

# Enhanced resolution profiling in twins reveals differential methylation signatures of type 2 diabetes with links to its complications



Colette Christiansen,<sup>a,p,\*</sup> Louis Potier,<sup>b</sup> Tiphaine C. Martin,<sup>c</sup> Sergio Villicaña,<sup>a</sup> Juan E. Castillo-Fernandez,<sup>a</sup> Massimo Mangino,<sup>a</sup> Cristina Menni,<sup>a</sup> Pei-Chien Tsai,<sup>a,d,e</sup> Purdey J. Campbell,<sup>f</sup> Shelby Mullin,<sup>f,g</sup> Juan R. Ordoñana,<sup>h</sup> Olga Monteagudo,<sup>h</sup> Perminder S. Sachdev,<sup>i</sup> Karen A. Mather,<sup>i</sup> Julian N. Trollor,<sup>i</sup> Kirsi H. Pietilainen,<sup>i,q</sup> Miina Ollikainen,<sup>k,r</sup> Christine Dalgård,<sup>l</sup> Kirsten Kyvik,<sup>l</sup> Kaare Christensen,<sup>l</sup> Jenny van Dongen,<sup>m</sup> Gonneke Willemsen,<sup>m</sup> Dorret I. Boomsma,<sup>m</sup> Patrik K. E. Magnusson,<sup>n</sup> Nancy L. Pedersen,<sup>n</sup> Scott G. Wilson,<sup>a,f,g</sup> Elin Grundberg,<sup>o</sup> Tim D. Spector,<sup>a</sup> and Jordana T. Bell<sup>p,\*\*</sup>



<sup>a</sup>King's College London, UK

<sup>b</sup>APHP, Paris Cité University, INSERM, Paris, France

<sup>c</sup>Icahn School of Medicine at Mount Sinai, USA

<sup>d</sup>Department of Biomedical Sciences, Chang Gung University, Taoyuan City, Taiwan

<sup>e</sup>Molecular Infectious Disease Research Center, Chang Gung Memorial Hospital, Taoyuan City, Taiwan

<sup>f</sup>Department of Endocrinology & Diabetes, Sir Charles Gairdner Hospital, Nedlands, WA, Australia

<sup>g</sup>School of Biomedical Sciences, University of Western Australia, Crawley, WA, 6009, Australia

<sup>h</sup>University of Murcia, Spain

<sup>i</sup>University of New South Wales, Australia

<sup>j</sup>Obesity Research Unit, Research Program for Clinical and Molecular Metabolism, Faculty of Medicine, University of Helsinki, Finland

<sup>k</sup>Minerva Foundation Institute for Medical Research, Helsinki, Finland

<sup>l</sup>University of Southern Denmark, Denmark

<sup>m</sup>Department of Biological Psychology, Vrije Universiteit Amsterdam, the Netherlands

<sup>n</sup>Karolinska Institute, Sweden

<sup>o</sup>Children's Mercy Kansas City, USA

<sup>p</sup>The Open University, Milton Keynes, UK

<sup>q</sup>HealthyWeightHub, Abdominal Center, Helsinki University Hospital and University of Helsinki, Finland

<sup>r</sup>Institute for Molecular Medicine Finland, FIMM, HiLIFE, University of Helsinki, Finland

## Summary

**Background** Type 2 diabetes (T2D) susceptibility is influenced by genetic and environmental factors. Previous findings suggest DNA methylation as a potential mechanism in T2D pathogenesis and progression.

**Methods** We profiled DNA methylation in 248 blood samples from participants of European ancestry from 7 twin cohorts using a methylation sequencing platform targeting regulatory genomic regions encompassing 2,048,698 CpG sites.

**Findings** We find and replicate 3 previously unreported T2D differentially methylated CpG positions (T2D-DMPs) at FDR 5% in *RGL3*, *NGB* and *OTX2*, and 20 signals at FDR 25%, of which 14 replicated. Integrating genetic variation and T2D-discordant monozygotic twin analyses, we identify both genetic-based and genetic-independent T2D-DMPs. The signals annotate to genes with established GWAS and EWAS links to T2D and its complications, including blood pressure (*RGL3*) and eye disease (*OTX2*).

**Interpretation** The results help to improve our understanding of T2D disease pathogenesis and progression and may provide biomarkers for its complications.

**Funding** Funding acknowledgements for each cohort can be found in the [Supplementary Note](#).

**Copyright** © 2024 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

**Keywords:** DNA methylation; Type 2 diabetes; Genetics; Twins

eBioMedicine  
2024;103: 105096  
Published Online xxx  
<https://doi.org/10.1016/j.ebiom.2024.105096>

\*Corresponding author. King's College London, UK.

\*\*Corresponding author.

E-mail addresses: [colette.christiansen@open.ac.uk](mailto:colette.christiansen@open.ac.uk) (C. Christiansen), [jordana.bell@kcl.ac.uk](mailto:jordana.bell@kcl.ac.uk) (J.T. Bell).

### Research in context

#### Evidence before this study

Type 2 Diabetes is the most common metabolic disease with rapidly rising prevalence worldwide. DNA methylation has been proposed as a mechanism to mediate genetic and environmental risk factors in development and progression of T2D. Multiple DNA methylation alterations have been identified in individuals with T2D, but efforts to date have been predominantly based on assays that only explore <2% of the human methylome.

#### Added value of this study

This study explored the blood methylome in T2D targeting regulatory and functional genomic regions. The approach allowed for characterisation of genomic regions that display highly variable methylation signatures, and a subset of variable signals differentiated individuals with T2D. The study was able to assess whether the T2D differential methylation

signatures had a genetic or non-genetic basis in genetically identical twins discordant for T2D. The peak T2D differential methylation signals targeted genes with links to clinical complications in T2D, and follow-up suggested that some methylation signals are altered specifically in individuals with T2D who develop complications.

#### Implications of all the available evidence

Distinct alterations to the human methylome occur with T2D. A subset of effects have a genetic basis or occur prior to T2D, suggesting a role for DNA methylation in T2D development. This study identifies evidence for a relationship between T2D methylation signals and development of T2D complications, which implicates DNA methylation in T2D progression. The results suggest that DNA methylation levels may be useful markers for risk of developing T2D complications.

## Introduction

Type 2 diabetes (T2D) prevalence is rising globally from an estimated 108 million cases in 1980<sup>1</sup> to projections of around 700 million by 2045.<sup>2</sup> It is a global health concern as chronic exposure to hyperglycemia induces clinical complications, including cardiovascular disease, visual impairment, lower limb amputation and renal disease. There is a clear genetic component to T2D with the most recent and largest multi-ancestry genome-wide association meta-analysis identifying 568 susceptibility loci, which account for around 50% of T2D heritability.<sup>3</sup> In the EURODISCOTWIN consortium, which included nearly 70,000 twins, the heritability for T2D was estimated at 72% (95% confidence interval 61–78%).<sup>4</sup> However, the rise in T2D also mirrors rising obesity rates globally, where the number of obese adults worldwide has tripled since 1975.<sup>1</sup> This in part reflects changing lifestyles, with diet and lack of exercise as two of the environmental factors that contribute to disease pathogenesis. Epigenetic alterations have been proposed as a candidate mechanism to mediate genetic and environmental effects leading to T2D. Although multiple studies have linked DNA methylation changes to T2D, the role of epigenetics in T2D remains incompletely clear, including whether methylation is causal or secondary to disease onset, and how it relates to T2D complications.

Previous work has identified many differentially methylated CpG sites between individuals with T2D and those without, suggesting that DNA methylation could be a possible epigenetic mechanism underlying T2D pathogenesis. Multiple studies have been carried out in whole blood<sup>5–9</sup> and in T2D relevant tissues including pancreatic islet cells,<sup>10–15</sup> liver<sup>16</sup> and adipose<sup>17,18</sup> tissue.

The largest study to date was a longitudinal study in whole blood in 13,535 Indian Asian and 7066 European participants,<sup>5</sup> contrasting DNA methylation profiles based on the Infinium HumanMethylation450 Bead-Chip array (450k) between individuals who developed T2D over 8 years (1608 Indian Asians and 306 Europeans) and those who did not. The findings replicated previous differential methylation T2D signals in *TCF7L2* and *KCNQ1*, and found previously unreported signals including in *TXNIP*, *ABCG1* and *SREBF1*, which have been subsequently replicated by other blood studies with more than 100 T2D cases.<sup>6–9</sup> More recent studies have also identified and replicated differential methylation in T2D at *CPT1A*.<sup>9</sup> It is difficult to disentangle cause and effect, but the findings by Chambers et al.<sup>5</sup> suggest that differential methylation at CpG sites in five regions (*TXNIP*, *SOCS3*, *PHOSPHO1*, *ABCG1*, *SREBF1*) plays a contributing role to future development of T2D over an 8 year follow up. Overall DNA methylation at *TXNIP*, *ABCG1*, *SREBF1*, *CPT1A*, *TCF7L2* and *KCNQ1* has been found to be associated with T2D in at least two independent whole-blood datasets. Except for signals in *TCF7L2* and *KCNQ1*, two genes that also harbour genetic variants associated with T2D, the T2D blood-based DNA methylation signals appear to be tissue specific. T2D blood-based DNA methylation studies have not identified consistent association results. The only gene for which most studies report differential DNA methylation in T2D is *TXNIP*.

Most T2D DNA methylation studies to date have been carried out in unrelated individuals, but there have been a small number of studies in families and within monozygotic (MZ) twin pairs. MZ twins in a pair have almost identical genetic variation profiles, as they arise

from a single fertilized egg, and are matched for age and sex, enabling analyses of the role of non-genetic factors on DNA methylation in T2D. These may include DNA methylation signals mediating T2D environmental risk factor effects, or signals that occur as a consequence of T2D. Four studies of MZ twin pairs who were discordant for T2D have examined the association between T2D and DNA methylation. These include two studies in whole blood, where the first was conducted in 17 MZ twin pairs discordant for T2D<sup>19</sup> and identified a hypermethylated region in the promoter of *MALT1*, a gene with a key role in energy and insulin pathways. The second blood-based study of 11 MZ twin pairs discordant for T2D<sup>20</sup> found differential methylation at cg18681426 in *ELOVL5*, a gene involved in the elongation of long-chain polyunsaturated fatty acids. Two further studies in adipose (14 twin pairs<sup>17</sup>), and in adipose and skeletal tissues (11 twin pairs<sup>21</sup>) found no significant results genome-wide after multiple testing correction. In addition to the studies in twins, a study of 39 families was carried out<sup>22</sup>, which replicated the results from studies in unrelated samples of *TXNIP* and *ABCG1*, and also found differential methylation in *SAMD12*.

If the associated differential methylation signals occur as a consequence of T2D, they could unveil mechanisms through which T2D-related clinical complications develop. In contrast to the previous studies on T2D and DNA methylation, there has not been a consistent exploration of the effect of DNA methylation variation on T2D disease progression, especially in the development of T2D complications. In previous work when T2D cases were grouped into 4 subtypes (severe insulin-deficient diabetes, severe insulin-resistant diabetes, mild obesity-related diabetes, and mild age-related diabetes) reflecting different risks for developing T2D complications, the four T2D groups had different DNA methylation signatures in blood.<sup>23</sup> For example, of 95 methylation sites found to discriminate between these four groups, 39 were annotated to genes previously linked to diabetes and related traits, including *TXNIP*.<sup>23</sup> Increased expression of *TXNIP* has also been shown to cause oxidative stress, inflammation and apoptosis in retinal cells,<sup>24</sup> and its inhibition blocks the early stages of diabetic retinopathy.<sup>25</sup> Given these findings a more comprehensive study of DNA methylation changes and T2D complications may yield insights and provide biomarkers for developing complications in T2D progression.

Nearly all previous DNA methylation studies of T2D in blood, with the exception of Yuan et al.,<sup>19</sup> have been carried out utilising the 450k array. However, this methylation array design does not take into account variation in DNA methylation levels across individuals and targets many CpGs that exhibit little inter-individual variability.<sup>26</sup> In particular, the 450k array has a bias for promoter regions, which have lower variability and are

less dynamic in response to exposures.<sup>27</sup> Yuan et al.,<sup>19</sup> explored whole blood DNA methylation profiles determined using methylated DNA immunoprecipitation sequencing (MeDIP-seq). This sequencing technology offers wider genome coverage, but lacks single base pair resolution and shows bias towards methylated regions. In contrast, a recently developed DNA methylation sequencing technology, methyl-C capture sequencing (MCC-seq), targets enhancers and functional areas of the genome leading to an enrichment in regulatory regions. Allum et al.<sup>28</sup> demonstrate that MCC-seq is comparable in accuracy to the 450k array, but better targets disease-relevant regions. In addition to targeting disease-relevant regions, MCC-seq profiles 3.7 million sites, nearly an 8-fold increase on the 450k array. Allum et al.<sup>29</sup> have also utilised the power of MCC-seq profiling to identify differential methylation in functional regions relating to cardiometabolic traits using circulating plasma lipid levels as proxies for cardiometabolic health.

