Title

Epigenomic profiling reveals an association between persistence of DNA methylation and metabolic memory in the DCCT/EDIC type 1 diabetes cohort

Journal

Proceedings of the National Academy of Sciences of the United States of America, 113(21)

ISSN

0027-8424

Authors

Chen, Zhuo
Miao, Feng
Paterson, Andrew D
et al.

Publication Date

2016-05-24

DOI

10.1073/pnas.1603712113

Peer reviewed

Epigenomic profiling reveals an association between persistence of DNA methylation and metabolic memory in the DCCT/EDIC type 1 diabetes cohort

Zhuo Chen[a,1], Feng Miao[a,1], Andrew D. Paterson[b], John M. Lachin[c], Lingxiao Zhang[a], Dustin E. Schones[a], Xiwei Wu[d], Jinhui Wang[d], Joshua D. Tompkins[a], Saul Genuth[e], Barbara H. Braffett[c], Arthur D. Riggs[a,2], DCCT/EDIC Research Group[3], and Rama Natarajan[a,2]

[a]Department of Diabetes Complications and Metabolism, Beckman Research Institute of City of Hope, Duarte, CA 91010; [b]Genetics and Genome Biology, The Hospital for Sick Children, Toronto, ON M5G 0A4, Canada; [c]The Biostatistics Center, The George Washington University, Rockville, MD 20852-3943; [d]Department of Molecular and Cellular Biology, Beckman Research Institute of City of Hope, Duarte, CA 91010; and [e]Case Western Reserve University, Cleveland, OH 44106

Contributed by Arthur D. Riggs, March 11, 2016 (sent for review January 24, 2016; reviewed by John Cijiang He, Renu A. Kowluru, and Kumar Sharma)

We examined whether persistence of epigenetic DNA methylation (DNA-me) alterations at specific loci over two different time points in people with diabetes are associated with metabolic memory, the prolonged beneficial effects of intensive vs. conventional therapy during the Diabetes Control and Complications Trial (DCCT) on the progression of microvascular outcomes in the long-term follow-up Epidemiology of Diabetes Interventions and Complications (EDIC) Study. We compared DNA-me profiles in genomic DNA of whole blood (WB) isolated at EDIC Study baseline from 32 cases (DCCT conventional therapy group subjects showing retinopathy or albuminuria progression by EDIC Study year 10) vs. 31 controls (DCCT intensive therapy group subjects without complication progression by EDIC year 10). DNA-me was also profiled in blood monocytes (Monos) of the same patients obtained during EDIC Study years 16–17. In WB, 153 loci depicted hypomethylation, and 225 depicted hypermethylation, whereas in Monos, 155 hypomethylated loci and 247 hypermethylated loci were found (fold change ≥1.3; $P < 0.005$; cases vs. controls). Twelve annotated differentially methylated loci were common in both WB and Monos, including thioredoxin-interacting protein (*TXNIP*), known to be associated with hyperglycemia and related complications. A set of differentially methylated loci depicted similar trends of associations with prior HbA1c in both WB and Monos. In vitro, high glucose induced similar persistent hypomethylation at *TXNIP* in cultured THP1 Monos. These results show that DNA-me differences during the DCCT persist at certain loci associated with glycemia for several years during the EDIC Study and support an epigenetic explanation for metabolic memory.

metabolic memory | epigenetics | DNA methylation | *TXNIP* | diabetic complications

The landmark Diabetes Control and Complications Trial (DCCT; 1983–1993) clearly showed that intensive (INT) glycemic control profoundly reduces the development and progression of microvascular complications in type 1 diabetes (T1D). The DCCT participants were subsequently followed in the Epidemiology of Diabetes Interventions and Complications (EDIC) Study (1994 to present), during which all subjects were advised to practice INT treatment. Surprisingly, those previously assigned to conventional (CONV) therapy continued to develop complications, such as nephropathy, retinopathy, and macrovascular diseases, at significantly higher rates than the previous INT therapy group, despite nearly similar HbA1c levels during the EDIC Study (1–3). This persistence of benefit from early application of INT therapy, called "metabolic memory," is an enigma in the field of T1D: recent studies have suggested the involvement of epigenetic mechanisms (4–9).

Epigenetics is the study of mostly heritable changes in gene expression and phenotype that occur without alterations in the underlying DNA sequence. Epigenetic states are affected by environmental factors, such as aberrant nutrition and metabolic states (4, 6–8, 10). DNA methylation (DNA-me; the classic epigenetic mark) and posttranslational modifications (PTMs) of histone tails in chromatin (e.g., acetylation and methylation) form an epigenome layer and play key roles in controlling gene expression (11–14). Promoter DNA-me is usually associated with gene repression and usually inversely correlated with histone lysine acetylation, which is generally associated with active genes. Histone lysine methylation can lead to gene activation or repression (14, 15). Abnormal alterations in DNA-me and histone PTMs at distinct regulatory regions in the genome can eventually lead to dysregulated gene transcription and disease progression (13–16). DNA-me levels are controlled by both methylation catalyzed by DNA methyltransferases (DNMTs), and demethylation caused by oxidation of methylcytosine by the ten-eleven translocation (TET) enzymes or deamination by deaminases. DNA-me

Significance

Vascular complications are the main cause of morbidity and mortality in the diabetic population. Clinical trials of diabetic complications show a persistence of benefit from early application of intensive therapy for glycemic control in diabetic patients, a phenomenon referred to as metabolic memory. The mechanisms underlying metabolic memory are not fully understood. In this study, using two groups of type 1 diabetic patients with and without complications development and two sets of genomic DNAs collected 16–17 y apart from the same patients, we showed a persistency of DNA methylation over time at key genomic loci associated with diabetic complications. These data provide direct evidence of a relationship between epigenetics (DNA methylation variations) and human metabolic memory, supporting an epigenetic mechanism.

Author contributions: Z.C., F.M., A.D.P., J.M.L., X.W., S.G., A.D.R., D.E.R.G., and R.N. designed research; Z.C., F.M., L.Z., and J.W. performed research; Z.C., F.M., A.D.P., J.M.L., D.E.S., X.W., J.D.T., B.H.B., A.D.R., and R.N. analyzed data; and Z.C., F.M., A.D.P., J.M.L., X.W., and R.N. wrote the paper.

Reviewers: J.C.H., Mount Sinai School of Medicine; R.A.K., Wayne State University; and K.S., University of California, San Diego.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

Data deposition: DNA methylation profilings of all samples (including both whole-blood samples and monocyte samples) by Infinium Human DNA Methylation 450K Array have been deposited in the Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo (accession no. GSE76171). Trial Registration is at clinicaltrials.gov (NCT00360815 and NCT00360893).

[1]Z.C. and F.M. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: ariggs@coh.org or rnatarajan@coh.org.

[3]A complete list of the DCCT/EDIC Research Group can be found in *SI Appendix S2*.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1603712113/-/DCSupplemental.

is relatively long-lasting and may function to stably lock in gene silencing and/or reflect some aspects of cellular history (17–19).

Several laboratories, including our laboratory, using cell and animal models have suggested roles for key histone PTMs in high glucose (HG)-mediated effects, diabetic complications, and metabolic memory (9, 20–31). Histone PTM profiling of vascular and inflammatory cells cultured with HG vs. normal glucose (NG) or WBCs from T1D patients vs. healthy volunteers or T1D patients with complications vs. those without depicted significant PTM differences and candidate differentially methylated or acetylated genes relevant to diabetes or its complications (23, 32–34). Because DNA-me can be inherited during cell division, persistence of DNA-me changes might reflect metabolic history. This hypothesis is supported by experimental models showing persistent DNA-me changes (global or at specific loci related to complications) in zebrafish with hyperglycemia (35), diabetic patient-derived fibroblasts (36), and retinas from diabetic rats under poor glucose control (37). However, the cause–effect relationship between epigenetics and human metabolic memory is not yet fully clear.

We recently did histone PTMs profiling (31) in peripheral blood monocytes (Monos) and lymphocytes collected from 30 former DCCT CONV group subjects (cases; mean DCCT HbA1c > 9.1% and progression of retinopathy or macroalbuminuria by EDIC Study year 10 of follow-up) vs. 30 former DCCT INT subjects (controls; mean DCCT HbA1c < 7.3% and without progression through EDIC Study year 10). We showed that Monos from cases were significantly enriched with histone H3-lysine9-acetylation (H3K9Ac) relative to controls at promoter regions. Interestingly, the top 38 hyperacetylated promoters in cases included 15 genes associated with the nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB) inflammatory pathway, which is strongly associated with diabetic complications. Moreover, Monos H3K9Ac, a mark of gene activation, showed significantly positive association with mean HbA1c during the DCCT and the EDIC Study. Together, these results suggested the potential mediatory role of epigenetics in metabolic memory in humans. However, this first study did not evaluate the persistence of epigenetic marks over time in the DCCT/EDIC Study cohort.

Although aberrant DNA-me has been reported to be associated with cancer progression (13, 16, 38), only a few studies have examined the role of DNA-me in T1D complications (23, 32–34, 39). Notably, it is unknown if DNA-me persists over time in people with diabetes and is associated with metabolic memory. Here, we addressed this issue by using DNA collected at two different time points (EDIC Study baseline and EDIC Study years 16 and 17) from the same DCCT/EDIC Study cohort in which we previously evaluated histone PTMs (31). To our knowledge, this study is the first one in which two sets of DNAs collected at least 16–17 y apart from the same participants are used to show the persistency of DNA-me over time. Our goal is to provide more direct evidence of a relationship between epigenetics and metabolic memory in human T1D.

Results

Experimental Design. Participants from the DCCT/EDIC Study cohort were selected based on the same criteria as those reported in our previous histone PTM study (31). Briefly, the case group included 32 EDIC Study participants from the former DCCT CONV group who had mean HbA1c > 9.1% (76 mmol/mol) during the DCCT and significant progression of retinopathy and/or nephropathy from the DCCT closeout to EDIC Study year 10. The control group consisted of 31 EDIC Study participants from the former DCCT INT group. They had mean DCCT HbA1c below 7.3% (56 mmol/mol) and no progression of retinopathy and/or nephropathy through EDIC Study year 10. The characteristics of these 63 participants are shown in *SI Appendix S1*, Table S1, and they are very similar to those previously reported (31).

