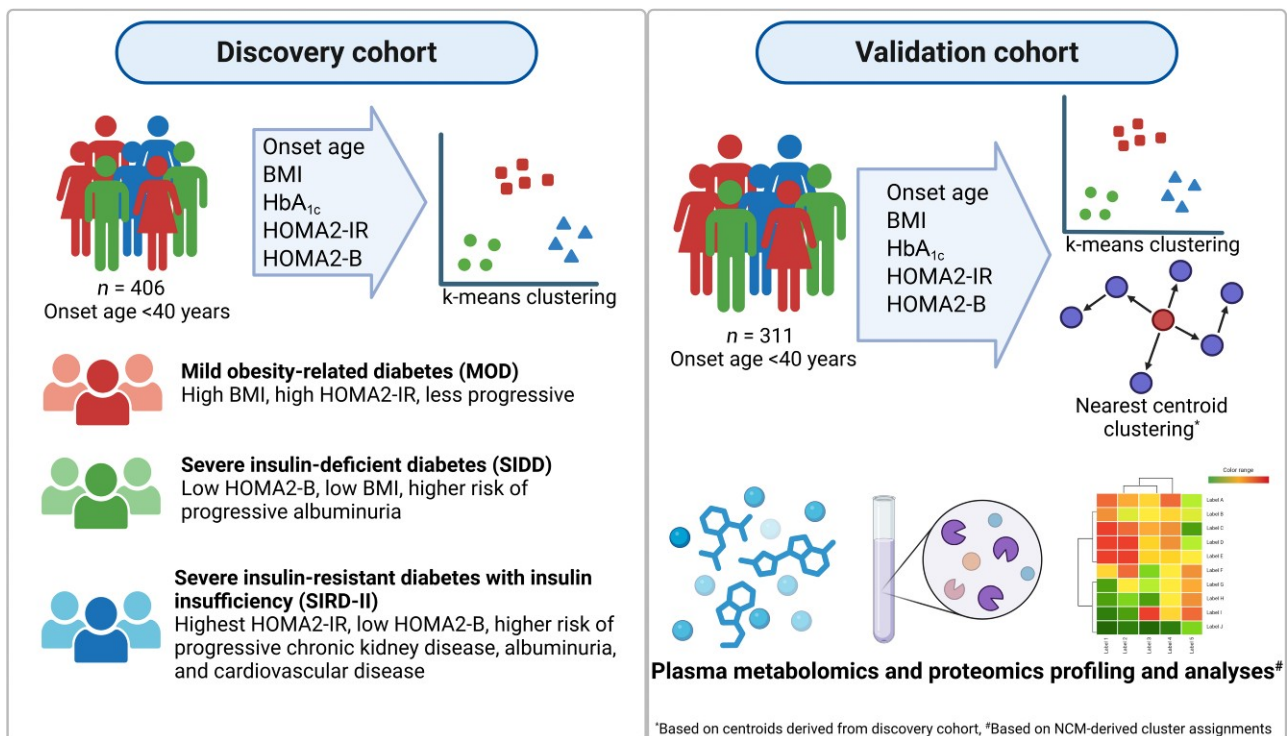


## Cluster Analysis of Younger-Onset Type 2 Diabetes in an Asian Cohort Reveals Distinct Subgroups With Differential Pathophysiology and Outcomes

Clara Si Hua Tan, Kai Xiang Kee, Huili Zheng, Kay Wye Sabrina Wong, Wan Ting Lovynn Chan, Yuzhen Song, Keven Ang, Tavintharan Subramaniam, Chee Fang Sum, and Su Chi Lim

*Diabetes* 2026;75(3):1–12 | <https://doi.org/10.2337/db25-0871>



HOMA2-B, HOMA for  $\beta$ -cell function; HOMA2-IR, HOMA for insulin resistance; NCM, nearest centroid method.

Created in BioRender. Tan, C. (2025) <https://BioRender.com/m04ynqw>



# Cluster Analysis of Younger-Onset Type 2 Diabetes in an Asian Cohort Reveals Distinct Subgroups With Differential Pathophysiology and Outcomes

Clara Si Hua Tan,<sup>1</sup> Kai Xiang Kee,<sup>1</sup> Huili Zheng,<sup>1</sup> Kay Wye Sabrina Wong,<sup>2,3</sup> Wan Ting Lovynn Chan,<sup>1</sup> Yuzhen Song,<sup>1</sup> Keven Ang,<sup>1</sup> Tavintharan Subramaniam,<sup>1,4</sup> Chee Fang Sum,<sup>4</sup> and Su Chi Lim<sup>1,3,4,5</sup>

<https://doi.org/10.2337/db25-0871>

Younger-onset type 2 diabetes (T2D, onset <40 years) represents a growing global health challenge, characterized by heterogenous pathophysiology and accelerated complications. Current ‘one-size-fits-all’ treatment approaches may be inadequate for this population. To address this heterogeneity, we performed clinical variable-based clustering using BMI, onset age, HbA<sub>1c</sub> and HOMA2 indices in 717 participants across discovery and validation cohorts. Three distinct subgroups were identified: mild obesity-related diabetes (MOD), severe insulin-deficient diabetes (SIDD) and severe insulin-resistant diabetes with insulin insufficiency (SIRD-II). Over median follow-up of 2.8 years, SIRD-II demonstrated 11-fold increased risk of progressive chronic kidney disease, while both SIDD and SIRD-II showed threefold increased risk for progressive albuminuria compared with MOD. SIRD-II also demonstrated 3.5-fold and 2.3-fold higher 10-year cardiovascular risk compared with SIDD and MOD respectively. Metabolomic analysis revealed distinct signatures: SIDD exhibited lower levels of lipids, amino acids, and inflammatory markers, while SIRD-II demonstrated elevated glucose, lipids, and branched-chain amino acids, suggesting glucolipotoxicity. Proteomics analysis validated previously reported biomarkers (IGFBP1, RTN4R, PLXNB2) and identified additional molecules (CDHR2, ERBB4, DPP6) that may shed light on disease mechanisms. In conclusion, younger-onset T2D exhibits distinct subgroups with differential pathobiology, molecular signatures, and clinical outcomes, suggesting the need for personalised precision diabetes care.

## ARTICLE HIGHLIGHTS

- To understand the heterogeneity of younger-onset type 2 diabetes, clinical data-driven clustering was performed, which identified three distinct subgroups that were replicated in an independent cohort.
- Compared with the mild obesity-related diabetes (MOD) subgroup, both severe insulin-deficient diabetes (SIDD) and severe insulin-resistant diabetes with insulin insufficiency (SIRD-II) subgroups had higher risk of developing diabetes-related complications.
- Differential molecular signatures confirmed the biological distinctiveness of younger-onset T2D subgroups and highlight potential mechanisms, such as glucolipotoxicity stress that may drive complications in the SIRD-II subgroup.
- Proteomic analyses validated previously reported biomarkers and identified novel candidates, providing a foundation for future mechanistic studies.

The global incidence of younger-onset type 2 diabetes (T2D), pragmatically defined as onset age before 40 years (1), has been increasing, with particularly pronounced trends among Asian populations (2–4). This phenomenon is attributed to the increasing propensity of obesity among the young (5). Younger-onset T2D is primarily driven by

<sup>1</sup>Clinical Research Unit, Khoo Teck Puat Hospital, Singapore

<sup>2</sup>Clinical Research Unit, National Healthcare Group Polyclinics, Singapore

<sup>3</sup>Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore

<sup>4</sup>Diabetes Centre, Admiralty Medical Centre, Singapore

<sup>5</sup>Saw Swee Hock School of Public Health, Singapore

Corresponding author: Su Chi Lim, [lim.su.chi@nhghealth.com.sg](mailto:lim.su.chi@nhghealth.com.sg)

Received 24 September 2025 and accepted 1 December 2025

This article contains supplementary material online at <https://doi.org/10.2337/figshare.30752879>.

© 2025 by the American Diabetes Association. Readers may use this work for educational, noncommercial purposes if properly cited and unaltered. This publication and its contents may not be reproduced, distributed, or used for text or data mining, machine learning, or other similar technologies without prior written permission. More information is available at <https://diabetesjournals.org/journals/pages/license>.

obesity-related insulin resistance (6) and concomitant  $\beta$ -cell dysfunction (7,8). Compared with older-onset T2D, individuals with younger-onset T2D demonstrate a more aggressive disease trajectory, characterized by earlier onset of diabetes-related complications and premature mortality, imposing significant burden on health-care systems (9,10). Moreover, current models of care are typically derived from older adult populations and do not adequately address the age-specific pathophysiology of younger-onset T2D, including rapid  $\beta$ -cell deterioration, severe insulin resistance, and accelerated hyperglycemic progression, nor do they account for phenotypic heterogeneity and varying risk profiles (11,12). Therefore, stratifying younger-onset T2D into more homogenous subgroups may be clinically helpful for early, targeted interventions that could modify disease trajectory and improve outcomes.

