



# Subtypes of Type 2 Diabetes Determined From Clinical Parameters

Emma Ahlqvist,<sup>1</sup> Rashmi B. Prasad,<sup>1</sup> and Leif Groop<sup>1,2</sup>

<https://doi.org/10.2337/dbi20-0001>

**Type 2 diabetes (T2D) is defined by a single metabolite, glucose, but is increasingly recognized as a highly heterogeneous disease, including individuals with varying clinical characteristics, disease progression, drug response, and risk of complications. Identification of subtypes with differing risk profiles and disease etiologies at diagnosis could open up avenues for personalized medicine and allow clinical resources to be focused to the patients who would be most likely to develop diabetic complications, thereby both improving patient health and reducing costs for the health sector. More homogeneous populations also offer increased power in experimental, genetic, and clinical studies. Clinical parameters are easily available and reflect relevant disease pathways, including the effects of both genetic and environmental exposures. We used six clinical parameters (GAD autoantibodies, age at diabetes onset, HbA<sub>1c</sub>, BMI, and measures of insulin resistance and insulin secretion) to cluster adult-onset diabetes patients into five subtypes. These subtypes have been robustly reproduced in several populations and associated with different risks of complications, comorbidities, genetics, and response to treatment. Importantly, the group with severe insulin-deficient diabetes (SIDD) had increased risk of retinopathy and neuropathy, whereas the severe insulin-resistant diabetes (SIRD) group had the highest risk for diabetic kidney disease (DKD) and fatty liver, emphasizing the importance of insulin resistance for DKD and hepatosteatosis in T2D. In conclusion, we believe that subclassification using these highly relevant parameters could provide a framework for personalized medicine in diabetes.**

## Heterogeneity of Diabetes

### **Current Classifications for Diabetes**

Diabetes is a heterogeneous disease with varying manifestation and risk of complications (1). While diabetes is

diagnosed on the basis of a single metabolite, glucose, hyperglycemia can arise due to multiple complex etiological processes that can vary between individuals (2). These processes influence the clinical characteristics, progression, drug response, and development of complications. Diabetes is therefore traditionally divided into different types (Fig. 1A). Type 1 diabetes (T1D) and latent autoimmune diabetes of the adult (LADA) both result from autoimmune destruction of  $\beta$ -cells, often, but not always, reflected by presence of pancreatic autoantibodies in the blood that can be used as a diagnostic marker (3). Identification of such antibodies is a strong indicator that the patient will eventually need insulin treatment to maintain glucose homeostasis (4). Rare monogenic diabetes types, such as maturity-onset diabetes of the young (MODY) and neonatal diabetes, account for about 3% of diabetes diagnosed in individuals <30 years of age. Diagnosis requires sequencing of known monogenic diabetes genes, and the consequences for those affected are life changing, since a correct diagnosis has major implications on choice of treatment (5).

After exclusion of these and a few other subtypes, such as diabetes secondary to steroid use, cystic fibrosis, and hemochromatosis, the remaining patients, 75–85%, are considered to have type 2 diabetes (T2D). Because autoantibodies are not always measured and genetic diagnostics are often not available, the T2D group may include patients with undiagnosed autoimmune or monogenic diabetes. While the two main types of diabetes have been recognized for thousands of years, the names and definitions have changed and there is still no clear-cut definition that will allow all patients to be classified as either T1D or T2D (2). Some patients show signs of both autoimmune destruction of  $\beta$ -cells and profound insulin resistance with features of the “insulin resistance syndrome.” The large remaining group of true T2D is still highly heterogeneous

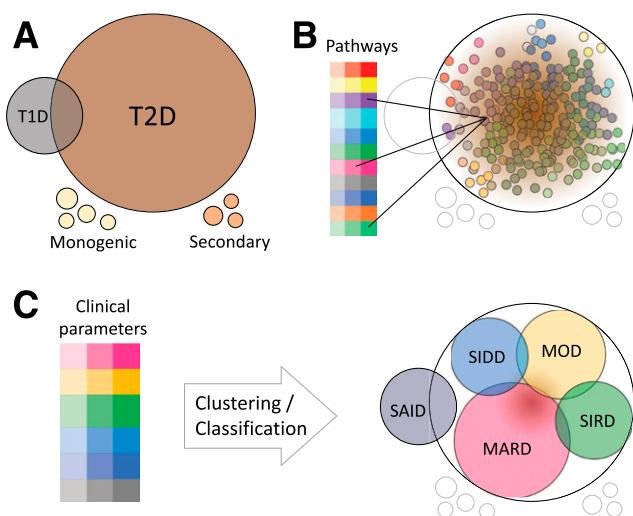
<sup>1</sup>Department of Clinical Sciences, Genomics, Diabetes and Endocrinology, Lund University Diabetes Centre, Lund University, Malmö, Sweden

<sup>2</sup>Finnish Institute of Molecular Medicine Finland (FIMM), Helsinki University, Helsinki, Finland

Corresponding author: Emma Ahlqvist, [emma.ahlqvist@med.lu.se](mailto:emma.ahlqvist@med.lu.se)

Received 6 May 2020 and accepted 14 July 2020

© 2020 by the American Diabetes Association. Readers may use this article as long as the work is properly cited, the use is educational and not for profit, and the work is not altered. More information is available at <https://www.diabetesjournals.org/content/license>.