Two sets of DNA samples from the same 63 EDIC Study participants [including all 60 participants from the EDIC Study histone PTM study (31)] collected at least 16–17 y apart (Table 1) were used in this study. The first set was comprised of whole-blood (WB) genomic DNAs collected at the DCCT closeout/EDIC Study baseline (1991–1993) for genetic studies (40). These samples were used to examine the DNA-me status in patients when glucose control, as measured by HbA1c, differed between the two groups, and microvascular complications were more common in cases than in controls (*SI Appendix S1*, Table S1). The second set of samples analyzed was genomic DNAs isolated from Monos of 61 subjects, including 31 cases and 30 controls (two Monos samples could not be analyzed because of insufficient DNA). These Monos samples were collected during EDIC Study years 16 and 17 (2009 and 2010) when histone PTM profiling was conducted (31), and were used to profile DNA-me 16–17 y later in the same patients, when diabetic complications in cases became more pronounced during the EDIC Study, despite reduction in the HbA1c difference (*SI Appendix S1*, Table S1).

As shown in Fig. 1, all genomic DNA samples from WB and Monos were subjected to DNA-me profiling using Illumina Infinium Human Methylation 450 Bead Chip Arrays. The 485,512 probes in the array cover 99% RefSeq genes (including promoter, body, and 5′/3′ UTRs), 3′-cytosine-phosphate-guanine-5′ (CpG) islands (CGIs), CpG shores, and shelves. Following our data analysis pipeline (*SI Appendix S1*, Fig. S1*A*), we identified differentially methylated CpGs (DMLs) between cases and controls in both WB and Monos, validated several of these using different platforms, and investigated persistence of DNA-me changes over time by comparing DMLs identified from the two sets of samples. Finally, in vitro cell culture experiments were performed to shed light on the molecular mechanism(s) underlying DNA-me changes and gene expression differences and further explore the association of DNA-me with diabetic complications and metabolic memory.

Identification and Characteristics of DMLs in WB Genomic DNA Samples. DNA-me at 213,620 probes (after CpG site exclusions) from 450K arrays (*SI Appendix S1*, Fig. S1*B*) were compared between cases and controls in WB DNAs using multiple linear regression models with and without adjustments for covariates including age, gender, WBC type composition, and complications (both proliferative diabetic retinopathy and microalbuminiuria) at the EDIC Study baseline. As indicated in *SI Appendix S1*, *SI Methods*, because epigenetics is cell-specific, to address the issue of heterogeneity of DNA in WB, we adopted a data deconvolution (41) approach to estimate cell type composition in each WB DNA sample and then, added the estimated cell compositions as covariates in the analysis. The compositions of each cell type in all 63 WB samples were within the expected range (*SI Appendix S1*, Fig. S1*C*). Moreover, there were no significant differences in compositions for each cell type, except in Monos, between cases and controls (*SI Appendix S1*, Table S2). This approach was further verified by our Monos dataset, showing an estimated average of 84% Monos across 61 Mono samples (*SI Appendix S1*, Fig. S1*D*). These data confirmed the feasibility and reliability of using deconvolution to estimate cell type composition in WB.

DMLs in WB of cases vs. controls were thus identified based on fold changes (FCs) ≥1.3 and nominal $P < 0.005$. As shown in Fig. 2*A*, without adjustments for covariates, 131 hypomethylated loci (hypo-MLs) and 141 hypermethylated loci (hyper-MLs) in cases vs. controls were identified. After adjustment for covariates, 153 hypo-MLs (Dataset S1) and 225 hyper-MLs (Dataset S2) were found. In comparing the lists with and without adjustment, about one in three (51 hypo-MLs and 63 hyper-MLs) were in common, indicating that these covariates did, to some extent, account for

Table 1. DCCT/EDIC Study genomic DNA sample information

Samples	Case	Control	Time of sample collection
WB DNA	32	31	EDIC Study baseline (1991–1993)
Monos	31	30	EDIC Study years 16–17 (2009–2010)

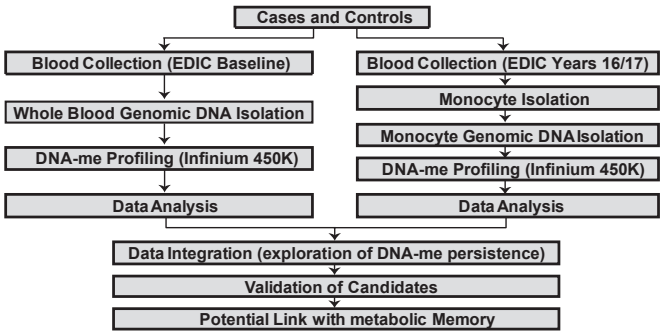

Fig. 1. Experimental design. The experiments and related data analyses are summarized in the depicted pipeline.

variation in DNA-me. In silico analyses were subsequently done to further characterize the DMLs identified with adjustment in WB.

Using unsupervised hierarchical clustering analysis on the DNA-me at DMLs across 63 patients, a heat map (Fig. 2*B*) shows clear DNA-me differences between the two groups. Based on genomic locations relative to RefSeq genes, each DML was classified to one of the following eight regions: transcription start site (TSS) 1500 (1,500 bp upstream to 200 bp upstream of RefSeq genes), TSS200 (200 bp upstream to TSS), 5′ UTR, first exon, 3′ UTR, other exon (excluding first exon), and intron and intergenic regions (Fig. 2*C*). No obvious location differences were found between hypo- and hyper-MLs ($P = 0.90$; Fisher exact test). In both, about 60% of DMLs were located within gene bodies. Interestingly, the numbers of DMLs located in introns were relatively high (37.3% in hypo-MLs and 39.1% in hyper-MLs), and those in 3′ UTR also accounted for 6.5% and 7.1% in hypo- and hyper-MLs, respectively. Moreover, nearly 30% of DMLs were located at intergenic regions, whereas those located in promoter regions (TSS200-proximal and TSS1500-distal promoters) accounted for only 11.1% and 8.9% in hypo- and hyper-MLs, respectively. Thus, it is clear that the majority of the DMLs were located in potential regulatory regions (including introns, intergenic, and 3′ UTRs) rather than promoters. Six genomic locations of the hypo- and hyper-MLs relative to CGIs were also examined (Fig. 2*D*), including CGIs (island), N_shore and S_shore (up to 2 kb upstream and downstream of CGIs, respectively), N_Shelf and S_Shelf (2–4 kb upstream and downstream of CGIs, respectively), and open sea (remaining genomic regions). More than 60% of hypo- and hyper-MLs were in open sea and not related to CGIs or their nearby shores or shelves, whereas those located in CGIs accounted for only 11.8% and 6.7% in hypo- and hyper-MLs, respectively.

We then focused on DMLs located in RefSeq genes or their flanking regions. Although the pie charts in Fig. 2*C* show a high percentage of DMLs in gene bodies and intergenic regions, they did not reflect the location "density," because the genomic lengths of gene bodies and intergenic regions are much longer than other regions (including promoters), and the array design may have bias toward these regions. Thus, we calculated the percentage of hypo- and hyper-MLs among the CpGs covered on the array mapping to the same regions across gene bodies and their 5-kb flanking regions (Fig. 3*A* and *SI Appendix S1*, *SI Methods*). Compared with the overall percentage among all 213,620 CpG sites (0.07%) (Fig. 3*A*, dashed line), hypo-MLs were enriched in gene bodies, 3′ UTRs, and downstream intergenic regions (Fig. 3*A*, *Left*). The hyper-MLs (Fig. 3*A*, *Right*) were also enriched in similar regions plus in upstream intergenic regions. Both hypo- and hyper-MLs were not enriched in promoter regions as shown in Fig. 2*C*.

Next, we used bean plots to examine the distribution of DNA-me in cases and controls of hyper- and hypo-MLs located at different locations relative to RefSeq genes, including promoters (up to 1,500 bp upstream of TSS), 3′ UTRs, 5′ UTRs, gene bodies, and intergenic regions (Fig. 3*B*). The DNA-me levels at all of the CpG sites used in the comparison except DMLs (others in Fig. 3*B*) were analyzed as references. Excepting hyper-MLs in 3′ UTR, mean DNA-me levels of CpGs in hyper- and hypo-MLs were higher than others, especially those in the hypo-MLs (FC compared with others ranged from 2.46 to 4.29 in different regions, with P values ranging from 3.72×10^{-3} to 5.32×10^{-15}; t tests). The largest difference between cases and controls was at hypo-MLs in 3′ UTRs (FC = 1.67; $P = 1.68 \times 10^{-4}$; paired t test).

To explore the potential biological functions of these DMLs, Ingenuity Pathway Analysis (IPA) was conducted on the RefSeq genes containing either hypo- or hyper-MLs at their promoters (up to 1,500 bp upstream of TSS) or gene bodies. The ensuing top network (Fig. 3*C*) revealed that many of these genes had connections to NF-κB and PI3K-Akt kinase, which are well-known to be related to the development of diabetic complications (7, 42).

Identification and Characteristics of DMLs in Monos Genomic DNA Samples. Similar analyses were performed on Monos. After CpG site exclusions (*SI Appendix S1*, Fig. S1*B*), DNA-me levels at 205,639 CpG sites were compared between 31 cases and 30 controls