In recent years, data-driven cluster analyses have been adopted in several studies to subtype individuals with T2D into relatively more homogenous subgroups for disease risk stratification (13). The landmark study by Ahlqvist et al. (14) applied clinical variables encompassing glutamic acid decarboxylase, diabetes onset age, BMI, HbA<sub>1c</sub>, and HOMA2 (homeostatic model assessment 2) indices of  $\beta$ -cell function (HOMA2-B) and insulin resistance (HOMA2-IR) derived from fasting C-peptide to subclassify diabetes in adult-onset individuals. Five subgroups with distinct clinical pathophysiology and disease progression were identified. Using similar approach, many studies were able to replicate comparable subgroups in other diabetic populations with differential molecular signatures (15–17). However, similar to Ahlqvist et al., most diabetes subtyping studies have largely focused on older-onset diabetes, leaving uncertainty about their applicability to younger-onset T2D.

In this study, using similar clinical variables as Ahlqvist et al. (14), we applied de novo k-means clustering to younger-onset T2D patients from a multiethnic Asian population. We hypothesize that systematic subtyping of individuals with younger-onset T2D may reveal novel distinct subgroups characterized by differential clinical profiles, molecular signatures (thereby suggesting differential underlying pathobiology) and renal trajectories.

## RESEARCH DESIGN AND METHODS

### Study Cohorts

The discovery cohort consists of 406 patients (diabetes onset <40 years old) recruited between 2013–2024 (Supplementary Fig. 1, Supplementary Table 1). Participants were recruited from outpatient clinics within secondary care institutions in Singapore. Participants with estimated glomerular filtration rate (eGFR) <15 mL/min/1.73m<sup>2</sup> and/or incomplete data available for the clustering variables were excluded from the study. All participants underwent targeted 16-gene panel sequencing to exclude monogenic diabetes and comprehensive islet autoantibody testing (including GAD, ZnT8, and IA2) to rule out type 1 diabetes, following previously described protocols (18). Follow-up data were

collected by reviewing electronic medical records. The National Healthcare Group (NHG) Health Domain Specific Review Board of Singapore provided ethical approval for this study (ECOS 2019/00822), which was conducted following Declaration of Helsinki principles. Informed consent was obtained from all study participants.

The validation cohort comprised 311 younger-onset T2D participants (diabetes onset <40 years) from two independent studies: SMART2D (Study of Macro-Angiopathy and microvascular reactivity in type 2 diabetes) and DORIS (Diabetic Kidney Disease-Onset and progressive risk factors). Full cohort descriptions have been reported elsewhere (19,20). These participants were enrolled from secondary care outpatient clinics and a neighbouring primary care center between 2011 and 2020, yielding a combined study population of 717 participants across discovery and validation cohorts. T2D diagnosis was established by primary care physicians after ruling out T1D and other specific diabetes aetiologies. Inclusion required eGFR  $\geq$  15 mL/min/1.73m<sup>2</sup> and complete clustering variable datasets. Ethical approval for SMART2D and DORIS was attained from the NHG Health Domain Specific Review Board (ECOS 2024/3805 and 2017/00561, respectively).

### Clinical and Biochemical Assays

Following overnight fasting, blood and urine specimens were obtained. Data on diabetes duration and medication usage were determined through postenrollment medical record review. Anthropometric parameters were measured by trained research coordinators. BMI was derived using weight (kg) divided by height (m<sup>2</sup>). Age at diabetes onset was determined through review of medical records to identify the date of initial diabetes diagnosis. Laboratory analyses for plasma glucose, HbA<sub>1c</sub>, lipid profile, high-sensitive C-reactive protein (hsCRP), and urinary albumin-to-creatinine ratio (uACR) were conducted by the Department of Laboratory Medicine, Khoo Teck Puat Hospital, Singapore, an American College of Pathology-accredited lab. C-peptide levels were quantified in duplicates using sandwich enzyme-linked immunosorbent assay (ELISA) whereby both intra- and inter-assay coefficients of variation are below 5% (Mercodia, Uppsala, Sweden). eGFR was calculated using the CKD-EPI formula.

### Clinical Variable-Driven Cluster Analysis

We applied clustering analysis to 406 participants in the discovery cohort using clinical variables including BMI, age of diabetes onset, HbA<sub>1c</sub>, HOMA2-B and HOMA2-IR. HOMA2 estimates were computed based on fasting C-peptide using the HOMA calculator (<https://www.dtu.ox.ac.uk/homacalculator/>, version 2.2.3). HOMA2 estimates were log-transformed and z-score standardized before cluster analysis. Using the R package 'NbClust (version 3.0.1)', we determined optimal number of clusters based on consensus voting across 26 indices. Cluster stability was evaluated using the Jaccard index with bootstrapping through the R package 'fpc (version

2.2–13) (21). Each cluster was labeled by examining cluster variable median values. To explore potential sex-dependent differences, k-means clustering was performed separately for men and women using identical clustering parameters. To validate the three primary metabolic subgroups, similar de novo k-means clustering was performed on 311 participants from SMART2D and DORIS cohorts.

For clinical implementation, patients need to be assigned to clusters independently without de novo clustering of complete patient cohorts. Therefore, using cluster centroids derived from the discovery cohort, we assigned participants in the validation cohort to their nearest clusters based on minimal Euclidean distance (nearest centroid method, NCM). Subsequent omics analyses were based on the NCM-derived cluster assignments.

Statistical analyses were performed using R (version 4.1.1) and Stata/SE version 14 (StataCorp LLC, College Station, TX, USA). We reported continuous variables as mean  $\pm$  SD or median with interquartile range (IQR) according to data distribution determined by Shapiro-Wilk test. For normally distributed continuous variables, we used one-way ANOVA for between-group comparisons. For nonparametric data, we used Kruskal-Wallis test. Categorical variables were summarized as frequencies and percentages, with group differences evaluated using Pearson's  $\chi^2$  test.

### Cluster Analysis Using TG/HDL Ratio as a Proxy for Insulin Resistance

In patients with long-standing diabetes, progressive  $\beta$ -cell deterioration can lead to underestimation of insulin resistance when using HOMA2-IR, while C-peptide measurements for HOMA2 calculations may not be routinely accessible. We therefore investigated TG/HDL ratio as an alternative clustering variable. Alternative k-means clustering was performed retaining original variables (BMI, age of diabetes onset, HbA<sub>1c</sub>) and substituting HOMA2-IR with TG/HDL ratio, excluding HOMA2-B. 395 participants in the discovery cohort with available TG/HDL ratio were included. Consensus voting across 26 clustering validation indices confirmed optimal cluster number as  $k = 3$ . Agreement was assessed using specificity, sensitivity and Rand Index.

### Definition of Renal Outcomes and Statistical Analysis

Progressive chronic kidney disease (CKD) was defined as sustained eGFR decline of  $>40\%$  from baseline (22), over minimum one-year follow-up. Participants with baseline eGFR  $\leq 60$  mL/min/1.73m<sup>2</sup> were omitted from analysis. Progressive albuminuria was established as worsening albuminuria category according to Kidney Disease: Improving Global Outcomes (KDIGO) guidelines: progression from A1 to A2, A2 to A3 or A1 to A3 (representing transitions from normal/mildly increased to moderately increased, moderately increased to severely increased or normal/mildly increased to severely increased). The follow-up was censored on 15 May 2024.

Incidence rates for progressive CKD and albuminuria were presented as event number per 1000 person-years. The Kaplan-Meier approach was used to estimate cumulative risk, and survival differences between subgroups were evaluated using log-rank analysis. Cox proportional hazard regression models were fitted using R package 'survival (version 3.7.0)' to study the associations of subgroups identified in the discovery cohort with renal outcomes. Models were adjusted for index age, sex, and ethnicity in all analyses. Additional covariates included baseline eGFR for progressive CKD models and baseline diabetic kidney disease (eGFR  $<60$  mL/min/1.73m<sup>2</sup> or uACR  $>30$  mg/mmol) for progressive albuminuria models. Schoenfeld residuals were used to assess the proportional hazards assumption, with no violations identified.

### Cardiovascular Risk Assessment

Cardiovascular disease (CVD) risk was assessed using the SCORE2-Diabetes risk score (23) for participants with complete data ( $n = 343$ ). SCORE2-Diabetes is a risk prediction algorithm specifically designed for patients with diabetes, providing 10-year CVD risk estimates. The algorithm incorporates age, sex, smoking status, systolic blood pressure, total cholesterol, HDL-cholesterol, HbA<sub>1c</sub> and eGFR. We applied the low-risk region calibration, corresponding to Singapore's most recent age- and sex-standardized CVD mortality risk (76 per 100,000) (data source: <https://world-heart-federation.org/world-heart-observatory/countries/singapore/>). Risk categories were defined as low-to-moderate risk ( $<2.5\%$ ), high-risk (2.5% to  $<7.5\%$ ), and very high-risk ( $\geq 7.5\%$ ) based on the European Society of Cardiology guidelines (24). Differences in cardiovascular risk between subgroups were assessed using Kruskal-Wallis test due to nonnormal distribution of risk scores. Post hoc pairwise comparisons were performed using Wilcoxon rank-sum tests with Bonferroni correction for multiple comparisons.