**Figure 1**—Schematic representations of diabetes classification and models of heterogeneity. *A*: The traditional classification of diabetes where T1D and T2D are viewed as distinct subgroups but with an overlap including intermediate diabetes types such as LADA. Monogenic and secondary diabetes are rare, etiologically distinct subtypes that are not included in the novel subclassification. *B*: The palette model where the many molecular pathways involved in diabetes pathogenesis are represented as colors and each individual with T2D is represented by the mixed colors of all the pathways affected in that individual. *C*: The ANDIS model, where colors instead represent clinical parameters, reflecting the underlying mechanisms, that are easy to measure in the clinic. With this approach, individuals are grouped based on their main color into subtypes with different clinical characteristics.

with respect to clinical characteristics, progression, and risk of complications. Even before diabetes onset, the two pre-diabetes states, impaired glucose tolerance and impaired fasting glucose, only show partial overlap, suggesting they may result from different pathophysiological mechanisms (6). T2D is thus clearly a multifactorial disease with multiple underlying etiologies that results from the combined effects of numerous genetic and environmental risk factors (7,8).

#### Proposed Models of Heterogeneity Within T2D

The nature and origin of the heterogeneity in T2D has been discussed, contrasting a model where T2D is seen as a mixture of patients with homogeneous phenotypes and distinct mechanisms with a model that assumes that each patient develops diabetes due to a combination of many small defects in different pathways placing patients on a quantitative spectrum of metabolic disturbance (9,10). It is clear that the first alternative is an overly simplistic model, even for the division of patients into T1D and T2D, which has served the clinic well for a long time (11).

A version of the second alternative was described as the palette model (Fig. 1B), in which each pathway is imagined as a color and each patient given the hue of the combined pathways that are defect in that individual, which for most patients is assumed to be brown (10). As a strategy for personalized medicine, small archetypal groups dominated by one mechanism would be identified to study

the mechanism in isolation. While this model fits well with the commonly held view of complex diseases, it does not offer many actionable new avenues for research or clinical implementation, and it does not recognize the easily measurable differences in clinical features between patients.

Instead, we propose an intermediary model (Fig. 1C), painting with a broader brush and colors that reflect major clinical parameters instead of individual molecular pathways. This model still assumes that diabetes is caused by many overlapping mechanisms, but it postulates that most patients have a predominant color and that dividing patients into shades of red, green, and blue is more useful and informative than thinking of them all as shades of brown, even if some have a muddled or intermediary color. Some mechanisms might play roles, to different extents, in all individuals with diabetes, but it seems reasonable to assume that different pathways dominate, or are even uniquely involved, in leaner patients with insulin deficiency and obese patients with severe insulin resistance and that studying them separately offers advantages similar to studying archetypal groups, with a trade-off of less specificity for larger, more inclusive groups.

#### Subclassification of Adult-Onset Diabetes

ANDIS (All New Diabetics In Scania) is a large diabetes cohort started in 2008, with the purpose of studying diabetes heterogeneity. ANDIS aims to include all newly diagnosed individuals with any type of diabetes in the Scania region in southern Sweden, within 1 year from diagnosis. To date, it includes >20,000 individuals representing >90% of newly diagnosed patients. At registration, two blood measurements are added to standard measurements: glutamate decarboxylase autoantibodies (GADA) and C-peptide.

In this cohort, we stratified adult individuals into subtypes using a data-driven approach based on the most relevant, easily available, clinical variables for individuals with diabetes: GADA, BMI, glycosylated hemoglobin (HbA<sub>1c</sub>), age at diabetes onset,  $\beta$ -cell function (HOMA2-B), and insulin resistance (HOMA2-IR) estimated from fasting glucose and C-peptide (12). These variables were selected based on clinical experience and current knowledge of T2D, postulating that diabetes develops when insulin secretion fails to meet the demands of insulin resistance (2). Hyperglycemia can thus result from either a deficient insulin secretion, the extreme being T1D, or from severe insulin resistance, with T2D patients seen across this spectrum. Including key measures of the pathogenesis of diabetes, i.e., measures of insulin secretion and action in the definition of diagnosis therefore seems logical.

We applied two different methods for clustering (12). The first is TwoStep, which determines the optimal number of clusters based on silhouette width, followed by hierarchical clustering. In separate analyses for males and females, the optimal number of clusters was five, which was replicated in the Scania Diabetes Registry (SDR) cohort. One

cluster was completely defined by GADA positivity and was therefore referred to as severe autoimmune diabetes (SAID). Using the k-means method, we could identify clusters of similar sizes with the same combination of clinical characteristics in four independent cohorts, including ANDIS and SDR.