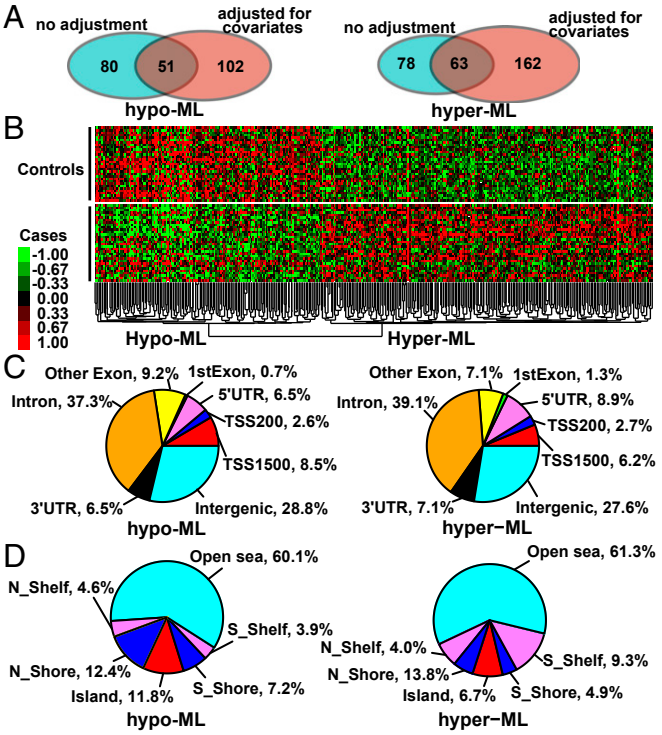

Fig. 2. DMLs between cases and controls identified in WB genomic DNAs collected at the EDIC Study baseline. DMLs with FC ≥ 1.3 and $P < 0.005$ in cases vs. controls were identified. (*A*) Venn diagrams depicting hypo- and hyper-MLs in cases with and without adjustment for covariates, including age, gender, WBC compositions, and complications. (*B*) Heat map of DMLs in cases vs. controls. DNA-me levels of each DML (rows) across all of the cases and controls (columns) after unsupervised hierarchical clustering analysis (*SI Appendix S1*, *SI Methods*) are shown. Green indicates DNA-me below the average of all of the samples, and red indicates DNA-me above the average level. (*C*) Genomic locations of (*Left*) hypo-ML and (*Right*) hyper-ML relative to RefSeq genes. Locations were classified to one of the following eight regions: TSS1500 (1,500 bp upstream to 200 bp upstream of TSS), TSS200 (200 bp upstream of TSS), 5′ UTR, the first exon (1st Exon), 3′ UTR, other exon (excluding the first exon), and intron and intergenic regions. (*D*) Locations of (*Left*) hypo-MLs and (*Right*) hyper-MLs in regions defined relative to CGI (Island). These regions include island, N_shore and S_shore (up to 2 kb upstream and downstream of CGI), N_Shelf and S_Shelf (2–4 kb upstream and downstream of CGI), and open sea (the remaining genomic regions). DMLs after adjustment for covariates were analyzed in *B*–*D*.

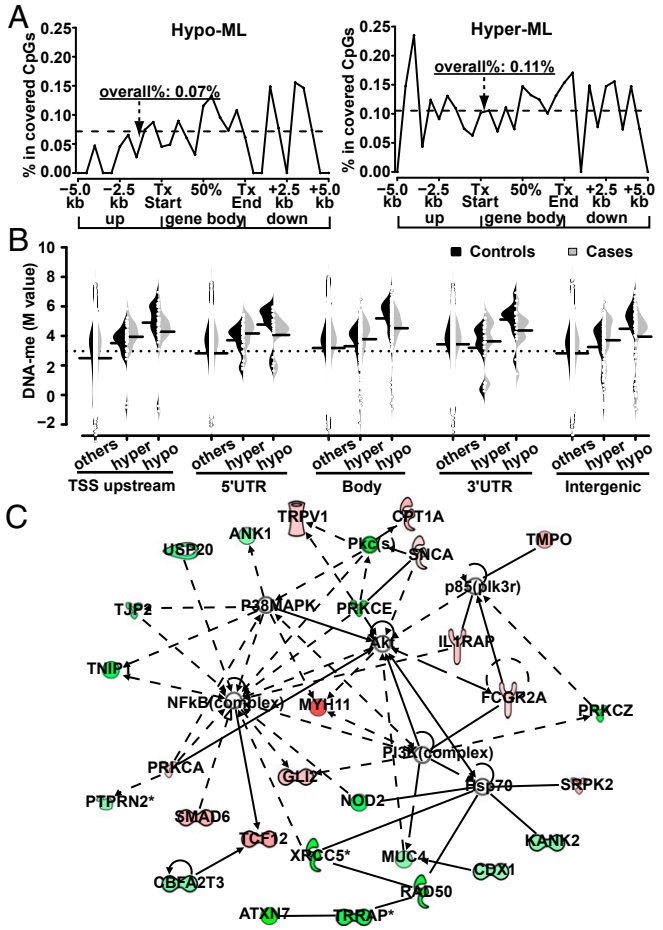

Fig. 3. DMLs located in RefSeq genes or their flanking regions in WB. (*A*) Percentage of hypo- and hyper-MLs in CpG sites covered in the array mapping to the same regions across RefSeq genes and their 5-kb upstream and downstream regions. Each CpG was annotated to one RefSeq transcript if it is located in 1 of 30 subregions that covered the gene body [from TSS (Tx Start) to end site (Tx End)] or its 5-kb flanking regions. The percentages of hypo- and hyper-MLs in CpG sites covered in the array at each subregion were calculated and plotted (*SI Appendix S1, SI Methods*). (*B*) Bean plots of DNA-me levels of three sets of CpGs including hypo-MLs (hypo), hyper-MLs (hyper), or all other CpGs after exclusion steps (others) in cases and controls. Based on CpG sites' location, each set was further classified into five groups as labeled at the bottom. Body represents the groups located in gene bodies but not in UTRs. The distribution of DNA-me levels (M values) of each group in cases and controls is presented by two-sided bean plots (left-side density plot in black is for controls, and right-side density plot in gray is for cases). The overall average of DNA-me for each group is depicted as horizontal short lines. (*C*) Top network identified among the genes containing DMLs in their promoters (up to 1,500 bp upstream of TSS) or gene bodies by IPA. Red nodes represent the genes containing hyper-MLs, whereas green ones represent the genes with hypo-MLs, and color intensity indicates FC in cases vs. controls. Solid lines indicate direct interactions, and dashed lines refer to indirect interactions.

using multiple linear regression models with and without adjustment for age and gender. As shown in Fig. 4*A*, 196 hypo- and 226 hyper-MLs were identified without any adjustment, whereas 155 hypo-MLs (Dataset S3) and 247 hyper-MLs (Dataset S4) were found after adjustment, with FC $\geq$ 1.3 and nominal $P < 0.005$. Comparing the lists with and without adjustment, about 80% (128 hypo-MLs and 175 hyper-MLs) were in common (Fig. 4*A*). The Monos DMLs after adjustment were used in the subsequent analyses.

Hierarchical cluster, location, and IPA network analyses of these DMLs revealed similar results to those found in WB (*SI Appendix S1*, Fig. S2), except that Monos hypo-MLs were enriched in promoters. DNA-me at hypo- and hyper-MLs in various regions of RefSeq genes (Fig. 4*B*) did not show the universal higher DNA-me levels in hyper- and hypo-MLs than others, which were found in WB DMLs. In Monos, only the hyper-MLs in gene bodies including 3′ and 5′ UTR showed higher than average DNA-me. Notably, again, the average DNA-me differences at DMLs in 3′ UTRs were the largest, highlighting the importance of DNA-me variation at 3′ UTRs.

Persistency of DNA-me Differences Between Cases and Controls from the EDIC Study Baseline to EDIC Study Years 16 and 17. To examine whether the DNA-me differences in cases vs. controls at one time point are also present 16–17 y later, the hyper- and hypo-MLs identified in WB and Monos after adjustment for covariates and annotated to a gene or its promoter were compared (Fig. 5). Notably, four hypo-MLs (Fig. 5*A*, bright green intersection of two solid green ellipses) and eight hyper-MLs (Fig. 5*C*, bright red intersection of two solid red ellipses) were similarly differentially methylated in WB and Monos. The DNA-me differences in cases vs. controls at each of the hypo-MLs [located in GATA zinc finger domain containing 1 (*GATAD1*), bassoon presynaptic cytomatrix protein (*BSN*), protein kinase C epsilon type (*PRKCE*), and thioredoxin-interacting protein (*TXNIP*)] and hyper-MLs [located in LIM/homeobox protein 6 (*LHX6*), microtubule-associated protein (*MAP7*), cadherin-3 (*CDH3*), SET and MYND domain containing 1 (*SMYD1*), chloride channel 7 alpha subunit (*CLCN7*, two loci), zinc finger protein 167 (*ZNF167*), and CUE domain containing 1 (*CUED1*)] in both sample sets are shown side by side in the heat map (Fig. 5*B*, *b*). For each of these 12 DMLs, Pearson's correlation analyses were applied to DNA-me between WB and Monos collected from each of the same 61 participants. Among these DMLs, five located at *PRKCE*, *BSN*, *CUEDC1*, and *CLCN7* were highly correlated with $r > 0.9$, whereas five located at *TXNIP*, *MAP7*, *LHX6*, *ZNF167*, and *CDH3* were moderately correlated ($0.56 < r < 0.79$) (Fig. 6). Of note, DNA-me at three DMLs (*PRKCE*, *BSN*, and *CUEDC1*) formed two or three distinct clusters on scatterplots (Fig. 6*A*), suggesting that DNA-me at these three CpGs was likely to be interfered by SNPs. Data from the 1000 Genomes Project Phase 3 (www.1000genomes.org) confirmed that all three contain common SNPs (minor allele frequency in European population > 1%) at the CpG sites (*SI Appendix S1*, Table S3). Thus, some but not all of the consistent DNA-me differences identified between the two time points may be caused by SNP variants.

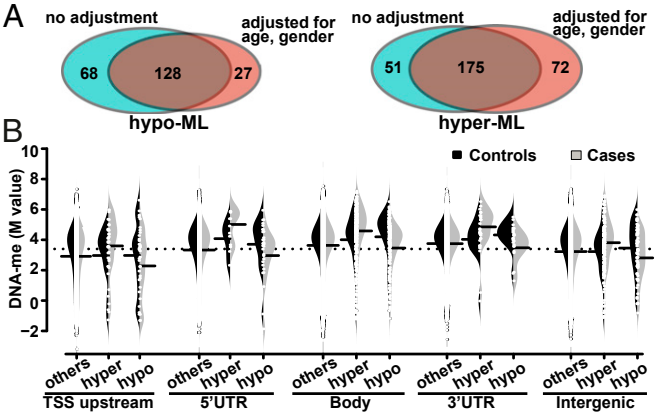

Fig. 4. DMLs between cases and controls identified in genomic DNAs isolated from Monos collected in the EDIC Study years 16 and 17. DMLs with FC $\geq$ 1.3 and $P < 0.005$ in cases vs. controls were identified by multiple linear regression models. (*A*) Venn diagrams depicting hypo- and hyper-MLs in cases identified with and without adjustment for age and gender. (*B*) Bean plots for DNA-me levels of all CpG sites except DMLs (others), hypo-MLs (hypo), and hyper-MLs (hyper) in cases and controls. Details are in Figs. 2 and 3.

