glucose control (n = 4, 67%) 1) CVD event (n = 3, 37.5%) 2) Physiological measure (brachial-ankle PWV, metabolic syndrome; n = 2, 25%) 3) Glucose control (n = 1, 12.5%) 4) More than one of the above (n = 2, 25%)	Moderate (n = 5, 83%) Very low (n = 1, 17%) Moderate (n = 2, 25%) Low (n = 4, 50%) Very low (n = 2, 25%)	5,25-27,62,63 16-19,64-67

Table 1 (continued)

Approach Total n participants % Female Mean age	Race/ethnicity breakdown, %	% Prevalent or new onset T2D, n (%) Prospective or Cross- sectional, n (%)	T2D subclassified Groups identified	Outcome category, n (%)	Overall quality, n (%)	References
Simple: OGTT data (4 studies) n = 1399 31.0% female 50.7 years	34% White 12% Black 12% Hispanic 39% East Asian (Japanese)	Prevalent (n = 1, 25%) New onset (n = 3, 75%) Longitudinal (n = 2, 50%) Cross-sectional (n = 2, 50%)	1) Categorised by fasting and 2 h glucose 2) Treatment assignment and OGTT profile (monophasic, biphasic, and upward curve) 3) OGTT profiles: monophasic curve; biphasic curve & upward curve	1) CVD event (n = 1, 25%) 2) Glucose control (n = 1, 25%) 3) Physiological measure (glucose response, insulin sensitivity, insulinogenic index; n = 2, 50%)	Moderate (n = 2, 50%) Low (n = 2, 50%)	28,68-70
Simple: Alternative approach to stratification (5 studies) n = 1,572,942 45.2% female 58.6 years	99.7% East Asian (Korean) 0.28% European/ White 0.03% Black	Prevalent (n = 4, 80%) New-onset (n = 1, 20%) Longitudinal (n = 3, 60%) Cross-sectional (n = 2, 40%)	1) Total illness burden index 2) Diabetic retinopathy (yes/no) 3) Ketosis-prone diabetes defined as new-onset diabetes without precipitating events, 4) Severe hypoglycaemia (yes/no) 5) Diabetes duration	1) CVD event (n = 3, 60%) 2) Clinical biomarker (n = 1, 20%) 3) Physiological measure (n = 1, 20%)	Moderate (n = 1, 20%) Low (n = 4, 80%)	48,71-74
Simple: Cardiovascular features (3 studies) n = 8793 53.4% female 62.3 years	75% European/ White 16.4% Asian (Chinese, Malay, Indian) 8.6% Black	Prevalent (n = 3, 100%) Longitudinal (n = 2, 66.6%) Cross-sectional (n = 1, 33.3%)	1) Clinic pulse pressure quartiles 2) Pulse wave velocity tertiles 3) Cardiomyopathy (yes/no)	1) CVD event (n = 2, 66.6%) 2) T2D complication other than CVD (CKD progression, n = 1, 33.3%)	Moderate (n = 2, 66.6%) Low (n = 1, 33.3%)	75-77
Simple: Other blood or urine test (11 studies) n = 71,609 %47.4 female 62.2 years	70.6% European/ White 15.5% East Asian (Chinese, Taiwanese, Japanese) 1.2% Black 0.6% Hispanic 0.1% Middle- Eastern (Egyptian) 11.3% non-White 0.7% Other	Prevalent (n = 8, 72.7%) New-onset (n = 3, 27.3%) Longitudinal (n = 6, 54.5%) Cross-sectional (n = 5, 45.5%)	1) UACR ranges (n = 2) 2) eGFR (n = 1) 3) Glycosuria categories (n = 1) 4) C-reactive protein quartiles (n = 1) 5) Vaspin and adiponectin (n = 1) 6) Quartiles of fibrosis index & albuminuria (n = 1) 7) Total bile acid levels (n = 1) 8) log(bilirubin) (n = 1) 9) Hp2-2 phenotype (n = 1) 10) Multiple (HbA1c, total cholesterol, blood pressure) (n = 1)	1) Clinical biomarker (n = 4, 36.4%) 2) CVD event (n = 3, 27.3%) 3) T2D complication other than CVD (n = 2, 18.2%) 4) Death (n = 2, 18.2%)	Moderate (n = 5, 46%) Low (n = 3, 27.3%) Very low (n = 3, 27.3%)	78-88

T2D type 2 diabetes, GAD glutamic decarboxylase antibody, UCP2CR urine C-peptide to creatinine ratio, BMI body mass index, CVD cardiovascular disease, OGTT oral glucose tolerance test, CKD chronic kidney disease, LDL low-density lipoprotein cholesterol.

baseline) was assessed in four studies either as primary or secondary outcomes^{6,25,26}. We graded the quality of evidence as very low to moderate, and no consistency of effect was observed across all studies. In one secondary analysis of a randomised control trial, higher BMI at baseline was associated with faster progression to adverse renal outcomes, however, this was not replicated in any other study²⁷.

Age at diagnosis was assessed as a stratification tool in two studies; younger age (mean age 33 years) was associated with higher rates of proliferative retinopathy in an observational study with 12 months follow-up versus older age (mean 50 years)⁴. In a second study, patients aged 60–75 versus those >75 years had a high risk of CVD and mortality when stratified by cholesterol levels⁶. Neither study was replicated to confirm findings.

Four studies used results from oral glucose tolerance tests (OGTT) as exposures. The specific stratification approach applied to OGTT profiles was different in each study and based on cut-offs of fasting glucose levels, glucose gradients after stimulation and responses to different drug treatments. Outcomes included clamp-derived insulin sensitivity and differences in the shape of glucose profiles between youths and adults²⁸.

Measures of estimated beta-cell function were assessed in six studies including C-peptide levels and homeostasis model assessment-2 indices for beta-cell function (HOMA2-B) or insulin resistance (HOMA2-IR), which require measurement of fasting insulin and glucose levels. C-peptide was defined using variable cut-offs. Outcomes included clinical phenotype data, response to medication, and microvascular or macrovascular complications. For example, hyperinsulinaemia and higher urine C-peptide were independently associated with cardiovascular disease.

Other exposure variables included less routine biomarkers, pulse wave velocity, ketosis/ketoacidosis and other disease indices, but these were each single studies precluding grouping. All data are summarised in Table 1.

Use of complex approaches to subclassify type 2 diabetes

Description of extracted studies: There were 62 studies of complex/ML approaches to type 2 diabetes subclassification in a total of 793,291 participants (Table 2). Over half of the studies included non-European ancestry in relevant proportions (>20%). Only ~30% (19 out of 62) of the studies analysed participants with new-onset diabetes. Mean diabetes duration ranged from recent onset (within 1 year) to over 36 years. Most data were from observational studies (46 out of 62), with some post-hoc analyses of clinical trials (10), survey data (4) and mixed study types (2). Half of the studies had prospective design (31 out of 62) with a mean follow-up duration ranging from 1 year to 11.6 years. K-means clustering was the most applied ML approach (30 out of 62). Eight studies used established centroids⁸ to assign participants to clusters. Two studies decomposed combinations of genetic variants and their association with clinical and laboratory phenotypes into genotype-phenotype clusters by using Bayesian non-negative matrix factorisation.

Description of the categorised subgroups: Following the seminal work by Ahlqvist et al.⁸, multiple studies used the variables derived at time of diabetes diagnosis: age, HbA1c, BMI, HOMA2-B, HOMA2-IR and GAD-65 antibody (Table 2). The majority of these studies employed C-peptide-based homeostasis model assessment indices (HOMA, or its updated variant, HOMA2, using fasting insulin and glucose), as surrogates for insulin resistance (HOMA2-IR) and insulin secretion (HOMA2-B). In different contexts and populations, 22 studies replicated identification of the four non-autoimmune diabetes subtypes first described by Ahlqvist et al.⁸: severe insulin-deficient diabetes (SIDD), severe insulin-resistant diabetes (SIRD), mild obesity-

related diabetes (MOD), and mild age-related diabetes (MARD). The subset of studies including measurements of GAD antibody also identified the fifth cluster, severe autoimmune diabetes (SAID). Associations of these subtypes with clinical outcomes, including glycaemia, microvascular and macrovascular outcomes, and death, were replicated in 12 studies (Table 3).

Thirteen additional papers used variations of the original set of variables from Ahlqvist et al.⁸ by substituting HOMA with C-peptide, adding lipid traits, e.g. HDL-cholesterol, or approximating the clusters from different/simplified variable sets by applying advanced statistical learning approaches such as self-normalising neural networks. These approaches identified some type 2 diabetes subgroups resembling the clusters from Ahlqvist et al. and also novel subgroups related to the additional variables (Fig. 3). Several of the novel subgroups were associated with clinical outcomes. However, these findings have not been replicated in other studies (Table 2).

Additional papers ($n = 27$) assessed various sets of phenotypic inputs for ML approaches. Grouped into five categories of inputs, studies identified many subtypes and associations with clinical outcomes, however, they all lacked replication (Table 2). Four papers applied complex ML methods to a set of less than ten clinical variables such as systolic blood pressure, waist circumference, BMI, fasting plasma glucose, and age at diabetes diagnosis, and resulting subgroups were variably associated with outcomes, such as mortality. Eleven studies used a larger set of more than ten clinical features as inputs for classification, including data from electronic health records^{29,30}, and identified subgroups variably associated with clinical outcomes, including risk of cardiovascular disease. Two other studies specifically employed cardiovascular traits, including ECG³¹ and echocardiographic³² for ML algorithm inputs, and each identified subgroups with different associations with risk of cardiovascular disease. Finally, four studies involved inputs of change of glycaemic variables (HbA1c trajectories, glycaemia during a mixed meal test, continuous glucose monitoring features)^{33–35}, one study focused on fasting GLP-1, GIP and ghrelin levels³⁶, and two studies focused on behavioural traits such as novelty seeking, harm avoidance, and hospital anxiety and depression scale.

Human genetic risk information is rapidly penetrating clinical medicine. Two sets of papers utilised genomic data to identify diabetes subtypes, either in the form of inherited common genetic variation^{10,37} or gene expression data from muscle biopsies³⁸ (Table 2). The first approach clustered genetic variants with clinical traits associated with type 2 diabetes to identify subsets of variants predicted to act in shared mechanistic processes. Using these sets of genetic variants, process-specific or partitioned polygenic scores were constructed in individuals with type 2 diabetes and were associated with differences in clinical features and prevalence of metabolic outcomes, with replication across multiple cohorts. The muscle gene expression study has not been replicated. Overall, half of the studies had cross-sectional designs, and the other half involved prospective follow-up (Table 2).

Quality assessment. For simple approaches, of the 51 studies assessed, 55% were quality graded as very low-, or low-GRADE certainty, 45% had moderate certainty and none achieved high certainty. For complex approaches, around 70% of the studies had moderate evidence certainty. In both approaches, the majority of the studies had moderate or lower GRADE certainty on account of the (1) study design not addressing precision medicine objectives (not an RCT testing differential treatment effects in subclassified type 2 diabetes groups), (2) lack of a meaningful clinical outcome (i.e. although subgroups of type 2 diabetes were found, the measured outcome had little clinical significance because the study was not designed to study this) (3) Confidence in the findings were low

Table 2 Summary of published studies using complex approaches to type 2 diabetes classification.

