**Supplementary Fig 1. Workflow of the study**

- Both type 2 diabetes cases/non-diabetic control are used
- Only non-diabetic controls are used

<table>
<thead>
<tr>
<th>Methylation data available</th>
<th>MWAS</th>
<th>Risk Prediction</th>
<th>Metabolic traits analysis</th>
<th>Mendelian Randomization</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNUH</td>
<td>N=429 (232/197)</td>
<td>N=197</td>
<td>N=129*</td>
<td></td>
</tr>
<tr>
<td>KoGES</td>
<td>N=356 (79/277)</td>
<td>N=356</td>
<td>N=356</td>
<td></td>
</tr>
<tr>
<td>HTS</td>
<td></td>
<td>N=463</td>
<td>N=463</td>
<td></td>
</tr>
</tbody>
</table>

*129 controls are only available with genotype data out of 197 controls.*
Supplementary Fig 2. Visualization of data recruitment

Data visualization of T2D case/Non-diabetic Controls (N=429)

Duplicated Samples were used for batch effect adjustment.
(N=2 (Experimented in both 2016 and 2018) + 2 (Experimented in both 2017, 2018))

T2D Case (N=232)
Control (N=197)

Experimented in 2016 (N=147)
Experimented in 2018 (N=145+4)
Experimented in 2017 (N=137)

N=

145
2
85
60
2
135

Data visualization of DKD cases/Non-DKD diabetic controls (N=167)

DKD cases
Non-DKD diabetic controls
45
34

DKD cases
Non-DKD diabetic controls
42
46
Supplementary Fig 3. QC procedure of DNA methylation data

**SNUH**

- QC procedures for Type 2 diabetes MWAS (N=429)
- **Illumina EPIC chip**
  - Probe N=866,895
  - 749,315
- **Normalization**
  - 796,599
- **excluded**
  - Sex chromosome 17,220
  - Non-CpG sites 1,241
  - Missing value CpG sites 28,822

**KoGES**

- QC procedures for DKD MWAS (N=167)
- **Illumina 450K chip**
  - Probe N=485,577
  - 442,598
- **Normalization**
  - 427,808
- **excluded**
  - Sex chromosome 10,124
  - Non-CpG sites 1,397
  - Missing value CpG sites 7,014

**HTS**

- **1st (N=370)**
  - **Illumina 450K chip**
    - Probe N=485,577
    - 480,817
  - **Normalization**
    - 843,036
- **excluded**
  - Sex chromosome 11,037
  - Non-CpG sites 3,073
  - Missing value CpG sites 3,707
  - 462,980
- **2nd (N=140)**
  - **Illumina EPIC chip**
    - Probe N=866,895
    - 794,474
  - **Normalization**
    - 843,036
- **excluded**
  - Sex chromosome 18,712
  - Non-CpG sites 2,854
  - Missing value CpG sites 26,996
  - 462,980
Supplementary Fig 4. QQ-plot of T2D/DKD MWAS

a. Type 2 diabetes

b. DKD
Supplementary Fig 5. Correlation between results from EPIC-Norfolk study and from T2D MWAS

Red dot indicates that P from EPIC-Norfolk study is under methylome-wide significance level ($P < 9 \times 10^{-8}$); blue dot indicates that P from EPIC-Norfolk study is under nominal significance level ($P < 0.05$) and Black dot indicates that P from EPIC-Norfolk study is over nominal significance level.

Two DMRs (cg26823705 and cg26974062) were not available in EPIC-Norfolk study.
Supplementary Note 1. Process of batch effect correction

1. Duplicate Correlation analysis
   Using different batch correction combination
   [Experiment slide number (A), Experiment year (B),
   person who experimented (C)]

<table>
<thead>
<tr>
<th>Batch Combination</th>
<th>no batch</th>
<th>A</th>
<th>B</th>
<th>A+B</th>
<th>A+B+C</th>
<th>B+A</th>
<th>B+A+C</th>
<th>B+C</th>
<th>B+C+A</th>
<th>C</th>
<th>C+B</th>
<th>C+B+A</th>
<th>C+A</th>
<th>C+B+A+B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duplicate correlation</td>
<td>0.213</td>
<td>0.094</td>
<td>0.262</td>
<td>0.116</td>
<td>0.113</td>
<td>0.166</td>
<td>0.163</td>
<td>0.262</td>
<td>0.166</td>
<td>0.226</td>
<td>0.261</td>
<td>0.166</td>
<td>0.132</td>
<td>0.138</td>
</tr>
</tbody>
</table>

2. Selected Best batch effect correction

3. Principle Component Analysis (PCA)

3. EWAS using only controls from different experiment year
   Phenotype: experiment year

4. Differentially methylated at statistical significant level $9 \times 10^{-8}$

<table>
<thead>
<tr>
<th>Position</th>
<th>Coefficients</th>
<th>P-value</th>
<th>Nearby gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>cg09547767</td>
<td>0.015</td>
<td>8.65E-08</td>
<td>BMPRA</td>
</tr>
</tbody>
</table>

Cg09547764 was reported to be associated with hepatocellular carcinoma according to EWAS Atlas.
As our samples were recruited and extracted 3 times through 2016-2018, removing batch effect is an important part of quality control. We selected the best batch effect correction method using duplicated samples that experimented in different year and tried to eliminate any confounders using the R package SVA\textsuperscript{1}.