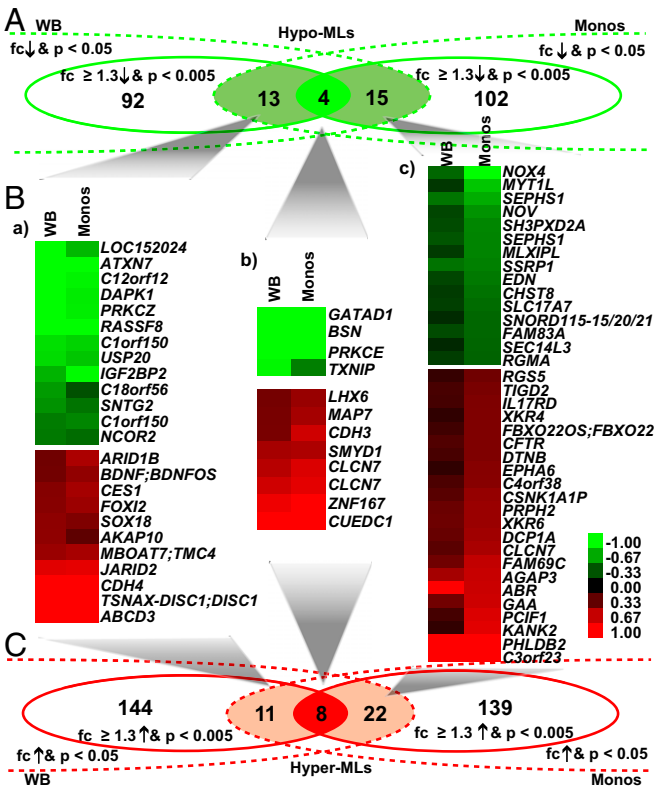

Fig. 5. Persistence of DNA-me over 16–17 y by comparing DMLs identified in WB (the EDIC Study baseline) with those in Monos (the EDIC Study years 16 and 17) and vice versa. (*A*) Venn diagram of hypo-MLs identified in WB and Monos. Only the hypo-MLs located in RefSeq gene bodies/promoters were counted. The two solid green ellipses represent hypo-MLs identified in WB and Monos, and the two dotted green half ellipses represent loci depicting trends of hypo-me in cases vs. controls at a lower confidence level ($P < 0.05$). The overlapping region (bright green) of the two solid lines depicts the hypo-MLs found in both WB and Monos, whereas the overlapping regions (light green) between the solid ellipse of one sample set and the dotted ellipse of the other sample set represent the hypo-MLs in one sample set (WB or Monos) with similar trends of hypo-me in the other sample set. Similar Venn diagrams in red were generated for the hyper-MLs in *C*. (*B*) The mean DNA-me differences between cases and controls at the hypo-MLs or hyper-MLs similarly modified in WB and Monos (represented in the overlapping regions of the Venn diagrams in *A* and *C*) are shown in heat maps *a*–*c*, with the corresponding annotated gene symbols depicted on the right side. Green is used to represent hypo-MLs, and red is used to represent hyper-MLs.

The relatively small overlap between WB and Monos could be because Monos form only a small fraction of WB (~10%) (*SI Appendix S1*, Fig. S1*C*). To address this issue, we next examined whether there are DMLs identified at one time point (either the EDIC Study baseline or years 16 and 17) showing similar trends of DNA-me difference at the other time point. These DMLs were defined as hypo-MLs identified in one sample set also displaying hypomethylation (hypo-me) in cases vs. controls with $P < 0.05$ in the other set (dotted green lines in Fig. 5*A*) or hyper-MLs in one sample set showing similar trends of hypermethylation in the other set (dotted red lines in Fig. 5*C*). Thus, 13 WB hypo-MLs (left intersection in pale green in Fig. 5*A*) and 11 WB hyper-MLs (left intersection in light red in Fig. 5*C*) were found to depict similar trends of hypo-me and hypermethylation in Monos, which are clearly shown by the heat map in Fig. 5*B*, *a*. Similarly, 15 Monos hypo-MLs and 22 Monos hyper-MLs depicted similar trends of DNA-me differences in WB (Fig. 5*B*, *c*).

Together, these data show that DNA-me at certain loci depicts persistence over 16–17 y in blood cells of cases vs. controls and thus, could play an important role in metabolic memory.

Validations of DNA-me Variations in Cases and Controls. Notably, among the DMLs identified in WB or Monos, a CpG site (probe cg19693031; CpG3) (Fig. 7*A*) located at chromosome 1: 145,441,552 (hg19) in 3′ UTR of *TXNIP* was hypomethylated in cases [$P = 2.4 \times 10^{-14}$; false discovery rate (FDR) = 1.2×10^{-8}] in WB. After adjustments for covariates, this hypo-me remained significant, with $P = 3.03 \times 10^{-11}$ and FDR = 1.45×10^{-5} (Fig. 7*B*, left comparison). Of 19 CpGs covered by 450K array in *TXNIP* and its promoter, this site was the only one showing DNA-me difference (*SI Appendix S1*, Fig. S3 *A* and *B*). Furthermore, this CpG3 showed similar hypo-me in Monos with an age- and gender-adjusted P value of 6.84×10^{-4} (Fig. 7*B*, right comparison).

Persistence of DNA hypo-me in *TXNIP* at two time periods was particularly interesting, because *TXNIP* is a reported sensor of glucose stress, a prooxidant, and a proapoptotic protein, with gene expression that is highly associated with hyperglycemia and diabetic complications, including retinopathy and nephropathy (43–45), both of which were prevalent in our case group. *TXNIP* expression is highly induced by HG in various cell types (46–48), which by inhibiting thioredoxin, subsequently causes oxidative stress and apoptosis. Recently, *TXNIP* up-regulation in the diabetic kidney was reported to be associated with changes in certain histone PTMs (49). Because hypo-me at *TXNIP* could likely lead to increased expression in cases vs. controls and thus, serve as an epigenetic mechanism for its dysregulation and metabolic memory, we further validated DNA-me at CpG3.

We first performed DNA bisulfite pyrosequencing in WB of 30 cases and 30 controls and confirmed the hypo-me identified at CpG3 (Fig. 7*A*) by 450K array. The average DNA-me was 68.3% in controls vs. 52.2% in cases ($P = 1.95 \times 10^{-10}$) (Fig. 7*C*). Interestingly, similar to CpG3, its two adjacent CpG sites covered by the PCR-amplified DNA bisulfite fragments for pyrosequencing (CpG1-2, Fig. 7*A*) also depicted significant hypo-me in cases vs. controls (Fig. 7*C* and *SI Appendix S1*, Fig. S3*B*), and their DNA-me levels (including CpG3) were significantly correlated within a person (*SI Appendix S1*, Fig. S3*C*). These results show that WB DNAs exhibit significant hypo-me at the *TXNIP* 3′ UTR in cases vs. controls.

To expand the scale of validation to several DMLs, we adopted a targeted DNA bisulfite sequencing approach (amplicon-seq) that directly sequences the PCR products of bisulfite-treated DNA to target multiple candidate regions of interest from numerous subjects simultaneously using next generation sequencing (*SI Appendix S1*, *SI Methods*). DNA from 18 cases and 18 controls was randomly selected for validation. We first examined three CpG sites at *TXNIP* 3′ UTR in WB samples as a positive control, and all three exhibited significant hypo-me (Fig. 7*D*), thus verifying amplicon-seq as an efficient method to validate DNA-me variations in multiple subjects. We further examined DNA-me at these three CpGs in both Monos and lymphocytes (key components of WBCs) from the same 36 participants (collected in the EDIC Study years 16 and 17) and validated that, similar to WB, they exhibited significant hypo-me (cases vs. controls) in both cell types (Fig. 7 *E* and *F*).

Analyzing the agreement between DNA-me measured by 450K array vs. validation platform within these participants, high Spearman coefficient (ρ) was shown at 0.93 ($P < 2.2 \times 10^{-16}$) between pyrosequencing and array or 0.76 ($P < 1.4 \times 10^{-6}$) between amplicon-seq and 450K array in WB and 0.85 ($P < 6.7 \times 10^{-8}$) between amplicon-seq and 450K array in Monos. Bland–Altman plots (*SI Appendix S1*, Fig. S4) also showed only small differences between the two platforms in both WB and Monos. Together, these results clearly show a persistence (memory) of the *TXNIP* hypo-me in cases over 16–17 y in multiple WBCs.

We next used amplicon-seq to validate additional CpGs (Table 2) located in *CLCN7*, cytochrome B5 type B (*CYB5B*), *PRKCE*, membrane bound O-acyltransferase domain containing 7 (*MBOAT7*), and long intergenic nonprotein coding RNA 649 (*LINC00649*) and confirmed their differential DNA-me ($P <$ 0.05; Wilcoxon rank-sum test) between cases vs. controls (36 samples for each DML) in WB and/or Monos, with small

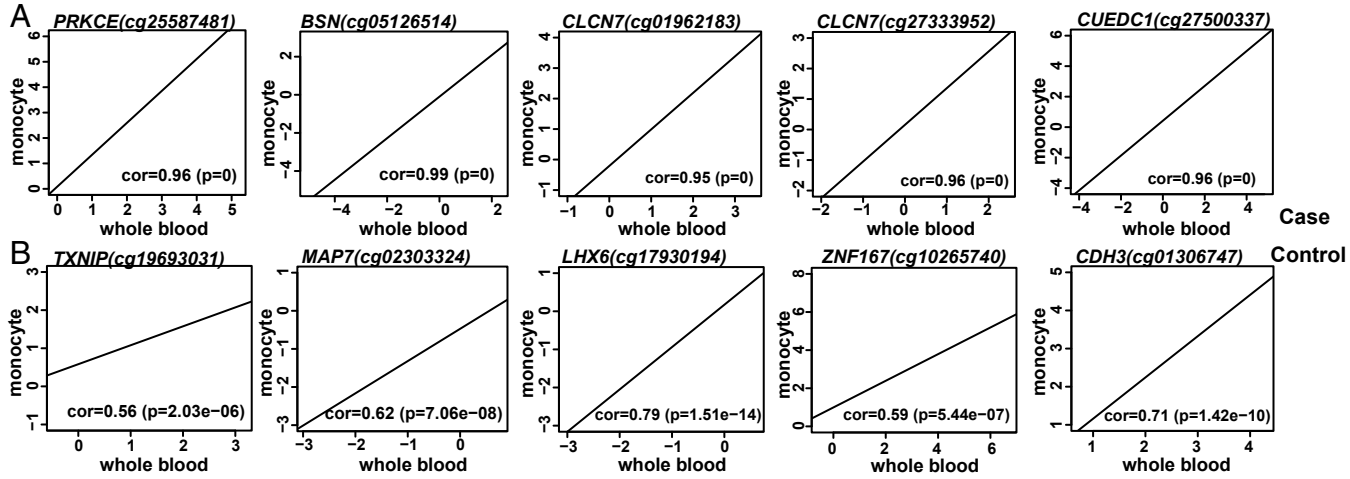

Fig. 6. Correlations on DNA-me levels at DMLs identified in both WB and Monos. The Pearson correlations of DNA-me at 12 annotated DMLs identified in both WB and Monos DNA samples were analyzed between the two sample sets across 61 participants. Among them, 10 were significantly correlated, including (*A*) 5 highly correlated (correlation coefficients; $r > 0.9$) and (*B*) 5 moderately correlated ($0.5 < r < 0.8$). For each of these DMLs, scatterplots generated with DNA-me (M values) of WB are presented on the *y* axis, and those generated with DNA-me of Monos are presented on the *x* axis. The Infinium probe identification and its annotated gene are shown on the top of each plot. Coefficients with *P* values are listed at the bottom of each box. Cases are presented as red dots, and controls are presented as blue dots.

differences observed between 450K array and amplicon-seq (*SI Appendix S1*, Fig. S4). These results also validated similar variations at DMLs around not only *TXNIP* but also, *CLCN7* and *PRKCE* in both WB and Monos.

