nature research

https://doi.org/10.1038/s41574-021-00475-4

Supplementary information

Clustering for a better prediction of type 2 diabetes mellitus

In the format provided by the authors and unedited

Supplementary Table 1. Recent studies based on clustering strategies for deciphering the heterogeneity of (pre)diabetes mellitus

Study	Initial clustering 1/ Total number 2/ Phenotypes; 3/ Used variables)	Replication study 1/ Total number 2/ Phenotypes 3/ Used variables	Clusters	Outcomes
Wagner *et al*. (2021)[1]	**1/** *n* = 899 (including 421 participants with longitudinal data) **2/** history of prediabetes, family history of diabetes, BMI ≥ 27 kg/m² or history of gestational diabetes **3/** glycemia during OGTT (5-point OGTT AUC), ISI (Matsuda), insulin secretion ($AUC_{0\text{-}30}$ C-peptide / $AUC_{0\text{-}30}$ Glucose), HDLc, liver fat content, subcutaneous fat volume, visceral fat volume, PRS for T2D (including 484,788 SNPs)	**1/** *n* = 6,810 with longitudinal data **2/** exclusion of participants with prevalent or incident diabetes at baseline and those of a non-white ethnicity **3/** glycemia during OGTT (2-point OGTT AUC), ISI (Matsuda), insulin secretion (Stumvoll's 1st phase-index), FI, fasting TG, WC, HC, BMI, HDLc	**6 clusters** **1:** low risk **2:** very low risk **3:** beta-cell failure **4:** low risk obese **5:** high risk insulin-resistant fatty liver **6:** high risk visceral fat nephropathy	- Clusters **3**, **5** and **6** associated with increased glycemia (fasting glucose, postchallenge glucose, HbA1c), lower DI, higher risk of kidney disease - Clusters **3** and **5** associated with imminent T2D risk - Clusters **5** and **6** associated with all-cause mortality - Cluster **6** associated with lower GRS for beta-cell function (according to Udler et al.) - Cluster **3** associated with higher PRS for T2D - Cluster **3** associated with the T2D-associated SNP in *MTNR1B*
Ahlqvist *et al*. (2018)[2]	**1/** *n* = 8,990 with longitudinal data **2/** individuals with newly diagnosed diabetes; exclusion of patients with secondary diabetes and extreme outliers **3/** GADA, age at diagnosis, BMI, HbA1c, HOMA2-B and HOMA2-IR based on C-peptide concentrations.	**1/** 3 cohorts: *n* = 1,466; *n* = 844; and *n* = 3,485 with longitudinal data **2/** individuals with diabetes; exclusion of patients with secondary diabetes and extreme outliers **3/** GADA, age at diagnosis, BMI, HbA1c, HOMA2-B and HOMA2-IR based on C-peptide concentrations.	**5 clusters** **1:** severe autoimmune diabetes **2:** severe insulin-deficient diabetes **3:** severe insulin-resistant diabetes **4:** mild obesity-related diabetes **5:** mild age-related diabetes	- Clusters **1** and **2** associated with higher HbA1c, more frequent ketoacidosis, shortest time to sustained insulin use - Cluster **3** associated with highest risk of NAFLD, CKD and diabetic kidney disease - Cluster **2** associated with highest risk of retinopathy - Clusters **1**, **2**, **4** and **5** associated with GRS for T2D (including 64 SNPs) - Clusters **2**, **4** and **5** associated with GRS for insulin secretion (including 15 SNPs)
Udler *et al*. (2018)[3]	**1/** and **2/** based on GWAS summary statistics (from MAGIC, GIANT, EGG, VATGen, ICBP, ENGAGE, CHARGE consortia) [mostly general populations] **3/** 94 independent T2D-associated SNPs and 47 diabetes-related traits	Application in: 4 cohorts including patients with T2D (*n* = 487, *n* = 509, *n* = 2,065 and *n* = 14,813)	**5 clusters** **1:** beta-cell **2:** proinsulin **3:** obesity **4:** lipodystrophy **5:** liver/lipid	- Cluster **1** associated with decreased CIR, DI, Ins30 adj BMI, HOMA-B, and increased PI adj for FI; T2D-associated SNPs in *MTNR1B*, *HHEX*, *TCF7L2*, *SLC30A8*, *HNF1A*, and *HNF1B*; increased risk for ischemic stroke; in T2D: decreased BMI, % body fat, fasting C-peptide - Cluster **2** associated with decreased Ins30, HOMA-B, PI adj for FI; T2D-associated SNPs in *ARAP1/STARD10* and *SPRY2*; increased blood pressure; in T2D: decreased BMI, fasting C-peptide - Cluster **3** associated with increased WC, HC, BMI, and % body fat; T2D-associated SNPs in *FTO* and *MC4R*; in T2D, increased BMI, % body fat, HC, and WC - Cluster **4** associated with decreased ISI adj BMI, adiponectin, HDLc, and increased TG; increased UACR; in T2D: decreased BMI, % body fat, HDLc - Cluster **5** associated with decreased TG, palmitoleic acid, urate, and linolenic acid, and increased gamma-linolenic acid; reduced eGFR; in T2D: decreased TG - Individuals with T2D and a GRS uniquely at the top 10th of one cluster as a group had representative trait characteristics distinguishing them from all other individuals with T2D

AUC, area under the curve; ***BMI***, body mass index; ***CIR***, corrected insulin response; ***CKD***, chronic kidney disease; ***DI***, disposition index; ***eGFR***, estimated glomerular filtration rate; ***FI***, fasting insulin; ***GADA***, glutamic acid decarboxylase antibodies; ***GRS***, genetic risk score; ***GWAS***, genome-wide association study; ***HbA1c***, glycated hemoglobin A1c; ***HC***, hip circumference; ***HDLc***, high-density lipoprotein cholesterol; ***HOMA-B***, homoeostasis model assessment estimate of beta-cell function; ***HOMA2-B***, homoeostasis model assessment 2 estimate of beta-cell function; ***HOMA2-IR***, homoeostasis model assessment 2 estimate of insulin resistance; ***Ins30 adj BMI***, insulin at 30 minutes of OGTT adjusted for BMI; ***ISI adj BMI***, insulin sensitivity index adjusted form BMI; ***NAFLD***, non-alcoholic fatty liver disease; ***OGTT***, oral glucose tolerance test; ***PI adj FI***, proinsulin levels adjusted for fasting insulin; ***PRS***, polygenic risk score; ***SNP***, single nucleotide polymorphism; ***T2D***, type 2 diabetes; ***TG***, triglycerides; ***UACR***, urine albumin-creatinine ratio; ***WC***, waist circumference.

References

1. Wagner, R. *et al.* Pathophysiology-based subphenotyping of individuals at elevated risk for type 2 diabetes. *Nat Med* (2021) doi:10.1038/s41591-020-1116-9.
2. Ahlqvist, E. *et al.* Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables. *Lancet Diabetes Endocrinol* **6**, 361–369 (2018).
3. Udler, M. S. *et al.* Type 2 diabetes genetic loci informed by multi-trait associations point to disease mechanisms and subtypes: A soft clustering analysis. *PLOS Medicine* **15**, e1002654 (2018).