### Metabolomics Assay and Data Analysis

Plasma metabolite profiling was performed using the Nightingale Nuclear Magnetic Resonance (NMR) platform, quantifying 249 metabolites (Nightingale Health, Ltd., Helsinki, Finland). All metabolites were included for analyses as  $<10\%$  of samples had metabolites below the limit of detection. Complete metabolomic data were available for all participants in the discovery cohort and 304 participants in the validation cohort. Prior to analysis, metabolite concentrations were log-transformed and standardized (mean = 0, SD = 1). Differences in metabolite concentrations across subgroups were evaluated using one-way ANOVA. A Bonferroni-adjusted significance threshold of  $P < 2.01 \times 10^{-4}$  (0.05/249) was applied to control for multiple comparisons.

### Proteomics Assay and Data Analysis

Plasma protein profiling was performed using proximity extension assay technology on the Olink platform (Olink Proteomics AB, Uppsala, Sweden). Four panels were analyzed:

cardiometabolic, inflammation, neurology, and oncology, measuring a total of 1,472 proteins. Complete proteomic data were available for 277 of the 311 participants in the validation cohort. After quality control, proteins that failed quality metrics or present in multiple panels were excluded, leaving 1,448 proteins for subsequent analyses. Protein concentrations were reported as normalized protein expression (NPX) values, representing the log<sub>2</sub>-transformed relative protein concentrations. The Kruskal-Wallis test was employed to assess between-group differences in protein levels. Statistical significance was set at  $P < 3.45 \times 10^{-5}$  (0.05/1448) to account for multiple testing using Bonferroni correction. For pairwise comparisons between subgroups, Wilcoxon rank-sum tests were performed, and the adjusted significance threshold was set at  $P < 1.15 \times 10^{-5}$  (0.05/(1448  $\times$  3)). Protein fold changes were calculated as log<sub>2</sub>(comparison group/reference group) from normalized expression values.

### Data and Resource Availability

All data generated during this study are included in the published article and its online supplementary files. Individual-level proteomics and metabolomics data are available from the corresponding author upon reasonable request and subject to appropriate data sharing agreements. No applicable resources were generated or analyzed during the current study.

## RESULTS

### Clinical Variable-Driven Cluster Analysis

In the discovery cohort, de novo cluster analysis by k-means identified three subgroups. Cluster stability was assessed using bootstrap resampling ( $n = 2,000$ ). The mean Jaccard indices were 0.63, 0.76, and 0.77 for clusters 1, 2 and 3 respectively, indicating that clusters 2 and 3 were stable ( $>0.75$ ), while cluster 1 showed a consistent pattern ( $>0.6$ ). The coordinates for cluster centers are provided in Supplementary Table 2. Cluster 1 (19% of participants) was characterized by insulin deficiency (median HOMA2-B 30.1%, IQR 14.8–44.3), early-onset diabetes (median age of onset 19 years old, IQR 13–24), normal BMI (median 23.8 kg/m<sup>2</sup>, IQR 21.4–26.2), and elevated HbA<sub>1c</sub> (median 7.6% [60 mmol/mol], IQR 7.0–8.5% [53–69 mmol/mol]). Based on these characteristics, particularly the pronounced insulin deficiency, we labeled this cluster as severe insulin-deficient diabetes (SIDD). Cluster 2, the largest subgroup (52% of participants), was characterized by obesity (median BMI 30.6 kg/m<sup>2</sup>, IQR 27.2–35.5), insulin resistance (median HOMA2-IR 2.5, IQR 1.8–3.3), and relatively good glycaemic control (median HbA<sub>1c</sub> 7.1% [54 mmol/mol], IQR 6.3–7.8% [45–62 mmol/mol]). Based on these features, we labeled this cluster as mild obesity-related diabetes (MOD). Cluster 3 (29% of participants) exhibited the most severe metabolic profile, characterized by both insulin resistance (median HOMA2-IR 3.0, IQR 2.0–4.4) and insulin deficiency (median HOMA2-B 31.5%, IQR 21.2–48.0), accompanied by obesity (median BMI 27.6 kg/m<sup>2</sup>, IQR 25.1–31.7), and poor

glycaemic control (median HbA<sub>1c</sub> 10.3% [89 mmol/mol], IQR 9.6–11.6% [81–103 mmol/mol]). This subgroup also had the longest diabetes duration (median 8 years, IQR 3–15) and most adverse lipid profile, with the highest total cholesterol (median 5.1 mmol/L, IQR 4.3–6.2) and triglyceride levels (median 1.9 mmol/L, IQR 1.4–3.2). Given the presence of marked insulin resistance and relative insulin deficiency, we labeled this cluster as severe insulin-resistant diabetes with insulin insufficiency (SIRD-II). The baseline characteristics of all study participants in the discovery cohort are presented in Table 1 and Supplementary Fig. 2. In additional analyses exploring sex differences, we observed higher median HOMA2-B in the male SIDD subgroup compared with the other two subgroups, likely reflecting relatively preserved  $\beta$ -cell function owing to shorter diabetes duration (median 2 years, IQR 1–8) in men. The direction of association for other clinical variables remained generally consistent with the discovery cohort (Supplementary Fig. 3).

The validation cohort differed from the discovery cohort in key characteristics, notably older age of diabetes onset (median 35 vs 27 years) and longer diabetes duration (median 10 vs 5 years) (Supplementary Table 1). Nonetheless, de novo k-means clustering similarly identified three stable subgroups, with all clusters achieving Jaccard indices above 0.75 across 2,000 bootstrap iterations. Detailed baseline characteristics are presented in Supplementary Table 3 and Supplementary Fig. 4. Similar to the discovery cohort, MOD emerged as the largest subgroup (46% of participants), characterized by obesity (median BMI 28.6 kg/m<sup>2</sup>, IQR 25.4–32.3), insulin resistance (median HOMA2-IR 2.1, IQR 1.7–2.8) and relatively good glycaemic control (median HbA<sub>1c</sub> 7.2% [55 mmol/mol], IQR 6.6–8.0% [49–64 mmol/mol]). The SIDD subgroup (32% of participants) exhibited insulin insufficiency with low HOMA2-B (median 24.0%, IQR 17.0–29.6), minimal insulin resistance (median 1.0, IQR, 0.7–1.4), and elevated HbA<sub>1c</sub> (median 8.8% [73 mmol/mol], IQR 7.9–9.6% [63–81 mmol/mol]). The SIRD-II subgroup demonstrated the most severe metabolic profile despite shortest diabetes duration (8 years, IQR 2–14), with the highest values for BMI (median 34.4 kg/m<sup>2</sup>, IQR 31.0–40.3), HOMA2-IR (median 2.8, IQR 2.2–3.9), and HbA<sub>1c</sub> (median HbA<sub>1c</sub> 9.0% [75 mmol/mol], IQR 8.1–10.2% [65–88 mmol/mol]). This subgroup also showed an adverse lipid profile with elevated total cholesterol (median 4.5 mmol/L, IQR 3.8–5.1), LDL-cholesterol (median 3.0 mmol/L, IQR 2.3–3.4), and triglycerides (median 1.8 mmol/L, IQR 1.4–2.4).

Using the nearest centroid approach to assign participants in the validation cohort yielded similar cluster distributions to the de novo clustering (Supplementary Fig. 5, Supplementary Table 4).

### Concordance between TG/HDL and HOMA2-Based Clustering Methods in Identifying Diabetes Subtypes

As an exploratory analysis, we tested TG/HDL ratio as an alternative clustering variable to HOMA2-IR. Comparing