Approach Total <i>n</i> % Female Mean age/age range	% Race/ethnicity breakdown	% Prevalent or new onset T2D, <i>n</i> (%) % Prospective or Cross sectional, <i>n</i> (%)	Machine learning approach	T2D subclassified Groups identified	Outcome category, <i>n</i> (%)	Overall quality, <i>n</i> (%)	References
Complex: Ahlqvist and directly replicated Ahlqvist clusters (22 studies) <i>n</i> = 88,197 43% female 55.3 years	81% Non-Hispanic White, 11% East Asian, 4% Hispanic, 3% South Asian, <1% Black, <1% Other	Prevalent (<i>n</i> = 11, 50%), New onset (<i>n</i> = 11, 50%) Longitudinal (<i>n</i> = 8, 36.6%), Cross- sectional (<i>n</i> = 14, 63.6%)	100% k-means	SAID, SIDD, SIRD, MOD, MARD	Microvascular & macrovascular events (<i>n</i> = 9, 41%), Clinical and biochemical traits (<i>n</i> = 4, 18%) Microvascular events only (<i>n</i> = 3, 13%), Glycaemia (<i>n</i> = 2, 9%), Macrovascular events only (<i>n</i> = 1, 5%), omic (<i>n</i> = 1, 5%), Other (<i>n</i> = 2, 9%)	Very Low (<i>n</i> = 1, 5%), Low (<i>n</i> = 3, 13%), Moderate (<i>n</i> = 18, 82%)	8,40,89-108
Complex: Similar to Ahlqvist clusters (13 studies) <i>n</i> = 214,093 45% female 58.6 years	72% Non-Hispanic White, 12% Hispanic, 10% South Asian, 4% East Asian, 2% Black, <1% Native American, <1% Other	Prevalent (<i>n</i> = 11, 85%), New onset (<i>n</i> = 2, 15%) Longitudinal (<i>n</i> = 7, 54%), Cross sectional (<i>n</i> = 6, 46%)	a) Addition of complementary clinical variables (i.e., HDL, TG, waist circumference, uric acid, etc.) b) Incorporating new clustering, i.e., self-normalising neural networks trained on k-means clustering. c) Addition of ethnic-specific thresholds for BMI.	SAID, SIDD, SIRD, MOD, and MARD. Five clusters: Older onset, Severe hyperglycaemia, Severe obesity, Younger Onset, and Insulin use. Four clusters: 42% (older onset), 14% (poor glucose control), 24% (severe obesity), and 20% (younger- onset). New subgroups MD, EOIDD, EOIRD, LOIDD, LOIRD, variable	Microvascular & macrovascular (<i>n</i> = 5, 38%), Microvascular events only (<i>n</i> = 3, 23%), Macrovascular events (<i>n</i> = 1, 8%), Glycaemia (<i>n</i> = 1, 8%), Other (<i>n</i> = 3, 23%)	Very Low (<i>n</i> = 1, 8%), Low (<i>n</i> = 5, 38%) Moderate (<i>n</i> = 7, 54%)	41,52,54,59,109-117
Complex: Simple clinical features (4 studies) <i>n</i> = 22,296 46.5% female 56 years	34% Non-Hispanic White, 25% Asian, 25% Middle Eastern, 9% Black, 5% Hispanic, 1% Native American/American Indian/Alaskan Native, <1% Other	Prevalent T2D (<i>n</i> = 3, 75%), New onset (<i>n</i> = 1, 25%) Longitudinal (<i>n</i> = 3, 75%), Cross-sectional (<i>n</i> = 1, 25%)	50% k-means (<i>n</i> = 2), 50% other (<i>n</i> = 2)	Mortality (<i>n</i> = 2, 50%), Cardiovascular events (<i>n</i> = 1, 25%), Clinical and biochemical traits (<i>n</i> = 1, 25%)	Moderate (<i>n</i> = 3, 75%), Low (<i>n</i> = 1, 25%)	118-121	
Complex: Complex clinical features (11 studies) <i>n</i> = 386,889 46.8% female 60 years	57% Non-Hispanic White, 24% Asian, 8% Black, 6% Hispanic, <5% Other	Prevalent T2D (<i>n</i> = 9, 82%), New onset (<i>n</i> = 2, 18%) Longitudinal (<i>n</i> = 7, 64%), Cross-sectional (<i>n</i> = 4, 36%)	variable	Cardiovascular events (<i>n</i> = 4, 36%), Glycaemic control (<i>n</i> = 3, 27%), Complications other than CVD event (<i>n</i> = 2, 18%), Mortality (<i>n</i> = 1, 9%), Other (<i>n</i> = 1, 9%)	Moderate (<i>n</i> = 9, 82%), Low (<i>n</i> = 2, 18%)	29,30,122-130	
Complex: Cardiovascular features (2 studies) <i>n</i> = 974 55.4% female 63 years	100% Non-Hispanic White	Prevalent T2D (<i>n</i> = 2, 100%) Longitudinal (<i>n</i> = 2, 100%)	Factor analysis (clustering) (<i>n</i> = 1, 50%), Hierarchical clustering (<i>n</i> = 1, 50%)	3 and 4 clusters	Moderate (<i>n</i> = 1, 50%), Low (<i>n</i> = 1, 50%)	31,32	
Complex: Behavioural features (2 studies) <i>n</i> = 653 40.5% female 63.5 years	48% Non-Hispanic White, 50% Asian, 2% others	Prevalent (<i>n</i> = 1, 50%), New onset (<i>n</i> = 1, 50%) Longitudinal (<i>n</i> = 1, 50%), Cross sectional (<i>n</i> = 1, 50%)	clustering, hierarchical clustering	2 and 4 clusters	Glycaemic control (<i>n</i> = 2, 100%)	Moderate (<i>n</i> = 1, 50%), Low (<i>n</i> = 1, 50%)	131,132

Table 2 (continued)

Approach Total <i>n</i> % Female Mean age/age range	% Race/ethnicity breakdown	% Prevalent or new onset T2D, <i>n</i> (%) % Prospective or Cross sectional, <i>n</i> (%)	Machine learning approach	T2D subclassified Groups identified	Outcome category, <i>n</i> (%)	Overall quality, <i>n</i> (%)	References
Complex: Glycemic features (4 studies) <i>n</i> = 67,064 42.8% female 62.5 years	40.6% Non-Hispanic White, 25% Asian, 25% Middle Eastern, 9.4% Other	Prevalent (<i>n</i> = 3, 75%), New onset (<i>n</i> = 1, 25%) Longitudinal (<i>n</i> = 3, 75%), Cross sectional (<i>n</i> = 1, 25%) Prevalent T2D (<i>n</i> = 3, 100%) Cross sectional (<i>n</i> = 3, 100%)	50% k-means, 25% latent-class analysis, 25% hierarchical clustering	3 and 4 clusters	Glycaemic control (<i>n</i> = 2, 50%), Cardiovascular events (<i>n</i> = 2, 50%)	Moderate (<i>n</i> = 3, 75%), Very low (<i>n</i> = 1, 25%)	33-35,133
Complex: Genetics (3 studies) <i>n</i> = 42,952	100% Non-Hispanic White	New onset (<i>n</i> = 3, 100%) Cross sectional (<i>n</i> = 3, 100%)	Bayesian Non- negative Matrix Factorisation (<i>n</i> = 2, 67%), Hierarchical clustering (<i>n</i> = 1, 33%) Two-step cluster analysis using log-likelihood distance measures	5 clusters of variant- trait associations: 3 clusters of skeletal dysregulated genes/ pathways in people with diabetes	Coronary artery disease, stroke, renal disease	Moderate (<i>n</i> = 2, 67%), Low (<i>n</i> = 1, 33%)	10,37,38
Complex: Hormonal (1 study) <i>n</i> = 96 participants 53% female 62 years	100% Non-Hispanic White	New onset (<i>n</i> = 1, 100%) Cross sectional (<i>n</i> = 1, 100%)	Two-step cluster analysis using log-likelihood distance measures	Two clusters (cluster 1: low GLP-1 and Ghrelin; cluster 2: high GLP-1 and Ghrelin)	Glycemia (<i>n</i> = 1, 100%)	Moderate (<i>n</i> = 1, 100%)	36

T2D type 2 diabetes, GAD glutamic decarboxylase antibody, LUCPGR urine C-peptide to creatinine ratio, BMI body mass index, CVD cardiovascular disease, OGTT oral glucose tolerance test, CKD chronic kidney disease, LDL low density lipoprotein cholesterol, SAID severe autoimmune diabetes, SIDD severe insulin deficient diabetes, SIRD severe insulin resistant diabetes, MOD mild obesity-related diabetes, MOD mid obesity-related diabetes, IMARD mid age-related diabetes, EOAD early-onset insulin resistant diabetes, LOIRD late-onset insulin deficient diabetes, LOIRD late-onset insulin resistant diabetes.

due to small sample sizes, lack of replication or lack of diversity of studied subgroups and (4) the potential for bias was large due to lack of adjustment for possible confounders.

Discussion

Summary of findings. This systematic review analysed two broad approaches to the subclassification of type 2 diabetes to identify clinically meaningful subtypes that may advance precision diagnostics. We found many simple stratification approaches using, for example, clinical features such as BMI, age at diagnosis, and lipid levels, but none had been replicated and many lacked associations with clinical outcomes. Complex stratification models using ML approaches with and without genetic data showed reproducible subtypes of type 2 diabetes associated with outcomes. Both approaches require a higher grade of evidence but support the premise that type 2 diabetes can be subclassified into clinically meaningful subtypes.

Simple approaches to subclassification included urine and blood biomarkers, anthropometric measures, clinical data such as age at diagnosis, surrogate beta-cell metrics derived from blood C-peptide or insulin along with other less diabetes-related biomarkers such as bilirubin levels or pulse wave velocity. Approaches to subclassification were diverse. Some studies dichotomised continuous variables based on clinical cut-points. Other studies used a composite exposure (two or more criteria each with cut-points) or analysed changes in continuous variables over time e.g. change in eGFR over time.

The study designs, specific cut-offs and outcomes were heterogenous, and no studies met high-quality GRADE certainty. No study evaluating a simple approach to type 2 diabetes subtyping has been adequately reproduced, although some studies identified biologically plausible subgroups. For example, subclassifications derived using BMI, beta-cell function, lipid profiles and age appeared to be associated with some outcomes which could be helpful in clinical practice. These potential subclassifications need to be replicated in better-designed studies (see section on additional supporting literature). Other evidence not included in our systematic review (either due to the study population including people without diabetes or the analysis was only performed in people with the exposure without a comparison group), support the role of simple variables in stratifying diabetes; for example, younger age at diagnosis is reproducibly associated with worse cardiorenal outcomes in a number of studies³⁹.

Machine learning approaches yielded some reproducible subtypes of type 2 diabetes using a variety of clinical and genetic variables. The best-replicated subtypes were the clusters first described by Ahlqvist et al.⁸, which were replicated in 22 studies, including ~88,000 individuals of diverse ancestry. There also was replication of genetic subtypes of type 2 diabetes from Udler et al.¹⁰ with associations with clinical features seen in multiple cohorts across almost 454,000 individuals³⁶. However, the latter associations involved small absolute effects with unclear clinical utility for individual patient management, and studies were restricted to individuals of European ancestry. While there was replication of the clusters from Ahlqvist et al. across studies, the generated clusters appeared to be dependent on the characteristics of the underlying populations, especially factors such as distribution of ancestry, age, duration of diabetes, anthropometric trait variability as in BMI, and the variety of variable terms included in learning models. Nevertheless, at least some of the resulting subtypes appeared to be robust to differences in specific ML method, input variables, and populations (Fig. 3).

Many of the input variables for the complex ML subtyping approaches were also used in studies involving simple approaches to subclassification, recapitulating the biological plausibility of specific

Table 3 Association of the Ahlqvist-clusters with outcomes from 22 reviewed studies using consistent cluster assignment methods.

	Glycemic outcomes	Microvascular outcomes	Macrovascular outcomes	Other outcomes	Death
SIDD	Insulin requirement ^{8,97}	Retinopathy ^{8,97,102-104} (n = 3160) Nephropathy ^{97,103} (n = 1074)	Lower extremity arterial disease ¹⁰³ (n = 223)	Distal symmetric polyneuropathy, Cardiac autonomic neuropathy ⁹⁰ (n = 28), Erectile dysfunction ⁹³ (n = 4) Prevalence of NAFLD ⁹⁰ (n = 121), Erectile dysfunction ⁹³ (n = 7), Diabetic peripheral neuropathy ¹⁰³ (n = 225) Neuropathy ⁹⁷ (n = 1258)	
SIRD	Glycemic benefit with thiazolidinedione therapy ⁴⁰ (n = 800)	Diabetic kidney disease ^{8,90,98,100,103,107} (n = 2686)	MACE (confounded by age and sex ⁸ , n = 1373)		
MOD MARD	Glycemic benefit with sulfonylurea therapy ⁴⁰ (n = 1361)	Retinopathy ¹⁰⁰ (n = 1487)	MACE ^{8,97} (n = 3513)		CVD mortality ¹⁰² (n = 56)

SIDD severe autoimmune diabetes, SIDD severe insulin deficient diabetes, SIRD severe insulin resistant diabetes, MOD mild obesity-related diabetes, MARD mild age-related diabetes.

clustering variables in defining type 2 diabetes subtypes. One study directly compared a simple clinical approach to the clustering approach from Ahlqvist et al.⁸ and found that simple single clinical measures analysed in a quantitative (rather than categorical) framework could better predict relevant clinical outcomes, such as incidence of chronic kidney disease and glycaemic response to medications⁴⁰. Thus, further research is needed to determine whether assigning a patient to one of the clusters from Ahlqvist et al.⁸ offers additional clinical benefit beyond evaluation of simple clinical measures and also beyond current standard of care. For example, high quality randomised controlled trial evidence is needed to demonstrate that knowledge of a patient’s clinical or genetic cluster membership could meaningfully guide treatment and/or clinical care and improve outcomes.

Study quality. No studies included in our systematic review had above moderate certainty of evidence. Some strengths of included studies were the large sample sizes, the diversity of variables considered, and inclusion of both prevalent and new-onset cases of type 2 diabetes. However, the varied study designs and lack of replication limits our ability to draw firm conclusions about the most effective approaches to subclassification. Most variables used for subclassification capture momentary metabolic states, which limits their long-term utility as cluster assignment is likely to change over time^{41,42}. Most studies were retrospective analyses of established cohorts, and there were, at the time of the search, no data available involving subtype-stratified clinical trials or real world implementation of approaches. Finally, most studies focused on European-ancestry populations, and the clinical value of these approaches may vary across different ancestries. While East Asian ancestries had representation in some studies, research in Black, South Asian and Hispanic populations remains sparse. This is particularly important, as four out of five people with type 2 diabetes come from marginalised groups or live in low- or middle-income countries. Future precision diagnostic interventions should address and narrow inequalities.