### Characteristics of Subtypes

The five identified clusters had different clinical characteristics, disease progression, and outcomes (Fig. 2). The SAID cluster, defined by presence of GADA, includes antibody-positive individuals traditionally referred to as T1D and LADA and represented 6% of adult individuals in ANDIS. SAID was characterized by relatively early disease onset, low insulin secretion, relatively low BMI, and poor metabolic control (high HbA<sub>1c</sub>). Measuring additional autoantibodies may increase the prevalence of this subtype, as it has been shown that up to 7% of patients from the GADA-negative clusters have positive islet cell antibodies and/or insulin autoantibodies (13). The second cluster, severe insulin-deficient diabetes (SIDD), comprised 18% of included individuals, with similar characteristics as SAID but without GADA. Patients in the third cluster (15%), severe insulin-resistant diabetes (SIRD), were characterized by very high HOMA2-IR and HOMA2-B and high BMI but low HbA<sub>1c</sub>. The fourth cluster, mild obesity-related diabetes (MOD) (22%) was also characterized by high BMI but not by insulin resistance. The fifth and largest cluster (39%), mild age-related diabetes (MARD), had late-onset diabetes but otherwise no extreme characteristics. Based on these results, we suggested a new subclassification system for T2D (12,14,15).

### Risk of Complications

The clusters were not only different in their characteristics, but they also had different risks of diabetes related outcomes (12).

The SAID and SIDD clusters had very high HbA<sub>1c</sub> at diagnosis (31% and 25%, respectively, presented with ketoacidosis) and progressed more rapidly to insulin treatment compared with the other clusters. The SIDD group had the highest prevalence of diabetic retinopathy, with

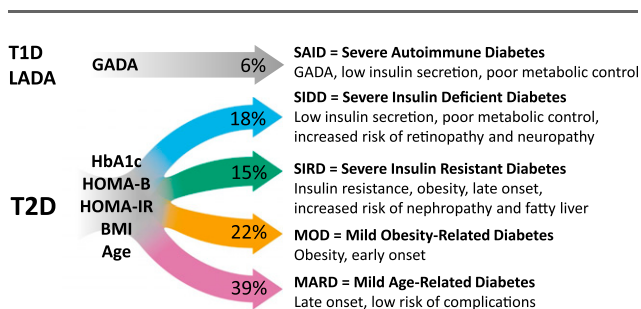
23% showing signs of at least mild retinopathy even when scanned soon after diagnosis.

The new diabetes clusters were further studied in the German Diabetes Study (GDS), where the low C-peptide secretory capacity of SIDD patients was confirmed using intravenous glucose tolerance tests (13). In GDS, SIDD patients also had higher prevalence of diabetic sensorimotor polyneuropathy and cardiac autonomic neuropathy at diagnosis. In spite of restored glucose homeostasis at the 5-year follow-up, neuronal signaling and nerve function were not restored (13). These results indicate that SIDD patients would benefit from early, intense treatment, frequent monitoring of complications, and sensitive diagnostic methods for early prediction.

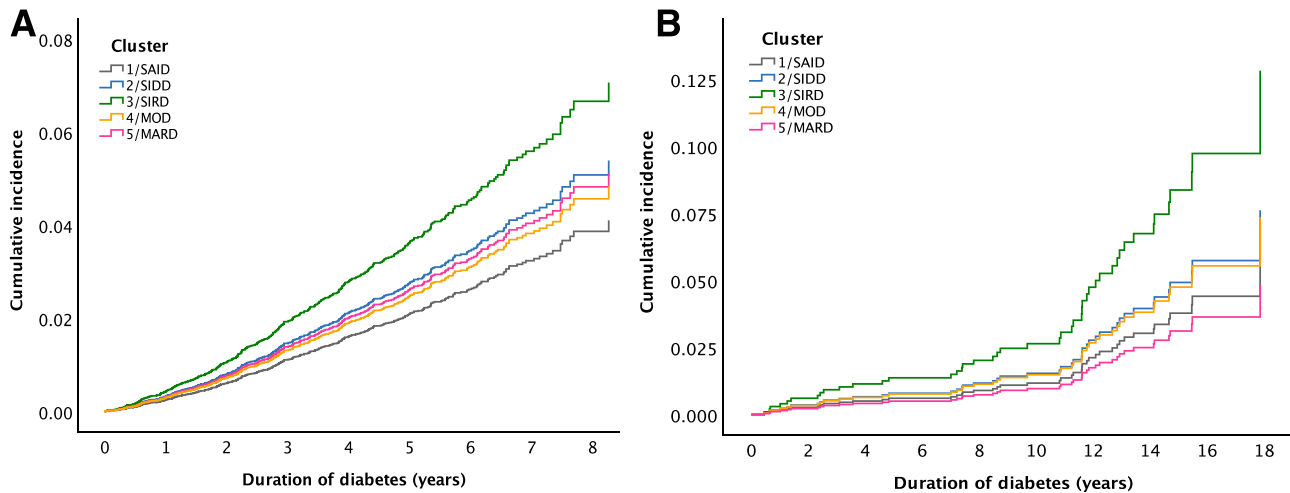
Based on extrapolation from T1D, many clinicians have been misled to think that microangiopathic complications, such as retinopathy, neuropathy, and nephropathy, coincide also in T2D. However, while retinopathy and neuropathy clustered in SIDD individuals, the SIRD subtype had the highest risk of developing diabetic kidney disease (DKD) (12). The SIRD subtype had the lowest mean estimated glomerular filtration rate (eGFR) at diabetes diagnosis and were at increased risk of developing chronic kidney disease (CKD), macroalbuminuria, and end-stage renal disease (ESRD). In the SDR cohort, SIRD patients had two times higher risk of developing CKD and macroalbuminuria and almost five times higher risk of ESRD after adjustment for age and sex than MARD patients. Increased DKD in early stages of SIRD was also replicated in GDS (13).