**Table 1—Baseline characteristics of younger-onset T2D subgroups in the discovery cohort (n = 406)**

Clinical characteristics	MOD	SIDD	SIRD-II	P value
No. of participants	210 (52%)	77 (19%)	119 (29%)	—
Index age (years)	33 (27–38)	24 (17–34)	36 (30–43)	<b>&lt;0.001</b>
Age at diabetes onset (years)	29 (23–33)	19 (13–24)	28 (21–32)	<b>&lt;0.001</b>
Male	125 (60%)	29 (38%)	71 (60%)	<b>0.002</b>
Ethnicity				<b>0.003</b>
Chinese	122 (58%)	51 (66%)	58 (49%)	
Malay	54 (26%)	17 (22%)	35 (29%)	
Indian	27 (13%)	4 (5%)	26 (22%)	
Others	7 (3%)	5 (6%)	0 (0%)	
Diabetes duration (years)	3 (1–8)	5 (2–11)	8 (3–15)	<b>&lt;0.001</b>
BMI (kg/m <sup>2</sup> )	30.6 (27.2–35.5)	23.8 (21.4–26.2)	27.6 (25.1–31.7)	<b>&lt;0.001</b>
Waist circumference (cm)	98.0 (90.0–108.0)	83.0 (73.0–90.0)	92.8 (86.0–104.5)	<b>&lt;0.001</b>
Fasting plasma glucose (mmol/L)	7.0 (5.8–8.6)	8.7 (6.1–11.4)	12.9 (9.9–15.6)	<b>&lt;0.001</b>
HbA <sub>1c</sub> (%)	7.1 (6.3–7.8)	7.6 (7.0–8.5)	10.3 (9.6–11.6)	<b>&lt;0.001</b>
HbA <sub>1c</sub> (mmol/mol)	54.1 (45.4–61.7)	59.6 (53.0–69.4)	89.1 (81.4–103.8)	—
Cholesterol (mmol/L)	4.5 (3.9–5.2)	4.6 (4.0–5.6)	5.1 (4.3–6.2)	<b>&lt;0.001</b>
HDL-cholesterol (mmol/L)	1.1 (1.0–1.2)	1.3 (1.1–1.5)	1.1 (1.0–1.3)	<b>&lt;0.001</b>
LDL-cholesterol (mmol/L)	2.8 (2.3–3.5)	2.8 (2.4–3.6)	3.1 (2.4–3.9)	0.129
Triglyceride (mmol/L)	1.6 (1.2–2.3)	1.0 (0.8–1.5)	1.9 (1.4–3.2)	<b>&lt;0.001</b>
C-peptide (pmol/L)	992 (732–1366)	363 (196–480)	897 (653–1179)	<b>&lt;0.001</b>
TG/HDL ratio	1.5 (1.0–2.3)	0.8 (0.5–1.3)	1.7 (1.3–3.0)	<b>&lt;0.001</b>
HOMA2-B (%) <sup>§</sup>	87.2 (58.2–119.0)	30.1 (14.8–44.3)	31.5 (21.1–48.0)	<b>&lt;0.001</b>
HOMA2-IR <sup>§</sup>	2.5 (1.8–3.3)	1.0 (0.5–1.3)	3.0 (2.0–4.4)	<b>&lt;0.001</b>
hsCRP (mg/L)	2.0 (1.0–3.9)	1.2 (0.5–2.5)	2.5 (1.2–4.6)	<b>&lt;0.001</b>
Systolic blood pressure (mmHg)	128 (119–140)	120 (110–128)	126 (117–137)	<b>&lt;0.001</b>
Diastolic blood pressure (mmHg)	80 ± 12	74 ± 11	80 ± 12	<b>0.025</b>
CKD-Epi-eGFR (mL/min/1.73m <sup>2</sup> )	116 (101–125)	126 (116–135)	116 (105–129)	<b>&lt;0.001</b>
Urinary ACR (mg/mmol)	1.9 (0.9–5.9)	1.7 (0.8–3.8)	4.3 (1.4–34.7)	<b>&lt;0.001</b>
Diabetes treatment				
Insulin	61 (29%)	47 (61%)	61 (51%)	<b>&lt;0.001</b>
Metformin	187 (89%)	49 (64%)	106 (89%)	<b>&lt;0.001</b>
Sulfonylurea	54 (26%)	17 (22%)	51 (43%)	<b>0.001</b>
DPP4i	50 (24%)	10 (13%)	36 (30%)	<b>0.021</b>
SGLT2i	86 (41%)	8 (10%)	36 (30%)	<b>&lt;0.001</b>
Lipid-lowering treatment				
Statins	88 (42%)	16 (21%)	58 (49%)	<b>&lt;0.001</b>
Fibrates	13 (6%)	3 (4%)	20 (17%)	<b>0.002</b>
Ezetimibe	2 (2%)	0 (0%)	3 (3%)	0.454
Antihypertensive treatment				
ACE inhibitors	22 (10%)	4 (5%)	19 (16%)	0.063
ARBs	38 (18%)	8 (10%)	17 (14%)	0.277

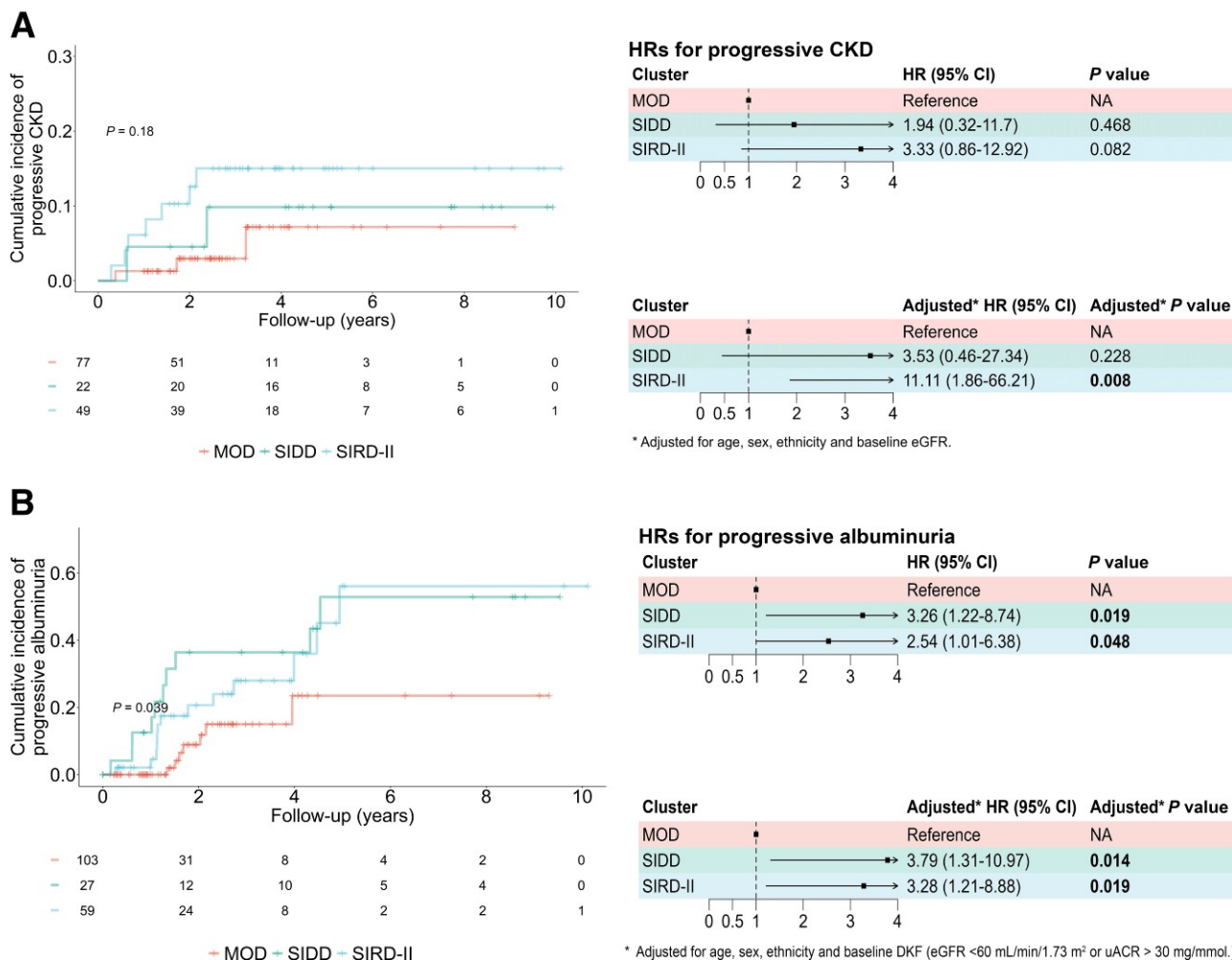
Data are presented as mean ± SD, median (IQR), or n (%). Bold values indicate  $P < 0.05$ . ACR, albumin-to-creatinine ratio; ACE inhibitors, angiotensin-converting enzyme inhibitors; ARB, angiotensin II receptor blockers; BMI, body mass index; CKD-Epi-eGFR, chronic kidney disease epidemiology collaboration – estimated glomerular filtration rate; DPP4i, dipeptidyl peptidase 4-inhibitors; HbA<sub>1c</sub>, hemoglobin A1C; HDL, high-density lipoprotein; hs-CRP, high-sensitivity C-reactive protein; LDL, low-density lipoprotein; MOD, mild obesity-related diabetes; SGLT2i, sodium-glucose transport protein 2 (SGLT2) inhibitors; SIRD-II, severe insulin-resistant diabetes with insulin insufficiency, TG/HDL, triglyceride/HDL ratio. <sup>§</sup>Homeostasis model assessment (HOMA) estimates of  $\beta$ -cell function (HOMA2-B) and insulin resistance (HOMA2-IR) were calculated based on fasting plasma blood glucose and fasting C-peptide (<https://www.dtu.ox.ac.uk/homacalculator/>, version 2.2.3).

subgroup assignments between the two approaches (using HOMA2 indices versus substituting HOMA2-IR with TG/HDL) showed moderate agreement, with a Rand index of 0.77 (Rand index after adjustment for chance was 0.46) (Supplementary Table 5). Using HOMA2-based clustering as the reference standard, the classification performance varied across subgroups: SIDD showed the highest sensitivity (0.93) and good specificity (0.79), indicating effective identification of the insulin-deficient subgroup by TG/HDL clustering. The SIRD-II subgroup also demonstrated good performance with moderate sensitivity (0.78) and high specificity (0.94). However, MOD showed notably lower sensitivity (0.68) despite good specificity (0.94), suggesting that approximately one-third of obesity-predominant cases were not captured by TG/HDL-based approach.

### Risk of Adverse Renal Outcomes

In the discovery cohort, participants were followed for a median of 2.8 years (IQR 2.0–4.3) for progressive CKD outcomes and 1.4 years (IQR 0.4–2.7) for albuminuria outcomes. Incidence rates for progressive CKD and albuminuria varied across subgroups (Supplementary Table 6). For progressive CKD, the SIRD-II subgroup showed the highest incidence rate (36.7 per 1000 person-years; 95%CI 14.8, 75.6). Progressive albuminuria was more common, with particularly high rates in SIDD (118.9 per 1000 person-years, 95% CI 57.0, 218.7) and SIRD-II (107.0 per 1000 person-years, 95% CI 57.0, 183.0) subgroups, while the MOD subgroup showed relatively lower risk (42.4 per 1000 person-years, 95% CI 17.0, 87.0).