Additional supporting literature. Since our literature search was conducted, four new publications have advanced our understanding of type 2 diabetes subclassification.

Two recent studies applied ML approaches to stratify diabetes heterogeneity, both considering continuous approaches rather than with discrete clusters^{43,44}. Nair et al. used a non-linear transformation and visualisation of nine variables onto a tree-like structure⁴⁴ and with replication in two large datasets. This approach linked underlying disease heterogeneity to risk of complications; those at risk of cardiovascular disease had a different phenotype to those with microvascular complications and to drug response and demonstrated associations of gradients across the tree using genetic process-specific scores from Udler et al.¹⁰ Wesolowska-Andersen et al. performed soft-clustering from 32 clinical variables which yielded 4 diabetes archetypes comprising a third of the study population. The remaining study population was deemed as mixed-phenotype. This study has not been replicated⁴³. A third study re-identified the genetic subtypes and their clinical associations from Udler et al.⁴⁵.

Additionally, one of the first clinical trials to assess precision medicine approaches for diabetes management was published after our literature search. The TriMaster Study tested dichotomised BMI and eGFR strata in a three-period crossover trial using three pharmacologic interventions with the primary hypothesis being stratum-specific differences in HbA1c⁴⁶. Participants with obesity (BMI > 30 kg/m²) showed a glycaemic benefit on pioglitazone versus sitagliptin and participants with lower eGFR (60–90 ml/min/1.73 m²) responded with lower HbA1c to sitagliptin as compared

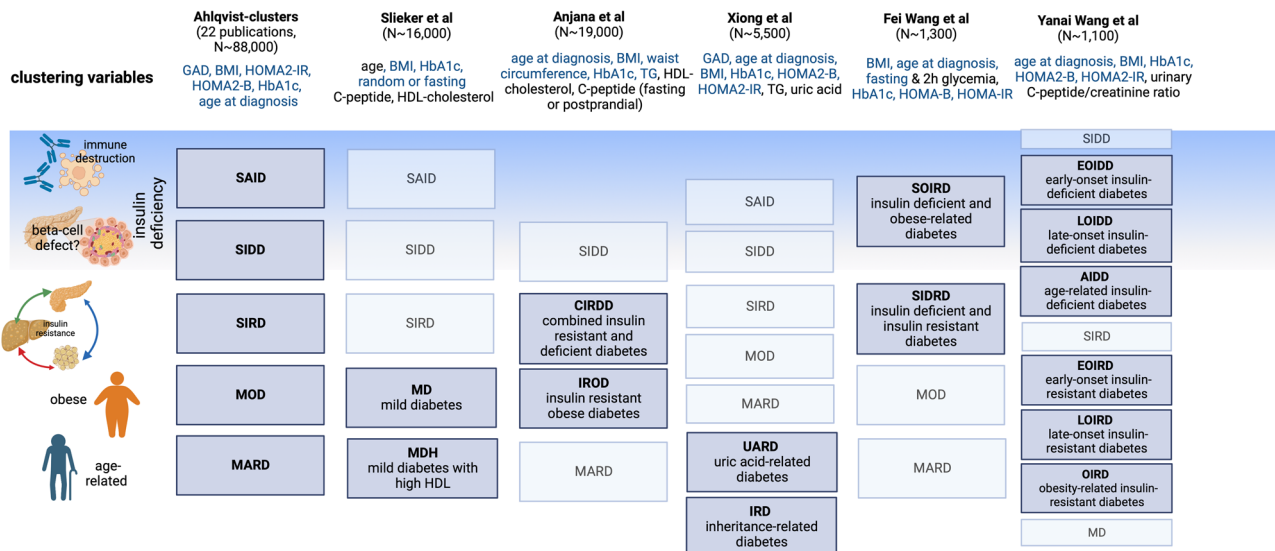


Fig. 3 Main characteristics of diabetes clusters derived using a modified set of clustering variables, compared to original ‘Ahlqvist’ clusters. Clustering variables denoted in blue are consistent across the different studies, those in black are unique to the particular study outlined. A greyed-out box indicates that the indicated diabetes cluster was replicated from the Ahlqvist study, a dark blue box indicates a new diabetes cluster. GAD, glutamic decarboxylase antibody; BMI, body mass index; HDL, high-density lipoprotein cholesterol; HOMA2-IR/B, homoeostasis model assessment-2 insulin resistance/beta cell function. SAID, severe autoimmune diabetes; SIDD, severe insulin-deficient diabetes; SIRD, severe insulin resistant diabetes; MOD, mild obesity-related diabetes; MARD, mild age-related diabetes.

to canagliflozin. In a secondary analysis, drug-choice corresponding to patient preferences yielded lower glycemias than a random allocation, suggesting that listening to patients is critical in informing therapeutic decisions⁴⁷. Ramifications of this study are limited by the non-comparable pharmacologic doses used, and the primary focus on glycaemia which may not be indicative of long-term therapeutic success and/or prevention of complications. Yet these studies have generated higher quality evidence linking type 2 diabetes heterogeneity to treatment and disease outcomes. It remains to be seen if these can be replicated in other ancestries and translated into ‘usable products’ for healthcare professionals.

It is worth noting that ketosis-prone type 2 diabetes, an established type 2 diabetes subtype, was not captured adequately in our systematic review: only one study included ketosis-prone type 2 diabetes as an exposure⁴⁸. Study designs for ketosis-prone type 2 diabetes were usually analyses of cohorts with diabetic ketoacidosis at presentation with type 1 diabetes as the outcome, rather than as an exposure in people with type 2 diabetes. Since our search was designed to identify studies stratifying type 2 diabetes, this literature was not captured. Like many other ‘simple’ criteria for classification, the characteristics of people with diabetic ketoacidosis at presentation of type 2 diabetes have been studied, but with few prospective studies that have been replicated⁴⁹.

Age at diagnosis as a simple approach to stratification also did not feature strongly in our search results. The body of literature that outlines higher risk of microvascular or macrovascular complications in early-onset type 2 diabetes has focussed on comparing people with type 2 diabetes to those without diabetes in different age groups^{39,50} or studied cohorts of early-onset cases in isolation⁵¹ and, thus, would not have been captured in our search strategy. Recent epidemiological studies have compared outcomes between early and late age onset strata^{52,53} showcasing higher risks of cardiorenal outcomes with early age at onset, but these were retrospective analyses of health record databases, potentially confounded by age-related risk of complications and duration of diabetes. To move forward, prospective studies stratifying different interventions (e.g., tighter treatment targets or better cardiovascular risk reduction) in those diagnosed at younger age, are needed.

Findings in context. We found that simple features have not been precisely and reproducibly evaluated to a high enough standard to subclassify type 2 diabetes into subtypes. This is not surprising, as many studies were not necessarily conducted for the purpose of ‘precision diagnosis’, but rather as studies of clinical phenotypes spanning a time period that preceded the current research focus on precision medicine. It is important to re-emphasise that many of the simple clinical criteria studied, do have other bodies of evidence supporting associations with outcomes, like age -at -diagnosis. While these studies have set the scene, the field needs more robust evidence.

‘Complex’ methods for diabetes subclassification have shown better reproducibility, have been linked to a variety of meaningful clinical outcomes more consistently, and more recently have been able to demonstrate differential treatment responses related to stratification.

What do these findings mean for a precision medicine approach to type 2 diabetes diagnosis? Ideally, subclassification strategies should be deployed at diagnosis of type 2 diabetes on the basis of measured clinical characteristics such that people in different subgroups of type 2 diabetes could be treated differently. One key question is whether such efforts would cost-effectively improve clinical outcomes, compared to the current standard of care. However, another more fundamental question is whether subclassification approaches at diagnosis alone are enough? For example, another approach may be to iteratively subclassify longitudinal disease trajectories. Such an approach is supported by studies that have shown cluster-based assignments of type 2 diabetes at diagnosis are not robust and may change over time⁵⁴. It may be argued that subclassification at one-time point is overly simplistic and should be regularly reviewed based on trajectory.

Irrespective of the subclassification approach studied, they need replication in independent datasets, assessment in diverse populations, in people with both new-onset and prevalent diabetes, and investigation using prospective data, ideally in the form of randomised clinical trials. Clinical trials of treatment approaches tailored to diabetes subtypes will be necessary to understand the clinical benefits of clinical subtyping. Ideally, sub-phenotyping

should lead to benefits for patients in real-world clinical settings. Conducting these studies will be challenging due to the necessity for extensive follow-up, large sample sizes, and substantial resource requirements. There is a pressing need for innovative strategies to generate high-quality evidence on treatment options tailored to specific diabetes subtypes in diverse populations. These data will be critical to determine generalisability of findings and amenability for clinical translation including in resource-constrained settings.

Clinical applicability. The current evidence supports distinguishable subtypes of type 2 diabetes and that these subtypes are associated with variation in clinical outcomes. However, the very low to moderate quality of existing studies and the need for replication in ancestry-diverse studies make it difficult to identify a strongly evidence-based, universally applicable approach.

The most clinically valuable methods are likely to be those that are easy and inexpensive to implement. For more complex approaches, computer decision support tools will need to be developed and assessed for feasibility and utility. Although the evidence supporting complex approaches has leap-frogged the evidence in favour of more simplified approaches, there is still likely a place for simple approaches that can be more accessible at diverse clinical interfaces. Meanwhile how cluster assignment could be translated into actionable data for the individual remains unclear; will for example, a given person with type 2 diabetes exist in a distinct subgroup with associated outcomes or will the subtype of type 2 diabetes have associated probabilities or risks of certain outcomes? While stratifying people with type 2 diabetes into discrete subtypes might result in information loss, compared to continuous risk modelling⁴⁰, discrete clusters might inform clinical decisions⁴².

Limitations. The limitations of this review reflect the limitations of the literature. To manage the breadth of literature analysed in this systematic review, focussed on genomic data and did not include proteomic or metabolomic data as these are potentially more premature for clinical use. We also did not include studies on participants at risk of type 2 diabetes, although we recognise that a body of evidence is emerging to stratify type 2 diabetes incidence risk using multiple approaches that are similar to those for established type 2 diabetes. Since we focused on studies that attempted to subgroup type 2 diabetes, we also did not capture analyses of independent cohorts with a particular type 2 diabetes phenotype at baseline, for example, studies of young people with type 2 diabetes or those with ketosis-prone type 2 diabetes, as outlined.

Next steps and recommendations. Future research should aim to identify and validate clinically useful and cost-effective methods for type 2 diabetes subclassification that can be applied across diverse populations. Such research will involve replication of a given approach in independent datasets, including from diverse ancestral populations, to ensure generalisability that doesn't widen health disparities. For simple stratification approaches, there is still much that can be done—agreement on standardised study designs for precision diagnostics studies could be a first step. For ML requiring real-time computation, the development of strategies to overcome local resource constraints in implementing these methods could be explored.

Conclusion

In this first systematic review of the evidence underpinning type 2 diabetes diagnostic subclassification, multiple approaches were identified. Among them are strategies that used simple criteria based on fundamental categorisation of mostly routine measures, and complex approaches with multi-trait or genetic inputs that required ML or other computation. While simple approaches are more easily

deployed, the study designs and level of evidence currently limits any firm conclusions regarding the utility of such approaches. The clinical variables and data incorporated into 'complex' approaches have yielded reproducible subclassifications and a growing body of evidence supports clinically meaningful associations of subtypes with outcomes and treatment responses. This is a rapidly evolving field with higher quality evidence emerging. It will be crucial to develop interventions that target diverse populations and be feasible in all resource settings to prevent widening existing inequalities in the precision medicine era of diabetes care.

Data availability

The extracted data from full-text articles included in this systematic review are available in Supplementary Data 1.

