Evaluating the Role of Epigenetic histone modifications in the Metabolic Memory of Type 1 Diabetes.

Miao F\textsuperscript{1*}, Chen Z\textsuperscript{1*}, Genuth S\textsuperscript{2*}, Paterson AD\textsuperscript{3}, Zhang L\textsuperscript{1}, Wu X\textsuperscript{5}, Li SM\textsuperscript{6}, Cleary P\textsuperscript{4}, Riggs AD\textsuperscript{1}, Harlan DM\textsuperscript{7}, Lorenzi G\textsuperscript{8}, Kolterman O\textsuperscript{8}, Sun W\textsuperscript{4}, Lachin JM\textsuperscript{4}, Natarajan R\textsuperscript{1} and the DCCT/EDIC Research Group

\textsuperscript{1} Departments of Diabetes and Metabolic Diseases Research, \textsuperscript{5}Molecular and Cellular Biology and \textsuperscript{6}Biostatistics, Beckman Research Institute of City of Hope, Duarte, CA., \textsuperscript{2}Case Western Reserve University, Cleveland, OH, \textsuperscript{3}The Hospital for Sick Children, Toronto, Canada, \textsuperscript{4}The Biostatistics Center, The George Washington University, \textsuperscript{7}University of Massachusetts School of Medicine, Worcester, MA, \textsuperscript{8}University of California, San Diego, CA.

*These authors contributed equally

\textit{Correspondence to:}
Rama Natarajan, PhD
Endowed Professor and Associate Chair
Department of Diabetes,
Beckman Research Institute of the City of Hope
1500 East Duarte Road, Duarte, CA 91010
Tel. 626-256-4673, ext 62289, Fax: 626-301-8136
Email: RNatarajan@coh.org

\textbf{Keywords:} DCCT/EDIC, diabetes complications, metabolic memory, chromatin histone modifications, epigenomics, HbA1c

\textbf{Running title:} Epigenomics in Metabolic Memory

\textbf{Abstract:} 201 words

\textbf{Text: (introduction through discussion):} 4759 words

\textbf{Number of Figures/Tables:} 8
Abstract:
We assessed whether epigenetic histone post-translational modifications are associated with the prolonged beneficial effects (metabolic memory) of intensive versus conventional therapy during the Diabetes Control and Complications Trial (DCCT) on the progression of microvascular outcomes in the long-term Epidemiology of Diabetes Interventions and Complications (EDIC) study. We performed chromatin immunoprecipitation linked to promoter tiling arrays to profile H3lysine-9 acetylation (H3K9Ac), H3K4-trimethylation (H3K4Me3) and H3K9Me2 in blood monocytes and lymphocytes obtained from 30 DCCT conventional treatment group subjects [cases, mean DCCT HbA1c >9.1% (76 mmol/mol) and progression of retinopathy or nephropathy by EDIC year 10 of follow-up] versus 30 DCCT intensive treatment subjects [controls, mean DCCT HbA1c <7.3% (56 mmol/mol) and without progression]. Monocytes from cases had statistically greater numbers of promoter regions with enrichment in H3K9Ac (active chromatin mark) as compared to controls (p=0.0096). Among the two groups combined, monocyte H3K9Ac was significantly associated with the mean HbA1c during DCCT and EDIC (each p<2.2E-16). Of note, the top 38 case hyperacetylated promoters (p<0.05) included over fifteen genes related to the NF-κB inflammatory pathway, and were enriched in genes related to diabetic complications. These results suggest an association between HbA1c and H3K9Ac and a possible epigenetic explanation for metabolic memory in humans.
Introduction

Long-term micro- and macro-vascular complications are a major cause of morbidity and mortality in type 1 diabetes patients. Several genetic studies (1-3) have identified loci associated with diabetic nephropathy and recent evidence suggests that diabetes and its complications may also involve epigenetic factors (4-7). Epigenetic mechanisms triggered by hyperglycemia and related diabetic conditions may be particularly important since some patients with diabetes appear to develop long-term complications even after improved glucose control.