It should be noted that, because of the large number of CpG sites analyzed (>0.2 million) and relatively low sample size (32 cases vs. 31 controls), identification of DMLs using FDR to control type I error would be too stringent and cause high type II error (high false negatives). To avoid missing any true positives, the criteria for DML identification (FC $\geq$ 1.3; nominal $P <$ 0.005) were empirically chosen based on these technical validations on a different platform. Along with *TXNIP*, these results show that the DNA-me differences between cases and controls by 450K array are real and can be validated within this cohort. They also uncover potential roles for DNA-me at specific sites important in diabetes complications for additional investigations.

Association of DNA-me with HbA1c at WB or Monos DMLs. We next examined the associations between glycemia (HbA1c) and epigenetic DNA-me. For each DCCT/EDIC Study participant, HbA1c was measured periodically at and after entry into the study as described (31). To facilitate comparisons of the associations between HbA1c and DNA-me at different time periods, mean HbA1c levels over three time periods (the DCCT mean, the EDIC Study mean, and pre-DCCT/DCCT/EDIC Study combined mean) and at two specific time points (the EDIC Study baseline, when WB samples were collected, and the EDIC Study years 16 and 17, when Monos were collected) were used for each subject in the analysis (*SI Appendix S1*, Fig. S5*A*). For each DML identified in WB or Monos, Spearman correlations between HbA1c over these different time periods and DNA-me levels in both WB and Monos were assessed. Because only the time periods earlier than the sample collection would be used, HbA1c levels at two time periods (the DCCT mean and the EDIC Study baseline) were used in the correlation analysis with WB DNA-me, whereas HbA1c levels at all of the five time periods were included in the analysis of Monos DNA-me. Thus, two matrices of Spearman correlation coefficients were generated for DMLs identified in WB or Monos, which were then subjected to unsupervised hierarchical clustering analysis and are shown as heat maps in Fig. 8.

For WB DMLs (Fig. 8*A*, *Left*), we observed a generally negative correlation of DNA-me in WB with HbA1c of two time periods at hypo-MLs (yellow in Fig. 8*A*, *Center*) and a positive correlation at hyper-MLs (blue in Fig. 8*A*, *Center*). This result was as expected because of the differences in HbA1c between cases and controls by design (*SI Appendix S1*, Table S1). However, interestingly, some of these DMLs also show similar trends of associations with HbA1c in Monos at these WB DMLs (Fig. 8*A*, *Right*). Specifically, negative correlations were observed in Monos at some WB hypo-MLs (upper red dashed box in Fig. 8*A*), and positive correlations were observed in Monos at WB hyper-MLs (lower red dashed box in Fig. 8*A*). Similarly, for Monos DMLs (Fig. 8*B*, *Left*), as expected, negative or positive associations with HbA1c can also be observed in Monos hypo- and hyper-MLs (Fig. 8*B*, *Center*). Notably, similar to our previous observations of associations between H3K9Ac and HbA1c (31), the associations of DNA-me with long-term HbA1c (the DCCT mean, the EDIC Study mean, or the pre-DCCT/DCCT/EDIC Study mean) were higher than the associations with single HbA1c measured at the EDIC Study years 16 and 17. Furthermore, some Monos DMLs were also found to have similar trends of associations with HbA1c in WB as seen in Monos (red dashed boxes in Fig. 8*B*). We further identified 50 WB DMLs (*SI Appendix S1*, Fig. S5*B*) or 52 Monos DMLs (*SI Appendix S1*, Fig. S5*C*) depicting similar trends of association between DNA-me and HbA1c in the other sample set that showed association ($P < 0.05$) with HbA1c in at least one time period in the other set.

These data suggest persistence of the association between DNA-me and HbA1c at certain DMLs over 16 y and again, support our speculation that DNA-me differences in cases vs. controls at these DMLs may be introduced by HG during or even before the DCCT and may play an important role in complications development/progression in cases during the EDIC Study.

Induction of *TXNIP* by HG Is Correlated to *TXNIP* 3′ UTR Hypo-me in THP1 Cells. We next addressed the potential underlying mechanism by which these persistent and HbA1c-associated DNA-me differences in cases vs. controls, might be involved in diabetic complication progression. For this question, we specifically studied the most significant DML and its nearby CpGs located in *TXNIP* 3′ UTR, due to its significant associations with oxidative stress, and diabetic complications.

Because only limited RNA from the patients was available, to understand the relationships between DNA-me, gene expression, and effect of hyperglycemic exposure, and to explore the potential links between hypo-me at *TXNIP* 3′ UTR and complication progression, we performed mechanistic in vitro

Table 2. Validation of the DNA-me differences between EDIC Study case and control groups using amplicon-seq

Gene	Cell	CpG location*	Case mean (%)	Control mean (%)	Difference (%)	*P* value†
CYB5B	WB	16:69,457,645	91.8	94.7	−2.9	7.9E-03
CYB5B	WB	16:69,457,652	88.7	91.5	−2.8	4.7E-02
CYB5B	WB	16:69,457,738	86.1	89.1	−3.0	1.4E-02
TXNIP	WB	1:145,441,517	55.5	72.4	−16.9	9.2E-05
TXNIP	WB	1:145,441,526	56.2	72.7	−16.5	1.3E-04
TXNIP	WB	1:145,441,552	58.8	69.7	−10.9	3.3E-05
LINC00649	WB	21:35,320,667	41.0	48.4	−7.4	1.6E-03
LINC00649	WB	21:35,320,677	20.4	28.2	−7.9	1.6E-03
MBOAT7	WB	19:54,677,166	57.7	44.6	13.1	6.6E-03
CLCN7	WB	16:1,510,620	76.7	60.9	15.8	4.4E-04
CLCN7	WB	16:1,510,813	54.1	43.4	10.8	1.3E-02
CLCN7	WB	16:1,510,849	71.1	64.8	6.3	3.2E-02
CLCN7	WB	16:1,510,879	54.6	43.8	10.7	4.4E-02
CLCN7	WB	16:1,510,897	32.3	25.0	7.3	3.4E-02
CLCN7	WB	16:1,510,922	80.2	73.8	6.4	3.7E-02
PRKCE	WB	2:46,396,825	73.8	87.8	−14.0	3.2E-02
PRKCE	Mo	2:46,396,825	71.8	93.4	−21.6	1.1E-02
CLCN7	Mo	16:1,509,221	62.6	53.2	9.5	4.4E-02
CLCN7	Mo	16:1,509,206	58.2	46.3	12.0	2.7E-02
CLCN7	Mo	16:1,507,865	92.8	83.3	9.5	5.2E-02
CLCN7	Mo	16:1,507,958	23.3	15.8	7.4	1.8E-03
TXNIP	Mo	1:145,441,517	60.1	73.1	−12.0	2.9E-05
TXNIP	Mo	1:145,441,526	61.7	73.7	−12.0	2.4E-05
TXNIP	Mo	1:145,441,552	60.3	67.8	−7.5	2.8E-04

Mo, Monos.
*Chromosome:genomic location.
†Wilcoxon rank-sum test.

experiments with human THP1 Monos grown under various glycemic conditions: NG (5.5 mM), HG (25 mM), reversal, and restimulation with HG. Briefly, experiments were performed in three steps. (*i*) Cells were cultured under NG or HG for 3 d (a and b, respectively, in Fig. 9). (*ii*) We continued to culture NG cells (from step 1) in NG (c in Fig. 9), whereas the cells in HG (from step 1) were switched to NG (d in Fig. 9), and all cells were cultured for 4 more d. (*iii*) Cells in NG (from steps 1 and 2) were maintained in NG (e in Fig. 9), whereas those switched from HG to NG (at step 2) were returned to HG (f in Fig. 9), and all cells were cultured for 3 more d. Total RNA and genomic DNAs were then prepared from these six samples and subjected to RT-PCR for measuring *TXNIP* expression (Fig. 9*A*) and amplicon-seq for DNA-me at the three CpG sites (CpG1–3) in *TXNIP* 3′ UTR (Fig. 9 *B*–*D*). Interestingly, Fig. 9*A* showed that HG treatment markedly increased *TXNIP* expression (b vs. a in Fig. 9), which was clearly attenuated after return to NG (d vs. c in Fig. 9). Subsequent switch and restimulation with HG produced an even more dramatic increase in expression (f vs. e and f vs. b in Fig. 9). Notably, with respect to DNA-me on the same samples, HG treatment decreased DNA-me (hypo-me) at all three CpG sites (b vs. a in Fig. 9) from 95.1% to 86.3% (Fig. 9*B*), from 95.0% to 86.6% (Fig. 9*C*), and from 90.8% to 82.0% (Fig. 9*D*), respectively. Interestingly, this HG-induced hypo-me remained at 89.5% (Fig. 9*B*), 89.8% (Fig. 9*C*), and 86.7% (Fig. 9*D*), respectively, in sample d when the cells were switched from HG to NG, although gene expression returned to normal under the same condition. Moreover, this HG-induced hypo-me continued to persist when cells were recultured in HG at step 3, this time with greater magnitude at all three CpGs (85.9%, 84.6.1%, and 82.8%, respectively) (f vs. e in Fig. 9 *B*–*D*), and this hypo-me inversely correlated with the markedly augmented increase in gene expression under this condition (Fig. 9*A*, f vs. e).