Here we utilise MCC-seq to profile DNA methylation in whole blood samples from European ancestry participants from 7 international twin cohorts,<sup>4</sup> including 113 T2D cases and 135 controls, of which a subset were 74 MZ twin pairs discordant for T2D. The aim was to determine associations between DNA methylation genome-wide and retrospective T2D status (T2D-DMPs). We explored the genetic drivers of T2D, and inclusion of 74 T2D-discordant MZ pairs enabled us to test for non-genetic drivers of T2D. T2D-DMP replication was pursued in an independent dataset of 573 individuals, including 33 T2D cases. T2D-DMPs were also validated using blood 450k profiles in 49 cases and 681 controls for association with T2D, and in 978 individuals for association with fasting blood glucose also profiled in the 450k array. T2D-DMPs were further investigated for their functional relevance by exploring their association with blood metabolomic profiles, and for clinical relevance with respect to the development of T2D-related complications. The results of this study have potential to improve our understanding of the molecular changes mediating genetic and environmental risk effects for T2D, and their interactions, and potentially act as biomarkers for T2D complications.

## Methods

### Study participants

The primary analysis in the study included 248 participants (135 controls and 113 T2D) from 7 international twin study cohorts including the Danish Twin Registry (DTR), The Finnish Twin Cohort Study (FTCS), Murcia Twin Registry (MTR), Netherlands Twin Register (NTR), The Older Australian Twins Study (OATS), The Swedish Twin Registry (STR), and TwinsUK (Table 1, Supplementary Note). The 248 individuals included in the main study were of European ancestry with the exception of two individuals of Indonesian ancestry in

Cohorts	n	Controls	Control BMI <sup>b</sup> (kg.m <sup>b</sup> )	Control Age <sup>a</sup>	Control FBG	T2D	T2D BMI <sup>b</sup> (kg.m <sup>b</sup> )	T2D Age <sup>a</sup>	T2D FBG
TwinsUK	93	48	27.5 ± 5.0	63.5 ± 13.0	4.9 ± 0.5	45	29.8 ± 6.0	60.9 ± 12.8	7.6 ± 4.3
Older Australian Twins Study	11	6	25.6 ± 2.7	69.4 ± 4.2	4.8 ± 1.0	5	30.5 ± 5.2	68.8 ± 4.1	6.7 ± 1.3
Finnish Twin Cohort Study	20	10	28.1 ± 4.8	60.4 ± 11	5.6 ± 0.4	10	33.9 ± 5.6	60.4 ± 11	7.6 ± 3.3
Netherlands Twin Register	46	33	26.2 ± 4.6	50.9 ± 15	5.3 ± 0.5	13	27.2 ± 3.8	59.2 ± 19	9.2 ± 2.8
Murcia Twin Registry	9	5	32.4 ± 2.6	61.1 ± 5.1	n/a	4	31 ± 5.4	62.1 ± 5.3	7.3 ± 1.4
Danish Twin Registry	22	11	28.1 ± 1.9	55.6 ± 6.7	5.9 ± 0.5	11	32 ± 5.7	55.7 ± 6.7	7.4 ± 1.0
Swedish Twin Registry	47	23	25.1 ± 3.9	69.7 ± 6	5.3 ± 0.3	24	28.8 ± 4.3	69.8 ± 5.8	7.6 ± 1.4
<b>Total</b>	<b>248</b>	<b>135</b>	<b>27 ± 4.5</b>	<b>60.7 ± 13.3</b>		<b>113</b>	<b>30 ± 5.4</b>	<b>62.5 ± 12.1</b>	

<sup>a</sup>No significant difference in age overall P = 0.23. <sup>b</sup>Significant overall difference in BMI P < 0.001.

**Table 1: Discovery sample participant characteristics.**

the NTR cohort. Overall, 85% of the sample consisted of monozygotic (MZ) twin pairs, including 74 MZ twin pairs discordant for T2D. The sex proportion varied across cohort samples, ranging from 18% to 100% female, and overall 61% of participants were female. Participants were predominantly sampled in mid-life or at an older age, from the youngest cohort samples in DTR (mean age 55 ± 7) to the oldest cohort samples from OATS (mean age 69 ± 4) and STR (mean age 70 ± 6). Whole blood DNA samples from the 248 participants were profiled for DNA methylation using MCC-seq as described below.

The validation dataset for the study included 978 individuals from the TwinsUK cohort (mean age 57.8 ± 10.7, mean BMI 26.7 kg/m<sup>2</sup> ± 4.7) whose whole blood DNA methylomes were profiled using an alternate method, the 450k array. Of these, 730 participants were analysed as a case-control study of T2D (49 T2D cases and 681 controls), and all 978 individuals were included in the DNA methylation association analysis with fasting blood glucose. Altogether 44 of these participants (19 T2D cases and 25 controls) overlapped with the discovery participants in the primary analysis from the TwinsUK MCC-seq sample.

The replication dataset consisted of 573 TwinsUK individuals with whole blood MCC-seq profiles, of whom 33 had T2D at the time of blood collection (age 54.2 ± 13.4, BMI 25.7 kg/m<sup>2</sup> ± 4.1). These individuals were predominantly female (93%) and the majority were MZ twins. They were free from major diseases at the time of sampling and metabolic health phenotypes in this sample were representative of the population. There was no overlap between the discovery and replication samples.

All study participants provided informed consent. Ethical approval was granted for all the cohort samples. Further details and a description of each cohort are provided in [Supplementary Note 1](#).

### DNA methylation profiles

#### Methyl C capture sequencing profiles

Methylation was profiled using Methyl C Capture Sequencing (MCC-seq), a targeted bisulfite sequencing

epigenome profiling approach, which has been previously described.<sup>28</sup> In this study a capture panel designed to target functional regions in whole blood was used. This panel has been used in three other studies to date.<sup>29–31</sup> Specifically, regions incorporated in the panel included 1) low and unmethylated regions (LMR/UMR) identified from merged whole genome bisulfite sequencing (WGBS) data from 30 whole blood samples derived from the TwinsUK cohort,<sup>32</sup> 2) Immune/circulating leukocyte-specific regulatory regions identified by ChIP-Seq or DHS mapping and 3) Illumina 450k array CpG sites. In all, the panel captures 4,861,805 CpGs *via* 607,984 targeted regions. All targeted regions were merged, and a panel was generated by Roche NimbleGen.

The protocol for using the panel has been described previously.<sup>28</sup> Briefly, following library preparation, bisulfite conversion and amplification of the MCC-seq library, targeted DNA fragments are captured, amplified again, and sequenced. Reads are then aligned to a bisulfite-converted hg19/GRCh37 reference genome, with the removal of poor-quality reads and mismatches. CpGs were also removed if not covered by at least five reads and at least two reads per strand. Further, CpGs were selected for less than 20% difference in methylation between strands. The methylation value was determined as total (forward and reverse) non-converted C over total (forward and reverse) reads. Sites were only included in the downstream analysis where missing values were limited to less than 20% of the individuals. Overall, this included 2,048,698 sites. Further details are provided in the [Supplementary Note](#).

For many of the sequenced sites there was little variability between individuals. The standard deviation for the DNA methylation variability between all sites had a median of 2% (lower quartile 1.4% and upper quartile 4.2%). To ensure the analysis focused on sites which were variable between individuals, and as such more likely to be related to disease pathogenesis, we restricted our analysis to the top quartile of most variable CpG sites. Therefore, the downstream analyses focused on the 512,175 most variable CpG sites.

#### 450k DNA methylation profiles

DNA extraction in the whole blood TwinsUK validation samples has been described previously.<sup>33</sup> DNA methylation profiling of these samples using the Illumina HumanMethylation 450kBeadChip array (450k array) has also been previously described.<sup>33</sup> Briefly DNA methylation levels used here are Illumina beta-values, defined as the ratio of the methylated probe intensity over the sum of methylated and unmethylated probe intensities plus 100.<sup>34</sup> Methylation beta-values range between 0 at unmethylated CpG-sites, and 1 at fully methylated CpG sites. Methylation 450k data were processed, and quality control assessment was carried out using Enmix.<sup>35</sup> Exclusions were determined using Minfi<sup>36</sup> with samples with median methylated and unmethylated signals below 10.5 excluded. In addition, cross-reactive probes and probes containing >2 alignment mismatches were excluded. Altogether, 438,594 450k probes were included in the downstream analysis.

#### MeDIP-seq methylation profiles

DNA methylation in whole blood was profiled using methylated DNA immunoprecipitation sequencing (MeDIP-seq) in the TwinsUK cohort, as previously described.<sup>37,38</sup> TwinsUK participants included the current MCC-seq study datasets were excluded from the TwinsUK MeDIP-seq dataset, resulting to a final independent TwinsUK MeDIP-seq dataset consisting of 116 T2D cases and 3318 controls. As previously described,<sup>38</sup> sonication was used to fragment DNA, following which libraries were prepared using Illumina's DNA Sample Prep kit for single-end sequencing. Immunoprecipitation was then carried out using anti-5mC antibody (Diagenode) and qPCR used for validation. Captured DNA underwent purification and amplification, following which 200-500bp fragments were selected. Sequencing was then carried out using the Illumina Platform and aligned using BWA.<sup>39</sup> Methylation levels were quantified using MEDIPS v1.0.<sup>40</sup> After processing, MeDIPseq data was quantified in bins of 500bps with a 250bp overlap.

#### Type 2 diabetes status and fasting blood glucose

Type 2 Diabetes (T2D) status was determined for all participants in the study, using a combination of approaches. We determined self-reported T2D status from questionnaire data, as well as based on circulating fasting glucose measurement of at least 7 mmol/L, or use of blood glucose lowering medication. T2D status information was obtained at the same clinical visit during which the blood sample for DNA methylation profiling was collected, with the exception of the OATS samples. In the OATS dataset if T2D status is based on self-reported status, then this was information was obtained within 3 months of the blood sample collection date. Full details of T2D status determination for each cohort are provided in the [Supplementary Note](#).

#### Peripheral blood cell proportions

Blood cell type proportions were estimated for monocytes, granulocytes, Natural Killer (NK) cells, CD8, CD4 and plasmablasts using the approach proposed by Houseman et al.<sup>41</sup> The R package "Flow-Sorted.Blood.450k"<sup>42</sup> was used to estimate blood cell type proportions based on the subset of 450k signals included in the MCC-seq panel. As expected, the estimated cell type proportions showed high levels of correlation and therefore not all estimated cell types were included in the model. The final model of analysis included covariates for granulocytes, NK cells and CD4 T cells, due to these cell types showing the lowest correlations between them. All other cell types were correlated ( $r > 0.7$ ) with one of these variables.

#### Epigenome-wide association analysis

An epigenome-wide association study (EWAS) was carried out to determine associations between whole blood DNA methylation variation and T2D case control status. DNA methylation values for each CpG-site were normalised to  $N(0,1)$  prior to fitting linear models. For the case-control analysis, a mixed effect linear model was fitted (using lme4 and lmerTest in R). Methylation at the CpG site was the response variable with T2D status as the predictor. Covariates included age, smoking status, sex, blood cell composition as fixed effects and family and zygosity as separate random effects. Methylation effect sizes were calculated using the same linear models, but without normalizing DNA methylation levels to  $N(0,1)$  prior to data analysis. A sensitivity analysis was carried out including BMI as an additional covariate. Multiple testing adjustment was carried out using Benjamini and Hochberg False Discovery Rate (FDR) thresholds of FDR 5%, and a more relaxed threshold of FDR 25%.

This mixed effects linear model was also applied in the replication analysis in testing the association of MCC-seq variation at each candidate CpG site with T2D status, and similarly in the validation analyses assessing the association of 450k methylation profiles at candidate sites and metabolic phenotypes.

We explored the genes that the T2D-DMPs arising from the case-control analysis mapped to using gene set enrichment analysis (GSEA),<sup>43</sup> as previously described.<sup>44</sup> We also compared the 16 T2D-DMP genes to nine major collections of gene sets within the Human Molecular Signatures Database (MSigDB (<http://www.gsea-msigdb.org/gsea/msigdb/index.jsp>)). We report enrichment results at a significance value of FDR 5%.

#### Genetic independent (discordant MZ twin) analysis

A subset of the data was used to carry out a T2D-discordant MZ twin pair analysis, focusing on MZ twin pairs alone. Ascertainment into the study prioritized inclusion of 74 pairs of MZ twins who were discordant for T2D in the main MCC-seq dataset.