Received: 9 May 2023; Accepted: 15 September 2023;

Published online: 05 October 2023

References

1. Diabetes Atlas 10th Edition, World Health Organisation. <https://diabetesatlas.org/atlas/tenth-edition/>, accessed 6 Jun 2023 (2021).
2. ElSayed, N. A. et al. Classification and diagnosis of diabetes: standards of care in diabetes-2023. *Diabetes Care* **46**, S19–S40 (2023).
3. Davies, M. J. et al. Management of hyperglycemia in type 2 diabetes, 2022. A consensus report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). *Diabetes Care* **45**, 2753–2786 (2022).
4. Parameswarappa, D. C. et al. Severity of diabetic retinopathy and its relationship with age at onset of diabetes mellitus in India: a multicentric study. *Indian J. Ophthalmol.* **69**, 3255–3261 (2021).
5. Prando, R. et al. Is type 2 diabetes a different disease in obese and nonobese patients? *Diabetes Care* **21**, 1680–1685 (1998).
6. van Hateren, K. J. J. et al. The lipid profile and mortality risk in elderly type 2 diabetic patients: a ten-year follow-up study (ZODIAC-13). *PLoS One* **4**, e8464 (2009).
7. Shepherd, M. H. et al. A UK nationwide prospective study of treatment change in MODY: genetic subtype and clinical characteristics predict optimal glycaemic control after discontinuing insulin and metformin. *Diabetologia* **61**, 2520–2527 (2018).
8. Ahlqvist, E. et al. Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables. *Lancet Diabetes Endocrinol.* **6**, 361–369 (2018).
9. Mahajan, A. et al. Refining the accuracy of validated target identification through coding variant fine-mapping in type 2 diabetes. *Nat. Genet.* **50**, 559 (2018).
10. Udler, M. S. et al. Type 2 diabetes genetic loci informed by multi-trait associations point to disease mechanisms and subtypes: a soft clustering analysis. *PLoS Med.* **15**, e1002654 (2018).
11. Nolan, J. J. et al. ADA/EASD precision medicine in diabetes initiative: an international perspective and future vision for precision medicine in diabetes. *Diabetes Care* **45**, 261–266 (2022).
12. Tobias, D. Second international consensus report on gaps and opportunities for the clinical translation of precision diabetes medicine. *Nat. Med.* <https://doi.org/10.1038/s41591-023-02502-5>. (2023).
13. Page, M. J. et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* **372**, n71 (2021).
14. Guyatt, G. H. et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* **336**, 924–926 (2008).
15. What is GRADE? | BMJ Best Practice. <https://bestpractice.bmj.com/info/us/toolkit/learn-ebm/what-is-grade/>.
16. Valensi, P. et al. Atherogenic dyslipidemia and risk of silent coronary artery disease in asymptomatic patients with type 2 diabetes: a cross-sectional study. *Cardiovasc. Diabetol.* **15**, 104 (2016).
17. Huang, J. et al. Elevated serum small dense low-density lipoprotein cholesterol may increase the risk and severity of coronary heart disease and predict cardiovascular events in patients with type 2 diabetes mellitus. *Dis. Mark.* **2021**, 5597028 (2021).
18. Zhao, Q. et al. Comparison of various insulin resistance surrogates on prognostic prediction and stratification following percutaneous coronary intervention in patients with and without type 2 diabetes mellitus. *Cardiovasc. Diabetol.* **20**, 190 (2021).
19. Wang, S. et al. Stronger association of triglyceride glucose index than the HOMA-IR with arterial stiffness in patients with type 2 diabetes: a real-world single-centre study. *Cardiovasc. Diabetol.* **20**, 82 (2021).

20. Turner, R. et al. UKPDS 25: autoantibodies to islet-cell cytoplasm and glutamic acid decarboxylase for prediction of insulin requirement in type 2 diabetes. UK Prospective Diabetes Study Group. *Lancet* **350**, 1288–1293 (1997).
21. Lee, S. A. et al. Progression to insulin deficiency in Korean patients with Type 2 diabetes mellitus positive for anti-GAD antibody. *Diabet. Med.* **28**, 319–324 (2011).
22. Tuomi, T. et al. Clinical and genetic characteristics of type 2 diabetes with and without GAD antibodies. *Diabetes* **48**, 150–157 (1999).
23. Buzzetti, R. et al. High titer of autoantibodies to GAD identifies a specific phenotype of adult-onset autoimmune diabetes. *Diabetes Care* **30**, 932–938 (2007).
24. Ipadeola, A., Adeleye, J. O. & Akinlade, K. S. Latent autoimmune diabetes amongst adults with type 2 diabetes in a Nigerian tertiary hospital. *Prim. Care Diabetes* **9**, 231–236 (2015).
25. Dennis, J. M. et al. Precision medicine in type 2 diabetes: clinical markers of insulin resistance are associated with altered short- and long-term glycaemic response to DPP-4 inhibitor therapy. *Diabetes Care* **41**, 705–712 (2018).
26. Faraz, A., Ashraf, H. & Ahmad, J. Clinical features, biochemical profile, and response to standard treatment in lean, normal-weight, and overweight/obese Indian type 2 diabetes patients. *Rev. Diabet. Stud.* **17**, 68–74 (2021).
27. Mohammedi, K. et al. Associations between body mass index and the risk of renal events in patients with type 2 diabetes. *Nutr. Diabetes* **8**, 7 (2018).
28. Arslanian, S. et al. The shape of the glucose response curve during an oral glucose tolerance test: forerunner of heightened glycemic failure rates and accelerated decline in β -cell function in TODAY. *Diabetes Care* **42**, 164–172 (2019).
29. Li, L. et al. Identification of type 2 diabetes subgroups through topological analysis of patient similarity. *Sci. Transl. Med.* **7**, 311ra174 (2015).
30. Kirk, I. K. et al. Linking glycaemic dysregulation in diabetes to symptoms, comorbidities, and genetics through EHR data mining. *eLife* **8**, e44941 (2019).
31. Kuusisto, J., Lempiäinen, P., Mykkänen, L. & Laakso, M. Insulin resistance syndrome predicts coronary heart disease events in elderly type 2 diabetic men. *Diabetes Care* **24**, 1629–1633 (2001).
32. Ernande, L. et al. Clinical implications of echocardiographic phenotypes of patients with diabetes mellitus. *J. Am. Coll. Cardiol.* **70**, 1704–1716 (2017).
33. Karpati, T. et al. Patient clusters based on HbA1c trajectories: a step toward individualized medicine in type 2 diabetes. *PLoS ONE* **13**, e0207096 (2018).
34. Tao, R. et al. Multilevel clustering approach driven by continuous glucose monitoring data for further classification of type 2 diabetes. *BMJ Open Diabetes Res. Care* **9**, e001869 (2021).
35. Mariam, A. et al. Type 2 diabetes subtype responsive to ACCORD intensive glycemia treatment. *Diabetes Care* <https://doi.org/10.2337/dc20-2700> (2021).
36. Amato, M. C., Pizzolanti, G., Torregrossa, V., Pantò, F. & Giordano, C. Phenotyping of type 2 diabetes mellitus at onset on the basis of fasting incretin tone: results of a two-step cluster analysis. *J. Diabetes Investig.* **7**, 219–225 (2016).
37. DiCorpo, D. et al. Type 2 diabetes partitioned polygenic scores associate with disease outcomes in 454,193 individuals across 13 cohorts. *Diabetes Care* **45**, 674–683 (2022).
38. Khoshnejat, M., Kavousi, K., Banaei-Moghaddam, A. M. & Moosavi-Movahedi, A. A. Unraveling the molecular heterogeneity in type 2 diabetes: a potential subtype discovery followed by metabolic modeling. *BMC Med. Genom.* **13**, 119 (2020).
39. Sattar, N. et al. Age at diagnosis of type 2 diabetes mellitus and associations with cardiovascular and mortality risks. *Circulation* **139**, 2228–2237 (2019).
40. Dennis, J. M., Shields, B. M., Henley, W. E., Jones, A. G. & Hattersley, A. T. Disease progression and treatment response in data-driven subgroups of type 2 diabetes compared with models based on simple clinical features: an analysis using clinical trial data. *Lancet Diabetes Endocrinol.* **7**, 442–451 (2019).
41. Lugner, M. et al. Comparison between data-driven clusters and models based on clinical features to predict outcomes in type 2 diabetes: nationwide observational study. *Diabetologia* **64**, 1973–1981 (2021).
42. Herder, C. & Roden, M. A novel diabetes typology: towards precision diabetology from pathogenesis to treatment. *Diabetologia* <https://doi.org/10.1007/s00125-021-05625-x> (2022).
43. Wesolowska-Andersen, A. et al. Four groups of type 2 diabetes contribute to the etiological and clinical heterogeneity in newly diagnosed individuals: an IMI DIRECT study. *Cell Rep. Med.* **3**, 100477 (2022).
44. Nair, A. T. N. et al. Heterogeneity in phenotype, disease progression and drug response in type 2 diabetes. *Nat. Med.* **28**, 982–988 (2022).
45. Kim, H. et al. High-throughput genetic clustering of type 2 diabetes loci reveals heterogeneous mechanistic pathways of metabolic disease. *Diabetologia* **66**, 495–507 (2023).
46. Shields, B. M. et al. Patient stratification for determining optimal second-line and third-line therapy for type 2 diabetes: the TriMaster study. *Nat. Med.* <https://doi.org/10.1038/s41591-022-02120-7> (2022).
47. Shields, B. M. et al. Patient preference for second- and third-line therapies in type 2 diabetes: a prespecified secondary endpoint of the TriMaster study. *Nat. Med.* <https://doi.org/10.1038/s41591-022-02121-6> (2022).
48. Lontchi-Yimagou, E. et al. Ketosis-prone atypical diabetes in Cameroonian people with hyperglycaemic crisis: frequency, clinical and metabolic phenotypes. *Diabet. Med.* **34**, 426–431 (2017).
49. Mauvais-Jarvis, F. et al. Ketosis-prone type 2 diabetes in patients of sub-Saharan African origin: clinical pathophysiology and natural history of beta-cell dysfunction and insulin resistance. *Diabetes* **53**, 645–653 (2004).
50. Wright, A. K. et al. Age-, sex- and ethnicity-related differences in body weight, blood pressure, HbA1c and lipid levels at the diagnosis of type 2 diabetes relative to people without diabetes. *Diabetologia* **63**, 1542–1553 (2020).
51. TODAY Study Group et al. Long-term complications in youth-onset type 2 diabetes. *N. Engl. J. Med.* **385**, 416–426 (2021).
52. Paul, S. K., Shaw, J. E., Fenici, P. & Montvida, O. Cardiorenal complications in young-onset type 2 diabetes compared between white Americans and African Americans. *Diabetes Care* **45**, 1873–1881 (2022).
53. Dibato, J. E. et al. Association of cardiometabolic multimorbidity and depression with cardiovascular events in early-onset adult type 2 diabetes: a multiethnic study in the U.S. *Diabetes Care* **44**, 231–239 (2021).
54. Bello-Chavolla, O. Y. et al. Clinical characterization of data-driven diabetes subgroups in Mexicans using a reproducible machine learning approach. *BMJ Open Diabetes Res. Care* **8**, e001550 (2020).
55. O’Gorman, D. J. et al. In vivo and in vitro studies of GAD-antibody positive subjects with Type 2 diabetes: a distinct sub-phenotype. *Diabetes Res. Clin. Pract.* **80**, 365–370 (2008).
56. Tajiri, Y., Kimura, M., Mimura, K. & Umeda, F. Variation of fasting serum C-peptide level after admission in Japanese patients with type 2 diabetes mellitus. *Diabetes Technol. Ther.* **11**, 593–599 (2009).
57. Stidsen, J. V. et al. Pathophysiology-based phenotyping in type 2 diabetes: a clinical classification tool. *Diabetes/Metab. Res. Rev.* **0**, e3005 (2018).
58. Anoop, S., Misra, A., Bhatt, S. P., Gulati, S. & Mahajan, H. High fasting C-peptide levels and insulin resistance in non-lean & non-obese (BMI > 19 to <25 kg/m²) Asian Indians with type 2 diabetes are independently associated with high intra-abdominal fat and liver span. *Diabetes Metab. Syndr.* **13**, 708–715 (2019).
59. Wang, Y. et al. Urinary C-peptide/creatinine ratio: a useful biomarker of insulin resistance and refined classification of type 2 diabetes mellitus. *J. Diabetes* **13**, 893–904 (2021).
60. Saxena, T., Maheshwari, S. & Goyal, R. K. Serum insulin assay: an important therapeutic tool in management of freshly diagnosed type 2 diabetes mellitus. *J. Assoc. Physicians India* **48**, 815–817 (2000).
61. Selvaraj, R., Rajappa, M. & Shamanna, S. B. Phenotypic characterization of patients with type 2 diabetes mellitus. *Int. J. Diabetes Dev. Ctries* **40**, 242–247 (2020).
62. Dennis, J. M. et al. Sex and BMI alter the benefits and risks of sulfonylureas and thiazolidinediones in type 2 diabetes: a framework for evaluating stratification using routine clinical and individual trial data. *Diabetes Care* **41**, 1844–1853 (2018).
63. Feng, W. et al. Lixisenatide is effective and safe as add-on treatment to basal insulin in Asian individuals with type 2 diabetes and different body mass indices: a pooled analysis of data from the GetGoal Studies. *BMJ Open Diabetes Res. Care* **9**, e002290 (2021).
64. Miñambres, I. et al. Characterization of the hypertriglyceridemic waist phenotype in patients with type 2 diabetes mellitus in Spain: an epidemiological study. *Rev. Clin. Esp.* **221**, 576–581 (2021).
65. Komptoi, M. et al. Elevated serum triglycerides is the strongest single indicator for the presence of metabolic syndrome in patients with type 2 diabetes. *Cardiovasc. Diabetol.* **5**, 21 (2006).
66. Radenković, S. P. et al. The hypertriglyceridemic waist phenotype and metabolic syndrome by differing criteria in type 2 diabetic patients and their relation to lipids and blood glucose control. *Endokrynol. Pol.* **62**, 316–323 (2011).
67. Fu, L. et al. Atherogenic index of plasma is associated with major adverse cardiovascular events in patients with type 2 diabetes mellitus. *Cardiovasc Diabetol.* **20**, 201 (2021).
68. Færch, K. et al. Trajectories of cardiometabolic risk factors before diagnosis of three subtypes of type 2 diabetes: a post-hoc analysis of the longitudinal Whitehall II cohort study. *Lancet Diabetes Endocrinol.* **1**, 43–51 (2013).
69. Arslanian, S. A. et al. OGTT glucose response curves, insulin sensitivity, and β -cell function in rise: comparison between youth and adults at randomization and in response to interventions to preserve β -cell function. *Diabetes Care* **44**, 817–825 (2021).
70. Mitsui, R. et al. Factors responsible for deteriorating glucose tolerance in newly diagnosed type 2 diabetes in Japanese men. *Metabolism* **55**, 53–58 (2006).
71. Greenfield, S. et al. Comorbidity affects the relationship between glycaemic control and cardiovascular outcomes in diabetes: a cohort study. *Ann. Intern. Med.* **151**, 854–860 (2009).
72. Roy Chowdhury, S. et al. Diabetic retinopathy in newly diagnosed subjects with type 2 diabetes mellitus: contribution of β -cell function. *J. Clin. Endocrinol. Metab.* **101**, 572–580 (2016).
73. Yun, J.-S. et al. Severe hypoglycemia and the risk of cardiovascular disease and mortality in type 2 diabetes: a nationwide population-based cohort study. *Cardiovasc. Diabetol.* **18**, 103 (2019).
74. Echouffo-Tcheugui, J. B. et al. Duration of diabetes and incident heart failure: the ARIC (Atherosclerosis Risk In Communities) study. *JACC Heart Fail.* **9**, 594–603 (2021).

