Novel subgroups of type 2 diabetes display different epigenetic patterns which associate with future diabetic complications

Short running title: Epigenetics in type 2 diabetes subgroups

Silja Schrader, MSc[1], Alexander Perfilyev, PhD[1], Emma Ahlqvist, PhD[2], Leif Groop, MD, PhD[2], Allan Vaag, MD, PhD[3], Mats Martinell, MD, PhD[4, 5], Sonia García-Calzón, PhD[1,6*] and Charlotte Ling, PhD[1*]

[*]Equal contribution

[1] Epigenetics and Diabetes Unit, Department of Clinical Sciences, Lund University Diabetes Centre, Lund University, Scania University Hospital, 205 02 Malmö, Sweden

[2] Genomics, Diabetes and Endocrinology Unit, Department of Clinical Sciences, Lund University, 205 02 Malmö, Sweden

[3]Type 2 Diabetes Biology Research, Steno Diabetes Center, Copenhagen, Denmark

[4]Department of Public Health and Caring Sciences, Uppsala University, Uppsala, Sweden

[5]Academic Primary Care Centre, Region Uppsala, Sweden

[6]Department of Food Science and Physiology, University of Navarra, Pamplona, Spain

Corresponding Authors and Lead Contacts: Sonia García-Calzón, e-mail: sgcalzon@unav.es, Tel: +34948425600 and Charlotte Ling, e-mail: charlotte.ling@med.lu.se, Tel: +46-(0)-391213

Word count: 4912

Number of tables: 2

Number of figures: 2

Abstract

Objective: Type 2 diabetes (T2D) was recently reclassified into Severe Insulin Deficient Diabetes (SIDD), Severe Insulin Resistant Diabetes (SIRD), Mild Obesity-related Diabetes (MOD), and Mild Age-Related Diabetes (MARD), which have different risk of complications. We explored whether DNA methylation differs between these subgroups and if subgroup-unique methylation risk scores (MRSs) predict diabetic complications.

Methods: Genome-wide DNA methylation was analysed in blood from newly diagnosed T2D subjects in discovery and replication cohorts. Subgroup-unique MRSs were built including top subgroup-unique DNA methylation sites. Regression models examined whether MRSs associated with subgroups and future complications.

Results: We found epigenetic differences between the T2D subgroups. Subgroup-unique MRSs were significantly different in those patients allocated to each respective subgroup compared to the combined group of all other subgroups. These associations were validated in an independent replication cohort, showing that subgroup-unique MRSs associate with individual subgroups (OR 1.6-6.1 per 1SD increase, $p<0.01$). Subgroup-unique MRSs were also associated with future complications. Higher MOD-MRS was associated with lower risk of cardiovascular (HR=0.65, $p=0.001$) and renal (HR=0.50, $p<0.001$) disease, whereas higher SIRD-MRS and MARD-MRS were associated with increased risk of these complications (HR 1.4-1.9 per 1SD increase, $p<0.01$). 39 of 95 methylation sites included in subgroup-unique MRSs were annotated to genes previously linked to diabetes-related traits, including *TXNIP* and *ELOVL2*. Methylation in blood of 18 subgroup-unique sites mirror epigenetic patterns in tissues relevant for T2D, muscle and adipose tissue.

Conclusions: We identified differential epigenetic patterns between T2D subgroups, which associated with future diabetic complications. These data support a reclassification of diabetes and the need for precision medicine in T2D subgroups.

Diabetes is responsible for over 1.5 million deaths per year (WHO 2021). A better prediction, prevention and targeted treatment of diabetes and its complications may decrease mortality rates and reduce the burden of this disease. Diabetes is mainly classified into type 1 diabetes and type 2 diabetes (T2D). With this traditional classification, T2D includes >85% of all diabetic patients. T2D is, however, a complex and heterogeneous disease, influenced by genetic, epigenetic and environmental factors, and characterized by several pathological conditions, including insulin resistance, beta-cell dysfunction, and elevated hepatic glucose production (1). Classifying T2D as one group has been shown to be insufficient to adequately treat diabetes and predict related complications (2). New reclassifications of T2D were therefore carried out in the All New Diabetics in Scania (ANDIS) cohort and in several other cohorts (3-8). In the original paper, four different subgroups of T2D were identified based on unsupervised data-driven clustering analysis of six phenotypes: age at onset of diabetes, BMI, HbA1c at diagnosis, homeostasis model assessment 2B (HOMA2-B, measure of beta-cell function), HOMA2-IR (measure of insulin resistance) and glutamic acid decarboxylase autoantibodies (GADA). These novel subgroups were labelled as: Severe Insulin Deficient Diabetes (SIDD), Severe Insulin Resistant Diabetes (SIRD), Mild Obesity-related Diabetes (MOD), and Mild Age-Related Diabetes (MARD) (3). The subgroups have different patient characteristics and risk of diabetic complications (3). Differences in their genetic, metabolomic, and proteomic signatures further support that diverse aetiologies exist between the subgroups (9; 10). This reclassification may hence provide a better basis for understanding differences in T2D patients, representing an important step towards precision medicine in diabetes.

Our group and others have found epigenetic differences in tissues from patients with T2D versus controls (11-15), demonstrating that epigenetic mechanisms contribute to the pathogenesis of T2D. Moreover, there has been an increasing interest in identifying blood-based epigenetic biomarkers for risk assessment in patients with diabetes. For example, DNA methylation in blood was associated with future T2D, insulin secretion and response to therapy (11; 16-18). However, it remains unknown if the epigenetic patterns differ between the novel subgroups of T2D and whether these epigenetic differences may predict complications in newly diagnosed patients with diabetes.

Therefore, we analysed the methylome in blood of newly diagnosed T2D patients from the prospective ANDIS cohort and our first goal was to investigate whether DNA methylation differs between the four T2D subgroups identified by Ahlqvist *et al.* (3) and to find "subgroup-unique" methylation sites, i.e. sites that show different methylation levels in one subgroup compared with each of the other subgroups (**Figure S1**). We then tested whether combined subgroup-unique methylation risk scores (MRSs) generated from identified top subgroup-unique methylation sites in the *ANDIS discovery cohort* **i)** associated with T2D subgroups in replication cohorts and **ii)** associated with future diabetic complications such as cardiovascular disease (CVD), chronic kidney disease (CKD), and retinopathy (**Figure S1**).

Methods

Study populations

ANDIS is an ongoing prospective study of newly-diagnosed diabetes patients which aims to document all new incidences of diabetes within the Scania region in Southern Sweden (http://andis.ludc.med.lu.se) (3; 9; 18). Blood samples for DNA extraction are taken at registration, i.e. within one year from diagnosis of diabetes. The ANDIS protocol was approved by the Regional Ethical Review Board in Lund (584/2006, 2011/354, 2014/198).

All New Diabetics in Uppsala County (ANDiU) is a similar study to ANDIS but includes newly-diagnosed diabetes patients living in the Uppsala region (http://www.andiu.se/) (18). The ANDiU study protocol was approved by the regional ethics review committee in Uppsala (2011/155).

Discovery and replication cohorts

To study the association between DNA methylation in blood and the recently defined novel subgroups of T2D (3), we included 280 T2D patients from ANDIS who were previously assigned to the four novel T2D subgroups based on unsupervised clustering (3), and who had available DNA from blood at diagnosis in the *ANDIS discovery cohort*. We selected these 280 patients while blind to their subgroup information based on a power calculation from a previous study showing 80% power to detect differences in DNA methylation of 4,000 sites with false discovery rate (FDR)<5% (12).

We are only aware of a few cohorts other than ANDIS (e.g. ANDiU), with available blood samples at T2D diagnosis and available phenotypes required for clustering of subgroups in newly-diagnosed subjects with T2D. We therefore selected one replication cohort from ANDIS, the *ANDIS replication cohort*, including 76 additional T2D patients and one replication cohort from ANDiU, the *ANDiU replication cohort* including 197 T2D patients.

Clinical characteristics of these cohorts are shown in **Table S1** and **Figure 1**. The inclusion criteria and the flowchart of the selection of patients in these cohorts is displayed in **Figure S2**.

