# Identification of epigenetic biomarkers associated with the development of diabetic kidney disease in individuals with type 2 diabetes: a nested cohort study

**Author:** Marian Paiva Marchiori

**Supervisor:** Charlotte Ling, Epigenetics and Diabetes Unit at Lund University Diabetes Center, Malmö, Sweden.

**Co-supervisor:** Sonia Garcia-Calzon, Department of Food Science and Physiology, Center for Nutrition Research, University of Navarra, Pamplona, Spain

**May 2023**

**Abstract**

**Introduction.** Type 2 diabetes (T2D) is the most common type of diabetes, and it can cause complications such as diabetic kidney disease (DKD), the main cause of end-stage renal disease worldwide. There is an urgent need to develop new biomarkers that would improve DKD's risk prediction in patients with T2D. This research aimed at identifying novel epigenetic biomarkers associated with future DKD in patients with T2D. **Methods.** Nested cohort study of 487 newly diagnosed individuals with T2D within the ANDIS and ANDIU cohorts, followed-up over a period of 11.5 years. Genome-wide DNA methylation analysis was performed in blood samples taken at registration in the cohorts, and its association with future DKD was assessed through weighted-cox regression models. **Results.** DNA methylation of 37 CpG sites was found to be significantly associated with future DKD in the sample, with an effect size of 5% ($q<0.05$). The HR ranged from 0.0005 (95% CI 0 - 0.07) to 37.05 (95% CI 4.88 - 281.28) per 1% of methylation increase. 20 CpG sites (54%) were hypermethylated in patients with diabetes who developed DKD, in relation to those who did not develop DKD. Some of these 37 CpG sites are annotated to genes with important reported biological processes, such as *ADAMTS16*, involved in regulation of arterial blood pressure; *RPH3AL* in positive regulation of insulin secretion, and *EHMT1* in methylation of histones. **Discussion.** Our study found that DNA methylation in blood taken at baseline is associated with future DKD in the study's population, indicating that DNA methylation markers can be potential valuable biomarkers for predicting DKD in T2D. **Conclusions.** To the author's knowledge, no study has investigated the association between DNA methylation and future DKD. There is a need, however, to refine the study design and to validate the results in different populations.

## Table of Contents

## 1. Introduction/Background

Diabetes is a chronic condition characterized by elevated levels of blood glucose due to insufficient production of insulin in the pancreas, or the body's inability to use the produced insulin effectively (1). According to the International Diabetes Federation, around 537 million people have been diagnosed with diabetes globally, and its prevalence is projected to increase to 643 million by 2030 (2). The disease has caused 6.7 million deaths, only in the year of 2021, and it represents a significant economic burden, corresponding to 9% of health expenditure on adults worldwide (2).

Diabetes can be divided into two main categories, type 1 and type 2 diabetes (T1D and T2D, respectively). T1D diabetes is considered an auto-immune disease, in which the body's immune system attacks the insulin producing beta-cells of the pancreas. As a result, the body produces very little or no insulin. The causes of this destructive process are not fully understood, but a likely explanation is that the combination of genetic susceptibility (conferred by a large number of genes) and an environmental trigger such as a viral infection, initiates the autoimmune reaction (3). The condition can develop at any age, although T1D occurs most frequently in children and young adults (2).

T2D is the most common type of diabetes, accounting for about 90% of cases (2). T2D is a multifactorial condition dependent on both genetic and environmental factors, in which the body's cells start being unable to fully respond to insulin, a condition known as insulin resistance. As a result, the hormone becomes less effective, eventually leading to an increased insulin production. However, over time, the pancreatic beta cells may fail to keep up with the demand for insulin, which can result in insufficient insulin production (2). This mixture of defective insulin secretion, insulin resistance, added to a state of chronic inflammation, affects the ability of the body to reduce the glucose levels in the blood. Chronic hyperglycemia damages the blood vessels, which can cause micro- and macrovascular complications, such as cardiovascular diseases (CVD), nephropathy, neuropathy, and retinopathy (4). These diabetes-related complications increase the necessity for medical assistance, reduce quality of life, generate stress in families and caregivers, and lead to more hospital admissions, weighting heavily in the individual's health and in the health systems as a whole. Ultimately, the risk for premature death in patients with diabetes is higher than the general population, putting diabetes between the 10 most common death causes in the world (2). The change towards a more unhealthy lifestyle observed in the recent years, followed by a rise in obesity and sedentarism,

contributed to the rapid growth of T2D (5), making it one of the most prevalent metabolic diseases nowadays.

Among the most common complications of T2D, occurring in around 40% of patients, is diabetic kidney disease (DKD), the main cause of chronic kidney disease (CKD) worldwide (6). Moreover, almost 50% of cases of end-stage renal disease (ESRD) occur due to DKD, making diabetes the main cause of ESRD globally (7). DKD contributes to high disease burden, and significant individual and societal costs. In the last decade of the 20$^{th}$ century, deaths attributed to DKD increased by 94%. For patients with diabetes, the majority of the excess risk of all-cause and CVD mortality is related to DKD (8).

Clinically, DKD is characterized by a persistently high urinary albumin over urinary creatinine ratio (UACR) and/or a sustained reduction in Estimated Glomerular Filtration Rate (eGFR). While a renal biopsy can confirm the diagnosis, it is typically unnecessary unless atypical features are present. Laboratories usually report serum creatinine along with eGFR, which is calculated through validated formulas, such as MDRD-4 and CKD-EPI, using age, sex, serum creatinine and race as parameters (9).

The risk factors for DKD include hypertension, hyperglycemia, advanced age, male sex, race/ethnicity (African American, Hispanic/Latino, American Indian, and Asian American/Pacific Islander descents are at a higher risk compared to non-Hispanic White people), family history of kidney disease, bad dietary habits, and obesity. Additionally, DKD occurs more often in patients who have had diabetes for a prolonged period and have also developed other microvascular complications, such as retinopathy (6). The pathophysiology of DKD involves several complex mechanisms that ultimately lead to kidney damage. The initial step in the development of DKD is maintained hyperglycemia, which causes damage to the small blood vessels in the kidneys. This leads to a reduction in blood flow and an increase in pressure within the blood vessels. Over time, this increased pressure causes the blood vessels to malfunction, allowing protein to escape into the urine, causing albuminuria. As the disease progresses, the damage to the blood vessels becomes more severe and the kidneys become less able to filter waste products from the blood, leading to an accumulation of waste products in the blood stream. Ultimately, diabetic kidney disease can progress to ESRD, which requires dialysis or kidney transplantation. The exact mechanisms that lead to ESRD are not fully understood, but they likely involve a complex interplay between genetic, environmental, and lifestyle factors (6).

Additionally, systemic hypertension and obesity contribute to glomerular hyperfiltration through mechanisms like high transmitted systemic blood pressure and glomerular

enlargement. Glomerular hyperfiltration is a well-established consequence of early diabetes, occurring in up to 40% of patients with T2D. The normal course of DKD evolves from glomerular hyperfiltration, albuminuria, reduction in estimated eGFR, followed by ESRD (6). However, it is noteworthy that a significant part of patients does not fall under this classic pattern. A study performed in the UK, following patients with T2D, found, for instance, that 60% of patients with impaired kidney function did not have previous albuminuria, and 40% never presented albuminuria during the follow-up (10).

In individuals with diabetes, prevention of DKD is fundamental to reduce morbidity and mortality, including rigid control of glucose, blood pressure, body weight, and attention to the use of nephrotoxic medications (11). For that to occur, patients with T2D must be correctly and timely diagnosed. But despite the burden T2D presents, approximately 30% of people who have the disease are not diagnosed (12). This leads to a lot of patients already having, or being in the path to having, comorbidities and complications such as DKD at diagnosis (13). In addition to early diabetes diagnosis, the identification of individuals with diabetes at a high risk of developing DKD is fundamental for its prevention, and it would represent a step forward towards precision medicine. However, there are currently no clinically useful biomarkers which could predict DKD ideally already at diabetes diagnosis. Identification of such biomarkers would allow for targeted treatments and intensive follow-up for this high-risk population, thus preventing DKD and reducing the burden of this complication for patients and society.

Epigenetic biomarkers may be developed and used for precision medicine (14). Epigenetics are changes in the DNA and its structure, that are not changes in the sequence of the genome, and which can alter gene expression. These changes can be brought about by various external or internal factors such as environment, lifestyle, and aging, and they can be passed on through cell division. The most common epigenetic modification is DNA methylation, in which methyl groups are added to the cytosine base of DNA. This modification typically occurs at CpG sites (CpGs) and is catalyzed by DNA methyltransferases (DNMTs). CpGs are regions of DNA where a cytosine nucleotide is followed by a guanine nucleotide, connected by a phosphate group, in the DNA sequence. CpGs are often found clustered in specific regions of the genome called CpG islands, which are typically located in or near the promoter regions of genes. DNA methylation can regulate gene expression by altering chromatin structure and inhibiting the binding of transcription factors and other DNA-binding proteins to the DNA sequence, ultimately leading to the repression of gene expression (15). Figure 1 shows a schematic representation of how DNA methylation works.

Blood is a readily accessible and non-invasive source of DNA, making it an attractive option for biomarker discovery and validation, including epigenetic biomarkers. Biomarkers are measurable indicators or signals of biological processes, disease states, or responses to treatments or interventions. They can be molecules, such as proteins or nucleic acids, or physical characteristics, such as blood pressure or imaging features. They can be used for a variety of purposes, including early disease detection, diagnosis, monitoring disease progression, predicting treatment response, and identifying potential therapeutic targets. For example, blood glucose levels are a biomarker for diabetes, and eGFR levels are biomarkers for DKD (16). Furthermore, it is already shown that DNA methylation is associated with different clinical manifestations and phenotypes, supporting its potential use as a biomarker for diagnosis and treatment of diseases (17).

Currently, identification of people with diabetes at a high risk for developing DKD is done through a subjective assessment based on traditional risk factors (8). With proper use, biomarkers from routine clinical data such as albuminuria and eGFR can modestly predict future renal status, but only when the pathophysiological processes have already started, and the disease is on its initial stages (16). Many novel biomarkers have been studied in relation to their ability to predict the progression of DKD, but most of them have no proven prognostic value, and/or are difficult to be measured in a practical clinical setting (18). Additionally, current biomarker studies insufficiently evaluate the incremental predictive value over routinely available clinical data (16). On a positive note, a study showed that clinical data with known risk factors, including age, sex, glycated hemoglobin (HbA1c), eGFR, and albuminuria, associated with progression of CKD in patients with T2D with a 0.706 area under the ROC curve, and that the addition of a panel of 14 serum protein and metabolite biomarkers increased it to 0.868 (19).

However, even with the inclusion of more biomarkers to prediction models, current risk assessment criteria and therapies are not enough to reduce the development and progression of DKD (19). For instance, it is known that an adequate glucose control is fundamental for the prevention of DKD, along with other microvascular complications. Despite that, it is observed that some patients still develop DKD even with good glycemic control, and a significant number will not develop it despite having poor glycemic control (20), highlighting the contribution of additional factors in the advancement of DKD. Furthermore, traditional risk factors and biomarkers may be delayed in identifying the disease onset, since before eGFR reduction and albuminuria appear, various cellular and molecular abnormalities already exist (6). It is now well known that the initiation and advancement of DKD are the outcome of intricate interplay

between genetic and environmental factors. Environmental cues such as diet, exercising, stress, environmental toxins, and sleep, may alter intracellular pathways through chromatin modifiers and regulate gene expression patterns, resulting in diabetes and its complications (21). As a result, there is a pressing need to discover simple, innovative, and accurate methods to detect DKD in its early stages, or even before its onset, in order to slow down or halt its progression (22), and epigenetic biomarkers are a very promising option for that.

As a first argument, we have several examples of how epigenetic changes are related to the development of T2D. For example, epigenetic differences between tissues from T2D patients and controls in different populations have been widely found in the literature (23). Moreover, there has been an increasing interest in identifying blood-based epigenetic biomarkers for risk assessment in patients with diabetes. For instance, DNA methylation in blood has already been associated with future T2D (24), dysregulation in pancreatic islets and insulin secretion (25) and response and tolerance to metformin therapy (26).

On a similar note, several studies have shown the relationship between DNA methylation and the development and progression of renal disfunction and kidney disease in the general population (27, 28). Human studies and animal models showed an association between DNA methylation and kidney function, including assotiation of methylation and renal fibrosis (29), inflammatory activity of peripheral immune cells (30), and albuminuria in patients with diabetes (31). Moving forward, differences in DNA methylation between CKD patients and controls have been found in non T2D patients (32) and in patients with T2D (33). Potential epigenetic biomarkers for CKD progression were also found in a study that compared methylation levels in patients with diabetes and ESRD and patients with diabetes without nephropathy, with the majority of the sample being of patients with T2D (34). Interestingly, a study showed difference in methylation between T2D subgroups: severe insulin-deficient diabetes (SIDD), severe insulin-resistant diabetes (SIRD), mild obesity-related diabetes (MOD), and mild age-related diabetes (MARD), and that subgroup-unique methylation risk scores were associated to different risks of DKD, with lower risk of renal disease in the MOD subgroup, and increased risk in the SIRD and MARD subgroups (35), evidencing the heterogeneity of T2D patients, and the need to find specific biomarkers in order to improve treatment and prevention of complications in this population. It is important to highlight, however, that the studies done so far have looked at the association of DNA methylation and the outcomes at the same point in time, where the disease was already in progress, so no causal or prediction value can be drawn from them.

In summary, even though research is developing regarding novel biomarkers for diagnosis and prognosis of diabetes complications (35) and kidney disease (36), there is an urgent need to develop new biomarkers, that, on the top of traditional risk factors, would improve DKD's risk prediction in patients with T2D (37), since it poses a public health challenge of global impact, with important comorbidities, reduced life expectancy, and increased mortality.

With that in mind, the overall objective of this research was to identify novel high-quality epigenetic biomarkers that can be associated with the development of DKD in patients with newly diagnosed T2D, alone and/or in addition to the traditional clinical risk factors. The study hypothesis is that CpGs in blood have different degrees of DNA methylation already at diabetes's diagnosis, comparing T2D patients who developed DKD and those who did not develop DKD during 11.5 years of follow-up. The specific aims are to describe a population of patients with newly diagnosed T2D who developed and did not develop DKD, and to assess if there are differentially methylated CpGs between the two groups at the diagnosis of diabetes while evaluating if methylation of these sites is significantly associated with future DKD in the study population.

## 2. Methods

### 2.1. Study design and population

This is a nested cohort study of newly diagnosed individuals with T2D within the ANDIS ("All New Diabetics in Scania") and ANDIU cohorts. ANDIS is an ongoing clinical study started in 2008 aimed at recruiting all new cases of diabetes in Scania region in Sweden, with a total sample of approximately 26962 patients (last update in December 12$^{th}$, 2022) (https://andis.ludc.med.lu.se) (38, 39), whereas ANDIU ("All New Diabetics in Uppsala County") started in 2012, including all newly diagnosed patients with diabetes from the Uppsala region (https://www.andiu.se/) (26). There's no update on the total population of the study, but the population of Uppsala County is around 300,000 individuals, with an estimated diabetes prevalence of 3.5% and an incidence of 3.8 per 10,000 inhabitants annually. The ANDIU study protocol was approved by the Regional Ethics Review Committee in Uppsala, Sweden (2011/155). The ANDIS cohort was approved by the Regional Ethical Review Board in Lund (numbers 584/2006, 2011/354, 2014/198). Both cohorts were performed according to the Declaration of Helsinki, and written informed consent was obtained from all participants.

Demographic data was collected through questionnaires for all registered patients. Blood sample was also collected for all patients, both for clinical analysis and DNA extraction, at the moment of registration. Clinical information was obtained from the patient's medical records, which in ANDIS were connected to the National Diabetes Registry (NDR) and 'Region Skånes Vårddatabas', or Region Skåne's Healthcare Database, and in ANDIU to the NDR. Information about medication came from the drug registry, with data on drug prescriptions and withdraws. All patients were followed up until the most recent data from the cohorts' databases for the outcome definition.

### 2.2. Sample selection

The flow chart of individuals' selection is shown in Figure 2. The inclusion criteria for this study were patients diagnosed with T2D, with DNA methylation data available, and older than 46 years old. Patients with no available information about baseline clinical variables, such as eGFR, six months before or six months after DNA methylation sample date were excluded from the analysis. Cases with onset of DKD before or within baseline were also excluded. Patients who developed DKD but had reported acute kidney failure, based on the ICD code N17 taken from their medical records, were also excluded. Patients with no follow up of clinical variables were excluded as well. At last, all samples who clustered with the wrong sex in the quality control step of the DNA methylation analysis were excluded to reduce uncertainty (see item 2.3 – DNA methylation analysis). At the end we had a total of 487 patients, putting together patients from both ANDIS and ANDIU cohorts, with 88 patients who developed DKD and 399 patients who did not develop DKD over a follow-up period of 11.5 years.

### 2.3. Baseline sample characterization and phenotypes

Baseline characterization of the sample was considered at the blood sample collection date (DNA methylation measurement), and baseline characteristics were collected within a period of one year, either six months before or six months after the blood sample collection date (Figure 3). This choice was made due to the fact that patients had their baseline measurements taken at different time points, because this information was taken from the registries rather than collected as a part of the study. 88% of the sample had all clinical and laboratory measurements within this interval, and most of them very close to the baseline, as

shown in figure 4. The 12% of the sample with no baseline values within this time period was excluded, as explained above.

Information on age, sex, weight, height, and glycated hemoglobin (HbA1C) measurements were taken at the participants' registration in the cohorts. Body mass index (BMI) was calculated from weight and height measurements using standardized protocol. Duration of diabetes at baseline was calculated from T2D diagnosis date to DNA methylation sample date.

All eGFR measurements (baseline and follow-up eGFR), were calculated with the MDR-4 equation using serum creatinine values (40). Baseline eGFR was calculated with serum creatinine values measured at registration. Creatinine, urinary creatinine, urinary albumin, UACR, triglycerides, total Cholesterol, HDL, and LDL values were taken from the blood test with date closest to the DNA methylation sample date. HOMA2-IR for the ANDIU patients was taken from the registry, and for the ANDIS patients, was calculated using the HOMA model (41) through the HOMA2 Oxford Calculator (42).

Data on medication was taken from the national drugs registry. If the patients bought the medications at least once within baseline (six months before or six months after DNA methylation sample date), we considered that they were using the medication - antihypertensives, statins or metformin - at baseline.

CVD was defined as either having had coronary events (defined by ICD-10 codes I20-, I21, I24, and I25) or stroke (defined by ICD-10 codes I60, I61, I63, and I64) before or during baseline (up to six months after DNA methylation date).

### 2.4. Outcome definition and follow-up measurements

During follow up, patients were divided accordingly, as either having developed the outcome of the study, which is DKD, and not having developed DKD. DKD was defined by at least two measurements of eGFR below 60ml/min/1.73m2 with 90 days or more of interval between them, corresponding to patients belonging to the G3a, G3b, G4 and G5 groups in the KDIGO classification (43).

First eGFR below 60 corresponds to the first altered eGFR measurement during follow-up. Second eGFR below 60 corresponds to the eGFR at the moment the criteria for the disease (second measurement of eGFR lower than 60 with more than three months of interval from the first altered measurement) is fulfilled. Time to event was calculated only for patients who developed DKD, from DNA methylation sample date to the date of the second altered eGFR

measure, according to the disease definition previously stated. Total follow up time was calculated only for patients who did not develop DKD, from DNA methylation sample date to the date of the last blood test in the registry.

### 2.5. DNA methylation analysis

The Gentra Puregene Blood kit (Qiagen) was used to extract DNA from whole blood, following the manufacturer's instructions. Subsequently, the nucleic acid concentration and purity were evaluated with a NanoDrop 1000 spectrophotometer (NanoDrop Technologies). Bisulfite treatment of either 1000ng or 500ng of genomic DNA was carried out using the EZ DNA methylation kit from Zymo Research.

Genome-wide DNA methylation analysis was then performed for all samples. The resulting DNA methylation data was analyzed with the Infinium MethylationEPIC BeadChip array which covers 853,307 CpG sites, from which 333,265 are situated in regulatory regions, 2,880 are non-CpG sites, and 59 are SNP sites. It encompasses 99% of all RefSeq genes and 96% of CpG islands. Moreover, out of the 439,562 CpGs featured on the 450K Infinium array (Illumina), 91% (400,496 CpGs) are also present on the 850K array (44).

Bisulfite-converted human blood DNA samples were randomized across the chips. The DNA was subsequently amplified, fragmented, and hybridized to the BeadChips using the Infinium HD assay methylation protocol. After single base extension and staining, the BeadChips were imaged with the Illumina iScan. The raw fluorescence intensities from the scanned images were extracted using the GenomeStudio methylation module software, which provided the raw methylation score for each DNA methylation site. This score is represented as a β-value, computed using the following equation: β-value = intensity of the Methylated allele (M) / (intensity of the Unmethylated allele (U) + intensity of the Methylated allele (M) +100). The β-values range from 0 (completely unmethylated) to 1 (completely methylated). All the samples cleared the quality control checks in GenomeStudio, which involved the built-in control probes for staining, hybridization, extension, and specificity, and exhibited high-quality bisulfite conversion efficiency (intensity signal >4,000).

DNA methylation data was exported from GenomeStudio to be analyzed by the Bioconductor package (45). The first dataset contained 865918 probes, from which 537 probes annotated to the Y chromosome, 59 rs-probes, 2932 ch-probes, 42328 cross-reactive probes, 894 polymorphic probes, 2280 probes based on average detection p-values equal or higher than 0.01 based on McCartney et al. annotation (46) were filtered out. Dataset contained 816888

probes after filtering. β-values were then converted into M-values to eliminate heteroscedasticity in the data distribution (47). Quantile normalization to correct for intra-array variation and background correction were performed (48). Beta-Mixture Quantile normalization method (BMIQ) was used to correct for probe type bias and ComBat was performed to correct for batch effect (49, 50). Principal component analysis (PCA) was done before and after ComBat, to make sure that between-array batch effects were removed. PCA analysis plot and tree diagram of the hierarquical cluster analysis after combat both show the 487 selected samples clustered by gender, with no sample clustered in the wrong group (Figure 5).

Since whole blood contains various cell types, a reference-based method was employed to correct for any potential effects of cellular heterogeneity (51). Finally, for easier biological interpretation, the M-values were converted back into β-values, which were used to describe the data and generate figures.

### 2.6. Statistical Analysis

The data analyses were done using the R Project for Statistical Computing Software version 4.2.1 (2022-06-23, Funny-Looking Kid). At first, descriptive statistics was generated to characterize the sample. Normality was tested using the Shapiro-Wilk test. Differences between groups were analyzed through the appropriate test, according to the variables' distribution.

An *a priori* power analysis and sample size calculation was done using the Power and Sample size free online calculator (http://powerandsamplesize.com/Calculators/Test-Time-To-Event-Data/Cox-PH-2-Sided-Equality) through the Cox PH, 2-Sided Equality test to calculate the sample size needed to test time-to-event data (52), as done and explained by others (53). 24% was used as the probability of the disease in the general population, as it was shown in a Nature Review paper (54) as the prevalence of CKD in patients with T2D in the European population. The proportion of one patient with DKD for each four patients without DKD was then estimated (0.25 or 1:4) for the present study. The goal was an alpha of 0.05, or 5%, and a power of 0.8, or 80%. The formula used the standard normal distribution for the population variance. For the treatment effect size, we considered a study which proposed two differentially expressed genes as biomarkers for the development of DKD (55). The first gene proposed (*RUNX3*) showed a hazard ratio (HR) of 2.5, which was used as a reference for this study. Applying all parameters in the calculator, the sample size needed in the present study would be

at least 208 patients. Since the current sample was 487 patients, we believe that we had enough statistical power for our analyses.

After that, general logistic regression models were done to assess the association between each clinical variable (covariate) and the development of DKD (models followed the formula DKD ~ phenotype). Then, the Receiver Operator Characteristic Curve (ROC curve) was plotted and the area under the curve (AUC) calculated for each association.