Overall, these results show a differential response between *TXNIP* expression and DNA-me variations induced under the same NG/HG conditions. *TXNIP* expression is dynamic and changes with environmental stimuli (HG), whereas the loss of DNA-me triggered by the initial HG stimulus does not recover to the original level, even when cells were returned to NG, at least within the timeframe studied. Instead, DNA-me seems to create poised and persistent states through additional reduction in DNA-me, which can sensitize cells and augment gene expression response when restimulated with HG. These data can explain, at least in part, the connections between DNA-me and the metabolic memory observed in the EDIC Study, wherein progression of complications in the CONV group continued at a greater rate than that in INT, despite similar glycemic control during the EDIC Study.

Discussion

It is imperative to evaluate the mechanisms underlying metabolic memory to reduce morbidity/mortality in the diabetic population and also, identify new therapeutic targets for complications that progress, despite improved glycemic control. We examined the DNA methylome of two groups of subjects with T1D from the DCCT/EDIC Study at two different time points, namely the EDIC Study baseline and the EDIC Study years 16 and 17. The case group included former DCCT CONV participants who experienced progression of complications compared with the control group, which included former DCCT INT participants with minimal complication development. The objective was to gain information about persistence of epigenetic differences between the two groups at two time points to determine potential associations between DNA-me and metabolic memory. During the identification of DMLs in cases vs. controls, several confounding variables were adjusted for in the comparisons. For WB samples (EDIC Study baseline), these variables included age, gender, and cell type compositions. Because some of the cases already had developed complications (proliferative diabetic retinopathy and microalbuminuria) at the EDIC Study baseline, the complication status was also adjusted to avoid DMLs caused by complications, although this adjustment might reduce the power. Hence, we expected that the DMLs identified in WB would be mainly associated with the history of high blood glucose in cases vs. controls at that time, because glycemic separation was the goal during the DCCT (*SI Appendix S1*, Table S1). This relationship is also shown by the association of DNA-me with HbA1c history. Furthermore, because complications progressed in all of the cases during EDIC but not in any of the controls, the WB DMLs can also be associated with future complications progression during the EDIC Study. However, the DNA-me differences between cases and controls in Monos collected at the EDIC Study years 16 and 17 should reflect the effects of either history of blood glucose differences (if DNA-me changes persist; discussed below), consequences of diabetic complications (because cases but not controls experienced complications progression during the EDIC Study), or both. Using pyrosequencing and amplicon-seq, we validated several DMLs in both WB and Monos (Table 2), supporting our criteria for DML identification.

IPA of DMLs identified in both WB and Monos revealed strong connections with networks associated with diabetic complications. The locations of these DMLs were also extensively analyzed. Major changes in DNA-me distribution at promoters are a hallmark of cancer (13, 16). However, in chronic diseases, such as diabetes and its complications, increasing evidence suggests the importance of DNA-me in regulatory regions, including putative enhancer regions in kidney tissues from patients with diabetic kidney disease (50), intronic enhancers in blood cells type 2 diabetes patients (51), and gene bodies/intergenic regions rather than promoters in adipose tissues from twins (52). The high percentage of DMLs in intergenic regions, introns, and 3′ UTRs in our study is in line with such reports. Furthermore, using ENCODE Monos ChIP-seq data on Histone H3-Lysine4-monomethylation (a well-accepted mark of enhancers), 64 (15.92%) of 402 Monos

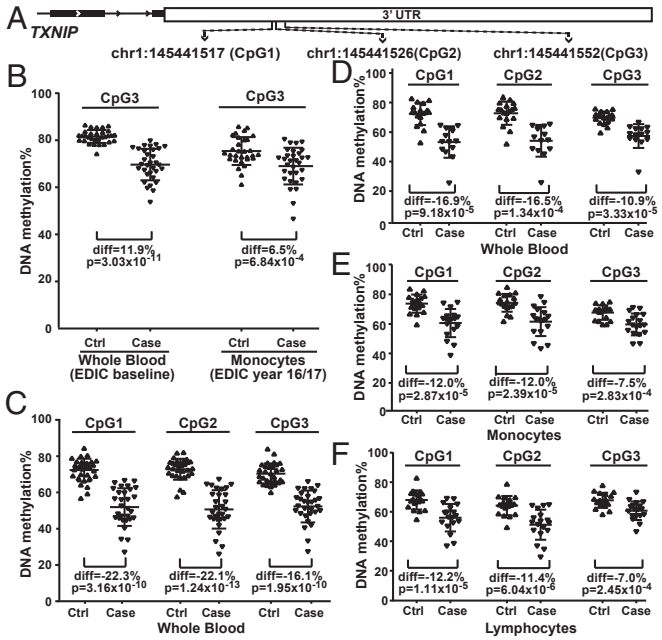

Fig. 7. Hypo-me at the *TXNIP* 3′ UTR and its validations in WB, Monos, and lymphocytes. (*A*) DNA-me at three CpG sites in 3′ UTR of *TXNIP* named as CpG1 (chromosome 1: 145,441,517), CpG2 (chromosome 1: 145,441,526), and CpG3 (chromosome 1: 145,441,552; also, the DML identified at CG19693031 in Infinium 450K array) were measured. (*B*) Highly statistically significant hypo-me in WB (32 cases vs. 31 controls) and Monos (31 cases vs. 30 controls) at CpG3 by 450K array. Covariates-adjusted *P* values by multiple linear regressions are presented. (*C*) Validations by pyrosequencing of hypo-me at CpG1–3 in WB (30 cases vs. 30 controls). (*D*–*F*) Validations by amplicon-seq of hypo-me (16 cases vs. 16 controls) at the CpG1–3 in (*D*) WB, (*E*) Monos, and (*F*) lymphocytes as indicated at the bottom of each plot. Wilcoxon rank-sum tests on β-values were performed on validation data. Mean DNA-me difference in β-values (diff) and *P* values for each comparison are depicted at the bottom. Details of the validation methods were described in *SI Appendix S1, SI Methods*. Ctrl, control.

DMLs, including CpG3 in *TXNIP*, were located in Histone H3-Lysine4-monomethylation-enriched regions, which are potential regulatory regions (53). However, the functions of DNA-me at these regions and how they might contribute to gene expression are not yet clear. Only ∼10% of DMLs were located in CGI, whereas the majority were in open seas in both WB and Monos, consistent with findings in WB from monozygotic twins (54), β-cells of type 2 diabetes (55), and WB of chronic kidney disease patients (56). This result is in the contrast to high enrichment in CGIs and their shores in cancer (13, 16). Together, these data suggest functional differences in site-specific DNA-me in different diseases.

We profiled DNAs from the same DCCT/EDIC Study patients collected 16–17 y apart, and comparison of both sample sets allowed us to identify 14 common DMLs (12 annotated). Among these DMLs, apart from DMLs possibly associated with genetic variations, DNA-me differences in cases vs. controls were likely associated with the differences in their history of blood glucose levels during the DCCT. The persistence of DNA-me variations at the two time points was validated at *TXNIP* (hypo-me) and *CLCN7* (hypermethylation) in both sample sets by pyrosequencing and/or amplicon-seq. Because Monos constitute only about 10% of blood cells (*SI Appendix S1*, Fig. S1*C*), it is possible we could have identified more common DMLs (persistency) at two time points if we had compared the same cell type; however, samples were unavailable. We therefore performed analyses using less stringent criteria, which resulted in the identification of another 28 annotated hypo-MLs and 33 annotated hyper-MLs in WB or Monos where the same trends of DNA-me changes were present in the other sample set (Fig. 5). We also identified over 50 WB DMLs and 52 Monos DMLs that showed trends of association with HbA1c history and had DNA-me levels that depicted similar trends of association with HbA1c in the other sample set (Mono and WB, respectively) (*SI Appendix S1*, Fig. S5). In summary, this study allowed us to uncover the persistency of DNA-me variations over several years that may be associated with metabolic memory and the progression of diabetes complications.

Among the common DMLs identified in both WB and Monos, hypo-me in cases vs. controls at *TXNIP* 3′ UTR was highly significant in WB, suggesting potential association with prior history of HG at the EDIC Study baseline. This finding was further validated in WB by pyrosequencing and amplicon-seq. Interestingly, association of *TXNIP* hypo-me at the same CpG3 with type 2 diabetes and related traits (including fasting glucose, HbA1C, and insulin resistance) (57–60), serum metabolites related to diabetes (61), and high blood triglycerides (62) in WB DNA from various cohorts was recently reported. Because *Txnip* was associated with lipids in a mouse model (63), we reanalyzed methylation differences in cases vs. controls at CpG3 after adding triglycerides at the EDIC Study baseline as a covariate. The resulting hypo-me ($\log_2$ FC = −0.95) remained highly significant ($P = 2.46 \times 10^{-9}$) in WB, indicating that blood glucose (HbA1c) is still a significant factor associated with

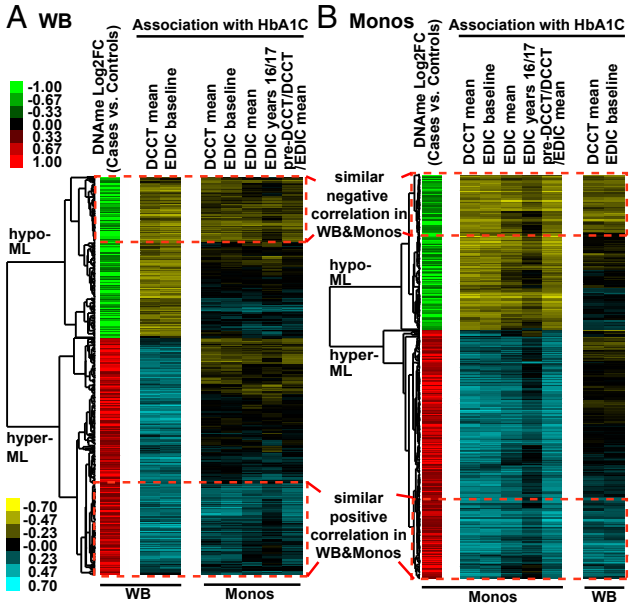

Fig. 8. Association of DNA-me with patient HbA1c at DMLs identified in WB and Monos. (*A*) The associations between DNA-me and HbA1c levels at the indicated times periods of the DCCT/EDIC Study at WB DMLs. For each DML, Spearman correlation was calculated between its DNA-me levels and mean HbA1c levels at different time periods across all patients in both WB and Monos (only the time periods that are earlier than the sample collection are shown). The HbA1c values at different time periods for each participant are presented by five values: the DCCT mean, the EDIC Study baseline, the EDIC Study mean, the EDIC Study years 16 and 17, and pre-DCCT/DCCT/EDIC Study mean (details are in *SI Appendix S1*, Fig. S5*A*). The matrix of Spearman correlation coefficients (ρ) was subjected to unsupervised clustering and is shown as heat maps, in which each row represents one DML and each column represents ρ at a specified time period as shown at the top of the heat map. Yellow represents negative correlation, and blue represents positive correlation. Mean DNA-me differences between cases and controls in WB are shown in *Left*, with green for hypo-MLs and red for hyper-MLs. (*B*) A similar figure was generated for Monos DMLs. The red dashed boxed areas highlight the DMLs depicting similar trends of association between DNA-me and HbA1c in both WB and Monos.