We used ‘duplicate correlation’ function in R package limma, which was originally designed to estimate the correlation between duplicate spots (regularly spaced replicate spots on the same array) or between technical replicates from a series of arrays. We adjusted each combination of experiment slide, experiment year and experimented person and investigated duplicated correlation. Duplicate correlation function is mainly designed to find tissue specific difference of the individual, where both tissue is experimented in same time. Though getting duplicate correlation value between the samples experimented separately is not fitted for the purpose of the function and therefore correlation value is not higher, we just used the correlation value as the reference of selecting best batch effect correction combination. Only adjusting experiment year showed best duplicate samples correlation. To ascertain that the batch effect was well removed, we verified in two ways. First, principle component analysis (PCA) was applied to data before adjusted and data after adjusted. As seen in SN1.Fig1, no longer batch effect was observed after adjusting batch effect correction.
Secondly, we tried to find if there are differentially methylated sites between control groups from 2017 recruited samples and from 2018’s. We found cg09547764 was differentially methylated in 9 × 10⁻⁸ significance level. Reported CpG sites were searched in EWAS Atals². cg09547767 was previously reported to be associated with hepatocellular carcinoma³, not associated with T2D or related traits. We concluded that batch effect was successfully removed to conduct T2D and DKD MWAS.
SN1. Reference


**Supplementary Note 2. Evaluation imputation accuracy of DNAm data**

Imputation performance was first evaluated using HTS 1 dataset which is assayed by Illumina EPIC beadchip platform. First of all, HTS 1 dataset was divided into two parts: reference set and test set. For test set, DNAm markers only included in EPIC platform was deleted. Next, imputation of deleted markers of test set using reference sets performed. To evaluate imputation performance of ‘methyLImp’, difference of original β-value from EPIC chip and imputed β-value of test set were compared. Root mean square error was 0.0087 and Pearson’s correlation was 0.99 (**SN2. Figure 1**). As ‘methyLImp’ had showed remarkable imputation performance, remain datasets are imputed by ‘methyLImp’ using HTS 1 as a reference.

**SN2. Figure 1. Scatter plot of methylation beta-value**
Supplementary Note 3. Selecting genetic markers for the MR analysis

The participants were genotyped using Illumina Omni1 (KoGES, rural), Affymatrix 6.0 (KoGES, urban; HTS) and 5.0 (KoGES, KARE) genotyping arrays. The single nucleotide polymorphisms (SNPs) were filtered by following genotype quality control (QC) criteria: (1) genotyping call rate >0.95 (2) minor allele frequency (MAF) >0.01, (3) P value in Hardy–Weinberg equilibrium (HWE) testing >10e-6 and individual QC criteria: sample call rate >0.9. Imputation for variants that were not directly genotyped was performed using IMPUTE 2.0 software. Genotype data after QC was imputed using the Korean data from Korean Reference Genome (KRG) which initiated by Center of Genome Science (CGS) of Korea National Instituted of Health (KNIH) in 2012 and East Asian (EAS) data from the 1,000 Genomes Project Phase 3 (NCBI build 37) as the reference.

SN3. Table 1. Number of SNPs used for IV in Mendelian randomization analysis

<table>
<thead>
<tr>
<th></th>
<th>BMI</th>
<th>Glucose</th>
<th>HbA1c</th>
<th>eGFR</th>
<th>T2D</th>
<th>DKD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Number of SNP previously reported</td>
<td>5834</td>
<td>126</td>
<td>180</td>
<td>1,117</td>
<td>2280</td>
<td>16</td>
</tr>
<tr>
<td>2. Available in our data</td>
<td>2477</td>
<td>115</td>
<td>175</td>
<td>838</td>
<td>1207</td>
<td>12</td>
</tr>
<tr>
<td>3. Replicated with metabolic traits/T2D and DKD in our data</td>
<td>218</td>
<td>32</td>
<td>33</td>
<td>33</td>
<td>192</td>
<td>2</td>
</tr>
<tr>
<td>4. Not associated with T2D/metabolic traits in our data</td>
<td>218</td>
<td>27</td>
<td>18</td>
<td>33</td>
<td>64</td>
<td>1</td>
</tr>
</tbody>
</table>

* Reported SNPs on GWAS catalogue was examined on March, 2020.

All SNPs reported to be associated with metabolic traits and T2D/DKD on GWAS catalogue was extracted. (SN3. Table 1, Stage 1) Next, availability of SNPs in Korean data after QC was checked (Stage 2). We only selected SNPs that replicated with target traits (Stage 3), and not associated with consequence traits (For metabolic trait, T2D/DKD; For for T2D/DKD, metabolic traits) in nominal significance (P <0.05). Selected SNPs was not showed association with potential confounders (age,sex and BMI).