Dennis et al. (15) looked at risk for CKD stage 3A during the first 4 years after diagnosis in two clinical trials, ADOPT (A Diabetes Outcome Progression Trial) and RECORD (Rosiglitazone Evaluated for Cardiac Outcomes and Regulation of Glycaemia in Diabetes). Despite inclusion criteria that exclude most of the typical SIDD and SIRD individuals (16), they confirmed early signs of kidney disease in SIRD but could not replicate differences in progression after adjustment for initial eGFR (15). In contrast, the increased risk of kidney disease in SIRD remained significant for all CKD stages after adjustment for initial eGFR as well as sex and age in ANDIS (12). In SDR, with an average follow-up of 11 years, the odds ratio for ESRD in SIRD (compared with MARD) was 3.61 ( $P = 3.9 \times 10^{-5}$ ) after adjustment for initial eGFR (12). Even after exclusion of individuals with CKD at diagnosis, the SIRD group had increased risk after adjustment for eGFR and other commonly used variables (Fig. 3).

The relationship between insulin resistance and kidney disease is complex. Insulin resistance is a common and early alteration in CKD and almost universal in ESRD (17). In healthy individuals, more than half of the plasma insulin is cleared by the kidney (18). Zaharia et al. (13) used hyperinsulinemic-euglycemic clamp to estimate whole-body insulin resistance showing that whole-body insulin



**Figure 2**—Novel diabetes subtype characteristics. Overview of distribution and characteristics of subtypes generated by clustering based on clinical parameters in the Swedish ANDIS cohort.



**Figure 3**—Adjusted risk of diabetic kidney complications. Cumulative incidence of CKD stage 3A in ANDIS ( $HR_{SIRDvsMARD} = 1.39, P = 0.018$ ) (A) and ESRD in SDR ( $HR_{SIRDvsMARD} = 2.53, P = 0.047$ ) (B) in individuals with eGFR  $>60$  at the first measurement after diagnosis, adjusted for sex, age, BMI, HbA<sub>1c</sub>, and first eGFR. HR, hazard ratio.

sensitivity was similar to HOMA2-IR estimates, with the lowest M-values in SIRD patients both at diagnosis and after 5 years of follow-up. This shows that SIRD patients are truly insulin resistant and that the increased HOMA2-IR is not merely a result of impaired C-peptide clearance due to reduced kidney function (19).

SIRD patients also had the highest frequency of non-alcoholic fatty liver disease (NAFLD), as described in ANDIS, defined by two pathological measurements of the liver enzyme ALT and high BMI (12). In line with these findings, individuals assigned to the SIRD cluster in GDS had the highest hepatocellular lipid content (19% compared with  $<7\%$  for other clusters) and the highest NAFLD fibrosis scores, fatty liver index, and AST-to-platelet ratio index (13).

#### Subtypes of Diabetes in Other Populations

An important question is whether the subclassification can be replicated and used also in other populations. India and China are the most populated countries with the fastest growing economies in the world, and the prevalence of diabetes has tripled in China and doubled in India in less than two decades (20). Both Indian and Chinese populations develop diabetes at a younger age and at lower BMI than Caucasians (21,22). The body composition of South Asians in general is also different compared with the Caucasian population. For a given BMI, the fat percentage is significantly higher in South Asians, accompanied by higher insulin resistance (23).

An effort to identify the new subtypes of diabetes in the Chinese population was published recently (24). The novel diabetes clustering based on commonly available phenotypes was robustly replicated in 2,316 participants with newly diagnosed diabetes from the China National Diabetes and Metabolic Disorders Study (CNDMDS) and 685 participants from the National Health and Nutrition Examination Survey (NHANES III) from China and the

U.S. In the absence of GADA, four clusters could be identified by k-means clustering using age at diagnosis, BMI, HbA<sub>1c</sub> (or alternatively mean plasma glucose), HOMA2-B, and HOMA2-IR. MARD comprised nearly half the participants in both studies, followed by MOD, whereas SIRD and SIDD were least prevalent. However, SIDD was more prevalent in Chinese than in Caucasian populations. The cluster distribution was similar in other ethnic groups including non-Hispanic white and non-Hispanic black participants (24). The four clusters recapitulated the cluster characteristics defined in the Swedish population and other follow-up studies based on European populations, suggesting at least some generalizability of the classification in non-European populations.