MOD was used as the reference group for subsequent renal outcome risk analyses due to its largest representation



**Figure 1**—Cumulative incidence and cox regression analysis of adverse renal outcomes by subgroups: (A) progressive chronic kidney disease (CKD), (B) progressive albuminuria. CKD progression defined as >40% eGFR decline from baseline with at least 1 year follow-up data and baseline eGFR >60ml/min/1.73m<sup>2</sup>. Progressive albuminuria defined as worsening of albuminuria category based on Kidney Disease: Improving Global Outcomes (KDIGO) (A1 to A2, A2 to A3 or A1 to A3). DKD, diabetic kidney disease; eGFR, estimated glomerular filtration rate; MOD, mild obesity-related diabetes; SIDD, severe insulin-deficient diabetes; SIRD-II, severe insulin-resistant diabetes with insulin insufficiency; uACR, urinary albumin-to-creatinine ratio.

**Table 2—SCORE2-Diabetes cardiovascular risk by diabetes subgroups**

Risk	Subgroup	n	Median (%)	IQR (%)	P value*
Low Risk	MOD	183	2.20 <sup>a</sup>	1.15–4.34	<0.001
	SIDD	51	1.48 <sup>a</sup>	0.77–3.55	
	SIRD-II	109	5.14 <sup>b</sup>	3.29–8.34	

\*Kruskal-Wallis test for overall group differences. Post hoc Wilcoxon rank-sum tests with Bonferroni correction: <sup>a</sup>Both MOD and SIDD are not significantly different from each other ( $P = 0.096$ ). <sup>b</sup>SIRD-II is significantly higher than both MOD and SIDD ( $P < 0.001$ ).

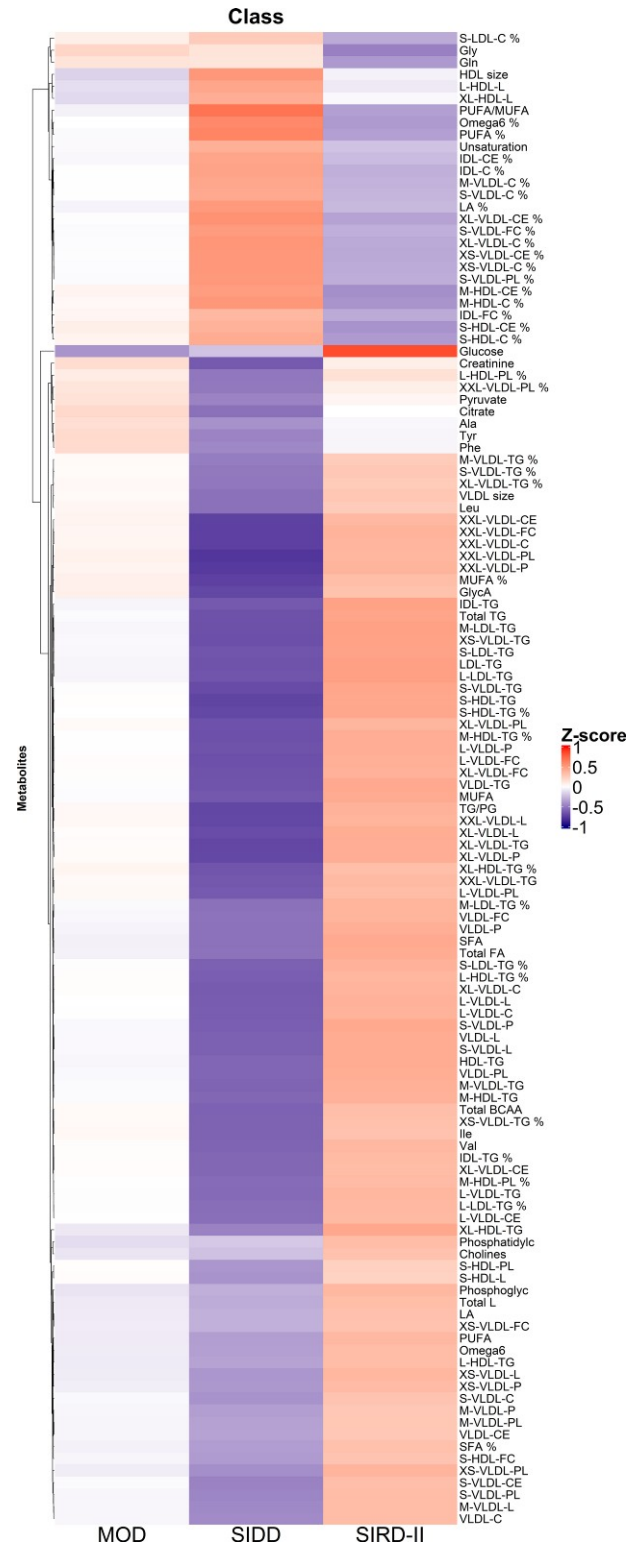
within the study population. Kaplan-Meier curves illustrated cumulative incidences across subgroups (Fig. 1). Cox proportional hazards analysis revealed that the SIRD-II subgroup demonstrated a 3.3-fold elevated risk for progressive CKD relative to MOD (unadjusted HR 3.33; 95% CI 0.86, 12.9). The association was substantially strengthened following adjustment for index age, sex, ethnicity, and baseline eGFR (adjusted HR 11.1; 95%CI 1.86, 66.2,  $P = 0.008$ ). For progressive albuminuria, both SIRD-II and SIDD subgroups showed significantly elevated risks. The SIRD-II subgroup exhibited a 2.5-fold elevated risk relative to MOD (unadjusted HR 2.54; 95% CI 1.01, 6.38,  $P = 0.048$ ), which increased following adjustment for index age, sex, ethnicity, and baseline DKD (adjusted HR 3.28; 95% CI 1.21–8.88,  $P = 0.019$ ). Similarly, the SIDD subgroup showed a 3.3-fold higher risk (unadjusted HR 3.26; 95% CI 1.22, 8.74,  $P = 0.019$ ) which persisted after adjustment (adjusted HR 3.79, 95% CI 1.31, 11.0,  $P = 0.014$ ).

**Risk of Cardiovascular Disease**

SCORE2-Diabetes cardiovascular risk assessment revealed significant differences between diabetes subgroups ( $P < 0.001$ ). SIRD-II had the highest median 10-year CVD risk at 5.14% (IQR 3.29–8.34%), followed by MOD at 2.20% (IQR 1.15–4.34%) and SIDD at 1.48% (IQR 0.77–3.55%) (Table 2). Post hoc analysis confirmed significant differences between SIRD-II and the other two subgroups ( $P < 0.001$  for MOD and SIDD), while SIDD and MOD showed no significant difference after correction for multiple comparisons ( $P = 0.096$ ). SIRD-II demonstrated a 3.5-fold and 2.3-fold higher cardiovascular risk compared with SIDD and MOD respectively. With a median risk of 5.14%, participants in SIRD-II predominantly fell within the high-risk category, while participants in SIDD and MOD subgroups were primarily in the low-to-moderate risk category.

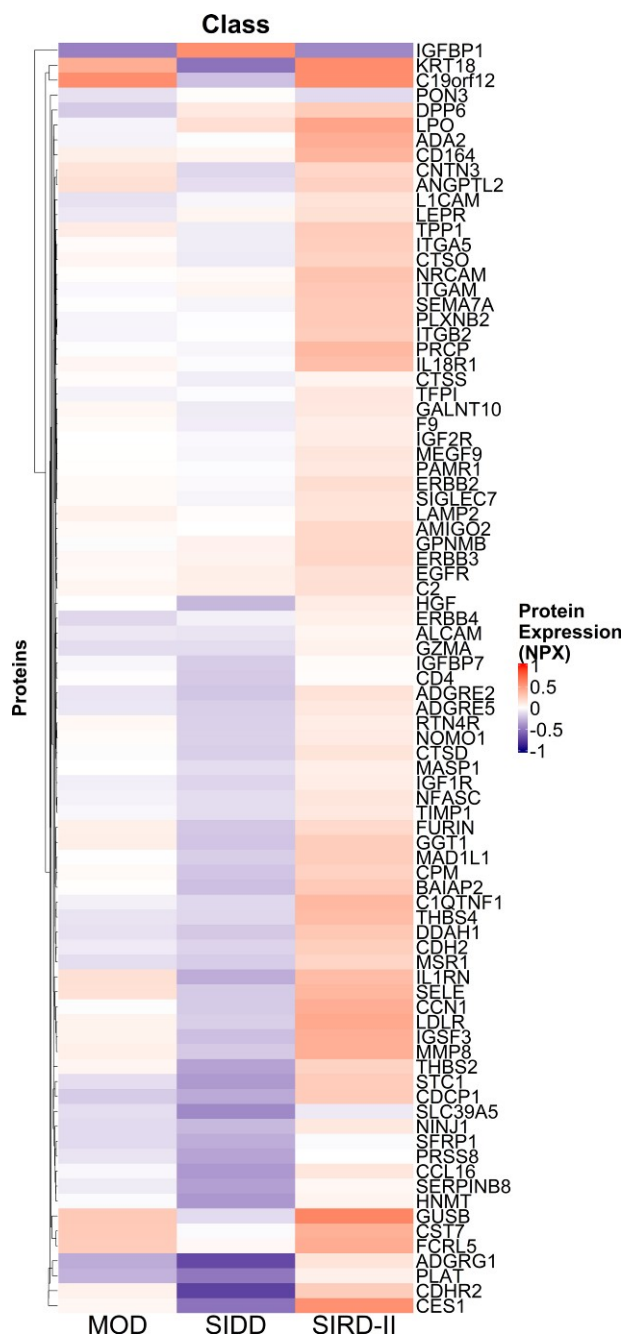
**Metabolomics**

Metabolomic profiling of 249 plasma metabolites identified 130 metabolites that were differentially expressed across diabetes subgroups in the discovery cohort  $P < 2.01 \times 10^{-4}$ , Bonferroni-corrected) (Fig. 2). SIDD, characterized by insulin