75. Januszewicz, A. et al. Office and ambulatory pulse pressure–association with clinical characteristics and cardiovascular risk factors in normoalbuminuric patients with type 2 diabetes (ROADMAP study). *J. Hum. Hypertens.* **25**, 679–685 (2011).
76. Liu, J.-J. et al. Aortic pulse wave velocity, central pulse pressure, augmentation index and chronic kidney disease progression in individuals with type 2 diabetes: a 3- year prospective study. *BMC Nephrol.* **21**, 359 (2020).
77. Segar, M. W. et al. Prevalence and prognostic implications of diabetes with cardiomyopathy in community-dwelling adults. *J. Am. Coll. Cardiol.* **78**, 1587–1598 (2021).
78. Ihara, K. et al. Profibrotic circulating proteins and risk of early progressive renal decline in patients with type 2 diabetes with and without albuminuria. *Diabetes Care* **43**, 2760–2767 (2020).
79. Linnemann, B., Voigt, W., Nobel, W. & Janka, H. U. C-reactive protein is a strong independent predictor of death in type 2 diabetes: association with multiple facets of the metabolic syndrome. *Exp. Clin. Endocrinol. Diabetes* **114**, 127–134 (2006).
80. Fukui, M. et al. Relationship between serum bilirubin and albuminuria in patients with type 2 diabetes. *Kidney Int.* **74**, 1197–1201 (2008).
81. Jian, W. et al. Role of serum vaspin in progression of type 2 diabetes: a 2-year cohort study. *PLoS One* **9**, e94763 (2014).
82. Gong, S. et al. Clinical and genetic features of patients with type 2 diabetes and renal glycosuria. *J. Clin. Endocrinol. Metab.* **102**, 1548–1556 (2017).
83. Saleem, T. et al. The profile of plasma free amino acids in type 2 diabetes mellitus with insulin resistance: association with microalbuminuria and macroalbuminuria. *Appl. Biochem. Biotechnol.* **188**, 854–867 (2019).
84. Usman, M., Khunti, K., Davies, M. J. & Gillies, C. L. Association and relative importance of multiple risk factor control on cardiovascular disease, end-stage renal disease and mortality in people with type 2 diabetes: a population-based retrospective cohort study. *Prim. Care Diabetes* **15**, 218–226 (2021).
85. Feng, X. et al. Myocardial infarction and coronary artery disease in menopausal women with type 2 diabetes mellitus negatively correlate with total serum bile acids. *Front. Endocrinol.* **12**, 754006 (2021).
86. Warren, R. A. et al. Haptoglobin phenotype modifies the effect of fenofibrate on risk of coronary event: ACCORD lipid trial. *Diabetes Care* **45**, 241–250 (2022).
87. Yu, H. et al. Serum chromogranin A correlated with albuminuria in diabetic patients and is associated with early diabetic nephropathy. *BMC Nephrol.* **23**, 41 (2022).
88. Chang, Y.-S., Li, Y.-H. & Lee, I.-T. A synergistic effect of variability in estimated glomerular filtration rate with chronic kidney disease on all-cause mortality prediction in patients with type 2 diabetes: a retrospective cohort study. *Cardiovasc. Diabetol.* **20**, 209 (2021).
89. Mansour Aly, D. et al. Genome-wide association analyses highlight etiological differences underlying newly defined subtypes of diabetes. *Nat. Genet.* **53**, 1534–1542 (2021).
90. Zaharia, O. P. et al. 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 (2019).
91. Saito, K. et al. Usefulness of subclassification of adult diabetes mellitus among inpatients in Japan. *J. Diabetes Investig.* **13**, 706–713 (2022).
92. Saatmann, N. et al. Physical fitness and cardiovascular risk factors in the novel diabetes subgroups. *J. Clin. Endocrinol. Metab.* <https://doi.org/10.1210/clinem/dgab810> (2021).
93. Maalmi, H. et al. Differences in the prevalence of erectile dysfunction between novel subgroups of recent-onset diabetes. *Diabetologia* **65**, 552–562 (2022).
94. Zou, X., Zhou, X., Zhu, Z. & Ji, L. Novel subgroups of patients with adult-onset diabetes in Chinese and US populations. *Lancet Diabetes Endocrinol.* **7**, 9–11 (2019).
95. Pigeyre, M. et al. Validation of the classification for type 2 diabetes into five subgroups: a report from the ORIGIN trial. *Diabetologia* **65**, 206–215 (2022).
96. Herder, C. et al. Differences in biomarkers of inflammation between novel subgroups of recent-onset diabetes. *Diabetes* <https://doi.org/10.2337/db20-1054> (2021).
97. Prasad, R. B. et al. Subgroups of patients with young-onset type 2 diabetes in India reveal insulin deficiency as a major driver. *Diabetologia* <https://doi.org/10.1007/s00125-021-05543-y> (2021).
98. Peng, X. et al. Roles of plasma leptin and resistin in novel subgroups of type 2 diabetes driven by cluster analysis. *Lipids Health Dis.* **21**, 7 (2022).
99. Xie, J. et al. Validation of type 2 diabetes subgroups by simple clinical parameters: a retrospective cohort study of NHANES data from 1999 to 2014. *BMJ Open* **12**, e055647 (2022).
100. Christensen, D. H. et al. Type 2 diabetes classification: a data-driven cluster study of the Danish Centre for Strategic Research in Type 2 Diabetes (DD2) cohort. *BMJ Open Diabetes Res. Care* **10**, e002731 (2022).
101. Gao, H. et al. Cardiorenal risk profiles among data-driven type 2 diabetes subphenotypes: a post-hoc analysis of the China health and nutrition survey. *Front. Endocrinol.* **13**, 828403 (2022).
102. Li, P.-F. & Chen, W.-L. Are the different diabetes subgroups correlated with all-cause, cancer-related, and cardiovascular-related mortality? *J. Clin. Endocrinol. Metab.* **105**, e4240–e4251 (2020).
103. Xing, L. et al. Clinical characteristics and risk of diabetic complications in data-driven clusters among type 2 diabetes. *Front. Endocrinol.* **12**, 617628 (2021).
104. Wang, W. et al. Application of new international classification of adult-onset diabetes in Chinese inpatients with diabetes mellitus. *Diabetes Metab. Res.* **37**, e3427 (2021).
105. Ratter-Rieck, J. et al. Leukocyte counts and T-cell frequencies differ between novel subgroups of diabetes and are associated with metabolic parameters and biomarkers of inflammation. *Diabetes* **70**, 2652–2662 (2021).
106. Aoki, Y. et al. Assessing reproducibility and utility of clustering of patients with type 2 diabetes and established CV disease (SAVOR -TIMI 53 trial). *PLoS One* **16**, e0259372 (2021).
107. Tanabe, H. et al. Factors associated with risk of diabetic complications in novel cluster-based diabetes subgroups: a Japanese retrospective cohort study. *JCM* **9**, 2083 (2020).
108. Fedotkina, O. et al. Novel reclassification of adult diabetes is useful to distinguish stages of β -cell function linked to the risk of vascular complications: the DOLCE study from Northern Ukraine. *Front. Genet.* **12**, 637945 (2021).
109. Wang, F. et al. Novel subgroups and chronic complications of diabetes in middle-aged and elderly Chinese: a prospective cohort study. *Front. Endocrinol.* **12**, 802114 (2022).
110. Antonio-Villa, N. E. et al. Prevalence trends of diabetes subgroups in the United States: a data-driven analysis spanning three decades from NHANES (1988–2018). *J. Clin. Endocrinol. Metab.* dgab762 <https://doi.org/10.1210/clinem/dgab762> (2021).
111. Safai, N., Ali, A., Rossing, P. & Ridderstråle, M. Stratification of type 2 diabetes based on routine clinical markers. *Diabetes Res. Clin. Pract.* **141**, 275–283 (2018).
112. Slieker, R. C. et al. Distinct molecular signatures of clinical clusters in people with type 2 diabetes: an IMI-RHAPSODY study. *Diabetes* **70**, 2683–2693 (2021).
113. Slieker, R. C. et al. Replication and cross-validation of type 2 diabetes subtypes based on clinical variables: an IMI-RHAPSODY study. *Diabetologia* **64**, 1982–1989 (2021).
114. Bancks, M. P. et al. Association of diabetes subgroups with race/ethnicity, risk factor burden and complications: the MASALA and MESA studies. *J. Clin. Endocrinol. Metab.* **106**, e2106–e2115 (2021).
115. Xiong, X. et al. Identification of two novel subgroups in patients with diabetes mellitus and their association with clinical outcomes: A two-step cluster analysis. *J. Diabetes Investig.* **12**, 1346–1358 (2021).
116. Anjana, R. M. et al. 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 Diabet. Res. Care* **8**, e001506 (2020).
117. Kahkoska, A. R. et al. Validation of distinct type 2 diabetes clusters and their association with diabetes complications in the DEVOTE, LEADER and SUSTAIN-6 cardiovascular outcomes trials. *Diabetes Obes. Metab.* **22**, 1537–1547 (2020).
118. Lin, C.-C. et al. Three-year trajectories of metabolic risk factors predict subsequent long-term mortality in patients with type 2 diabetes. *Diabetes Res. Clin. Pract.* **179**, 108995 (2021).
119. Bancks, M. P. et al. Type 2 diabetes subgroups, risk for complications, and differential effects due to an intensive lifestyle intervention. *Diabetes Care* **44**, 1203–1210 (2021).
120. Gunaid, A. A., Al-Kebsi, M. M., Bamashmus, M. A., Al-Akily, S. A. & Al-Radaei, A. N. Clinical phenotyping of newly diagnosed type 2 diabetes in Yemen. *BMJ Open Diabetes Res. Care* **6**, e000587 (2018).
121. Basu, S., Raghavan, S., Wexler, D. J. & Berkowitz, S. A. Characteristics associated with decreased or increased mortality risk from glycemic therapy among patients with type 2 diabetes and high cardiovascular risk: machine learning analysis of the ACCORD trial. *Diabetes Care* **41**, 604–612 (2017).
122. Sharma, A. et al. Cluster analysis of cardiovascular phenotypes in patients with type 2 diabetes and established atherosclerotic cardiovascular disease: a potential approach to precision medicine. *Diabetes Care* **45**, 204–212 (2022).
123. Aravindakshan, M. R. et al. Distinct pathoclinical clusters among patients with uncontrolled type 2 diabetes: results from a prospective study in rural India. *BMJ Open Diabetes Res. Care* **10**, e002654 (2022).
124. Raghavan, S. et al. Generalizability of heterogeneous treatment effects based on causal forests applied to two randomized clinical trials of intensive glycemic control. *Ann. Epidemiol.* **65**, 101–108 (2022).
125. Landi, I. et al. Deep representation learning of electronic health records to unlock patient stratification at scale. *NPJ Digit. Med.* **3**, 96 (2020).
126. Eby, E. L. et al. Evaluating the relationship between clinical and demographic characteristics of insulin-using people with diabetes and their health outcomes: a cluster analysis application. *BMC Health Serv. Res.* **21**, 669 (2021).