The Diabetes Control and Complications Trial (DCCT) (1983-93) demonstrated that intensive therapy aimed at near-normal levels of glycemia (HbA1c) profoundly reduced the development and progression of microvascular complications in patients with type 1 diabetes compared with conventional therapy aimed at maintaining clinical well-being (8). After DCCT end, the cohort was followed annually in the observational Epidemiology of Diabetes Interventions and Complications (EDIC) study (1994-present) when all subjects were encouraged to practice intensive therapy. Since EDIC year 4, both groups have maintained mean HbA1c levels of about 8% (64mmol/mol). Surprisingly, despite nearly equivalent glycemic control during EDIC, subjects in the original DCCT conventional therapy group continued to develop complications at significantly greater rates than participants in the original intensive therapy group (8-14). This phenomenon has been termed “metabolic memory”. Similar “legacy effects” of prior strict glycemic control were also observed in the United Kingdom Prospective Diabetes Study (15; 16).

Thus, an early period of better glycemic control appears to have sustained influence on long-term complications, and the adverse effects of early hyperglycemia may be ‘imprinted’ on target
tissues over time. Recently, epigenetic mechanisms have been implicated in the damaging effects of hyperglycemia and in experimental models of metabolic memory (17-25).

Epigenetics refers to mechanisms that can modify gene expression and phenotype without changes in the underlying DNA sequence (26; 27). Although different cell types in a person are assumed to have identical DNA sequence, they possess distinct differences in their epigenetic information, such as DNA methylation and post-translational modifications (PTMs) of histone proteins contained in the chromatin. Epigenetic modifications can occur when individuals are exposed to environmental factors such as infections and nutritional changes, and can predispose them to diseases such as diabetes (28). Nucleosomes, the basic subunits of chromatin, consist of octamers of histones H2A, H2B, H3 and H4, wrapped by DNA. PTMs of histones (e.g. acetylation, methylation, phosphorylation and ubiquitylation) form an epigenetic layer together with DNA methylation (29-31) which affects gene transcription. Acetylation of histones at lysine residues is generally associated with transcriptionally active genes, whereas lysine methylation leads to gene activation or repression, depending on the specific site and level of methylation (29; 30; 32). Alterations in histone PTMs and their interactions with other nuclear proteins at gene promoters or other regulatory regions can lead to relatively stable epigenetic changes that alter chromatin structure. In turn, this can lead to long-term dysregulated gene expression and disease progression.

Recent studies using vascular and inflammatory cells treated in vitro with high glucose (HG), or target cells and tissues derived from models of diabetic complications, provide strong evidence that alterations in epigenetic histone PTMs play key roles in diabetes-induced inflammation and vascular complications and potentially in the metabolic memory phenomenon (17-25; 33-41). However, studies have not yet been performed directly in humans with diabetes
and metabolic memory. To examine if epigenetic mechanisms are related to glycemic history, the progression of complications and metabolic memory in human diabetes, we explored variations in the profiles of key histone PTMs at promoter regions in peripheral blood lymphocytes and monocytes obtained from selected EDIC cohort subjects.

**Methods:**

The human research study protocol was approved by the Institutional Review Board (IRB) at City of Hope and the EDIC clinics.

**Subjects and Experimental Design.**

The DCCT enrolled 1441 subjects who were followed for an average of 6.5 years to study closeout in 1993. Consenting subjects were then followed annually during EDIC, of whom 60 were selected for this study. A case group of 30 subjects was selected from the former DCCT conventional therapy group who had a DCCT mean HbA1c of > 9.1% (76 mmol/mol), and progression of ≥ 5 steps from DCCT closeout to year 10 in EDIC on the Early Treatment Diabetic Retinopathy Study (EDTRS) scale of retinopathy severity, or scatter laser in EDIC, and/or albumin excretion rate (AER)>300 mg/24 hours or end stage renal disease (ESRD) at EDIC year 9/10 among those free of scatter laser and albuminuria during DCCT (some had both retinal and renal complications). A control group of 30 subjects was selected from the former DCCT intensive therapy group who had a DCCT mean HbA1c < 7.3% (56 mmol/mol), and no progression in ETDRS scale from DCCT closeout through EDIC year 9/10 and AER≤30 mg/24 hours in all DCCT-EDIC years (Table 1). All subjects were white, non-smokers. The epigenetics laboratory was masked to the case/control group identity of each patient until all profilings and initial data analyses were completed.
**Blood cell isolation.**

Blood samples (50 ml) were freshly collected from each patient at participating EDIC clinics and shipped overnight to the City of Hope using Thermasure Ambient Gel Packs (Sabra) to maintain the temperature between 15-25°C. Peripheral blood mononuclear cells were isolated using endotoxin-free Ficoll-Paque Plus (GE Healthcare 17-1440-03). Monocytes were purified by CD14+ positive selection (STEMCELL Technologies Cat. No.18088) and flow-through fractions were collected, washed and used as the lymphocyte fractions.