Phenotype measurements

Age and HbA1c were considered at diagnosis, and BMI, HOMA2-B, and HOMA2-IR were measured at registration in ANDIS and ANDiU. Standard protocols were applied for measuring weight and height to calculate BMI (kg/m^2). HbA1c was measured using the Variant II Turbo HbA1c Kit 2.0 (Bio-Rad Laboratories, Copenhagen, Denmark) (3). C-peptide concentrations for HOMA2 were measured using an electro-chemiluminescence immunoassay on Cobas e411 (Roche Diagnostics, Mannheim, Germany) or a radioimmunoassay (Human C-peptide RIA; Lincom St Charles, MO; or Peninsula Laboratories, CA, USA) and used with the HOMA calculator (19).

Diabetic complications

CVD was defined as having had either coronary events (defined by ICD-10 codes I20-I21, I24, I251 and I253-I259) or stroke (defined by ICD-10 codes I60, I61, I63 and I64). CKD was defined as a minimum of two measurements of estimated glomerular filtration rate (eGFR) below $60ml/min/1.73m^2$ for more than 90 days or a single measurement of eGFR below $15ml/min/1.73m^2$ (kidney failure) (3). eGFR was calculated with the Modification of Diet in Renal Disease (MDRD-4) Study equation (20). Diagnosis of diabetic retinopathy was based on ICD-10 codes E113 and H36.0. Patients with complications before DNA methylation samples were excluded for the respective analyses. Analyses related to complications were done in combined *ANDIS and ANDiU cohorts*, to improve statistical power due to the modest number of cases with complications in each individual cohort. Phenotypes of individuals with each complication are presented in **Tables S2-S4**.

Genome-wide DNA methylation analysis

Participants' whole blood samples were taken at ANDIS and ANDiU registration and DNA was then extracted using the Gentra Puregene Blood kit (Qiagen, Hilden, Germany). DNA methylation analysis of the *ANDIS discovery* and *replication cohorts* was carried out at two different time points and places; the SCIBLU genomics centre and at Lund University Diabetes Centre (LUDC), respectively. ANDiU samples were analysed at LUDC. 500-1000ng of genomic DNA was bisulphite treated using the EZ DNA methylation kit (Zymo Research, CA, USA). DNA methylation was analysed for all participants using Illumina MethylationEPIC BeadChip microarrays (Illumina, Inc., CA, USA) according to the Infinium HD assay methylation protocol. Detailed information about quality control and bioinformatic analyses of the genome-wide DNA methylation data is available in **Figure S3** and elsewhere (18).

DNA methylation in other tissues

To test whether DNA methylation of sites included in the subgroup-unique MRSs in blood mirror DNA methylation levels in other tissues, we used Illumina 450K array DNA methylation data from blood, skeletal muscle and adipose tissue taken from the same subject in *the Monozygotic Twin cohort* (13). Here, methylation data were extracted if methylation sites in MRSs were also covered by the 450K array. Twins with available methylation data were included (**Table S5**). Characteristics of the full twin cohort used for these analyses and additional information has been previously published (13; 17).

Statistical Analysis

Statistical analyses were performed using R software with the lowest p-value provided being $2x\ 10^{-16}$. Clinical patient data are presented as means (standard deviation (SD)) or percentages. Differences between the four subgroups regarding continuous clinical variables were analysed with a Kruskal-Wallis one-way analysis of variance and a Dunn's post hoc analysis corrected for multiple testing using Benjamini-Hochberg procedure. Differences in categorical variables were assessed using a Pearson's Chi-squared test.

To find differences in DNA methylation between the four T2D subgroups, we first performed an analysis of covariance (ANCOVA) adjusting for sex in *the ANDIS discovery cohort* (**Figure S1**). Here, a Benjamini-Hochberg was applied to correct for multiple testing and methylation sites with FDR below 5% ($q<0.05$) were included in further analyses. The X-chromosome was then removed to mitigate the effect of sex on DNA methylation data. Pairwise comparisons

were then used to identify “subgroup-unique methylation sites”, which were defined as sites with differences in methylation levels in one subgroup compared with the methylation levels in all other subgroups based on q<0.05. Here, we did six Bonferroni corrected pairwise comparisons using linear regression models adjusted for sex for each of the sites with q<0.05 in the ANCOVA.

To integrate epigenetic information across the identified subgroup-unique methylation sites, we calculated a weighted MRS for each subgroup, so called “subgroup-unique MRS”. To calculate these weighted MRSs, the sum of the standardised methylation values for each of the included subgroup-unique sites was multiplied by the ß-coefficient for the respective site (18; 21) as follows:

$$\text{MRS} = \frac{\text{methylation beta-value}_{\text{replication cohort}} - \text{mean methylation}_{\text{ANDIS discovery cohort}}}{\text{SD methylation}_{\text{ANDIS discovery cohort}}} \times \text{ß-coefficient}_{\text{ANDIS discovery cohort}}$$

These ß-coefficients were obtained from sex-adjusted linear regression models for each of the included subgroup-unique sites in the *ANDIS discovery cohort*. Here, the methylation value for each site was the dependent variable, while the respective subgroup versus the combined group of all other subgroups was the binary independent variable. To select the best combination of subgroup-unique methylation sites to be included in the MRSs, the subgroup-unique sites were: **i**) first, rank-ordered based on their significance using q-values in *the ANDIS discovery cohort*; **ii**) second, subgroup-unique sites included in each subgroup-unique MRS were then selected starting with the highest rank (lowest q-value), and going down in rank until the best possible combination of sites were included in the MRS based on its ability to discriminate between subjects with a particular diabetes subgroup and those without, performing separate analyses in the *ANDIS discovery cohort* and the *ANDIS replication cohort*. The ability to discriminate subjects was based on the best area under the curve (AUC) using C-statistics in both *ANDIS* cohorts (**Table S6**). Subsequently, four different subgroup-unique MRSs (i.e. SIDD-MRS, SIRD-MRS, MOD-MRS, MARD-MRS) were generated for each person independently of which subgroup they belonged to. MRSs were adjusted for cell composition using a reference-based method (22). MRSs were generated in the same way in the independent *ANDiU replication cohort* and their ability to discriminate between subjects with a particular diabetes subgroup and those without was examined using linear and logistic regression in crude models and when adjusting for the clinical variables defining the subgroups.

To evaluate whether the subgroup-unique MRSs associate with future diabetic complications, sex-adjusted weighted-cox regression models were applied. Hazard ratios (HR) are presented with 95% CI and $p<0.05$ was considered statistically significant. A statistical power of 85% (α-level=0.05) was achieved with a sample size of 500 (probability of events: 0.15), and assuming a HR of 0.5 or 2 and a standard deviation of 0.5. Regression models were not adjusted for age at onset, BMI, HbA1c, HOMA2-B or HOMA2-IR due to multi-collinearity with the subgroup-unique MRSs.

To examine the correlation between methylation in blood and methylation in other tissues of sites included in subgroup-unique MRSs, Pearson correlation tests were performed. Benjamini-Hochberg was used to correct for multiple testing and $q<0.05$ was considered significant.

Results

Different DNA methylation patterns in T2D subgroups

We explored whether DNA methylation in blood is associated with the four novel T2D subgroups (SIDD, SIRD, MOD and MARD) using the *ANDIS discovery cohort*, the *ANDIS replication cohort*, and the *ANDiU replication cohort*. These cohorts include newly diagnosed subjects with T2D who had previously been assigned to a subgroup using data-driven clustering and who have DNA methylation data available. In line with our previous study (3), significant differences in age at onset of diabetes, BMI, HbA1c, HOMA2-B and HOMA2-IR were found between the four T2D subgroups in all three cohorts. SIDD subjects had higher HbA1c levels, SIRD subjects had higher HOMA2-IR and HOMA2-B, MOD subjects had higher BMI and lower age, whereas MARD subjects were the oldest of all subgroups (**Figure 1** and **Table S1**).