Next step was to assess whether DNA methylation at baseline, the main exposure of the study, was associated with the development of DKD. Eight different weighted-cox regression models were performed. In a weighted Cox regression, the effect size of a predictor variable on the hazard rate (HR) is calculated by considering the distribution of that variable within the study population, which is the weighted mean. The weighted mean provides a more accurate measure of the predictor variable's effect on the HR, as it calculates an average effect size of the predictor variable weighting the effect size based on the proportion of individuals in the population who have that particular value of the predictor variable (56). Model 1a was adjusted for age, sex, BMI, and baseline eGFR, and model 1b was model 1a adjusted also for cell composition using the reference-based method (51). None of the blood cell types were associated with the outcome in this analysis, and therefore the following models were not adjusted for cell composition. Model 2 included model 1a and diabetes duration at baseline. Model 3 build up from model 2, including also HbA1c as a covariate. Model 4 included model 3 plus UACR. Model 5 was adjusted for use of statins and antihypertensives in addition to the covariates of model 3. Model 6 was based on model 3, with the addition of total cholesterol, LDL, HDLD, use of antihypertensive medication, and smoking. Finally, model 7 was built from model 6, adding also previous CVD event as a covariate. Since smoking information was lacking for both cohorts, we used DNA methylation of the CpG site cg05575921 to adjust for smoking, since it has been shown that its methylation in whole blood DNA strongly correlates with smoking status (AUC of 0.98) (57).

The models were adjusted for multiple testing by utilizing false discovery rate (FDR) analysis (Benjamini-Hochberg), where significance was considered at FDR <5% ($q < 0.05$). Model 1a was used as the reference model. The CpGs were further filtered based on an effect size higher than 0.02, or 2%. In the present study, effect size was based on the difference in methylation between patients who developed DKD and patients who did not develop DKD. We checked the overlap of significant CpGs in Model 1a with the other models using $p < 0.05$. At last, we increased the effect size to 5%, and checked again for overlap with other models. The resulting CpGs were then used for further analyses.

The ROC curves and AUCs were calculated for each of the final identified CpGs (CpGs with a 5% difference in methylation between groups and q<0.05). Additionally, the association between DNA methylation levels for each identified CpG site and all the covariates in Model 1a (Sex, Age, BMI and baseline eGFR) was assessed through generalized multiple linear regression models, and the association between DNA methylation levels for each CpG with each clinical variable in isolation through Spearman correlation tests for the continuous variables (age, BMI and baseline eGFR) and through point-biserial correlation tests for the binary variable (sex).

The "GeneCards – the human gene" database (www.genecards.org) (58) was used to search for the genes to which the CpGs of interest were annotated to, in order to better understand their biological function. Finally, all beta values, representing the methylation level for each patient in each CpG site, were plotted as box plots comparing patients who developed and who did not develop DKD.

Data are presented as mean ± SD or counts and percentages, unless stated otherwise. Normalized methylation β-values were used for the Weighted-cox regression models.

## 3. Results

The total sample constituted of 487 individuals from the ANDIS and ANDIU cohorts. The mean follow-up time was 2380 days for patients who did not develop DKD and time to event was 1566.5 days for the patients who developed DKD (Table 1).

For patients who did not develop DKD, 59.4% were male, mean age was 60.38 years, and mean BMI 31.41 kg/m$^2$. Average diabetes duration at baseline was 53.35 days. Mean baseline eGFR was 96.79 mL/min/1.73m$^2$. 55.6% of them were using antihypertensive medication, 43.1% statins, and 84.5% metformin at baseline. 8% had reported cardiovascular event before follow-up started (Table 1).

For patients who developed DKD, 57.9% were male, mean age was 67.69 years, and mean BMI 30.49 kg/m$^2$. Average diabetes duration at baseline was 62.98 days. Mean baseline eGFR was 80.33 mL/min/1.73m$^2$, and mean first eGFR below 60 and second eGFR below 60 were 53.5 mL/min/1.73m$^2$ and 51.25 mL/min/1.73m$^2$, respectively. 73.9% of them were using antihypertensive medication, 47.7% statins, and 81.8% metformin at baseline. 15.9% had reported cardiovascular event before follow-up started (Table 1).

Patients who developed DKD were significantly older than patients who did not developed DKD (p = 7.063e-11), had a higher duration of diabetes at baseline (p = 0.01498),

higher baseline eGFR (p =2.59e-13), higher creatinine (p = 1.075e-07), made more use of antihypertensive medication (p = 0.001679), and had a higher presence of CVD event (p = 0.02218) at baseline. All the information about clinical and demographical sample characterization for both groups is summarized in Table 1. The ROC curves, with the respective calculated AUC values for each general logistic regression model for the association between each variable and DKD, are shown in figure 6. The only variables with good AUCs were age (AUC of 0.72) and baseline eGFR (AUC of 0.74).

To assess if methylation of CpG sites was associated with future DKD showing a different methylation pattern between patients who developed DKD and patients who did not develop DKD at diagnosis of T2D, we analyzed DNA methylation from blood of a subset of patients from the ANDIS and ANDIU cohorts (n=487). We then performed eight different weighted-cox regression models. For models 1a, 1b, 2, 3 and 5, all the 487 participants had information on the covariates and were included in the analysis. For model 4, 461 participants were included due to missing UACR data, in model 6, 476 patients were kept due to missing Total cholesterol, HDL and LDL data, and for model 7, 465 had HOMA2-IR information for the analysis. With $p < 0.05$, there were 129,811, 117,598, 131,194, 133,821, 127,661, 136,946, 121,801, and 129,766 CpGs differentially methylated in Models 1a, 1b, 2, 3, 4, 5, 6 and 7, respectively. For $q < 0.05$, 56,411, 5,086, 56,968, 59,373, 56,144, 57,559, 2,862, and 3,833 differentially methylated CpGs were found in Models 1a, 1b, 2, 3, 4, 5, 6 and 7, respectively. All the eight models and the respective methylated CpG sites associated with DKD found in the analyses are summarized in Table 2.

Model 1a was defined as the reference model, with 129,811 CpGs associated with DKD with $p < 0.05$. Feature selection was done at first considering a level of significance of $q < 0.05$, after which there were 56411 significant CpGs. After further filtering for an effect size of 2%, there were 3,147 significant sites. From these 3,147 sites, the overlap with the other models based on $p < 0.05$ was checked. For model 1b, there was an overlap of 2,975 sites. In models 2, 3, 4 and 5, all sites overlapped. For model 6, 3,142 sites overlapped, and for model 7, 3,143. When increasing the desired effect size to 5% of difference in methylation between groups to increase the possible biological function of methylation sites, there were 37 significant sites in Model 1a. All 37 sites were present in the other models with $p < 0.05$. These 37 CpGs were then considered our CpGs of interest and were further examined.

In figure 7, all the 129,811 CpGs significant with $p < 0.05$ in Model 1a are shown in the Volcano and Manhattan plots. The CpGs significant with $q < 0.05$ appear above the red line,

and the 37 CpGs of interest are marked in red. Most of these 37 CpGs were located in the body region of the respective annotated gene, and in the shore region of the CpG island (Table 3).

The lowest HR for DNA methylation was 0.0005 (95% CI 0.00002 - 0.01) per 1% in increase of methylation for the CpGs cg08641951, annotated to *ADAMTS16*. The highest HR was 37.05 (95% CI 4.88 - 281.28) per 1% in increase of methylation for the cg12183150 site, annotated to *AC097713.4* and *AC097713.3* (Table 3).

The most hypermethylated CpG site for patients who did not develop DKD was the cg25621215 site (88% methylated), annotated to *RP11-445H22.3*, and the most hypomethylated was the cg08641951 site (24% methylated), annotated to *ADAMTS16*. For patients who developed DKD, the most hypermethylated CpG site was cg01079515 (86% methylated), annotated to *AC124944.5*, and the most hypomethylated was the cg08641951 site (19% methylated), same as the non DKD group. 20 of these 37 CpGs (54%) were hypermethylated in patients who developed DKD, in relation to the patients who did not develop DKD. The biggest difference in methylation was 12.9% of higher methylation in patients who developed DKD in the cg11420142 site (annotated to *RP11-122C21.1*) compared to non DKD individuals. The smallest difference was 5.04% of lower methylation in patients who developed DKD in the cg11065575 site (annotated to *SHISA3*) versus those who did not develop DKD. The description of these 37 sites is shown in table 3.

In relation to the association of the DNA methylation of these sites with DKD, all sites were significant with $q<0.05$ for the cox regression Model 1a, adjusted for age, sex, BMI, and baseline eGFR, as feature selection indicated. A significant association of age and baseline eGFR with the outcome (DKD) was observed in the same model for all the 37 CpGs. No significant associations of sex and BMI with DKD were seen. All the results of the weighted cox regression model 1a for these 37 CpGs are shown in table 4.

ROC curves were plotted individually for all the selected 37 CpGs in Model 1a, and the AUCs were calculated, as show in figure 8. No CpG in isolation could differentiate between patients who developed and patients who did not develop DKD in this sample, since all AUCs were below 0.60. To further characterize the 37 CpGs of interest, the associations between DNA methylation of these sites and clinical phenotypes were tested through general multiple linear regression models for each CpG site, with DNA methylation, represented by the beta values, as the dependent variable, and age, sex, BMI and baseline eGFR as independent covariates, and through Spearman correlation tests and point-biserial correlation tests, testing the direct association of DNA methylation of each CpG site with each variable/phenotype in isolation. For the linear regression models, age was significantly correlated with 6 CpGs (cg23480021, p

= 0.00129; cg18826637, p <2e-16; cg10187601, p = 0.0141; cg07784975, p = 9.23e-05; cg13989295, p = 0.04378; and cg16788857, p = 0.0153). Sex was significantly correlated with 4 CpGs, (cg11065575, p <2e-16; cg25291037, p = 0.0437; cg18826637, p = 0.0123, and cg07784975, p = 1.36e-06). BMI correlated significantly with 6 CpGs (cg00169354, p = 0.0388; cg26365090, p = 0.0037; cg11420142, p = 0.000598; cg19618634, p = 00693; cg11128983, p = 0.028124; and cg07917191, p = 0.0315). At last, Baseline eGFR correlated significantly only with 3 CpGs (cg25291037, p = 0.044; cg10187601, p = 0.0287; and cg11128983, p = 0.006296). In the point-biserial correlation test for the correlation between DNA methylation for each of the 37 CpGs and sex, only three sites were significant (cg11065575, p = 2.692e-16; cg25291037, p = 0.02943; and cg07784975, p = 3.287e-07). Only one site had a moderate strength of correlation (cg11065575, coefficient = -0.36). In the Spearman correlation tests, age was significantly correlated with 9 CpGs, but only showed one moderate strength of correlation (cg18826637, p < 2.2e-16, coefficient = -0.42). BMI correlated significantly with 5 CpGs, but all the correlations were weak, with the strongest being cg11420142 (p = 0.000151, coefficient = -0.17). Baseline eGFR was the phenotype that was correlated with the most CpGs, 12 in total. The strength of the correlations was also weak, with the strongest being a negative correlation with cg00169354 (p = 0.002194, coefficient = -0.14) and a positive correlation with cg04885581 (p = 0.00214, coefficient = 0.14). Table 5 shows the summary of these results.

Furthermore, since these 37 CpGs were annotated to 27 unique genes, a preliminary search was performed in the GeneCards database (58) for more information on their biological role in the context of T2D and DKD. When looking at the phenotypes associated with these genes in previous GWAS studies, *RPH3AL* was associated with insulin sensitivity measurement; *ZNF385D* with, among other traits, T2D, BMI, and CKD; *SHISA3* with T2D, creatinine measurement, and eGFR; *DZIP1* with eGFR; *SLC43A2* with eGFR and UACR; *TOX2* with eGFR and arterial blood pressure; *SKA2* with eGFR, and T2D; *PAPL* with T2D; *EHMT1* and *SLC9A4* with eGFR and creatinine measurements; and a number of others with blood pressure, such as *TEC, KCNH6* and *FAM13B*. Regarding Gene Ontology (GO) for biological processes, the *ADAMTS16* gene was found to be involved in regulation of systemic arterial blood pressure; *RPH3AL* in positive regulation of insulin secretion, and *EHMT1* in methylation of histones. The summary of the results of the GeneCards search relevant to the present study is shown on table 6.

Finally, to better visualize the 37 CpGs and understand the previous results, box plots were created comparing the methylation level, expressed in beta values, for patients divided

into those who developed and who did not develop DKD, as shown in figure 9. The box plots show that, for most of the CpGs, the distribution of beta values is quite similar when comparing patients who developed DKD and patients who did not develop DKD, even if their means are different, which indicates that the methylation levels of these CpGs are not very good to discriminate between these groups. Additionally, it is noteworthy that for many of the CpGs, a pattern of either low (close to 0%), intermediate (around 50%) or high (higher than 75%) methylation levels is evident, with clusters of patients clearly shown with gaps in between them. This pattern of distribution of beta values in boxplots indicates the presence of Single Nucleotide Polymorphisms (SNPs) (59). SNPs are the most common type of genetic variation found in the human genome. They are single nucleotide variations that occur at specific locations in the DNA sequence, where one nucleotide is replaced with another (60).

Only five out of the 37 CpGs do not have SNPs associated with them according to the Illumina manifest (cg08641951, cg23480021, cg12183150, cg07784975, and cg25621215) (Table 7). From the remaining 32 CpGs, 24 have SNPs at a distance of 10 base pairs or less from the probe sequence. Among those 24 CpGs, there are 15 SNPs with minor allele frequency higher than 0.05, which is considered frequent (61) (rs45554435, rs4773321, rs10270156, rs2021033, rs11700304, rs843071, rs7208505, rs73299210, rs57896542, rs12365323, rs5821824, rs57733150, rs59145932, rs35528310, rs59694233) (Table 7). All of these 15 CpGs show a SNP pattern in the box plots (figure 9).

From the 8 CpGs with SNPs at a bigger distance than 10 base pairs from the probe sequence, cg11065575, cg00169354, and cg23246911 show a good pattern on the box plots. cg11065575 is annotated to the *SHISA3* gene, which previous GWAS studies reported having SNPs associated to T2D, creatinine measurement, and glomerular filtration rate phenotypes. cg23246911 is annotated to the *RPH3AL* gene, which associates in previous GWAS studies to insulin sensitivity measurement and appear in the GO biological processes search as being involved in positive regulation of insulin secretion, glucose homeostasis, intracellular protein transport, and exocytosis, all important mechanisms involved in the development of DKD (Table 6). Only one SNP, rs35528310, reported in the Illumina manifest at one base pair distance from the cg14924512 probe with an allele frequency of approximately 6%, appears in the phenotypes described by previous GWAS, as showed in the GeneCard search, associated with BMI (Table 6). At last, four out of five previously mentioned CpGs which do not have SNPs associated with them have a good pattern on their boxplots (cg08641951, cg23480021, cg12183150, and cg07784975). The cg08641951 site is annotated to the *ADAMTS16*, and the cg23480021 site to the *ZNF385D* gene, both relevant genes for DKD pathogenesis, as

previously described. The information regarding the CpGs of interest and their associated SNPs is shown in table 7.

## 4. Discussion

The present study followed 487 recently diagnosed individuals with T2D for 11.5 years, which is a longer follow-up time when comparing with prospective cohorts studying patients with T2D for clinical outcomes (62, 63, 64). The majority of the individuals were male (59%), and patients who developed DKD were on average older than patients who did not develop DKD. They also had higher duration of diabetes, higher eGFR and creatinine values, a higher use of antihypertensive medication, and a higher presence of CVD event at baseline. The findings are in accordance with what is said in the literature about the characteristics of patients with DKD (54, 65), which are also related to the risk factors for the disease, as previously explained.

Since DKD is a multifactorial disease and has many risk factors and associated conditions which predispose its development, eight different weighted-cox regression models were generated, adjusting for different covariates, described above in the methods section. In the end, Model 1 a, adjusted for age, sex, BMI and baseline eGFR was considered the model of reference, because it contained fewer covariates, and originated similar results compared to the other models. This was done to simplify the analysis and not compromise statistical power. The rest of the models were also considered for the final selection of the methylation markers, thus identifying those methylation sites associated with future DKD independently of known clinical risk factors.

DNA methylation of a large number of CpGs significantly associated with DKD was found in this study (129,811 CpGs in our reference model). It is known that the quantity of CpG sites that are associated with a specific clinical outcome can vary considerably, usually ranging from a few to thousands. This variation is influenced by several factors such as the analysis method, level of statistical significance used, and complexity of the disease phenotype (66), but the high number of CpGs found shows that most probably DNA methylation plays an important role in the development of DKD, worth of further investigation. We then had to narrow down the CpGs sites of interest to understand better the biological relevance of the findings, and to get to clinically relevant results. Our feature selection to get to the final CpGs in the present study included selecting the sites who were significant in the model of reference with $q < 0.05$, but still significant in all the other models with $p < 0.05$. Besides that, we considered a

difference in DNA methylation levels of a at least 5% between the groups. That led us to the final 37 CpGs that were studied in depth in the subsequent analyses.

There has been increasing interest in investigating the role DNA methylation plays in the development of T2D complications, including its association with DKD and its features, although literature is still scarce in the matter. For instance, a cross-sectional study measured UACR of 96 patients with T2D and divided them into two groups: with DKD (n = 69) and without DKD (n = 27), based on their albumin excretion. They showed that hypomethylation of two target genes, *TIMP-2* and *AKR1B1*, was associated with albuminuria in early DKD (31). Another cross-sectional study performed in an East Asian population of 232 T2D patients found 3 differentially methylated CpGs in COMMD1, TMOD1, and FHOD1, when comparing 87 DKD patients and 80 non-DKD controls (33). Also, to corroborate the relevance of studying DNA methylation and its association with DKD, a meta-analysis identified 32 differentially methylated CpGs associated with DKD, but in individuals with T1D (67). However, no study has looked at the association of DNA methylation and the future development of DKD in T2D, making the prospective nature of our study one of its biggest strengths. Prospective studies overall provide stronger evidence for causality and are less prone to bias than cross-sectional studies (68). Additionally, it is important to look for biomarkers who could predict the disease before its onset, since current research is focused on disease diagnosis, monitoring, and prognosis, and only mentioning prediction when it comes to therapeutic response, as this literature review from 2023 summarized (69).

We are aware that the analysis of the AUCs for the association of DNA methylation levels of each of the 37 CpGs of interest and DKD showed that DNA methylation had no ability of discriminating between the study groups. However, it is important to highlight that it does not diminish their potential role as biomarkers for DKD prediction, since even if a single methylated CpG site was strongly associated with DKD, there is the possibility that it would be a small part of a larger network of epigenetic changes and other molecular mechanisms that contribute to the disease (70). These findings could also be due to the heterogeneous characteristics of DKD, so it is unlikely that one methylated CpG site alone would be able to capture this complexity. On a similar note, a study proposing new biomarkers to improve prediction of DKD progression considered a panel of 14 biomarkers on top of traditional risk factors and saw an increase in the AUC from 0.706 (clinical data alone) to 0.868, not showing the AUC for the new proposed biomarkers in isolation (19). To exemplify the possible role of DNA methylation to predict disease related outcomes in patients with T2D, there is a study which showed an association of higher methylation levels of 11 CpGs with a higher risk of not

responding to metformin therapy in T2D patients. Methylation risk scores for these 11 CpGs differed between the groups with AUCs ranging from 0.80 to 0.98 in replication cohorts (26). In this context, the ideal next step for the current project would be to build risk scores considering the CpGs found in this study together, them alone and on top of traditional biomarkers and clinical variables, which would give a better representation of the practical use of DNA methylation in the prediction of DKD.

Another positive finding was that the GeneCard search for the genes to which they are annotated to was very promising, since many of them had been previously related to T2D or DKD phenotypes (71, 72, 73, 74, 75, 76, 77). Among the genes found, *ADAMTS16*, for instance, was already shown to be involved in the pathogenesis of CVDs and hypertension (71). Despite its role in T2D and DKD not being yet understood, we can see an overlap of pathophysiological processes between the development of CVDs and DKD. The *RPH3AL* gene was associated with T2D (72) and methylation of CpGs annotated to this gene in obese people was associated with T2D (73). *ZNF385D* has evidence of association with eGFR and kidney failure in the literature (77). *TGF-β1* has been linked to the pathogenesis of many forms of CKD, including DKD, through renal fibrosis (74, 75). *CCR6* receptors play an important role in the pro inflammatory state associated with kidney disfunction (76). These results indicate the importance of continuing researching the association of DNA methylation and DKD in T2D, since many of these genes seem to have a key biological function in these diseases.

Even though most of the 37 CpGs of interest were shown to be associated to SNPs, we saw that 14 of them have a good pattern in the boxplots (Figure 9). The cg08641951 and cg23480021 sites, annotated to the *ADAMTS16* and to the *ZNF385D* genes, respectively, two genes previously related to DKD development (71, 76), could be interesting candidates to be used in methylation risk scores and as biomarkers for the disease prediction. In the possibility that the SNPs associated with some CpGs were affecting the difference in DNA methylation found between groups, a different feature selection may be done to identify more relevant sites which are not driven by genetic alterations and that can reflect the actual relationship between DNA methylation and DKD. The final 37 CpGs explored in this work were selected based mainly on a higher effect size, or difference in methylation between the groups. Maybe future feature selection could focus on a lower, but still relevant, effect size. A possibility is to consider 2% of difference in methylation between groups, and to focus on higher HRs or q values.

Another point raised in the beginning of data analysis was the fact that both groups were significantly different regarding age. As previously explained, age was included as a covariate in the adjustment for confounders in all the regression models, so the main associations between

methylation and DKD were found independently of age. It is known that including age as a covariate in the regression models is a common strategy to account for its effect, and that adjusting for age as a confounder in DNA methylation studies is fundamental due to the fact that age is strongly associated with DNA methylation patterns and related to many other factors that may influence DNA methylation (78).

However, age was still significantly associated with the outcome for most of the CpGs, including the 37 CpGs of interest, in all models. On the one hand, matching by age on top of covariate adjustment can be a useful strategy when the effect of age on the outcome variable is not linear and can be difficult to model accurately. By matching participants based on their age, the distribution across groups can be balanced, which can help to reduce confounding, and provide additional control for the age's effect. In addition, matching can increase the efficiency and precision of a study by reducing the sample variation, and by consequence, the sample size required to detect a significant difference between the groups (79). On the other hand, matching can also introduce selection bias in our case, because age is associated both with our exposure (DNA methylation) and the outcome (DKD), which can result in underestimation of the true effect of age. Matching by age can also increase the complexity of the study design and analysis, reducing the power of the study if more tests need to be performed. Additionally, it can reduce the generalizability and the external validity of the study, as it may not accurately reflect the real-world distribution of the matching variable and other confounding variables (68). It was decided, for the present study, to not match the sample, but it is something that could be done in the future in order to ensure that the results found are truly due to the effect of DNA methylation on the development of DKD, and that age is not driving the associations.

Baseline eGFR was also strongly correlated with future DKD, which was expected, since follow-up eGFR was used to define the disease. eGFR, together with albuminuria, is often used to diagnose and monitor the progression of DKD. Nevertheless, there is still no single biomarker that can definitively predict the appearance and progression of DKD. Clinicians typically use a combination of biomarkers, including albuminuria, eGFR, and other markers of kidney injury and function, to diagnose and monitor DKD in patients with diabetes (69), but still with only moderately good AUCs of around 0.7 (16).

That being said, eGFR is a measure of kidney function and its decline is a common feature of DKD, which already indicates the presence, or at least the beginning, of disease. The strong correlation found between baseline eGFR and future DKD in our study could be due to the fact that patients who developed DKD were already in the initial stages of the disease, although still clinically not diagnosed. The mean baseline eGFR for the diseased group was 80.33

mL/min/1.73m2, which corresponds to the G2 group of the KDIGO classification (3), significantly lower compared to 96.79 mL/min/1.73m2 for the group who will not develop CKD, which corresponds to the G1 KDIGO group. Furthermore, the AUCs for the associations of each phenotype in the study and the outcome showed that both age and baseline eGFR were the only ones capable of discriminating between patients who developed and who did not develop DKD, but only moderately well. There's still a need for biomarkers which can, alone or in groups, detect the disease even earlier, and more accurately predict diagnosis and monitor prognosis of DKD. A possible next step would be to set the threshold for the disease at a higher eGFR value than the one chosen in the study, which was 60 mL/min/1.73m2. This would make possible to detect the disease even earlier, and maybe a stronger association of DKD with DNA methylation than with eGFR would be shown.