**Chromatin preparation, Chromatin Immunoprecipitation (ChIP) assays and ChIP-chip experiments.**

Lymphocytes and monocytes from each participant were cross-linked in 1% formaldehyde, sonicated to shear DNA and stored at -80°C for ChIP assays. ChIPs were performed as described (17; 21) using the following histone PTM antibodies: anti-histone H3 lysine-9 acetylation (H3K9Ac) (Abcam ab10812), H3 lysine-4 trimethylation (H3K4Me3) (Active motif 39169); H3K9Me2 (Active motif 39375). ChIP-enriched DNA samples and no antibody controls were purified from monocytes and lymphocytes of each subject. Five sets of ChIP-enriched DNA samples (H3K9Ac, H3K4Me3, and H3K9Me2 in lymphocytes, and H3K9Ac and H3K4Me3 in monocytes) from the 60 patients were obtained (299 samples total, one monocyte H3K4Me3 ChIP from the control group failed). Each sample was then labeled and hybridized to Roche Nimblegen human 720K RefSeq promoter tiling arrays (ChIP-chip) to profile these histone PTMs at promoter regions (-3,200 bp to 800 bp relative to the transcription start sites [TSS] of 22,542 Refseq genes) genome-wide. Hybridizations were performed at the Clinical Microarray Core, University of California, Los Angeles or the Functional Genomics Core, Beckman
Research Institute. Therefore, 5 ChIP-chip data sets were generated from 5 sample sets. The experimental design and data analysis pipeline are shown in the Supplemental Figs. S1 and S2.

**Real-Time Quantitative PCR (QPCR).**

ChIP-QPCRs were performed using the histone H3K9Ac ChIP-enriched DNA from each individual (n=60) and normalized to Input DNA as described (17; 21). Primers were designed to amplify the following promoter regions for ChIP-QPCR validations. *STAT1* (5’GGTCATTCAAGAGCCAGGGCAGGAGA3’ and 5’TTAAAGCCAGCCCCAAATAC3’); *TNF*α (5’ACCACAGCAATGGGTAGGAG3’ and 5’ GAGGTCCTGGAGGCTCTTTC3’); *IL1A* (5’ TCACCTGAGGCCCAGAGTTT3’ and 5’ AAGCTGCTTTCCTCCCAGAT3’) and *SPI1* (5’ GCTTCCCTACACCAGAAGA3’ and 5’ TGGGGAACACTGAGAAACCTG3’). Total RNA and polyA RNA were prepared as described (18).

**Bioinformatic Analysis of ChIP-chip data.**

Extensive quality assessments were first applied to each array/dataset using methods listed in Online Supplemental Figure S3. Probe-level Cy5/Cy3 log2 ratios on each array were shifted to render the Tukey biweight mean of log2 ratios to 0. The complete data from all the case and control samples in each dataset were then quantile normalized. For H3K4Me3 and H3K9Me2, Loess normalization was initially applied to each sample to correct dye bias found on some of the samples by MA plots. Regions enriched with histone modifications (i.e. containing at least 4 consecutive probes with log2 ratio greater than 85th percentile of all the probe-level log2 ratios on each array) were identified using TAMAL(42) in each of the 60 samples. Commonly-acetylated regions among all these samples were then designated as regions acetylated in at least 25% samples and at least 350 bp in length. The associations between acetylation and HbA1c levels at different time periods (Table 1) at these commonly-acetylated regions were analyzed.
using Pearson correlation coefficients (see Figure 2 legend). Null data were generated at the same 6248 regions after random sample permutation (N=100). Furthermore, case hyper-modified regions at commonly-modified regions in at least 50% of the cases were identified by multiple linear regression analysis with and without adjusting for age and gender (see Figure 3 legend). Hyper-modified regions in controls were similarly identified at commonly-modified regions in the controls. The comparisons were not adjusted for batches due to good technical reproducibility (Online Supplementary Figure S4), the sample processing (see Online Supplemental Figure S1 legend), as well as normalization applied for systemic bias correction. The false discovery rate (FDR) was calculated as the number of falsely discovered regions (estimated by the median number of regions of 100 sample permutations using the same analysis procedure described above) divided by the number of observed regions.