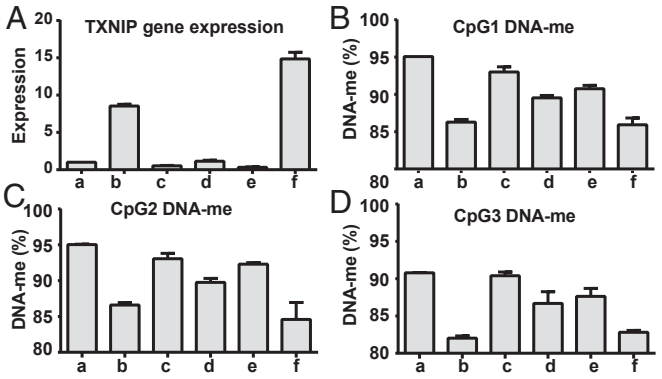

Fig. 9. *TXNIP* expression and DNA-me levels at the three CpG sites in *TXNIP* 3′ UTR induced by HG in cultured THP-1 Monos. *TXNIP* expression and DNA-me at three *TXNIP* 3′ UTRs in six samples were measured by RT-PCR and amplicon-seq, respectively. The six THP1 samples labeled a–f are a, cultured in NG for 72 h; b, cultured in HG for 72 h; c, sample a retained in NG for 96 more h; d, sample b switched from HG to NG and cultured in NG for 96 more h; e, sample c continued to be cultured in NG for 72 more h; and f, sample d switched from NG to HG and kept in HG for 72 more h. (*A*) *TXNIP* expression measured by RT-PCR in samples a–f. Details of RT-PCR are described in *SI Appendix S1, SI Methods*. Data shown are the averages of samples run in triplicates. Statistically significant increase in *TXNIP* expression ($P < 0.0001$; *t* test) was found in b vs. a and f vs. e. (*B–D*) DNA-me levels at three CpG sites (CpG1–3), respectively, in *TXNIP* 3′ UTR were obtained by amplicon-seq (details are in *SI Appendix S1, SI Methods*). The DNA-me levels (β-values) at each CpG site are the average from two separate experiments generated based on at least 10,000 sequences obtained by amplicon-seq in each of two experiments.

TXNIP hypo-me in our cohort. Moreover, the same trend of associations in our WB dataset (*SI Appendix S1*, Table S4) at all five top blood glucose-associated CpG sites (including *TXNIP* CpG3) previously reported (57) also indicates the reliability of our data. Notably, whereas these previous findings were in WB, in our study, the case *TXNIP* hypo-me was further validated in specific white cell types, namely CD14+ Monos and lymphocytes, which is important because of the cell type-specific nature of DNA-me. These findings also suggest potential common mechanisms of inducing DNA-me changes at specific loci (such as activation of TETs) in different cell types in peripheral blood or hematopoietic stem cells, which continue to divide and differentiate into various blood cells.

The functions of 3′ UTR DNA-me are not fully clear. A few studies showed that 3′ UTRs DNA-me is related to gene expression, with either positive or negative correlations (64–66), thus supporting connections between *TXNIP* hypo-me and its gene expression. Importantly, our experiments with THP1 cells cultured under various HG/NG conditions revealed an association between persistent DNA-me triggered by prior history of HG exposure and enhanced *TXNIP* expression in response to HG in vitro. We, thus, speculate that transient hyperglycemic episodes during the EDIC Study would induce more *TXNIP* overexpression and cellular dysfunction in cases vs. controls because of *TXNIP* hypo-me, which is in line with the reported adverse cellular roles of TXNIP (43–45, 48). This type of cellular response to episodic HG provides mechanistic information toward understanding the connections between persistent DNA hypo-me at *TXNIP* and its gene expression and hence, metabolic memory of diabetic complications.

We noticed that 3 of 12 annotated DMLs identified in both WB and Monos contain a common SNP at their CpGs (*SI Appendix S1*, Table S3), which might account for the observed DNA-me persistency at these three sites. To explore DNA-me changes possibly associated with genetic variations, we expanded our search of common SNPs (minor allele frequency in European population > 1%; 1000 Genomes Project Phase 3) near these DMLs. We found that 23 (6.08%) WB DMLs and 14 (3.48%) Monos DMLs contain common SNPs at CpG sites, whereas 81 (21.43%) WB DMLs and 77 (19.15%) Monos DMLs contain common SNPs in the probes. However, not all SNPs in CpGs or probes would result in DNA-me changes. By examining DMLs where the DNA-me levels can be classified into two to four distinct clusters (implying possible interference by nearby SNPs) (*SI Appendix S1, SI Methods*), 5.31% WB DMLs (*SI Appendix S1*, Fig. S6*A*) and 3.50% Monos DMLs (*SI Appendix S1*, Fig. S6*B*) were identified. The majority of them (16 of 19 in WB and 13 of 14 in Monos) have common SNPs at the CpG sites, which is in line with a previous report that genetic variations more distal to CpGs are unlikely to influence DNA-me at the site (57). Hence, we would not expect many more of our DMLs to be associated with genetic variation than the estimated rate in our cohort. Detailed information related to SNPs is included in Datasets S1 and S2 for WB and Datasets S3 and S4 for Monos.

As mentioned previously (31), cases and controls also differed in other clinical characteristics, such as blood pressure and lipids (*SI Appendix S1*, Table S1), which may have contributed to some of the observed DMLs between cases and controls. Because of the relatively small sample size, we did not adjust for these additional factors. In addition, DMLs not covered by the 450K array may have been missed. DNA-me profiling of a bigger DCCT/EDIC Study cohort using high-throughput sequencing can address these issues. It is clearly ideal to study DNA-me related to diabetic complications in relevant affected tissues. However, quite often, obtaining such tissues requires invasive approaches, and there is a paucity of such tissues. Because most diabetic complications are associated with inflammation, WBCs, including inflammatory Monos, which exhibit similar DNA-me changes as those in specific tissues, are good and easily accessible "surrogates" (56).

This study provides the first evidence, to our knowledge, of persistence of differential methylation at several loci over more than 16–17 y in the same cohort. Based on our data, DNA-me at key DMLs, especially 3′ UTR of *TXNIP*, in peripheral blood cells can be used as a biomarker for glycemia and metabolic memory along with other epigenetic factors and nonepigenetic factors, like oxidative stress. These findings together with our previous data on promoter histone PTMs support an epigenetic explanation for metabolic memory.

Methods

The human study protocols were approved by the Institutional Review Board at the City of Hope Medical Center and the participating EDIC clinics. Genomic DNA samples from WB and Monos were used to profile DNA-me using Infinium Human Methylation 450 Bead Chip Arrays (WG-314-1002). Multiple linear regressions with or without adjustment for covariates were performed to identify DMLs between cases and controls. Cell composition of each sample was estimated by deconvolution (41). Validation of DNA-me at specific sites was performed by pyrosequencing or amplicon-seq. Details are in *SI Appendix S1, SI Methods*.

ACKNOWLEDGMENTS. We thank the EDIC executive committee and study group for helpful discussions and all study participants for their generosity and interest. We thank Gayle Lorenzi (University of San Diego), Dr. Rose Gubitosi-Klug (Case Western University), Paddy Cleary (The George Washington University), and all of the EDIC coordinators for their valuable assistance. Research reported in this publication included work performed in the Integrative Genomics Core at the City of Hope supported by National Cancer Institute of the NIH Grant P30CA33572. This work was supported by Juvenile Diabetes Research Foundation Grants JDRF17-2008-900 and 17-2012-480 and NIH, National Institute of Diabetes and Digestive and Kidney Disease (NIDDK) Grants DP3 DK106917 (to R.N.), R01 DK065073 (to R.N.), and R01 DK081705 (to R.N.). The DCCT/EDIC has been supported by NIDDK cooperative agreement grants (1982–1993 and 2012–2017) and contracts (1982–2012; currently, Grants U01 DK094176 and U01 DK094157), the National Eye Institute, the National Institute of Neurologic Disorders and Stroke, the General Clinical Research Centers Program (1993–2007), and the Clinical Translational Science Center Program (2006 to present).