When we tested the associations between DNA methylation of these 37 CpGs and the clinical phenotypes used for adjustment in the reference model (age, sex, BMI, and baseline eGFR), we saw that no variable was associated with all or the majority of the CpGs, and even when correlation was present, it was nor strong, with coefficients smaller then 0.5 in both directions. This indicates that age and eGFR may be more associated with the outcome (DKD) than with the main exposure (DNA methylation), and it helps us to corroborate our adjustments, as well as to inform about the importance of considering DNA methylation as a biomarker in addition to traditional risk factors, as has been argued from the beginning of this study.

Advancing on the discussion of this study's strengths and limitations, it is known that the sample size, effect size, level of significance (or alpha level), the variability of the outcome, follow-up time, loss to follow-up, confounding factors and type of statistical test can influence statistical power in a cohort study (80). Consequently, failure to adjust for confounding factors can lead to erroneous conclusions about the true relationship between the exposure and the outcome, reducing the validity and reliability of the results, and the overall statistical power (81). Yet, adjusting for too many covariates in a Cox regression model can lead to overfitting, which can result in biased estimates and loss of predictive power of the model. Overadjustment can also lead to loss of statistical power, as the model may be too complex, may result in a smaller sample size, and may lead to collinearity, introducing new confounders to the analysis (82). That is why the covariates to adjust the reference model in this study were carefully selected, and we came down to most relevant ones in the context. This, in addition to considerable size of our sample, as indicated in the *a priori* power calculation performed before the study began, indicates that the present study has significant statistical power to corroborate the results.

Moreover, it was already shown that patients with T2D are very heterogeneous and could be divided in four subgroups, and that the risk for different complications vary among those subgroups (35). In a similar rationale, it was shown that the eGFR trajectories vary between individuals with CKD (83), and between individuals with early DKD, whom present non-linear trajectories of eGFR decline during 5 years of follow-up (84), which means that the speed and slope of eGFR decline varies greatly among different patients. That said, the accuracy of current diagnostic and prognostic biomarkers may be limited to specific follow-up periods and particular populations. It would be important then, in the future, to aim at dividing the study population and analyzing the predictive value of DNA methylation based on T2D subgroups characteristics. Maybe DNA methylation is better at predicting a specific profile of patients, and the incorporation of DNA methylation analysis as a precision medicine tool in the management of DKD can help to identify high-risk patients and guide personalized treatment interventions, improving patient outcomes and reducing the disease burden.

Before concluding, some other things need to be taken into consideration when interpreting the results of the present work. The first is related to inflation, which in DNA methylation analysis refers to the overestimation of the significance of methylation differences between groups, and can occur due to technical or biological factors that can confound the analysis, leading to false-positive findings (85). There was an active effort to limit inflation in this study, correcting for batch effects, and for cell-type heterogeneity (86). Even though, the intrinsic heterogeneity of DNA methylation, which can vary widely between individuals or tissues, can make it challenging to detect subtle methylation differences between groups, leading to the overestimation of significance. The large sample size of the present study also helps to account for that, and multiple replication cohorts to validate findings across different populations would represent a further step to mitigate these and other issues and improve the accuracy and reliability of the results presented. Secondly is the fact that the present sample included only northern European patients, highlighting the need of external validation of the findings, also in populations with different ethnicities and demographic characteristics.

In summary, despite the fact that recent studies have shown that DNA methylation profiles associate with DKD, and can accurately predict the progression of the disease and identify patients who are at high risk of ESRD (22), to the knowledge of the author, no study so far has investigated the association between DNA methylation and future development of DKD in newly-diagnosed individuals with T2D. Our study found that DNA methylation at baseline is associated with future DKD in a population of newly diagnosed patients with T2D. There is a need, however, to refine the study design, investigating other CpGs of interest,

building methylation risk scores, performing prediction analysis, maybe classifying the population into subgroups, and validating the results in different populations. Nevertheless, the present results indicate that DNA methylation analysis may be a valuable precision medicine tool and biomarker in predicting DKD and guiding personalized treatment interventions for T2D patients.

## 5. Conclusion

This is the first study, to the author's knowledge, to assess and find associations between DNA methylation at baseline and the future development of DKD in newly diagnosed patients with T2D in a prospective cohort, indicating that DNA methylation could be a potential valuable biomarker to predict DKD at T2D diagnosis and to inform personalized treatment for T2D patients. Nonetheless, there is still a need to refine the study design, improve feature selection of the CpGs of interest, build risk scores, and validate the results in different populations in order to assure accuracy and validity of the results.

## 6. References


1. Association AD. 1. Improving Care and Promoting Health in Populations: Standards of Medical Care in Diabetes—2021. Diabetes Care. 2020;44(Supplement_1):S7-S14.
2. IDF. IDF Diabetes Atlas 10th edition. 2021.
3. 2. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes-2018. Diabetes Care. 2018;41(Suppl 1):S13-s27.
4. Galicia-Garcia U, Benito-Vicente A, Jebari S, Larrea-Sebal A, Siddiqi H, Uribe KB, et al. Pathophysiology of Type 2 Diabetes Mellitus. International Journal of Molecular Sciences. 2020;21(17):6275.
5. Chatterjee S, Khunti K, Davies MJ. Type 2 diabetes. The Lancet. 2017;389(10085):2239-51.
6. Alicic RZ, Rooney MT, Tuttle KR. Diabetic Kidney Disease: Challenges, Progress, and Possibilities. Clinical Journal of the American Society of Nephrology. 2017;12(12):2032-45.
7. Tuttle KR, Bakris GL, Bilous RW, Chiang JL, de Boer IH, Goldstein-Fuchs J, et al. Diabetic Kidney Disease: A Report From an ADA Consensus Conference. Diabetes Care. 2014;37(10):2864-83.

8. Doshi SM, Friedman AN. Diagnosis and Management of Type 2 Diabetic Kidney Disease. Clinical Journal of the American Society of Nephrology. 2017;12(8):1366-73.
9. Stephens JW, Brown KE, Min T. Chronic kidney disease in type 2 diabetes: Implications for managing glycaemic control, cardiovascular and renal risk. Diabetes, Obesity and Metabolism. 2020;22(S1):32-45.
10. Retnakaran R, Cull CA, Thorne KI, Adler AI, Holman RR. Risk factors for renal dysfunction in type 2 diabetes: U.K. Prospective Diabetes Study 74. Diabetes. 2006;55(6):1832-9.
11. Anders H-J, Huber TB, Isermann B, Schiffer M. CKD in diabetes: diabetic kidney disease versus nondiabetic kidney disease. Nature Reviews Nephrology. 2018;14(6):361-77.
12. Defronzo RA, Ferrannini E, Groop L, Henry RR, Herman WH, Holst JJ, et al. Type 2 diabetes mellitus. Nature Reviews Disease Primers. 2015;1(1):15019.
13. Shahwan MJ, Gacem SA, Zaidi SK. Prevalence of Diabetic Nephropathy and associated risk factors among type 2 diabetes mellitus patients in Ramallah, Palestine. Diabetes & Metabolic Syndrome: Clinical Research & Reviews. 2019;13(2):1491-6.
14. Traversi D, Pulliero A, Izzotti A, Franchitti E, Iacoviello L, Gianfagna F, et al. Precision Medicine and Public Health: New Challenges for Effective and Sustainable Health. Journal of Personalized Medicine. 2021;11(2):135.
15. Kumari P, Khan S, Wani IA, Gupta R, Verma S, Alam P, et al. Unravelling the Role of Epigenetic Modifications in Development and Reproduction of Angiosperms: A Critical Appraisal. Front Genet. 2022;13:819941.
16. Colhoun HM, Marcovecchio ML. Biomarkers of diabetic kidney disease. Diabetologia. 2018;61(5):996-1011.
17. Prasad RB, Groop L. Precision medicine in type 2 diabetes. Journal of Internal Medicine. 2019;285(1):40-8.
18. MacIsaac RJ, Ekinci EI, Jerums G. Markers of and Risk Factors for the Development and Progression of Diabetic Kidney Disease. American Journal of Kidney Diseases. 2014;63(2, Supplement 2):S39-S62.
19. Looker HC, Colombo M, Hess S, Brosnan MJ, Farran B, Dalton RN, et al. Biomarkers of rapid chronic kidney disease progression in type 2 diabetes. Kidney Int. 2015;88(4):888-96.
20. Khurana I, Kaipananickal H, Maxwell S, Birkelund S, Syreeni A, Forsblom C, et al. Reduced methylation correlates with diabetic nephropathy risk in type 1 diabetes. The Journal of Clinical Investigation. 2023;133(4).

21. Wu YL, Lin ZJ, Li CC, Lin X, Shan SK, Guo B, et al. Epigenetic regulation in metabolic diseases: mechanisms and advances in clinical study. Signal Transduct Target Ther. 2023;8(1):98.

22. Singh S, Sonkar SK, Sonkar GK, Mahdi AA. Diabetic kidney disease: A systematic review on the role of epigenetics as diagnostic and prognostic marker. Diabetes/Metabolism Research and Reviews. 2019;35(5).

23. Fraszczyk E, Spijkerman AMW, Zhang Y, Brandmaier S, Day FR, Zhou L, et al. Epigenome-wide association study of incident type 2 diabetes: a meta-analysis of five prospective European cohorts. Diabetologia. 2022;65(5):763-76.

24. Chambers JC, Loh M, Lehne B, Drong A, Kriebel J, Motta V, 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(7):526-34.

25. Volkov P, Bacos K, Ofori JK, Esguerra JL, Eliasson L, Rönn T, et al. Whole-Genome Bisulfite Sequencing of Human Pancreatic Islets Reveals Novel Differentially Methylated Regions in Type 2 Diabetes Pathogenesis. Diabetes. 2017;66(4):1074-85.

26. García-Calzón S, Perfilyev A, Martinell M, Ustinova M, Kalamajski S, Franks PW, et al. Epigenetic markers associated with metformin response and intolerance in drug-naïve patients with type 2 diabetes. Sci Transl Med. 2020;12(561).

27. Ingrosso D, Perna AF. DNA Methylation Dysfunction in Chronic Kidney Disease. Genes (Basel). 2020;11(7).

28. Larkin BP, Glastras SJ, Chen H, Pollock CA, Saad S. DNA methylation and the potential role of demethylating agents in prevention of progressive chronic kidney disease. Faseb j. 2018;32(10):5215-26.

29. Chu AY, Tin A, Schlosser P, Ko Y-A, Qiu C, Yao C, et al. Epigenome-wide association studies identify DNA methylation associated with kidney function. Nature Communications. 2017;8(1):1286.

30. Chen XJ, Zhang H, Yang F, Liu Y, Chen G. DNA Methylation Sustains "Inflamed" Memory of Peripheral Immune Cells Aggravating Kidney Inflammatory Response in Chronic Kidney Disease. Front Physiol. 2021;12:637480.

31. Aldemir O, Turgut F, Gokce C. The association between methylation levels of targeted genes and albuminuria in patients with early diabetic kidney disease. Ren Fail. 2017;39(1):597-601.

32. Smyth LJ, McKay GJ, Maxwell AP, McKnight AJ. DNA hypermethylation and DNA hypomethylation is present at different loci in chronic kidney disease. Epigenetics. 2014;9(3):366-76.

33. Kim H, Bae JH, Park KS, Sung J, Kwak SH. DNA Methylation Changes Associated With Type 2 Diabetes and Diabetic Kidney Disease in an East Asian Population. J Clin Endocrinol Metab. 2021;106(10):e3837-e51.

34. Sapienza C, Lee J, Powell J, Erinle O, Yafai F, Reichert J, et al. DNA methylation profiling identifies epigenetic differences between diabetes patients with ESRD and diabetes patients without nephropathy. Epigenetics. 2011;6(1):20-8.

35. Schrader S, Perfilyev A, Ahlqvist E, Groop L, Vaag A, Martinell M, et al. Novel Subgroups of Type 2 Diabetes Display Different Epigenetic Patterns That Associate With Future Diabetic Complications. Diabetes Care. 2022;45(7):1621-30.

36. Tummalapalli L, Nadkarni GN, Coca SG. Biomarkers for predicting outcomes in chronic kidney disease. Curr Opin Nephrol Hypertens. 2016;25(6):480-6.

37. Dunkler D, Gao P, Lee SF, Heinze G, Clase CM, Tobe S, et al. Risk Prediction for Early CKD in Type 2 Diabetes. Clin J Am Soc Nephrol. 2015;10(8):1371-9.

38. LUDC. ANDIS - All New Diabetics in Skåne Sweden2022 [updated 2022-12-12. Available from: http://andis.ludc.med.lu.se/.

39. Ahlqvist E, Storm P, Käräjämäki A, Martinell M, Dorkhan M, Carlsson 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(5):361-9.

40. Levey AS, Coresh J, Greene T, Stevens LA, Zhang YL, Hendriksen S, 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(4):247-54.

41. Levy JC, Matthews DR, Hermans MP. Correct homeostasis model assessment (HOMA) evaluation uses the computer program. Diabetes Care. 1998;21(12):2191-2.

42. Oxford Uo. HOMA2 Calculator 2004 [Available from: https://www.dtu.ox.ac.uk/homacalculator/.

43. KDIGO. KDIGO 2012 Clinical Practice Guideline for the Evaluation and Management of

Chronic Kidney Disease. Official Journal of the International Society of Nephrology. 2013;3(1):163.

44. Moran S, Arribas C, Esteller M. Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhancer sequences. Epigenomics. 2016;8(3):389-99.

45. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome Biology. 2004;5(10):R80.

46. McCartney DL, Walker RM, Morris SW, McIntosh AM, Porteous DJ, Evans KL. Identification of polymorphic and off-target probe binding sites on the Illumina Infinium MethylationEPIC BeadChip. Genom Data. 2016;9:22-4.

47. Du P, Zhang X, Huang C-C, Jafari N, Kibbe WA, Hou L, et al. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. BMC Bioinformatics. 2010;11(1):587.

48. Du P, Kibbe WA, Lin SM. lumi: a pipeline for processing Illumina microarray. Bioinformatics. 2008;24(13):1547-8.

49. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics. 2007;8(1):118-27.

50. Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, et al. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. Bioinformatics. 2013;29(2):189-96.

51. Gervin K, Salas LA, Bakulski KM, van Zelm MC, Koestler DC, Wiencke JK, et al. Systematic evaluation and validation of reference and library selection methods for deconvolution of cord blood DNA methylation data. Clinical Epigenetics. 2019;11(1):125.

52. LLC HC. Calculate Sample Size Needed to Test Time-To-Event Data: Cox PH, 2-Sided Equality Atlanta, GA2013-2023 [cited 2023 Apr 14]. Available from: http://powerandsamplesize.com/Calculators/Test-Time-To-Event-Data/Cox-PH-2-Sided-Equality.

53. Madjarova SJ, Pareek A, Eckhardt CM, Khorana A, Kunze KN, Ollivier M, et al. Fragility Part I: a guide to understanding statistical power. Knee Surgery, Sports Traumatology, Arthroscopy. 2022;30(12):3924-8.

54. Thomas MC, Brownlee M, Susztak K, Sharma K, Jandeleit-Dahm KAM, Zoungas S, et al. Diabetic kidney disease. Nature Reviews Disease Primers. 2015;1(1):15018.

55. Wang X, Wu H, Yang G, Xiang J, Xiong L, Zhao L, et al. REG1A and RUNX3 Are Potential Biomarkers for Predicting the Risk of Diabetic Kidney Disease. Frontiers in Endocrinology. 2022;13.

56. Kleinbaum DG, Klein M. Survival Analysis: A Self-Learning Text. 3rd edition ed. London: Springer; 2012.
57. Dawes K, Andersen A, Reimer R, Mills JA, Hoffman E, Long JD, et al. The relationship of smoking to cg05575921 methylation in blood and saliva DNA samples from several studies. Scientific Reports. 2021;11(1):21627.
58. Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, et al. The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. Current Protocols in Bioinformatics. 2016;54(1):1.30.1-1..3.
59. LaBarre BA, Goncearenco A, Petrykowska HM, Jaratlerdsiri W, Bornman MSR, Hayes VM, et al. MethylToSNP: identifying SNPs in Illumina DNA methylation array data. Epigenetics & Chromatin. 2019;12(1):79.
60. Yang Y, Peng X, Ying P, Tian J, Li J, Ke J, et al. AWESOME: a database of SNPs that affect protein post-translational modifications. Nucleic Acids Research. 2018;47(D1):D874-D80.
61. Charames GS, Sabatini P, Watkins N. Chapter 5 - Clinical and genetic evidence and population evidence. In: Lázaro C, Lerner-Ellis J, Spurdle A, editors. Clinical DNA Variant Interpretation: Academic Press; 2021. p. 59-87.
62. Penno G, Solini A, Orsi E, Bonora E, Fondelli C, Trevisan R, et al. Insulin resistance, diabetic kidney disease, and all-cause mortality in individuals with type 2 diabetes: a prospective cohort study. BMC Medicine. 2021;19(1):66.
63. Gao B, Wu S, Wang J, Yang C, Chen S, Hou J, et al. Clinical features and long-term outcomes of diabetic kidney disease – A prospective cohort study from China. Journal of Diabetes and its Complications. 2019;33(1):39-45.
64. Hsieh Y-T, Hsieh M-C. Time-sequential correlations between diabetic kidney disease and diabetic retinopathy in type 2 diabetes – an 8-year prospective cohort study. Acta Ophthalmologica. 2021;99(1):e1-e6.
65. Yakush Williams JK. Management Strategies for Patients with Diabetic Kidney Disease and Chronic Kidney Disease in Diabetes. Nursing Clinics of North America. 2017;52(4):575-87.
66. Sepulveda AR, Jones D, Ogino S, Samowitz W, Gulley ML, Edwards R, et al. CpG methylation analysis--current status of clinical assays and potential applications in molecular diagnostics: a report of the Association for Molecular Pathology. J Mol Diagn. 2009;11(4):266-78.

67. Smyth LJ, Dahlström EH, Syreeni A, Kerr K, Kilner J, Doyle R, et al. Epigenome-wide meta-analysis identifies DNA methylation biomarkers associated with diabetic kidney disease. Nature Communications. 2022;13(1):7891.
68. Rothman KJ, Greenland S, Lash TL. Modern Epidemiology. 3rd edition ed. Philadelphia, USA: LIPPINCOTT WILLIAMS & WILKINS; 2008.
69. Rico-Fontalvo J, Aroca-Martínez G, Daza-Arnedo R, Cabrales J, Rodríguez-Yanez T, Cardona-Blanco M, et al. Novel Biomarkers of Diabetic Kidney Disease. Biomolecules. 2023;13(4):633.
70. Zoghbi HY, Beaudet AL. Epigenetics and Human Disease. Cold Spring Harb Perspect Biol. 2016;8(2):a019497.
71. Santamaria S, De Groot R. ADAMTS proteases in cardiovascular physiology and disease. Open Biology. 2020;10(12):200333.
72. Farooqui A, Alhazmi A, Haque S, Tamkeen N, Mehmankhah M, Tazyeen S, et al. Network-based analysis of key regulatory genes implicated in Type 2 Diabetes Mellitus and Recurrent Miscarriages in Turner Syndrome. Sci Rep. 2021;11(1):10662.
73. Mirzaeicheshmeh E, Zerrweck C, Centeno-Cruz F, Baca-Peynado P, Martinez-Hernandez A, García-Ortiz H, et al. Alterations of DNA methylation during adipogenesis differentiation of mesenchymal stem cells isolated from adipose tissue of patients with obesity is associated with type 2 diabetes. Adipocyte. 2021;10(1):493-504.
74. Wang L, Wang HL, Liu TT, Lan HY. TGF-Beta as a Master Regulator of Diabetic Nephropathy. Int J Mol Sci. 2021;22(15).
75. Voelker J, Berg PH, Sheetz M, Duffin K, Shen T, Moser B, et al. Anti-TGF-β1 Antibody Therapy in Patients with Diabetic Nephropathy. J Am Soc Nephrol. 2017;28(3):953-62.
76. Tavener SK, Jewell DE, Panickar KS. The Increase in Circulating Levels of Pro-Inflammatory Chemokines, Cytokines, and Complement C5 in Canines with Impaired Kidney Function. Curr Issues Mol Biol. 2022;44(4):1664-76.
77. Simon P, Krajcoviechová A, Harvey F, Haloui M, Marois-Blanchet F-C, Chalmers J, et al. YIA 03-01 VALIDATION OF GENETIC DETERMINANTS OF UNMET NEEDS IN THE TREATMENT OF KIDNEY DISEASE IN TYPE-2 DIABETES. Journal of Hypertension. 2016;34:e204.
78. Horvath S. DNA methylation age of human tissues and cell types. Genome Biology. 2013;14(10):3156.
79. Iwagami M, Shinozaki T. Introduction to Matching in Case-Control and Cohort Studies. Annals of Clinical Epidemiology. 2022;4(2):33-40.

80. Norton BJ, Strube MJ. Understanding statistical power. J Orthop Sports Phys Ther. 2001;31(6):307-15.
81. VanderWeele TJ. Explanation in causal inference: developments in mediation and interaction. Int J Epidemiol. 2016;45(6):1904-8.
82. E. F, Jr. H. Regression Modeling Strategies. 2 ed. Switzerland: Springer Cham; 2015.
83. Weng SC, Chen CM, Chen YC, Wu MJ, Tarng DC. Trajectory of Estimated Glomerular Filtration Rate and Malnourishment Predict Mortality and Kidney Failure in Older Adults With Chronic Kidney Disease. Front Med (Lausanne). 2021;8:760391.
84. Kerschbaum J, Rudnicki M, Dzien A, Dzien-Bischinger C, Winner H, Heerspink HL, et al. Intra-individual variability of eGFR trajectories in early diabetic kidney disease and lack of performance of prognostic biomarkers. Scientific Reports. 2020;10(1):19743.
85. van Iterson M, van Zwet EW, Heijmans BT, the BC. Controlling bias and inflation in epigenome- and transcriptome-wide association studies using the empirical null distribution. Genome Biology. 2017;18(1):19.
86. Rahmani E, Zaitlen N, Baran Y, Eng C, Hu D, Galanter J, et al. Correcting for cell-type heterogeneity in DNA methylation: a comprehensive evaluation. Nat Methods. 2017;14(3):218-9.
87. Inc. BT. BioRender c2021 [cited 2023 22 Apr]. Available from: https://biorender.com.