127. Segar, M. W. et al. Development and validation of optimal phenomapping methods to estimate long-term atherosclerotic cardiovascular disease risk in patients with type 2 diabetes. *Diabetologia* **64**, 1583–1594 (2021).
128. Seng, J. J. B. et al. Differential health care use, diabetes-related complications, and mortality among five unique classes of patients with type 2 diabetes in Singapore: a latent class analysis of 71,125 patients. *Diabetes Care* **43**, 1048–1056 (2020).
129. Nowakowska, M. et al. The comorbidity burden of type 2 diabetes mellitus: patterns, clusters and predictions from a large English primary care cohort. *BMC Med.* **17**, 145 (2019).
130. Hertrijns, D. F. L. et al. A risk score including body mass index, glycated haemoglobin and triglycerides predicts future glycaemic control in people with type 2 diabetes. *Diabetes Obes. Metab.* **20**, 681–688 (2018).
131. Yoda, N. et al. Classification of adult patients with type 2 diabetes using the Temperament and Character Inventory. *Psychiatry Clin. Neurosci.* **62**, 279–285 (2008).
132. Skinner, T. C. et al. Comparison of illness representations dimensions and illness representation clusters in predicting outcomes in the first year following diagnosis of type 2 diabetes: results from the DESMOND trial. *Psychol. Health* **26**, 321–335 (2011).
133. Obura, M. et al. Post-load glucose subgroups and associated metabolic traits in individuals with type 2 diabetes: an IMI-DIRECT study. *PLoS One* **15**, e0242360 (2020).

Acknowledgements

No specific funding was received to undertake this body of work. The authors acknowledge individual and institutional funding as follows: S.M. has a personal award from Wellcome Trust Career Development scheme (223024/Z/21/Z) and holds Institutional funds from the NIHR Biomedical Research Centre Funding Scheme; B.O. is supported by American Heart Association grant (205FRN35120152); M.S.G. is supported by the American Diabetes Association (9-22-PDFPM-04) and NIH (5UM1DK078616-14); R.J.K. is supported by NIGMS T32GM774844 and Pediatric Endocrine Society Rising Star Award; S.J.C. is supported by a Junior Faculty Development Award from the American Diabetes Association (7-21-JDFM-005); D.D. is supported by NIH grant K23DK133690; A.C.B.T., M.A. and T.H. acknowledge that The Novo Nordisk Foundation Center for Basic Metabolic Research is supported by and unrestricted grant from the Novo Nordisk Foundation (NNF18CC0034900); A.W. is supported by NIH/NHLBI grant T32HL007024; A.L. is supported by grant 2020096 from the Doris Duke Foundation and the American Diabetes Association Grant 7-22-ICTSPM-23; A.J.D. is supported by NIH/NIDDK grant T32DK007028; W.H.H.S. obtained funding from MOST, Taiwan (MOST 107-2314-B-075A-001 -MY3 and by MOST 109-2321-B-075A-001). M.G. is supported by the Eris M. Field Chair in Diabetes Research and NIH grant P30-DK063491; D.R. is supported by NIH/NIDDK grant R21DK125888, and other grants from the NIH; E.S. is supported by NIH/NHLBI grant K24 HL152440 and other grants from the NIH; J.C.F. is supported by NIH K24 HL157960; J.B.M. reports funding from NIH U01 DK078616, R01 HL151855; M.U. is supported by a NIH K23DK114551. The ADA/EASD Precision Diabetes Medicine Initiative, within which this work was conducted, has received the following support: The Covidence licence was funded by Lund University (Sweden) for which technical support was provided by Maria Björklund and Krister Aronsson (Faculty of Medicine Library, Lund University, Sweden). Administrative support was provided by Lund University (Malmö, Sweden), University of Chicago (IL, USA), and the American Diabetes Association (Washington D.C., USA). The Novo Nordisk Foundation (Hellerup, Denmark) provided grant support for in-person writing group meetings (PI: L Phillipson, University of Chicago, IL).

Author contributions

In this manuscript, SM and RW contributed equally as first authors. J.B.M. and MSU jointly supervised the work. B.O., M.S. and M.S.-G. contributed equally as second authors. Review Design: S.M., R.W., B.O., M.S., M.S.G., K.P., C.C.W., R.J.K., S.J.C., M.R.R., D.D., A.C.B.T., A.S.W., A.L., A.J.D., M.K.A., L.K.B., R.H.E., W.H.H.S., T.H., N.S., M.O.G., D.R., E.S., J.C.F., A.D.A./E.A.S.D. P.M.D.I. J.B.M. and M.S.U. Systematic Review Implementation: S.M., R.W., B.O., M.S., M.S.G., K.P., C.C.W., R.J.K., S.J.C., M.R.R., D.D., A.C.B.T., A.S.W., A.L., A.J.D., M.K.A., L.K.B., R.H.E., W.H.H.S., T.H., N.S., M.O.G., D.R., E.S., J.C.F., J.B.M. and M.S.U. Full-text data extraction: S.M., R.W., B.O., M.S., M.S.G., K.P., C.C.W., R.J.K., S.J.C., M.R.R., D.D., A.C.B.T., A.S.W., A.L., A.J.D., M.K.A., L.K.B.,

R.H.E., W.H.H.S., T.H., N.S., M.O.G., D.R., E.S., J.C.F., J.B.M. and M.S.U. Data synthesis: S.M., R.W., B.O., J.B.M., M.U., M.S., K.P., C.W., M.S.G., J.C.F., R.K., M.R.R., A.S.W. Manuscript writing: S.M., R.W., J.B.M. and M.U. Manuscript Review: S.M., R.W., B.O., M.S., M.S.G., K.P., C.C.W., R.J.K., S.J.C., M.R.R., D.D., A.C.B.T., A.S.W., A.L., A.J.D., M.K.A., L.K.B., R.H.E., W.H.H.S., T.H., N.S., M.O.G., D.R., E.S., J.C.F., J.B.M. and M.S.U. Project Management: S.M., R.W., A.D.A./E.A.S.D. P.M.D.I., J.B.M. and M.U.

Competing interests

The authors declare the following conflicts of interest. S.M. has investigator-initiated funding from DexCom, has received speaker fees (donated to institution) from Sanofi for a scientific talk over which she had full control of content and serves on the Board of Trustees for the Diabetes Research & Wellness Foundation (UK); R.W. declares lecture fees from Novo Nordisk, Sanofi and Eli Lilly. He served on an advisory board for Akcea Therapeutics, Daiichi Sankyo, Sanofi, Eli Lilly, and NovoNordisk; S.J.C. reports a close family member employed by a Johnson & Johnson company; R.H.S. reports fees from Novo Nordisk and Amgen; L.K.B. has received consulting honoraria from Bayer, Novo Nordisk, Sanofi, Lilly, and Xeris; W.H.H.S. reported as Advisor and/or Speaker for AstraZeneca, Bayer HealthCare, Boehringer Ingelheim Pharmaceuticals, Daiichi-Sankyo, Eli Lilly and Company, Merck Sharp & Dohme, Mitsubishi Tanabe Pharma Corporation, Novartis Pharmaceuticals, Novo Nordisk, Pfizer, Sanofi-Aventis, Takeda Pharmaceutical Company; N.S. is Senior Associate Editor of *Diabetes* and has received speaking honoraria from AstraZeneca, Boehringer Ingelheim, Eli Lilly, Merck Sharp & Dohme, Novartis, Novo Nordisk, Pfizer and Sanofi for scientific talks over which he had full control of content; M.G. has served on an advisory board for Nestle Health Science; E.S. is a Deputy Editor of *Diabetes Care* and a member of the editorial board of *Diabetologia* and receives payments from Wolters Kluwer for chapters and laboratory monographs in *UpToDate* on measurements of glycemic control and screening tests for type 2 diabetes; J.C.F. has received speaking honoraria from AstraZeneca and Novo Nordisk for scientific talks over which he had full control of content; J.B.M. is an Academic Associate for Quest Inc. Diagnostics R&D; M.U. reports an unpaid collaborator with AstraZeneca. All other authors have no disclosures. A.L. reports a close family member employed by Merck & Co., Inc.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s43856-023-00360-3>.

Correspondence and requests for materials should be addressed to Shivani Misra.

Peer review information *Communications Medicine* thanks Albert Salas-Huetos, Peter Rossing and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permission information is available at <http://www.nature.com/reprints>

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



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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023

¹Department of Metabolism, Digestion and Reproduction, Imperial College London, London, UK. ²Department of Diabetes and Endocrinology, Imperial College Healthcare NHS Trust, London, UK. ³Department of Endocrinology and Diabetology, University Hospital Düsseldorf, Heinrich Heine University Düsseldorf, Moorenstr. 5, 40225 Düsseldorf, Germany. ⁴Institute for Clinical Diabetology, German Diabetes Center, Leibniz Center for Diabetes Research at Heinrich Heine University Düsseldorf, Auf'm Hennekamp 65, 40225 Düsseldorf, Germany. ⁵German Center for Diabetes Research (DZD), Ingolstädter Landstraße 1, 85764 Neuherberg, Germany. ⁶Welch Center for Prevention, Epidemiology, and Clinical Research, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA. ⁷Ciccarone Center for the Prevention of Cardiovascular Disease, Johns Hopkins School of Medicine, Baltimore, MD, USA. ⁸Institute of Experimental Endocrinology, Biomedical Research Center, Slovak

Academy of Sciences, Bratislava, Slovakia. ⁹Clinical and Translational Epidemiology Unit, Massachusetts General Hospital, Boston, MA, USA. ¹⁰Programs in Metabolism and Medical & Population Genetics, The Broad Institute of MIT and Harvard, Cambridge, MA, USA. ¹¹Department of Medicine, Harvard Medical School, Boston, MA, USA. ¹²Diabetes Unit, Division of Endocrinology, Massachusetts General Hospital, Boston, MA, USA. ¹³Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA. ¹⁴Department of Pediatrics, Division of Endocrinology, Boston Children's Hospital, Boston, MA, USA. ¹⁵Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA. ¹⁶Division of Endocrinology, Diabetes and Metabolism, Johns Hopkins University School of Medicine, Baltimore, MD, USA. ¹⁷Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. ¹⁸Division of General Internal Medicine, Massachusetts General Hospital, 100 Cambridge St 16th Floor, Boston, MA, USA. ¹⁹Division of Endocrinology, Diabetes and Metabolism, NorthShore University Health System, Skokie, IL, USA. ²⁰Department of Medicine, Pritzker School of Medicine, University of Chicago, Chicago, IL, USA. ²¹Division of Endocrinology, Metabolism and Diabetes, University of Colorado School of Medicine, Aurora, CO, USA. ²²Institute of Molecular and Genomic Medicine, National Health Research Institute, Miaoli County, Taiwan, ROC. ²³Division of Endocrinology and Metabolism, Taichung Veterans General Hospital, Taichung, Taiwan, ROC. ²⁴Division of Endocrinology and Metabolism, Taipei Veterans General Hospital, Taipei, Taiwan, ROC. ²⁵University Hospital of Tübingen, Tübingen, Germany. ²⁶Institute of Diabetes Research and Metabolic Diseases (IDM), Helmholtz Center Munich, Neuherberg, Germany. ²⁷Division of Endocrinology, Diabetes and Metabolism, Department of Medicine, Cedars-Sinai Medical Center, Los Angeles, CA, USA. ²⁸Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA. ²⁰⁹These authors contributed equally: Shivani Misra, Robert Wagner. ²¹⁰These authors jointly supervised this work: James B. Meigs, Miriam S. Udler. *A list of authors and their affiliations appears at the end of the paper. ✉email: s.misra@imperial.ac.uk