1. Writing Team for the Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications Research Group (2002) Effect of intensive therapy on the microvascular complications of type 1 diabetes mellitus. *JAMA* 287(19):2563–2569.
2. DCCT/EDIC research group (2014) Effect of intensive diabetes treatment on albuminuria in type 1 diabetes: Long-term follow-up of the Diabetes Control and Complications Trial and Epidemiology of Diabetes Interventions and Complications study. *Lancet Diabetes Endocrinol* 2(10):793–800.
3. Nathan DM; DCCT/EDIC Research Group (2014) The diabetes control and complications trial/epidemiology of diabetes interventions and complications study at 30 years: Overview. *Diabetes Care* 37(1):9–16.
4. Villeneuve LM, Natarajan R (2010) The role of epigenetics in the pathology of diabetic complications. *Am J Physiol Renal Physiol* 299(1):F14–F25.
5. Ihnat MA, Thorpe JE, Ceriello A (2007) Hypothesis: The 'metabolic memory', the new challenge of diabetes. *Diabet Med* 24(6):582–586.
6. Pirola L, Balcerczyk A, Okabe J, El-Osta A (2010) Epigenetic phenomena linked to diabetic complications. *Nat Rev Endocrinol* 6(12):665–675.
7. Reddy MA, Natarajan R (2011) Epigenetic mechanisms in diabetic vascular complications. *Cardiovasc Res* 90(3):421–429.
8. Cooper ME, El-Osta A (2010) Epigenetics: Mechanisms and implications for diabetic complications. *Circ Res* 107(12):1403–1413.
9. Kowluru RA, Santos JM, Mishra M (2013) Epigenetic modifications and diabetic retinopathy. *BioMed Res Int* 2013:635284.
10. Jirtle RL, Skinner MK (2007) Environmental epigenomics and disease susceptibility. *Nat Rev Genet* 8(4):253–262.
11. Chen ZX, Riggs AD (2011) DNA methylation and demethylation in mammals. *J Biol Chem* 286(21):18347–18353.
12. Strahl BD, Allis CD (2000) The language of covalent histone modifications. *Nature* 403(6765):41–45.
13. Jones PA (2012) Functions of DNA methylation: Islands, start sites, gene bodies and beyond. *Nat Rev Genet* 13(7):484–492.
14. Kouzarides T (2007) Chromatin modifications and their function. *Cell* 128(4):693–705.
15. Berger SL (2007) The complex language of chromatin regulation during transcription. *Nature* 447(7143):407–412.
16. Shen H, Laird PW (2013) Interplay between the cancer genome and epigenome. *Cell* 153(1):38–55.
17. Riggs AD (1989) DNA methylation and cell memory. *Cell Biophys* 15(1-2):1–13.
18. Smith ZD, Meissner A (2013) DNA methylation: Roles in mammalian development. *Nat Rev Genet* 14(3):204–220.
19. Tompkins JD, et al. (2016) Mapping human pluripotent-to-cardiomyocyte differentiation: Methylomes, transcriptomes, and exon DNA methylation "memories." *EBioMedicine* 4(2016):74–85.
20. Miao F, et al. (2008) Lymphocytes from patients with type 1 diabetes display a distinct profile of chromatin histone H3 lysine 9 dimethylation: An epigenetic study in diabetes. *Diabetes* 57(12):3189–3198.
21. Miao F, et al. (2007) Genome-wide analysis of histone lysine methylation variations caused by diabetic conditions in human monocytes. *J Biol Chem* 282(18): 13854–13863.
22. Miao F, Gonzalo IG, Lanting L, Natarajan R (2004) In vivo chromatin remodeling events leading to inflammatory gene transcription under diabetic conditions. *J Biol Chem* 279(17):18091–18097.
23. Pirola L, et al. (2011) Genome-wide analysis distinguishes hyperglycemia regulated epigenetic signatures of primary vascular cells. *Genome Res* 21(10): 1601–1615.
24. El-Osta A, et al. (2008) Transient high glucose causes persistent epigenetic changes and altered gene expression during subsequent normoglycemia. *J Exp Med* 205(10): 2409–2417.
25. Villeneuve LM, et al. (2008) Epigenetic histone H3 lysine 9 methylation in metabolic memory and inflammatory phenotype of vascular smooth muscle cells in diabetes. *Proc Natl Acad Sci USA* 105(26):9047–9052.
26. Sun G, et al. (2010) Epigenetic histone methylation modulates fibrotic gene expression. *J Am Soc Nephrol* 21(12):2069–2080.
27. Yuan H, et al. (2013) Involvement of p300/CBP and epigenetic histone acetylation in TGF-β1-mediated gene transcription in mesangial cells. *Am J Physiol Renal Physiol* 304(5):F601–F613.
28. Miao F, et al. (2012) Profiles of epigenetic histone post-translational modifications at type 1 diabetes susceptible genes. *J Biol Chem* 287(20):16335–16345.
29. Reddy MA, et al. (2014) Losartan reverses permissive epigenetic changes in renal glomeruli of diabetic db/db mice. *Kidney Int* 85(2):362–373.
30. Zhong Q, Kowluru RA (2010) Role of histone acetylation in the development of diabetic retinopathy and the metabolic memory phenomenon. *J Cell Biochem* 110(6): 1306–1313.
31. Miao F, et al.; DCCT/EDIC Research Group (2014) Evaluating the role of epigenetic histone modifications in the metabolic memory of type 1 diabetes. *Diabetes* 63(5): 1748–1762.
32. Reddy MA, Tak Park J, Natarajan R (2013) Epigenetic modifications in the pathogenesis of diabetic nephropathy. *Semin Nephrol* 33(4):341–353.
33. Bell CG, et al. (2010) Genome-wide DNA methylation analysis for diabetic nephropathy in type 1 diabetes mellitus. *BMC Med Genomics* 3:33.
34. Agardh E, et al. (2015) Genome-wide analysis of DNA methylation in subjects with type 1 diabetes identifies epigenetic modifications associated with proliferative diabetic retinopathy. *BMC Med* 13:182.
35. Olsen AS, Sarras MP, Jr, Leontovich A, Intine RV (2012) Heritable transmission of diabetic metabolic memory in zebrafish correlates with DNA hypomethylation and aberrant gene expression. *Diabetes* 61(2):485–491.
36. Park LK, et al. (2014) Genome-wide DNA methylation analysis identifies a metabolic memory profile in patient-derived diabetic foot ulcer fibroblasts. *Epigenetics* 9(10): 1339–1349.
37. Tewari S, Zhong Q, Santos JM, Kowluru RA (2012) Mitochondria DNA replication and DNA methylation in the metabolic memory associated with continued progression of diabetic retinopathy. *Invest Ophthalmol Vis Sci* 53(8): 4881–4888.
38. Laird PW (2010) Principles and challenges of genomewide DNA methylation analysis. *Nat Rev Genet* 11(3):191–203.
39. Brennan EP, et al. (2010) DNA methylation profiling in cell models of diabetic nephropathy. *Epigenetics* 5(5):396–401.
40. Al-Kateb H, et al.; Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications Research Group (2008) Multiple superoxide dismutase 1/splicing factor serine alanine 15 variants are associated with the development and progression of diabetic nephropathy: The Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications Genetics study. *Diabetes* 57(1):218–228.
41. Houseman EA, et al. (2012) DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* 13:86–101.
42. Kato M, Natarajan R (2014) Diabetic nephropathy–emerging epigenetic mechanisms. *Nat Rev Nephrol* 10(9):517–530.
43. Singh LP (2013) Thioredoxin interacting protein (TXNIP) and pathogenesis of diabetic retinopathy. *J Clin Exp Ophthalmol* 4:4.
44. Shah A, et al. (2015) Thioredoxin-interacting protein deficiency protects against diabetic nephropathy. *J Am Soc Nephrol* 26(12):2963–2977.
45. Shah A, et al. (2013) Thioredoxin-interacting protein mediates high glucose-induced reactive oxygen species generation by mitochondria and the NADPH oxidase, Nox4, in mesangial cells. *J Biol Chem* 288(10):6835–6848.
46. Cheng DW, et al. (2006) An analysis of high glucose and glucosamine-induced gene expression and oxidative stress in renal mesangial cells. *Arch Physiol Biochem* 112(4-5): 189–218.
47. Miao F, et al. (2013) RNA-sequencing analysis of high glucose-treated monocytes reveals novel transcriptome signatures and associated epigenetic profiles. *Physiol Genomics* 45(7):287–299.
48. Chong CR, et al. (2014) Thioredoxin-interacting protein: Pathophysiology and emerging pharmacotherapeutics in cardiovascular disease and diabetes. *Cardiovasc Drugs Ther* 28(4):347–360.
49. Siddiqi FS, et al. (November 3, 2015) The histone methyltransferase enzyme enhancer of zeste homolog 2 protects against podocyte oxidative stress and renal injury in diabetes. *J Am Soc Nephrol*, ASN.201409089.
50. Ko YA, et al. (2013) Cytosine methylation changes in enhancer regions of core profibrotic genes characterize kidney fibrosis development. *Genome Biol* 14(10):R108.
51. Toperoff G, et al. (2012) Genome-wide survey reveals predisposing diabetes type 2-related DNA methylation variations in human peripheral blood. *Hum Mol Genet* 21(2):371–383.
52. Grundberg E, et al.; Multiple Tissue Human Expression Resource Consortium (2013) 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* 93(5):876–890.
53. Pennacchio LA, Bickmore W, Dean A, Nobrega MA, Bejerano G (2013) Enhancers: Five essential questions. *Nat Rev Genet* 14(4):288–295.
54. Kaminsky ZA, et al. (2009) DNA methylation profiles in monozygotic and dizygotic twins. *Nat Genet* 41(2):240–245.
55. Volkmar M, et al. (2012) DNA methylation profiling identifies epigenetic dysregulation in pancreatic islets from type 2 diabetic patients. *EMBO J* 31(6): 1405–1426.
56. Smyth LJ, McKay GJ, Maxwell AP, McKnight AJ (2014) DNA hypermethylation and DNA hypomethylation is present at different loci in chronic kidney disease. *Epigenetics* 9(3):366–376.
57. Kulkarni H, et al. (2015) Novel epigenetic determinants of type 2 diabetes in Mexican-American families. *Hum Mol Genet* 24(18):5330–5344.
58. Florath I, et al. (2016) Type 2 diabetes and leucocyte DNA methylation: An epigenome-wide association study in over 1,500 older adults. *Diabetologia* 59(1): 130–138.
59. Chambers JC, et al. (2015) 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* 3(7):526–534.
60. Soriano-Tárraga C, et al.; GENESTROKE Consortium (2016) Epigenome-wide association study identifies TXNIP gene associated with type 2 diabetes mellitus and sustained hyperglycemia. *Hum Mol Genet* 25(3):609–619.
61. Petersen AK, et al. (2014) Epigenetics meets metabolomics: An epigenome-wide association study with blood serum metabolic traits. *Hum Mol Genet* 23(2):534–545.
62. Pfeiffer L, et al. (2015) DNA methylation of lipid-related genes affects blood lipid levels. *Circ Cardiovasc Genet* 8(2):334–342.
63. Bodnar JS, et al. (2002) Positional cloning of the combined hyperlipidemia gene Hyplip1. *Nat Genet* 30(1):110–116.
64. Maussion G, et al. (2014) Functional DNA methylation in a transcript specific 3'UTR region of TrkB associates with suicide. *Epigenetics* 9(8):1061–1070.
65. Mayol G, et al. (2012) DNA hypomethylation affects cancer-related biological functions and genes relevant in neuroblastoma pathogenesis. *PLoS One* 7(11):e48401.
66. Smith JF, et al. (2007) Identification of DNA methylation in 3′ genomic regions that are associated with upregulation of gene expression in colorectal cancer. *Epigenetics* 2(3):161–172.
67. DCCT/EDIC Research Group; Aiello LP, et al. (2015) Intensive diabetes therapy and ocular surgery in type 1 diabetes. *N Engl J Med* 372(18):1722–1733.