**Figure 2—Differentially expressed plasma metabolites across diabetes subgroups in the discovery cohort. MOD, mild obesity-related diabetes; SIDD, severe insulin-deficient diabetes; SIRD-II, severe insulin-resistant diabetes with insulin insufficiency.**

deficiency, demonstrated lower levels of several metabolite classes compared with MOD and SIRD: lipid species (chylomicrons, extremely large VLDL, triglycerides), fatty acids



**Figure 3**—Differentially expressed plasma proteins across diabetes subgroups in the validation cohort. 85 plasma proteins were differentially expressed across the three subgroups with Bonferroni-corrected level of significance ( $p < 3.45 \times 10^{-5}$ ). MOD, mild obesity-related diabetes; SIDD, severe insulin-deficient diabetes; SIRD-II, severe insulin-resistant diabetes with insulin insufficiency.

(including saturated, monounsaturated and polyunsaturated fatty acids), branched-chain amino acids (BCAAs) (leucine, isoleucine and valine), aromatic amino acids (phenylalanine and tyrosine), and other metabolites including glycoprotein acetyls (GlycA) and pyruvate. In contrast, SIRD-II exhibited a markedly different metabolic signature characterized by elevated glucose, lipids (triglycerides, cholesterol, MUFA, PUFA), BCAAs, phosphatidylcholines (PCs)

and GlycA (a relatively novel biomarker for systemic inflammation). Conversely, amino acids including glycine and glutamine were significantly lower in SIRD-II in comparison with MOD and SIDD, highlighting the distinct metabolic dysregulation in this subgroup.

Among the 130 metabolites that were differentially expressed in the discovery cohort, 28 were also differentially expressed in the validation cohort ( $P < 2.01 \times 10^{-4}$ ) (Supplementary Fig. 6). The metabolomic patterns were consistent across both cohorts, with subgroup-specific metabolite levels showing similar direction. Mean metabolite concentrations, standard deviations, and  $P$  values for statistical comparisons across subgroups in both discovery and validation cohorts are provided in Supplementary Tables 7 and 8.

### Proteomics

Among 1,448 plasma proteins analyzed, 85 proteins were differentially expressed ( $P < 3.45 \times 10^{-5}$ , Bonferroni-corrected) across diabetes subgroups identified in the validation cohort, including 16 proteins in the inflammation panel (Fig. 3, Supplementary Table 9). Notably, all 16 inflammation-associated proteins were elevated in SIRD-II except for PON3. These included inflammatory markers such as IL18R1, HGF, and ADGRE2. Importantly, three proteins (IL18R1, HGF and CDCP1) overlapped with inflammatory markers previously identified by Herder et al. in recent-onset diabetes (25), demonstrating consistent elevation in the severe insulin-resistant subgroup across independent cohorts despite different age demographics and panel compositions.

Detailed pairwise comparisons revealed distinct proteomic signatures for each subgroup. Between SIDD and MOD, 3 proteins (IGFBP1, RTN4R and CDHR2) were differentially expressed ( $P < 1.15 \times 10^{-5}$ , corrected for multiple testing) (Supplementary Fig. 7, Supplementary Table 10). Insulin-like growth factor binding protein 1 (IGFBP1) showed significantly higher levels in SIDD ( $\log_2FC = 1.05$ ) and was moderately correlated with both HOMA2-B ( $\rho = -0.43$ ,  $P < 0.001$ ) and HOMA2-IR ( $\rho = -0.46$ ,  $P < 0.001$ ) (Supplementary Fig. 8). RTN4R (receptor for reticulon 4), was significantly lower in SIDD ( $\log_2FC = -0.21$ ). CDHR2, a member of the protocadherin family, showed significantly higher levels in MOD ( $\log_2FC = 0.81$ ) and correlated with HOMA2-IR ( $\rho = 0.43$ ,  $P < 0.001$ ).

Comparison between SIRD-II and MOD identified 26 differentially expressed proteins, all showing elevated levels in SIRD-II with  $\log_2FC$  ranging from 0.12 to 0.51. These included several inflammatory markers such as IL18R1, ERBB3, ADGRE2 ( $P < 1.15 \times 10^{-5}$ ) (Supplementary Fig. 7, Supplementary Table 11). Notably, nine of the top 10 proteins demonstrated moderate correlation with HbA<sub>1c</sub> ( $\rho > 0.40$ ,  $P < 0.001$ ), including PLXNB2, ERBB4, and DPP6. PLXNB2, significantly correlated with HOMA2-B ( $\rho = -0.32$ ,  $P < 0.001$ ) and HOMA2-IR ( $\rho = 0.20$ ,  $P < 0.001$ ), emerged

as the most significantly elevated protein in SIRD-II ( $\log_2\text{FC} = 0.32$ ,  $P = 2.13 \times 10^{-15}$ ).

The comparison between SIRD-II and SIDD yielded 68 differentially expressed proteins ( $P < 1.15 \times 10^{-5}$ ) (Supplementary Fig. 7, Supplementary Table 12). Of these, 67 showed higher levels in SIRD-II with  $\log_2\text{FC}$  ranging from 0.12 to 1.10 while PON3, an enzyme that associates with HDL, was negatively correlated with BMI ( $\rho = -0.47$ ,  $P < 0.001$ ) and showed lower abundance ( $\log_2\text{FC} = -0.14$ ). The top 5 differentially expressed proteins include PLXNB2, IGSF3, PRCP, CPM and SEMA7A, which were correlated with either insulin resistance or HbA<sub>1c</sub> ( $\rho > 0.40$ ,  $P < 0.001$ ) (Supplementary Fig. 8).

## DISCUSSION

This study provides the first comprehensive characterization of younger-onset T2D in a Southeast Asian population, revealing three reproducible subgroups with differential pathobiology, molecular signatures and risk of clinical outcomes. Firstly, our study demonstrated that younger-onset T2D comprises at least three reproducible subgroups. MOD represents the largest subgroup, reflecting obesity-predominant insulin resistance with relatively preserved glycaemic control. SIDD is characterized by profound insulin deficiency without autoimmunity, while SIRD-II displays the most adverse metabolic phenotype, combining severe insulin resistance and relative insulin insufficiency. Importantly, both SIRD-II and SIDD had threefold higher risks of progressive albuminuria, and SIRD-II showed an 11-fold increased risk of progressive CKD compared with MOD. Additionally, SIRD-II demonstrated 3.5-fold and 2.3-fold higher 10-year cardiovascular risk compared with SIDD and MOD respectively. These findings highlight the importance of early aggressive risk factor management in high-risk subgroups and support the clinical relevance of diabetes subtyping for risk stratification.

Our findings in younger-onset T2D show both notable similarities and important differences compared with the landmark Ahlqvist et al. study in older-onset diabetes (14). We identified three subgroups that correspond to three of their five clusters: MOD, SIDD, and a variant of their SIRD. Notably, the severe autoimmune diabetes (SAID) and mild age-related diabetes (MARD) subgroups were not identified in our cohort. The absence of SAID reflects our methodological approach: unlike Ahlqvist et al., we did not include GAD autoantibodies as a clustering variable. Consequently, any participants with autoimmune diabetes would be classified based solely on their metabolic phenotype. Our rigorous exclusion of clinically suspected autoimmune diabetes cases further reduced the likelihood of identifying this subgroup. The absence of MARD likely reflects the younger age profile of our cohort, as this subgroup typically emerges in older-onset diabetes cohorts. Importantly, our SIRD-II subgroup demonstrates a more severe metabolic phenotype than the original SIRD, characterized by both concurrent insulin resistance and

relative insulin insufficiency. This dual pathophysiology may be particularly relevant in younger Asian populations, where limited  $\beta$ -cell reserve predisposes to earlier  $\beta$ -cell decompensation under metabolic stress (8,26). Consistent with Ahlqvist et al., SIRD subgroups in both studies demonstrated the highest CKD risk, supporting the robustness of this association across different age populations.