ADA/EASD PMDI

Deirdre K. Tobias^{29,30}, Jordi Merino^{13,17,31}, Abrar Ahmad³², Catherine Aiken^{33,34}, Jamie L. Benham³⁵, Dhanasekaran Bodhini³⁶, Amy L. Clark³⁷, Kevin Colclough³⁸, Rosa Corcoy^{39,40,41}, Sara J. Cromer^{11,31,42}, Daisy Duan¹⁶, Jamie L. Felton^{43,44,45}, Ellen C. Francis⁴⁶, Pieter Gillard⁴⁷, Véronique Gingras^{48,49}, Romy Gaillard⁵⁰, Eram Haider⁵¹, Alice Hughes³⁸, Jennifer M. Ikle^{52,53}, Laura M. Jacobsen⁵⁴, Anna R. Kahkoska⁵⁵, Jarno L. T. Kettunen^{56,57,58}, Raymond J. Kreienkamp^{13,14,31,42}, Lee-Ling Lim^{59,60,61}, Jonna M. E. Männistö^{62,63}, Robert Massey⁵¹, Niamh-Maire McLennan⁶⁴, Rachel G. Miller⁶⁵, Mario Luca Morieri^{66,67}, Jasper Most⁶⁸, Rochelle N. Naylor⁶⁹, Bige Ozkan^{6,70}, Kashyap Amratlal Patel³⁸, Scott J. Pilla^{71,72}, Katsiaryna Prystupa^{4,5}, Sridaran Raghaven^{73,74}, Mary R. Rooney^{6,15}, Martin Schön^{4,5,75}, Zhila Semnani-Azad³⁰, Magdalena Sevilla-Gonzalez^{9,11,42}, Pernille Svalastoga^{76,77}, Wubet Worku Takele⁷⁸, Claudia Ha-ting Tam^{61,79,80}, Anne Cathrine B. Thuesen¹⁷, Mustafa Tosur^{81,82,83}, Amelia S. Wallace^{6,15}, Caroline C. Wang¹⁵, Jessie J. Wong⁸⁴, Jennifer M. Yamamoto⁸⁵, Katherine Young³⁸, Chloé Amouyal^{86,87}, Mette K. Andersen¹⁷, Maxine P. Bonham⁸⁸, Mingling Chen⁸⁹, Feifei Cheng⁹⁰, Tinashe Chikowore^{11,91,92,93}, Sian C. Chivers⁹⁴, Christoffer Clemmensen¹⁷, Dana Dabelea⁹⁵, Adem Y. Dawed⁵¹, Aaron J. Deutsch^{11,13,42}, Laura T. Dickens⁹⁶, Linda A. DiMeglio^{43,44,45,97}, Monika Dudenhöffer-Pfeifer³², Carmella Evans-Molina^{43,44,45,98}, María Mercè Fernández-Balsells^{99,100}, Hugo Fitipaldi³², Stephanie L. Fitzpatrick¹⁰¹, Stephen E. Gitelman¹⁰², Mark O. Goodarzi^{103,104}, Jessica A. Grieger^{105,106}, Marta Guasch-Ferré^{30,107}, Nahal Habibi^{105,106}, Torben Hansen¹⁷, Chuiguo Huang^{61,79}, Arianna Harris-Kawano^{43,44,45}, Heba M. Ismail^{43,44,45}, Benjamin Hoag^{108,109}, Randi K. Johnson^{110,111}, Angus G. Jones^{38,112}, Robert W. Koivula¹¹³, Aaron Leong^{11,31,114}, Gloria K. W. Leung⁸⁸, Ingrid M. Libman¹¹⁵, Kai Liu¹⁰⁵, S. Alice Long¹¹⁶, William L. Lowe Jr.¹¹⁷, Robert W. Morton^{118,119,120}, Ayesha A. Motala¹²¹, Suna Onengut-Gumuscu¹²², James S. Pankow¹²³, Maleesa Pathirana^{105,106}, Sofia Pazmino¹²⁴, Dianna Perez^{43,44,45}, John R. Petrie¹²⁵, Camille E. Powe^{11,31,42,126}, Alejandra Quinteros¹⁰⁵, Rashmi Jain^{127,128}, Debashree Ray^{15,28}, Mathias Ried-Larsen^{129,130}, Zeb Saeed¹³¹, Vanessa Santhakumar²⁹, Sarah Kanbour^{71,132}, Sudipa Sarkar⁷¹, Gabriela S. F. Monaco^{43,44,45}, Denise M. Scholtens¹³³, Elizabeth Selvin^{6,15}, Wayne Huey-Herng Sheu^{24,134,135}, Cate Speake¹³⁶, Maggie A. Stanislawski¹¹⁰, Nele Steenackers¹²⁴, Andrea K. Steck¹³⁷, Norbert Stefan^{5,25,138}, Julie Støy¹³⁹, Rachael Taylor¹⁴⁰, Sok Cin Tye^{141,142}, Gebresilasea Gendisha Ukke⁷⁸, Marzhan Urazbayeva^{82,143}, Bart Van der Schueren^{124,144}, Camille Vatier^{145,146}, John M. Wentworth^{147,148,149}, Wesley Hannah^{150,151}, Sara L. White^{94,152}, Gechang Yu^{61,79}, Yingchai Zhang^{61,79}, Shao J. Zhou^{106,153}, Jacques Beltrand^{154,155}, Michel Polak^{154,155}, Ingvild Aukrust^{76,156}, Elisa de Franco³⁸, Sarah E. Flanagan³⁸, Kristin A. Maloney¹⁵⁷, Andrew McGovern³⁸,

Janne Molnes^{76,156}, Mariam Nakabuye¹⁷, Pål Rasmus Njølstad^{76,77}, Hugo Pomares-Millan^{32,158}, Michele Provenzano¹⁵⁹, Cécile Saint-Martin¹⁶⁰, Cuilin Zhang^{161,162}, Yeyi Zhu^{163,164}, Sungyoung Auh¹⁶⁵, Russell de Souza^{119,166}, Andrea J. Fawcett^{167,168}, Chandra Gruber¹⁶⁹, Eskedar Getie Mekonnen^{170,171}, Emily Mixter¹⁷², Diana Sherifali^{119,173}, Robert H. Eckel¹⁷⁴, John J. Nolan^{175,176}, Louis H. Philipson¹⁷², Rebecca J. Brown¹⁶⁵, Liana K. Billings^{177,178}, Kristen Boyle⁹⁵, Tina Costacou⁶⁵, John M. Dennis³⁸, Jose C. Florez^{11,13,31,42}, Anna L. Gloyn^{52,53,179}, Maria F. Gomez^{32,180}, Peter A. Gottlieb¹³⁷, Siri Atma W. Greeley¹⁸¹, Kurt Griffin^{128,182}, Andrew T. Hattersley^{38,112}, Irl B. Hirsch¹⁸³, Marie-France Hivert^{31,184,185}, Korey K. Hood⁸⁴, Jami L. Josefson¹⁶⁷, Soo Heon Kwak¹⁸⁶, Lori M. Laffel¹⁸⁷, Siew S. Lim⁷⁸, Ruth J. F. Loos^{17,188}, Ronald C. W. Ma^{61,79,80}, Chantal Mathieu⁴⁷, Nestoras Mathioudakis⁷¹, James B. Meigs^{11,114,189}, Shivani Misra^{190,191}, Viswanathan Mohan¹⁹², Rinki Murphy^{193,194,195}, Richard Oram^{38,112}, Katharine R. Owen^{113,196}, Susan E. Ozanne¹⁹⁷, Ewan R. Pearson⁵¹, Wei Perng⁹⁵, Toni I. Pollin^{157,198}, Rodica Pop-Busui¹⁹⁹, Richard E. Pratley²⁰⁰, Leanne M. Redman²⁰¹, Maria J. Redondo^{81,82}, Rebecca M. Reynolds⁶⁴, Robert K. Semple^{64,202}, Jennifer L. Sherr²⁰³, Emily K. Sims^{43,44,45}, Arianne Sweeting^{204,205}, Tiinamaija Tuomi^{56,58,148}, Miriam S. Udler^{11,13,31,42}, Kimberly K. Vesco²⁰⁶, Tina Vilsbøll^{207,208}, Robert Wagner^{3,4,5,209}, Stephen S. Rich¹²² & Paul W. Franks^{30,32,113,120}

²⁹Division of Preventative Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA. ³⁰Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA, USA. ³¹Diabetes Unit, Endocrine Division, Massachusetts General Hospital, Boston, MA, USA. ³²Department of Clinical Sciences, Lund University Diabetes Centre, Lund University, Malmö, Sweden. ³³Department of Obstetrics and Gynaecology, the Rosie Hospital, Cambridge, UK. ³⁴NIHR Cambridge Biomedical Research Centre, University of Cambridge, Cambridge, UK. ³⁵Departments of Medicine and Community Health Sciences, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada. ³⁶Department of Molecular Genetics, Madras Diabetes Research Foundation, Chennai, India. ³⁷Division of Pediatric Endocrinology, Department of Pediatrics, Saint Louis University School of Medicine, SSM Health Cardinal Glennon Children's Hospital, St. Louis, MO, USA. ³⁸Department of Clinical and Biomedical Sciences, University of Exeter Medical School, Exeter, Devon, UK. ³⁹CIBER-BBN, ISCIII, Madrid, Spain. ⁴⁰Institut d'Investigació Biomèdica Sant Pau (IIB SANT PAU), Barcelona, Spain. ⁴¹Departament de Medicina, Universitat Autònoma de Barcelona, Bellaterra, Spain. ⁴²Programs in Metabolism and Medical & Population Genetics, Broad Institute, Cambridge, MA, USA. ⁴³Department of Pediatrics, Indiana University School of Medicine, Indianapolis, IN, USA. ⁴⁴Herman B Wells Center for Pediatric Research, Indiana University School of Medicine, Indianapolis, IN, USA. ⁴⁵Center for Diabetes and Metabolic Diseases, Indiana University School of Medicine, Indianapolis, IN, USA. ⁴⁶Department of Biostatistics and Epidemiology, Rutgers School of Public Health, Piscataway, NJ, USA. ⁴⁷University Hospital Leuven, Leuven, Belgium. ⁴⁸Department of Nutrition, Université de Montréal, Montreal, QC, Canada. ⁴⁹Research Center, Sainte-Justine University Hospital Center, Montreal, QC, Canada. ⁵⁰Department of Pediatrics, Erasmus Medical Center, Rotterdam, The Netherlands. ⁵¹Division of Population Health & Genomics, School of Medicine, University of Dundee, Dundee, UK. ⁵²Department of Pediatrics, Stanford School of Medicine, Stanford University, Stanford, CA, USA. ⁵³Stanford Diabetes Research Center, Stanford School of Medicine, Stanford University, Stanford, CA, USA. ⁵⁴University of Florida, Gainesville, FL, USA. ⁵⁵Department of Nutrition, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. ⁵⁶Helsinki University Hospital, Abdominal Centre/Endocrinology, Helsinki, Finland. ⁵⁷Folkhalsan Research Center, Helsinki, Finland. ⁵⁸Institute for Molecular Medicine Finland FIMM, University of Helsinki, Helsinki, Finland. ⁵⁹Department of Medicine, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia. ⁶⁰Asia Diabetes Foundation, Hong Kong SAR, China. ⁶¹Department of Medicine & Therapeutics, Chinese University of Hong Kong, Hong Kong SAR, China. ⁶²Departments of Pediatrics and Clinical Genetics, Kuopio University Hospital, Kuopio, Finland. ⁶³Department of Medicine, University of Eastern Finland, Kuopio, Finland. ⁶⁴Centre for Cardiovascular Science, Queen's Medical Research Institute, University of Edinburgh, Edinburgh, UK. ⁶⁵Department of Epidemiology, University of Pittsburgh, Pittsburgh, PA, USA. ⁶⁶Metabolic Disease Unit, University Hospital of Padova, Padova, Italy. ⁶⁷Department of Medicine, University of Padova, Padova, Italy. ⁶⁸Department of Orthopedics, Zuyderland Medical Center, Sittard-Geleen, The Netherlands. ⁶⁹Departments of Pediatrics and Medicine, University of Chicago, Chicago, IL, USA. ⁷⁰Ciccarone Center for the Prevention of Cardiovascular Disease, Johns Hopkins School of Medicine, Baltimore, MD, USA. ⁷¹Department of Medicine, Johns Hopkins University, Baltimore, MD, USA. ⁷²Department of Health Policy and Management, Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD, USA. ⁷³Section of Academic Primary Care, US Department of Veterans Affairs Eastern Colorado Health Care System, Aurora, CO, USA. ⁷⁴Department of Medicine, University of Colorado School of Medicine, Aurora, CO, USA. ⁷⁵Institute of Experimental Endocrinology, Biomedical Research Center, Slovak Academy of Sciences, Bratislava, Slovakia. ⁷⁶Mohn Center for Diabetes Precision Medicine, Department of Clinical Science, University of Bergen, Bergen, Norway. ⁷⁷Children and Youth Clinic, Haukeland University Hospital, Bergen, Norway. ⁷⁸Eastern Health Clinical School, Monash University, Melbourne, VIC, Australia. ⁷⁹Laboratory for Molecular Epidemiology in Diabetes, Li Ka Shing Institute of Health Sciences, The Chinese University of Hong Kong, Hong Kong, China. ⁸⁰Hong Kong Institute of Diabetes and Obesity, The Chinese University of Hong Kong, Hong Kong, China. ⁸¹Department of Pediatrics, Baylor College of Medicine, Houston, TX, USA. ⁸²Division of Pediatric Diabetes and Endocrinology, Texas Children's Hospital, Houston, TX, USA. ⁸³Children's Nutrition Research Center, USDA/ARS, Houston, TX, USA. ⁸⁴Stanford University School of Medicine, Stanford, CA, USA. ⁸⁵Internal Medicine, University of Manitoba, Winnipeg, MB, Canada. ⁸⁶Department of Diabetology, APHP, Paris, France. ⁸⁷Sorbonne Université, INSERM, NutriOmic team, Paris, France. ⁸⁸Department of Nutrition, Dietetics and Food, Monash University, Melbourne, VIC, Australia. ⁸⁹Monash Centre for Health Research and Implementation, Monash University, Clayton, VIC, Australia. ⁹⁰Health Management Center, The Second Affiliated Hospital of Chongqing Medical University, Chongqing Medical University, Chongqing, China. ⁹¹MRC/Wits Developmental Pathways for Health Research Unit, Department of Paediatrics, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa. ⁹²Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, MA, USA. ⁹³Sydney Brenner Institute for Molecular Bioscience, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa. ⁹⁴Department of Women and Children's health, King's College London, London, UK. ⁹⁵Lifecourse Epidemiology of Adiposity and Diabetes (LEAD) Center, University of