Similar efforts are underway in other populations. We have confirmed the generalizability of the classification system in an Indian population diagnosed at  $<45$  years of age (R.B.P., unpublished observations), where we applied the clustering method to 972 T2D patients  $\sim 10$  years post-diagnosis. Here we first mapped to cluster centers from the ANDIS cohort and subsequently confirmed the subtypes by applying de novo k-means clustering wherein we obtained a  $>80\%$  concordance. Using both de novo clustering and a reference is valuable since the de novo clustering allows validation of the stability and reproducibility of clusters in different populations, while using a reference allows comparison between populations. We identified the expected four clusters, albeit with different distributions. Contrary to popular belief, the insulin-deficient SIDD cluster was predominant followed by MOD. Higher degree of nephropathy, retinopathy, and neuropathy was seen in the SIDD cluster compared with the other subtypes (R.B.P., O. Asplund, S. Shukla, et al., unpublished observations). While these data offer insights into the generalizability of the clustering, further large-scale studies in unselected cohorts are needed before this can be taken into the clinic.



### **Important Considerations for the Subclassification**

While the clustering strategy has strong advantages in form of simplicity and utility, it also has limitations. We have shown that clustering is reproducible in unselected European diabetes cohorts using C-peptide and glucose measured at diagnosis and that individuals with these data can be individually classified using ANDIS as reference (12). This does not prove that the clustering method will produce reliable and comparable results in cohorts selected using different inclusion criteria, surrogate variables, or lower-quality data. Using a reference for classification can solve some of these problems but requires large high-quality population-specific cohorts as reference. For European populations, ANDIS can be used as a reference, and a simple online tool allows classification for single individuals as well as full cohorts directly from the clinical parameters already measured (O. Asplund, O. Hansson, R.B.P., et al., unpublished observations). Importantly, this eliminates the need for scaling full-cohort data that can cause problems in selected cohorts. This tool could easily be expanded to include a selection of reference populations of different ethnicity as well as adaptations for different treatment status and duration of diabetes. However, such reference data from large unselected cohorts is still missing and the validity of the methods in other populations needs to be tested. Once available and tested, such a tool could be made available to clinicians to allow direct classification of patients.

The importance of correct classification is illustrated by a recent study where Kahkoska et al. (25) attempted a validation of the subtype and assessed their association with diabetes complications in the DEVOTE (A Trial Comparing Cardiovascular Safety of Insulin Degludec Versus Insulin Glargine in Patients With Type 2 Diabetes at High Risk of Cardiovascular Events), LEADER (Liraglutide Effect and Action in Diabetes: Evaluation of Cardiovascular Outcome Results), and SUSTAIN-6 (Trial to Evaluate Cardiovascular and Other Long-term Outcomes With Semaglutide in Subjects With Type 2 Diabetes-6) cardiovascular outcomes trials. Instead of de novo clustering, they used ANDIS as reference to assign study participants into clusters based on age at diabetes diagnosis, HbA<sub>1c</sub>, and BMI only. They identified the highest risk for cardiovascular events in the cluster of participants with high HbA<sub>1c</sub> and low BMI that most closely resembles SIDD but could not replicate the increased risk of kidney disease in cluster B, which most resembles SIRD. This is likely explained by poor performance of the simplified clustering strategy for this subtype. Using the same method in ANDIS, cluster A would include 90% SIDD patients but cluster B only 40% SIRD patients. Given the differences in inclusion criteria, the proportions could be even smaller in the clinical trials. This emphasizes the importance of the HOMA measures for identification of the high-risk SIRD group and the importance of properly validating alternative clustering methods.

An important question is whether clustering gives the same results at diagnosis and after a longer duration of diabetes, and to what extent individuals will move from

one cluster to another. In the DIREVA (DIabetes REgistry VAse) study from Finland, we clustered both individuals registered within 2 years after diagnosis and individuals with longer duration of diabetes with similar results supporting the robustness of clusters and limited influence of disease duration. One of the few studies addressing this question was GDS, which showed that 5 years after diagnosis 23% could change cluster allocation (13). However, the generalizability of these results is uncertain given that it is a selected cohort and based on only 367 individuals.

### **Etiological Differences and Genetics**

Another question concerns etiological differences between clusters. Are they to some extent mechanistically distinct, or do the same mechanisms operate in all subtypes? Or do the clusters perhaps reflect different stages of the natural progression of T2D, with cluster differences depending on how long the individual had undiagnosed diabetes? Or are SIDD patients a group of GADA-negative autoimmune patients? One way to answer this question is using genetic information. More than 400 loci affecting risk of T2D have been identified; many of them are also associated with subphenotypes such as BMI, insulin secretion, and insulin resistance (8). These genetic associations help shed light on what pathways are involved in disease etiology.

Genetically, SIRD stands out by neither being associated with the well-established T2D locus in *TCF7L2* (26,27) nor with a weighed risk score for insulin secretion, which are both strongly associated with SIDD, MOD, and MARD (12), clearly showing that SIRD is a distinct subtype with at least partially different etiology. Another important conclusion from genetics is that only SAID is associated with T1D-associated single nucleotide polymorphisms in the HLA region, giving no indication of involvement of the adaptive immune system in the development of SIDD, and thereby clearly distinguishing it from autoimmune diabetes (12,28). Recently, a genetic variant in *PNPLA3* was shown to associate positively with hepatocellular lipid content and the SIRD subtype. Hepatocellular lipid content was higher in the SIRD group compared with the MOD, MARD, and SAID groups and a glucose-tolerant control group. Although the *PNPLA3* polymorphism did not directly associate with whole-body insulin sensitivity in SIRD, the G-allele carriers had higher circulating free fatty acid concentrations and greater adipose tissue insulin resistance compared with noncarriers (29).