SN3. Reference

Supplementary Note 4. Replication of known T2D associated DMRs

Table 1) Comparison of reported and replicated effect size of CpGs reported multiple times.

<table>
<thead>
<tr>
<th>CpG</th>
<th>Reference a</th>
<th>Reported effect size b</th>
<th>Reported $P$</th>
<th>Effect size in this study ($\beta$-values)</th>
<th>$P$ in this study</th>
</tr>
</thead>
<tbody>
<tr>
<td>cg19693031</td>
<td>[7]</td>
<td>-0.58</td>
<td>1.53x10^{-19}</td>
<td>-0.061</td>
<td>2.65x10^{-9}</td>
</tr>
<tr>
<td>cg19693031</td>
<td>[8]</td>
<td>-0.08</td>
<td>1.00x10^{-13}</td>
<td>-0.061</td>
<td>2.65x10^{-9}</td>
</tr>
<tr>
<td>cg19693031</td>
<td>[10]</td>
<td>-0.65</td>
<td>2.70x10^{-21}</td>
<td>-0.061</td>
<td>2.65x10^{-9}</td>
</tr>
<tr>
<td>cg19693031</td>
<td>[12]</td>
<td>-0.03</td>
<td>4.50x10^{-7}</td>
<td>-0.061</td>
<td>2.65x10^{-9}</td>
</tr>
<tr>
<td>cg19693031</td>
<td>[13]</td>
<td>-0.13</td>
<td>7.30x10^{-16}</td>
<td>-0.061</td>
<td>2.65x10^{-9}</td>
</tr>
<tr>
<td>cg19266329</td>
<td>[7]</td>
<td>0.50</td>
<td>9.98x10^{-11}</td>
<td>-0.023</td>
<td>0.001</td>
</tr>
<tr>
<td>cg00574958</td>
<td>[7]</td>
<td>0.38</td>
<td>4.29x10^{-9}</td>
<td>-0.016</td>
<td>5.40x10^{-10}</td>
</tr>
<tr>
<td>cg00574958</td>
<td>[10]</td>
<td>-0.37</td>
<td>5.20x10^{-9}</td>
<td>-0.016</td>
<td>5.40x10^{-10}</td>
</tr>
<tr>
<td>cg17058475</td>
<td>[7]</td>
<td>-0.37</td>
<td>3.89x10^{-9}</td>
<td>-0.018</td>
<td>1.10x10^{-6}</td>
</tr>
<tr>
<td>cg14597545</td>
<td>[7]</td>
<td>-0.69</td>
<td>1.50x10^{-8}</td>
<td>0.012</td>
<td>0.013</td>
</tr>
<tr>
<td>cg11024682</td>
<td>[8]</td>
<td>0.06</td>
<td>8.40x10^{-9}</td>
<td>0.016</td>
<td>2.89x10^{-4}</td>
</tr>
<tr>
<td>cg11024682</td>
<td>[10]</td>
<td>0.44</td>
<td>6.00x10^{-10}</td>
<td>0.016</td>
<td>2.89x10^{-4}</td>
</tr>
<tr>
<td>cg02650017</td>
<td>[8]</td>
<td>-0.06</td>
<td>2.10x10^{-9}</td>
<td>-0.005</td>
<td>2.83x10^{-4}</td>
</tr>
<tr>
<td>cg26836479</td>
<td>[7]</td>
<td>0.56</td>
<td>1.05x10^{-7}</td>
<td>-0.004</td>
<td>0.007</td>
</tr>
<tr>
<td>cg06500161</td>
<td>[7]</td>
<td>0.39</td>
<td>9.43x10^{-10}</td>
<td>0.024</td>
<td>2.65x10^{-9}</td>
</tr>
<tr>
<td>cg06500161</td>
<td>[8]</td>
<td>0.08</td>
<td>2.20x10^{-13}</td>
<td>0.024</td>
<td>2.65x10^{-9}</td>
</tr>
<tr>
<td>cg06500161</td>
<td>[10]</td>
<td>0.50</td>
<td>6.40x10^{-14}</td>
<td>0.024</td>
<td>2.65x10^{-9}</td>
</tr>
<tr>
<td>cg04816311</td>
<td>[7]</td>
<td>-0.50</td>
<td>5.47x10^{-8}</td>
<td>0.016</td>
<td>0.018</td>
</tr>
<tr>
<td>cg04816311</td>
<td>[10]</td>
<td>0.41</td>
<td>1.70x10^{-8}</td>
<td>0.016</td>
<td>0.018</td>
</tr>
</tbody>
</table>

a Reference number is based on the main text.

b Reported odds ratio (or risk ratio) was transformed to effect size.
We compared the previously reported effect size and replicated effect size in our Korean data. cg19693031 was constantly hypo-methylated in T2D patients. cg06500161 and cg11024682 was constantly hyper-methylated. Our replication was consistent with previous studies. Although the effect of cg00574958 and cg04816311 were controversial in previous studies.

SN4 Figure 1. QQ plot of reported T2D associated CpGs