Figure S1 presents our study design. First, we assessed if DNA methylation of any individual sites associated with the T2D subgroups in the *ANDIS discovery cohort*. Here, 22,034 sites showed differences in methylation between any of the four subgroups based on an ANCOVA adjusted for sex (FDR<5%, $q<0.05$, **Table S7**). We then performed post hoc pairwise comparisons to identify "subgroup-unique" sites among these sites, i.e. sites that showed different methylation levels in one subgroup compared with the level in each of the other subgroups in the *ANDIS discovery cohort*. We identified 4,465 subgroup-unique methylation

sites, including 56 sites unique for SIDD, 74 sites unique for SIRD, 4,135 sites unique for MOD and 200 sites unique for MARD in the *ANDIS discovery cohort* (**Table S8**).

Next, we selected top ranked subgroup-unique methylation sites from **Table S8** to build subgroup-unique MRSs (SIDD-MRS, SIRD-MRS, MOD-MRS, MARD-MRS) that best discriminate T2D subgroups in not only the *ANDIS discovery cohort*, but also in the *ANDIS replication cohort*. Based on this, we included 54 SIDD-unique sites to generate SIDD-MRS, 2 SIRD-unique sites for SIRD-MRS, 31 MOD-unique sites for MOD-MRS, and 8 MARD-unique sites for MARD-MRS (**Tables S9**). The subgroup-unique MRSs were significantly different in patients allocated to each respective subgroup compared to the combined group of all other subgroups in both *the ANDIS discovery* and *ANDIS replication cohorts* (**Figure 2A-B**). The MRSs did also differ between the four subgroups, showing a higher SIDD-MRS in SIDD individuals versus all other subgroups, a higher SIRD-MRS in SIRD individuals versus all other subgroups, a higher MOD-MRS in MOD individuals versus all other subgroups, and a higher MARD-MRS in MARD individuals versus all other subgroups (**Figures S4A-B**). Moreover, these subgroup-unique MRSs remained significant after adjusting for cell composition (**Table S10**). We then visualized the methylation level of the sites included in the subgroup-unique MRSs in relation to all sites analysed with the MethylationEPIC BeadChip. **Figure S5** shows that a large proportion of SIDD-MRS sites are hypermethylated whereas overall SIRD-MRS sites, MOD-MRS sites and MARD-MRS sites have intermediate levels of methylation.

We proceeded to validate if these subgroup-unique MRSs could discriminate between the four subgroups in the independent *ANDiU replication cohort*. Importantly, the respective MRSs were statistically significantly different between one subgroup and the combined group of all other subgroups (**Figure 2C**) and also differed between the four subgroups (SIDD-MRS p= 3.8e-04, SIRD-MRS p=5.4e-04, MOD-MRS p=9.8e-14, MARD-MRS p=1.5e-09) in a similar pattern observed in the ANDIS cohorts (**Figure S4C**). When a single subgroup corresponding to the respective MRS was taken as the reference category (i.e. SIDD for SIDD-MRS, SIRD for SIRD-MRS, MOD for MOD-MRS, MARD for MARD-MRS), all other subgroups had a significantly lower MRS in sex-adjusted models which shows the ability of each MRS to differentiate the reference subgroup from the others in an independent replication cohort (**Table 1**). The only one that did not reach significance was the comparison between SIDD

versus SIRD for the SIRD-MRS, although SIDD shows a lower MRS than SIRD. Moreover, higher values of all MRSs were associated with a higher probability of clustering to a particular subgroup, i.e. higher SIDD-MRS was associated with SIDD (OR=2.08, p=2e-04), higher SIRD-MRS was associated with SIRD (OR=1.61, p=0.011), higher MOD-MRS was associated with MOD (OR=6.06, p=1.85e-09) and higher MARD-MRS was associated with MARD (OR=2.52, p=6.72e-07) (**Figure 2D**). These associations remained with similar effect sizes after adjusting for the primary variable defining each respective subgroup, except for SIDD-MRS where the effect was lost after adjusting for HbA1c (**Table S11**). Overall, these results show that the four novel subgroups of T2D display different epigenetic patterns in discovery and validation cohorts.

Subgroup-unique methylation risk scores associate with future diabetic complications

It was previously found that patients allocated to certain T2D subgroups show higher risk of developing diabetic complications including CKD or diabetic retinopathy (3). We therefore tested whether subgroup-unique MRSs were associated with future complications. To increase statistical power and due to the modest number of patients that develop complications during follow-up, discovery and replication cohorts were combined for complication-related analyses. Their characteristics are presented in **Tables S2-S4**. Sex-adjusted weighted-cox regression models showed that all MRSs were associated with development of future CVD, all but SIDD-MRS were associated with future CKD and none of them were associated with future retinopathy during a mean follow-up of 4.5 years (**Figure 2E**). Higher SIDD-MRS (HR=0.72, p=0.032) and MOD-MRS (HR=0.65, p=0.001) were associated with lower risk of developing CVD, whereas higher SIRD-MRS (HR=1.47, p=0.002) and MARD-MRS (HR=1.41, p=0.007) were associated with higher risk of future CVD. Regarding CKD, higher MOD-MRS (HR=0.50, p=3.11e-07) was associated with lower risk, whereas higher SIRD-MRS (HR=1.55, p=0.007) and MARD-MRS (HR=1.90, p=1.72e-06) were associated with higher risk of developing renal disease. These associations remain significant after further adjustment for blood cell types, except for SIRD-MRS and the risk for CKD (p=0.09) (**Table S12**). Due to multi-collinearity, these associations could not be adjusted by subgroup-defining phenotypes. However, all these phenotypes, except age and HOMA2-B, did not associate with future vascular and kidney complications (**Table S13**). We also found associations between some T2D subgroups and complications in the present study (**Table S14**). However, none of the

subgroup-unique MRS, subgroup-defining phenotypes or T2D subgroups generated area under the curves (AUC) above 0.75 (**Table S15**).

Biological function of subgroup-unique sites

To better understand the biological function of the 72 genes annotated to the 95 differentially methylated sites included in any of the subgroup-unique MRSs, we performed a systematic literature search using each gene symbol and the following terms: *diabetes, insulin secretion/beta-cell function, insulin resistance, obesity, age* and *nonalcoholic fatty liver disease (NAFLD)*. Any study with any of these search terms and any of the 72 gene symbols was considered. This showed that 39 (54%) of the 72 genes have been associated with diabetes and/or with some characteristics defining the subgroups or NAFLD (**Table S16**). Furthermore, among the 72 genes, 26 genes (36%) have been associated with diabetes (e.g. *LMNB2, NBPF20, RREB1, IFIH1)*, 15 (21%) with insulin secretion/beta-cell function (e.g. *TXNIP, TFEB*), 9 (12%) with insulin resistance (e.g. *GRK5, SOD3*), 20 (28%) with obesity (e.g. *GSN, MOGAT1, RREB1, STK3, SLC6A4*), 16 (22%) with age (e.g. *ELOVL2, SOD3, TRIM59, TFEB*), and 11 (15%) with NAFLD, highlighting the relevance of identified methylation sites for stratification of the subgroups (**Figure 2F**). When looking at individual subgroup-unique MRSs, 23 (52%) of 44 genes included in SIDD-MRS, both genes (100%) included in SIRD-MRS, 12 (57%) of 21 genes included in MOD-MRS and 2 (40%) of 5 genes included in MOD-MRS were associated with any of the terms representing the subgroup-traits (**Figure 2G, Table S16**). Genes annotated to SIDD-MRS sites include *AATK, CPLX1, CTDSPL, GRK5, LMNB2, RREB1, SMARCA4, SOD3, SYT2,* and *TXNIP*, which play a role in insulin secretion/beta-cell function; genes annotated to SIRD-MRS sites include *RAB27B* and *RBL2,* previously associated with diabetes; genes annotated to MOD-MRS sites include *ELOVL2, PDGFC, SCN9A, SLC6A4*, and *TFEB,* previously associated with obesity; and genes annotated to MARD-MRS sites include *CRMP1* and *RNF170,* previously associated with age (**Table S16**).

Cross-tissue methylation in different tissues

Finally, we examined whether blood-based DNA methylation of the 95 sites included in any of the subgroup-unique MRSs mirror methylation in other central tissues for T2D. Methylation of these sites was compared between blood, skeletal muscle, and adipose tissue. 450K methylation array data from blood, muscle and adipose tissue of the Monozygotic Twin Cohort was used since there we had access to data from the same individuals for these cell types (**Table S5**) (13). Methylation data for 57 out of 95 sites were available in the 450K array. Among

these, blood methylation correlated positively with methylation of 18 sites in adipose tissue and 3 sites in muscle after correcting for multiple testing (**Table 2**), including cg14013597 and sites annotated to *AATK*, *CRMP1*, *ELOVL2*, *KCNQ2*, *MOGAT1*, *PGAM2*, and *SLC6A4* suggesting that methylation of some subgroup-unique sites may play a role in relevant tissues for T2D (**Figure S6**).

Discussion

We demonstrate for the first time that there are epigenetic differences between the novel T2D subgroups: SIDD, SIRD, MOD and MARD, already at diagnosis. Importantly, epigenetic markers differed between the four subgroups in an independent validation cohort, further establishing a clear heterogeneity of these T2D subgroups. The subgroup-unique epigenetic markers did also associate with future diabetic complications, supporting development of blood-based epigenetic biomarkers for precision medicine of diabetes.

T2D is a heterogeneous disease with individual variation in obesity, insulin resistance, insulin secretory defects and/or age between different patients. There is a need to improve treatment strategies for T2D and to better identify individuals with increased risk for complications. We recently reclassified T2D into four novel subgroups based on age at onset of diabetes, BMI, HbA1c at diagnosis, HOMA2-B, HOMA2-IR and GADA (3). We also demonstrated that genetic risk scores for diabetes-related traits associate with these subgroups (9). Interestingly, here, we identified subgroup-unique epigenetic modifications in newly diagnosed T2D patients, further supporting different underlying etiopathological processes for each subgroup. The fact that ~50% of genes annotated to the subgroup-unique sites included in the MRSs have been previously associated with diabetes, insulin secretion, insulin resistance, obesity and/or age, suggests that these epigenetic modifications have important functions in the pathogenesis of diabetes as well. This is further supported by methylation in blood of some subgroup-unique sites mirroring the methylation pattern in adipose tissue and muscle, two relevant tissues for T2D. One needs to consider that these are "*only*" correlations and based on 450k arrays which miss methylation data for several sites included in the subgroup-unique MRS. Future studies should further examine if there are epigenetic differences between subgroups also in adipose tissue, muscle, liver, and islets. Nevertheless, our data point to adipose tissue being important, as 18 subgroup-unique methylation sites showed positive correlations between blood and adipose tissue and interestingly 12 of these are SIDD-unique sites, while 5 sites are MOD-

unique sites, suggesting a potential role for methylation in adipose tissue in SIDD- and MOD-specific pathogenesis.

To better understand the biology of our findings, we performed a systematic literature search using each gene symbol annotated to the differentially methylated sites included in the subgroup-unique MRSs, and diabetes, insulin secretion/beta-cell function, insulin resistance, obesity, or age. For example, SIDD is characterised by beta-cell dysfunction, poor metabolic control and higher HbA1c. We found two SIDD-MRS sites annotated to *TXNIP*, encoding a thioredoxin-binding protein that protects against oxidative stress, regulates beta-cell function and *TXNIP* methylation is associates with HbA1c and incident T2D and is altered in muscle and islets from individuals with T2D (12; 17; 23; 24). Moreover, methylation of *CTDSPL*, a gene annotated to SIDD-MRS sites, was associated with incident T2D and islet dysfunction (25). Interestingly, we found that higher SIRD-MRS was associated with increased risk of CKD with certain SIRD-MRS sites annotated to *RAB27B* whose downregulation contributes to exosome dysfunction in diabetic kidney disease (26). MOD-MRS sites were annotated to genes previously associated with obesity, including *SLC6A4*, *TFEB* and *ELOVL2* (27-29). *SLC6A4* encodes a serotonin transporter regulating energy balance and methylation of *SLC6A4* has been linked to obesity (27). The transcription factor encoded by *TFEB* regulates lysosomal biogenesis and autophagy and is upregulated in adipose tissue during obesity (28), while *ELOVL2* regulates synthesis of very long polyunsaturated fatty acids (29). MOD is characterized by higher BMI, but also by younger age, and notably genes annotated to MOD-MRS sites are associated with ageing, including *ELOVL2, IFIH1, NAV2, TFEB* and *TRIM59* (30-34). Two genes annotated to MARD-MRS sites have been linked to aging, including *CRMP1,* whose expression decreased with age (35), and *RNF170* associated with age-dependent gait abnormalities (36). These results clearly show that the novel sites included in subgroup-unique MRS identified in this study are linked to diabetes and subgroup-defining phenotypes.

T2D patients have higher risk of CVD, CKD, and/or retinopathy (37). Nevertheless, T2D subgroups have different risks of these complications, confirming variability in T2D patients and justifying reclassification of diabetes (3). CVD is responsible for the majority of deaths among patients with diabetes, while CKD is the leading cause of hospitalization in individuals with diabetes (37), reflecting the need for reliable prediction tools in this at-risk population.

Interestingly, the identified subgroup-unique MRSs were associated with future risks of developing diabetes complications. Higher SIRD-MRS and MARD-MRS were associated with an increased risk of developing CVD and CKD, whereas a higher MOD-MRS was associated with lower risk for developing these two complications. Our results are supported by previous studies, where SIRD patients had higher risk of CKD and MOD patients had lower risk of CKD and coronary events (3; 10). In the present study, we also found that the associations between MOD and MARD patients and the risk for CVD and CKD were similar to the associations observed for MOD- and MARD-MRSs. On the other hand, SIRD-MRS and SIDD-MRS were able to predict CVD and/or CKD, whereas SIRD and SIDD patients were not associated with the risk of these diseases, suggesting a potentially better ability of these MRS on predicting complications compared to the T2D subgroups. Of note, our study compared one subgroup versus all remaining patients, while MARD has previously been used as a reference group (3). Overall, a new classification of T2D, now supported by epigenetic markers, could identify patients at high risk of developing complications already at diagnosis. This tool may be further developed to decrease suffering for patients and costs for the society.

This study has some limitations. The first discovery step identified a big number of methylation sites which may resulted in an inflated type I error rate, also supported by a lambda of 1.7. It should however be noted that lambda is not suitable to measure inflation in epigenome-wide associations studies (38). Nevertheless, we mitigated this potential issue by studying MRSs in three different cohorts including external validation, thus reducing possible bias and showing the reliability of the methylation sites included in the MRSs. At this point, we cannot fully conclude that the subgroup-unique sites presented here are the optimal combination of sites to differentiate between the individual subgroups and/or predict complications. Nevertheless, these MRSs seem to be robust since they remain significantly different between subgroups after adjusting for confounding factors, such as cell composition, sex and clinical variables defining the subgroups and could discriminate between the subgroups in an independent validation cohort. Moreover, taking one subgroup at a time as a reference group showed that the MRSs robustly differ between that specific subgroup from all the other subgroups, demonstrating the capability of the MRSs at characterizing each respective subgroup.

Importantly, all subgroup-unique MRSs, except for SIDD-MRS, were strongly associated with their respective subgroup, independently of adjustment for the corresponding subgroup-defining variable. SIDD-MRS was associated with a 2-fold higher probability of being SIDD,

but this result was partially driven by HbA1c which is the identifying variable of this subgroup. In contrast, similar effect sizes were observed for association between MARD-MRS and MARD after adjusting for age, and between SIRD-MRS and SIRD after adjusting for HOMA-IR. Notably, the effect size was even greater when the clinical variable was included in the model for association between SIRD-MRS and SIRD after adjustment for HOMA-B and association between MOD-MRS and MOD after adjusting for BMI. This suggests that our epigenetic markers are equally robust at discriminating these subgroups as the identifying trait of the subgroup, further supporting that epigenetics play a key role in the aetiology of these subgroups.

Due to multi-collinearity, analyses for complications were not adjusted for the subgroup-defining phenotypes. Therefore, we cannot rule out that the associations between MRSs and complications are influenced by these phenotypes. Hence, these associations may be both due to altered DNA methylation and clinical phenotypes, a conundrum difficult to disentangle. However, only age and HOMA2-B were associated with future CVD and CKD, suggesting a minor effect for the rest of subgroup-defining phenotypes. Nevertheless, age could have influenced some of the associations between MRSs and diabetic complications, especially for MOD-MRS and SIDD-MRS, associated with a low risk for CVD and/or CKD. MOD and SIDD were younger individuals and therefore less likely of developing events after 8 years of follow-up. Certain risk factors for diabetic complications (e.g. smoking, LDL, albumin-creatinine ratio) were completely or partially missing at baseline for patients included in this study, and we could therefore not assess whether MRSs affected the predictive value of such factors. Limited sample size and the short period of follow-up may affect associations with the studied complications. Moreover, future epigenetic studies focusing only on diabetic complications may identify additional markers to be included in scores for prediction of CVD, CKD and retinopathy, respectively. However, such analyses are not suitable in this study since they are unlikely to support reclassification of diabetes, the main goal of this study. Since this study was based on predominantly northern European subjects, the applicability to other ethnicities needs to be evaluated to establish the generalisability of associations between the identified methylation patterns and SIDD, SIRD, MOD and MARD.

Conclusion

A reclassification of T2D might help tailor prevention strategies to individual subgroups and personalise care for those affected by diabetes. The observed associations between DNA

methylation of certain sites and the novel T2D subgroups support the hypothesis that subgroups have epigenetic differences. Our study also suggests that epigenetic mechanisms may be more important for some of the subgroups, which should be further dissected in future studies. Identified epigenetic markers could successfully discriminate between diabetes subgroups and associated with future incidence of CVD and CKD, thus further validating the reclassification of T2D on an epigenetic level. These epigenetic markers may be developed for precision medicine to improve treatment of T2D subgroups and prevent their complications.

Article information

Acknowledgements

We thank Ylva Wessman, Per-Anders Jansson, Pernille Paulsen, Maria Sterner, Emma Nilsson and the SCIBLU genomics facility at Lund University.

We declare no competing interests.

Author Contributors

SS, EA, SG and CL contributed to the conception of the work. SS, AV, LG, EA, MM, SG and CL contributed to the data collection. SS, AP, LG, EA and SG contributed to the data analysis. SS, SG and CL drafted the article. All authors contributed to the interpretation of data and critical revision of the article. All authors gave final approval of the version to be published.

SG and CL are the guarantors of this manuscript.

Duality of Interest

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

Funding

The work was supported by grants from the EFSD, Swedish Heart Lung Foundation, the Novo Nordisk foundation, The Swedish Research Council, Region Skåne (ALF), ERC-Co Grant (PAINTBOX, No 725840), H2020-Marie Skłodowska-Curie grant agreement No 706081 (EpiHope), Exodiab, Swedish Foundation for Strategic Research for IRC15-0067 and Swedish Diabetes Foundation.

The funders had no role in design and conduct of the study, collection, management, analysis, and interpretation of the data, preparation, review, and approval of the manuscript, or the decision to submit the manuscript for publication.

References

1. Ling C, Ronn T. Epigenetics in Human Obesity and Type 2 Diabetes. *Cell Metab* 2019;29:1028-1044

2. Gloyn AL, Drucker DJ. Precision medicine in the management of type 2 diabetes. *Lancet Diabetes Endocrinol* 2018;6:891-900

3. Ahlqvist E, Storm P, Karajamaki 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

4. Dennis JM, Shields BM, Henley WE, Jones AG, Hattersley AT. Disease progression and treatment response in data-driven subgroups of type 2 diabetes compared with models based on simple clinical features: an analysis using clinical trial data. *Lancet Diabetes Endocrinol* 2019;7:442-451

5. Zaharia OP, Strassburger K, Strom A, et al. 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

6. Zou X, Zhou X, Zhu Z, Ji L. Novel subgroups of patients with adult-onset diabetes in Chinese and US populations. *The Lancet Diabetes & Endocrinology* 2019;7:9-11

7. Pigeyre M, Hess S, Gomez MF, et al. Validation of the classification for type 2 diabetes into five subgroups: a report from the ORIGIN trial. *Diabetologia* 2021

8. Slieker RC, Donnelly LA, Fitipaldi H, et al. Replication and cross-validation of type 2 diabetes subtypes based on clinical variables: an IMI-RHAPSODY study. *Diabetologia* 2021;64:1982-1989

9. Mansour Aly D, Dwivedi OP, Prasad RB, et al. Genome-wide association analyses highlight etiological differences underlying newly defined subtypes of diabetes. *Nat Genet* 2021;53:1534-1542

10. Slieker RC, Donnelly LA, Fitipaldi H, et al. Distinct Molecular Signatures of Clinical Clusters in People with Type 2 Diabetes: an IMIRHAPSODY Study. *Diabetes* 2021:db201281

11. Chambers JC, Loh M, Lehne B, et al. Epigenome-wide association of DNA methylation markers in peripheral blood from Indian Asians and Europeans with incident type 2 diabetes: a nested case-control study. *Lancet Diabetes Endocrinol* 2015;3:526-534

12. Dayeh T, Volkov P, Salo S, et al. Genome-wide DNA methylation analysis of human pancreatic islets from type 2 diabetic and non-diabetic donors identifies candidate genes that influence insulin secretion. *PLoS Genet* 2014;10:e1004160

13. Nilsson E, Jansson PA, Perfilyev A, et al. Altered DNA methylation and differential expression of genes influencing metabolism and inflammation in adipose tissue from subjects with type 2 diabetes. *Diabetes* 2014;63:2962-2976

14. Nilsson E, Matte A, Perfilyev A, et al. Epigenetic Alterations in Human Liver From Subjects With Type 2 Diabetes in Parallel With Reduced Folate Levels. *J Clin Endocrinol Metab* 2015;100:E1491-1501

15. Volkov P, Bacos K, Ofori JK, et al. Whole-Genome Bisulfite Sequencing of Human Pancreatic Islets Reveals Novel Differentially Methylated Regions in Type 2 Diabetes Pathogenesis. *Diabetes* 2017;66:1074-1085

16. Bacos K, Gillberg L, Volkov P, et al. Blood-based biomarkers of age-associated epigenetic changes in human islets associate with insulin secretion and diabetes. *Nat Commun* 2016;7:11089

17. Dayeh T, Tuomi T, Almgren P, et al. DNA methylation of loci within ABCG1 and PHOSPHO1 in blood DNA is associated with future type 2 diabetes risk. *Epigenetics* 2016;11:482-488

18. Garcia-Calzon S, Perfilyev A, Martinell M, et al. Epigenetic markers associated with metformin response and intolerance in drug-naive patients with type 2 diabetes. *Sci Transl Med* 2020;12

19. Levy JC, Matthews DR, Hermans MP. Correct homeostasis model assessment (HOMA) evaluation uses the computer program. *Diabetes Care* 1998;21:2191-2192

20. Levey AS, Coresh J, Greene T, et al. Using standardized serum creatinine values in the modification of diet in renal disease study equation for estimating glomerular filtration rate. *Ann Intern Med* 2006;145:247-254

21. Wahl S, Drong A, Lehne B, et al. Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. *Nature* 2017;541:81-86

22. Salas LA, Koestler DC, Butler RA, et al. An optimized library for reference-based deconvolution of whole-blood biospecimens assayed using the Illumina HumanMethylationEPIC BeadArray. *Genome Biol* 2018;19:64

23. Soriano-Tarraga C, Jimenez-Conde J, Giralt-Steinhauer E, et al. Epigenome-wide association study identifies TXNIP gene associated with type 2 diabetes mellitus and sustained hyperglycemia. *Hum Mol Genet* 2016;25:609-619

24. Shao W, Liu W, Liang P, et al. GABA requires GLP-1R to exert its pancreatic function during STZ challenge. *J Endocrinol* 2020;246:207-222

25. Ouni M, Saussenthaler S, Eichelmann F, et al. Epigenetic Changes in Islets of Langerhans Preceding the Onset of Diabetes. *Diabetes* 2020;69:2503-2517

26. Zeng M, Wen J, Ma Z, et al. FOXO1-Mediated Downregulation of RAB27B Leads to Decreased Exosome Secretion in Diabetic Kidneys. *Diabetes* 2021;70:1536-1548

27. Lillycrop KA, Garratt ES, Titcombe P, et al. Differential SLC6A4 methylation: a predictive epigenetic marker of adiposity from birth to adulthood. *Int J Obes (Lond)* 2019;43:974-988

28. Kim J, Kim SH, Kang H, et al. TFEB-GDF15 axis protects against obesity and insulin resistance as a lysosomal stress response. *Nat Metab* 2021;3:410-427

29. Maguolo A, Zusi C, Giontella A, et al. Influence of genetic variants in FADS2 and ELOVL2 genes on BMI and PUFAs homeostasis in children and adolescents with obesity. *Int J Obes (Lond)* 2021;45:56-65

30. Ronn T, Volkov P, Gillberg L, et al. Impact of age, BMI and HbA1c levels on the genome-wide DNA methylation and mRNA expression patterns in human adipose tissue and identification of epigenetic biomarkers in blood. *Hum Mol Genet* 2015;24:3792-3813

31. Lyu W, Zhou Y, Zhuang Y, et al. Surfactant protein D is associated with 3-month mortality of anti-MDA5 antibody-interstitial lung disease. *Clin Exp Rheumatol* 2020;38:1068-1074

32. Bae H, Lunetta KL, Murabito JM, et al. Genetic associations with age of menopause in familial longevity. *Menopause* 2019;26:1204-1212

33. Wang C, Niederstrasser H, Douglas PM, et al. Small-molecule TFEB pathway agonists that ameliorate metabolic syndrome in mice and extend C. elegans lifespan. *Nat Commun* 2017;8:2270

34. Dias HC, Cordeiro C, Pereira J, et al. DNA methylation age estimation in blood samples of living and deceased individuals using a multiplex SNaPshot assay. *Forensic Sci Int* 2020;311:110267

35. Miyazaki T, T TB, Mori M, Komori T. Collapsin Response Mediator Protein 1, a Novel Marker Protein for Differentiated Odontoblasts. *Acta Histochem Cytochem* 2018;51:185-190

36. Kim Y, Kim SH, Kim KH, et al. Age-dependent gait abnormalities in mice lacking the Rnf170 gene linked to human autosomal-dominant sensory ataxia. *Hum Mol Genet* 2015;24:7196-7206

37. Zheng Y, Ley SH, Hu FB. Global aetiology and epidemiology of type 2 diabetes mellitus and its complications. *Nat Rev Endocrinol* 2018;14:88-98

38. van Iterson M, van Zwet EW, Consortium B, Heijmans BT. Controlling bias and inflation in epigenome- and transcriptome-wide association studies using the empirical null distribution. *Genome Biol* 2017;18:19

TABLES

Table 1. Associations between the four subgroups and subgroup-unique methylation risk scores (MRSs) in the ANDiU replication cohort.

	ANDiU replication cohort	
	Adjusted for sex	
	B-coeff (SE)	p
SIDD-MRS		
SIDD	0 (Ref.)	
SIRD	-0.12 (0.05)	0.021
MOD	-0.13 (0.05)	0.008
MARD	-0.19 (0.04)	2.0 e-05
SIRD-MRS		
SIDD	-0.02 (0.01)	0.168
SIRD	0 (Ref.)	
MOD	-0.05 (0.01)	3.8 e-05
MARD	-0.02 (0.01)	0.037
MOD-MRS		
SIDD	-0.79 (0.15)	3.7 e-07
SIRD	-0.91 (0.16)	2.4 e-08
MOD	0 (Ref.)	
MARD	-1.14 (0.13)	6.3 e-15
MARD-MRS		
SIDD	-0.08 (0.02)	0.002
SIRD	-0.09 (0.02)	2.9 e-04
MOD	-0.17 (0.02)	2.0 e-11
MARD	0 (Ref.)	

Linear regression coefficients for the association between the four subgroups and the MRSs, taking the corresponding subgroup for each MRS as the reference group.

Table 2. **Cross-tissue DNA methylation of sites included in subgroup-unique methylation risk scores (MRSs) in different human tissues.** Correlations between DNA methylation of sites included in subgroup-unique MRSs in blood and DNA methylation of these sites in adipose tissue and skeletal muscle taken from the same subjects for these cell types from the Monozygotic Twin Cohort based on FDR below 5% (q<0.05).

			Blood ~ Adipose Tissue (n=32)		**Blood ~ Skeletal Muscle (n=28)**	
CpG site	**Subgroup**	**Annotated gene**	**r**	**q**	**r**	**q**
cg05963087	SIDD	*ENOX1*	0.96	5.53E-14		
cg14013597	SIDD		0.81	2.18E-06	0.78	9.82E-06
cg23616741	SIDD	*PGAM2*	0.82	2.18E-06		
cg25356393	SIDD	*KCNQ2*	0.73	0.0001		
cg16867657	MOD	*ELOVL2*	0.74	0.0001		
cg13379325	SIDD	*KCNQ2*	0.72	0.0001		
cg22891868	SIDD	*MOGAT1*	0.69	0.0003		
cg13930790	SIDD	*NCLN*	0.68	0.0005	0.56	0.016
cg15081033	SIDD	*SPSB4*	0.64	0.001		
cg26161329	MOD	*PPM1E*	0.62	0.002		
cg16276209	SIDD	*AATK*	0.59	0.004		
cg14578612	SIDD	*KCNAB2*	0.59	0.004		
cg02789526	SIDD	*A2BP1*	0.57	0.007		
cg01542019	MOD	*TECR*	0.56	0.007		
cg07963234	MARD	*CRMP1*	0.52	0.016		
cg14692377	MOD	*SLC6A4*	0.49	0.025		
cg06933824	MOD	*NEURL1B*	0.49	0.025		
cg15225267	SIDD		0.49	0.027	0.72	8.31E-05

Pearson correlation tests show significant correlations between DNA methylation of sites in blood and skeletal muscle and adipose tissue, respectively, for the subgroup-unique sites included in the subgroup-unique MRSs. A false discovery rate (FDR) analysis based on Benjamini-Hochberg was performed and FDR below 5% ($q < 0.05$) was considered significant. DNA methylation of 57 out of the 95 subgroup-unique sites included in any of the MRSs was available from the 450K array and used to analyze DNA methylation in blood, muscle and adipose tissue in subjects from the Monozygotic Twin Cohort. For 32 subjects, methylation data was available for blood and adipose tissue, and for 28 subjects, methylation data was available for blood and skeletal muscle. Here, DNA methylation in blood correlated positively with DNA methylation in adipose tissue of 18 sites and in skeletal muscle of 3 sites ($q < 0.05$).

FIGURE LEGENDS

Figure 1. Patient distribution and phenotype characteristics by type 2 diabetes subgroups in the discovery and replication cohorts. Phenotypes were measured in the ANDIS (All New Diabetics in Scania) and ANDiU (All New Diabetics in Upsala) cohorts. Included patients were previously defined as: SIDD: severe insulin-deficient diabetes, SIRD: severe insulin-resistant diabetes, MOD: mild obesity-related diabetes, MARD: mild age-related diabetes. Pie chart showing the subgroup distribution in the (A) *ANDIS discovery cohort* (n=280), (B) the *ANDIS replication cohort* (n=76), and (C) the *ANDiU replication cohort* (n=197). Boxplots are showing the distribution of age at diagnosis, BMI, HbA1c, HOMA2-B and HOMA2-IR and bar charts are showing the prevalence of male sex for each T2D subgroup in the respective cohort. Statistical differences between the subgroups were evaluated using Kruskal-Wallis for continuous variables, and Chi-squared test for categorical variables. $p < 0.05$ was considered significant. Post-hoc pairwise comparisons for continuous variables were done using Dunn's test including correction for multiple testing based on Benjamini-Hochberg. Significance is indicated as follows: q<0.05*, q<0.01**, q<0.001***. For detailed characteristics see Supplementary Table 1.