## 7. Tables

**Table 1.** Demographic and clinical characteristics of the sample. Total sample (ANDIU + ANDIS patients) of 487 individuals.

| Phenotype | Prevalence (%) / Mean (SD) | | Min. to Max. | | Missing values | | P value |
|---|---|---|---|---|---|---|---|
| | Patients who did not develop DKD (n=399) | Patients who developed DKD (n=88) | Patients who did not develop DKD (n=399) | Patients who developed CKD (n=88) | Patients who did not develop DKD (n=399) | Patients who developed DKD (n=88) | |
| **Time to event (days)** | - | 1566.5 (830.9) | - | 281-3581 | - | 0 | - |
| **Follow-up (days)** | 2380 (798.9) | - | 307-4202 | - | 0 | - | - |
| **Sex (male)** | 237 (59.4%) | 51 (57.9%) | - | - | 0 | 0 | 0.8036 |
| **Age (years)** | 60.38 (8.9) | 67.69 (8.5) | 46-85 | 46-90 | 0 | 0 | **7.063e-11** |
| **BMI (kg/m$^2$)** | 31.41 (5.4) | 30.49 (4.4) | 20.38-50.71 | 23.12-43.66 | 0 | 0 | 0.1276 |
| **Duration of diabetes at baseline (days)*** | 52.35 (118.8) | 62.98 (93.3) | 0-1342 | 0-519 | 0 | 0 | **0.01498** |
| **HbA1C (mmol/mol)** | 62.47 (17.6) | 60.40 (15.9) | 34-144 | 33.34-107 | 0 | 0 | 0.3922 |
| **Baseline eGFR (mL/min/1.73m$^2$)** | 96.79 (20.6) | 80.33 (16.4) | 60.82-182.47 | 60.09- 156.65 | 0 | 0 | **2.59e-13** |
| **First eGFR below 60 (mL/min/1.73m$^2$)** | - | 53.5 (6.5) | - | 28.82 – 59.97 | - | 0 | - |
| **Second eGFR below 60 (mL/min/1.73m$^2$)**** | - | 51.25 (9.3) | - | 10.06 -59.93 | - | 0 | - |
| **Urinary albumin/Urinary creatinine ratio (mg/mmol)** | 1.8 (5.2) | 5.4 (22.2) | 0-66 | 0-182 | 19 | 7 | 0.4242 |
| **Creatinine (mmol/L)** | 67.87 (13.9) | 77.49 (14.6) | 33-105 | 47-112 | 0 | 0 | **1.075e-07** |
| **Urinary Creatinine (mmol/L)** | 11 (5.4) | 9.9 (5.1) | 1-30.3 | 3-31.9 | 15 | 6 | 0.05495 |
| **Urinary Albumin (mg/L)** | 18.66 (73.9) | 37.78 (128.1) | 0-1225.9 | 0-1073.8 | 13 | 2 | 0.584 |
| **HOMA2-IR** | 3.9 (2.9) | 3.6 (1.3) | 0.99-47.62 | 1.1-9.8 | 11 | 0 | 0.8677 |
| **Triglycerides (mmol/L)** | 2.4 (3.5) | 2.1 (1.2) | 0.5-56.2 | 0.56-8.4 | 13 | 2 | 0.993 |
| **Total Cholesterol (mmol/L)** | 5.3 (1.2) | 5.1 (1.3) | 2.3-14.7 | 2.8-9.4 | 6 | 1 | 0.152 |
| **HDL (mmol/L)** | 1.2 (0.3) | 1.1 (0.3) | 0.41-2.8 | 0.39-2.4 | 6 | 1 | 0.3205 |
| **LDL ( mmol/L)** | 3.4 (1.05) | 3.3 (1.1) | 0.8-7.6 | 1.3-7.2 | 9 | 1 | 0.06739 |
| **Antihypertensives at baseline *** ** | 222 (55.6%) | 65 (73.9%) | - | - | 0 | 0 | **0.001679** |
| **Statins at baseline *** ** | 172 (43.1%) | 42 (47.7%) | - | - | 0 | 0 | 0.4301 |
| **Metformin at baseline *** ** | 337 (84.5%) | 72 (81.8%) | - | - | 0 | 0 | 0.5415 |
| **Previous CVD event ^** | 32 (8%) | 14 (15.9%) | - | - | 0 | 0 | **0.02218** |

Significant p values are shown in bold. P<0.05 show a statistically significance difference between groups. All variables refer to baseline status except of the first and second eGFR measurements and time to event for cases and follow-up for controls.

*Duration of diabetes at baseline: days from the diabetes diagnosis to the DNA methylation sample. ** Second eGFR below 60 is the second altered eGFR measurement with an interval of more than 3 months from the first (when we closed the criteria for the disease). ***Antihypertensives, statins and metformin at baseline: yes for all the patients who bought the medications at least once within baseline (six months before or six months after DNA methylation sample date. ^ Previous CVD event represents all the patients with reported CVD events before or during baseline, up to six months after DNA methylation sample date. SD: Standard deviation

**Table 2.** Weighted-cox regression models to assess the relationship between DNA methylation and the development of DKD in 487 type 2 diabetes patients from the ANDIU and ANDIS cohorts. The table shows the number of CpG sites that are significant for each model, considering p and q (FDR adjusted p) values. N indicates the sample size used for each specific model.

| Models (weighted-cox) | q<0.05 | q<0.1 | q<0.15 | p<0.05 | N |
|---|---|---|---|---|---|
| 1a - Surv( Time_to_DKD_Follow_up , DKD ) ~ methylation + Age + BMI + Baseline_eGFR + Sex | 56,411 | 71,460 | 84,952 | 129,811 | 487 |
| 1b - Surv( Time_to_DKD_Follow_up , DKD ) ~ methylation + Age + BMI + Baseline_eGFR + CD8T + CD4T + NK + Bcell + Mono + Sex | 5,086 | 11,040 | 19,605 | 117,598 | 487 |
| 2 - Surv( Time_to_DKD_Follow_up , DKD ) ~ methylation + Age + BMI + Baseline_eGFR + Diabetes_duration_at_baseline + Sex | 56,968 | 72,576 | 86,315 | 131,194 | 487 |
| 3 - Surv( Time_to_DKD_Follow_up , DKD ) ~ methylation + Age + BMI + Baseline_eGFR + Diabetes_duration_at_baseline + HBA1C + Sex | 59,373 | 75,444 | 89,358 | 133,821 | 487 |
| 4 - Surv( Time_to_DKD_Follow_up , DKD ) ~ methylation + Age + BMI + Baseline_eGFR + Diabetes_duration_at_baseline + HBA1C + UACR + Sex | 56,144 | 70,546 | 82,925 | 127,661 | 461 |
| 5 - Surv( Time_to_DKD_Follow_up , DKD ) ~ methylation + Age + BMI + Baseline_eGFR + Diabetes_duration_at_baseline + HBA1C + Sex + Statins_At_Baseline + Antihypertensive_At_Baseline | 57,559 | 74,925 | 90,370 | 136,946 | 487 |
| 6 - Surv( Time_to_DKD_Follow_up , DKD ) ~ methylation + Age + BMI + Baseline_eGFR + Diabetes_duration_at_baseline + HBA1C + Total_Cholesterol + LDL + HDL + Sex + Antihypertensive_At_Baseline + Smoking (cg05575921) | 2,862 | 6,794 | 43,504 | 121,801 | 476 |
| 7 - Surv( Time_to_DKD_Follow_up , DKD ) ~ methylation + Age + BMI + Baseline_eGFR + Diabetes_duration_at_baseline + HBA1C + Total_Cholesterol + LDL + HDL + HOMA2IR + Sex + Antihypertensive_At_Baseline + CVD_At_Baseline + Smoking (cg05575921) | 3,833 | 34,008 | 63,405 | 129,766 | 465 |

* Model 1b is the model 1a adjusted for cell composition.

**Table 3.** Description of the identified 37 CpGs with difference in methylation > 5% between patients with T2D who developed DKD and patients with T2D who did not develop DKD in a subset of 487 patients from the ANDIS and ANDIU cohorts, with a significance level of q < 0.05 in Model 1a, adjusted for age, sex, BMI and baseline eGFR, and p < 0.05 in all the other models.

| TargetID | CHR | Gene symbol, Illumina standard manifest | Relation to gene region, Illumina standard manifest | Relation to CpG Island, Illumina standard manifest | Gene symbol, hg38, GENCODE version 12 | Mean methylation beta values (patients who did not develop DKD, n = 399) | SD (patients who did not develop DKD) | Mean methylation beta values (patients who developed DKD, n= 88) | SD (patients who developed DKD) | Difference in methylation between groups |
|---|---|---|---|---|---|---|---|---|---|---|
| cg08641951 | 5 | *ADAMTS16* | Body | | | 0.24 | 0.08 | 0.19 | 0.07 | -0.05 |
| cg11940040 | 17 | *RPH3AL* | Body | S_Shore | *RPH3AL* | 0.38 | 0.10 | 0.33 | 0.10 | -0.05 |
| cg23480021 | 3 | | | N_Shore | *ZNF385D* | 0.61 | 0.14 | 0.66 | 0.12 | 0.05 |
| cg11065575 | 4 | *SHISA3* | 3'UTR | S_Shelf | *SHISA3* | 0.66 | 0.09 | 0.61 | 0.11 | -0.05 |
| cg25291037 | 2 | | | | | 0.37 | 0.20 | 0.48 | 0.21 | 0.11 |
| cg01079515 | 3 | | | | *AC124944.5;AC124944.5* | 0.79 | 0.18 | 0.86 | 0.12 | 0.072 |
| cg05074631 | 13 | *DZIP1;DZIP1* | TSS1500;TSS1500 | S_Shore | *DZIP1;DZIP1;DZIP1* | 0.60 | 0.14 | 0.54 | 0.17 | -0.06 |
| cg00169354 | 20 | | | S_Shelf | | 0.47 | 0.14 | 0.54 | 0.15 | 0.07 |
| cg18826637 | 2 | | | | | 0.42 | 0.08 | 0.36 | 0.08 | -0.06 |
| cg22297055 | 17 | *SLC43A2;SLC43A2; SLC43A2* | TSS200;Body;Body | N_Shore | *SLC43A2;SLC43A2* | 0.47 | 0.14 | 0.53 | 0.16 | 0.06 |
| cg17386240 | 5 | *TGFBI* | Body | | *TGFBI* | 0.55 | 0.24 | 0.64 | 0.19 | 0.09 |
| cg26705599 | 13 | | | Island | | 0.53 | 0.30 | 0.60 | 0.27 | 0.07 |
| cg12183150 | 2 | | | | *AC097713.4;AC097713.3* | 0.45 | 0.12 | 0.50 | 0.12 | 0.05 |
| cg10187601 | 7 | *ST7;ST7* | Body;Body | | *ST7* | 0.71 | 0.17 | 0.78 | 0.13 | 0.07 |
| cg07784975 | 12 | | | | | 0.31 | 0.08 | 0.26 | 0.08 | -0.05 |
| cg10482512 | 6 | *CCR6* | TSS1500 | | *CCR6* | 0.41 | 0.24 | 0.52 | 0.24 | 0.10 |
| cg26365090 | 20 | *TOX2;TOX2;TOX2; TOX2* | 5'UTR;Body;TSS200;5'UTR | | *TOX2;TOX2;TOX2;TOX2; TOX2* | 0.34 | 0.31 | 0.25 | 0.25 | -0.08 |
| cg22273830 | 17 | *SLC43A2;SLC43A2; SLC43A2* | TSS200;Body;Body | N_Shore | *SLC43A2;SLC43A2* | 0.40 | 0.17 | 0.46 | 0.18 | 0.07 |
| cg11076954 | 17 | *SLC43A2* | Body | N_Shore | *SLC43A2;SLC43A2* | 0.40 | 0.14 | 0.45 | 0.15 | 0.05 |
| cg11420142 | 8 | | | | *RP11-122C21.1* | 0.42 | 0.30 | 0.55 | 0.28 | 0.13 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| cg13989295 | 17 | *SKA2;SKA2* | 3'UTR;3'UTR | S_Shelf | *AC099850.1;SKA2;SKA2* | 0.38 | 0.29 | 0.46 | 0.30 | 0.08 |
| cg08230244 | 5 | *FAM13B;FAM13B; FAM13B* | Body;Body;Body | | | 0.83 | 0.22 | 0.77 | 0.23 | -0.06 |
| cg16788857 | 3 | *C3orf59* | Body | | | 0.52 | 0.24 | 0.46 | 0.24 | -0.07 |
| cg25621215 | 20 | | | | *RP11-445H22.3* | 0.88 | 0.13 | 0.81 | 0.22 | -0.07 |
| cg26423139 | 17 | *SLC43A2;SLC43A2; SLC43A2* | TSS200;Body;Body | N_Shore | *SLC43A2;SLC43A2* | 0.46 | 0.14 | 0.52 | 0.15 | 0.06 |
| cg04706137 | 11 | | | | *RP11-867G2.8;RP11-867G2.8* | 0.75 | 0.16 | 0.69 | 0.21 | -0.06 |
| cg19618634 | 17 | | | | | 0.40 | 0.22 | 0.47 | 0.20 | 0.06 |
| cg10499451 | 19 | *PAPL* | Body | S_Shelf | | 0.76 | 0.25 | 0.71 | 0.27 | -0.05 |
| cg23246911 | 17 | *RPH3AL* | Body | N_Shelf | *RPH3AL* | 0.37 | 0.15 | 0.32 | 0.14 | -0.05 |
| cg02427933 | 20 | | | | *RP11-445H22.3* | 0.25 | 0.21 | 0.33 | 0.24 | 0.08 |
| cg08586441 | 4 | *TEC* | Body | | | 0.52 | 0.21 | 0.58 | 0.19 | 0.06 |
| cg04885581 | 14 | *FRMD6* | 5'UTR | N_Shelf | *FRMD6* | 0.83 | 0.10 | 0.77 | 0.15 | -0.06 |
| cg12434901 | 17 | *KCNH6;KCNH6* | Body;Body | | | 0.62 | 0.28 | 0.71 | 0.25 | 0.09 |
| cg11128983 | 5 | *CPLX2;CPLX2* | TSS1500;5'UTR | N_Shore | *CPLX2;CPLX2;CPLX2;CPLX2* | 0.66 | 0.31 | 0.76 | 0.26 | 0.10 |
| cg07505631 | 3 | *CDCP1;CDCP1* | Body;Body | | *CDCP1* | 0.68 | 0.26 | 0.56 | 0.29 | -0.12 |
| cg14924512 | 9 | *EHMT1;EHMT1* | Body;Body | | *EHMT1* | 0.78 | 0.23 | 0.71 | 0.26 | -0.07 |
| cg07917191 | 2 | *SLC9A4* | Body | | | 0.82 | 0.20 | 0.77 | 0.24 | -0.05 |

* CHR: Chromosome. ** SD: standard deviation

**Table 4.** Results for the weighted cox regression model 1a, adjusted for age, sex, BMI, and baseline eGFR of the 37 CpGs with difference in methylation > 5% between patients with T2D who developed DKD and patients with T2D who did not develop DKD in a subset of 487 patients from the ANDIS and ANDIU cohorts, with a significance level of $q < 0.05$ in Model 1a, and $p < 0.05$ in all the other models.

| | Methylation | | | Age | | | BMI | | | Baseline eGFR | | | Sex | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Target ID** | **p-value** | **q-value** | **HR (95% CI) *** | **p-value** | **q-value** | **HR (95% CI) *** | **p-value** | **q-value** | **HR (95% CI) *** | **p-value** | **q-value** | **HR (95% CI) *** | **p-value** | **q-value** | **HR (95% CI) *** |
| cg08641951 | 6.0E-06 | 5.6E-04 | 0.0005 (0.00002 - 0.01) | 1.1E-03 | 1.1E-03 | 1.05 (1.02 - 1.09 ) | 5.1E-01 | 5.1E-01 | 0.98 ( 0.94 - 1.03 ) | 6.9E-05 | 7.0E-05 | 0.96 ( 0.93 - 0.98 ) | 5.0E-01 | 5.3E-01 | 1.19 ( 0.72 - 1.98 ) |
| cg11940040 | 3.6E-05 | 1.6E-03 | 0.003 (0.0002 - 0.05 ) | 2.7E-05 | 3.1E-04 | 1.07 ( 1.04 - 1.11 ) | 3.6E-01 | 3.6E-01 | 0.98 ( 0.93 - 1.02 ) | 9.3E-06 | 3.9E-05 | 0.95 ( 0.93 - 0.97 ) | 2.4E-01 | 5.1E-01 | 1.34 ( 0.82 - 2.2 ) |
| cg23480021 | 4.4E-05 | 1.8E-03 | 26.32 (5.48 - 126.3) | 1.3E-05 | 3.1E-04 | 1.07 ( 1.04 - 1.11 ) | 1.9E-01 | 3.1E-01 | 0.97 ( 0.93 - 1.01 ) | 5.1E-05 | 5.2E-05 | 0.96 ( 0.94 - 0.98 ) | 1.9E-01 | 5.1E-01 | 1.39 ( 0.85 - 2.26 ) |
| cg11065575 | 4.7E-05 | 1.9E-03 | 0.01 (0.001 - 0.08) | 5.6E-05 | 3.1E-04 | 1.06 ( 1.03 - 1.1 ) | 2.0E-01 | 3.1E-01 | 0.97 ( 0.92 - 1.02 ) | 6.8E-05 | 6.9E-05 | 0.96 ( 0.94 - 0.98 ) | 6.6E-01 | 6.7E-01 | 0.89 ( 0.53 - 1.5 ) |
| cg25291037 | 5.0E-05 | 2.0E-03 | 9.98 (3.28 - 30.34) | 2.2E-04 | 3.1E-04 | 1.06 ( 1.03 - 1.09 ) | 2.2E-01 | 3.1E-01 | 0.97 ( 0.93 - 1.02 ) | 3.0E-05 | 3.9E-05 | 0.95 ( 0.93 - 0.98 ) | 2.4E-01 | 5.1E-01 | 1.34 ( 0.82 - 2.19 ) |
| cg01079515 | 6.0E-05 | 2.2E-03 | 24.8 (5.17 - 118.98) | 1.8E-04 | 3.1E-04 | 1.07 ( 1.03 - 1.1 ) | 1.5E-01 | 3.1E-01 | 0.96 ( 0.92 - 1.01 ) | 2.3E-05 | 3.9E-05 | 0.96 ( 0.93 - 0.98 ) | 4.7E-01 | 5.1E-01 | 1.2 ( 0.73 - 1.97 ) |
| cg05074631 | 7.0E-05 | 2.5E-03 | 0.08 (0.02 - 0.28) | 4.5E-04 | 4.7E-04 | 1.06 ( 1.02 - 1.09 ) | 2.1E-01 | 3.1E-01 | 0.97 ( 0.93 - 1.02 ) | 1.1E-05 | 3.9E-05 | 0.95 ( 0.93 - 0.97 ) | 4.4E-01 | 5.1E-01 | 1.22 ( 0.74 - 1.99 ) |
| cg00169354 | 1.4E-04 | 4.1E-03 | 12.41 (3.4 - 45.32) | 2.1E-04 | 3.1E-04 | 1.06 ( 1.03 - 1.09 ) | 5.1E-01 | 5.1E-01 | 0.98 ( 0.94 - 1.03 ) | 8.6E-05 | 8.6E-05 | 0.96 ( 0.94 - 0.98 ) | 3.3E-01 | 5.1E-01 | 1.28 ( 0.78 - 2.1 ) |
| cg18826637 | 1.7E-04 | 4.8E-03 | 0.003 (0.0002 - 0.07) | 7.5E-03 | 7.5E-03 | 1.05 ( 1.01 - 1.08 ) | 4.1E-01 | 4.1E-01 | 0.98 ( 0.94 - 1.03 ) | 1.7E-05 | 3.9E-05 | 0.95 ( 0.93 - 0.97 ) | 6.2E-01 | 6.3E-01 | 1.13 ( 0.69 - 1.86 ) |
| cg22297055 | 2.9E-04 | 7.2E-03 | 13.5 (3.31 - 55.12) | 1.2E-04 | 3.1E-04 | 1.06 ( 1.03 - 1.1 ) | 2.4E-01 | 3.1E-01 | 0.97 ( 0.92 - 1.02 ) | 3.6E-05 | 4.0E-05 | 0.95 ( 0.93 - 0.98 ) | 2.7E-01 | 5.1E-01 | 1.32 ( 0.81 - 2.15 ) |
| cg17386240 | 4.1E-04 | 9.4E-03 | 8.44 (2.59 - 27.5) | 3.3E-05 | 3.1E-04 | 1.07 ( 1.04 - 1.1 ) | 3.4E-01 | 3.4E-01 | 0.98 ( 0.93 - 1.02 ) | 4.6E-05 | 4.8E-05 | 0.96 ( 0.93 - 0.98 ) | 3.8E-01 | 5.1E-01 | 1.25 ( 0.77 - 2.03 ) |
| cg26705599 | 4.5E-04 | 1.0E-02 | 4.17 (1.88 - 9.26) | 6.2E-05 | 3.1E-04 | 1.06 ( 1.03 - 1.09 ) | 1.8E-01 | 3.1E-01 | 0.97 ( 0.92 - 1.02 ) | 1.4E-05 | 3.9E-05 | 0.95 ( 0.93 - 0.97 ) | 2.9E-01 | 5.1E-01 | 1.3 ( 0.8 - 2.13 ) |
| cg12183150 | 4.8E-04 | 1.1E-02 | 37.05 (4.88 - 281.28) | 4.5E-04 | 4.7E-04 | 1.06 ( 1.03 - 1.09 ) | 1.9E-01 | 3.1E-01 | 0.97 ( 0.92 - 1.02 ) | 1.8E-05 | 3.9E-05 | 0.95 ( 0.93 - 0.97 ) | 6.1E-01 | 6.2E-01 | 1.14 ( 0.69 - 1.88 ) |
| cg10187601 | 4.9E-04 | 1.1E-02 | 25.93 (4.15 - 161.82) | 2.6E-05 | 3.1E-04 | 1.07 ( 1.04 - 1.11 ) | 3.3E-01 | 3.4E-01 | 0.98 ( 0.93 - 1.02 ) | 3.7E-05 | 4.1E-05 | 0.96 ( 0.94 - 0.98 ) | 3.5E-01 | 5.1E-01 | 1.27 ( 0.77 - 2.09 ) |
| cg07784975 | 5.3E-04 | 1.2E-02 | 0.01 (0.0005 - 0.12) | 9.8E-04 | 9.8E-04 | 1.06 ( 1.02 - 1.09 ) | 2.9E-01 | 3.1E-01 | 0.97 ( 0.93 - 1.02 ) | 4.1E-05 | 4.4E-05 | 0.96 ( 0.94 - 0.98 ) | 1.5E-01 | 5.1E-01 | 1.44 ( 0.87 - 2.38 ) |
| cg10482512 | 7.9E-04 | 1.6E-02 | 6.59 (2.19 - 19.8) | 1.7E-04 | 3.1E-04 | 1.06 ( 1.03 - 1.09 ) | 2.5E-01 | 3.1E-01 | 0.97 ( 0.93 - 1.02 ) | 1.0E-05 | 3.9E-05 | 0.95 ( 0.93 - 0.97 ) | 4.6E-01 | 5.1E-01 | 1.2 ( 0.73 - 1.98 ) |
| cg26365090 | 8.5E-04 | 1.7E-02 | 0.2 (0.08 - 0.51) | 2.0E-05 | 3.1E-04 | 1.07 ( 1.04 - 1.1 ) | 2.7E-01 | 3.1E-01 | 0.98 ( 0.93 - 1.02 ) | 9.5E-06 | 3.9E-05 | 0.95 ( 0.93 - 0.97 ) | 1.6E-01 | 5.1E-01 | 1.43 ( 0.86 - 2.36 ) |
| cg22273830 | 9.3E-04 | 1.8E-02 | 7.95 (2.33 - 27.15) | 1.0E-04 | 3.1E-04 | 1.06 ( 1.03 - 1.1 ) | 2.5E-01 | 3.1E-01 | 0.97 ( 0.92 - 1.02 ) | 3.1E-05 | 3.9E-05 | 0.95 ( 0.93 - 0.98 ) | 3.0E-01 | 5.1E-01 | 1.29 ( 0.79 - 2.1 ) |
| cg11076954 | 1.1E-03 | 2.1E-02 | 11.1 (2.61 - 47.19) | 9.7E-05 | 3.1E-04 | 1.06 ( 1.03 - 1.1 ) | 2.5E-01 | 3.1E-01 | 0.97 ( 0.93 - 1.02 ) | 3.2E-05 | 3.9E-05 | 0.96 ( 0.93 - 0.98 ) | 3.1E-01 | 5.1E-01 | 1.29 ( 0.79 - 2.1 ) |
| cg11420142 | 1.2E-03 | 2.1E-02 | 3.61 (1.66 - 7.83) | 3.8E-04 | 4.0E-04 | 1.06 ( 1.03 - 1.09 ) | 4.4E-01 | 4.4E-01 | 0.98 ( 0.93 - 1.03 ) | 2.7E-05 | 3.9E-05 | 0.95 ( 0.93 - 0.98 ) | 2.5E-01 | 5.1E-01 | 1.35 ( 0.81 - 2.25 ) |
| cg13989295 | 1.2E-03 | 2.2E-02 | 4.59 (1.83 - 11.51) | 1.3E-05 | 3.1E-04 | 1.07 ( 1.04 - 1.11 ) | 3.7E-01 | 3.7E-01 | 0.98 ( 0.93 - 1.03 ) | 9.1E-06 | 3.9E-05 | 0.95 ( 0.93 - 0.97 ) | 3.6E-01 | 5.1E-01 | 1.27 ( 0.76 - 2.09 ) |
| cg08230244 | 1.3E-03 | 2.3E-02 | 0.21 (0.08 - 0.55) | 1.2E-04 | 3.1E-04 | 1.06 ( 1.03 - 1.09 ) | 2.4E-01 | 3.1E-01 | 0.97 ( 0.93 - 1.02 ) | 1.1E-05 | 3.9E-05 | 0.95 ( 0.93 - 0.97 ) | 3.6E-01 | 5.1E-01 | 1.26 ( 0.77 - 2.08 ) |
| cg16788857 | 1.6E-03 | 2.7E-02 | 0.19 (0.07 - 0.53) | 1.3E-05 | 3.1E-04 | 1.07 ( 1.04 - 1.1 ) | 3.6E-01 | 3.7E-01 | 0.98 ( 0.94 - 1.02 ) | 2.1E-05 | 3.9E-05 | 0.95 ( 0.93 - 0.97 ) | 3.8E-01 | 5.1E-01 | 1.25 ( 0.76 - 2.07 ) |