Data analyses (Online Supplement Figure S2) were performed using custom R codes on top of Bioconductor and "stats" packages unless otherwise specified. Hierarchical clustering was generated by Cluster v3.0 and visualized by Java Treeview v2.1.

**Data Deposition.**

The ChIP-chip datasets have been deposited in the NCBI Gene Expression Omnibus (GEO) database (GEO accession number GSE47385).

**Results**

**Cases and Controls:** The characteristics of the selected 30 cases and 30 controls are summarized in Table 1. As a result of the design, the cases had higher levels of AER and eGFR (evidence of hyperfiltration) even at DCCT baseline, and higher prevalences of retinopathy,
nephropathy and neuropathy during the DCCT and EDIC. While cases and controls were not selected on the basis of lipids and blood pressure, the levels among cases were also higher. Since the cases and controls were selected in part on the basis of different levels of HbA1c during DCCT, the HbA1c was higher among cases than controls during DCCT (10.2 vs. 6.4%, i.e. 88 versus 46 mmol/mol) (Table 1). Owing to a modest within-person (intra-class) correlation of HbA1c levels of 0.62 over the DCCT and 0.64 over the 10 years of EDIC, the cases also had higher HbA1c at DCCT eligibility testing (10.1% versus 8.1%, 87 versus 65 mmol/mol), during the 10 years of EDIC (8.5 vs. 7.1%, 69 vs. 54 mmol/mol) and at the time of blood draw for this study (EDIC year 16/17) (8.0 vs. 7.4%, 64 vs. 57 mmol/mol).

Global comparison of key histone modifications between cases and controls.

ChIP-enriched DNA samples were obtained from monocytes and lymphocytes of each of the 60 subjects with antibodies to histone H3K9Ac (a modification usually associated with active gene expression), H3K4Me3 (normally associated with active promoters) and H3K9Me2 (associated with repressed genes). Five ChIP-chip datasets were thereby generated as described under Methods. From these data, regions with differences in histone modifications were identified as described under Methods. The distribution of the number of regions enriched with histone-modifications in the cases and controls is shown in Figure 1 for each of the five chromatin marks evaluated. Cases had a significantly higher average number of regions enriched with H3K9Ac in monocytes than did controls (6470 vs. 6105, Figure 1A, Wilcoxon rank-sum exact two-sided test p=0.0096; p=0.048 adjusting for 5 tests). For the other chromatin marks studied, the average number of enriched regions showed no significant difference between cases and controls: monocyte H3K4Me3 (6284 vs. 6310); lymphocyte H3K9Ac (6979 vs. 7153);
lymphocyte H3K4Me3 (7686 vs. 7699); and lymphocyte H3K9Me2 (4979 vs. 5150) (Figure 1B to 1E).

**Association of monocyte H3K9Ac levels with HbA1c at different time periods.**

For each of the 60 EDIC subjects, mean HbA1c levels were measured over four time periods and also at the time of blood draw (EDIC yrs 16-17) for this study (Table 1) (schematically depicted in Figure 2A). HbA1c was measured upon entry, then quarterly during DCCT, and annually during EDIC. The four periods are: DCCT mean, representing the mean level of quarterly HbA1c during DCCT; EDIC mean, representing the mean of annual HbA1c during EDIC up to the time of blood draw for the current epigenetics study; DCCT/EDIC combined mean, representing the mean level during DCCT and EDIC periods combined; and the Pre-DCCT/DCCT/EDIC combined mean, representing an estimate of actual mean HbA1c since diagnosis up to blood draw. The DCCT/EDIC combined mean is the time-weighted average of the quarterly DCCT HbA1c values (weighted by 3 months) and the annual EDIC values (weighted by 12 months). The Pre-DCCT/DCCT/EDIC HbA1c likewise is a time-weighted average including the baseline HbA1c weighted by the months duration of diabetes on entry.