Figure 2. Subgroup-unique methylation risk scores (MRSs) associate with type 2 diabetes (T2D) subgroups and future diabetic complications and play a biological function in the pathogenesis of T2D. (A-C): The respective subgroup-unique MRSs differ statistically significantly between patients with SIDD, SIRD, MOD and MARD and patients without the respective T2D subgroup in the *ANDIS discovery* (A), *ANDIS replication* (B), and in the independent *ANDiU replication cohort* (C). Patients within each subgroup had a statistically significantly ($p<0.05$) higher subgroup-unique MRS compared to the combined group of all other subgroups. Differences in MRSs were compared using Mann-Whitney U test. **(D):** Subgroup-unique MRSs associate with the T2D subgroups in the independent *ANDiU replication cohort* (n=197). Odds ratios (OR) are shown per 1SD increase in MRSs. In the logistic regression model, the dependent variable is the corresponding subgroup for each MRS versus the combined group of all other subgroups, so for SIDD-MRS it is SIDD versus Non-SIDD individuals, for SIRD-MRS it is SIRD versus Non-SIRD individuals, for MOD-MRS it is MOD versus Non-MOD individuals, and for MARD-MRS it is MARD versus Non-MARD individuals. **(E):** Associations between subgroups-unique MRSs and the risk of developing diabetic complications during 8 years of follow-up (mean ~4.5 years) in the combined *ANDIS discovery, ANDIS replication* and *ANDiU replication cohorts*. $p<0.05$ was considered significant. The results for the sex adjusted weighted-cox regression are presented as hazard ratios (HR) and 95% confidence intervals (CI). For cardiovascular disease (CVD) there are 410 controls and 76 cases (n=486), for chronic kidney disease (CKD) there are 444 controls and 73 cases (n=517), for diabetic retinopathy there are 490 controls and 54 cases (n=544). CVD was defined as having had either stroke (ICD-10 codes I60, I61, I63 and I64) or coronary events (ICD-10 codes I20-I21, I24, I251 and I253-I259). CKD was defined as having had an eGFR <60ml/min/1.73m2 for a minimum period of 90 days or a single measurement of eGFR <15ml/min/1.73m2. Diagnosis of diabetic retinopathy was based on fundus photographs (ICD-

10 code E113). MRSs were normalised to show the risk per 1SD increase. Patients with the respective complication before DNA methylation samples were excluded for the respective analyses. **(F-G):** Relevant genes annotated to the 95 sites included in the subgroup-unique MRSs associate with diabetes, non-alcoholic fatty liver disease (NAFLD) and/or with some subgroup-defining phenotypes and might therefore be important in the pathogenesis of T2D. We performed a systematic literature search using each gene symbol and the following terms: diabetes, insulin secretion/beta-cell function, insulin resistance, obesity, age and NAFLD. 39 (54%) of the 72 genes have been associated with diabetes and/or with some characteristics which defined the subgroups or NAFLD (F) and when looking at individual subgroup-unique MRSs, 23 (52%) of 44 genes included in SIDD-MRS, both genes (100%) included in SIRD-MRS, 12 (57%) of 21 genes included in MOD-MRS and 2 (40%) of 5 genes included in MOD-MRS were associated with any of the terms representing the subgroup-traits.

Fig 1.

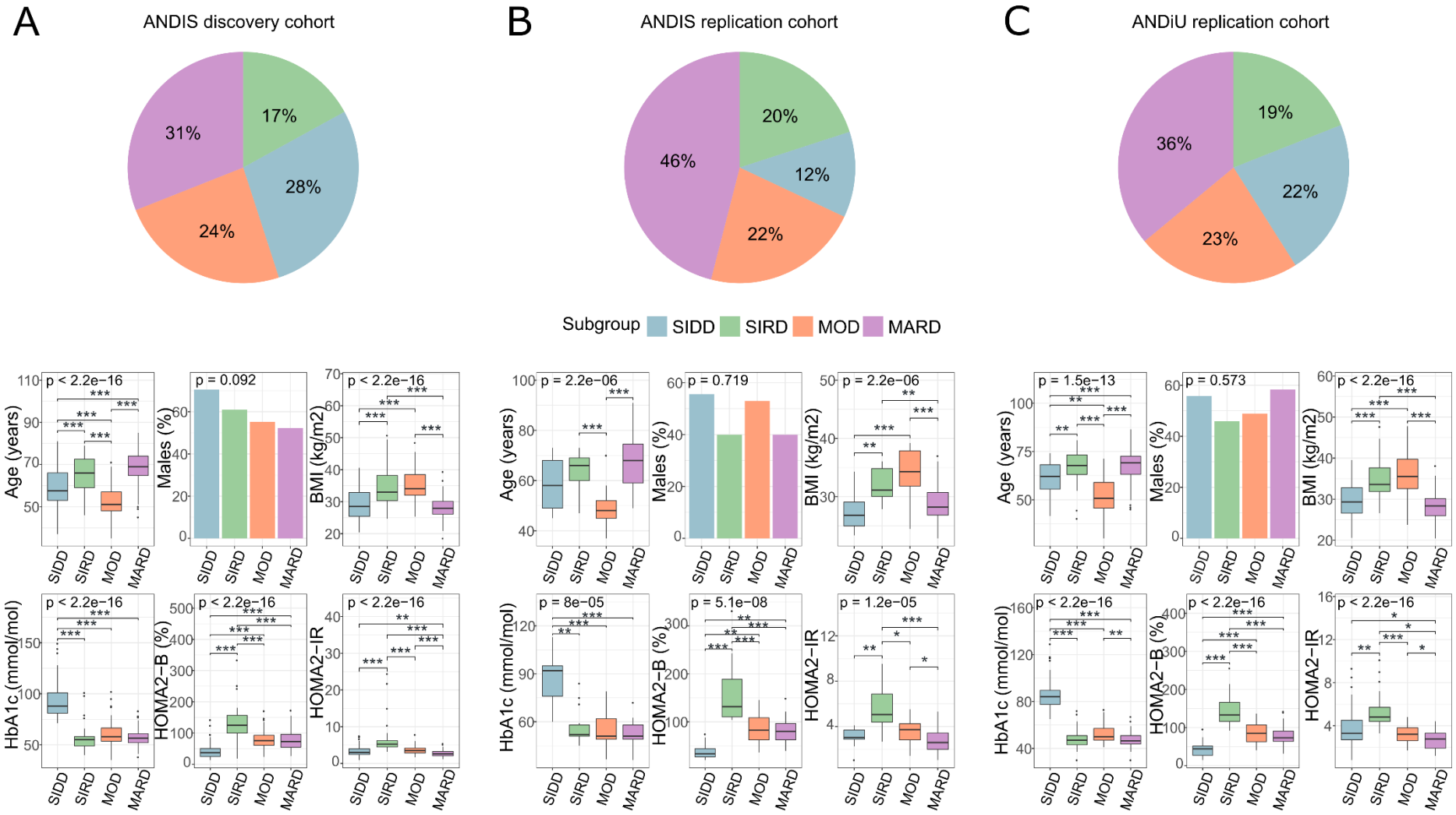
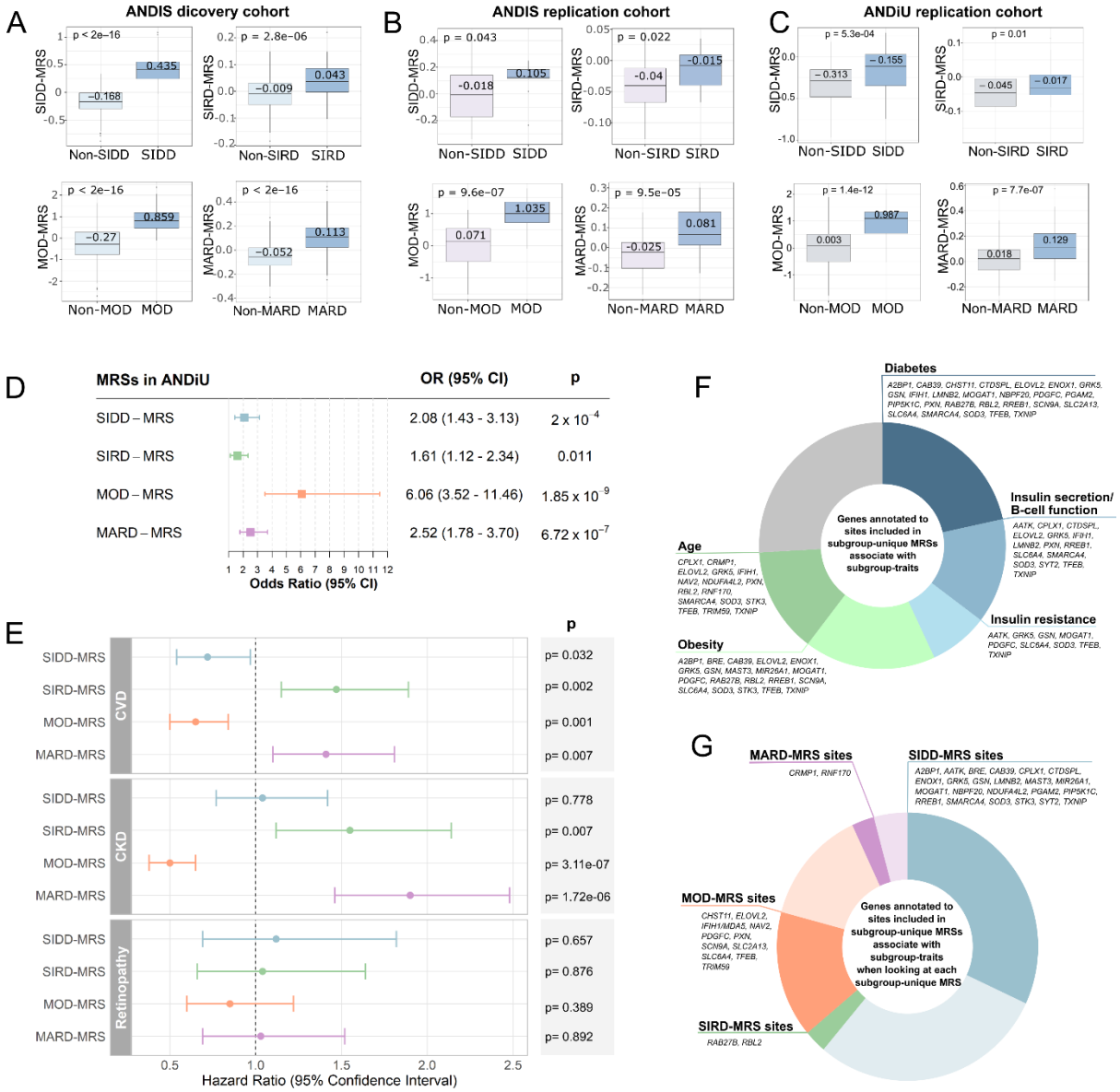