Colorado Anschutz Medical Campus, Aurora, CO, USA. ⁹⁶Section of Adult and Pediatric Endocrinology, Diabetes and Metabolism, Kovler Diabetes Center, University of Chicago, Chicago, IL, USA. ⁹⁷Department of Pediatrics, Riley Hospital for Children, Indiana University School of Medicine, Indianapolis, IN, USA. ⁹⁸Richard L. Roudebush VAMC, Indianapolis, IN, USA. ⁹⁹Biomedical Research Institute Girona, IDIBGi, Girona, Spain. ¹⁰⁰Diabetes, Endocrinology and Nutrition Unit Girona, University Hospital Dr Josep Trueta, Girona, Spain. ¹⁰¹Institute of Health System Science, Feinstein Institutes for Medical Research, Northwell Health, Manhasset, NY, USA. ¹⁰²University of California at San Francisco, Department of Pediatrics, Diabetes Center, San Francisco, CA, USA. ¹⁰³Division of Endocrinology, Diabetes and Metabolism, Cedars-Sinai Medical Center, Los Angeles, CA, USA. ¹⁰⁴Department of Medicine, Cedars-Sinai Medical Center, Los Angeles, CA, USA. ¹⁰⁵Adelaide Medical School, Faculty of Health and Medical Sciences, The University of Adelaide, Adelaide, SA, Australia. ¹⁰⁶Robinson Research Institute, The University of Adelaide, Adelaide, SA, Australia. ¹⁰⁷Department of Public Health and Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, 1014 Copenhagen, Denmark. ¹⁰⁸Division of Endocrinology and Diabetes, Department of Pediatrics, Sanford Children's Hospital, Sioux Falls, SD, USA. ¹⁰⁹University of South Dakota, School of Medicine, E Clark St, Vermillion, SD, USA. ¹¹⁰Department of Biomedical Informatics, University of Colorado Anschutz Medical Campus, Aurora, CO, USA. ¹¹¹Department of Epidemiology, Colorado School of Public Health, Aurora, CO, USA. ¹¹²Royal Devon University Healthcare NHS Foundation Trust, Exeter, UK. ¹¹³Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford, UK. ¹¹⁴Division of General Internal Medicine, Massachusetts General Hospital, Boston, MA, USA. ¹¹⁵UPMC Children's Hospital of Pittsburgh, Pittsburgh, PA, USA. ¹¹⁶Center for Translational Immunology, Benaroya Research Institute, Seattle, WA, USA. ¹¹⁷Department of Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA. ¹¹⁸Department of Pathology & Molecular Medicine, McMaster University, Hamilton, ON, Canada. ¹¹⁹Population Health Research Institute, Hamilton, ON, Canada. ¹²⁰Department of Translational Medicine, Medical Science, Novo Nordisk Foundation, Tuborg Havnevej 19, 2900 Hellerup, Denmark. ¹²¹Department of Diabetes and Endocrinology, Nelson R Mandela School of Medicine, University of KwaZulu-Natal, Durban, South Africa. ¹²²Center for Public Health Genomics, Department of Public Health Sciences, University of Virginia, Charlottesville, VA, USA. ¹²³Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, MN, USA. ¹²⁴Department of Chronic Diseases and Metabolism, Clinical and Experimental Endocrinology, KU Leuven, Leuven, Belgium. ¹²⁵School of Health and Wellbeing, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, UK. ¹²⁶Department of Obstetrics, Gynecology, and Reproductive Biology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA. ¹²⁷Sanford Children's Specialty Clinic, Sioux Falls, SD, USA. ¹²⁸Department of Pediatrics, Sanford School of Medicine, University of South Dakota, Sioux Falls, SD, USA. ¹²⁹Centre for Physical Activity Research, Rigshospitalet, Copenhagen, Denmark. ¹³⁰Institute for Sports and Clinical Biomechanics, University of Southern Denmark, Odense, Denmark. ¹³¹Department of Medicine, Division of Endocrinology, Diabetes and Metabolism, Indiana University School of Medicine, Indianapolis, IN, USA. ¹³²AMAN Hospital, Doha, Qatar. ¹³³Department of Preventive Medicine, Division of Biostatistics, Northwestern University Feinberg School of Medicine, Chicago, IL, USA. ¹³⁴Institute of Molecular and Genomic Medicine, National Health Research Institutes, Taipei City, Taiwan, ROC. ¹³⁵Division of Endocrinology and Metabolism, Taichung Veterans General Hospital, Taichung, Taiwan, ROC. ¹³⁶Center for Interventional Immunology, Benaroya Research Institute, Seattle, WA, USA. ¹³⁷Barbara Davis Center for Diabetes, University of Colorado Anschutz Medical Campus, Aurora, CO, USA. ¹³⁸Institute of Diabetes Research and Metabolic Diseases (IDM), Helmholtz Center Munich, Neuherberg, Germany. ¹³⁹Steno Diabetes Center Aarhus, Aarhus University Hospital, Aarhus, Denmark. ¹⁴⁰University of Newcastle, Newcastle upon Tyne, UK. ¹⁴¹Section of Genetics and Epidemiology, Joslin Diabetes Center, Harvard Medical School, Boston, MA, USA. ¹⁴²Department of Clinical Pharmacy and Pharmacology, University Medical Center Groningen, Groningen, The Netherlands. ¹⁴³Department of Gastroenterology, Baylor College of Medicine, Houston, TX, USA. ¹⁴⁴Department of Endocrinology, University Hospitals Leuven, Leuven, Belgium. ¹⁴⁵Sorbonne University, Inserm U938, Saint-Antoine Research Centre, Institute of Cardiometabolism and Nutrition, Paris 75012, France. ¹⁴⁶Department of Endocrinology, Diabetology and Reproductive Endocrinology, Assistance Publique-Hôpitaux de Paris, Saint-Antoine University Hospital, National Reference Center for Rare Diseases of Insulin Secretion and Insulin Sensitivity (PRISIS), Paris, France. ¹⁴⁷Department of Diabetes and Endocrinology, Royal Melbourne Hospital, Parkville, VIC, Australia. ¹⁴⁸Walter and Eliza Hall Institute, Parkville, VIC, Australia. ¹⁴⁹Department of Medicine, University of Melbourne, Parkville, VIC, Australia. ¹⁵⁰Deakin University, Melbourne, VIC, Australia. ¹⁵¹Department of Epidemiology, Madras Diabetes Research Foundation, Chennai, India. ¹⁵²Department of Diabetes and Endocrinology, Guy's and St Thomas' Hospitals NHS Foundation Trust, London, UK. ¹⁵³School of Agriculture, Food and Wine, University of Adelaide, Adelaide, Australia. ¹⁵⁴Institut Cochin, Inserm U, 10116 Paris, France. ¹⁵⁵Pediatric Endocrinology and Diabetes, Hôpital Necker Enfants Malades, APHP Centre, Université de Paris, Paris, France. ¹⁵⁶Department of Medical Genetics, Haukeland University Hospital, Bergen, Norway. ¹⁵⁷Department of Medicine, University of Maryland School of Medicine, Baltimore, MD, USA. ¹⁵⁸Department of Epidemiology, Geisel School of Medicine at Dartmouth, Hanover, NH, USA. ¹⁵⁹Nephrology, Dialysis and Renal Transplant Unit, IRCCS—Azienda Ospedaliero-Universitaria di Bologna, Alma Mater Studiorum University of Bologna, Bologna, Italy. ¹⁶⁰Department of Medical Genetics, AP-HP Pitié-Salpêtrière Hospital, Sorbonne University, Paris, France. ¹⁶¹Global Center for Asian Women's Health, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore. ¹⁶²Department of Obstetrics and Gynecology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore. ¹⁶³Division of Research, Kaiser Permanente Northern California, Oakland, CA, USA. ¹⁶⁴Department of Epidemiology and Biostatistics, University of California San Francisco, California, USA. ¹⁶⁵National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD, USA. ¹⁶⁶Department of Health Research Methods, Evidence, and Impact, Faculty of Health Sciences, McMaster University, Hamilton, ON, Canada. ¹⁶⁷Ann & Robert H. Lurie Children's Hospital of Chicago, Department of Pediatrics, Northwestern University Feinberg School of Medicine, Chicago, IL, USA. ¹⁶⁸Department of Clinical and Organizational Development, Children's Memorial Hospital, Chicago, IL, USA. ¹⁶⁹American Diabetes Association, Arlington, VA, USA. ¹⁷⁰College of Medicine and Health Sciences, University of Gondar, Gondar, Ethiopia. ¹⁷¹Faculty of Medicine and Health Sciences, Global Health Institute, University of Antwerp, 2160 Antwerp, Belgium. ¹⁷²Department of Medicine and Kovler Diabetes Center, University of Chicago, Chicago, IL, USA. ¹⁷³School of Nursing, Faculty of Health Sciences, McMaster University, Hamilton, ON, Canada. ¹⁷⁴Division of Endocrinology, Metabolism, Diabetes, University of Colorado, Boulder, CO, USA. ¹⁷⁵Department of Clinical Medicine, School of Medicine, Trinity College Dublin, Dublin, Ireland. ¹⁷⁶Department of Endocrinology, Wexford General Hospital, Wexford, Ireland. ¹⁷⁷Division of Endocrinology, NorthShore University HealthSystem, Skokie, IL, USA. ¹⁷⁸Department of Medicine, Pritzker School of Medicine, University of Chicago, Chicago, IL, USA. ¹⁷⁹Department of Genetics, Stanford School of Medicine, Stanford University, Stanford, CA, USA. ¹⁸⁰Faculty of Health, Aarhus University, Aarhus, Denmark. ¹⁸¹Departments of Pediatrics and Medicine and Kovler Diabetes Center, University of Chicago, Chicago, USA. ¹⁸²Sanford Research, Sioux Falls, SD, USA. ¹⁸³University of Washington School of Medicine, Seattle, WA, USA. ¹⁸⁴Department of Population Medicine, Harvard Medical School, Harvard Pilgrim Health Care Institute, Boston, MA, USA. ¹⁸⁵Department of Medicine, Université de Sherbrooke, Sherbrooke, QC, Canada. ¹⁸⁶Department of Internal Medicine, Seoul National University College of Medicine, Seoul National University Hospital, Seoul, Republic of Korea. ¹⁸⁷Joslin Diabetes Center, Harvard Medical School, Boston, MA, USA. ¹⁸⁸Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA. ¹⁸⁹Broad Institute, Cambridge, MA, USA. ¹⁹⁰Division of Metabolism, Digestion and Reproduction, Imperial College London, London, UK. ¹⁹¹Department of Diabetes & Endocrinology, Imperial College Healthcare NHS Trust, London, UK. ¹⁹²Department of Diabetology, Madras Diabetes Research Foundation & Dr. Mohan's Diabetes Specialities Centre, Chennai, India. ¹⁹³Department of Medicine, Faculty of Medicine and Health Sciences, University of Auckland,

Auckland, New Zealand. ¹⁹⁴Auckland Diabetes Centre, Te Whatu Ora Health New Zealand, Auckland, New Zealand. ¹⁹⁵Medical Bariatric Service, Te Whatu Ora Counties, Health New Zealand, Auckland, New Zealand. ¹⁹⁶Oxford NIHR Biomedical Research Centre, University of Oxford, Oxford, UK. ¹⁹⁷Metabolic Research Laboratories and MRC Metabolic Diseases Unit, University of Cambridge, Wellcome-MRC Institute of Metabolic Science, Cambridge, UK. ¹⁹⁸Department of Epidemiology & Public Health, University of Maryland School of Medicine, Baltimore, MD, USA. ¹⁹⁹Department of Internal Medicine, Division of Metabolism, Endocrinology and Diabetes, University of Michigan, Ann Arbor, MI, USA. ²⁰⁰AdventHealth Translational Research Institute, Orlando, FL, USA. ²⁰¹Pennington Biomedical Research Center, Baton Rouge, LA, USA. ²⁰²MRC Human Genetics Unit, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, UK. ²⁰³Yale School of Medicine, New Haven, CT, USA. ²⁰⁴Faculty of Medicine and Health, University of Sydney, Sydney, NSW, Australia. ²⁰⁵Department of Endocrinology, Royal Prince Alfred Hospital, Sydney, NSW, Australia. ²⁰⁶Kaiser Permanente Northwest, Kaiser Permanente Center for Health Research, Portland, OR, USA. ²⁰⁷Clinical Research, Steno Diabetes Center Copenhagen, Herlev, Denmark. ²⁰⁸Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark.