## DETAILED COHORT METHODS

### SIGMA-UIEM study

Laboratory samples were taken, and body composition was assessed using a SECA mBCA-515 medical body composition analyzer calibrated for the Hispanic population. Genotyping was also carried out to assess carrier status for the rs1334232 risk variant in SLC16A11 in all subjects. In a subset of subjects, the presence of diabetic kidney disease (DKD) was assessed using the albumin-to-creatinine ratio (ACR), diabetic neuropathy (DN) was assessed using the modified Michigan questionnaire with physical examination (n=1,123) and diabetic retinopathy was evaluated using a standardized ophthalmological examination (DR, n=353). To assess non-alcoholic fatty liver disease (NAFLD) the fatty-liver index (FLI) was used (14). In the overall UIEM-SIGMA cohort, SNNN-Model 2 was used for diabetes subgroup classification and the SNNN-Model 4 for subjects treated with insulin.

A subsample of study participants from the SIGMA-UIEM cohort (n=67) underwent additional phenotyping (15, Supplementary Material). Subjects had been diagnosed with T2D for<5 years, had HbA1c<8%, and were treated only with metformin. Insulin sensitivity was assessed using the M-values from euglycemic hyperinsulinaemic clamps (EHC), adjusted for body weight or insulin levels (M-I). To evaluate the acute insulin response to glucose (AIRg), a frequently-sampled intravenous glucose tolerance test (FSIVGTT) was performed. Subcutaneous and visceral adipose tissue areas (SFA, VFA) were quantified using magnetic resonance imaging (MRI); intra-pancreatic and intrahepatic triglyceride (IHTG) content was determined using MRI spectroscopy.

### Metabolic Syndrome cohort

We performed a prospective observational cohort study including Mexican adults living in large urban settings of central Mexico including Mexico City, Cuernavaca, Leon, Toluca and Aguascalientes to evaluate incidence of T2D, arterial hypertension and cardiovascular diseas. Study sample comprised apparently-healthy adults ≥20 years, who resided for >6 months in the evaluated city, and without plans to move to other city in the short term, whose grandparents and parents were born in Mexico. We excluded individuals with previously diagnosed diabetes, cardiovascular disease, cerebral vascular disease, pregnancy, alcoholism (≥10 servings of alcohol per week), acute stress event or any condition that could potentially endanger her life in the three following years. Participants (n=7, were identified and evaluated at their workplaces (offices of the federal government or private companies) (n=3246), homes (n=189) or during a visit of a relative to a medical unit (n=2709).

All assessments were performed after a 9-12hr fasting period. The evaluation consisted in a clinical examination using standardized questionnaires, anthropometric measurements and a blood draw. Demographic information and a medical history, including personal and family history of the most common chronic diseases, were obtained. The evaluation included a 24-hour diet recall, 7-day food frequency questionnaire, the three-factor eating questionnaire (16), the short version of the International physical activity questionnaire (IPAQ) (17) and for adults ≥50 years, an assessment of their functionality and depression. Participants were informed about their results and were advised to visit a primary care physician to seek for treatment if required. They were contacted after a three-year period (±6 months) and invited to repeat the evaluation using the same tools and methods. Multiple approaches were applied to cases that were not reachable at the place in which they were originally invited to participate, including phone calls, e-mail messages, telegrams, invitations through friends or relatives, and visits to the workplace. The response rate was 80.7% (n=6,166). Clinical data and blood samples were collected from 10,052 participants at baseline from 2007-2011. We excluded 2,416 individuals who had either undiagnosed T2D (n=429) or declined permission to be included in the follow-up (n=1987). Consequently, our study sample considered for the primary end-point was 7,636 participants. The follow-up visit was performed 29.5±9.7 months later (2010-2013); 6,166 patients were reached for the second evaluation. Twenty-two deaths were recorded among participants. The study was approved by the Ethics Committee of the Instituto Nacional de Ciencias Médicas y Nutrición and all participants signed an informed consent form.

All serum samples were kept frozen until processed in a central laboratory certified by the External Comparative Evaluation of Laboratories Program of the College of American Pathologists. Clinical chemistry parameters and the lipid profile were measured using commercially available reagents (Synchron CX5 delta, Beckman Coulter). Immunonephelometric methods were applied for the measurement of apolipoprotein B (IMMAGE, Beckman Coulter) and C reactive protein (BN ProSpec, Siemens). Insulin concentrations were measured using an ELISA method (AxSYM, Abbott).

Incident diabetes (ID) was defined if a previously healthy subject (fasting plasma glucose (FPG) <126mg/dL) at baseline had a medical diagnosis of T2D or started treatment with a glucose-lowering drug after follow-up and/or had a fasting glycemia ≥126mg/dL in the second visit. Incident impaired fasting glucose (IFG) was defined by FPG in the range 100-125mg/dL in the final visit for individuals that had the same variable <100 mg/dL at baseline. Hypercholesterolemia was defined by the presence of a total cholesterol concentration >200mg/dL or being under statin therapy. Metabolic syndrome and its components were defined according to IDF and ATP-III recommendations.

**SIGMA-UIEM subcohort**

The UIEM-SIGMA subcohort is an open-population cohort aimed at to characterize carriers and non-carriers of SLC16A11 variants associated with increased risk for type 2 diabetes (T2D) in Mexicans. The UIEM-SIGMA cohort includes individual with and without T2D (n=4,890). We included subjects with T2D and complete information for clustering (n=1521). In a subset of patients with T2D from this cohort, we assessed diabetic kidney disease (DKD) using albumin-to-creatinine ratio (ACR) and diabetic neuropathy (DN) using the modified Michigan questionnaire with physical examination (n=1,123) and diabetic retinopathy using a standardized ophthalmological examination (DR, n=353). Ophtalmologists evaluated vision acuity, ruled-out diabetic retinopathy and macular edema using a no-mydriatic camera for retinal review. Pupillary pharmacological dilation was performed when photographs had poor quality

For the deep-phenotyping subcohort we included Mexican-mestizo subjects with parents and grandparents born in Mexico, aged 20-79 years, with BMI 18-34.9 kg/m$^2$. Individuals with T2D, HbA1c concentration <8% and without insulin treatment were eligible for the study. No subject smoked tobacco had cardiovascular disease, diabetes complications, or an acute infection. Subjects with >3% weight loss in the last three months, taking medications or with conditions that could interfere with insulin secretion and action, high-performance athletes, with alcohol consumption more than 2 units per day in men or 1 unit in women were also excluded. Subjects provided written informed consent before participating in this study, which was approved by the Comité de Ética en Investigación of the Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán (INCMNSZ).

Intervention studies were conducted and completed within the course of a one-month period. First, to identify carriers of the risk haplotype of the SCL16A11 variant samples were genotyped using a Quant Studio 12K Flex Real-Time PCR platform from Thermo Fisher Scientific.

**Body composition analysis:** Body fat mass (FM) and fat-free mass (FFM) were determined using dual-energy X-ray absorptiometry (DXA, GE Healthcare) and were standardize using height squared in meters. Subcutaneous and visceral adipose tissue areas (SAA, VAA) were quantified using magnetic resonance imaging, and the subcutaneous/intra-abdominal fat ratio was calculated. Intra-pancreatic and intra-hepatic triglycerides content was determined using MRI spectroscopy (Philips Achieva 3 Teslas). In the overall UIEM-SIGMA cohort, we performed body composition analyses using a SECA mBCA-515 medical body composition analyzer calibrated for Hispanic population.

**Insulin sensitivity:** Participants were instructed to fast for at least 12 hours before the study and subjects with T2D were instructed to suspend oral treatment three days before the procedure. The study was not performed if FPG >250 mg/dL. A catheter was inserted into a forearm vein to infuse dextrose and insulin and a second catheter into a forearm vein in the contralateral hand was inserted in a retrograde fashion to obtain arterialized blood samples using a hot box. Insulin was infused at a rate of 50 mU/m$^2$ body surface area (BSA)/min

(initiated with a priming dose of 200 mU/m$^2$/min for 5 min and then 100 mU/m$^2$/min for 5 min). Euglycemia (~100 mg/dL) was maintained by a variable infusion of 20% dextrose. During the clamp procedure, blood samples were drawn every 10 minutes during the final 30 minutes to determine glucose and insulin concentrations. Insulin sensitivity was determined as the glucose infusion rate (M value) during the final 30 minutes adjusted for weight and for steady state insulin levels (M-I index).

**Insulin secretory response**: Participants were instructed to fast for 12 hours before the procedure. Two intravenous catheters were placed in antecubital veins (one in each arm). Blood samples were withdrawn at −10, −5, 0, 2, 3, 4, 5, 6, 8, 12, 14, 16, 19, 22, 24 and 25 minutes for measurement of serum glucose and insulin. Glucose was administered intravenously at a dose of 0.3 g/kg for 60 seconds beginning at time 0. The MINMOD Millenium computer package6, 7 was used to estimate the acute insulin response to glucose (AIRg). The AIRg represents the acute insulin response and is defined as the area under the serum insulin curve between 0 and 10 minutes.

**Laboratory methods**: Plasma glucose concentration was measured by an automated glucose analyzer (Yellow Springs Instruments Co.) Serum insulin concentration was measured by using a chemiluminescent immunoassay (Beckman Coulter Access 2) and HbA1c levels with HPLC (Variant II Turbo, BIORAD). Lipid concentrations (cholesterol, triglycerides, and HDL cholesterol), apolipoprotein AI, apolipoprotein B, uric acid, creatinine, hepatic enzymes and C reactive protein (CRP) were measured using colorimetric assays (Unicel DxC 600 Synchron Clinical System Beckman Coulter). Plasma adiponectin, leptin, and fibroblast growth factor (FGF)-21 concentrations were determined by performing ELISA assays (Merck Millipore).

**Population stratification**: A principal components analysis was performed on 32 ancestry informative markers genotypes, previously validated against whole genomic data using EIGENSTRAT software 9. The top two principal components were used as covariates in the linear regression model to correct for ancestry.

**ENSANUT Medio Camino**

ENSANUT Medio Camino 2016 is a population-based survey with probabilistic, multi-stage, stratified and cluster sampling, with regional representativeness (north, center, Mexico City and south) and by urban and rural strata. It comprised a sample of 9,474 homes, and the population interviewed was 29,795 individuals, of whom 8,824 were adults 20 years and older. The survey period was conducted from May to October 2016. The ENSANUT 2016 questionnaire is the same as the one applied in 2006 and 2012. In detail, the diagnostic questions for previous medical diagnosis of diabetes, hypertension and high cholesterol are similar. Weight and height measurements were taken using the same protocols as on previous occasions. A blood sample was taken in a stratified random representative subsample of this population (n=4,180).

For this evaluation, we included subjects with self-report of previous medical diagnosis of T2D or taking oral T2D medications or having either FPG ≥126mg/dL or HbA1c ≥6.5% regardless of time since T2D diagnosis. Plasma glucose concentration was measured by an automated glucose analyzer (Yellow Springs Instruments Co.) Serum insulin concentration was measured by using a chemiluminescent immunoassay (Beckman Coulter Access 2) and HbA1c levels with HPLC (Variant II Turbo, BIORAD). Lipid concentrations (cholesterol, triglycerides, and HDL cholesterol) and uric acid were measured using colorimetric assays (Unicel DxC 600 Synchron Clinical System Beckman Coulter).

**CAIPaDI cohort**

The CAIPaDi program consists of two phases. The first phase comprises an initial and 3 visits one month apart each one taking place in a single 7 hours shift. The interventions include medical care, diabetes education, nutrition, physical activity, psychological evaluation, psychiatric assessment, eye exam, foot and dental care. These are delivered by one nurse, two endocrinologists, a diabetes educator (DE), a nutritionist, an ophthalmologist, a psychologist, a psychiatrist, a physical activity instructor and a dentist. Each intervention follows a procedure manual and has: 1) a specific goal, 2) a self-management strategy and 3) prespecified indicators. Each session is 30 to 60 minutes long; some of them are group meetings in which a predesigned dynamic is executed. Blood test, EKG, weight and height are obtained at arrival; blood test

results are available in 2 hours and attached to every medical record so specialists can adapt and adjust the treatment according to their results.

The second phase consists of annual evaluations where all interventions from the initial phase are reinforced. A continuous at-distance support system was implemented to maintain communication with patients via e-mail, phone calls, text messages, and through the hospital's webpage. (http://innsz.mx/opencms/contenido/departamentos/CAIPaDi). The patients for this study were enrolled from November 1st, 2013 to September 30th, 2019. Inclusion criteria were: type 2 diabetes patients, ≤ 5 years of diagnosis, without disabling complications (blindness, renal failure, stroke, limb amputations, ischemic heart disease) and non-smokers; when smokers, patients attended a Smoking Cessation Clinic as part of the treatment for 6 months before entering the program given the negative impact of smoking in diabetes. If selected, patients received a phone call and an e-mail with the information of their first visit appointment. The Institutional Ethics and Research Committees from the National Institute of Medical Sciences and Nutrition Salvador Zubirán (INCMNSZ for its name in Spanish) approved this study (Ref 1198) and it was registered in ClinicalTrials.gov (NCT02836808). All patients signed an informed consent form. Each visit was held at the CAIPaDi Center.

**Procedures:** Participants could participate in groups of 10 people in individual sessions depending on the intervention, with a close relative being encouraged to participate with them. Every one of these interventions followed a procedure manual and included a checklist of the main actions to be implemented and variables to be measured. The aim of visit 1 was to obtain a complete assessment of the patient and provide basic information to start the required changes. On visit 2, patients underwent a problem-oriented evaluation, where the recommendations were selected based on patient´s profile. Visit 3 focused on the identification of potential barriers that may impede metabolic control achievement and visit 4 aimed to reinforce the knowledge already acquired and evaluate the initial results of interventions. During visits 5 and 6, the barriers and their proposed solutions were reviewed. In summary, a collaborative, iterative process was applied in each intervention. To evaluate the competencies acquired in every visit, a structured exam was applied to each patient asking them to undertake activities related to self-care. All interventions were applied in every

**Laboratory procedures:** Fasting concentrations of glucose, creatinine, lipids and HbA1c (Bio-Rad Variant II Turbo HbA1c Kit 2, with HPLC method) were assessed in each visit. Albuminuria/creatinuria ratio (ACR) (SYNCHRON CX system with colorimetric method) was used for screening diabetic nephropathy at baseline and annual visits. The laboratory is certified by ISO 90001:2015 and the College of American Pathologist. Body composition was assessed by bioimpedance (body composition analyzer JAWON medical ioi353).

## SUPPLEMENTARY TABLES

**Supplementary Table 1.** Self-normalizing neural network architechture and hyperparameters utilized for each final model after hyperparameter grid search. Abbreviations: IL: Input Layer, OL: Output Layer.

| Model | Layers | IL/OL neurons | Layer dropout rate | Kernel initializer | IL Activation function | OL Activation function | Optimizer | Loss function | Error function | Regularizaiton |
|---|---|---|---|---|---|---|---|---|---|---|
| HOMA2-IR, HbA1c, age of diabetes onset, HOMA2-β, BMI | 3 | 32/4 | 0.1 | LeCun normal | Scaled Exponential Linear (SeLU) | Softmax | ADAM | Sparse categorical cross-entropy | Accuracy | N/A |
| HOMA2-IR', HbA1c, age of diabetes onset, HOMA2-β', BMI | 3 | 32/4 | 0.2 | LeCun normal | Scaled Exponential Linear (SeLU) | Softmax | ADAMAX | Sparse categorical cross-entropy | Accuracy | NA |
| HOMA2-IR, glucose, age of diabetes onset, HOMA2-β, BMI | 3 | 32/4 | 0.1 | LeCun normal | Scaled Exponential Linear (SeLU) | Softmax | ADAM | Sparse categorical cross-entropy | Accuracy | $\lambda_{L1}$=0.001 $\lambda_{L2}$= 0.001 |
| HbA1c, age of diabetes onset, BMI, METS-VF, METS-IR | 3 | 32/4 | 0.2 | LeCun normal | Scaled Exponential Linear (SeLU) | Softmax | ADAM | Sparse categorical cross-entropy | Accuracy | NA |

**Supplementary Table 2.** Comparison of confusion matrix metrics with unsupervised classification through k-means clustering with k=4 and runs=100 compared to self-normalizing neural networks trained with the original data clustering in the validation cohorts derived from NHANES 1999-2004. Abbreviations: HbA1c: Glycated hemoglobin; BMI: Body-mass index; METS-VF: Metabolic Score for Visceral Fat; METS-IR: Metabolic Score for Insulin Resistance, NIR:No information rate, SNNN:Self-normalizing neural networks; SIRD: Severe Insulin Resistant Diabetes; SIDD: Severe Insulin Defficient Diabetes; MARD: Mild Age-Related Diabetes; MOD: Mild Obesity-Related Diabetes.

| Model | Method | Accuracy (95%CI) | NIR | Kappa | McNemar's test | Balanced accuracy SIRD | Balanced accuracy SIDD | Balanced accuracy MARD | Balanced accuracy MOD |
|---|---|---|---|---|---|---|---|---|---|
| HOMA2-IR', HbA1c, age of diabetes onset, HOMA2-β', BMI | K-means clustering | 0.686 (0.663-0.708) | 0.466 | 0.545 | <0.001 | 0.636 | 0.969 | 0.806 | 0.759 |
| | SNNN | 0.856 (0.826-0.883) | 0.374 | 0.801 | 0.007 | 0.7905 | 0.961 | 0.894 | 0.956 |
| HOMA2-IR', glucose, age of diabetes onset, HOMA2-β', BMI | K-means clustering | 0.339 (0.376-0.422) | 0.311 | 0.191 | <0.001 | 0.398 | 0.485 | 0.829 | 0.693 |
| | SNNN | 0.803 (0.770-0.834) | 0.375 | 0.728 | 0.001 | 0.786 | 0.845 | 0.875 | 0.933 |
| HbA1c, age of diabetes onset, BMI, METS-VF, METS-IR | K-means clustering | 0.451 (0.427-0.474) | 0.423 | 0.227 | <0.001 | 0.539 | 0.931 | 0.521 | 0.611 |
| | SNNN | 0.840 (0.809-0.868) | 0.390 | 0.778 | 0.147 | 0.769 | 0.953 | 0.880 | 0.953 |

**Supplementary Table 3.** Comparison of confusion matrix metrics and diagnostic performance for each SNNN model stratified by diabetes subgroups in the validation cohort. ***p<0.001 for comparison of Accuracy>NIR. Model 1: HOMA2-IR, HbA1c, age of diabetes onset, HOMA2-β, BMI; Model 2: HOMA2-IR', HbA1c, age of diabetes onset, HOMA2-β', BMI; Model 3: HOMA2-IR', glucose, age of diabetes onset, HOMA2-β', BMI; Model 4: HbA1c, age of diabetes onset, BMI, METS-IR, METS-VF. Abbreviations: HbA1c: Glycated hemoglobin; BMI: Body-mass index; METS-VF: Metabolic Score for Visceral Fat; METS-IR: Metabolic Score for Insulin Resistance, NIR:No Information Rate, SNNN:Self-normalizing neural networks; SIRD: Severe Insulin Resistant Diabetes; SIDD: Severe Insulin Defficient Diabetes; MARD: Mild Age-Related Diabetes; MOD: Mild Obesity-Related Diabetes.

| Diabetes Subgroup | Model | Accuracy | NIR | Kappa (p-value) | Sensitivity | Specificity | PPV | NPV | Balanced accuracy |
|---|---|---|---|---|---|---|---|---|---|
| SIRD | 1 | 0.9984 | 0.8243* | 0.9945 | 100.0% | 99.09% | 99.81% | 100.0% | 0.9955 |
| | 2 | 0.8914 | 0.7764* | 0.6454 | 95.9% | 65.6% | 90.1% | 81.7% | 0.7973 |
| | 3 | 0.8802 | 0.7859* | 0.6180 | 94.9% | 62.7% | 90.3% | 77.1% | 0.788 |
| | 4 | 0.8690 | 0.7875* | 0.5810 | 94.1% | 60.1% | 89.7% | 73.4% | 0.7713 |
| SIDD | 1 | 0.9936 | 0.8211* | 0.9784 | 99.4% | 99.1% | 99.8% | 97.4% | 0.9926 |
| | 2 | 0.9776 | 0.8147* | 0.9254 | 98.8% | 93.1% | 98.4% | 94.7% | 0.9596 |
| | 3 | 0.9105 | 0.8019* | 0.7096 | 95.4% | 77.4% | 93.5% | 79.8% | 0.8440 |
| | 4 | 0.9696 | 0.8227* | 0.8971 | 97.9% | 92.8% | 98.4% | 90.3% | 0.9533 |
| MARD | 1 | 0.9920 | 0.5911* | 0.9835 | 98.65% | 100.0% | 100.0% | 98.08% | 0.9932 |
| | 2 | 0.8898 | 0.6294* | 0.7696 | 87.6% | 91.4% | 94.5% | 81.2% | 0.8947 |
| | 3 | 0.8690 | 0.6342* | 0.7260 | 85.6% | 89.1% | 91.2% | 78.2% | 0.8136 |
| | 4 | 0.8802 | 0.6134* | 0.751 | 87.8% | 88.4% | 92.3% | 82.0% | 0.8810 |
| MOD | 1 | 0.9936 | 0.7968* | 0.9820 | 100.0% | 97.26% | 99.17% | 100.0% | 0.9863 |
| | 2 | 0.9633 | 0.7907* | 0.8923 | 96.6% | 95.4% | 98.8% | 88.0% | 0.9599 |
| | 3 | 0.9441 | 0.7971* | 0.8344 | 95.0% | 92.1% | 97.9% | 82.4% | 0.9356 |
| | 4 | 0.9633 | 0.7812* | 0.8939 | 97.1% | 93.4% | 98.1% | 90.1% | 0.9528 |

**Supplementary Table 4.** Comparison of measures obtained from euglycemic hyperinsulinaemic clamp, FSIVGTT, magnetic resonance imaging, body composition by DXA and adipokines across diabetes subgroups after application of the SNNN algorithm in the SIGMA cohort. *p<0.05 of SIRD compared to MOD, ** p<0.05 SIRD vs. MARD, ***p<0.05 SIDD vs. MOD/SIRD, #p<0.05 MARD vs. MOD, p<0.05 SIDD vs. SIRD, ###p<0.05 MARD/MOD vs. SIDD, ^p<0.05 SIID/MARD vs. MOD. Abbreviations: MRI: Magnetic Resonance Imaging; VAT: Visceral Adipose Tissue; DXA: Dual X-ray Absorciometry, SNNN: Self-normalizing neural networks; SIRD: Severe Insulin Resistant Diabetes; SIDD: Severe Insulin Deficient Diabetes; MARD: Mild Age-Related Diabetes; MOD: Mild Obesity-Related Diabetes.

| Parameters | MARD (n=19) | MOD (n=40) | SIDD (n=3) | SIRD (n=5) |
|---|---|---|---|---|
| Undadjusted M-value (mg/dL*min) | 409.5 (318.5-515.7) | 439.8 (273.0-542.2) | 379.2 (318.5 -439.8) | 248.70 (163.05-439.85)* |
| M-value adjusted by weight (mg/dL*min*kg) | 6.36 (4.23-7.25) | 5.66 (4.01-7.41) | 5.0 (4.5-5.4) | 4.02 (2.66-6.93) |
| M-I index | 7.86 (4.38-9.87) | 6.60 (3.95-8.30) | 5.36 (4.57-5.45) | 4.12 (3.10-6.63)** |
| AIRg (mU*L/min) | 46.4 (22.3-112.47)** | 137.0 (50.12-320.75)# | 37.0 (9.52-57.1) | 303.0 (59.1-780.5)*** |
| Liver fat content (%) | 3.84 (1.39-6.91) | 5.48 (1.60-10.59) | 6.08 (3.91-8.03) | 4.06 (2.63-6.94)*, ## |
| Pancreatic fat content (%) | 5.37 (3.0-9.60) | 4.81 (1.87-8.84) | 5.50 (2.94-6.97) | 8.04 (1.55-12.98) |
| Visceral fat area by MRI ($cm^2$) | 84.98 (72.75-127.62) | 92.6 (68.08-139.90) | 170.40 (166.80-253.8)### | 125.26 (66.63-205.35) |
| Subcutaneous fat area by MRI (cm2) | 194.22 (128.56-253.73) | 281.88 (211.0-371.40)# | 204.50 (188.0-210.4) | 125.26 (66.63-205.35) |
| Subcutaneous to visceral fat ratio | 2.00 (1.55-2.89) | 2.82 (2.00-3.77) | 1.20 (1.13-1.25)^ | 1.14 (1.10-3.55) |
| Total fat mass (kg) | 25.55 (20.27-27.40) | 27.94 (23.17-34.33)# | 24.08 (21.24-27.2) | 24.19 (21.75-30.29) |
| Total lean mass (kg) | 44.36 (37.55-50.52) | 40.21 (36.89-57.78) | 41.45 (37.13-45.78) | 33.44 (32.84-97.09)*, ** |
| Bone mineral content (kg) | 2.33 (2.03-2.73) | 2.23 (1.98-2.73) | 2.09 (1.68-2.51) | 1.60 (1.40-2.02) *, ** |
| VAT mass DXA (kg) | 1.45 (1.01-2.13) | 1.24 (1.07-1.99) | 1.94 (1.20-2.06) | 1.31 (1.61-1.67) |
| Adiponectin (ng/L) | 7.56 (5.47-9.23) | 6.25 (4.73-9.38) | 5.79 (3.37-9.50) | 6.49 (2.36-10.07) |
| Leptin | 12.07 (8.68-19.36) | 13.96 (7.06-22.66) | 17.36 (15.06-33.64) | 11.93 (9.42-18.39) |
| FGF-21 | 202.64 (67.73-353.77) | 127.14 (43.06-288.12) | 228.35 (154.84-308.99) | 276.76 (137.11-409.81) |

**Supplementary Table 5.** Comparisons of clinical measures from the Metabolic syndrome cohort before and after follow-up according to diabetes subgroup and compared to age, sex and BMI propensity score matched controls (n=991); *p<0.05 using paired tests, **p<0.05 compared to controls. Abbreviations: AHS, Arterial Hypertension; WC, Waist Cirumference; FPG, Fasting Plasma Glucose; TG, Triglycerides; TC, Total cholesterol; CRP, C-reactive protein; WHtR: Waist to height ratio; BMI, Body-mass index; METS-IR, Metabolic Score for Insulin Resistance; METS-VF, Metabolic Score for Visceral Fat; MARD, Mild Age-Related Diabetes; MOD: Mild Obesity-related diabetes; SIDD: Severe Insulin Deficient Diabetes; SIRD: Severe Insulin Resistant Diabetes.