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| cg25621215 | 1.6E-03 | 2.7E-02 | 0.15 (0.05 - 0.49) | 4.3E-04 | 4.5E-04 | 1.06 ( 1.03 - 1.09 ) | 3.8E-01 | 3.8E-01 | 0.98 ( 0.93 - 1.03 ) | 6.6E-05 | 6.7E-05 | 0.96 ( 0.93 - 0.98 ) | 4.9E-01 | 5.2E-01 | 1.19 ( 0.72 - 1.96 ) |
| cg26423139 | 1.6E-03 | 2.8E-02 | 11.37 (2.5 - 51.61) | 9.1E-05 | 3.1E-04 | 1.06 ( 1.03 - 1.1 ) | 2.6E-01 | 3.1E-01 | 0.97 ( 0.92 - 1.02 ) | 3.2E-05 | 3.9E-05 | 0.95 ( 0.93 - 0.98 ) | 3.3E-01 | 5.1E-01 | 1.27 ( 0.78 - 2.08 ) |
| cg04706137 | 2.1E-03 | 3.4E-02 | 0.22 (0.08 - 0.57) | 2.0E-04 | 3.1E-04 | 1.06 ( 1.03 - 1.09 ) | 1.9E-01 | 3.1E-01 | 0.97 ( 0.92 - 1.02 ) | 2.4E-05 | 3.9E-05 | 0.95 ( 0.93 - 0.98 ) | 3.8E-01 | 5.1E-01 | 1.25 ( 0.76 - 2.05 ) |
| cg19618634 | 2.2E-03 | 3.5E-02 | 7.12 (2.03 - 25) | 5.4E-05 | 3.1E-04 | 1.07 ( 1.04 - 1.11 ) | 6.2E-01 | 6.2E-01 | 0.99 ( 0.94 - 1.04 ) | 8.5E-06 | 3.9E-05 | 0.95 ( 0.93 - 0.97 ) | 3.1E-01 | 5.1E-01 | 1.29 ( 0.79 - 2.12 ) |
| cg10499451 | 2.2E-03 | 3.5E-02 | 0.27 (0.12 - 0.63) | 1.4E-04 | 3.1E-04 | 1.06 ( 1.03 - 1.09 ) | 2.1E-01 | 3.1E-01 | 0.97 ( 0.93 - 1.02 ) | 1.3E-05 | 3.9E-05 | 0.95 ( 0.93 - 0.97 ) | 4.8E-01 | 5.2E-01 | 1.2 ( 0.73 - 1.97 ) |
| cg23246911 | 2.2E-03 | 3.6E-02 | 0.05 (0.01 - 0.35) | 4.1E-05 | 3.1E-04 | 1.07 ( 1.04 - 1.11 ) | 3.1E-01 | 3.2E-01 | 0.98 ( 0.93 - 1.02 ) | 2.1E-05 | 3.9E-05 | 0.95 ( 0.93 - 0.97 ) | 3.4E-01 | 5.1E-01 | 1.28 ( 0.77 - 2.11 ) |
| cg02427933 | 2.3E-03 | 3.6E-02 | 5.62 (1.85 - 17.03) | 1.0E-03 | 1.0E-03 | 1.06 ( 1.02 - 1.09 ) | 3.8E-01 | 3.8E-01 | 0.98 ( 0.93 - 1.03 ) | 4.8E-05 | 4.9E-05 | 0.95 ( 0.93 - 0.98 ) | 5.7E-01 | 5.9E-01 | 1.15 ( 0.7 - 1.88 ) |
| cg08586441 | 2.4E-03 | 3.8E-02 | 5.55 (1.84 - 16.79) | 1.4E-04 | 3.1E-04 | 1.06 ( 1.03 - 1.09 ) | 2.3E-01 | 3.1E-01 | 0.97 ( 0.93 - 1.02 ) | 2.0E-05 | 3.9E-05 | 0.95 ( 0.93 - 0.97 ) | 3.3E-01 | 5.1E-01 | 1.28 ( 0.78 - 2.08 ) |
| cg04885581 | 2.8E-03 | 4.2E-02 | 0.08 (0.01 - 0.41) | 1.3E-04 | 3.1E-04 | 1.06 ( 1.03 - 1.09 ) | 4.0E-01 | 4.0E-01 | 0.98 ( 0.93 - 1.03 ) | 3.9E-05 | 4.2E-05 | 0.96 ( 0.93 - 0.98 ) | 2.1E-01 | 5.1E-01 | 1.42 ( 0.83 - 2.43 ) |
| cg12434901 | 3.0E-03 | 4.5E-02 | 4.69 (1.69 - 12.99) | 8.7E-05 | 3.1E-04 | 1.06 ( 1.03 - 1.1 ) | 2.5E-01 | 3.1E-01 | 0.97 ( 0.93 - 1.02 ) | 3.0E-05 | 3.9E-05 | 0.95 ( 0.93 - 0.98 ) | 3.9E-01 | 5.1E-01 | 1.24 ( 0.75 - 2.06 ) |
| cg11128983 | 3.0E-03 | 4.5E-02 | 3.75 (1.57 - 8.97) | 1.0E-04 | 3.1E-04 | 1.06 ( 1.03 - 1.1 ) | 1.7E-01 | 3.1E-01 | 0.97 ( 0.92 - 1.01 ) | 3.7E-05 | 4.1E-05 | 0.95 ( 0.93 - 0.98 ) | 5.0E-01 | 5.3E-01 | 1.19 ( 0.72 - 1.94 ) |
| cg07505631 | 3.1E-03 | 4.6E-02 | 0.25 (0.1 - 0.63) | 1.1E-04 | 3.1E-04 | 1.06 ( 1.03 - 1.09 ) | 4.8E-01 | 4.8E-01 | 0.98 ( 0.94 - 1.03 ) | 1.2E-05 | 3.9E-05 | 0.95 ( 0.93 - 0.97 ) | 3.2E-01 | 5.1E-01 | 1.29 ( 0.78 - 2.12 ) |
| cg14924512 | 3.1E-03 | 4.6E-02 | 0.28 (0.12 - 0.65) | 1.1E-04 | 3.1E-04 | 1.06 ( 1.03 - 1.1 ) | 2.8E-01 | 3.1E-01 | 0.97 ( 0.93 - 1.02 ) | 1.2E-05 | 3.9E-05 | 0.95 ( 0.93 - 0.97 ) | 2.1E-01 | 5.1E-01 | 1.38 ( 0.84 - 2.27 ) |
| cg07917191 | 3.2E-03 | 4.7E-02 | 0.23 (0.08 - 0.61) | 1.1E-04 | 3.1E-04 | 1.06 ( 1.03 - 1.1 ) | 3.9E-01 | 3.9E-01 | 0.98 ( 0.93 - 1.03 ) | 2.0E-05 | 3.9E-05 | 0.95 ( 0.93 - 0.97 ) | 5.2E-01 | 5.4E-01 | 1.18 ( 0.71 - 1.95 ) |

* HR: hazard ratio, calculated based on 1% increase in DNA methylation. CI: confidence interval

**Table 5.** Results from the general linear regression models and the spearmen correlation tests to test the association between DNA methylation of each of the finally selected 37 CpGs (with difference in methylation > 5% between patients with T2D who developed DKD and patients with T2D who did not develop DKD in a subset of 487 patients from the ANDIS and ANDIU cohorts, with a significance level of $q < 0.05$ in the cox regression Model 1a, and $p < 0.05$ in all the other models) and age, sex, BMI and baseline eGFR.

| | General linear regression models | | | | | | | | | | Spearman rank correlation tests | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Intercep | | Age | | Sex (male) | | BMI | | Baseline eGFR | | Age | | Sex (male) * | | BMI | | Baseline eGFR | |
| **TargetID** | **Coeff (± SD)** | **p-value** | **Coeff (± SD)** | **p-value** | **Coeff (± SD)** | **p-value** | **Coeff (± SD)** | **p-value** | **Coeff (± SD)** | **p-value** | **Coeff** | **p-value** | **Coeff** | **p-value** | **Coeff** | **p-value** | **Coeff** | **p-value** |
| cg08641951 | 0.21 (0.05) | **2.52e-05** | -0.0003 (0.0004) | 0.464 | -0.008 (0.008) | 0.303 | 0.0009 (0.0007) | 0.228 | 0.0002 (0.0002) | 0.181 | -0.108 | **0.01743** | -0.046 | 0.3151 | 0.07 | 0.1264 | 0.087 | 0.05482 |
| cg11940040 | 0.31 (0.06) | **7.38e-07** | 0.0003 (0.0006) | 0.586 | 0.015 (0.011) | 0.119 | 0.0003 (0.0009) | 0.735 | 0.00008 (0.0002) | 0.748 | 0.004 | 0.9227 | 0.068 | 0.131 | 0.017 | 0.7086 | 0.014 | 0.759 |
| cg23480021 | 0.87 (0.08) | **< 2e-16** | -0.002 (0.0007) | **0.00129** | -0.02 (0.012) | 0.07783 | -0.001 (0.001) | 0.37405 | -0.0005 (0.0003) | 0.09100 | -0.09 | **0.04151** | -0.07 | 0.1451 | -0.002 | 0.958 | -0.04 | 0.3833 |
| cg11065575 | 0.76 (0.05) | **<2e-16** | -0.0003 (0.0005) | 0.532 | -0.07 (0.009) | **<2e-16** | 0.001 (0.0008) | 0.906 | 0.0003 (0.0002) | 0.204 | -0.02 | 0.7219 | -0.36 | **2.692e-16** | 0.02 | 0.6503 | 0.04 | 0.4096 |
| cg25291037 | 0.47 (0.12) | **9.52e-05** | 0.0009 (0.001) | 0.407 | -0.04 (0.02) | **0.0437** | 0.0002 (0.002) | 0.9119 | -0.001 (0.0005) | **0.044** | 0.07 | 0.1142 | -0.1 | **0.02943** | 0.04 | 0.4133 | -0.12 | **0.00935** |
| cg01079515 | 0.78 (0.1) | **1.21e-13** | 0.0003 (0.001) | 0.761 | 0.02 (0.02) | 0.212 | -0.0007 (0.0015) | 0.637 | -0.00001 (0.0004) | 0.977 | 0.05 | 0.2464 | 0.06 | 0.2061 | -0.05 | 0.259 | -0.04 | 0.3639 |
| cg05074631 | 0.7 (0.09) | **6.79e-15** | -0.0005 (0.0008) | 0.525 | -0.01 (0.01) | 0.285 | -0.001 (0.001) | 0.242 | -0.00003 (0.0003) | 0.915 | 0.00005 | 0.9991 | -0.04 | 0.3465 | -0.03 | 0.4505 | 0.04 | 0.3611 |
| cg00169354 | 0.63 (0.08) | **8.24e-14** | 0.00005 (0.0008 ) | 0.9504 | -0.018 (0.013) | 0.1829 | -0.0026 (0.0013) | **0.0388** | -0.00063 (0.00033 ) | 0.0572 | 0.078 | 0.08513 | -0.057 | 0.2073 | -0.106 | 0.01912 | -0.14 | **0.002194** |
| cg18826637 | 0.62 (0.05) | **<2e-16** | -0.004 (0.0004) | **<2e-16** | -0.02 (0.007) | **0.0123** | 0.0005 (0.0007) | 0.45 | 0.0002 (0.0002) | 0.1767 | -0.42 | **< 2.2e-16** | -0.07 | 0.1175 | 0.09 | **0.04778** | 0.2 | **1.124e-05** |
| cg22297055 | 0.54 (0.09) | **1.99e-09** | 0.0001 (0.0008) | 0.899 | -0.01 (0.01) | 0.46 | -0.0002 (0.001) | 0.899 | -0.0005 (0.0003) | 0.143 | 0.04 | 0.4079 | -0.04 | 0.4247 | 0.04 | 0.3793 | -0.13 | **0.004802** |
| cg17386240 | 0.77 (0.14) | **3.63e-08** | -0.0006 (0.001) | 0.6161 | -0.01 (0.02) | 0.4802 | -0.002 (0.002) | 0.396 | -0.001 (0.0005) | 0.0715 | 0.02 | 0.701 | -0.03 | 0.5053 | -0.02 | 0.6222 | -0.06 | 0.1579 |
| cg26705599 | 0.67 (0.17) | **0.000127** | -0.003 (0.001) | 0.07136 | -0.03 (0.03) | 0.231922 | 0.001 (0.002) | 0.64805 | 0.0007 (0.0007) | 0.30494 | -0.14 | **0.002625** | -0.05 | 0.3045 | 0.06 | 0.1819 | 0.09 | **0.04091** |
| cg12183150 | 0.4 (0.07) | **9.96e-08** | 0.0003 (0.0006) | 0.596 | 0.007 (0.01) | 0.511 | 0.0008 (0.001) | 0.438 | -0.00002 (0.0003) | 0.942 | 0.03 | 0.466 | 0.02 | 0.5783 | 0.02 | 0.6226 | -0.008 | 0.8651 |
| cg10187601 | 1.1 (0.1) | **<2e-16** | -0.002 (0.0009) | **0.0141** | 0.006 (0.01) | 0.7134 | -0.002 (0.001) | 0.1056 | -0.0009 (0.0004) | **0.0287** | -0.07 | 0.1256 | 0.03 | 0.5255 | -0.1 | **0.02273** | -0.08 | 0.08622 |
| cg07784975 | 0.3 (0.05) | **6.28e-11** | -0.002 (0.0004) | **9.23e-05** | 0.04 (0.007) | **1.36e-06** | 0.0004(0.0007 ) | 0.547 | 0.00009 (0.0002) | 0.636 | -0.22 | **4.748e-07** | 0.23 | **3.287e-07** | 0.04 | 0.349 | 0.13 | **0.003049** |
| cg10482512 | 0.47 (0.14) | **0.00115** | -0.0005 (0.001) | 0.71222 | 0.01 (0.02) | 0.55841 | -0.0007 (0.002) | 0.73233 | -0.0001 (0.0006) | 0.80611 | -0.01 | 0.8086 | 0.03 | 0.5213 | 0.001 | 0.9759 | -0.004 | 0.9297 |
| cg26365090 | 0.35 (0.18) | **0.0450** | 0.002 (0.002) | 0.1548 | 0.007 (0.03) | 0.7962 | -0.008 (0.003) | **0.0037** | 0.0006 (0.0007) | 0.3894 | 0.09 | 0.05341 | 0.02 | 0.7217 | -0.13 | **0.005268** | -0.02 | 0.6773 |

| | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| cg22273830 | 0.48 (0.1) | **4.58e-06** | -0.0002 (0.0009) | 0.814 | -0.005 (0.02) | 0.767 | -0.0008 (0.001) | 0.954 | -0.0005 (0.0004) | 0.185 | 0.008 | 0.8508 | -0.01 | 0.7499 | 0.04 | 0.3359 | -0.11 | **0.01649** |
| cg11076954 | 0.48 (0.08) | **3.49e-08** | -0.0002 (0.0007) | 0.807 | -0.009 (0.01) | 0.509 | -0.0001 (0.001) | 0.924 | -0.0004 (0.0003) | 0.192 | 0.02 | 0.5704 | -0.03 | 0.4959 | 0.05 | 0.2715 | -0.12 | **0.007402** |
| cg11420142 | 0.63 (0.17) | **0.000371** | 0.002 (0.002) | 0.223935 | -0.001 (0.03) | 0.971462 | -0.009 (0.003) | **0.000598** | -0.0002 (0.0007) | 0.746852 | 0.07 | 0.1059 | 0.003 | 0.9371 | -0.17 | **0.000151** | -0.04 | 0.4256 |
| cg13989295 | 0.51 (0.18) | **0.00365** | -0.003 (0.002) | **0.04378** | -0.005 (0.03) | 0.86537 | 0.001 (0.003) | 0.56942 | 0.0004 (0.0007) | 0.5471 | -0.11 | **0.01226** | 0.00001 | 0.9998 | 0.04 | 0.4259 | 0.07 | 0.1366 |
| cg08230244 | 0.9 (0.1) | **8.69e-12** | -0.0001 (0.001) | 0.899 | -0.02 (0.02) | 0.449 | -0.002 (0.002) | 0.344 | -0.00008 (0.0005) | 0.878 | -0.06 | 0.1969 | -0.03 | 0.4899 | -0.07 | 0.1094 | 0.05 | 0.2242 |
| cg16788857 | 0.32 (0.14) | **0.0255** | 0.003 (0.001) | **0.0153** | 0.01 (0.02) | 0.6553 | -0.002 (0.002) | 0.3469 | 0.0004 (0.0006) | 0.4709 | 0.11 | **0.01665** | 0.01 | 0.7566 | -0.06 | 0.1794 | 0.006 | 0.9014 |
| cg25621215 | 0.89 (0.09) | **<2e-16** | -0.0006 (0.0008) | 0.461 | 0.01 (0.01) | 0.471 | -0.002 (0.001) | 0.242 | 0.0005 (0.0004) | 0.192 | -0.01 | 0.8017 | 0.04 | 0.3569 | -0.09 | 0.05354 | 0.07 | 0.1132 |
| cg26423139 | 0.5 (0.09) | **2.44e-09** | -0.00005 (0.0008) | 0.948 | 0.006 (0.01) | 0.679 | -0.0003 (0.001) | 0.803 | -0.0005 (0.0003) | 0.117 | 0.02 | 0.5895 | 0.02 | 0.6991 | 0.02 | 0.596 | -0.12 | **0.008839** |
| cg04706137 | 0.77 (0.1) | **1.06e-13** | -0.0007 (0.0009) | 0.444 | -0.01 (0.02) | 0.482 | 0.0005 (0.001) | 0.728 | 0.0001 (0.0004) | 0.698 | -0.03 | 0.541 | -0.03 | 0.5172 | 0.02 | 0.6152 | 0.03 | 0.4497 |
| cg19618634 | 0.66 (0.13) | **7.01e-07** | -0.001 (0.001) | 0.40974 | -0.03 (0.02) | 0.1055 | -0.005 (0.002) | **0.00693** | 0.0003 (0.0005) | 0.50686 | -0.05 | 0.2481 | -0.06 | 0.1823 | -0.11 | **0.01812** | 0.05 | 0.2714 |
| cg10499451 | 0.71 (0.15) | **4.56e-06** | 0.0006 (0.001) | 0.637 | -0.01 (0.02) | 0.573 | 0.002 (0.002) | 0.42 | -0.0004 (0.0006) | 0.501 | -0.01 | 0.7941 | -0.03 | 0.4883 | 0.07 | 0.1399 | -0.007 | 0.8769 |
| cg23246911 | 0.24 (0.09) | **0.00919** | 0.001 (0.0008) | 0.19411 | 0.008 (0.01) | 0.57728 | 0.0007 (0.001) | 0.58408 | 0.0003 (0.0003) | 0.39791 | 0.03 | 0.5595 | 0.02 | 0.6669 | 0.03 | 0.566 | 0.04 | 0.3985 |
| cg02427933 | 0.3 (0.1) | **0.0173** | -0.0002 (0.001) | 0.8332 | 0.004 (0.02) | 0.8236 | -0.0006 (0.002) | 0.7624 | -0.0002 (0.0005) | 0.7397 | -0.02 | 0.6308 | 0.01 | 0.7989 | -0.04 | 0.4066 | 0.006 | 0.8989 |
| cg08586441 | 0.56 (0.1) | **5.56e-06** | 0.0009 (0.001) | 0.3966 | -0.03 (0.02) | 0.0657 | -0.001 (0.002) | 0.5506 | 0.00004 (0.0005) | 0.9274 | 0.06 | 0.2025 | -0.09 | 0.05855 | -0.02 | 0.6848 | -0.04 | 0.3511 |
| cg04885581 | 0.78 (0.07) | **<2e-16** | -0.001 (0.0006) | 0.106 | 0.02 (0.01) | 0.112 | 0.0009 (0.001) | 0.357 | 0.0004 (0.0003) | 0.82 | -0.15 | **0.0007728** | 0.08 | 0.08267 | 0.02 | 0.6858 | 0.14 | **0.00214** |
| cg12434901 | 0.66 (0.16) | **5.6e-05** | 0.00008 (0.001) | 0.954 | 0.01 (0.02) | 0.568 | 0.0007 (0.002) | 0.765 | -0.0008 (0.0006) | 0.206 | -0.04 | 0.4181 | 0.02 | 0.6131 | -0.009 | 0.837 | -0.01 | 0.7313 |
| cg11128983 | 0.62 (0.18) | **0.000482** | 0.0001 (0.002) | 0.940341 | 0.03 (0.03) | 0.26 | 0.006 (0.003) | **0.028124** | -0.002 (0.0007) | **0.006296** | 0.02 | 0.6021 | 0.04 | 0.3845 | 0.06 | 0.218 | -0.13 | **0.003743** |
| cg07505631 | 0.7 (0.2) | **9.96e-06** | -0.002 (0.001) | 0.217 | 0.02 (0.02) | 0.515 | 0.0009 (0.002) | 0.702 | 0.00008 (0.0006) | 0.897 | -0.12 | **0.01051** | 0.03 | 0.4534 | 0.001 | 0.9739 | 0.04 | 0.4307 |
| cg14924512 | 0.73 (0.14) | **2.98e-07** | -0.0004 (0.001) | 0.717 | 0.001 (0.02) | 0.952 | 0.0003 (0.002) | 0.886 | 0.0005 (0.0005) | 0.336 | -0.06 | 0.1646 | 0.005 | 0.9059 | 0.02 | 0.7067 | 0.09 | **0.03861** |
| cg07917191 | 0.66 (0.13) | **2.75e-07** | 0.0006 (0.001) | 0.589 | -0.01 (0.02) | 0.504 | 0.004 (0.002) | **0.0315** | 0.0001 (0.0005) | 0.8327 | -0.04 | 0.3577 | -0.04 | 0.3867 | 0.08 | 0.06029 | 0.02 | 0.66 |

Significant p-values are highlighted in bold

* Point-biserial correlation test

**Table 6.** Description of the 27 unique genes annotated to the 37 CpGs with difference in methylation > 5% between patients with T2D who developed DKD and patients with T2D who did not develop DKD in a subset of 487 patients from the ANDIS and ANDIU cohorts, with a significance level of $q < 0.05$ in Model 1a, and $p < 0.05$ in all the other models, after a preliminary GeneCards search for GWAS phenotypes and GO biological processes related to those genes in the context of T2D and DKD.