We assessed the correlation between HbA1c over these different time periods with the H3K9Ac levels at the 6248 commonly-acetylated regions across all 60 samples (as defined under Methods). The empirical distributions of correlation coefficients with HbA1c at different time periods (colored solid lines), as well as under null condition (black dashed line) at all these regions are shown in Figure 2B (blue curve for EDIC phase is not visible due to overlap by the curves for DCCT/EDIC and pre-DCCT/DCCT/EDIC). Notably, when compared to null conditions, H3K9Ac levels showed highly significant, positive associations (right shifts) with HbA1c at all time periods (p < 2.2E-16, Kolmogorov–Smirnov two-sided test).
Identification of group-specific promoter regions displaying the largest differences in the H3K9Ac levels.

We then searched for specific genomic regions that had the largest differences in monocyte H3K9Ac levels in cases versus controls. We first identified 4,644 H3K9 commonly-acetylated regions in cases, and 4,286 regions in the controls respectively (see Methods and Figure 3 legend for details). Among these, 3780 case commonly-acetylated regions overlap with 3784 control regions.

For each of the 4,644 commonly-acetylated among the cases, the H3K9Ac levels were compared in cases versus controls using a one-sided t-test. Of these, 35 regions had nominally significantly higher H3K9Ac levels in the cases vs. controls (nominal p < 0.05 without correction for the 4,644 tests, and fold change ≥ 1.1) (Figure 3A). The median FDR of these identified regions was estimated to be 0.31 by sample permutation (n=100), which means that about 70% of the genes in this list are true positives. Because of the potential effect of demographic factors (principally age and gender) on H3K9Ac, we also analyzed the 4,644 regions that were acetylated among the cases using linear regression models adjusting for age and gender. Using the same criteria, 39 regions (Table 2) were identified that had higher H3K9Ac levels in the cases vs. controls (nominal p<0.05) with estimated FDR at 33%. These regions were located in the promoters of 38 genes (-3200bp to 800bp relative to TSS) compared to the 35 genes identified without adjustment, 26 being in common (Figure 3C).

Similarly, among the 4,286 control commonly-acetylated regions, six (located in the promoters of 7 Refseq genes) and 10 regions (including 8 covering 8 Refseq promoters, Table 2) were hyperacetylated in controls (relative to cases, nominal p < 0.05) without and with adjusting
for age and gender (Figure 3B). Among the two lists, five genes are in common (Figure 3C). However, the median FDR estimated by permutation was 100%.

Similar analyses were also performed on the other four datasets including monocyte H3K4Me3, and lymphocyte H3K9Ac, H3K4Me3 and H3K9Me2. Using the same criteria applied to the monocyte H3K9Ac, the case and control hyper-modified regions were identified and summarized for these four datasets (Supplementary Table S1). Because the estimated FDR was 100% for each of these four identified lists, we focused our subsequent analyses only on the monocyte H3K9Ac dataset. There was also no overlap between the regions found in the five datasets, suggesting that in general the differences in PTMs occur independent of each other.

**Analysis of monocyte histone H3K9 hyperacetylated regions/promoters in cases and controls.**

To visualize the differences between the case and control groups, and examine whether these regions share similar acetylation patterns among these samples, we performed hierarchical clustering analysis on the case or control hyperacetylated regions identified after adjusting for age and gender (which include 39 annotated case hyperacetylated and 8 annotated control hyperacetylated regions) (Figure 4A). As seen, the acetylation levels were higher in general in the cases than controls at the case hyperacetylated regions. Among these regions, many were hyperacetylated in the same set of cases. Some heterogeneity can also be noted.