Secondly, our exploratory analysis of TG/HDL ratio as an alternative clustering variable demonstrated moderate overall agreement with HOMA2-based clustering (Rand index 0.77) and showed good performance for SIDD (sensitivity 0.93) and SIRD-II (sensitivity 0.78). However, the reduced sensitivity for identifying MOD (sensitivity 0.68) highlights important limitations of this approach, where approximately one-third of MOD cases were not adequately captured by lipid-based clustering. While TG/HDL ratio effectively identifies distinct metabolic phenotypes, the cross-classification between subgroups underscores that it cannot fully substitute for direct  $\beta$ -cell function assessment, which remains the key distinguishing feature between insulin-deficient and insulin-resistant phenotypes. This finding reinforces that comprehensive diabetes subtyping requires both insulin resistance and  $\beta$ -cell function measures, and suggests that lipid-based surrogates, while potentially useful in resource-limited settings, require further validation and refinement before broader clinical application.

Thirdly, our metabolomic findings reveal distinct patterns of metabolic dysregulation across subgroups. In SIDD, low levels of triglyceride-rich lipoproteins, free fatty acids, BCAAs, inflammatory markers, and glycolytic intermediates may suggest a healthier metabolic phenotype (27,28), but more likely reflect insulinopenia (29). MOD showed moderate elevations in lipid species, consistent with mild dysmetabolic burden. SIRD-II exhibited the most adverse metabolic profile, characterized by marked hyperglycemia, dyslipidaemia, elevated BCAAs and increased inflammatory burden (GlycA). This constellation suggests concomitant glucotoxicity and lipotoxicity, creating a synergistic “glucolipotoxic” milieu that may accelerate  $\beta$ -cell decompensation through maladaptive interorgan crosstalk (30,31). The pronounced depletion of glycine and glutamine in SIRD-II provides additional mechanistic insight, as both amino acids possess anti-inflammatory properties and support insulin sensitivity (32–34). Their deficiency in SIRD-II may therefore contribute to the observed metabolic dysfunction and insulin resistance in this subgroup. Importantly, the consistency of these metabolomic signatures across discovery and validation cohorts validates the robustness of our subgroup classifications and their underlying biological basis.

Proteomic analysis further elucidated molecular heterogeneity underlying these diabetes subgroups. The differential expression of IGFBP1, RTN4R, and CDHR2 between SIDD and MOD reflects distinct pathophysiological processes, with IGFBP1 and RTN4R previously implicated in

insulin sensitivity (35) and glucose-stimulated insulin secretion (36), respectively. CDHR2, primarily known for its role in calcium-dependent cell-cell adhesion (37,38), emerged as a novel protein associated with insulin resistance, warranting further mechanistic investigation. Between SIRD-II and MOD, the three most differentially expressed proteins, PLXNB2, ERBB4 and DPP6, provide insights into the severe metabolic dysfunction characterizing SIRD-II. ERBB4 encodes a receptor tyrosine kinase of the ErbB family, previously associated with neurodevelopment and cellular growth pathways (39). More recently, studies in mouse models have demonstrated that ErbB4 is essential for maintaining adipocyte function, promoting glucose uptake, and protecting against diet-induced obesity and insulin resistance (40). In view of these findings, the elevated levels of ERBB4 observed in SIRD-II may reflect a compensatory response to adipose tissue dysfunction or perturbed ErbB4 signaling in the setting of insulin resistance. The elevated DPP6 (dipeptidyl peptidase 6) in SIRD-II presents a particularly intriguing finding, as this recently identified pancreatic endocrine cell biomarker (41) was increased despite reduced  $\beta$ -cell function. Given its positive correlation with HbA<sub>1c</sub> and negative correlation with HOMA2-B, DPP6 may serve as a marker of  $\beta$ -cell stress or maladaptive responses to chronic glucotoxicity rather than functional  $\beta$ -cell mass, providing molecular evidence for progressive  $\beta$ -cell decompensation in this subgroup. Most notably, Plexin-B2 (PLXNB2) emerged as the most significantly altered protein, with marked elevation in SIRD-II compared with both MOD and SIDD. Although primarily known for neuronal guidance functions, PLXNB2 is expressed in pancreatic islets and has been linked to T2D through Mendelian randomization and proteomic studies (42–45), suggesting previously unrecognized roles in metabolic regulation that merit further investigation. The elevated inflammatory proteins observed in SIRD-II, including IL18R1, HGF, and CDCP1, align remarkably well with findings by Herder et al. (25), who investigated 74 inflammatory biomarkers across diabetes subgroups in recent-onset diabetes. Despite differences in cohort characteristics (younger-onset vs recent-onset diabetes) and panel scope (368 vs 92 markers), both studies consistently identified the highest inflammatory burden in severe insulin-resistant subgroups. The replication of three specific inflammatory proteins (IL18R1, HGF, CDCP1) with identical directional changes in SIRD-II across independent cohorts strengthens the evidence for a distinct inflammatory signature in severe insulin-resistant diabetes. Our broader proteomic analysis identified 14 additional inflammatory proteins (ADGRE2, NFASC, PRSS8, IL1RN, TPP1, ANGPTL2, ERBB3, SERPINB8, CD4, SLC39A5, PON3, CST7, CTSO, GZMA) that were differentially expressed among subgroups, expanding the known inflammatory landscape of diabetes subgroups beyond previous targeted inflammatory panels. Of particular interest is PON3, which uniquely showed the opposite pattern – being lowest in SIRD-II rather than highest. As an HDL-associated antioxidant enzyme (46), reduced PON3 levels in SIRD-II may indicate depletion of

protective antioxidant systems, potentially explaining the heightened vulnerability to oxidative stress and cardiovascular complications in this subgroup. Collectively, these molecular signatures not only validate our clustering approach but also provide novel insights into subgroup-specific pathophysiology. The associations identified in this study lay the groundwork for future mechanistic studies and may highlight potentially actionable therapeutic targets for subgroup-specific personalised intervention strategies.

Our study has several strengths. Firstly, participants in the discovery cohort underwent deep phenotyping including gene sequencing to exclude monogenic diabetes and comprehensive islet autoantibody testing to identify autoimmune-related diabetes. Secondly, this study represents the first application of cluster analysis to younger-onset T2D in a Southeast Asian population, addressing an important knowledge gap in diabetes stratification for this under-represented patient group. While our sample size was limited, the consistency of our findings was demonstrated by successful replication of the diabetes subgroups in an independent validation cohort. Nonetheless, the interpretation of our findings should consider several important limitations. This study was based on secondary use of existing cohort data, primarily from participants attending secondary care institutions, including those with longstanding diabetes (duration >5 years) who were already on glucose-lowering treatment, potentially influencing the subgroup distributions. Furthermore, some individuals might shift between subgroups over time, although previous studies have shown that cluster allocations generally remain stable during long-term follow-up (47,48). Nonetheless, the primarily secondary care-based sampling frame may limit the generalizability of our findings to the broader younger-onset T2D population, particularly those managed in primary care. Validation of the diabetes subgroups identified in this study will be required across diverse Asian populations to determine whether they represent consistent phenotypes across different ethnic backgrounds, environmental exposures, and healthcare systems across the region.

In conclusion, this study addresses the heterogeneity of T2D within the younger-onset population by identifying distinct subgroups with differential clinical profiles, molecular signatures and clinical outcomes including renal progression and cardiovascular risk. The characterization of SIDD and SIRD-II as high-risk phenotypes, each with unique molecular signatures and elevated renal risks, provides novel and strategic insights for risk stratification of younger-onset T2D in Asian populations. These findings challenge the “one-size fits all” treatment approach for younger-onset T2D and support subgroup-specific management strategies to optimize clinical outcomes.

---

**Acknowledgments.** We extend our appreciation to all research participants and our colleagues at the Clinical Research Unit, Khoo Teck Puat Hospital, for their valuable support and contributions to this study. Graphical abstract was created in BioRender, <https://BioRender.com/m04ynqw>.

**Funding.** This study is supported by the Khoo Teck Puat Health Fund through the Science – Translational & Applied Research (STAR) grants (STAR19204, STAR20106, STAR21203, STAR24203). The corresponding author receives research support from Singapore National Medical Research Council (NMRC) MOH-001688-00, MOH-001704-00 and NMRC/CIRG/1398/2014.

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

**Author Contributions.** C.S.H.T. contributed to the conceptualization, methodology, formal analysis, interpretation of data, and writing of the manuscript. K.X.K. and H.Z. contributed to methodology, formal analysis and interpretation of the data. K.W.S.W. participated in the investigation and editing of the manuscript. W.T.L.C., Y.S., and K.A. acquired the data. T.S., C.F.S., and S.C.L. participated in reviewing and editing the manuscript. S.C.L. provided supervision and is the guarantor of this work. All authors reviewed the manuscript and approved it for publication. S.C.L. is the guarantor of this work, had full access to all the data, and takes full responsibility for the integrity of data and the accuracy of data analysis.

**Prior Presentation.** Parts of this study were presented in a poster for the Singapore Health Biomedical Congress, Singapore on 23 September 2025.