A more comprehensive genome-wide analysis of genetic differences between clusters has also been performed, suggesting stronger heritability for the SIDD and MOD groups (D. Mansour-Aly, O. Prakash Dwivedi, R.B.P., et al., unpublished observations). Thus far, hundreds of genetic loci have been implicated in various diabetes syndromes. Deep phenotyping and a more precise definition can lead to identification of more precise pathways and mechanisms, going beyond what the palette model proposes.

### **Clinical Importance**

The fundamental clinical differences between the subtypes suggest that they could benefit from a pathogenetically defined treatment, addressing their main dysfunction (e.g., insulin deficiency in SIDD and insulin resistance in SIRD). Identifying individuals with autoimmune diabetes, in itself a heterogeneous syndrome, at an early stage also has relevance for choice of medication, providing an opportunity to prevent  $\beta$ -cell failure and avoid T2D medication that might even exacerbate the autoimmune process (4).

The recommendations for management of T2D issued by the American Diabetes Association and the European Association for the Study of Diabetes emphasize the importance of personalization of treatment depending on patient characteristics and comorbidities (30). For patients with established cardiovascular disease or CKD, treatment with glucagon-like peptide 1 receptor agonists (GLP-1RA) or sodium–glucose cotransporter 2 inhibitors (SGLT2i) is recommended, but for other patients, selection between existing treatment options is focused on glucose reduction by stepwise addition of medication on top of metformin, and guided by severity of hyperglycemia, obesity, or cost restrictions. Apart from obesity, the causes and nature of underlying defects are not considered.

Individuals with SIRD show lower eGFR at diagnosis, indicating that the pathological process starts before diabetes diagnosis. Insulin resistance has long been recognized as a risk factor for DKD and has been shown to contribute to the development of disease through multiple mechanisms (31). Podocyte-specific deletion of the insulin receptor in mice causes albuminuria, together with histological features that recapitulate DKD, even in a normoglycemic environment (32). The fact that the SIRD subtype develops DKD in spite of relatively good metabolic control and low HbA<sub>1c</sub> levels indicates that treatment of these patients should not aim solely at lowering HbA<sub>1c</sub> but that improving insulin sensitivity could be beneficial to prevent complications.

The ADOPT and RECORD trials explored three different treatments in GADA-negative patients (15). The insulin sensitizer rosiglitazone, a thiazolidinedione, showed the strongest effect in reducing HbA<sub>1c</sub> in the SIRD subgroup (15). Unfortunately, the effect on eGFR was not tested, but it has previously been shown that thiazolidinediones are effective in reducing albuminuria, even when achieving the same HbA<sub>1c</sub> targets (33). Given the apparent key role of insulin resistance in development of DKD, this effect could be even stronger in SIRD patients.

Another finding in the ADOPT trial was that age-related (MARD) patients responded best to the insulin secretagogue sulfonylurea (15). In the ANDIS cohort, MARD and SIRD patients had been prescribed similar treatment in spite of considerable differences in risk of complications.

In the same article, Dennis et al. suggested that simple clinical variables (sex, age, BMI, and HbA<sub>1c</sub>) perform at least as well as clusters for selecting therapy. However, because there are limitations of the study, e.g., exclusion of the most typical patients from the severe subtypes, the types of medication tested, and the use of HbA<sub>1c</sub> reduction

as the only outcome, this remains to be further tested. It will be interesting to see corresponding studies in clinical trials testing treatments that address the insulin deficiency in SAID and SIDD, such as insulin or GLP-1RA, and effects of insulin sensitizers on complications in SIRD since HOMA measurements are important for the identification of this subtype.

Another clinical benefit is the possibility to focus resources, including more frequent screening and immediate intensive treatment, to the individuals most likely to develop complications, i.e., the three severe subtypes, whereas the milder forms could perhaps be managed safely by lifestyle interventions and standard care and given lower priority for screenings for microvascular complications.

### **Other Clustering Efforts**

There have been a few other efforts to cluster T2D patients using clinical variables. For example, Li et al. (34) used patient electronic medical records, identifying three patient-patient networks with different characteristics, genetic associations, and comorbidities (including everything from micro- and macrovascular complications to allergies and HIV infections). Vaccaro et al. (35) used decision tree-based clustering based on sex, urinary albumin-to-creatinine ratio, and BMI in the TOSCA. IT (Thiazolidinediones or Sulfonylureas and Cardiovascular Accidents Intervention Trial) trial. A group including male patients with albumin-to-creatinine ratio  $>9$  mg/g and BMI  $>28.7$  kg/m<sup>2</sup> had several conditions associated with insulin resistance including high waist circumference, blood pressure, triglycerides, and HDL cholesterol as well as increased risk of cardiovascular end points, which was better prevented by pioglitazone than sulfonylurea (35).