| Parameters | Controls (n=991) | | MARD (n=44) | | MOD (n=139) | | SIDD (n=19) | | SIRD (n=118) | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Baseline | Follow-up | Baseline | Follow-up | Baseline | Follow-up | Baseline | Follow-up | Baseline | Follow-up |
| Age (years) | 45.7±11.4 | 52.0±12.6 | 57.2±8.7 | 60.09±8.4 | 39.2±7.3 | 41.9±7.1 | 46.9±7.1 | 49.9±7.1 | 49.8±10.4 | 52.5±10.5 |
| Female sex (%) | 645 (65.1) | | 30 (66.7) | | 97 (65.1%) | | 11 (57.9) | | 81 (68.6) | |
| HAAS (%) | 301 (30.4) | 309 (31.2) | 26 (57.8) | 26 (57.8) | 38 (25.5) | 58 (38.9)* | 6 (31.6) | 8 (42.1) | 55 (46.6) | 61 (51.7)* |
| HOMA2-β (%) | 142.32±60.23 | 141.15±68.08 | **86.11±30.60** | 69.76±19.19* | 132.37±56.66 | 97.89±40.13* | 116.06±51.52 | 33.02±18.32* | 157.16±102.88 | 169.13±75.69 |
| HOMA2-IR | 1.63±1.02 | 1.69±1.21 | 1.20±0.51 | 1.29±0.63 | 2.13±1.34 | **2.00±1.41** | 2.09±1.04 | **2.35±1.04** | 2.21±1.20 | 2.95±2.37* |
| HOMA2-S (%) | 83.53±57.69 | 101.29±115.92 | 100.81±52.96 | 112.84±74.14 | 70.67±99.42 | 86.28±86.56 | 59.56±27.85 | 54.55±34.54 | 59.37±31.14 | 53.34±35.43 |
| WC (cm) | 96.22±11.86 | 96.62±11.31 | 93.18±9.95 | 91.84±9.50 | 98.48±13.42 | 100.18±13.59* | 103.44±10.41 | 103.29±9.90 | 98.84±11.73 | 99.40±37 |
| FPG (mg/dL) | 86.77±10.77 | 87.41±11.91 | 101.12±14.02 | 109.36±21.46* | 97.94±12.69 | 111.24±26.46* | 105.26±12.98 | 219.58±59.35* | 94.21±13.35 | 97.75±18.85* |
| Insulin (uU/mL) | 12.98±8.17 | 13.38±9.58 | 8.87±3.58 | 8.93±4.09 | 16.54±10.38 | 14.95±10.71 | 15.80±8.19 | 14.44±7.26 | 17.40±11.15 | 23.66±20.64* |
| TG (mg/dL) | 167.0 (118.3-242.5) | | 178.1 (119.0-223.7) | 169.0 (107.5-220.7) | 198.6 (138.5-272.2) | 168.0 (125.0-250.0) | 204.0 (151.9-279.2) | 263.0 (173.0-385.0)* | 198.6 (132.0-283.4) | 183.0 (139.0-257.5) |
| TC (mg/dL) | 209.09±41.06 | 200.23±40.46 | 224.52±43.97 | 212.45±45.52 | 207.50±39.55 | 201.03±36.81* | 210.79±35.65 | 209.79±31.06 | 210.85±41.11 | 199.94±40.93* |
| HDL-C (mg/dL) | 44.40±11.45 | 41.46±11.77 | 46.09±11.62 | 44.71±12.50 | 41.12±8.87 | 39.84±9.87 | 42.53±10.84 | 38.63±10.72 | 42.32±12.68 | 39.25±11.18* |
| ApoB (mg/dL) | 111.24±26.97 | 105.13±25.72 | 117.87±28.15 | 113.24±34.61 | 113.62±26.28 | 107.04±24.02* | 118.52±17.06 | 122.34±23.56 | 114.81±26.13 | 106.15±28.87* |
| CRP | 3.28±2.72 | 3.10±3.088 | 2.62±1.99 | 2.63±3.11 | 4.23±3.23 | 3.64±2.86 | 4.53±3.85 | 3.32±1.13 | 6.08±3.52 | 5.61±4.48 |
| WHtR | 0.606±0.076 | 0.608±0.072 | 0.582±0.063 | 0.574±0.059 | 0.615±0.080 | 0.626±0.086* | 0.657±0.077 | 0.655±0.069 | 0.625±0.067 | 0.630±0.071 |
| BMI (kg/m2) | 30.40±4.79 | 30.31±4.87 | 26.99±2.96 | 26.20±3.09* | 31.67±5.45 | 31.60±5.54 | 32.09±4.48 | 31.44±3.37 | 31.29±4.65 | 31.11±4.50 |
| METS-VF | 6.74±0.46 | 6.80±0.45 | 6.75±0.44 | 6.75±0.44 | 6.75±0.45 | 6.82±0.49* | 7.01±0.27 | 7.10±0.25* | 6.93±4.65 | 6.99±0.39*,** |
| METS-IR | 47.88±9.98 | 48.32±10.12 | 42.14±6.53 | 41.37±6.73 | 52.03±11.36 | 52.67±11.68 | 54.00±8.62 | 58.76±9.25* | 50.42±9.41 | 51.42±9.26* |

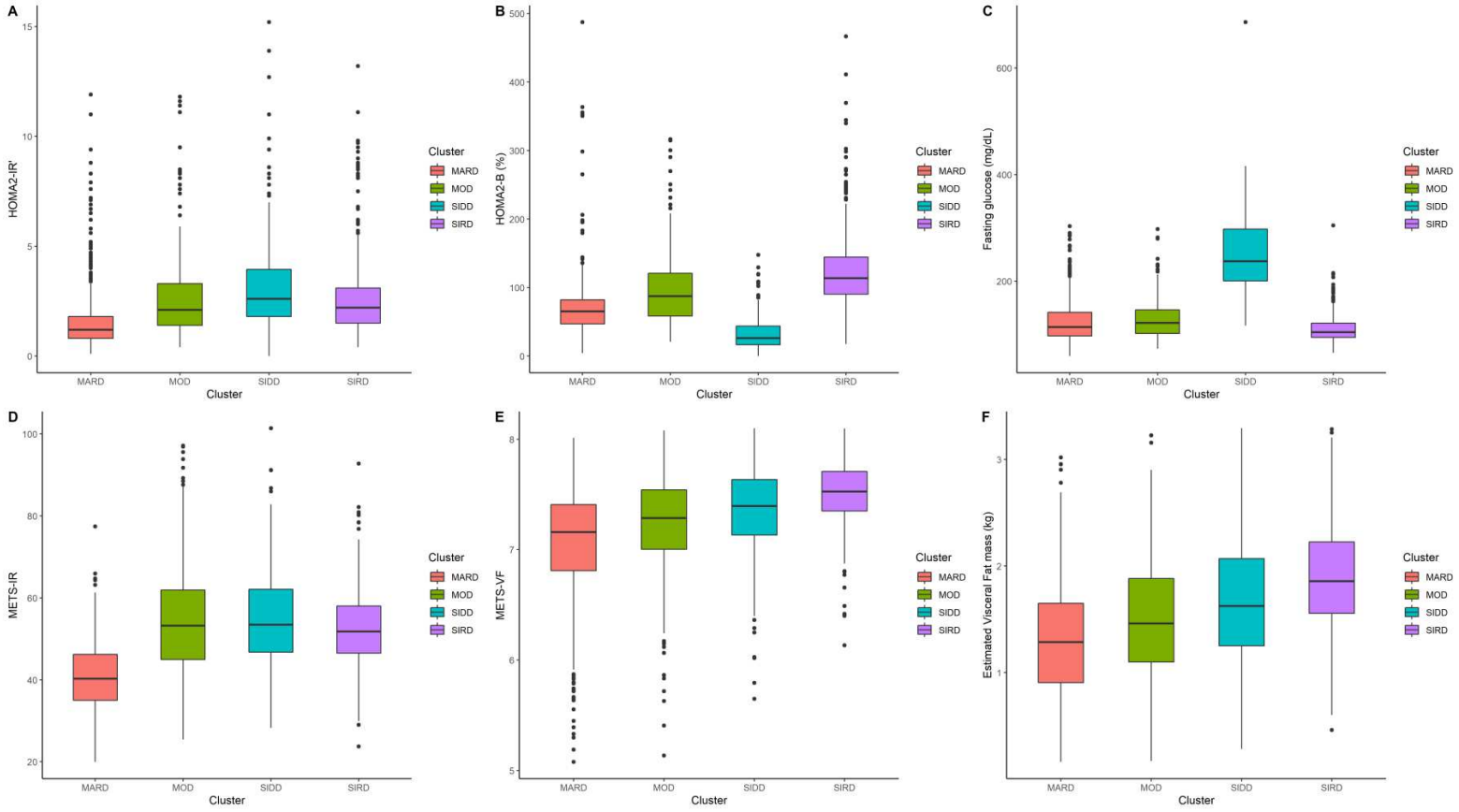
**Supplementary Table 6.** Comparison of biochemical measures, surrogate estimates and electrical bioimpedance-derived measures across diabetes subgroups in the UIEM cohort (n=1,521), classified using the SNNN model 2. Abbreviations: FPG, Fasting Plasma Glucose; TG, Triglycerides; TC, Total cholesterol; CRP, C-reactive protein, Body-mass index; METS-VF, Metabolic Score for Visceral Fat; GFR: Glomerular Filtration Rate; MARD, Mild Age-Related Diabetes; MOD: Mild Obesity-related diabetes; SIDD: Severe Insulin Deficient Diabetes; SIRD: Severe Insulin Resistant Diabetes.

| Parameters | MARD (n=255) | MOD (n=518) | SIDD (n=627) | SIRD (n=121) |
|---|---|---|---|---|
| Female sex (%) | 162 (63.5%) | 339 (65.4%) | 350 (55.8%)* | 80 (66.1%) |
| *SLC16A11* carriers (% from overall) | 125 (52.5%)** | 314 (64.1%)** | 370 (62.0%) | 68 (59.4%) |
| Age (years) | 2.0 (0.0-8.0) | 5.0 (0.0-13.0) | 7.0 (1.0-15.0) | 2.0 (0.0-10.0) |
| HbA1c (%) | 6.65±0.90 | 6.92±0.94 | 10.58±1.91 | 6.82±1.19 |
| FPG (mg/dL) | 122.09±25.77 | 126.81±25.77 | 227.08±75.48 | 99.21±21.41 |
| HOMA2B (%) | 51.4 (35.8-66.0) | 68.9 (49.4-93.1) | 21.1 (10.8-37.2) | 144.0 (100.4-227.2) |
| HOMA2S (%) | 111.50 (80.2-163.0) | 64.6 (41.6-100.0) | 66.90 (42.4-109.3) | 42.8 (24.1-61.0) |
| HOMA2-IR | 0.90 (0.60-1.20) | 1.50 (1.0-2.40) | 1.50 (0.90-2.40) | 2.30 (1.60-4.15) |
| TG (mg/dL) | 148.0 (112.0-210.0) | 169.50 (125.7-237.0) | 209.0 (143.0-299.0)** | 155.0 (107.5-233.0) |
| TC (mg/dL) | 195.49±40.72 | 192.07±44.21 | 208.66±56.13** | 192.60±47.17 |
| HDL-C (mg/dL) | 47.52±13.93** | 43.26±11.05 | 43.04±10.90 | 45.18±13.45 |
| Uric acid (mg/dL) | 5.59±1.39 | 5.77±1.48 | 4.98±1.48** | 5.89±1.48 |
| Apolipoprotein B (mg/dL) | 107.10±26.41 | 108.57±27.68 | 118.11±29.74** | 107.73±29.79 |
| BMI (kg/m2) | 26.26±3.35 | 32.31±6.99 | 29.09±4.86 | 31.66±4.72 |
| Fat-mass index (kg/m2) | 11.29±7.61 | 14.37±8.09** | 12.36±7.13 | 14.19±4.87** |
| Fat-free mass index (kg/m2) | 17.11±5.80** | 19.13±6.56 | 18.18±4.30 | 20.31±21.77 |
| Phase angle | 4.85 (4.30-5.42) | 5.60 (5.10-6.10)** | 5.10 (4.60-5.60) | 5.40 (4.60-5.80)** |
| Visceral fat (L) | 3.05 (2.50-3.80) | 3.50 (2.50-4.95) | 3.50 (2.60-4.40) | 3.70 (2.95-5.00)** |
| Fatty liver index | 46.62±24.70 | 69.98±26.32 | 67.36±25.17 | 73.04±22.21** |
| METS-VF | 7.14±1.40 | 7.18±0.46 | 7.17±0.42 | 7.36±0.32* |
| Albumin-creatinine ratio | 7.0 (5.0-14.7) | 10.1 (4.9-35.7) | 20.5 (8.9-103.4)** | 9.2 (4.8-27.7) |
| GFR (mL/min/1.73m2) | 84.95±24.14** | 98.11±28.62 | 95.10±26.44 | 82.81±29.93** |

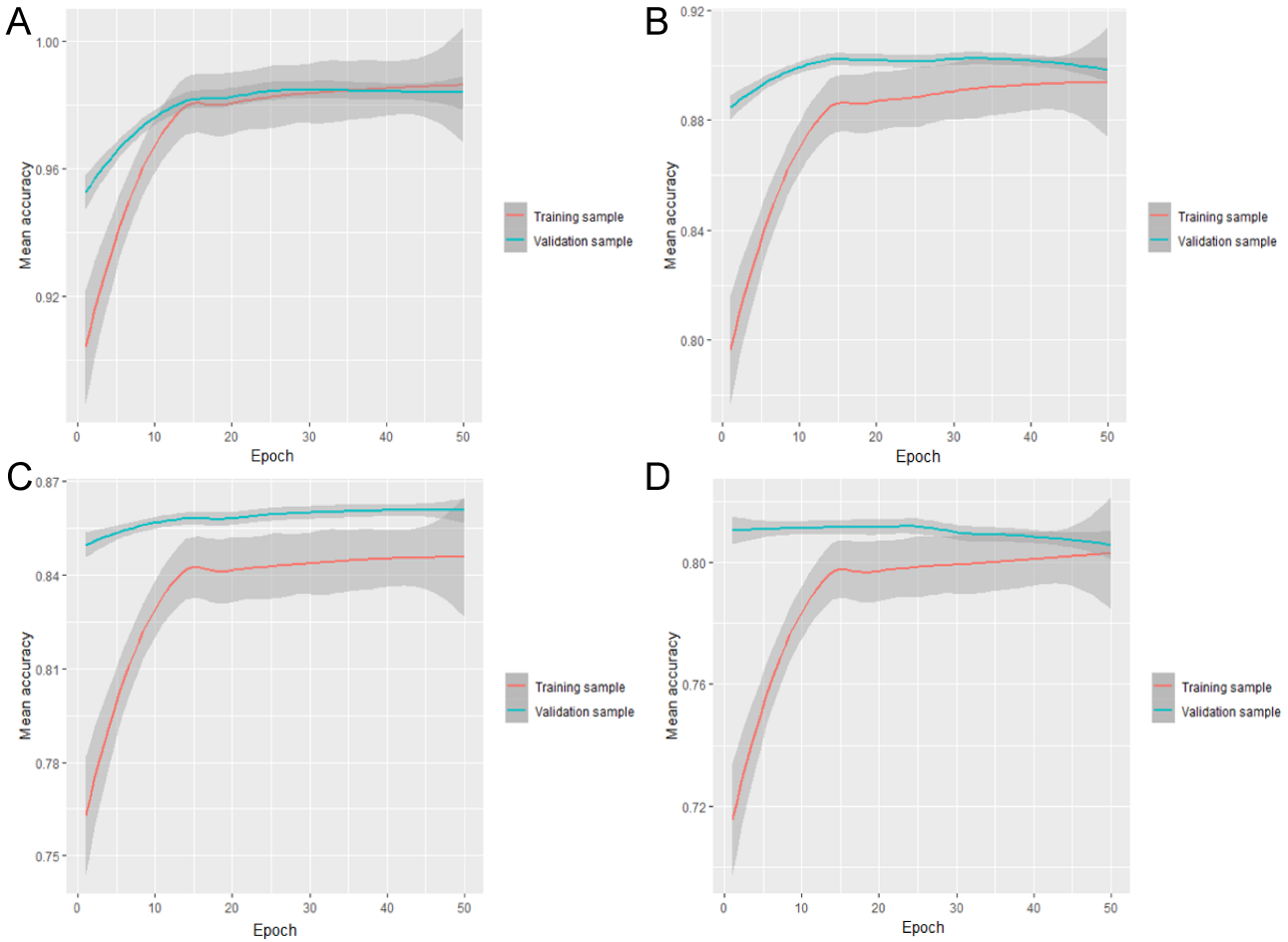
**Supplementary Table 7.** Fixed effects logistic regression models to investigate the association of specific diabetes subgroups identified by the SNNN algorithm with chronic diabetes complications in the SIGMA genotype cohort (n=1,516) adjusted for sex and years since T2D onset. NAFLD was defined as FLI >60. Abbreviations: SNNN: Self-normalizing neural networks; SIRD: Severe Insulin Resistant Diabetes; SIDD: Severe Insulin Deficient Diabetes; MARD: Mild Age-Related Diabetes; MOD: Mild Obesity-Related Diabetes; NAFLD: Non-Alcoholic fatty liver disease; FLI: Fatty Liver index; OR: Odds Ratio; 95%CI: 95% Confidence Interval.

| Diabetes subgroup | Parameter | Prevalence | Beta | OR (95%CI) | p-value |
|---|---|---|---|---|---|
| MOD | Diabetic Kidney Disease | 24.6% | -0.630 | 0.532 (0.389-0.723) | <0.001 |
| | Diabetic retinopathy | 35.1% | -0.699 | 0.512 (0.287-0.893) | 0.020 |
| | Diabetic neuropathy | 15.7% | -0.514 | 0.598 (0.408-0.863) | 0.007 |
| | NAFLD | 67.2% | 0.356 | 1.427 (1.122-1.816) | 0.004 |
| MARD | Diabetic Kidney Disease | 25.4% | -0.121 | 0.886 (0.600-1.291) | 0.534 |
| | Diabetic retinopathy | 13.6% | -1.299 | 0.272 (0.111-0.601) | 0.002 |
| | Diabetic neuropathy | 24.8% | 0.408 | 1.503 (0.943-2.349) | 0.789 |
| | NAFLD | 33.0% | -1.660 | 0.190 (0.139-0.261) | <0.001 |
| SIRD | Diabetic Kidney Disease | 27.7% | -0.098 | 0.907 (0.535-1.498) | 0.708 |
| | Diabetic retinopathy | 32.4% | -0.133 | 0.875 (0.356-2.047) | 0.764 |
| | Diabetic neuropathy | 17.4% | -0.171 | 0.843 (0.421-1.562) | 0.606 |
| | NAFLD | 76.4% | 0.721 | 2.056 (1.287-3.285) | 0.003 |
| SIDD | Diabetic Kidney Disease | 40.0% | 0.627 | 1.871 (1.420-2.470) | <0.001 |
| | Diabetic retinopathy | 54.4% | 1.184 | 3.269 (1.970-5.510) | <0.001 |
| | Diabetic neuropathy | 23.7% | 0.280 | 1.323 (0.950-1.846) | 0.982 |
| | NAFLD | 67.2% | 0.444 | 1.559 (1.234-1.970) | <0.001 |