## References

- Misra S, Ke C, Srinivasan S, et al. Current insights and emerging trends in early-onset type 2 diabetes. *Lancet Diabetes Endocrinol* 2023;11:768–782
- Perng W, Conway R, Mayer-Davis E, Dabelea D. Youth-onset type 2 diabetes: the epidemiology of an awakening epidemic. *Diabetes Care* 2023;46:490–499
- Wagenknecht LE, Lawrence JM, Isom S, et al.; SEARCH for Diabetes in Youth study. Trends in incidence of youth-onset type 1 and type 2 diabetes in the USA, 2002–18: results from the population-based SEARCH for Diabetes in Youth study. *Lancet Diabetes Endocrinol* 2023;11:242–250
- International Diabetes Federation. IDF Diabetes Atlas, 11th edition. International Diabetes Federation, 2025
- Wright AK, Welsh P, Gill JMR, 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* 2020;63:1542–1553
- Twig G, Zucker I, Afek A, et al. Adolescent obesity and early-onset type 2 diabetes. *Diabetes Care* 2020;43:1487–1495
- Lin B, Coleman RL, Bragg F, Maddaloni E, Holman RR, Adler AI. Younger-onset compared with later-onset type 2 diabetes: an analysis of the UK Prospective Diabetes Study (UKPDS) with up to 30 years of follow-up (UKPDS 92). *Lancet Diabetes Endocrinol* 2024;12:904–914
- Fan Y, Fan B, Lau ESH, et al. Comparison of beta-cell function between Hong Kong Chinese with young-onset type 2 diabetes and late-onset type 2 diabetes. *Diabetes Res Clin Pract* 2023;205:110954
- Bjornstad P, Drews KL, Caprio S, et al. Long-term complications in youth-onset type 2 diabetes. *N Engl J Med* 2021;385:416–426
- Sattar N, Rawshani A, Franzén S, et al. Age at diagnosis of type 2 diabetes mellitus and associations with cardiovascular and mortality risks. *Circulation* 2019;139:2228–2237
- Luk A, Wild SH, Jones S, et al. Early-onset type 2 diabetes: the next major diabetes transition. *Lancet* 2025;405:2313–2326
- Misra S, Khunti K, Goyal A, et al. Managing early-onset type 2 diabetes in the individual and at the population level. *Lancet* 2025;405:2341–2354
- Misra S, Wagner R, Ozkan B, et al.; ADA/EASD PMDI. Precision subclassification of type 2 diabetes: a systematic review. *Commun Med (Lond)* 2023;3:138
- Ahlqvist 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
- Slieker RC, Donnelly LA, Fitipaldi H, et al. Distinct molecular signatures of clinical clusters in people with type 2 diabetes: an IMI-RHAPSODY study. *Diabetes* 2021;70:2683–2693
- Zaghlool SB, Halama A, Stephan N, et al. Metabolic and proteomic signatures of type 2 diabetes subtypes in an Arab population. *Nat Commun* 2022;13:7121
- Pigeyre M, Gerstein H, Ahlqvist E, Hess S, Paré G. Identifying blood biomarkers for type 2 diabetes subtyping: a report from the ORIGIN trial. *Diabetologia* 2023;66:1045–1051
- Ang SF, Tan CSH, Chan LWT, et al. Clinical experience from a regional monogenic diabetes referral centre in Singapore. *Diabetes Res Clin Pract* 2020;168:108390
- Pek SLT, Tavintharan S, Wang X, et al. Elevation of a novel angiogenic factor, leucine-rich- $\alpha$ 2-glycoprotein (LRG1), is associated with arterial stiffness, endothelial dysfunction, and peripheral arterial disease in patients with type 2 diabetes. *J Clin Endocrinol Metab* 2015;100:1586–1593
- Moh MC, Low S, Shao Y-M, Subramaniam T, Sum CF, Lim SC. Association between neutrophil/lymphocyte ratio and kidney impairment in type 2 diabetes mellitus: a role of extracellular water/total body water ratio. *Diabetes Res Clin Pract* 2023;199:110634
- Henning C. Cluster-wise assessment of cluster stability. *Computational Statistics & Data Analysis* 2007;52:258–271
- Nelson RG, Grams ME, Ballew SH, et al.; CKD Prognosis Consortium. Development of Risk Prediction Equations for Incident Chronic Kidney Disease. *JAMA* 2019;322:2104–2114
- SCORE2-Diabetes Working Group and the ESC Cardiovascular Risk Collaboration. SCORE2-Diabetes: 10-Year Cardiovascular Risk Estimation in Type 2 Diabetes in Europe. *Eur Heart J* 2023;44:2544–2556
- Visseren FLJ, Mach F, Smulders YM, et al. 2021 ESC guidelines on cardiovascular disease prevention in clinical practice: developed by the task force for cardiovascular disease prevention in clinical practice with representatives of the European Society of Cardiology and 12 medical societies with the special contribution of the European Association of Preventive Cardiology (EAPC). *Eur Heart J* 2021;42:3227–3337
- Herder C, Maalmi H, Strassburger K, et al.; GDS Group. Differences in biomarkers of inflammation between novel subgroups of recent-onset diabetes. *Diabetes* 2021;70:1198–1208
- Siddiqui MK, Anjana RM, Dawed AY, et al. Young-onset diabetes in Asian Indians is associated with lower measured and genetically determined beta cell function. *Diabetologia* 2022;65:973–983
- Ahola-Olli AV, Mustelin L, Kalimeri M, et al. Circulating metabolites and the risk of type 2 diabetes: a prospective study of 11,896 young adults from four Finnish cohorts. *Diabetologia* 2019;62:2298–2309
- Würtz P, Havulinna AS, Soininen P, et al. Metabolite profiling and cardiovascular event risk: a prospective study of 3 population-based cohorts. *Circulation* 2015;131:774–785
- Dutta T, Chai HS, Ward LE, et al. Concordance of changes in metabolic pathways based on plasma metabolomics and skeletal muscle transcriptomics in type 1 diabetes. *Diabetes* 2012;61:1004–1016
- Xourafa G, Korbmacher M, Roden M. Inter-organ crosstalk during development and progression of type 2 diabetes mellitus. *Nat Rev Endocrinol* 2024;20:27–49
- Wang TJ, Larson MG, Vasani RS, et al. Metabolite profiles and the risk of developing diabetes. *Nat Med* 2011;17:448–453
- Yan-Do R, MacDonald PE. Impaired “Glycine”-mia in Type 2 Diabetes and Potential Mechanisms Contributing to Glucose Homeostasis. *Endocrinology* 2017;158:1064–1073
- Dollet L, Kuefner M, Caria E, et al. Glutamine regulates skeletal muscle immunometabolism in type 2 diabetes. *Diabetes* 2022;71:624–636
- Abboud KY, Reis SK, Martelli ME, et al. Oral glutamine supplementation reduces obesity, pro-inflammatory markers, and improves insulin sensitivity in DIO Wistar rats and reduces waist circumference in overweight and obese humans. *Nutrients* 2019;11:536
- Bae J-H, Song D-K, Im S-S. Regulation of IGFBP-1 in Metabolic Diseases. *J Lifestyle Med* 2013;3:73–79

36. Hardy OT, Hohmeier HE, Becker TC, et al. Functional genomics of the beta-cell: short-chain 3-hydroxyacyl-coenzyme A dehydrogenase regulates insulin secretion independent of K<sup>+</sup> currents. *Mol Endocrinol* 2007;21:765–773
37. Okazaki N, Takahashi N, Kojima S-I, Masuho Y, Koga H. Protocadherin LKC, a new candidate for a tumor suppressor of colon and liver cancers, its association with contact inhibition of cell proliferation. *Carcinogenesis* 2002;23:1139–1148
38. Crawley SW, Shifrin DA, Grega-Larson NE, et al. Intestinal brush border assembly driven by protocadherin-based intermicrovillar adhesion. *Cell* 2014;157:433–446
39. Olayioye MA, Neve RM, Lane HA, Hynes NE. The ErbB signaling network: receptor heterodimerization in development and cancer. *Embo J* 2000;19:3159–3167
40. Zeng F, Wang Y, Kloepfer LA, Wang S, Harris RC. ErbB4 deletion predisposes to development of metabolic syndrome in mice. *Am J Physiol Endocrinol Metab* 2018;315:E583–E593
41. Balhuizen A, Massa S, Mathijs I, et al. A nanobody-based tracer targeting DPP6 for non-invasive imaging of human pancreatic endocrine cells. *Sci Rep* 2017;7:15130
42. Zielonka M, Xia J, Friedel RH, Offermanns S, Worzfeld T. A systematic expression analysis implicates Plexin-B2 and its ligand Sema4C in the regulation of the vascular and endocrine system. *Exp Cell Res* 2010;316:2477–2486
43. Gudmundsdottir V, Zaghlool SB, Emilsson V, et al. Circulating protein signatures and causal candidates for type 2 diabetes. *Diabetes* 2020;69:1843–1853
44. Zaghlool SB, Halama A, Stephan N, et al. Metabolic and proteomic signatures of type 2 diabetes subtypes in an Arab population. *Nat Comm* 2022;13:7121
45. Cronjé HT, Mi MY, Austin TR, et al. Plasma proteomic risk markers of incident type 2 diabetes reflect physiologically distinct components of glucose-insulin homeostasis. *Diabetes* 2023;72:666–673
46. Aviram M, Rosenblat M. Paraoxonases 1, 2, and 3, oxidative stress, and macrophage foam cell formation during atherosclerosis development. *Free Radic Biol Med* 2004;37:1304–1316
47. Li X, Donnelly LA, Sliker RC, et al. Trajectories of clinical characteristics, complications and treatment choices in data-driven subgroups of type 2 diabetes. *Diabetologia* 2024;67:1343–1355
48. 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