T2D-discordant MZ twin pair analysis was performed for differential methylation at the T2D-DMPs identified by the case control analysis at a FDR 25% threshold. Statistical significance was determined using a one-sample parametric t-test for paired samples. This tested whether the mean methylation difference within each twin pair of the covariate-adjusted methylation data was significantly different from zero. Covariate-adjusted methylation data were generated using a linear mixed effects model including methylation as the response variable, and covariates from the case control analysis (age, sex, smoking status, blood cell composition) as fixed effect variables and family and zygosity as random effects.

### Validation analysis

The MCC-seq validation analyses using 450k data targeted the FDR 25% results from the MCC-seq T2D case-control analysis. We used 450k DNA methylation levels at the target CpG site, or at a 450k CpG-site nearest to the target, and assessed their association with T2D and fasting blood glucose. In comparing MCC-seq to 450k data we used overlapping CpG sites where possible. Only 5 of the 20 T2D-DMPs were profiled on the 450k array, with the majority being unique to the MCC-seq platform. In instances where there was no matching site available, we selected the nearest 450k CpG site for validation testing, where the furthest distance between nearest 450k CpG to target CpG was 2.7 kb. To assess significance accounting for multiple testing, a Bonferroni threshold of  $P = 0.0025$  ( $P = 0.05/20$ ) was applied. Validation analyses were also carried out using the independent MeDIP-seq dataset from TwinsUK. For MeDIP-seq validation we analysed T2D-DMPs at FDR 5% and the signal annotated to TXNIP. We averaged DNA methylation levels for the two bins that spanned each of the tested T2D-DMP base-pair location. The MeDIP-seq validation results are presented at nominal significance.

### Replication analysis

We pursued replication of the previously unreported T2D associated MCC-seq signals in an independent sample of 573 individuals from the TwinsUK cohort, with 33 T2D cases and 540 controls. The 573 participants had whole blood DNA methylation profiled using the same methylation profiling platform (MCC-seq). Quality control and analysis followed the same procedures as outlined for the discovery sample. To assess evidence for replication a Bonferroni threshold was applied to the number of tests ( $P = 0.05/20 \approx 0.0025$ ), and only results showing the same direction of the association as in the discovery sample were considered to replicate.

### Genetic data and meQTL analyses

To assess evidence for genetic impacts on DNA methylation levels we tested for genetic variants that

were associated with DNA methylation at the T2D-DMPs (meQTLs). Genotype data were explored in 277 TwinsUK twin pairs from both the discovery and replication datasets with MCC-seq data, and were used for the identification of meQTLs. Genotyping of the full TwinsUK genetic dataset has been described previously.<sup>45</sup> Briefly, genotyping was carried out using HumanHap300, HumanHap610Q, HumanHap1M Duo, and HumanHap1.2 M Duo 1 M arrays. Haplotypes were derived from pre-phasing using IMPUTE2 without a reference panel. Fast imputation was performed using these haplotypes and the 1000 Genomes phase 1 dataset. SNPs were excluded as part of quality control where they failed Hardy Weinberg equilibrium ( $P < 10^{-6}$ ), had a MAF  $< 0.01$ , had missingness of more than 5% or an info score  $< 0.8$ . In addition, individuals with discordant sex were removed. PLINK 2.0 was used to remove outliers in unrelated participants and GENESIS used for related participants. A deviation of more than 7 SD from the mean was considered an outlier. Pruning was undertaken to reduce relatedness, with participants with IBS  $> 0.125$  (calculated using PLINK 2.0) removed.

For identification of SNPs that were meQTLs a linear model was fitted using the MatrixEQTL R package,<sup>46</sup> where the methylation was the response variable, and dosage of minor allele was the predictor. Covariates included age, sex, BMI, smoking status and blood cell composition as fixed effects. Both *cis* me-QTL and *trans* meQTL SNPs were included, where the *cis* interval was defined as  $\pm 1$  Mb from the CpG site. A stringent *cis* me-QTL P-value threshold was used to test for significance ( $P = 1 \times 10^{-5}$ ), as previously described<sup>47</sup> and in *trans* a stricter threshold was applied ( $P = 5 \times 10^{-8}$ ). The most significantly associated SNP per CpG site was reported as the meQTL for the T2D-DMP. In addition, the GoDMC meQTL database<sup>48</sup> based on the 450k array was also used to investigate meQTLs at the nearest CpG in the database.

### Enrichment in GWAS regions

We carried out an analysis to determine whether the epigenetic associations were enriched in regions previously found to be associated with T2D in genome wide association studies (GWAS). Firstly, we determined GWAS regions, which were defined as a 50 kb window around the lead SNPs identified as being associated with T2D in two recent large European ancestry GWAS studies.<sup>49,50</sup> For the purpose of determining whether there was an enrichment of T2D-DMPs in GWAS regions, we utilised T2D-DMPs identified at a more relaxed threshold of FDR 50% in the case-control analysis (which resulted in 5057 T2D-DMP signals, around 1% of the total sites). We then assessed whether there was a difference in the number of CpG sites in and out of the identified GWAS regions. A Fisher's Exact Test was used to determine significance at a nominal threshold ( $P < 0.05$ ). Secondly, we assessed

whether the 20 genes to which the FDR 25% T2D-DMPs were annotated had been previously found in GWAS studies to have variants associated with T2D, its complications, metabolic health or inflammatory diseases. The GWAS catalog<sup>51</sup> was used for this purpose.

### Metabolomic profiling

To assess the functional relevance of T2D-associated MCC-seq signals, we explored their blood metabolomic signature. Blood metabolomic profiles were analysed for 181 individuals in TwinsUK in both the discovery and replication samples, where metabolomic data were obtained in samples collected within 5 years of the MCC-seq sample data. Fasting serum metabolite levels for 592 metabolites were detected and quantified by Metabolon, Inc (Durham, USA), as described in detail previously.<sup>52</sup> Briefly, the metabolomics analysis was carried out on a platform of four ultra-high performance liquid chromatography-tandem mass spectrometry (UPLC-MS/MS) instruments. Sample preparation used the MicroLab STAR system and samples underwent a multistep preparation process including removal of protein, division into five fractions for analysis using different methods and one backup, removal of organic solvent and storage overnight under nitrogen. Raw data underwent quality control using Metabolon's hardware and software. Metabolites included amino-acids, peptides, carbohydrates, energy intermediates, lipids, nucleotides, cofactors and vitamins, and xenobiotics. Linear models assessed the association between methylation levels at each T2D-DMP and blood metabolite levels. The linear model was similar to the model used in the main analysis, that is, methylation was the response variable and metabolite was the predictor. The model included covariates for smoking, age, BMI, family structure and cell counts, and an additional covariate reflecting the time difference between the time of blood draw for methylation profiling and for metabolomic profiling. Two Bonferroni multiple testing thresholds were assessed, the main focus being on FDR 5% significant CpGs (3 signals at  $P = 0.05/(592*3) \sim 2.8 \times 10^{-5}$ ), and secondly, we also considered 20 FDR 25% significant CpGs ( $P = 0.05/(592*20) \sim 4.2 \times 10^{-6}$ ).

### T2D complications

To assess the potential impact of T2D-DMPs on T2D progression and development of complications, we explored additional phenotype data in the TwinsUK discovery and replication samples. Individuals with T2D who had developed diabetic retinopathy were identified through self-reported questionnaire responses. The association between diabetic retinopathy and DNA methylation at chr14:77736811 (*NGB*) and chr14:57265055 (*OTX2*) was investigated in 9 T2D retinopathy cases (which included 4 individuals who were related as part of 2 twin pairs) and 66 T2D controls

using a two-tailed t-test at a nominal significance threshold ( $P = 0.05$ ).

Blood pressure measured at clinic visits was available for 77 individuals with T2D from the discovery and replication samples. For the purpose of this analysis, high blood pressure was defined as systolic pressure greater than 130 or diastolic pressure greater than 80. The association between DNA methylation at chr19:11529587 (*RGL3*) and high blood pressure in 77 T2D individuals was assessed using a two tailed t-test at a nominal significance threshold ( $P = 0.05$ ).

We further explored both retinopathy and blood pressure in the MeDIP methylation data at the same sites using the same approach. In this non overlapping dataset there were 34 T2D retinopathy cases and 82 T2D controls and 110 individuals with T2D with blood pressure data.

### Role of funders

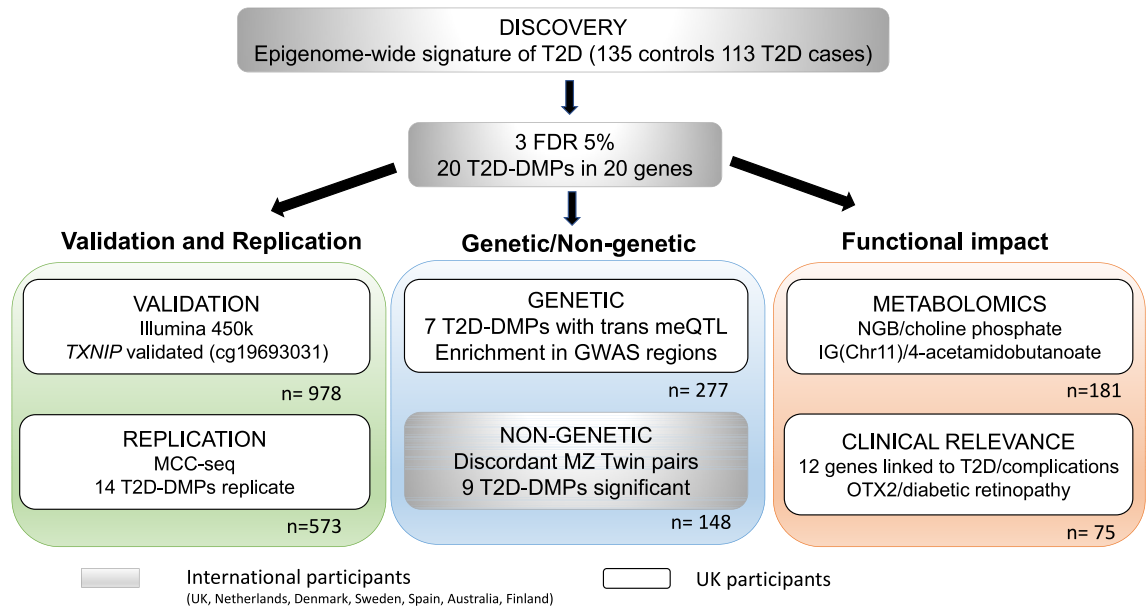
Funders had no role in the study design, data collection, data analyses, interpretation, or writing of the manuscript.

### Results

Associations between T2D and DNA methylation levels profiled using the MCC-seq platform were explored in 248 whole blood samples from 7 international twin cohorts in the discovery stage (135 controls and 113 T2D cases). Participants were predominantly of European ancestry with an average age of  $60 \pm 13$  and mostly female (Table 1). The primary analysis sought to identify previously unreported sites of differential methylation associated with T2D (T2D-DMPs) in a T2D case-control analysis. Integrating genotype data along with analysis of 74 T2D-discordant MZ twin pairs, provided the opportunity to explore the genetic basis of T2D-DMPs, that is, distinguishing genetically-based from genetically-independent T2D associated differential methylation signals. The findings were subject to replication, validation, functional follow-up and enrichment analyses. Fig. 1 shows an overview of the study design.

### T2D case control epigenome wide signals

A T2D case-control analysis, between 113 individuals with T2D and 135 individuals without T2D, identified 3 previously unreported genome-wide significant sites of differential DNA methylation annotated to the body of *RGL3* (chr19:11529587), the body of *NGB* (chr14:77736811) and downstream from *OTX2* (chr14:57265055) at FDR 5%. *RGL3* harbours genetic variants associated with blood pressure,<sup>53</sup> *NGB* is a neuroprotective protein expressed strongly in the retina<sup>54</sup> and *OTX2* is involved in the development of the eyes,<sup>55</sup> and therefore the function of these genes suggests potential links to T2D complications. At FDR 25% altogether 20 differential methylation positions



**Fig. 1: Study design and results summary.** Epigenome-wide analysis of T2D in 113 T2D cases and 135 controls. Downstream analyses explored genetic-based and genetic-independent associations, functional follow-ups, and clinical relevance to T2D complications.

(T2D-DMPs) were identified (Table 2, Fig. 2a) in 20 genes (Table 2, Fig. 3). These 20 T2D-DMPs are previously unreported, with one around 1 kB from a previously identified T2D-associated CpG site in *TXNIP* (cg19693031). In addition to *TXNIP*, differential methylation has previously been found in regions

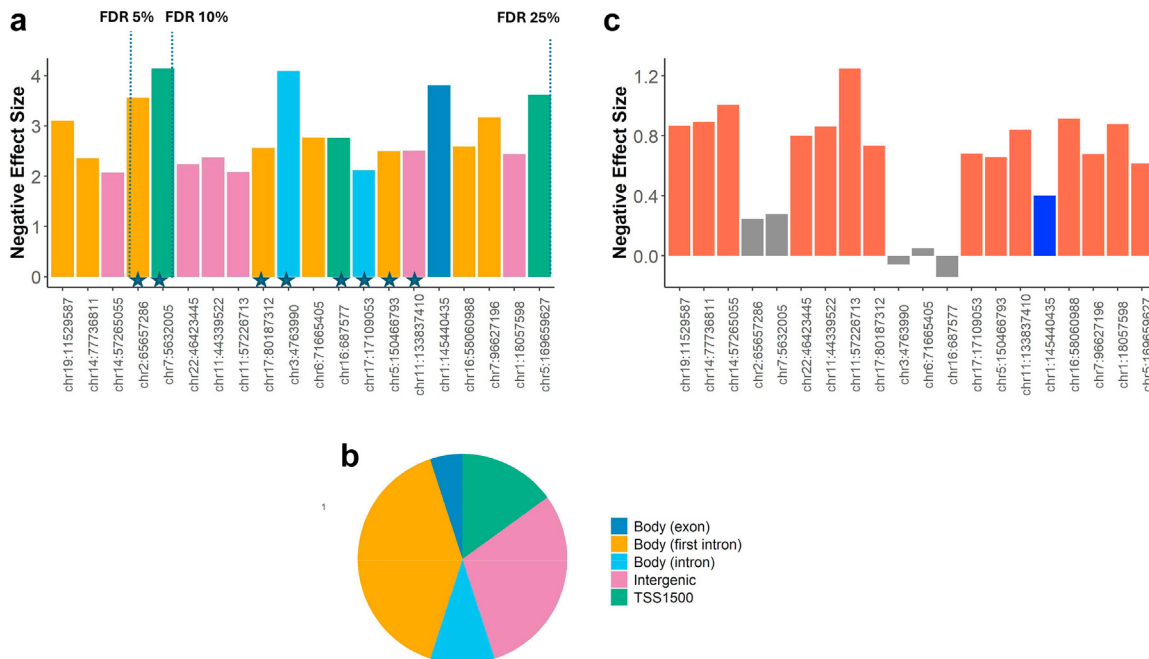
annotated to *SPRED2* and *ITPR1*.<sup>9</sup> All T2D-DMPs are hypomethylated in T2D with similar effect sizes (Table 2). Altogether 13 of the 20 T2D-DMPs are in gene promoters, with three in transcription start sites (chr7:5632005 *FSCN1*; chr16:687577 *METTL26*; chr5:169659627 *C5orf58*), and 8 of the 20 T2D-DMPs are

Site	Gene	Position	Effect <sup>b</sup>	P-value	FDR	Replication effect	Replication P-value
chr19:11529587-11529588	<i>RGL3</i>	Body (first intron)	-3.10	1.83E-07	0.05	-0.867	6.59E-06
chr14:77736811-77736812	<i>NGB</i>	Body (first intron)	-2.36	2.55E-07	0.05	-0.893	1.16E-05
chr14:57265055-57265056	<i>OTX2</i>	Downstream	-2.07	2.94E-07	0.05	-1.007	8.36E-08
chr2:65657286-65657287	<i>SPRED2</i>	Body (first intron)	-3.56	7.03E-07	0.07	-0.244	2.30E-01
chr7:5632005-5632006	<i>FSCN1</i>	TSS1500	-4.14	7.32E-07	0.07	-0.278	1.44E-01
chr22:46423445-46423446			-2.24	1.80E-06	0.13	-0.800	3.19E-05
chr11:44339522-44339523			-2.37	1.91E-06	0.13	-0.861	9.83E-06
chr11:57226713-57226714	<i>RTN4RL2</i>	Upstream	-2.08	1.96E-06	0.13	-1.248	1.08E-11
chr17:80187312-80187313	<i>SLC16A3</i>	Body (first intron)	-2.56	2.67E-06	0.15	-0.732	2.62E-04
chr3:4763990-4763991 <sup>a</sup>	<i>ITPR1</i>	Body (middle intron)	-4.09	3.67E-06	0.19	0.060	7.63E-01
chr6:71665405-71665406 <sup>a</sup>	<i>B3GAT2</i>	Body (first intron)	-2.77	4.34E-06	0.19	-0.051	7.84E-01
chr16:687577-687578	<i>Ak201549/LOC10028175/METTL26</i>	TSS1500 (METTL26)	-2.76	4.45E-06	0.19	0.140	4.44E-01
chr17:17109053-17109054	<i>PLD6</i>	Body (intron)	-2.12	4.82E-06	0.19	-0.680	3.32E-04
chr5:150466793-150466794	<i>TNIP1</i>	Body (first intron)	-2.50	5.76E-06	0.20	-0.658	3.79E-04
chr11:133837410-133837411			-2.51	6.02E-06	0.20	-0.840	2.74E-05
chr1:145440435-145440436	<i>NBFF10/LOC100288142/TXNIP</i>	<i>TXNIP</i> (body exon) LOC/NBFF10 (intron)	-3.81	6.24E-06	0.20	-0.400	3.81E-02
chr16:58060988-58060989 <sup>a</sup>	<i>MMP15</i>	Body (first intron)	-2.59	7.34E-06	0.22	-0.915	2.41E-06
chr7:96627196-96627197 <sup>a</sup>	<i>DLX6-AS1</i>	Body (first intron)	-3.17	9.16E-06	0.24	-0.679	8.06E-04
chr1:18057598-18057599			-2.44	9.19E-06	0.24	-0.879	3.71E-06
chr5:169659627-169659628 <sup>a</sup>	<i>C5orf58</i>	TSS1500	-3.62	9.44E-06	0.24	-0.615	2.30E-03

<sup>a</sup>This CpG site is in the 450 k array. <sup>b</sup>Effect size is shown without normalisation.

**Table 2: Differentially Methylated CpG sites associated with T2D at FDR 25%.**





**Fig. 2: Differential methylation signature of T2D in whole blood.** **a.** Differentially methylated sites ranked by FDR, showing effect size (y axis, expressed as negative effect size as all effects are hypomethylated in T2D) and genomic annotation (colour coded). T2D-DMPs in enhancer regions are indicated with a star. Vertical dotted lines show FDR 5%, 10% and 25% thresholds. **b.** Proportion of T2D-DMPs in each genomic annotation category. **c.** Effect sizes in the replication sample, where results replicating at multiple testing threshold are shown in red, and results nominally significant are shown in blue.

located in enhancer regions (Fig. 2); suggesting a role of T2D-DMPs in gene regulation. Furthermore, twelve of the 20 T2D-DMPs are located in gene bodies, which is consistent with previous studies.<sup>19</sup> All of the T2D-DMPs located in gene bodies are in intronic regions, with the exception of the signal in *TXNIP* (chr1:145440435). The T2D-DMPs are mainly located either in or near CpG Islands (Supplementary Table S9).

Gene set enrichment analysis (GSEA) of the genes that the FDR 25% T2D-DMPs annotate to was carried out in the Human Molecular Signatures Database (MSigDB), and identified a significant (FDR <5%) enrichment for two sets of genes. The first set included genes upregulated in CD4+ T cells overexpressing *FOXP3* and *PPARG1* (4 genes). The second set included genes, which contained binding sites for the transcription factor CHAF1B within their promoter sequence (6 genes, including *FSCN1*, *TXNIP*, *ITPR1*, *SLC16A3*, *SPRED2* and *METTL26*).

### Sensitivity analyses including BMI

BMI is a major risk factor for T2D, therefore, we assessed its impact by including it as a covariate in sensitivity analyses. In analyses incorporating BMI the resulting T2D-DMP P-values and effect sizes were similar to the primary analysis excluding BMI, however, the strength of association was attenuated

(Supplementary Table S3). Nine of the 20 T2D-DMPs still met the FDR 25% threshold including the three original FDR 5% results (now at FDR 9% when BMI is included in the model). Two additional sites met the FDR 25% threshold including chr5:133449700 (*TCF7*) and chr15:58085147 (intergenic). *TCF7* plays a role in beta cell function,<sup>56,57</sup> activates immune system genes, and genetic variants in this gene are associated with Type 1 Diabetes.<sup>58</sup>

### Validation of T2D methylation signature

We sought to validate the peak MCC-seq T2D-DMPs using whole blood DNA methylation profiled on the 450k array platform, as well as a subset of the peak T2D-DMPs using the MeDIPseq platform. In the 450k validation, altogether 978 participants from TwinsUK were selected for validation, and these included 44 individuals (19 cases and 25 controls) from the primary analysis. The validation analysis tested the association between DNA methylation with both T2D status (49 cases and 681 controls) and with fasting blood glucose (FBG; 978 individuals). There was an exact overlap between the MCC-seq T2D-DMPs CpG site and the 450 k array CpG site for 5 T2D-DMPs, while the remainder of the MCC-seq T2D-DMP sites were not profiled on the 450 k array. In this instance we used the nearest 450 k CpG site up to a distance of 2.7 kB. Site cg19693031 (in *TXNIP*)

validated with significant association at Bonferroni multiple testing threshold with both T2D ( $P = 3.4 \times 10^{-8}$ ) and FBG ( $P = 2.7 \times 10^{-7}$ ), with a consistent direction of association. Two further sites showed nominal significance in the FBG analysis, including cg15989436 (*TNIP1*;  $P = 0.0074$ ) and cg01088070 (*B3GAT2*;  $P = 0.0268$ ) (Supplementary Table S2). In addition, whilst not meeting nominal significance, 10 further T2D-DMPs showed the same direction of effect in the T2D association analysis (in cg02985292 (*METTL26*), cg01088070 (*B3GAT2*), cg10256249 (*SPRED2*), cg27175093 (*OTX2*), cg15989436 (*TNIP1*), cg05468064 (chr22:46423445), cg21545390 (*DLX6*), cg15131789 (chr1:18057598), cg02400572 (*NGB*), cg08995368 (*FSCN1*)) and 2 further T2D-DMPs matched direction of effect in only the FBG analysis (in cg21545390 (*DLX6*), cg15131789 (chr1:18057598)).

Validation analyses were also carried out using MeDIP-seq data. In the MeDIP-seq validation 3434 participants from TwinsUK (116 cases and 3318 controls) were selected for validation, and these did not include any participants in the primary analysis. Validation was tested for *RGL3* (chr19:11529587), *NGB* (chr14:77736811), *OTX2* (chr14:57265055), and *TXNIP* (chr1:145440435-145440436). Direction of effect matched across MCC-seq and MeDIP-seq data for all signals tested, and *NGB* (chr14:77736811) showed nominally significant results (Supplementary Table S11).

#### Replication of T2D methylation signature

The 20 FDR 25% T2D-DMPs were evaluated for replication in an independent blood MCC-seq methylation sample from TwinsUK. Of the 20 FDR 25% T2D-DMPs reported in this study 19 are previously unreported and only the 450 k CpG site in *TXNIP* (cg19693031; 1.1 kb away from MCC-seq T2D-DMP chr1:145440435-145440436) has been robustly replicated in previous studies. Here, we pursued replication of all 20 T2D-DMPs in an independent sample of 573 TwinsUK participants profiled using MCC-seq, including 33 individuals with T2D at the time of blood collection. Fourteen signals replicated at a Bonferroni corrected threshold ( $P = 0.0025$ ) and 1 further signal replicated at nominal significance ( $P < 0.05$ ; Fig. 2c) and with a consistent direction of association. The 14 signals included all three FDR 5% results (*RGL3*, *NGB* and *OTX2*) and signals annotated to *RTN4L2*, *SLC16A3*, *PLD6*, *TNIP1*, *MMP15*, *DLX6-AS1*, *C5orf58* and 4 further intergenic regions (Chr22:46423445, Chr11:44339522, Chr11:133837410, Chr1:18057598) (Supplementary Table S4).

DNA methylation is a potential mechanism for mediating genetic and environmental impacts on T2D, or the differential methylation signals observed could be secondary to disease. The next series of analyses consider this in more detail, exploring putative non-genetic and genetic sources of variation at the observed T2D methylation signature.

#### Genetic independent (MZ twin pair) analysis

Using a discordant MZ twin pair model including 74 MZ twin pairs discordant for T2D, we explored whether the T2D-DMPs showed genetically independent effects by contrasting DNA methylation profiles between genetically identical T2D-discordant twins. Genetically independent sites of differential methylation likely reflect environmental effects on T2D and/or may arise as a consequence of T2D. These signals can also be informative for T2D disease progression and development of T2D complications. A t-test of the methylation residuals, allowing for covariates, showed that all of the 20 T2D-DMPs except one (intergenic site on chr11:133837410) were nominally significant and with the same direction of effect (Supplementary Table S1) in the genetically independent analyses. Of the 20 FDR 25% tested T2D-DMPs, 9 sites (in *OTX2*, *SPRED2*, *SLC16A3*, *B3GAT2*, *METTL26*, *TNIP1*, *TXNIP*, *MMP15* and an intergenic regions) had significant associations in the T2D-discordant MZ twin analysis after multiple testing correction (Supplementary Table S1). These target genes with a range of functions including regulating growth (*SPRED2*), catalysing lactic acid and pyruvate transport (*SLC16A3*, a member of the MCT family), catalysing glycosaminoglycan metabolism (*B3GAT2*) and autoimmunity (*TNIP1*). *METTL26* function is unknown, however, it has been observed to be expressed at a high level in cancerous cells and associated with poor prognosis.<sup>59</sup> In addition to the 20 T2D-DMPs the two sites identified in the BMI sensitivity analysis (chr5:133449700 (*TCF7*) and chr15:58085147 (intergenic)) were also nominally significant ( $P < 0.05$ ) for genetically independent effects.

#### Genetic variation underlying T2D methylation signatures

We assessed if the differential methylation signature of T2D may be influenced by genetic variants, or DNA methylation quantitative trait loci (meQTLs). Using two approaches we observed evidence for a genetic basis at 8 of the 20 FDR 25% sites when considering only the target MCC-seq T2D-DMP CpG site, and at 15 of the 20 FDR 25% sites when considering the target CpG or the nearest 450 k CpG site. First, we explored the association between genetic variation and DNA methylation levels at the 20 T2D-DMPs in 277 twin families from TwinsUK. We identified 7 T2D-DMPs with significant *trans* meQTL associations ( $P = 5 \times 10^{-8}$ , Supplementary Table S5) in *NGB*, *OTX2*, *SLC16A3*, *ITPR1*, *TNIP1*, and two intergenic regions (chr1:18057598 and chr11:44339522). The most significant association was between a genetic variant in the *HPN* gene (chr19:35561358) and DNA methylation in *ITPR1* (chr3:4763990-4763991). *ITPR1* has been shown to have a role in diet induced diabetes<sup>60</sup> and *HPN* encodes hepsin and it has been shown in mouse models that hepsin deficiency lowers blood glucose.<sup>61</sup>

Second, using the GoDMC meQTL database<sup>48</sup> we investigated meQTLs for the target CpG T2D-DMP site, if this was profiled on the 450k array, or for the nearest 450k CpG site to the T2D-DMP using the same thresholds. Five of the 7 T2D-DMP CpGs identified above also had evidence for a genetic basis in GoDMC, and a further 8 sites were not profiled on the 450k array, but the nearest 450k CpG had meQTLs in GoDMC. Altogether these 13 450k CpGs had GoDMC meQTLs in *cis* only (11), *trans* only (1), or both *cis* and *trans* (1) (Supplementary Table S6). There were 147 *trans* meQTLs for cg19693031 in *TXNIP*, of which 133 are in *SLC2A1* (*GLUT1*) which is a major glucose transporter located over 100 MB away from *TXNIP* (Fig. 4). *TXNIP* expression causes a reduction in *GLUT1*,<sup>62</sup> which in turn transports more blood sugar when blood glucose is high. Suppressing *GLUT1* has been shown to reverse high retinal blood glucose levels in diabetic mice.<sup>63</sup> The remaining SNPs are in a 125 kB window of *SLC2A1*.

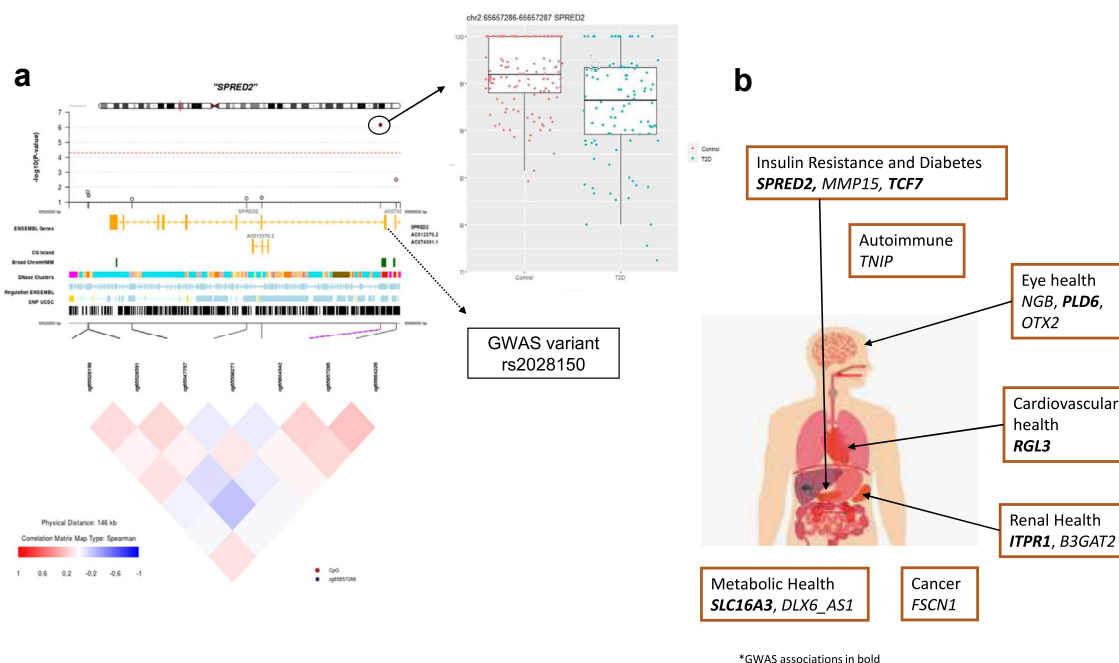
Of the 8 T2D-DMPs that had direct evidence for a genetic basis of the target MCC-seq CpG site, five (in *OTX2*, *SLC16A3*, *B3GAT2*, *TNIP1* and intergenic region chr11:44339522) were also identified in the genetically-independent results from the T2D-discordant MZ twin analysis. Thus, both genetic and non-genetic effects

contribute to DNA methylation variation at these 8 T2D-DMPs. It is therefore likely that DNA methylation levels at these five signals reflect consequences of T2D, which may potentially be risk or effects of developing T2D complications.

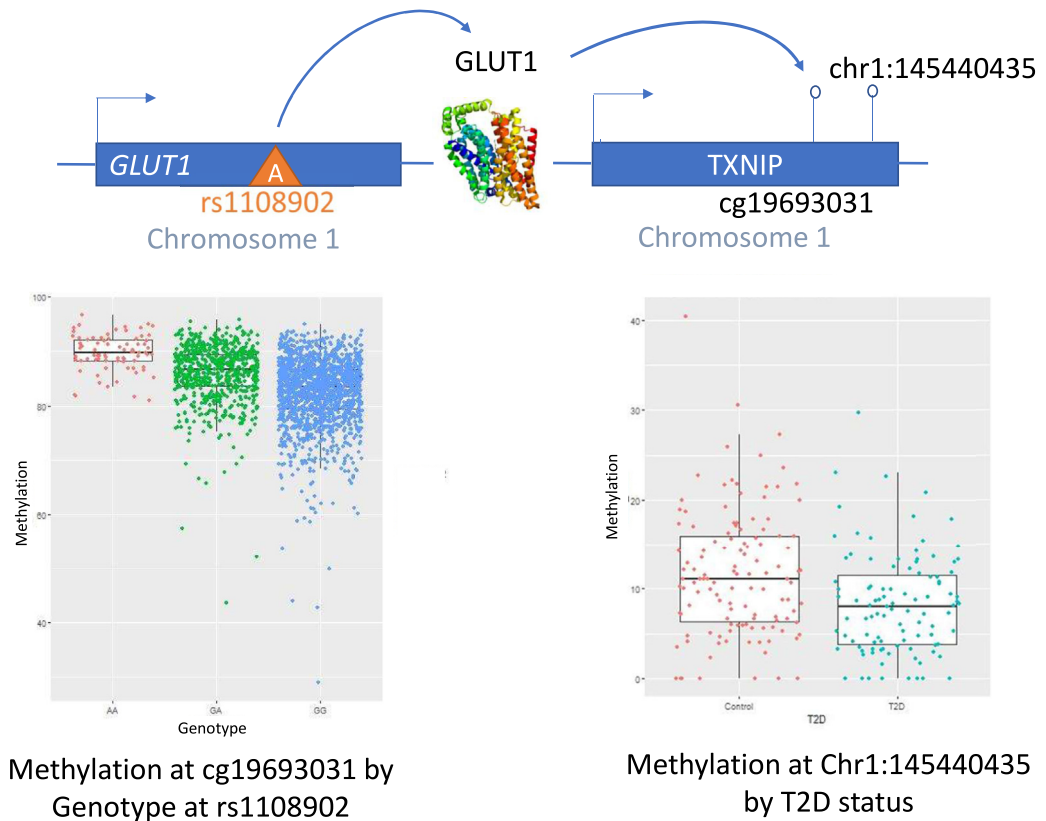
### Enrichment of T2D methylation signature in GWAS regions

Previous studies have found that there is an overlap between T2D GWAS results and T2D-DMPs.<sup>19</sup> Here, we also find an enrichment of T2D-DMPs in T2D GWAS regions based on two different approaches. First, we tested whether there was an enrichment of FDR 25% T2D-DMPs in T2D GWAS regions (defined as a window of 50 kb either side of a lead SNP in recent large GWAS). One of the 20 T2D-DMPs in *SPRED2* is in a T2D GWAS region (Fig. 3a), as well as one of the T2D-DMP results with BMI as a covariate (in *TCF7*). The differential methylation in *SPRED2* (chr2:65657286) and the GWAS variant (rs2028150) are located in the same first intron region of the gene.

Due to the small number of T2D-DMPs at FDR 25% we also explored GWAS enrichment for T2D-DMPs at a more relaxed FDR 50% threshold, resulting in 4452 T2D-DMPs. We also found a significant enrichment of



**Fig. 3: Genes annotated to T2D-DMPs and their relevance to human health.** a. Epigenetic association between T2D and *SPRED2* DNA methylation displayed in a coMET plot,<sup>34</sup> including T2D-methylation association results (top panel) along with functional annotation of the region (middle panel), and pattern of co-methylation at the CpG sites in the 450 k array annotated to *SPRED2* (bottom panel). Broad ChromHMM regions are displayed using UCSC genome browser colour schemes. GWAS variant rs2028150 is labelled and is located 2.27 kb from the MCC-seq T2D-DMP methylation site chr2:65657286. b. Biological relevance of genes annotated to FDR25% T2D-DMPs to human phenotypes, with published GWAS associations shown in bold.



**Fig. 4: DNA methylation in *TXNIP* is under distal genetic control.** SNP rs1108902 in *GLUT1* is a trans-meQTL for DNA methylation levels at CpG site cg19693031 in *TXNIP*, located around 1 kb away from the MCC-seq T2D-DMP on chromosome 1 (chr1:145440435). Shown are boxplots of the variation in DNA methylation levels by T2D status at chr1:145440435 in *TXNIP* for the discovery cohort, and change in *TXNIP* cg19693031 DNA methylation levels by genotype at rs1108902 in *GLUT1* in TwinsUK participants.

these FDR 50% T2D-DMPs in previously published T2D GWAS regions (Fisher Exact Test  $P = 3.5 \times 10^{-8}$ , [Supplementary Table S8](#)).

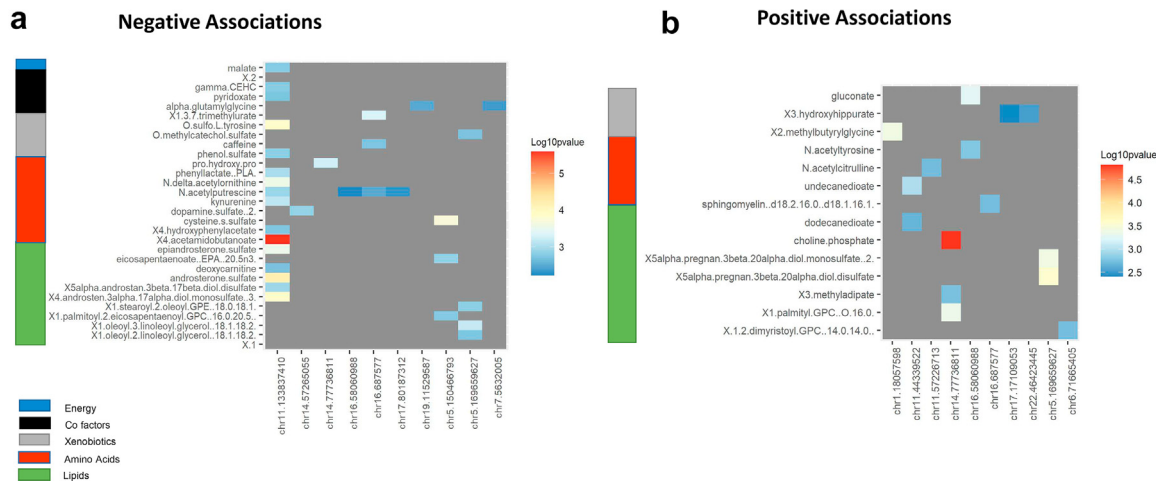
Second, we assessed if the genes to which the FDR 25% T2D-DMPs were annotated contained genetic variants that had previously been associated with either T2D or related phenotypes in GWAS studies. We found GWAS links between 7 T2D-DMP genes and T2D or related phenotypes ([Fig. 3b](#)), including *SPRED2* (T2D), *ITPR1* and *PLD6* (Diabetic complications), *SLC16A3* (BMI), *TCF7* (Type 1 Diabetes), *RGL3* (blood pressure) and *TNIP1* (autoimmune traits).

#### Metabolomic follow-up of T2D methylation signature

To explore the functional relevance of the T2D-associated differential methylation variation, we tested the association between DNA methylation levels at the 20 T2D-DMPs and blood metabolites. Circulating blood metabolites included amino acids, peptides, carbohydrates, energy intermediates, lipids, nucleotides, cofactors and vitamins, and xenobiotics. Of the 592

metabolites tested, two were significantly associated with DNA methylation levels at T2D-DMPs at a Bonferroni multiple testing threshold ([Fig. 5, Supplementary Table S7](#)). The two signals included choline phosphate levels, which were associated with methylation in *NGB* (chr14:77736811,  $P = 1.5 \times 10^{-5}$ ). Choline phosphate is a reactant in the formation of citicoline. There was a positive association between methylation levels and choline phosphate, with a negative association found between DNA methylation and T2D. There is evidence that citicoline use can alleviate symptoms of diabetic complications, including in diabetic retinopathy and diabetic neuropathy.<sup>64,65</sup>

The second signal was in 4-acetamidobutanoate, where metabolite levels were associated with DNA methylation in an intergenic region (chr11:133837410,  $P = 2.6 \times 10^{-6}$ ). Metabolite 4-acetamidobutanoate is a precursor to ornithine, a component of the urea cycle.<sup>66</sup> Previous work has shown 4-acetamidobutanoate to be the most enriched urinary metabolite when individuals with low liver and kidney disease severity were compared with those with high disease severity.<sup>67</sup>



**Fig. 5: Association between T2D methylation signature and metabolomics.** Heatmap showing significant T2D-DMP metabolite associations (FDR <5%) and their effect size. Only significant (a) negative and (b) positive associations are shown, with grey areas reflecting correlations that did not meet the multiple testing threshold, and T2D-DMPs not shown where there were no significant associations (FDR <5%).

### T2D methylation signature and diabetic complications

A systematic analysis of the association between DNA methylation and T2D complications has not been thoroughly explored to date. In this study, the 3 T2D-DMPs at FDR 5% annotated to genes which have biological relevance to T2D complications, and metabolomic follow-up strengthens the link to T2D complications. Therefore, we further explored links between T2D-DMP DNA methylation levels and targeted T2D complications, specifically, diabetic retinopathy and high blood pressure.

Both *NGB* and *OTX2* have biological links to eye health, therefore we focused on investigating the association between DNA methylation at these two genes and diabetic retinopathy status in T2D cases. We explored DNA methylation levels at chr14:77736811 (*NGB*) and chr14:57265055 (*OTX2*) in 9 T2D retinopathy cases and 66 T2D retinopathy controls, selected from the TwinsUK discovery and replication T2D cases. A nominally significant difference in DNA methylation was observed at chr14:57265055 in *OTX2* ( $P = 0.025$ ), but not at chr14:77736811 in *NGB* ( $P = 0.62$ ). We did not observe a nominally significant difference in the MeDIPseq data for chr14:57265055 in *OTX2*, but both sites showed consistent direction of effects. That is, we observed lower DNA methylation levels in T2D cases with diabetic retinopathy compared to T2D cases without retinopathy, and this was consistent with the results from the MCCseq analysis.

Genetic variation in *RGL3* has previously been associated with blood pressure variation (Wang and Wang, 2019). We therefore investigated the association between DNA methylation at chr19:11529587 (*RGL3*) and blood pressure levels in 77 T2D cases from the

TwinsUK discovery and replication sample, but did not observe a nominally significant difference. We also did not observe a significant difference in the MeDIPseq data.

### Discussion

In this study we sought to apply a DNA methylation profiling approach targeting functional regions of the genome (MCC-seq) in whole blood to identify previously unreported CpG sites that were significantly differentially methylated with T2D and assess their functional significance and biological relevance for T2D complications. Participants in the study were selected from 7 international twin cohorts and included the largest number of T2D-discordant MZ twins pairs studied to date. We identified and replicated 3 FDR 5% signals for T2D in 248 discovery phase participants (113 T2D cases, 135 controls) and 573 replication phase participants (33 T2D cases, 540 controls; 978 individuals with FBG). Altogether 20 CpG sites in 20 genes were differentially methylated at a more relaxed FDR 25%, of which 14 replicated. These 20 sites were then analysed in 74 T2D-discordant MZ twin pairs. We identified genetic-independent effects at 9 T2D-DMPs, which were significantly differentially methylated within T2D-discordant MZ twin pairs. At these sites DNA methylation are likely either consequences of T2D, or they are mediating the effects of non-genetic T2D risk factors. We further assessed differential methylation for links to genetic variation, testing for enrichment in GWAS regions, comparing genes to which the T2D-DMPs were annotated to results from GWAS, and investigating genetic variants influencing methylation levels. We found many links to genetic associations with T2D and related

phenotypes, highlighting the strong genetic contribution both to methylation levels and to T2D. A subset of five signals, including *OTX2*, had both a robust genetic basis and also showed differences in T2D-discordant MZ twins, suggesting that these signals may be secondary to T2D and could potentially reflect development of T2D complications. We further found strong links between the T2D-DMP genes identified in our study and T2D complications, which were in part supported in metabolomic follow ups. Further analyses in a sample subset identified a nominally significant difference in methylation at one of the FDR 5% T2D-DMPs in *OTX2* among individuals with T2D, between those with diabetic retinopathy and those without.

Using the MCC-seq panel we were able to target around 3.7 million CpG sites and selected sites that showed the most variation in methylation levels between individuals for further analysis. Additionally, this panel was designed to target functional regions of the genome.<sup>28</sup> This resulted in the identification of 20 previously unreported T2D-DMPs including one CpG site in *TXNIP*, which was near to a previously identified T2D-associated CpG site. Of the 20 T2D-DMPs, 8 were located in the first intron of the gene body. Previous work<sup>68</sup> has shown a consistent correlation across both tissues and species between DNA methylation at the first intron and gene expression, indicating that the methylation changes in these sites may also be linked to changes in gene expression. Furthermore, 8 of the 20 T2D-DMPs were located in enhancer regions suggesting potential regulatory impacts. Three of the 20 T2D-DMPs reached genome wide significance at FDR 5% and were located in genes with possible links to diabetic complications such as blood pressure, cardiovascular disease and diabetic retinopathy (*RGL3*, *NGB* and *OTX2*). *RGL3* has been shown to be a potential effector of MRas and Ras.<sup>69</sup> The *MRAS* gene is involved in a number of cell processes including signal transduction and cell growth and is suggested to have a key role in cardiovascular function.<sup>70</sup> Additionally, *RGL3* contains a significant number of SNPs which have been found to be associated with blood pressure.<sup>53</sup> *NGB* encodes a relatively recently discovered neuroglobin which is a neuroprotective protein for the retinal ganglion cells.<sup>54</sup> This protein is strongly expressed in the retina and has been shown to have a protective effect against damage due to high pressure in the eye.<sup>71</sup> It is unclear how this relates to diabetic retinopathy, which is caused by glucose damage to blood vessels in the retina, but a loss of methylation and a potential upregulation of *NGB* during diabetes suggests that more neuroprotective protein is produced in T2D cases in response to the high levels of blood glucose. Furthermore, our follow-up metabolomic analysis found a significant association between *NGB* methylation and levels of choline phosphate in serum. Choline phosphate is used to form citicoline which has been shown in combination with vitamin B12 to

stabilize and decrease functional impairment in diabetic retinopathy.<sup>65</sup> It has also been found to be effective in reducing diabetic neuropathic pain.<sup>64</sup> A significant metabolic association was also found with 4-acetaminobutanoate and methylation at an intergenic site. Metabolite 4-acetaminobutanoate is a precursor to the urea cycle and has been previously linked with kidney health.<sup>67</sup> *OTX2* is also related to eye health and an increase in *OTX2* levels in mouse models has been shown to promote the survival of retinal ganglion cells, the loss of which leads to visual impairment in diabetic retinopathy.<sup>72</sup> Further, we found a nominally significant difference in *OTX2* DNA methylation between individuals with diabetic retinopathy and individuals with diabetes who did not have retinopathy based on a limited sample size. Lastly, gene set enrichment analysis of the genes annotated to the T2D-DMPs identified an enrichment of target genes of binding sites for the transcription factor CHAF1B, which is required for haematopoiesis.<sup>73</sup> Haematopoiesis dysregulation is a consequence of diabetes, suggesting further links between our T2D-DMPs and T2D complications.<sup>74</sup>

The next most significant result was in *SPRED2* where we observed significant differential methylation within 2.5 kB of a SNP (rs2028150), which has been previously associated with T2D in GWAS.<sup>51</sup> This SNP is also associated with levels of methylation in *SPRED2* in GoDMC datasets,<sup>48</sup> along with many other SNPs in the same gene, which could suggest that the differential methylation we observe is due solely to genetic variation between individuals with diabetes and those without. However, we additionally found that the *SPRED2* methylation signal remained significant in the T2D-discordant MZ twin analysis, suggesting that there are both genetic and non-genetic influences on the T2D differential methylation signal in this gene. This is of particular note due to *SPRED2* having been identified as a potential therapeutic tool for the prevention of insulin resistance.<sup>75</sup> A knockout mice study showed that removal of *SPRED2* increased insulin resistance from a high fat diet suggesting *SPRED2* to be a negative regulator of insulin resistance *via* its suppression of the ERK/MAPK pathway.<sup>75</sup>

Almost all the remaining T2D-DMPs annotated to further genes of interest with either functional effects potentially explaining increased propensity of T2D cases for complications, or highlighting genes with a role in metabolic health. Genes linked to T2D complications or comorbidities include *FSCN1*, *ITPR1*, *B3GAT2* and *PLD6*. There is a positive association between diabetes and a number of cancers<sup>76</sup> and we observed differential methylation in the transcription start site of *FSCN1*, which has been shown to be upregulated in cancer.<sup>77</sup> *B3GAT2* encodes a protein which catalyses the formation of glycosaminoglycan-protein linkage. Abnormalities in glycosaminoglycan metabolism have been previously

found in people with diabetes<sup>78</sup> with a study in animal models showing that administration of glycosaminoglycans prevents deterioration in renal health in diabetic rats and reverts established lesions<sup>79</sup>. Diabetes is one of the leading causes of chronic kidney disease with people with diabetes at increased risk of end stage renal disease. The differential methylation is observed in our T2D-discordant analyses and also *B3GAT2* is associated with obesity related traits in GWAS studies<sup>51</sup> suggesting possible genetic and environmental contributors to disease relating to this gene. Some T2D-DMP genes show a link to both T2D complications and genetic associations to T2D or BMI. These include *ITPR1* which has been shown to have a role in both the development of diet induced diabetes<sup>60</sup> and a genetic association with kidney disease. Previous GWAS studies have also found associations between variants in *PLD6* and both BMI and diabetic retinopathy.<sup>51</sup>

Genes with links to metabolic health include *SLC16A3*, *DLX6-AS1* and *MMP15*. *SLC16A3* (also known as *MCT4*) is a member of the monocarboxylate transporter family and has a role in the movement of lactate across the plasma membrane. Lactate is an intermediate metabolite of glucose, and an excess of lactate has been shown to be a risk factor for insulin resistance.<sup>80</sup> *SLC16A3* was also observed to be differentially methylated in the genetically identical discordant twin pairs analysis, suggesting that this methylation signal may be partly environmentally driven or secondary to T2D onset. *DLX6-AS1* is part of a complex that activates transcription of *Dlx-5* and *Dlx-6*. Mouse models show that expression of these genes in GABAergic neurons has significant effects on metabolism, healthy ageing and longevity.<sup>81</sup> Finally, there is evidence that *MMP15* gene expression alters in the development of insulin resistance.<sup>82</sup>

We observed an overlap between the genes annotated to T2D-DMPs and genes identified in GWAS, in line with previous work.<sup>19</sup> The enrichment of T2D-DMPs in T2D GWAS regions was also observed at a slightly more relaxed P-value threshold of FDR 50% indicating it is not only the most significant results that overlap. This supports previous work indicating that the genes that the twin-based T2D-DMPs identify have biological relevance to T2D, where some of these effects are detected at the level of genetic variation in these genes.

We validated one of the T2D-DMPs (chr1:145440435 in *TXNIP*) in the 450k array in a partially independent sample, observing significant differential methylation at cg19693031 associated with both T2D and FBG. This site has been previously reported and replicated as a differentially methylated signal for T2D<sup>9</sup> and is one of the strongest T2D methylation signals independent of BMI (replicated in 5 of 10 blood methylation studies to date<sup>9</sup>). Recent large-scale analyses from over 27,000 samples within the GoDMC study<sup>48</sup> identified many genetic variants influencing DNA methylation levels on

the 450k array, and we used this GoDMC results database to investigate genetic variation influencing *TXNIP* methylation levels. We found that genetic variation at 133 different SNPs in *GLUT1* is associated with methylation levels in *TXNIP*. *TXNIP* expression causes a reduction in GLUT1, which in turn transports more blood sugar when blood glucose is high. A knock down of GLUT1 in mice using siRNA has been shown to reduce retinal blood glucose in diabetic mice to the level of non-diabetic mice providing the potential for mitigating diabetic retinopathy.<sup>63</sup> GoDMC datasets in over 30,000 samples show that genetic variation in *GLUT1* affects the level of methylation observed in *TXNIP*. As such *TXNIP* methylation levels may provide an indicator for how well the GLUT1 transporter is functioning and potentially identify individuals who are susceptible to T2D complications. *TXNIP* hypomethylation with T2D also shows parallels with *AHRR* hypomethylation in response to smoke exposure. In both cases upregulation of the gene inhibits the body's ability to detoxify. *TXNIP* binds to thioredoxin (Trx) inhibiting its ability to neutralise oxidants, and prolonged overexpression of *TXNIP* has been shown to lead to premature cell death in diabetic retinopathy.<sup>83</sup>

There are several limitations to this study. The MCC-seq read depth cut-off of 5 reads is relatively low. On the other hand, the average read depth for the T2D DMP signals identified is around 20 or greater at most signals (Supplementary Table S10). A relatively low read depth threshold could lead to a greater false positive rate and lower accuracy compared to previous microarray based studies. Whilst the number of participants is larger than many DNA methylation studies of T2D and includes the largest number of T2D-discordant MZ twins, the sample size is still relatively modest and provides only moderate power to detect the small changes in methylation associated with the disease. Sex-specific effects were not explored due to the relatively modest sample size. The participants are nearly all of European ancestry and the results may not translate to other ethnicities. This study included only DNA methylation and not gene expression follow-ups. We are therefore not able to determine whether the differential methylation we observe also has an impact on gene expression, however, we were able to integrate circulating metabolic levels as a functional follow up. DNA methylation is a dynamic marker, and longitudinal studies are needed to explore the longitudinal stability, or potential reversibility, of the identified T2D-DMPs. Such longitudinal studies would be particularly beneficial in diseases like T2D, which progress over time and result in risk of developing clinical complications, which increases with longer periods of sustained high blood sugar. It would also be insightful to study methylation profiles in individuals whose blood glucose levels normalise following lifestyle interventions, in contrast to those where T2D progresses.

In conclusion, we observe a number of previously unreported CpG sites differentially methylated with T2D, which are annotated to genes with functional links to diabetic health. There is a strong interlink between genetics and methylation, and it is likely that most T2D associated DNA methylation signals arise as a consequence of disease because a number of sites are in genes linked to T2D complications or consequences of high blood glucose. Methylation levels at these sites may be potential biomarkers of disease progression and may help to understand mechanisms for the development of complications.

#### Contributors

JTB and TDS designed the study. EG oversaw MCC-seq methylation profiling. CC led the data analysis with major inputs from LP and TM, and further support from SVM, JEC-F, MM, PC-T, SW, SC and PC. JTB, TDS, JO, OM, PS, KM, KHP, MO, CD, KK, KC, DB, GW, JvD, PM, NP, SW, CM, JT contributed samples and/or data. CC and JTB wrote the manuscript. All authors read and approved the manuscript. JTB and TM verified the underlying data.

#### Data sharing statement

Discovery phase DNA methylation profiles with the exception of Danish cohort data are available from the European Genome-phenome Archive (EGA; Accession no. EGAD5000000286). Covariate and phenotype data are available upon request from each twin cohort as described below.

TwinsUK: Additional individual-level data are not permitted to be publicly shared or deposited due to the original consent given at the time of data collection, where access to these data is subject to governance oversight. However, these data can be applied for through a data access application, where all data access requests are overseen by the TwinsUK Resource Executive Committee (TREC). For information on access to these genotype and phenotype data and how to apply see <https://twinsuk.ac.uk/resources-for-researchers/access-our-data/>

Older Australian Twins Study: Data provided by the participants can be shared with other research groups by application to the CHeBA Research Bank and scientific review by the OATS Governance Committee (see <https://cheba.unsw.edu.au/research-projects/older-australian-twins-study/>). A data application form can be obtained by emailing [CHeBAData@unsw.edu.au](mailto:CHeBAData@unsw.edu.au).

Finnish Twin Cohort Study: Data are deposited with the Biobank of the Finnish Institute for Health and Welfare (<https://thl.fi/en/web/thl-biobank/for-researchers/sample-collections/twin-study>) with identification number THLBB2021\_001. Data are available upon request, after approval of application by the biobank. For details on accessing the data, see <https://thl.fi/en/web/thl-biobank/for-researchers/application-process>.

Netherlands Twin Register: Data are available upon request, after approval of application by the NTR Data Access Committee. For information on procedures please see [https://tweelingenregister.vu.nl/information\\_for\\_researchers/information-for-researchers](https://tweelingenregister.vu.nl/information_for_researchers/information-for-researchers) and (<https://ntr-data-request.psy.vu.nl/>).

Murcia Twin Registry: Data are available upon request, after approval of application by the Murcia Twin Registry Committee. Information on data access procedures can be obtained from [geminis@um.es](mailto:geminis@um.es).

Danish Twin Registry: Individual-level data are not permitted to be publicly shared or deposited due to the original consent given at the time of data collection, as well as GDPR. All data access requests are overseen by the Danish Twin Research Center. For information on access to data and how to apply, see [https://www.sdu.dk/en/om\\_sdu/institutter\\_centre/ist\\_sundhedstjenesteforsk/centre/dtr/researcher](https://www.sdu.dk/en/om_sdu/institutter_centre/ist_sundhedstjenesteforsk/centre/dtr/researcher).

Swedish Twin Registry: As a national research infrastructure the STR is open to applications also from abroad. Confidentiality review determines data access. However, linkage of additional personal data to large data repositories such as European Genome-phenome Archive

(EGA) is generally not permitted. For information about how to apply, see <https://ki.se/en/research/the-swedish-twin-registry>.

#### Declaration of interests

The authors have no conflicts to declare.

#### Acknowledgements

We thank all the participants in the study.

Funding acknowledgements for each cohort can be found in the Supplementary Note. CC was supported by the National Institute for Health and Care Research (NIHR) Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust and King's College London and/or the NIHR Clinical Research Facility. This project also received support from the JPI ERA-HDHL DIMENSION project and UK Biological Sciences Research Council (BBSRC, BB/S020845/1 and BB/T019980/1 to JTB). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

#### Appendix A. Supplementary data

Supplementary data related to this article can be found at <https://doi.org/10.1016/j.ebiom.2024.105096>.

#### References

- World Health Organisation. *Global report on type 2 diabetes*. 2016.
- Cho N, Shaw J, Karuranga S, et al. IDF Diabetes Atlas: global estimates of diabetes prevalence for 2017 and projections for 2045. *Diabetes Res Clin Pract*. 2018;138:271–281.
- Vujkovic M, Keaton JM, Lynch JA, et al. Discovery of 318 new risk loci for type 2 diabetes and related vascular outcomes among 1.4 million participants in a multi-ancestry meta-analysis. *Nat Genet*. 2020;52:680–691.
- Willemsen G, Ward KJ, Bell CG, et al. The concordance and heritability of type 2 diabetes in 34,166 twin pairs from international twin registers: the discordant twin (DISCOTWIN) consortium. *Twin Res Hum Genet*. 2015;18(6):762–771.
- Chambers JC, Loh M, Lehne B, et al. Epigenome-wide association of DNA methylation markers in peripheral blood from Indian Asians and Europeans with incident type 2 diabetes: a nested case-control study. *Lancet Diabetes Endocrinol*. 2015;3(7):526–534.
- Al Muflih WA, Al-Shafai M, Zaghlool SB, et al. Epigenetic associations of type 2 diabetes and BMI in an Arab population. *Clin Epigenet*. 2016;8:13.
- Soriano-Tárraga C, Jiménez-Conde J, Giralte-Steinhauer E, et al. GENESTROKE Consortium. Epigenome-wide association study identifies TXNIP gene associated with type 2 diabetes mellitus and sustained hyperglycemia. *Hum Mol Genet*. 2016;25(3):609–619.
- Walaszczyk E, Luijten M, Spijkerman AMW, et al. DNA methylation markers associated with type 2 diabetes, fasting glucose and HbA1c levels: a systematic review and replication in a case-control sample of the Lifelines study. *Diabetologia*. 2018;61(2):354–368.
- Juviniao-Quintero DL, Marioni RE, Ochoa-Rosales C, et al. DNA methylation of blood cells is associated with prevalent type 2 diabetes in a meta-analysis of four European cohorts. *Clin Epigenet*. 2021;13:40.
- Ling C, Del Guerra S, Lupi R, et al. Epigenetic regulation of PPARGC1A in human type 2 diabetic islets and effect on insulin secretion. *Diabetologia*. 2008;51(4):615–622.
- Yang BT, Dayeh TA, Kirkpatrick CL, Taneera J, Kumar R, Groop L. Insulin promoter DNA methylation correlates negatively with insulin gene expression and positively with HbA(1c) levels in human pancreatic islets. *Diabetologia*. 2011;54:360–367.
- Yang BT, Dayeh TA, Volkov PA, et al. Increased DNA methylation and decreased expression of PDX-1 in pancreatic islets from patients with type 2 diabetes. *Mol Endocrinol*. 2012;26:1203–1212.
- Volkmar M, Dedeurwaerder S, Cunha DA, et al. “DNA methylation profiling identifies epigenetic dysregulation in pancreatic islets from type 2 diabetic patients”. *EMBO J*. 2012;31:1405–1426.
- Dayeh T, Volkov P, Saló S, et al. Genome-wide DNA methylation analysis of human pancreatic islets from type 2 diabetic and non-diabetic donors identifies candidate genes that influence insulin secretion. *PLoS Genet*. 2014;10:e1004160.



- 15 Volkov P, Bacos K, Ofori JK, et al. Whole-genome bisulfite sequencing of human pancreatic islets reveals novel differentially methylated regions in type 2 diabetes pathogenesis. *Diabetes*. 2017;66(4):1074–1085.
- 16 Nilsson E, Matte A, Perflyev A, et al. Epigenetic alterations in human liver from subjects with type 2 diabetes in parallel with reduced folate levels. *J Clin Endocrinol Metab*. 2015;100(11):1491–1501.
- 17 Nilsson E, Jansson PA, Perflyev A, et al. Altered DNA methylation and differential expression of genes influencing metabolism and inflammation in adipose tissue from subjects with type 2 diabetes. *Diabetes*. 2014;63:2962–2976.
- 18 Rönn T, Volkov P, Gillberg L, et al. Impact of age, BMI and HbA1c levels on the genome-wide DNA methylation and mRNA expression patterns in human adipose tissue and identification of epigenetic biomarkers in blood. *Hum Mol Genet*. 2015;24(13):3792–3813.
- 19 Yuan W, Xia Y, Bell CG, et al. An integrated epigenomic analysis for type 2 diabetes susceptibility loci in monozygotic twins. *Nat Commun*. 2014;5:5719.
- 20 Hwang J-Y, Lee H-J, Go MJ, et al. Genome-wide methylation analysis identifies ELOVL5 as an epigenetic biomarker for the risk of type 2 diabetes mellitus. *Nature Sci Rep*. 2018;8:14862.
- 21 Ribel-Madsen R, Fraga MF, Jacobsen S, Bork-Jensen J, Lara E, Calvanese V. Genome-wide analysis of DNA methylation differences in muscle and fat from monozygotic twins discordant for type 2 diabetes. *PLoS One*. 2012;7:e51302.
- 22 Kulkarni H, Kos MZ, Neary J, et al. Novel epigenetic determinants of type 2 diabetes in Mexican-American families. *Hum Mol Genet*. 2015;24(18):5330–5344.
- 23 Schrader S, Perflyev A, Ahlqvist E, et al. Novel subgroups of type 2 diabetes display different epigenetic patterns that associate with future diabetic complications. *Diabetes Care*. 2022;45(7):1621–1630.
- 24 Devi T, Somayajulu M, Kowluru R, et al. TXNIP regulates mitophagy in retinal Müller cells under high-glucose conditions: implications for diabetic retinopathy. *Cell Death Dis*. 2017;8:e2777.
- 25 Perrone L, Devi TS, Hosoya KI, Terasaki T, Singh LP. Inhibition of TXNIP expression in vivo blocks early pathologies of diabetic retinopathy. *Cell Death Dis*. 2010;1:e65.
- 26 Gunasekara CJ, MacKay H, Scott CA, et al. Systemic interindividual epigenetic variation in humans is associated with transposable elements and under strong genetic control. *Genome Biol*. 2023;24(2).
- 27 Ziller MJ, Gu H, Müller F, et al. Charting a dynamic DNA methylation landscape of the human genome. *Nature*. 2013;500:477–481.
- 28 Allum F, Shao X, Guénard F, et al. Characterization of functional methylomes by next-generation capture sequencing identifies novel disease-associated variants. *Nat Commun*. 2015;6:7211.
- 29 Allum F, Hedman ÅK, Shao X, et al. Dissecting features of epigenetic variants underlying cardiometabolic risk using full-resolution epigenome profiling in regulatory elements. *Nat Commun*. 2019;10(1):1209.
- 30 Li X, Shao X, Bazzano LA, et al. Blood DNA methylation at TXNIP and glycemic changes in response to weight-loss diet interventions: the POUNDS lost trial. *Int J Obes*. 2022;46(6):1122–1127.
- 31 Cheung WA, Shao X, Morin A, et al. Functional variation in allelic methylomes underscores a strong genetic contribution and reveals novel epigenetic alterations in the human epigenome. *Genome Biol*. 2017;18(1):50.
- 32 Busche S, Shao X, Caron M, et al. Population whole-genome bisulfite sequencing across two tissues highlights the environment as the principal source of human methylome variation. *Genome Biol*. 2015;16:290.
- 33 Kurushima Y, Tsai PC, Castillo-Fernandez J, et al. Epigenetic findings in periodontitis in UK twins: a cross-sectional study. *Clin Epigenet*. 2019;11(1):27.
- 34 Du P, Zhang X, Huang C-C, et al. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinf*. 2010;11:587.
- 35 Xu Z, Niu L, Li L, Taylor J. ENmix: a novel background correction method for Illumina HumanMethylation 450 BeadChip. *Nucleic Acids Res*. 2016;44(3):e20.
- 36 Aryee M, Jaffe A, Corrada-Bravo H, et al. Minfi: a flexible and comprehensive bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics*. 2014;30:1363–1369.
- 37 Bell CG, Gao F, Yuan W, et al. Obligatory and facilitative allelic variation in the DNA methylome within common disease-associated loci. *Nat Commun*. 2018;9(1):1–13.
- 38 Bell CG, Xia Y, Yuan W, et al. Novel regional age-associated DNA methylation changes within human common disease-associated loci. *Genome Biol*. 2016;17(1):1–14.
- 39 Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754–1760.
- 40 Chavez L, Jozefczuk J, Grimm C, et al. Computational analysis of genome-wide DNA methylation during the differentiation of human embryonic stem cells along the endodermal lineage. *Genome Res*. 2010;20(10):1441–1450.
- 41 Houseman EA, Molitor J, Marsit CJ. Reference-free cell mixture adjustments in analysis of DNA methylation data. *Bioinformatics*. 2014;30:1431–1439.
- 42 Jaffe AE, Irizarry RA. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biol*. 2013;15:R31.
- 43 Mootha V, Lindgren C, Eriksson KF, et al. PGC-1 $\alpha$ -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet*. 2003;34:267–273.
- 44 Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA*. 2005;102(43):15545–15550.
- 45 Tsai P, Glastonbury C, Eliot M, et al. Smoking induces coordinated DNA methylation and gene expression changes in adipose tissue with consequences for metabolic health. *Clin Epigenet*. 2018;10:26.
- 46 Shabalin AA. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics*. 2012;28:1353–1358.
- 47 Grundberg E, Meduri E, Sandling JK, et al. Global analysis of DNA methylation variation in adipose tissue from twins reveals links to disease-associated variants in distal regulatory elements. *Am J Hum Genet*. 2013;93(5):876–890.
- 48 Min JL, Hemani G, Hannon E, et al. Genomic and phenotypic insights from an atlas of genetic effects on DNA methylation. *Nat Genet*. 2021;53:1311–1321.
- 49 Xue A, Wu Y, Zhu Z, et al. Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. *Nat Commun*. 2018;9:2941.
- 50 Mahajan A, Taliun D, Thurner M, et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat Genet*. 2018;50(11):1505–1513.
- 51 Buniello A, MacArthur JAL, Cerezo M, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics. *Nucleic Acids Res*. 2019;47(D1):D1005–D1012.
- 52 Long T, Hicks M, Yu HC, et al. Whole-genome sequencing identifies common-to-rare variants associated with human blood metabolites. *Nat Genet*. 2017;49(4):568–578.
- 53 Wang Y, Wang JG. Genome-wide association studies of hypertension and several other cardiovascular diseases. *Pulse*. 2019;6(3–4):169–186.
- 54 Tun SBB, Barathi VA, Luu CD, Lynn MN, Chan ASY. Effects of Exogenous Neuroglobin (Ngb) on retinal inflammatory chemokines and microglia in a rat model of transient hypoxia. *Sci Rep*. 2019;9(1):18799.
- 55 Beby F, Lamonerie T. The homeobox gene *Otx2* in development and disease. *Exp Eye Res*. 2013;111:9–16.
- 56 Campbell JE, Ussher JR, Mulvihill EE, et al. TCF1 links GIPR signaling to the control of beta cell function and survival. *Nat Med*. 2016;22(1):84–90.
- 57 Kaur K, Vig S, Srivastava R, et al. Elevated hepatic miR-22-3p expression impairs gluconeogenesis by silencing the wnt-responsive transcription factor Tcf 7. *Diabetes*. 2015;64(11):3659–3669.
- 58 Erlich H, Valdes A, Julier C, et al. Evidence for association of the TCF7 locus with type 1 diabetes. *Genes Immun*. 2009;10(Suppl 1):S54–S59.
- 59 Campeanu IJ, Jiang Y, Liu L, et al. Multi-omics integration of methyltransferase-like protein family reveals clinical outcomes and functional signatures in human cancer. *Sci Rep*. 2021;11:14784.
- 60 Ye R, Ni M, Wang M, et al. Inositol 1,4,5-trisphosphate receptor 1 mutation perturbs glucose homeostasis and enhances susceptibility to diet-induced diabetes. *J Endocrinol*. 2011;210(2):209–217.
- 61 Li S, Wang L, Sun S, Wu Q. Hepsin (2021) “a multifunctional transmembrane serine protease in pathobiology”. *FEBS J*. 2021;288(18):5252–5264.
- 62 Waldhart AN, Dykstra H, Peck AS, et al. Phosphorylation of TXNIP by AKT mediates acute influx of glucose in response to insulin. *Cell Rep*. 2017;19(10):2005–2013.
- 63 Lu L, Seidel CP, Iwase T, et al. Suppression of GLUT1: a new strategy to prevent diabetic complications. *J Cell Physiol*. 2013;228(2):251–257.

- 64 Ahlawat A, Sharma S. A new promising simultaneous approach for attenuating type II diabetes mellitus induced neuropathic pain in rats: iNOS inhibition and neuroregeneration. *Eur J Pharmacol.* 2018;818:419–428.
- 65 Parravano M, Scarinci F, Parisi V, et al. Citicoline and vitamin B<sub>12</sub> eye drops in type 1 diabetes: results of a 3-year pilot study evaluating morpho-functional retinal changes. *Adv Ther.* 2020;37(4):1646–1663.
- 66 Tsalik EL, Willig LK, Rice BJ, et al. Renal systems biology of patients with systemic inflammatory response syndrome. *Kidney Int.* 2015;88(4):804–814.
- 67 Mindikoglu AL, Opekun AR, Putluri N, et al. Unique metabolomic signature associated with hepatorenal dysfunction and mortality in cirrhosis. *Transl Res.* 2018;195:25–47.
- 68 Anastasiadi D, Esteve-Codina A, Piferrer F. Consistent inverse correlation between DNA methylation of the first intron and gene expression across tissues and species. *Epigenet Chromatin.* 2018;11:37.
- 69 Ehrhardt G, Korherr C, Wieler J, et al. A novel potential effector of M-Ras and p21 Ras negatively regulates p21 Ras-mediated gene induction and cell growth. *Oncogene.* 2001;20:188–197.
- 70 Song Y, Ma R, Zhang H. The influence of MRAS gene variants on ischemic stroke and serum lipid levels in Chinese Han population. *Medicine.* 2019;98(48):e18065.
- 71 Wei X, Yu Z, Cho KS, et al. Neuroglobin is an endogenous neuroprotectant for retinal ganglion cells against glaucomatous damage. *Am J Pathol.* 2011;179(6):2788–2797.
- 72 Torero Ibad R, Rhee J, Mrejen S, et al. Otx2 promotes the survival of damaged adult retinal ganglion cells and protects against excitotoxic loss of visual acuity in vivo. *J Neurosci.* 2011;31(14):5495–5503.
- 73 Volk A, Liang K, Suraneni P, et al. A CHAF1B-dependent molecular switch in hematopoiesis and leukemia pathogenesis. *Cancer Cell.* 2018;34(5):707–723.e7.
- 74 Paolo Fadini G, Albiero M. Impaired hematopoietic stem/progenitor cell traffic and multi-organ damage in diabetes. *Stem Cell.* 2022;40(8):716–723.
- 75 Ohkura T, Yoshimura T, Fujisawa M, et al. Spred2 regulates high fat diet-induced adipose tissue inflammation, and metabolic abnormalities in mice. *Front Immunol.* 2019;10:17.
- 76 Giovannucci E, Harlan DM, Archer MC, et al. Diabetes and cancer: a consensus report. *Diabetes Care.* 2010;33(7):1674–1685.
- 77 Wang CQ, Tang CH, Wang Y, et al. FSCN1 gene polymorphisms: biomarkers for the development and progression of breast cancer. *Sci Rep.* 2017;7(1):15887.
- 78 Gambaro G, Baggio B. Role of glycosaminoglycans in diabetic nephropathy. *Acta Diabetol.* 1992;29:149–155.
- 79 Gambaro G, Venturini AP, Noonan DM, et al. Treatment with a glycosaminoglycan formulation ameliorates experimental diabetic nephropathy. *Kidney Int.* 1994;46(3):797–806.
- 80 Metz L, Mercier J, Tremblay A, Alméras N, Joannisse DR. Effect of weight loss on lactate transporter expression in skeletal muscle of obese subjects. *J Appl Phys.* 2008;104:633–638.
- 81 de Lombares C, Heude E, Alfama G, et al. *Dlx5* and *Dlx 6* expression in GABAergic neurons controls behavior, metabolism, healthy aging and lifespan. *Aging.* 2019;11(17):6638–6656.
- 82 Tinahones FJ, Coín-Aragüez L, Mayas MD, et al. Obesity-associated insulin resistance is correlated to adipose tissue vascular endothelial growth factors and metalloproteinase levels. *BMC Physiol.* 2012;12:4.
- 83 Singh J. Thioredoxin interacting protein (TXNIP) and pathogenesis of diabetic retinopathy. *Clin Exp Ophthalmol.* 2013;4(4).