Given the robustness of genetics compared with phenotypes, efforts have also been made to identify subtypes using genetics. In one such approach, multivariant-trait association patterns obtained from GWAS across various traits were leveraged to identify shared disease mechanisms based on the assumption that variants that act along a shared pathway will have similar directional impact on subphenotypes (36). A soft clustering method was applied to group variant-trait associations from publicly available GWAS for 94 known T2D variants and 47 T2D-related traits. Five clusters were obtained, of which two were related to insulin deficiency and three to insulin resistance. Interestingly, seven out of the ten variants associated with SIDD in ANDIS had the strongest weights in the proinsulin/ $\beta$ -cell genetic clusters in the study by Udler et al.; the genetic obesity cluster seemed to correspond to MOD, whereas the genetically obtained insulin resistance cluster shared genetic associations with SIRD, with four variants from Ahlqvist et al. having the strongest weights in the liver/lipid and lipodystrophy clusters in Udler et al. While genetics is certainly more robust over time, clinical application is more problematic. Moreover, it does not reflect the environmental risk factors; heritability is only a minor

determinant of risk of T2D. Further, the identified risk loci for T2D only explain a small proportion of the heritability.

None of these methods have the same reproducibility and utility as the ANDIS-based subtypes.

## Conclusions

Easily measurable clinical parameters can be used to subclassify individuals with T2D into five groups with differing characteristics and disease progression. These clusters have been shown to be stable and reproducible in several populations. Classification can be done also for single individuals and selected cohorts using a reference population, greatly expanding the utility of the method and opening avenues for personalized, or at least stratified, medicine.

A key component in the classification is measurement of C-peptide. Estimates of insulin secretion and action based on fasting C-peptide and glucose concentrations allow identification of the severely insulin-resistant group, SIRD, that is currently receiving similar treatment as the much milder MARD subtype, enabling treatment focused on protecting the kidneys by increasing insulin sensitivity rather than lowering glucose. The insulin indices also help to identify the SIDD subtype, highlighting the need for early intensive treatment that addresses the insulin deficiency and the need for close monitoring of microvascular complications. GADA measurements can also guide clinicians toward treatment addressing the resulting insulin deficiency, focusing on  $\beta$ -cell preservation.

Another great advantage of defining subtypes with less variability and a more homogenous pathology is that it could facilitate and improve power to study individual disease pathways even if the groups are not completely distinct.

It might also be possible to refine the ANDIS stratification further through inclusion of additional cluster variables, such as biomarkers, genotypes, or genetic risk scores. More advanced methods using artificial intelligence also offer a promising avenue toward subtypes with even higher resolution and better predictive power.

We therefore believe that subclassification using these clinically relevant parameters provides a framework for the next step toward personalized medicine in diabetes.

**Funding.** This work was supported by grants from the Swedish Research Council (2017-02688), Diabetes Research and Wellness Network Sweden (25-420 PG), and Swedish Heart-Lung Foundation.

**Duality of Interest.** No potential conflicts of interest relevant to this article were reported.

**Prior Presentation.** Parts of this work were presented at the 80th Scientific Sessions of the American Diabetes Association, 12–16 June 2020.