| | | GWAS Phenotype ** | | | | | | Gene Ontology (GO) – Biological Processes | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Differentially methylated sites** | **Gene symbol** | **Description** | **Gene Relation** | **Mean Score** | **SNP Count** | **Studies Count** | **SNPs** | **Description** | **Evidence** | **Accession** | **Sources** | **PubMed ID** |
| cg08641951 | *ADAMTS16* | BMI | GWAS | 5.10 | 1 | 1 | rs871122 | Involved in regulation of systemic arterial blood pressure | IEA | GO:0003073 | NCBI Entrez Gene\|\|Ensembl | |
| | | | | | | | | Involved in extracellular matrix organization | IBA | GO:0030198 | NCBI Entrez Gene\|\|Ensembl | 21873635 |
| cg11940040, cg23246911 | *RPH3AL* | insulin sensitivity measurement, response to polyunsaturated fatty acid supplementation | GeneHancer; GWAS | 5.77 | 2 | 1 | rs35621498;rs55842940 | Involved in positive regulation of insulin secretion | IEA | GO:0032024 | NCBI Entrez Gene\|\|Ensembl | |
| | | | | | | | | Involved in glucose homeostasis | IEA | GO:0042593 | NCBI Entrez Gene\|\|Ensembl | |
| | | | | | | | | Involved in intracellular protein transport | IEA | GO:0006886 | NCBI Entrez Gene\|\|Ensembl | |
| | | | | | | | | Involved in exocytosis | IBA, TAS | GO:0006887 | NCBI Entrez Gene\|\|Ensembl | 21873635 |
| cg23480021 | *ZNF385D* | waist-hip ratio | GWAS | 5.70 | 1 | 1 | rs1388551 | | | | | |
| | | T2D | GeneHancer; GWAS | 10.30 | 2 | 2 | rs2688419;rs6791903 | | | | | |
| | | body mass index, visceral:subcutaneous adipose tissue ratio | GeneHancer | 7.70 | 1 | 1 | rs7374732 | | | | | |
| | | hip circumference | GeneHancer | 5.15 | 1 | 1 | rs749519604 | | | | | |
| | | chronic kidney disease | GeneHancer | 5.70 | 1 | 1 | rs9310709 | | | | | |
| cg11065575 | *SHISA3* | T2D | GWAS | 5.70 | 1 | 1 | rs17447640 | | | | | |
| | | creatinine measurement | GWAS | 8.41 | 2 | 2 | rs11722932;rs74464322 | | | | | |
| | | glomerular filtration rate | GWAS | 8.70 | 1 | 1 | rs11722932 | | | | | |
| cg25291037 | x | | | | | | | | | | | |
| cg01079515 | *AC124944.5* * | | | | | | | | | | | |
| cg05074631 | *DZIP1* | glomerular filtration rate, cystatin C measurement | GWAS | 7.22 | 1 | 1 | rs12428035 | | | | | |
| | | glomerular filtration rate | GeneHancer; GWAS | 8.15 | 1 | 1 | rs77930391 | | | | | |
| cg00169354 | x | | | | | | | | | | | |
| cg18826637 | x | | | | | | | | | | | |
| cg22297055, cg22273830, cg11076954 | *SLC43A2* | BMI-adjusted waist circumference | GeneHancer; GWAS | 9.05 | 2 | 3 | rs58360980;rs78135061 | Involved in transmembrane transport | IEA | GO:0055085 | Ensembl | |
| | | serum albumin measurement | GeneHancer; GWAS | 26.89 | 4 | 4 | rs11078596;rs11078597;rs4790712;rs62090014 | Involved in amino acid transmembrane transport | IEA | GO:0003333 | Ensembl | |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | glomerular filtration rate | GeneHancer; GWAS | 9.77 | 2 | 1 | rs57937546;rs62088028 | Involved in amino acid transport | TAS | GO:0006865 | NCBI Entrez Gene\|\|Ensembl | |
| | | UACR | GeneHancer | 9.19 | 1 | 1 | rs11078597 | Involved in neutral amino acid transport | IBA | GO:0015804 | NCBI Entrez Gene\|\|Ensembl | 21873635 |
| | | BMI-adjusted waist-hip ratio | GeneHancer | 7.52 | 1 | 1 | rs67735224 | Involved in L-amino acid transport | IEA | GO:0015807 | NCBI Entrez Gene\|\|Ensembl | |
| | | blood protein measurement | GeneHancer | 69.04 | 4 | 3 | rs2272011;rs58697961;rs62088172;rs62090014 | | | | | |
| | | waist-hip ratio, | GeneHancer | 8.40 | 1 | 1 | rs11078594 | | | | | |
| | | systolic blood pressure | GeneHancer | 14.30 | 1 | 1 | rs8076897 | | | | | |
| | | lipid measurement | GeneExon;GeneHancer;GWAS | 5.70 | 1 | 1 | rs17027255 | | | | | |
| cg17386240 | *TGFBI* | blood protein measurement | GeneHancer; GWAS | 57.19 | 5 | 3 | rs13159365;rs17689879;rs248160;rs2526145;rs756463 | Involved in angiogenesis | IEP | GO:0001525 | NCBI Entrez Gene\|\|Ensembl | 11866539 |
| | | | | | | | | Involved in cell adhesion | IBA, IEA | GO:0007155 | NCBI Entrez Gene\|\|Ensembl | 21873635 |
| | | | | | | | | Involved in cell population proliferation | IEA | GO:0008283 | NCBI Entrez Gene\|\|Ensembl | |
| | | | | | | | | Involved in extracellular matrix organization | IBA | GO:0030198 | NCBI Entrez Gene\|\|Ensembl | 21873635 |
| cg26705599 | x | | | | | | | | | | | |
| cg12183150 | *AC097713.4;AC097713.3 ** | | | | | | | | | | | |
| cg10187601 | *ST7* | BMI-adjusted waist-hip ratio | GWAS | 8.65 | 3 | 3 | rs38902;rs38906;rs5886848 | Involved in extracellular matrix organization | NAS | GO:0030198 | NCBI Entrez Gene\|\|Ensembl | 16474848 |
| | | BMI-adjusted waist-hip ratio, sex interaction measurement, age at assessment | GWAS | 6.52 | 1 | 1 | rs38902 | | | | | |
| | | blood protein measurement | GeneHancer; GWAS | 22.67 | 2 | 2 | rs141636670;rs1858830 | | | | | |
| cg07784975 | x | | | | | | | | | | | |
| cg10482512 | *CCR6* | blood protein measurement | GeneHancer; GWAS | 11.22 | 1 | 1 | rs3093023 | | | | | |
| | | lipid measurement | GeneHancer; GWAS | 6.30 | 1 | 1 | rs113085122 | | | | | |
| cg26365090 | *TOX2* | metabolically healthy obesity | GWAS | 5.30 | 1 | 1 | rs6093921 | | | | | |
| | | systolic blood pressure | GeneHancer | 9.41 | 4 | 8 | rs6017279;rs6017281;rs6031431;rs6031435 | | | | | |
| | | creatinine measurement, glomerular filtration rate | GeneHancer | 6.52 | 1 | 1 | rs2235808 | | | | | |
| | | mean arterial pressure | GeneHancer | 8.38 | 3 | 3 | rs6017281;rs6031431;rs6031435 | | | | | |
| | | diastolic blood pressure | GeneHancer | 9.70 | 1 | 1 | rs6031431 | | | | | |
| | | cardiovascular disease | GeneHancer | 10 | 1 | 1 | rs6031431 | | | | | |
| | | glomerular filtration rate | GeneHancer | 11.30 | 1 | 1 | rs2235808 | | | | | |
| cg11420142 | *RP11-122C21.0 ** | | | | | | | | | | | |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| cg13989295 | *SKA2* | glomerular filtration rate | GWAS | 5.70 | 1 | 1 | rs8073316 | | | | | | |
| | | type 2 diabetes mellitus | GeneHancer | 7.61 | 2 | 2 | rs1451506;rs302864 | | | | | | |
| | | BMI-adjusted waist circumference | GeneHancer | 7.40 | 1 | 1 | rs5821283 | | | | | | |
| cg08230244 | *FAM13B* | systolic blood pressure | GeneExon; GWAS | 7.70 | 1 | 1 | rs33956817 | | | | | | |
| | | cardioembolic stroke | GeneHancer; GWAS | 25.40 | 1 | 1 | rs17171711 | | | | | | |
| | | lipid measurement | GWAS | 5 | 1 | 1 | rs1124472 | | | | | | |
| cg16788857 | *C3orf59* | | | | | | | | | | | | |
| cg25621215 | *RP11-445H22.2* * | | | | | | | | | | | | |
| cg26423139 | *SLC43A2* | | | | | | | | | | | | |
| cg04706137 | *RP11-867G2.8* * | | | | | | | | | | | | |
| cg19618634 | x | | | | | | | | | | | | |
| cg10499451 | *PAPL* | T2D | GWAS | 5.05 | 1 | 1 | rs472265 | | | | | | |
| | | blood protein measurement | GeneHancer | 36.96 | 2 | 2 | rs2229259;rs4802890 | | | | | | |
| | | body weight | GWAS | 9 | 1 | 1 | rs112691865 | | | | | | |
| | | waist circumference | GeneHancer | 5.15 | 1 | 1 | rs66959926 | | | | | | |
| | | retinal vasculature measurement | GeneHancer | 12.70 | 1 | 1 | rs1808382 | | | | | | |
| cg02427933 | *RP11-445H22.2* * | | | | | | | | | | | | |
| cg08586441 | *TEC* | systolic blood pressure | GWAS | 6.05 | 1 | 1 | rs17471509 | | | | | | |
| cg04885581 | *FRMD6* | triglyceride measurement, response to diuretic | GWAS | 5.87 | 1 | 1 | rs2790503 | | | | | | |
| | | Ischemic stroke | GeneHancer | 5.52 | 1 | 1 | rs139706713 | | | | | | |
| | | lipid measurement | GeneHancer | 5.10 | 1 | 1 | rs4432175 | | | | | | |
| cg12434901 | *KCNH6* | systolic blood pressure | GeneHancer | 8.5 | 2 | 2 | rs4291;rs4295 | Involved in transmembrane transport | IEA | GO:0055085 | Ensembl | | |
| | | diastolic blood pressure | GeneHancer | 7.40 | 1 | 1 | rs4295 | | | | | | |
| | | lipid measurement | GeneExon;GWAS | 5.05 | 1 | 1 | rs7225568 | | | | | | |
| cg11128983 | *CPLX2* | | | | | | | Involved in regulation of exocytosis | TAS | GO:0017157 | NCBI Entrez Gene\|\|Ensembl | 15217342 | |
| cg07505631 | *CDCP1* | blood protein measurement | GeneHancer; GWAS | 16.64 | 3 | 3 | rs35605067;rs7621542;rs8318 | | | | | | |
| | | urinary metabolite measurement | GeneHancer | 40.10 | 1 | 1 | rs17279437 | | | | | | |
| cg14924512 | *EHMT1* | BMI | GWAS | 8.15 | 1 | 1 | rs35528310 | Involved in DNA methylation | ISS, IBA, IBA | GO:0006306 | NCBI Entrez Gene\|\|Ensembl | | |
| | | systolic blood pressure | GeneHancer | 8 | 1 | 1 | rs821317 | | | | | | |
| | | blood protein measurement | GeneHancer | 27.48 | 7 | 2 | rs11145940;rs2071006;rs28578007;rs41309980;rs48880078;rs7850844;rs908830 | | | | | | |
| | | glomerular filtration rate | GeneHancer | 8.5 | 2 | 2 | rs6606564;rs6606567 | | | | | | |
| | | creatinine measurement | GeneHancer | 8.70 | 1 | 1 | rs6606567 | | | | | | |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| cg07917191 | *SLC9A4* | glomerular filtration rate | GeneHancer; GWAS | 12.03 | 2 | 3 | rs72995641;rs77375846 | Involved in transmembrane transport | IEA | GO:0055085 | Ensembl | |
| | | blood protein measurement | GeneHancer; GWAS | 139.85 | 2 | 2 | rs1420106;rs4851589 | | | | | |
| | | creatinine measurement | GeneHancer; GWAS | 14.70 | 1 | 1 | rs77375846 | | | | | |

* Uncharacterized non-coding RNA genes ** Source: GWAS catalog

**Table 7.** Description of the SNPs located nearby the 37 CpGs with difference in methylation > 5% between patients with T2D who developed DKD and patients with T2D who did not develop DKD in a subset of 487 patients from the ANDIS and ANDIU cohorts, with a significance level of $q < 0.05$ in Model 1a, and $p < 0.05$ in all the other models, taken from the Illumina Infinium MethylationEPIC v1.0 B5 Manifest File (https://emea.support.illumina.com/downloads/infinium-methylationepic-v1-0-product-files.html).

| Target ID | Chromossome | Gene symbol, Illumina standard manifest | Gene symbol, hg38, GENCODE version 12 | SNP ID | SNP Distance | SNP Minor Allele Frequency |
|---|---|---|---|---|---|---|
| cg08641951 | 5 | *ADAMTS16* | | | | |
| cg11940040 | 17 | *RPH3AL* | *RPH3AL* | rs148422036; rs62053678; rs377214475; rs535875419 | 51;44;12;10 | 0.005; 0.05; 0.0006; 0.0006 |
| cg23480021 | 3 | | *ZNF385D* | | | |
| cg11065575 | 4 | *SHISA3* | *SHISA3* | rs142050054; rs529887150 | 20;42 | 0.0002; 0.0002 |
| cg25291037 | 2 | | | rs148598135; rs2315380 | 48;44 | 0.002; 0.18 |
| cg01079515 | 3 | | *AC124944.5;AC124944.5* | rs569207004 | 1 | 0.0002 |
| cg05074631 | 13 | *DZIP1;DZIP1* | *DZIP1;DZIP1;DZIP1* | rs3759446 | 1 | 0.03 |
| cg00169354 | 20 | | | rs527854346 | 21 | 0.0002 |
| cg18826637 | 2 | | | rs576210844; rs543272142; rs561928581 | 9;23;51 | 0.0002; 0.0002; 0.0002 |
| cg22297055 | 17 | *SLC43A2;SLC43A2;SLC43A2* | *SLC43A2;SLC43A2* | rs562818385; rs531810546; rs548715581; rs568625960; rs534392668 | 24;34;43;47;50 | 0.0002; 0.0006; 0.0004; 0.000; 0.0002 |
| cg17386240 | 5 | *TGFBI* | *TGFBI* | rs535789363; **rs45554435** | 37;1 | 0.003; 0.33 |
| cg26705599 | 13 | | | rs575747933; **rs4773321** | 9;1 | 0.0002; 0.41 |
| cg12183150 | 2 | | *AC097713.4;AC097713.3* | | | |
| cg10187601 | 7 | *ST7;ST7* | *ST7* | rs567883351; **rs10270156** | 10;1 | 0.0002; 0.26 |
| cg07784975 | 12 | | | | | |
| cg10482512 | 6 | *CCR6* | *CCR6* | **rs2021033** | 1 | 0.24 |
| cg26365090 | 20 | *TOX2;TOX2;TOX2;TOX2* | *TOX2;TOX2;TOX2;TOX2 ;TOX2* | **rs11700304**; rs532615837; rs552433433; rs559654933 | 0;4;15;43 | 0.42; 0.0002; 0.0002; 0.0002 |
| cg22273830 | 17 | *SLC43A2;SLC43A2;SLC43A2* | *SLC43A2;SLC43A2* | rs562818385; rs531810546; rs548715581; rs568625960; rs534392668; rs376438864; rs17626180 | 1;11;20;24;27;33;51 | 0.0002; 0.0006; 0.0004; 0.0002; 0.0002; 0.0004; 0.33 |
| cg11076954 | 17 | *SLC43A2* | *SLC43A2;SLC43A2* | rs562818385; rs531810546; rs548715581; rs568625960 | 28;38;47;51 | 0.0002; 0.0006; 0.0004; 0.0002 |

| | | | | | | |
|---|---|---|---|---|---|---|
| cg11420142 | 8 | | *RP11-122C21.1* | rs538253857; rs547324927; **rs843071** | 41;9;1 | 0.0004; 0.0002; 0.19 |
| cg13989295 | 17 | *SKA2;SKA2* | *AC099850.1;SKA2;SKA2* | **rs7208505** | 1 | 0.41 |
| cg08230244 | 5 | *FAM13B;FAM13B;FAM13B* | | rs17171647; rs564191885; **rs73299210** | 13;2;1 | 0.21; 0.0002; 0.09 |
| cg16788857 | 3 | *C3orf59* | | rs569935473; rs542555526; **rs57896542** | 13;12;1 | 0.0004; 0.003; 0.35 |
| cg25621215 | 20 | | *RP11-445H22.3* | | | |
| cg26423139 | 17 | *SLC43A2;SLC43A2;SLC43A2* | *SLC43A2;SLC43A2* | rs562818385; rs531810546 | 40;50 | 0.0002; 0.0006 |
| cg04706137 | 11 | | *RP11-867G2.8;RP11-867G2.8* | **rs12365323**; rs75321423 | 1;16 | 0.22; 0.005 |
| cg19618634 | 17 | | | rs201226439; **rs5821824** | 2;1 | 0.0002; 0.40 |
| cg10499451 | 19 | *PAPL* | | **rs57733150**; rs544470869; rs562681261; rs533152162 | 0;13;31;47 | 0.18; 0.0002; 0.001; 0.0002 |
| cg23246911 | 17 | *RPH3AL* | *RPH3AL* | rs77845261; rs577284567; rs544356234 | 46;29;12 | 0.11; 0.0002; 0.0002 |
| cg02427933 | 20 | | *RP11-445H22.3* | rs191659354 | 46 | 0.0004 |
| cg08586441 | 4 | *TEC* | | rs140195097; rs191235412; rs184660155; rs2704433; rs3842586 | 1;11;13;21;39 | 0.0002; 0.0002; 0.005; 0.09; 0.48 |
| cg04885581 | 14 | *FRMD6* | *FRMD6* | rs73292745; rs577536589; rs113737961; rs562581435 | 0;22;28;43 | 0.04; 0.0002; 0.0024; 0.0002 |
| cg12434901 | 17 | *KCNH6;KCNH6* | | **rs4257270**; rs192218733 | 1;44 | 0.49; 0.0002 |
| cg11128983 | 5 | *CPLX2;CPLX2* | *CPLX2;CPLX2;CPLX2;CPLX2* | rs192385285; rs534263655; rs1366116 | 40;10;2 | 0.0002; 0.0002; 0.44 |
| cg07505631 | 3 | *CDCP1;CDCP1* | *CDCP1* | **rs59145932**; rs7612333; rs146318958 | 0;32;51 | 0.15; 0.25; 0.0002 |
| cg14924512 | 9 | *EHMT1;EHMT1* | *EHMT1* | **rs35528310** | 1 | 0.06 |
| cg07917191 | 2 | *SLC9A4* | | **rs59694233** | 1 | 0.14 |

* SNPs with a distance of less than 10 base pairs from the CpG probe, and with a minor allele frequency of more than 0.05 are highlighted in bold.

## 8. Figures

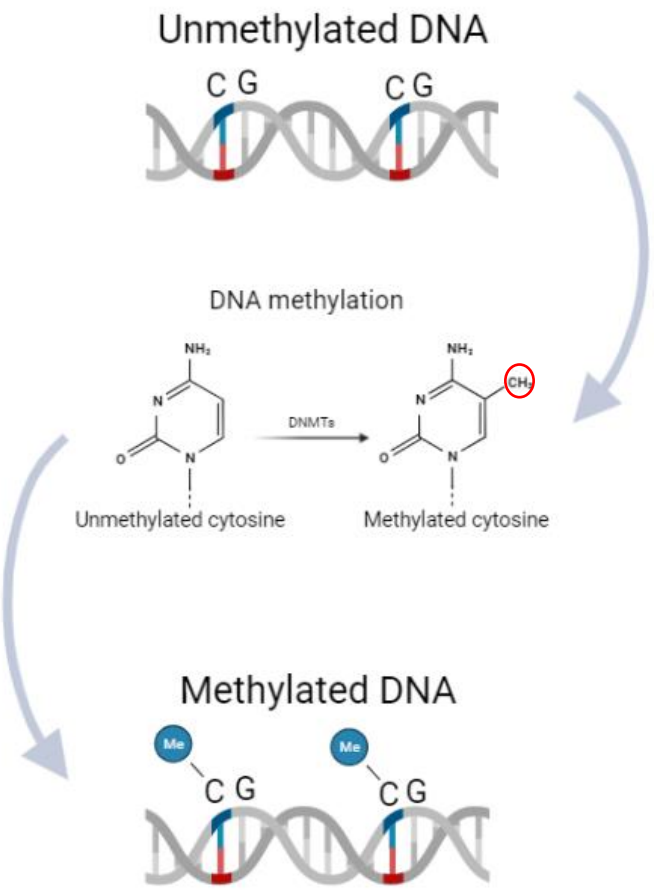



**Figure 1.** Schematic representation of the DNA methylation process of two CpGs. This figure was created using BioRender (https://biorender.com) (87).

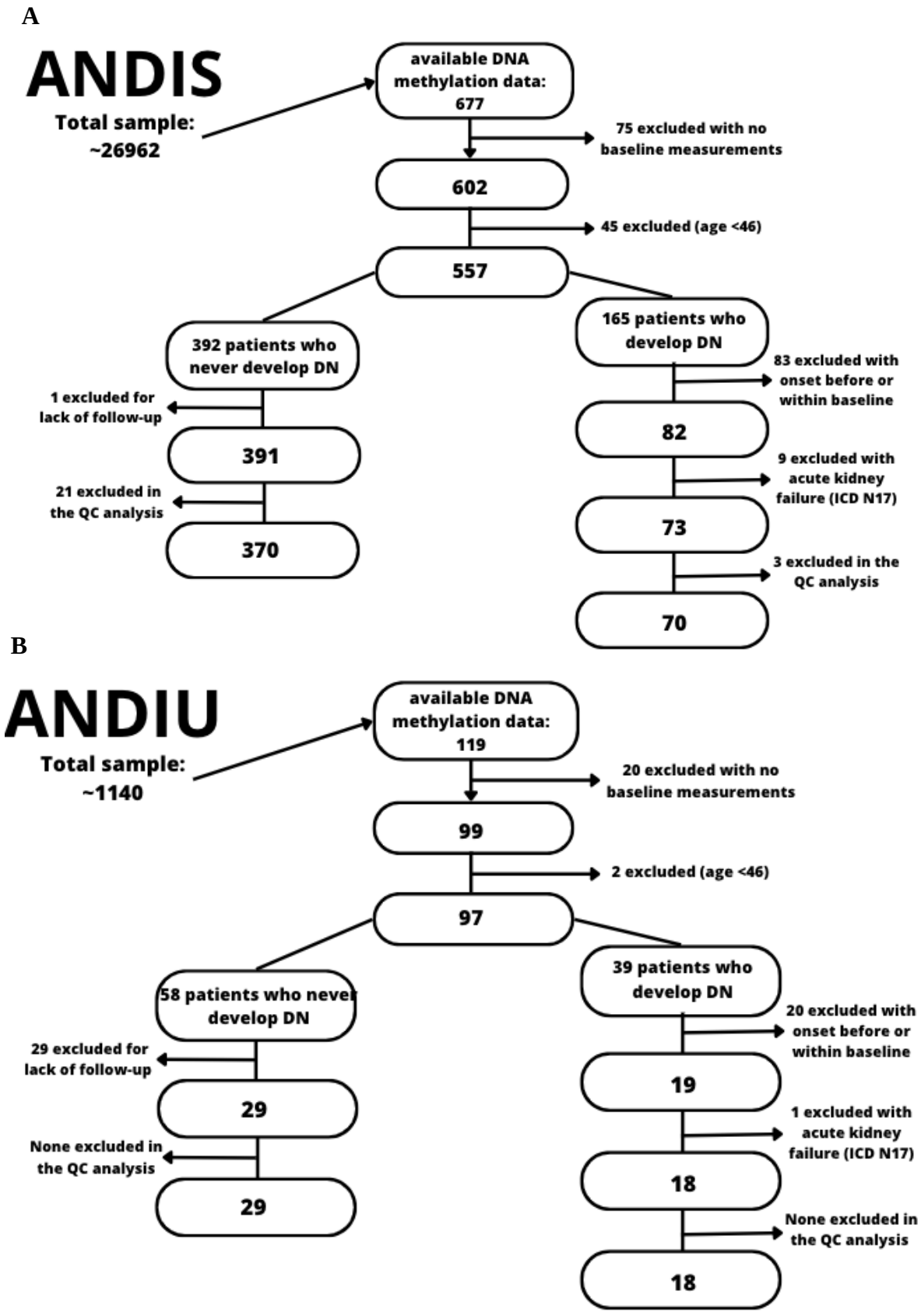



**Figure 2.** Flow chart for sample selection in the ANDIS (A) and ANDIU (B) cohorts.

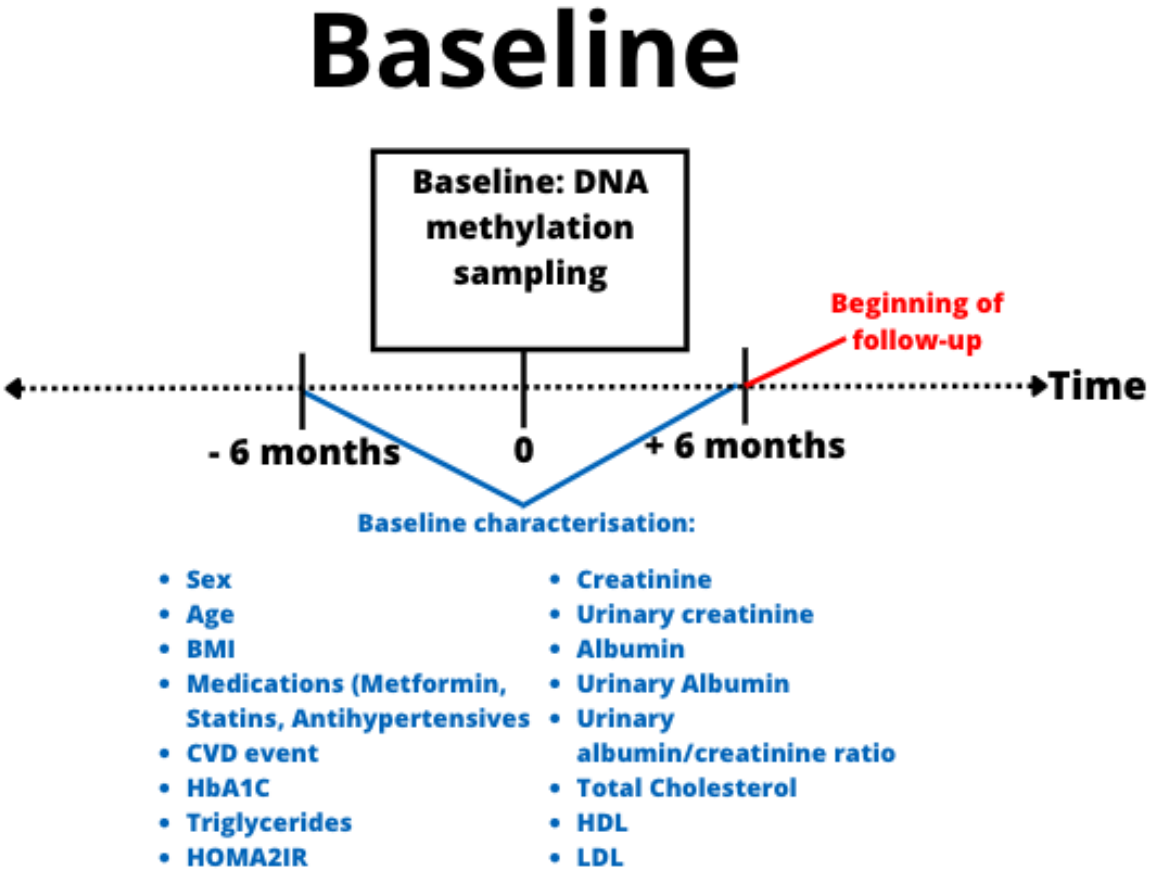



**Figure 3.** Baseline definition and characterization of the clinical variables used for the sample of a subset of 487 individuals from the ANDIS and ANDIU cohorts.