We used *in silico* analysis to further evaluate the potential functional relevance of the 38 genes which have 39 case hyperacetylated regions in their promoters (Table 2). Ingenuity pathway analysis (IPA) revealed that these genes were enriched for several diabetes and diabetes-complication related pathways, including TNFR2 signaling, macrophage and dendritic cell functions, interferon regulator factor (IRF) signaling and pattern recognition, apoptosis, reactive oxygen species, and antiviral responses (Figure 4B). Of note, genes related to the NF-κB
pathway, which is well known to be associated with inflammation, immunity and diabetic complications, were clearly enriched in this gene list. These included NF-κB pathway or target genes (TNF, IL1A, DENND4B, CYLD, IER3, TRAF1), transcription factors (NFKB1, NFKBIZ, STAT1) and regulatory genes (CYLD, PU.1, IER3, TRAF1). Furthermore, motif analysis by Transcription factor Affinity Prediction (43) showed that the top five enriched motifs (BH adjusted p<0.00005) with high affinities to the promoters of the 38 case-hyperacetylated genes (-500bp to 100bp relative to TSS) are all related to NF-κB.

**Evaluation of association between group-specific hyperacetylated regions and HbA1c at different time periods**

We then assessed the relationship between HbA1c over different time periods and the 39 case and 8 control hyperacetylated regions located in Refseq promoters (Table 2). Among the 60 case and control group subjects combined, we calculated the Pearson’s correlation coefficients of HbA1c measures with the H3K9Ac levels of these 47 regions, and visualized the resulting correlation coefficients by heatmaps (Figure 5). We found that the acetylation levels at most of these regions were positively (at case hyperacetylated regions) or negatively (at control hyperacetylated regions) correlated with average HbA1c levels. The associations with mean HbA1c during EDIC, DCCT/EDIC combined or the Pre-DCCT/DCCT/EDIC combined periods were much greater than at the time of blood draw or during DCCT only. HbA1c from the Pre-DCCT/DCCT/EDIC combined period had the highest association, while association at the time of blood draw was lowest. Notably, these results suggest that H3K9Ac levels of case/control hyperacetylated regions are associated with long-term HbA1c levels over the lifetime of diabetes including the DCCT and EDIC study phases.
Additional analyses also assessed correlations of HbA1c with the H3K9Ac levels at these same regions separately among the case and the control groups, with predominantly positive correlations among cases at case hyperacetylated regions, and inverse correlations among cases at control hyperacetylated regions. Correlations among controls were somewhat weaker. Details are shown by heatmaps in Online Supplemental Figure S5 and its Legend.

Validation of hyperacetylated region at the STAT1 gene promoter.

To validate the case hyperacetylated regions (Table 2), we chose STAT1, TNF, IL1A and SPI1. The H3K9Ac differences between cases and controls in the STAT1 promoter region from the ChIP-chip experiments are illustrated as a representative example in Figure 6A by plotting acetylation levels of the 30 cases (red dots) and 30 controls (blue dots) within the STAT1 promoter. The hyperacetylated region located at 2800 to 2400bp upstream of the STAT1 TSS is highlighted. The average acetylation enrichment is clearly higher in cases (red line) than in controls (blue line). To validate the ChIP-chip microarray data, we performed follow-up ChIP-QPCRs with ChIP-enriched DNA samples from each of the 60 participants. There was a significant difference between cases versus controls in the H3K9Ac level at the STAT1 promoter region (one-sided p=0.018, Wilcoxon’s Rank Sum exact test; Figure 6B). Similar significant differences were also obtained for TNF, IL1A and SPI1 promoter acetylations (Fig. 6C, 6D, 6E) (p<0.001, p=0.007 and p=0.002 respectively). A positive correlation was also seen between data from ChIP-chips and ChIP-QPCRs for STAT1 (Pearson’s r = 0.46, p = 1.9×10⁻⁴, Online Supplemental Figure S6).
Discussion

The objective of this study was to explore whether epigenetic differences could in part be associated with glycemic history and complications, and metabolic memory in type 1 diabetes by studying patients from the DCCT/EDIC cohort. As a first step toward understanding the link between epigenetics and metabolic memory in humans, we examined whether there were specific differences in genome-wide PTMs of promoter histones in monocytes and lymphocytes between subjects whose complications progressed versus those whose did not (cases versus controls). Among cases drawn from the conventional group with a mean HbA1c of 10.2% (88mmol/mol) during the DCCT, complications progressed markedly in EDIC during which the mean HbA1c was 8.7% (72mmol/mol). Conversely, among controls drawn from the intensive group with a mean HbA1c of 6.4% (46mmol/mol) during the DCCT, complications showed no progression in EDIC during which the mean HbA1c rose to 7.1% (54mmol/mol). The conventional group of the entire DCCT/EDIC cohort showed progression of complications during EDIC even though the HbA1c improved from about 9% (75mmol/mol) during the DCCT to about 8% (64mmol/mol) during EDIC, and the intensive group showed substantially less progression even though the HbA1c rose from about 7% (53mmol/mol) during the DCCT to about 8% (64mmol/mol) during EDIC. Thus, our cases and controls represent the extremes of the respective original conventional and intensive groups in which the metabolic memory effect was initially documented (9; 12; 13) and thus provided a good platform for this study.