A
ANDIS discovery cohort
17%
28%
24%
31%
B
ANDIS replication cohort
20%
12%
22%
46%
C
ANDiU replication cohort
19%
22%
23%
36%
Subgroup SIDD SIRD MOD MARD
Age (years)
Males (%)
BMI (kg/m2)
HbA1c (mmol/mol)
HOMA2−B (%)
HOMA2−IR
p < 2.2e−16
p = 0.092
p < 2.2e−16
p = 2.2e−06
p = 0.719
p = 2.2e−06
p = 1.5e−13
p = 0.573
p < 2.2e−16
p < 2.2e−16
p = 8e−05
p = 5.1e−08
p = 1.2e−05
p < 2.2e−16
p < 2.2e−16
p < 2.2e−16
SIDD
SIRD
MOD
MARD

Fig 2.


A
ANDIS dicovery cohort
SIDD-MRS
p < 2e−16
−0.168
0.435
Non-SIDD SIDD
SIRD-MRS
p = 2.8e−06
−0.009
0.043
Non-SIRD SIRD
MOD-MRS
p < 2e−16
−0.27
0.859
Non-MOD MOD
MARD-MRS
p < 2e−16
−0.052
0.113
Non-MARD MARD
B
ANDIS replication cohort
SIDD-MRS
p = 0.043
-0.018
0.105
Non-SIDD SIDD
SIRD-MRS
p = 0.022
-0.04
-0.015
Non-SIRD SIRD
MOD-MRS
p = 9.6e−07
0.071
1.035
Non-MOD MOD
MARD-MRS
p = 9.5e−05
-0.025
0.081
Non-MARD MARD
C
ANDiU replication cohort
SIDD-MRS
p = 5.3e-04
− 0.313
− 0.155
Non-SIDD SIDD
SIRD-MRS
p = 0.01
− 0.045
− 0.017
Non-SIRD SIRD
MOD-MRS
p = 1.4e-12
0.003
0.987
Non-MOD MOD
MARD-MRS
p = 7.7e-07
0.018
0.129
Non-MARD MARD

D
MRSs in ANDiU
OR (95% CI)
p
SIDD – MRS
2.08 (1.43 - 3.13)
2 x 10 -4
SIRD – MRS
1.61 (1.12 - 2.34)
0.011
MOD – MRS
6.06 (3.52 - 11.46)
1.85 x 10 -9
MARD – MRS
2.52 (1.78 - 3.70)
6.72 x 10−7
1 2 3 4 5 6 7 8 9 10 11 12
Odds Ratio (95% CI)

E
p
CVD
SIDD-MRS
p= 0.032
SIRD-MRS
p= 0.002
MOD-MRS
p= 0.001
MARD-MRS
p= 0.007
CKD
SIDD-MRS
p= 0.778
SIRD-MRS
p= 0.007
MOD-MRS
p= 3.11e-07
MARD-MRS
p= 1.72e-06
Retinopathy
SIDD-MRS
p= 0.657
SIRD-MRS
p= 0.876
MOD-MRS
p= 0.389
MARD-MRS
p= 0.892
0.5 1.0 1.5 2.0 2.5
Hazard Ratio (95% Confidence Interval)

F
Diabetes
A2BP1, CAB39, CHST11, CTDSPL, ELOVL2, ENOX1, GRK5, GSN, IFIH1, LMNB2, MOGAT1, NBPF20, PDGFC, PGAM2, PIP5K1C, PXN, RAB27B, RBL2, RREB1, SCN9A, SLC2A13, SLC6A4, SMARCA4, SOD3, TFEB, TXNIP
Insulin secretion/ B-cell function
AATK, CPLX1, CTDSPL, ELOVL2, GRK5, IFIH1, LMNB2, PXN, RREB1, SLC6A4, SMARCA4, SOD3, SYT2, TFEB, TXNIP
Insulin resistance
AATK, GRK5, GSN, MOGAT1, PDGFC, SLC6A4, SOD3, TFEB, TXNIP
Obesity
A2BP1, BRE, CAB39, ELOVL2, ENOX1, GRK5, GSN, MAST3, MIR26A1, MOGAT1, PDGFC, RAB27B, RBL2, RREB1, SCN9A, SLC6A4, SOD3, STK3, TFEB, TXNIP
Age
CPLX1, CRMP1, ELOVL2, GRK5, IFIH1, NAV2, NDUFA4L2, PXN, RBL2, RNF170, SMARCA4, SOD3, STK3, TFEB, TRIM59, TXNIP
Genes annotated to sites included in subgroup-unique MRSs associate with subgroup-traits

G
MARD-MRS sites
CRMP1, RNF170
SIDD-MRS sites
A2BP1, AATK, BRE, CAB39, CPLX1, CTDSPL, ENOX1, GRK5, GSN, LMNB2, MAST3, MIR26A1, MOGAT1, NBPF20, NDUFA4L2, PGAM2, PIP5K1C, RREB1, SMARCA4, SOD3, STK3, SYT2, TXNIP
MOD-MRS sites
CHST11, ELOVL2, IFIH1/MDA5, NAV2, PDGFC, PXN, SCN9A, SLC2A13, SLC6A4, TFEB, TRIM59
SIRD-MRS sites
RAB27B, RBL2
Genes annotated to sites included in subgroup-unique MRSs associate with subgroup-traits when looking at each subgroup-unique MRS