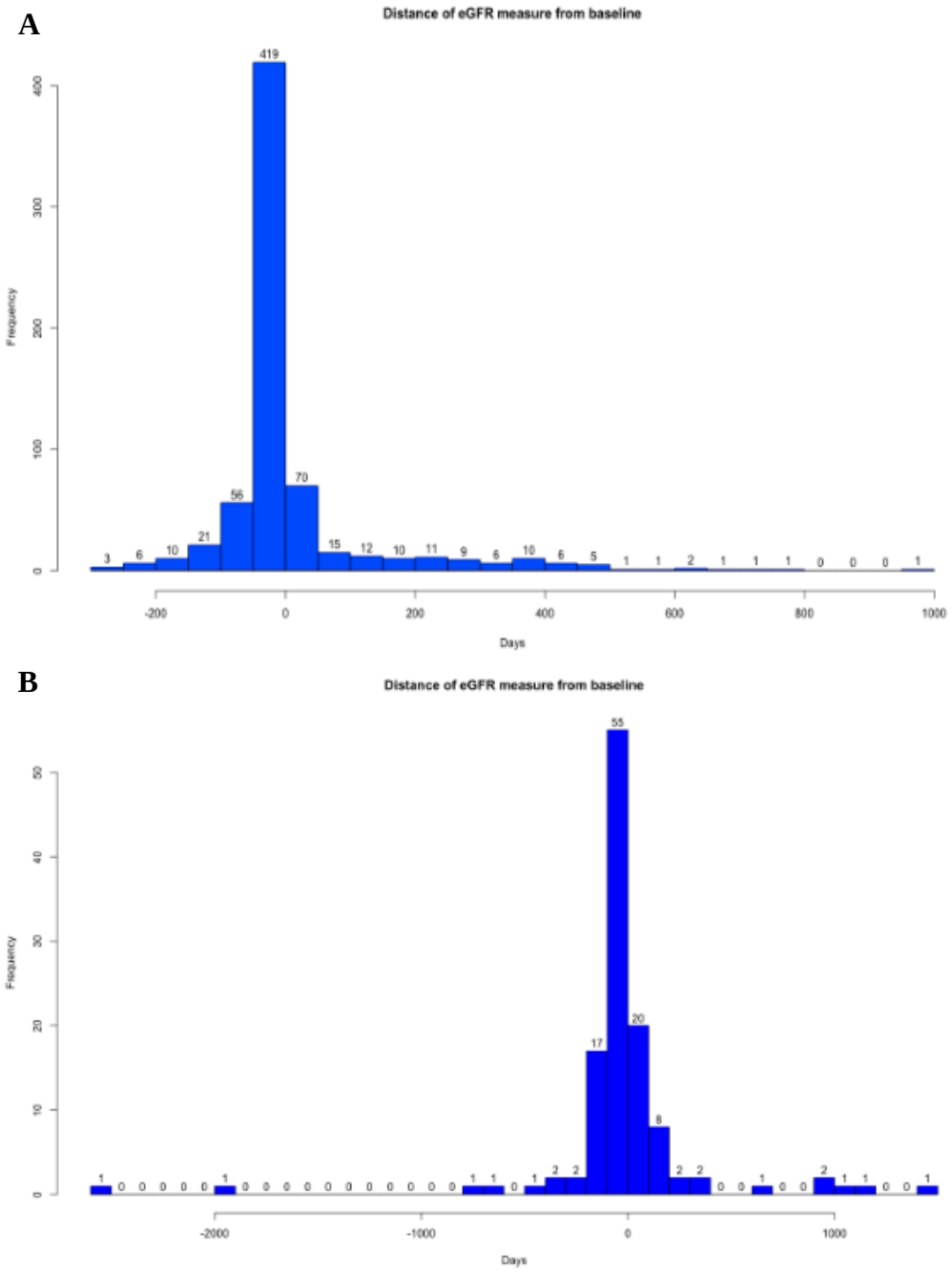



**Figure 4.** Distance (in days) of the measurement of the eGFR value and the baseline (DNA methylation sample date) for a subset of 487 individuals from the ANDIS (A) and ANDIU (B) cohorts. eGFR was taken as a reference variable because it was used to define the outcome. It was calculated from creatinine values using the MDR-4 study equation. The distance was calculated individually for each patient, from the day of the blood test closest to the DNA methylation sample date, considered as baseline blood test, to the DNA methylation sample date. A negative value indicates that the blood sample was taken before DNA methylation sample date, and a positive value, afterwards.

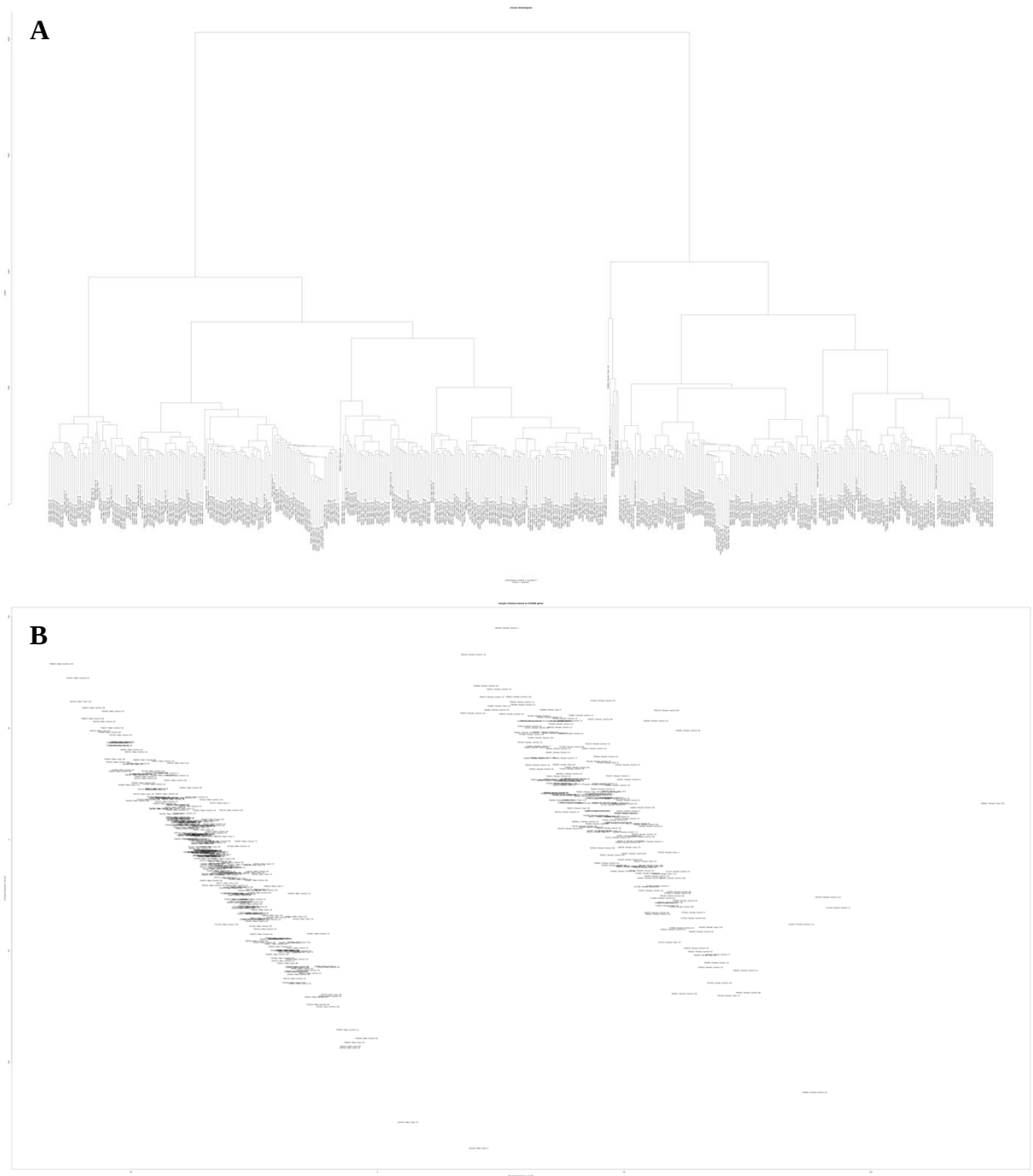
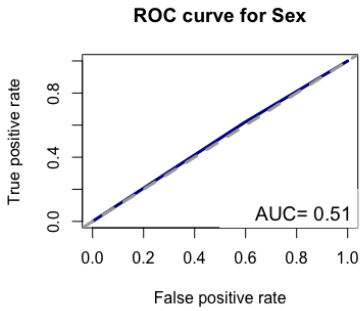
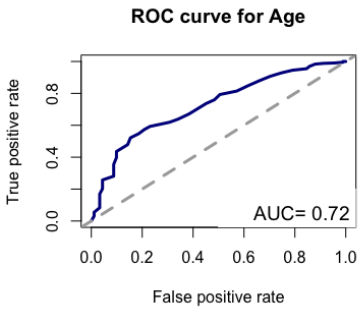
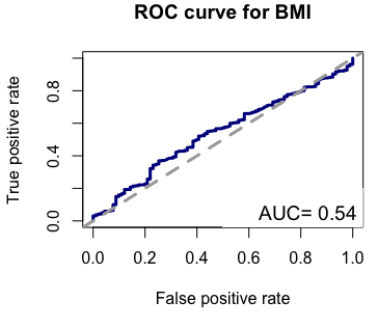
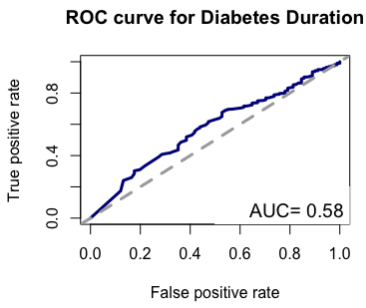


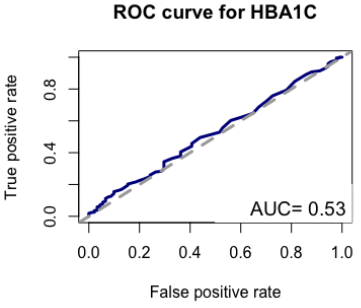

**Figure 5.** Tree diagram of the hierarquical cluster analysis by distance between samples (A) and PCA plot (B) from the genome-wide DNA methylation analyses, both performed after combat, showing adequate clustering of samples (487 individuals from ANDIS and ANDIU cohorts) by sex.

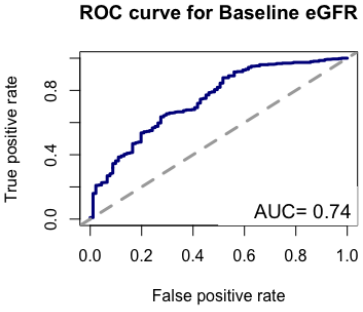

ROC curve for Sex
True positive rate
False positive rate
AUC= 0.51

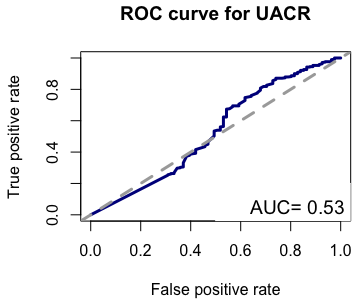

ROC curve for Age
True positive rate
False positive rate
AUC= 0.72

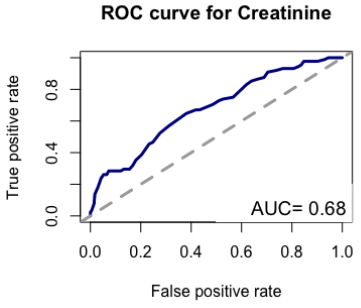

ROC curve for BMI
True positive rate
False positive rate
AUC= 0.54

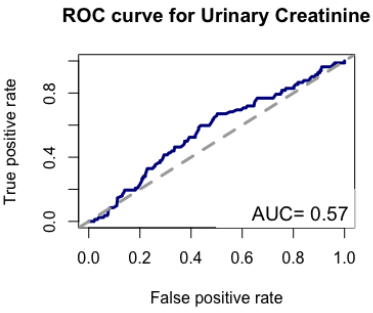

ROC curve for Diabetes Duration
True positive rate
False positive rate
AUC= 0.58

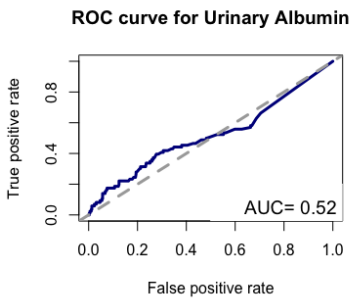

ROC curve for HBA1C
True positive rate
False positive rate
AUC= 0.53

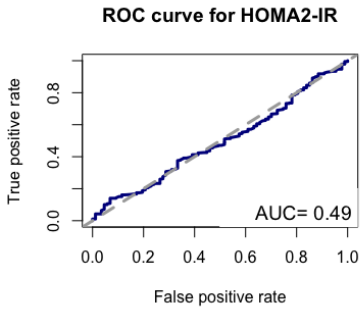

ROC curve for Baseline eGFR
True positive rate
False positive rate
AUC= 0.74

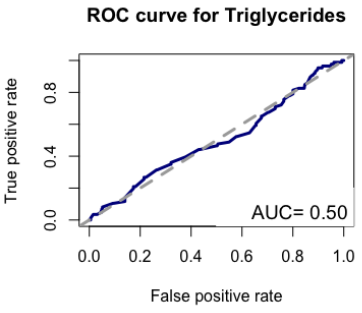

ROC curve for UACR
True positive rate
False positive rate
AUC= 0.53


ROC curve for Creatinine
True positive rate
False positive rate
AUC= 0.68


ROC curve for Urinary Creatinine
True positive rate
False positive rate
AUC= 0.57


ROC curve for Urinary Albumin
True positive rate
False positive rate
AUC= 0.52


ROC curve for HOMA2-IR
True positive rate
False positive rate
AUC= 0.49


ROC curve for Triglycerides
True positive rate
False positive rate
AUC= 0.50

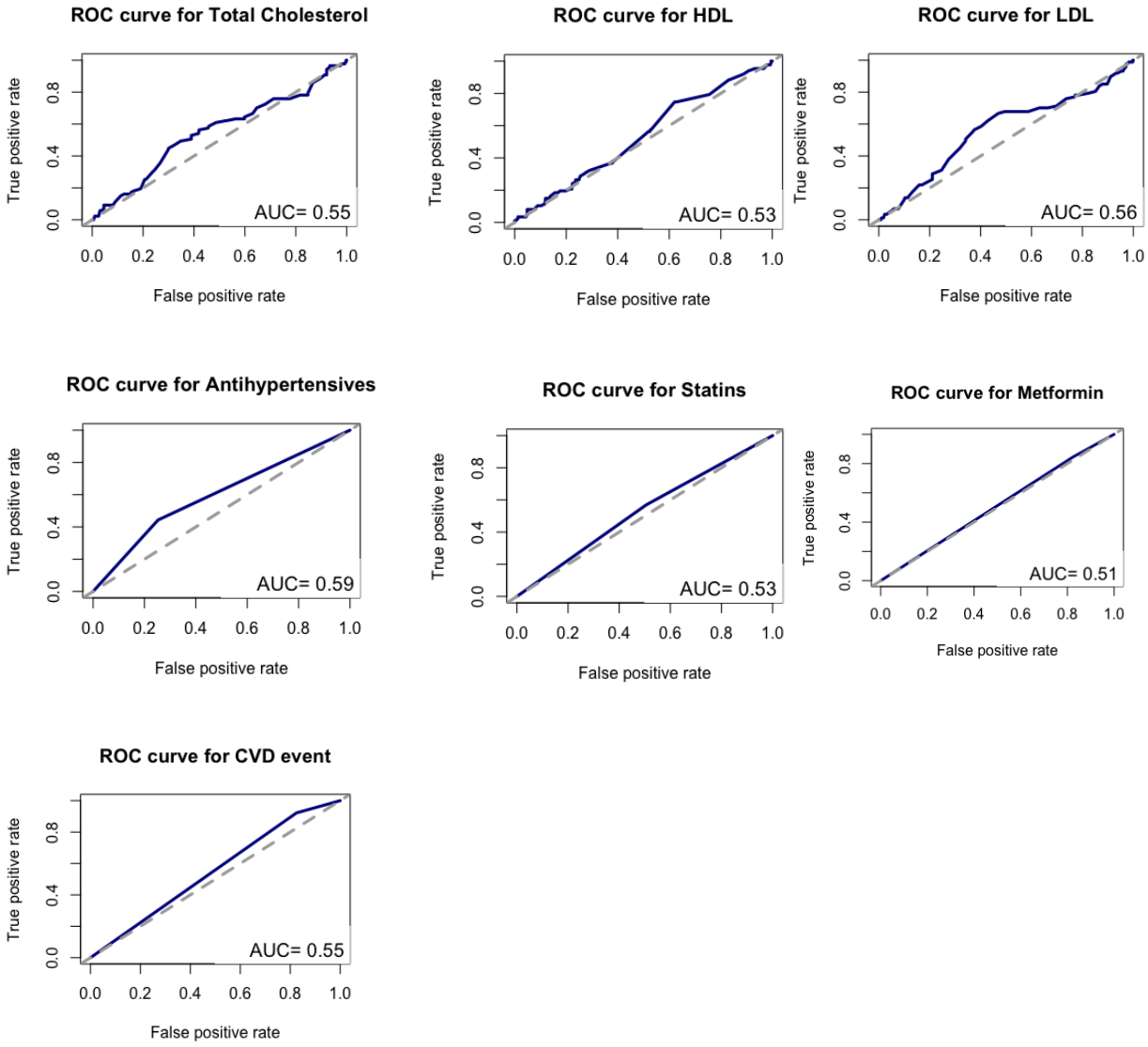



**Figure 6.** ROC curves for the association of each clinical variable at baseline and the outcome (DKD development after 11.5 years of follow-up), based on general logistic models in the study population composed by 487 individuals from ANDIS and ANDiU cohorts.

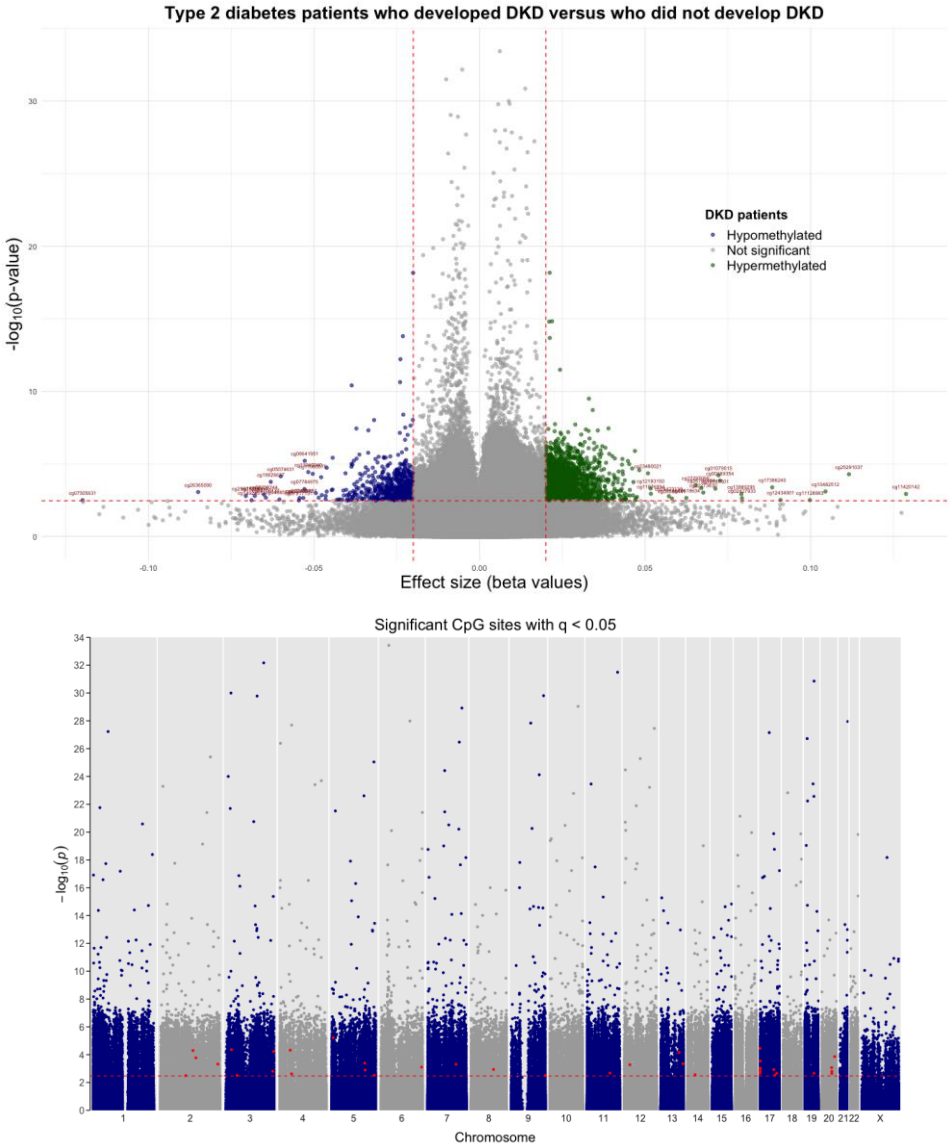
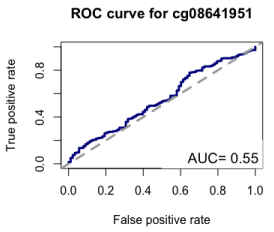
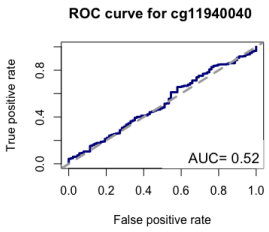
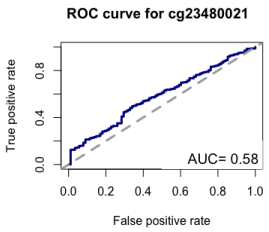
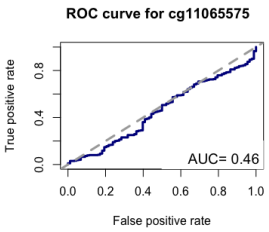


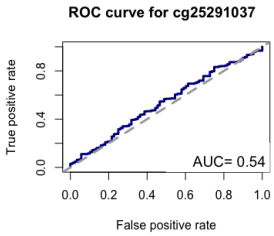

**Figure 7.** Volcano and Manhattan plots for the associations between DNA methylation at baseline and the development of DKD after 11.5 years of follow-up in the population composed by 487 individuals from ANDIS and ANDIU cohorts, adjusted for age, sex, BMI, and baseline eGFR, showing 129,811 significant CpG sites with $p < 0.05$. Red dashed horizontal lines indicate methylome-wide significance ($q < 0.05$), and red dashed vertical lines indicate cutoff on effect size (difference in DNA methylation). Above the horizontal red line in the Manhattan plot are the 56411 significant CpG sites (FDR < 5%, $q < 0.05$). To the left of the first vertical red line are the CpGs which are hypomethylated in patients who developed DKD in relation to patients who did not develop DKD, and to the right of the second vertical red line are the CpGs which are hypermethylated in patients who developed DKD in relation to patients who did not develop DKD Beta-coefficients in the volcano plot are shown as the difference from patients who developed DKD compared to patients who did not develop DKD. The 37 CpGs with effect size higher than 5% in both directions and $q < 0.05$ are identified in red.