## References

- Tuomi T, Santoro N, Caprio S, Cai M, Weng J, Groop L. The many faces of diabetes: a disease with increasing heterogeneity. *Lancet* 2014;383:1084–1094
- American Diabetes Association. 2. Classification and diagnosis of diabetes: *Standards of Medical Care in Diabetes—2019*. *Diabetes Care* 2019;42(Suppl. 1):S13–S28
- Regnell SE, Lernmark Å. Early prediction of autoimmune (type 1) diabetes. *Diabetologia* 2017;60:1370–1381
- Buzzetti R, Zampetti S, Maddaloni E. Adult-onset autoimmune diabetes: current knowledge and implications for management. *Nat Rev Endocrinol* 2017;13:674–686
- Pearson ER, Flechtner I, Njølstad PR, et al.; Neonatal Diabetes International Collaborative Group. Switching from insulin to oral sulfonylureas in patients with diabetes due to Kir6.2 mutations. *N Engl J Med* 2006;355:467–477
- Unwin N, Shaw J, Zimmet P, Alberti KG. Impaired glucose tolerance and impaired fasting glycaemia: the current status on definition and intervention. *Diabet Med* 2002;19:708–723
- Franks PW, McCarthy MI. Exposing the exposures responsible for type 2 diabetes and obesity. *Science* 2016;354:69–73
- Mahajan A, Wessel J, Willems SM, et al.; ExomeBP Consortium; MAGIC Consortium; GIANT Consortium. Refining the accuracy of validated target identification through coding variant fine-mapping in type 2 diabetes. *Nat Genet* 2018;50:559–571
- Pearson ER. Type 2 diabetes: a multifaceted disease. *Diabetologia* 2019;62:1107–1112
- McCarthy MI. Painting a new picture of personalised medicine for diabetes. *Diabetologia* 2017;60:793–799
- Leslie RD, Grant SFA. The dynamic origins of type 1 diabetes. *Diabetes Care* 2018;41:2441–2443
- Ahlgqvist E, Storm P, Käräjämäki A, et al. Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables. *Lancet Diabetes Endocrinol* 2018;6:361–369
- Zaharia OP, Strassburger K, Strom A, et al.; German Diabetes Study Group. Risk of diabetes-associated diseases in subgroups of patients with recent-onset diabetes: a 5-year follow-up study. *Lancet Diabetes Endocrinol* 2019;7:684–694
- Ahlgqvist E, Prasad RB, Tuomi T, Rosengren A, Groop L. Novel diabetes subgroups - Authors' reply. *Lancet Diabetes Endocrinol* 2018;6:440–441
- Dennis JM, Shields BM, Henley WE, Jones AG, Hattersley AT. 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* 2019;7:442–451
- Ahlgqvist E, Tuomi T, Groop L. Clusters provide a better holistic view of type 2 diabetes than simple clinical features. *Lancet Diabetes Endocrinol* 2019;7:668–669
- Spoto B, Pisano A, Zoccali C. Insulin resistance in chronic kidney disease: a systematic review. *Am J Physiol Renal Physiol* 2016;311:F1087–F1108
- Rabkin R, Ryan MP, Duckworth WC. The renal metabolism of insulin. *Diabetologia* 1984;27:351–357
- Wallace TM, Levy JC, Matthews DR. Use and abuse of HOMA modeling. *Diabetes Care* 2004;27:1487–1495
- International Diabetes Federation. *IDF Diabetes Atlas, 9th edition*, 2019. Accessed 5 May 2020. Available from <https://diabetesatlas.org>
- Shelgikar KM, Hockaday TD, Yajnik CS. Central rather than generalized obesity is related to hyperglycaemia in Asian Indian subjects. *Diabet Med* 1991;8:712–717
- Boffetta P, McLerran D, Chen Y, et al. Body mass index and diabetes in Asia: a cross-sectional pooled analysis of 900,000 individuals in the Asia cohort consortium. *PLoS One* 2011;6:e19930
- Yajnik CS, Yudkin JS. The Y-Y paradox. *Lancet* 2004;363:163
- Zou X, Zhou X, Zhu Z, Ji L. Novel subgroups of patients with adult-onset diabetes in Chinese and US populations. *Lancet Diabetes Endocrinol* 2019;7:9–11
- Kahkoska AR, Geybels MS, Klein KR, 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*. [Epub ahead of print 20 April 2020]. DOI: 10.1111/dom.14063
- Grant SF, Thorleifsson G, Reynisdottir I, et al. Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. *Nat Genet* 2006;38:320–323
- Saxena R, Gianniny L, Burt NP, et al. Common single nucleotide polymorphisms in TCF7L2 are reproducibly associated with type 2 diabetes and reduce the insulin response to glucose in nondiabetic individuals. *Diabetes* 2006;55:2890–2895

28. Cousminer DL, Ahlqvist E, Mishra R, et al.; Bone Mineral Density in Childhood Study. First genome-wide association study of latent autoimmune diabetes in adults reveals novel insights linking immune and metabolic diabetes. *Diabetes Care* 2018;41:2396–2403
29. Zaharia OP, Strassburger K, Knebel B, et al. Role of patatin-like phospholipase domain-containing 3 gene for hepatic lipid content and insulin resistance in diabetes. *Diabetes Care* 2020;43:2161–2168
30. Davies MJ, D'Alessio DA, Fradkin J, et al. Management of hyperglycemia in type 2 diabetes, 2018. A consensus report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). *Diabetes Care* 2018;41:2669–2701
31. Mitrofanova A, Sosa MA, Fornoni A. Lipid mediators of insulin signaling in diabetic kidney disease. *Am J Physiol Renal Physiol* 2019;317:F1241–F1252
32. Welsh GI, Hale LJ, Eremina V, et al. Insulin signaling to the glomerular podocyte is critical for normal kidney function. *Cell Metab* 2010;12:329–340
33. Scherthner G, Matthews DR, Charbonnel B, Hanefeld M, Brunetti P; Quartet Study Group. Efficacy and safety of pioglitazone versus metformin in patients with type 2 diabetes mellitus: a double-blind, randomized trial. *J Clin Endocrinol Metab* 2004;89:6068–6076
34. Li L, Cheng WY, Glicksberg BS, et al. Identification of type 2 diabetes subgroups through topological analysis of patient similarity. *Sci Transl Med* 2015;7:311ra174
35. Vaccaro O, Lucisano G, Masulli M, et al.; TOSCA.IT Investigators. Cardiovascular effects of pioglitazone or sulfonylureas according to pretreatment risk: moving toward personalized care. *J Clin Endocrinol Metab* 2019;104:3296–3302
36. Udler MS, Kim J, von Grotthuss M, et al.; METASTROKE and the ISGC. Type 2 diabetes genetic loci informed by multi-trait associations point to disease mechanisms and subtypes: a soft clustering analysis. *PLoS Med* 2018;15:e1002654