The rationale for studying monocytes and lymphocytes was as follows. These cells have inflammatory and immune response properties and can be obtained in a relatively non-invasive fashion during a regular visit. Next, the epigenome, unlike the genome, is cell type specific. Moreover, evidence shows that inflammation and abnormal immune responses accompanied by
monocyte and lymphocyte infiltration play major pathogenic roles in diabetic retinopathy and nephropathy (5-7), two complications evident in our case but not control group. Finally, reports (18; 34; 44; 45) show that hyperglycemia itself can induce histone lysine acetylation and methylation in monocytes at inflammatory- and diabetes-related gene loci.

As the first study to profile multiple histone modifications at promoters genome-wide in lymphocytes and monocytes of several participants from a diabetes clinical trial, we analyzed the initial overall correlations across the five histone PTM datasets. This revealed high correlations between the same active marks (H3K9Ac and H3K4Me3) in different cell types (monocytes and lymphocytes), modest correlations between the two active marks in same cell types, and poor correlations between repressive mark (H3K9Me2) and either of the active marks (Online Supplementary Figures S7 and S8). Notably, through identification of promoter histone-modified regions genome-wide, we found significantly greater H3K9Ac-enriched regions in monocytes of cases than controls, whereas the other PTMs showed no significant differences. Furthermore, we identified 38 H3K9 hyperacetylated promoters in cases, none of which was differentially-modified in the other PTM datasets. Exploratory evaluation of the functional relevance of these 38 hyperacetylated hotspots revealed they were enriched for several diabetes and diabetes complication-related pathways, including TNFR2 signaling. Of note, recent studies show that circulating TNFR1 and TNFR2 levels predict renal function decline in type 1 diabetes(46; 47). Of the affected gene promoters detected in cases, many were regulated by the NF-κB pathway, including key inflammatory genes, transcriptional factors, and others involved in inflammatory and immunological pathways known to operate in diabetic complications (5-7). This suggests that systemic deregulation induced by histone acetylation at multiple levels in the NF-κB pathway could be related to metabolic memory.
The associations of HbA1c with H3K9Ac levels of the case-control groups combined were further analyzed. The mean HbA1c over extended periods (e.g. DCCT and EDIC combined) was significantly associated with H3K9Ac levels in monocytes among over 6000 commonly-acetylated regions across 60 samples. Although this may be expected since we observed hyperacetylation in cases, and the case-control groups were defined on the basis of differential HbA1c levels during DCCT, interestingly, additional analyses separately among cases and controls at case hyperacetylated regions showed somewhat stronger correlations with HbA1c among cases (see Online Supplemental Figure S5). These results are consistent with prior analyses from the DCCT (48; 49) showing a non-linear relationship between HbA1c and complications progression, such that the risk increases exponentially with increasing HbA1c.

Our current studies link diabetic complications to the histone acetylation of a set of genes related to inflammation in vivo. Moreover, the acetylation level was correlated with the previous history of Hb1Ac levels. These results, together, support a possible epigenetic mechanism for metabolic memory, whereby prolonged hyperglycemia may have a cumulative effect over time on H3K9 hyper-acetylation at key genome regions that once acetylated, may remain relatively stable and transmitted by unknown mechanisms. One obvious question here is whether the hyper-acetylation of the promoters correlates with increased expression of the corresponding genes. We did not observe significant correlations between them when we profiled gene expression (Affymetrix) in 12 monocyte samples (7 cases and 5 controls) (data not shown). This could be due to the limited number of samples tested. However, because we also did not find any significant differences in circulating cytokines (serum) between 30 cases and 30 controls at blood draw (data not shown), we speculate that, unlike the demonstrated epigenetic H3K9Ac levels, gene expression levels at the time of blood draw may not clearly reflect differences
between cases and controls. But, the acetylation could sensitize or render genes more susceptible to self-perpetuating changes in expression in response to external stimuli. Because H3K9Ac is associated with relaxed chromatin and active genes, higher H3K9Ac occurring around a set of complication-related gene promoter regions may lead to a more open or poised chromatin state at these regions. Support also comes from our recent transcriptomic profiling of monocytes which showed that most genes induced by HG have high levels of promoter H3K9Ac even in the normoglycemic state and are thus pre-selected or poised for activation by hyperglycemia (45). Interestingly, among our 38 case hyperacetylated promoters, the expression of \textit{TNF, LTA, STAT1} and \textit{CARD8} were all induced by HG (45).