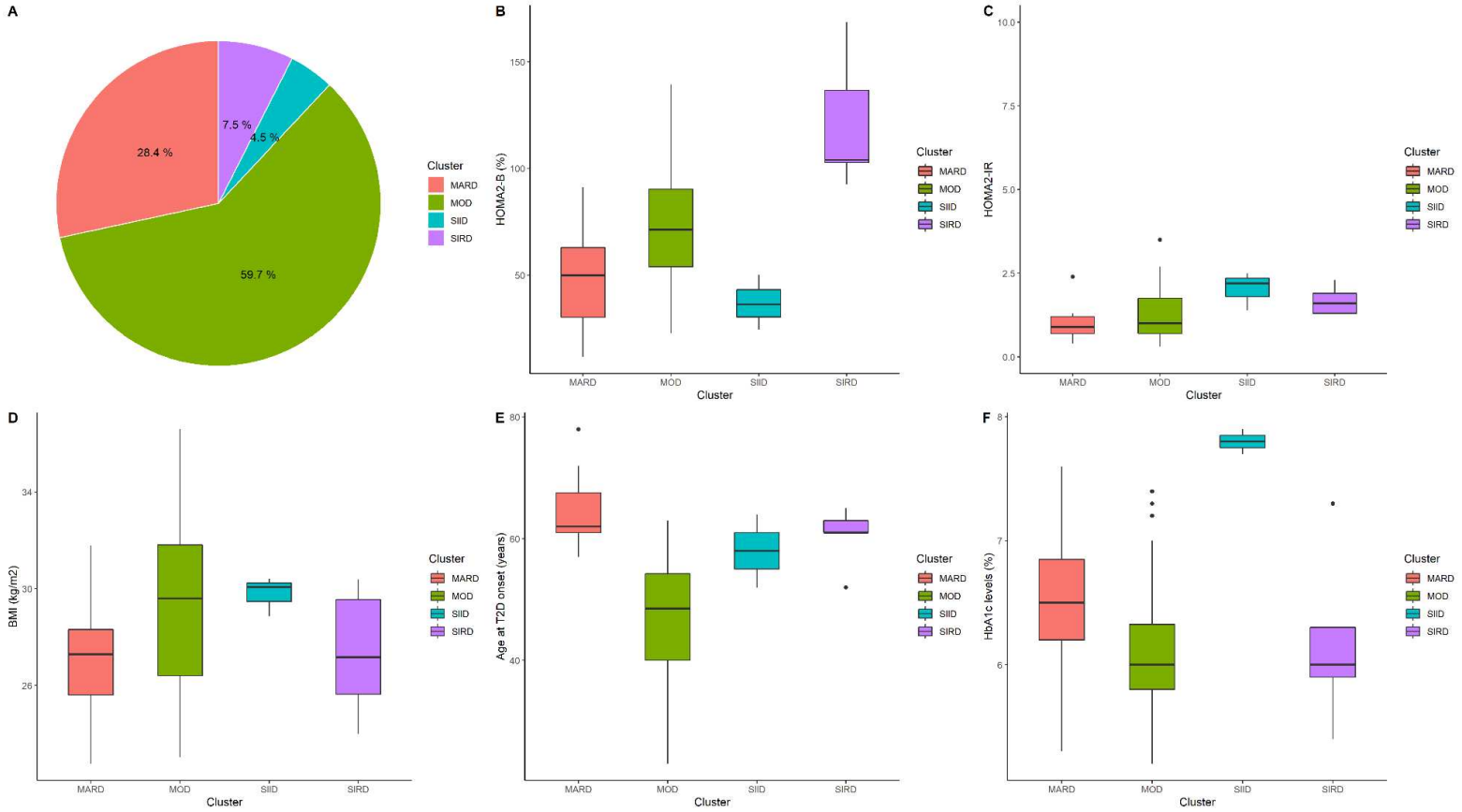
**Supplementary Figure 1.** Distribution of HOMA2-IR', HOMA2-β', fasting glucose, METS-IR, METS-VF and estimated visceral fat in the frour diabetes subgroup in the combined NHANES cohort.



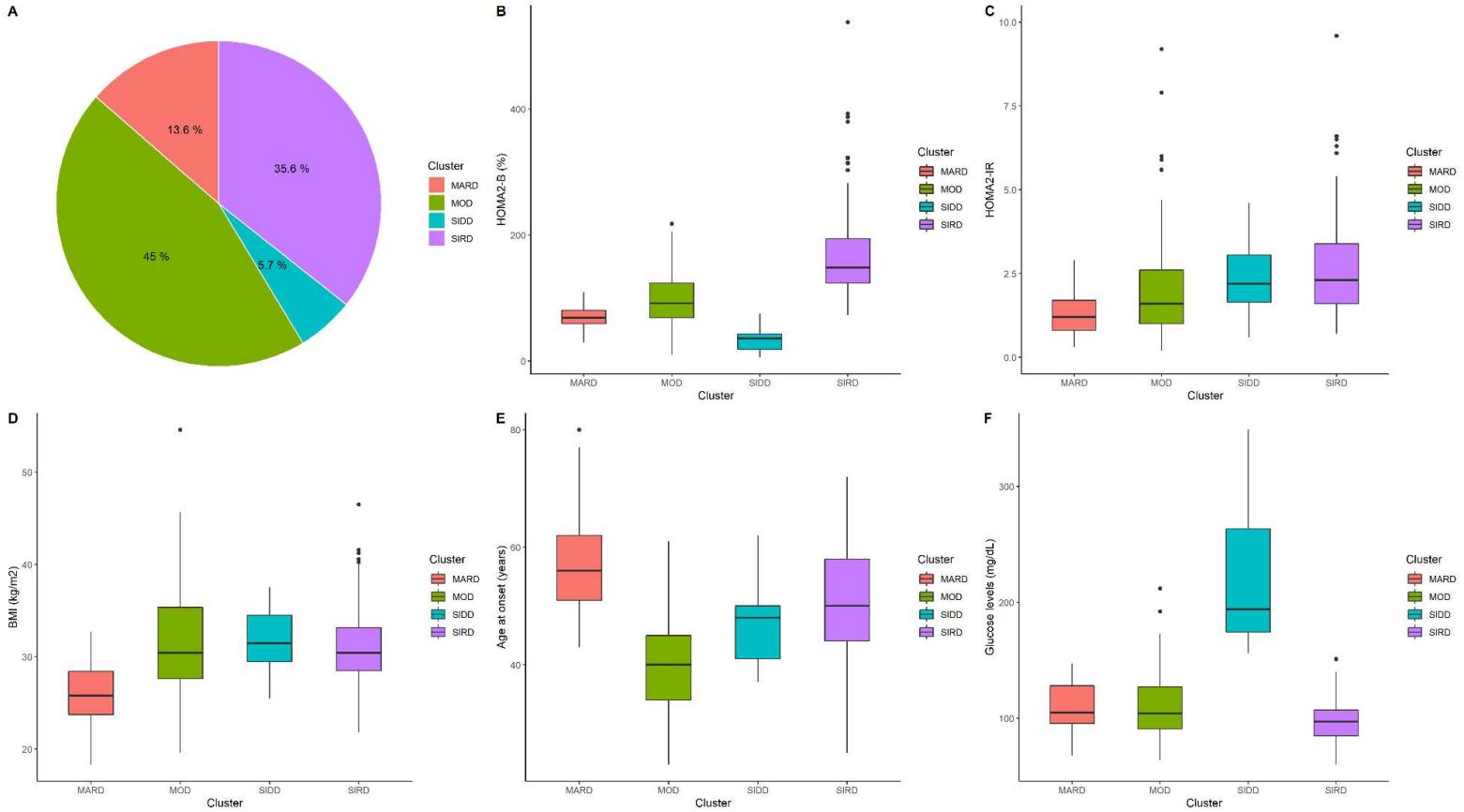
**Supplementary figure 2.** Performance of the four trained self-normalizing neural network models contrasting accuracy in training and validation simple after k-fold cross-validation (k=10).