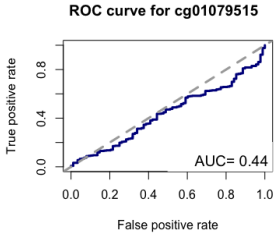

ROC curve for cg08641951
True positive rate
False positive rate
AUC= 0.55

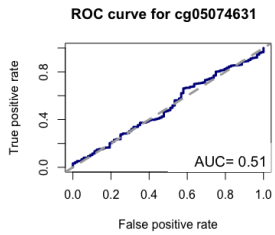

ROC curve for cg11940040
True positive rate
False positive rate
AUC= 0.52

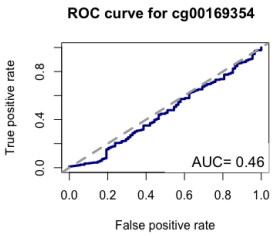

ROC curve for cg23480021
True positive rate
False positive rate
AUC= 0.58

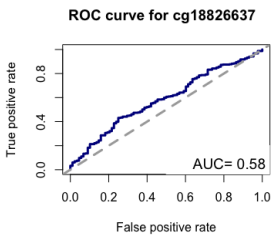

ROC curve for cg11065575
True positive rate
False positive rate
AUC= 0.46

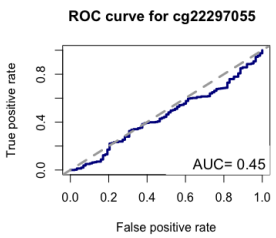

ROC curve for cg25291037
True positive rate
False positive rate
AUC= 0.54

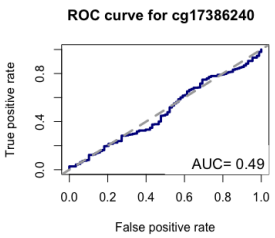

ROC curve for cg01079515
True positive rate
False positive rate
AUC= 0.44

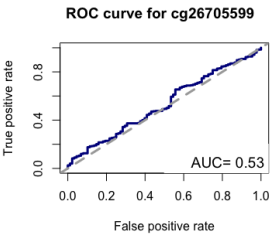

ROC curve for cg05074631
True positive rate
False positive rate
AUC= 0.51

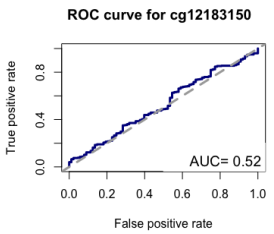

ROC curve for cg00169354
True positive rate
False positive rate
AUC= 0.46

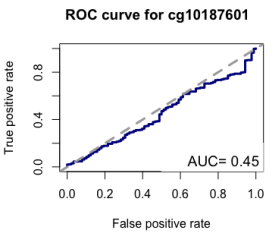

ROC curve for cg18826637
True positive rate
False positive rate
AUC= 0.58

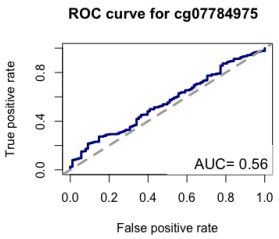

ROC curve for cg22297055
True positive rate
False positive rate
AUC= 0.45

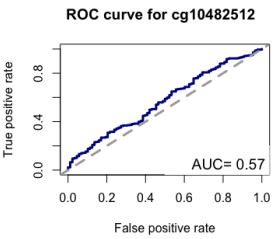

ROC curve for cg17386240
True positive rate
False positive rate
AUC= 0.49

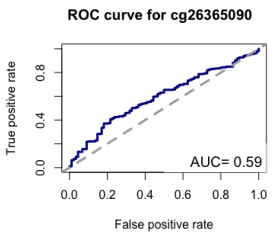

ROC curve for cg26705599
True positive rate
False positive rate
AUC= 0.53

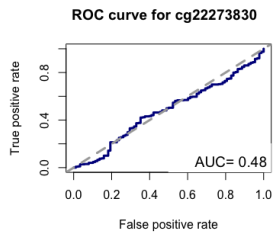

ROC curve for cg12183150
True positive rate
False positive rate
AUC= 0.52

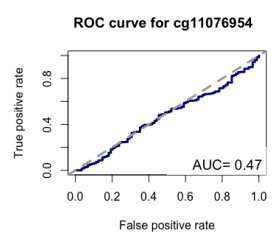

ROC curve for cg10187601
True positive rate
False positive rate
AUC= 0.45

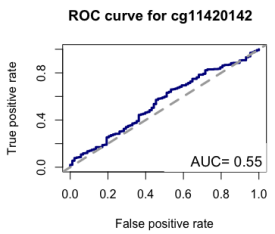

ROC curve for cg07784975
True positive rate
False positive rate
AUC= 0.56

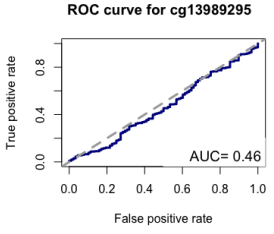

ROC curve for cg10482512
True positive rate
False positive rate
AUC= 0.57

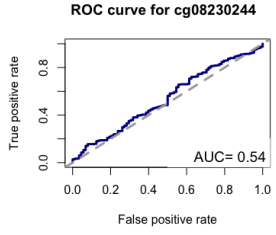

ROC curve for cg26365090
True positive rate
False positive rate
AUC= 0.59

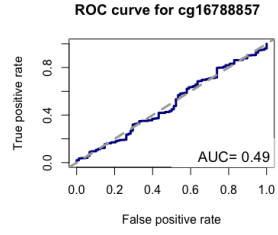

ROC curve for cg22273830
True positive rate
False positive rate
AUC= 0.48

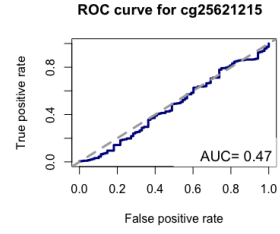

ROC curve for cg11076954
True positive rate
False positive rate
AUC= 0.47


ROC curve for cg11420142
True positive rate
False positive rate
AUC= 0.55


ROC curve for cg13989295
True positive rate
False positive rate
AUC= 0.46


ROC curve for cg08230244
True positive rate
False positive rate
AUC= 0.54


ROC curve for cg16788857
True positive rate
False positive rate
AUC= 0.49


ROC curve for cg25621215
True positive rate
False positive rate
AUC= 0.47

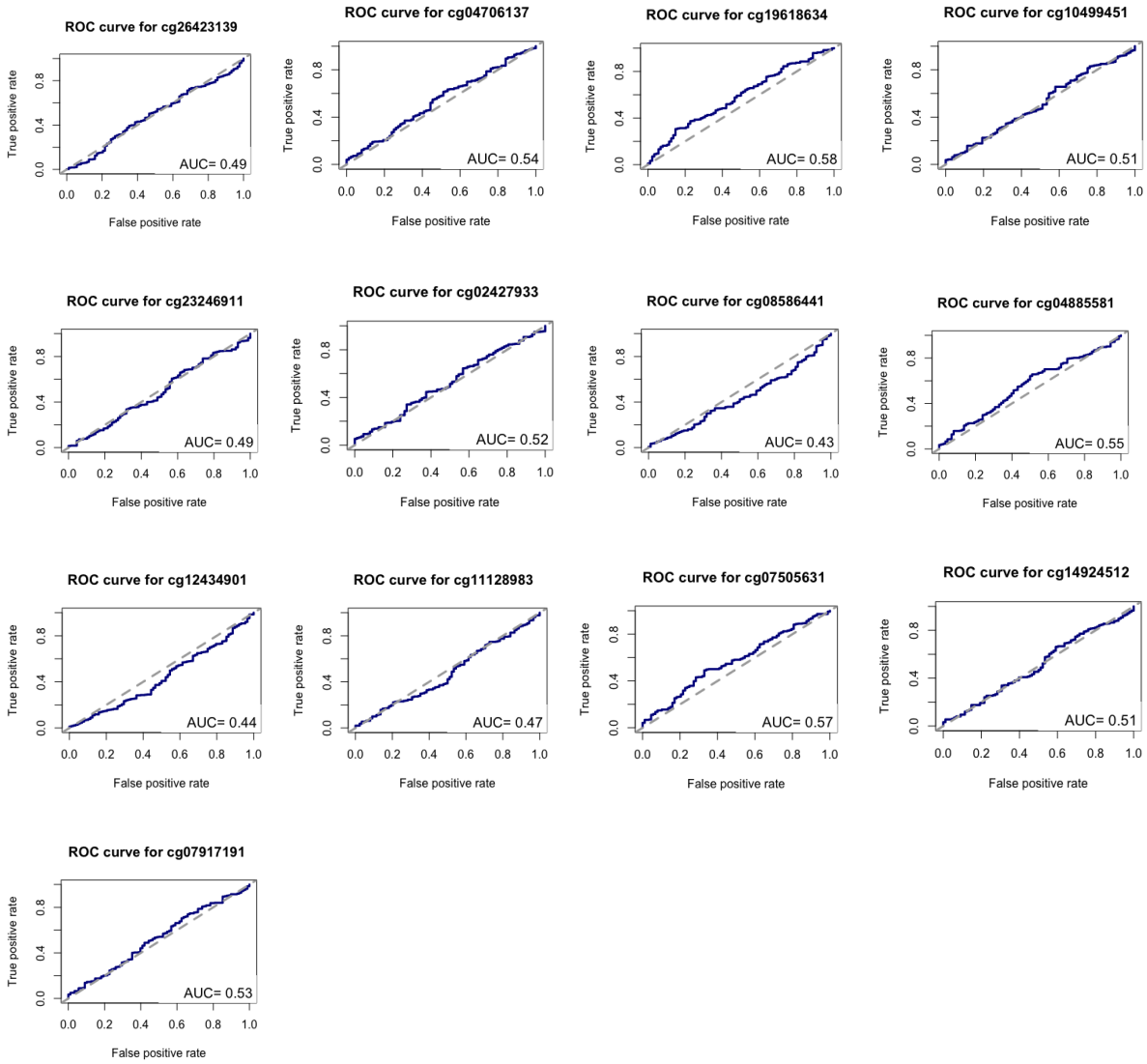
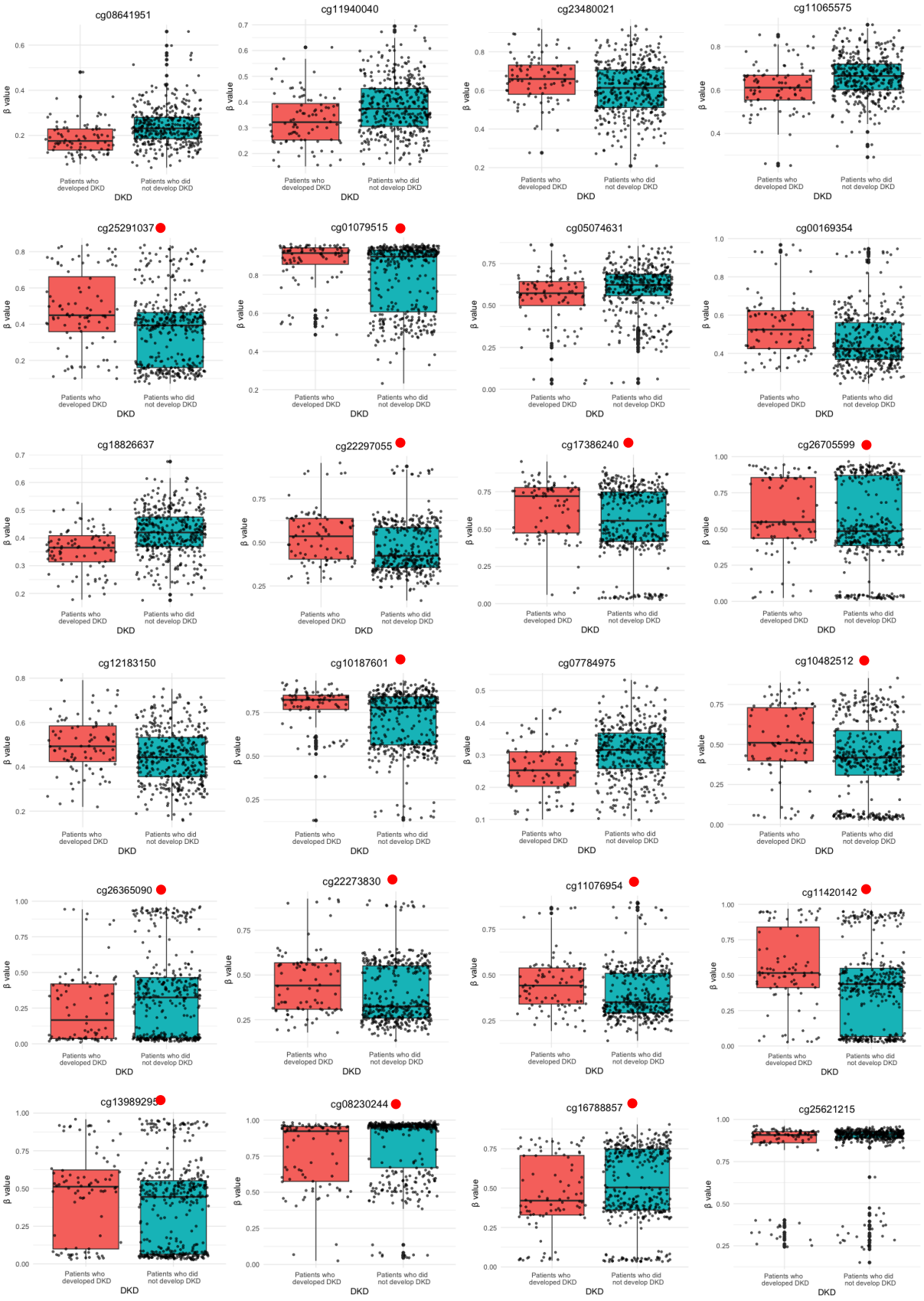



**Figure 8.** ROC curves for the association of DNA methylation levels, measured by beta values, for each of the 37 CpGs with difference in methylation > 5% between patients with T2D who developed DKD and patients with T2D who did not develop DKD, with a significance level of $q < 0.05$ in Model 1a, and $p < 0.05$ in all the other models, and the outcome (DKD), in 487 individuals from the ANDIS and ANDIU cohorts after 11.5 years of follow-up.


cg08641951
cg11940040
cg23480021
cg11065575
cg25291037
cg01079515
cg05074631
cg00169354
cg18826637
cg22297055
cg17386240
cg26705599
cg12183150
cg10187601
cg07784975
cg10482512
cg26365090
cg22273830
cg11076954
cg11420142
cg13989295
cg08230244
cg16788857
cg25621215
β value
DKD
Patients who developed DKD
Patients who did not develop DKD

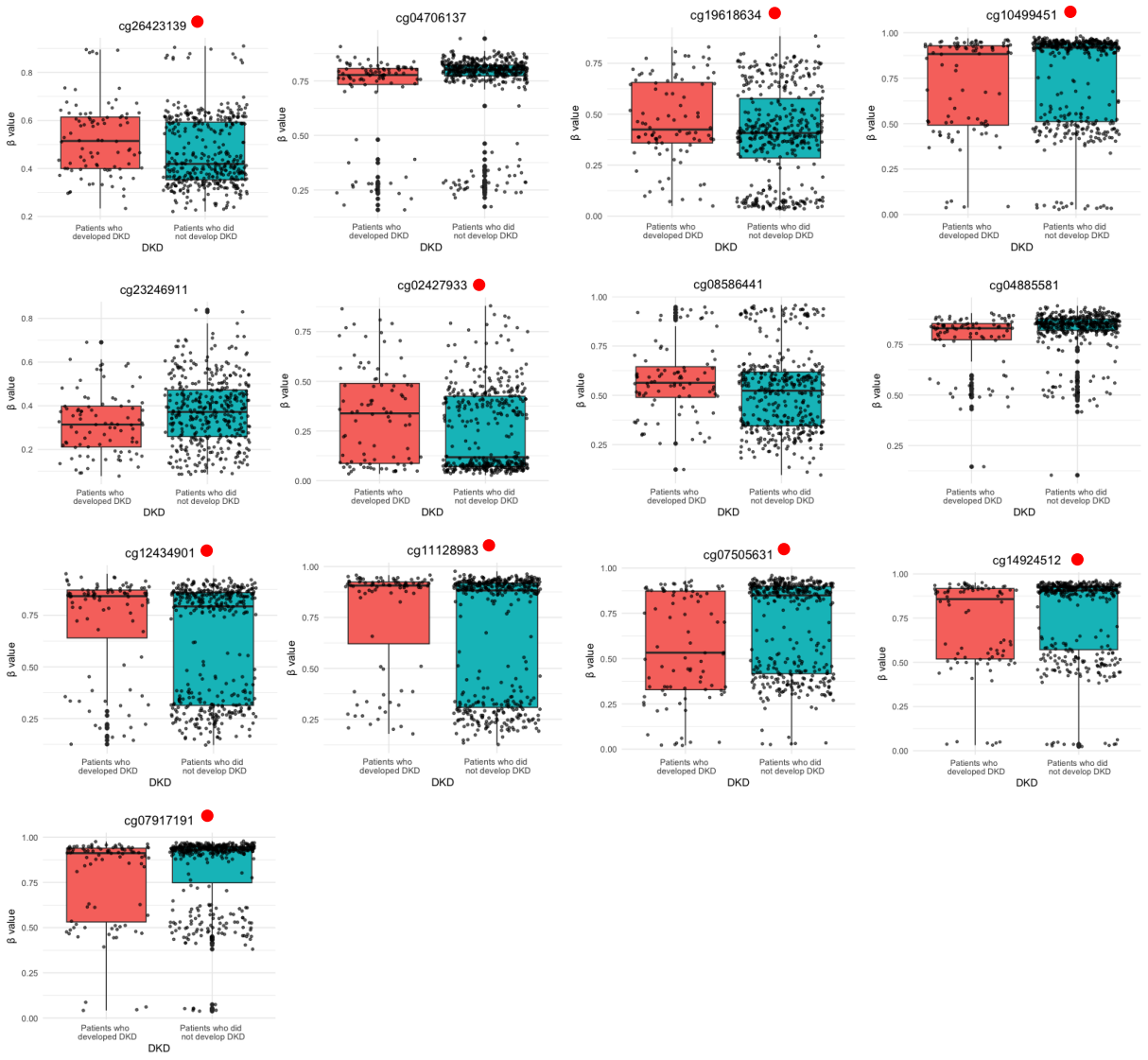



**Figure 9.** Box plots for each of the identified 37 CpGs with difference in methylation > 5% between patients with T2D who developed DKD and patients with T2D who did not develop DKD in a subset of 487 patients from the ANDIS and ANDIU cohorts, with a significance level of $q < 0.05$ in Model 1a, adjusted for age, sex, BMI, and baseline eGFR, and $p < 0.05$ in all the other models. ● CpGs that present a SNPs pattern.

## 9. Popular Science summary

Diabetes is a global health crisis, with over 537 million people diagnosed worldwide, causing 6.7 million deaths in 2021 alone. Type 2 diabetes (T2D) is the most common type and can lead to serious complications such as diabetic kidney disease (DKD), the main cause of end-stage renal disease globally. DKD is a significant health and economic burden. To improve DKD's prediction in T2D patients, there is a critical need to develop new biomarkers.

Our study aimed to identify novel epigenetic biomarkers associated with the development of DKD in 487 newly diagnosed T2D patients who were followed up for 11.5 years. Epigenetic biomarkers, such as DNA methylation, refer to modifications on DNA molecules that affect how genes are expressed, which can be used to predict an individual's risk of developing certain diseases.

Genome-wide DNA methylation analysis, which analyses 850,000 methylation markers, was performed for all patients at baseline, and its association with the development of DKD was assessed.

37 differentially methylated CpGs (regions of DNA) significantly associated with the development of DKD were found. Hypermethylation in 20 of these CpGs was observed in patients who developed DKD. These CpGs are annotated to genes involved in critical biological processes, including *ADAMTS16*, involved in the regulation of systemic arterial blood pressure; *RPH3AL* in positive regulation of insulin secretion, and *EHMT1* in DNA methylation.

The results suggest that DNA methylation can be a potential valuable biomarker in predicting DKD for T2D patients. This study is the first to investigate the association between DNA methylation and future DKD. While the findings are promising, the study design needs to be refined and validated in different populations. The identification of epigenetic biomarkers could help to improve the risk prediction of DKD in T2D patients, ultimately leading to earlier diagnosis and better management of this debilitating condition.

## 10. Acknowledgements

I would like to take this opportunity to express my sincere gratitude to several individuals who have contributed to the completion of my thesis.

First and foremost, I would like to thank my supervisor Charlotte Ling and co-supervisor Sonia Garcia-Calzon for their invaluable guidance and support throughout the research process. Their expertise, insights, and encouragement have been instrumental in shaping the direction of my work.

I would also like to extend my heartfelt appreciation to my teachers in the MPH course, who have provided me with an outstanding education and a strong foundation for my research. Their dedication and commitment to excellence have been an inspiration to me.

To my dear colleagues, thank you for your unwavering support, encouragement, and laughter throughout my academic journey. Our diverse background and cultural exchange made this experience all the more meaningful and enjoyable.

To my family, I am forever grateful for your love, support, and understanding. The distance was felt strongly, but your belief in me has given me the strength to overcome every obstacle and achieve my goals.

Lastly, but most importantly, I would like to express my deepest gratitude to my wife Jessica. Your unwavering love, support, and encouragement have been the foundation of my success. Your patience, understanding, and advice have been a constant source of motivation and inspiration. I cannot thank you enough for being my best friend throughout this journey.