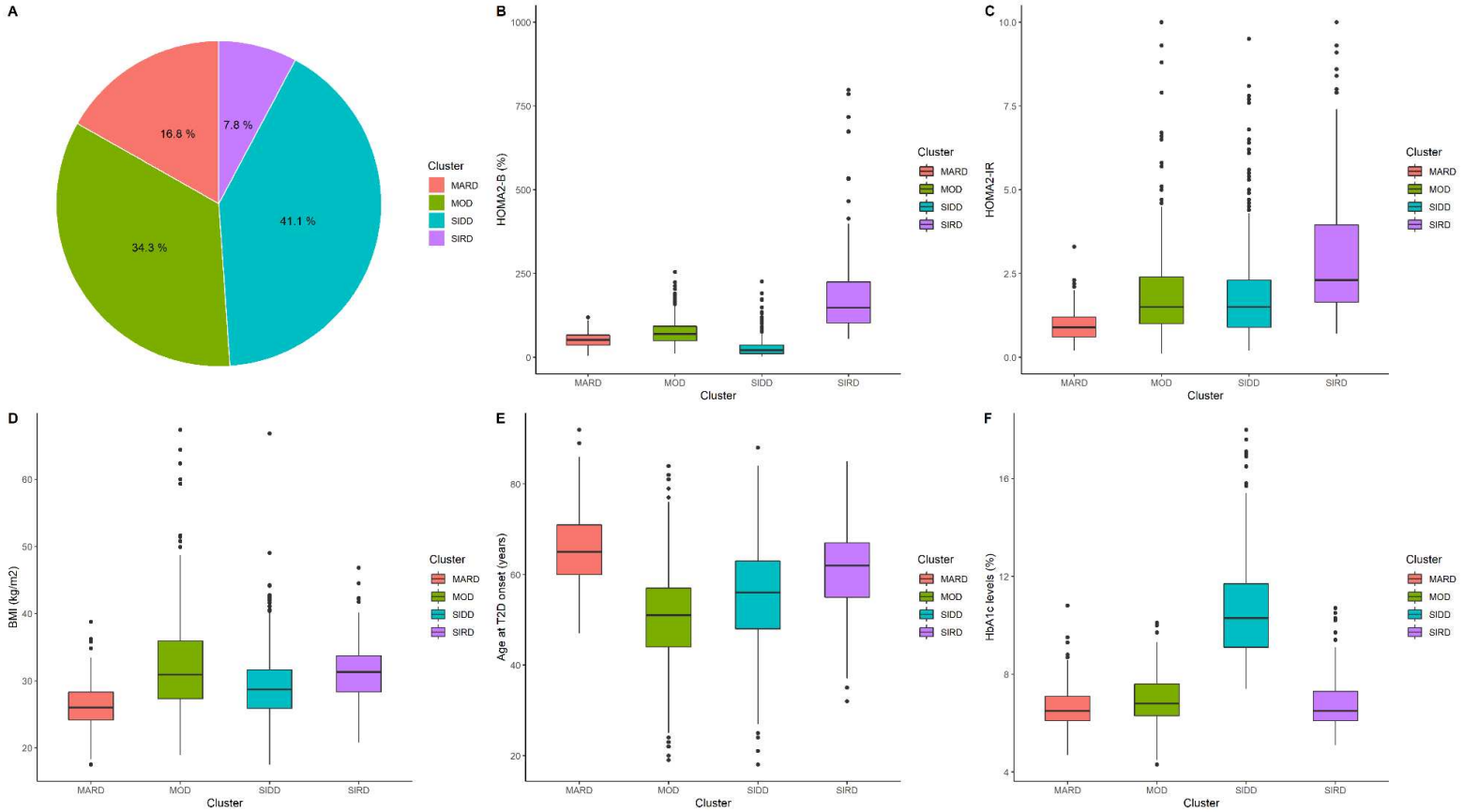
**Supplementary Figure 3.** Distribution of diabetes subgroups in the SIGMA cohort (n=67) after application of the SNNN algorithm to identify diabetes subtypes as well as a comparison of metabolic variables related to each diabetes subgroup. **Abbreviations:** HbA1c: Glycated hemoglobin; BMI: Body-mass index; HOMA: Homeostasis Model Assessment, SNNN: Self-normalizing neural networks; SIRD: Severe Insulin Resistant Diabetes; SIDD: Severe Insulin Deficient Diabetes; MARD: Mild Age-Related Diabetes; MOD: Mild Obesity-Related Diabetes



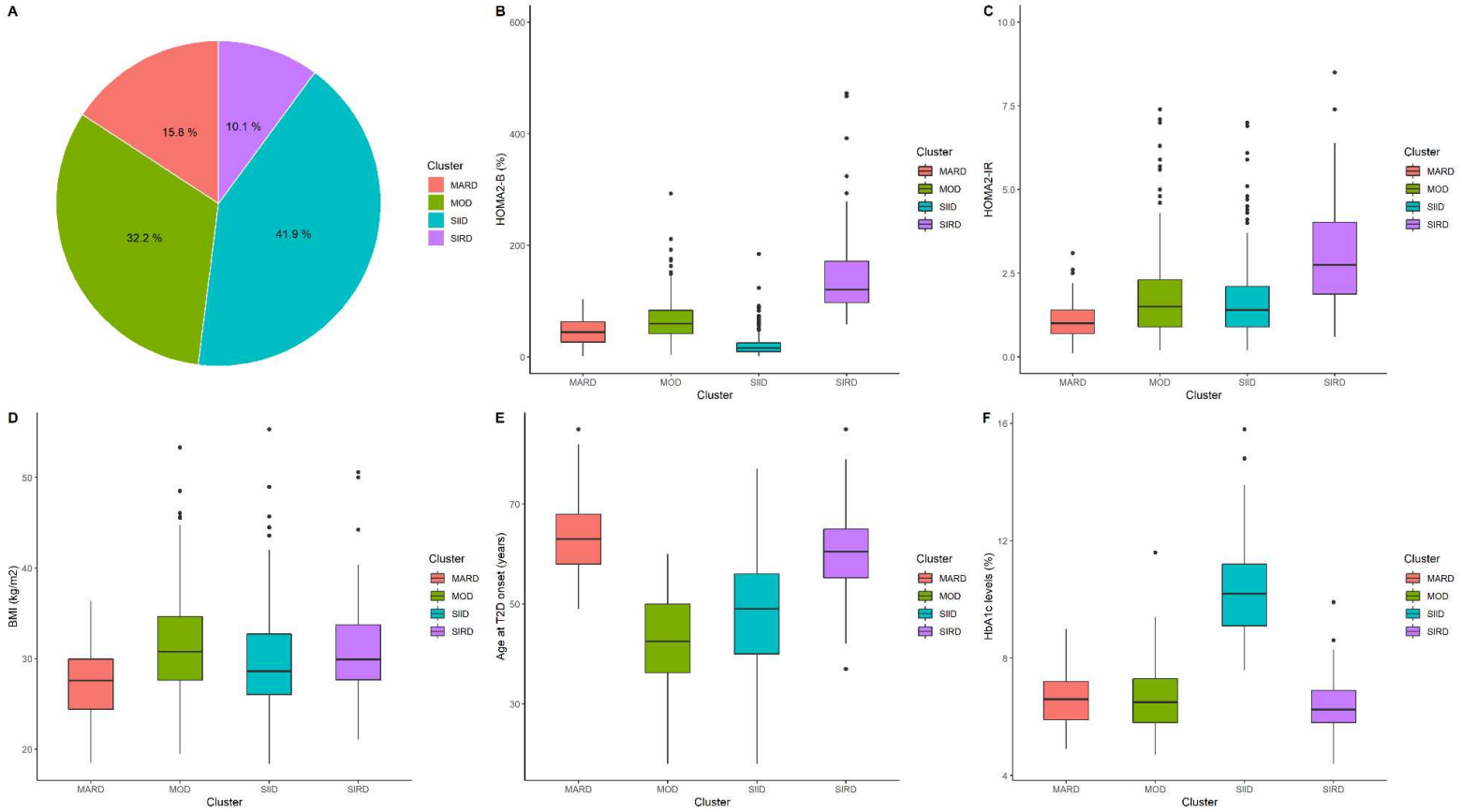
**Supplementary Figure 4.** Distribution of incident diabetes subgroups in the Metabolic Syndrome cohort (n=331) after application of the SNNN algorithm to identify diabetes subtypes after 3-years of follow-up of healthy but at-risk individuals. **Abbreviations:** HbA1c: Glycated hemoglobin; BMI: Body-mass index; HOMA: Homeostasis Model Assessment, SNNN: Self-normalizing neural networks; SIRD: Severe Insulin Resistant Diabetes; SIDD: Severe Insulin Deficient Diabetes; MARD: Mild Age-Related Diabetes; MOD: Mild Obesity-Related Diabetes



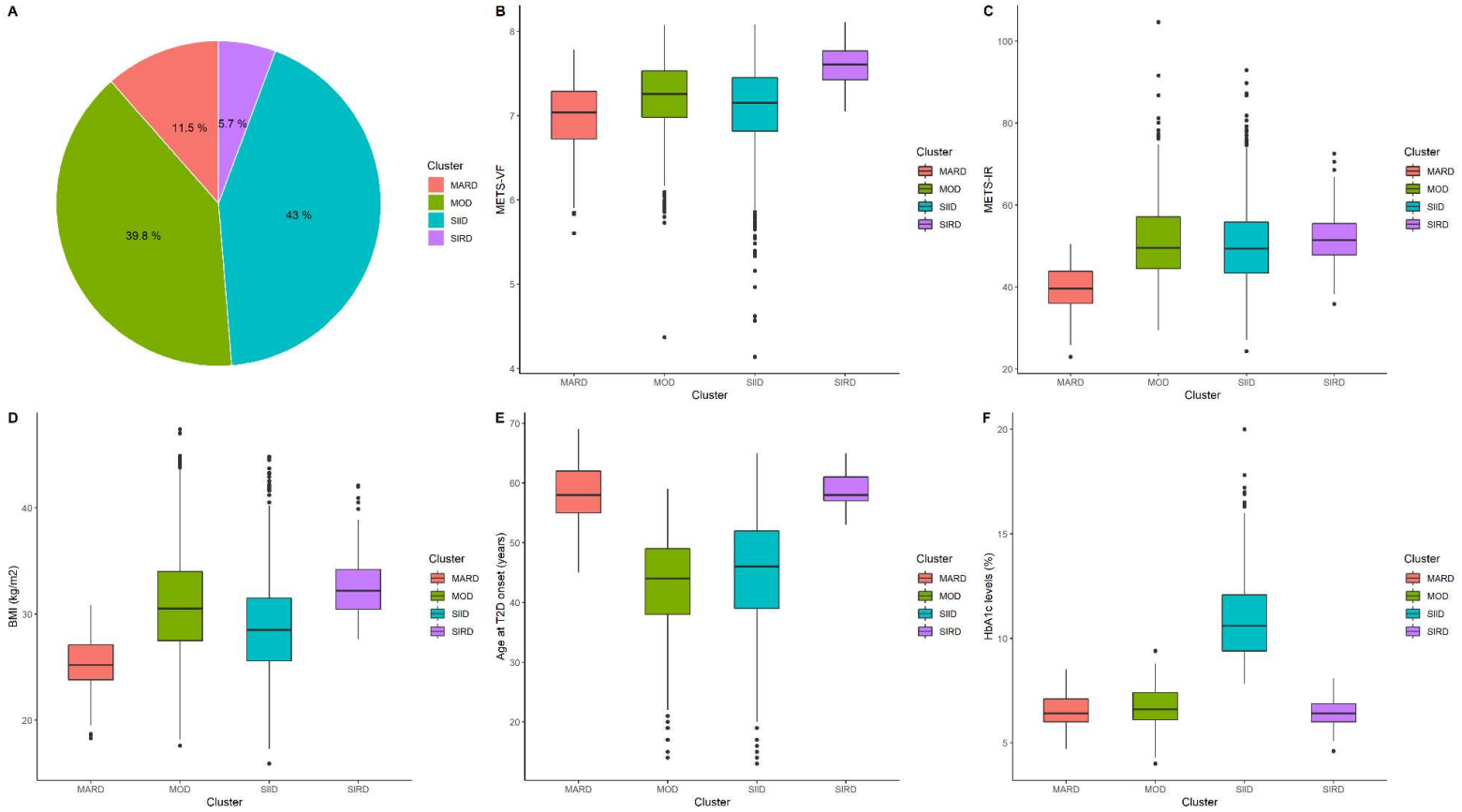
**Supplementary Figure 5.** Distribution of diabetes subgroups in the SIGMA genotype cohort (n=1,521) after application of the SNNN algorithm to identify diabetes subtypes in the evaluated population. **Abbreviations:** HbA1c: Glycated hemoglobin; BMI: Body-mass index; HOMA: Homeostasis Model Assessment, SNNN: Self-normalizing neural networks; SIRD: Severe Insulin Resistant Diabetes; SIDD: Severe Insulin Deficient Diabetes; MARD: Mild Age-Related Diabetes; MOD: Mild Obesity-Related Diabetes



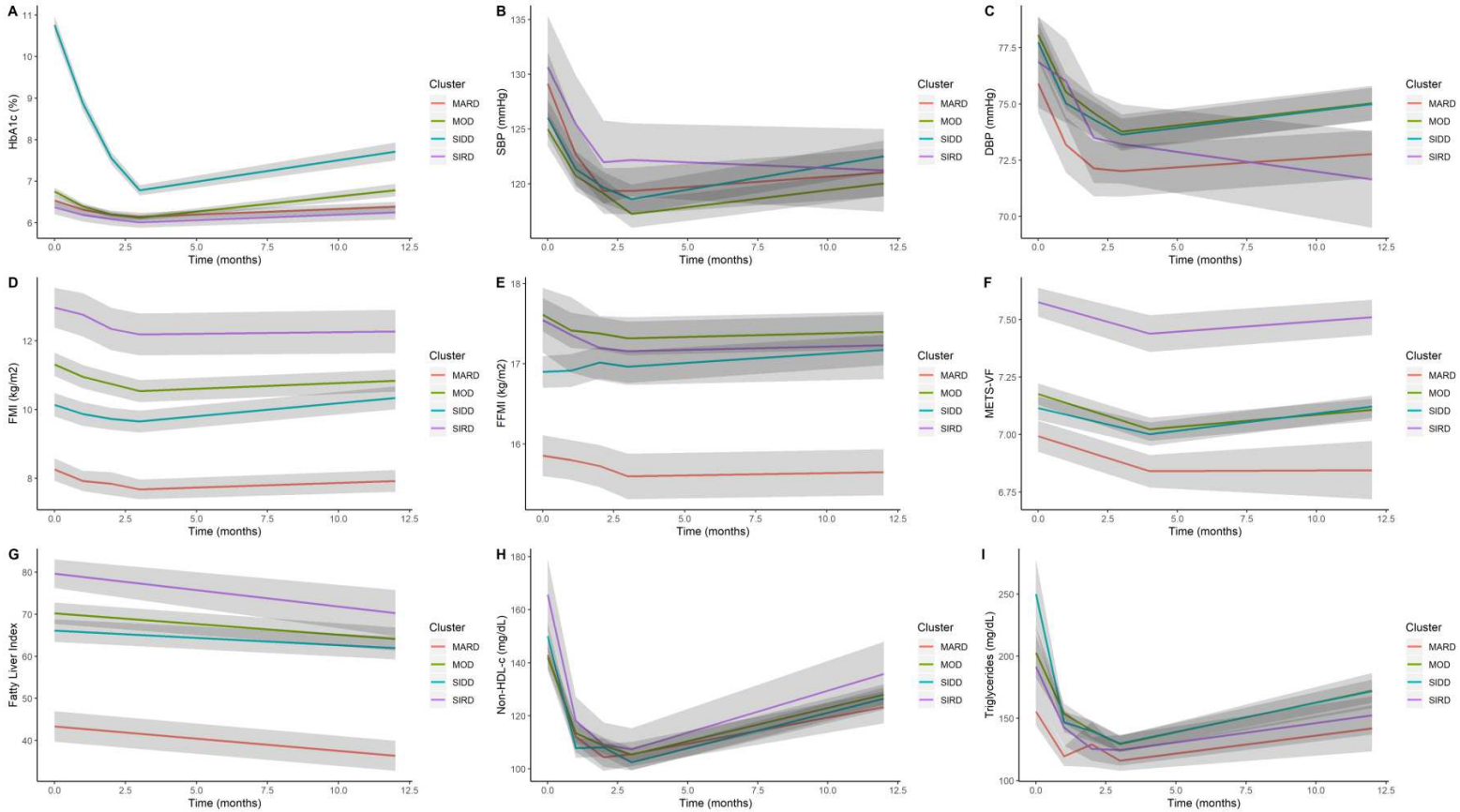
**Supplementary Figure 6.** Distribution of diabetes subgroups in ENSANUT 2016 (n=614) after application of the SNNN algorithm to identify diabetes subtypes in this population-based simple. **Abbreviations:** HbA1c: Glycated hemoglobin; BMI: Body-mass index; HOMA: Homeostasis Model Assessment, SNNN: Self-normalizing neural networks; SIRD: Severe Insulin Resistant Diabetes; SIDD: Severe Insulin Deficient Diabetes; MARD: Mild Age-Related Diabetes; MOD: Mild Obesity-Related Diabetes



**Supplementary Figure 7.** Distribution of diabetes subgroups in the CAIPaDi cohort (n=1,608) after application of the SNNN algorithm to identify diabetes subtypes in this population-based simple. **Abbreviations:** METS-VF: Metabolic Score for Visceral Fat; METS-IR: Metabolic Score for Insulin Resistance; BMI: Body-mass index; SNNN: Self-normalizing neural networks; SIRD: Severe Insulin Resistant Diabetes; SIDD: Severe Insulin Deficient Diabetes; MARD: Mild Age-Related Diabetes; MOD: Mild Obesity-Related Diabetes.



**Supplementary Figure 8.** Plot of mean trajectories of HbA1c, systolic blood pressure (SBP, B), fat-mass index (FMI, C), fat-free mass index (FFMI, D) across 3 months using the SNNN algorithm for classification using measures at baseline. **Abbreviations:** HbA1c: Glycated hemoglobin; SNNN: Self-normalizing neural networks; SIRD: Severe Insulin Resistant Diabetes; SIDD: Severe Insulin Deficient Diabetes; MARD: Mild Age-Related Diabetes; MOD: Mild Obesity-Related Diabetes



**Supplementary Figure 9.** Plot of mean trajectories of HbA1c, systolic blood pressure (SBP, B), fat-mass index (FMI, C), fat-free mass index (FFMI, D) across 3 months using the SNNN algorithm for classification using measures at the 3-month point. **Abbreviations:** HbA1c: Glycated hemoglobin; SNNN: Self-normalizing neural networks; SIRD: Severe Insulin Resistant Diabetes; SIDD: Severe Insulin Deficient Diabetes; MARD: Mild Age-Related Diabetes; MOD: Mild Obesity-Related Diabetes.