Thus, the most likely scenario is these genes are more poised for activation in the case group compared to control, contributing to various diabetic complications in the long term. This could be a consequence of the early exposure to hyperglycemia (measured by HbA1c), which is known to be associated with increased rates of long-term diabetic complications. Recent evidence indicates that the enzymes mediating histone modifications (writers) are more likely to be inherited epigenetically than the modification itself (50), suggesting that in our study, key histone H3K9-acetyltransferases may be playing a role. It is also likely that observed acetylation differences are due to either DNA methylation, or other heritable epigenetic changes at distant transcription factors or microRNA-encoding regions. An epigenetic mechanism is also supported by previous experimental models of metabolic memory (17-25;34-40). Further investigation is needed to better understand the biological meaning of these phenotypes (hyperacetylation) and their association with diabetes complications and metabolic memory.

Our results suggest an association between epigenetic changes and microvascular complications and that hyperglycemia induced epigenetic changes can potentially explain part of
the metabolic memory phenomenon. However, they do not alone prove it, because from a cross-sectional study of this nature, one can only infer correlations, not causation. A limitation is that we did not select samples of cases and controls with different levels of HbA1c during the DCCT but equivalent levels of HbA1c at DCCT baseline and during EDIC. Thus, hyperglycemia prior to the DCCT or during EDIC could be partly responsible for epigenetic changes that affected the course of complications during DCCT (or even earlier) and EDIC. Secondly, given that cases and controls were selected on the basis of HbA1c levels during the DCCT (high versus low) and progression of complications during EDIC (yes versus no), any case-control differences in epigenetic changes would expected to be also accompanied by correlations of those changes with HbA1c. Thirdly, the cases and controls also differ with respect to other characteristics, such as blood pressure and lipids (Table 1) that in turn may have caused some epigenetic differences, and the resulting associations with HbA1c may be secondary to the association of HbA1c with these factors. Unfortunately, the sample size is relatively small to conduct multiple linear regression analyses adjusting for any impact of these different factors to provide a more definitive assessment of these relationships. Ideally, an epigenetic profiling of a much larger fraction of the cohort would be needed to address these issues.

In conclusion, we conducted comprehensive epigenomic profiling using cells from two selected sub-sets of DCCT/EDIC participants who experienced different rates of complications following a period with different levels of hyperglycemia to explore an epigenetic mechanism for metabolic memory in type 1 diabetes. Our results suggest that this metabolic memory phenomenon can in part be explained by increased epigenetic differences at key complication-related genes among individuals with higher HbA1c levels that may contribute to further progression of complications during EDIC.
Acknowledgements

This work was supported by grants from the Juvenile Diabetes Research Foundation (JDRF17-2008-900 and 17-2012-480) and the National Institutes of Health R01 DK065073 and R01 DK058191 (to RN). The DCCT/EDIC has been supported by U01 Cooperative Agreement grants (1982-93, 2011-2016), and contracts (1982-2011) with the Division of Diabetes Endocrinology and Metabolic Diseases of the National Institute of Diabetes and Digestive and Kidney Disease (U01 DK094176 and U01 DK094157).

FM, ZC, SG, ADP, and JML designed experiments, researched data, wrote and edited the manuscript; LZ and FM performed experiments; LZ, XW, SML, PC, GL, OK, and WS researched data; GL reviewed and edited the manuscript; ADR and DMH contributed to discussion and reviewed the manuscript; RN designed experiments, researched data, wrote and edited the manuscript, and supervised the project; DCCT/EDIC Research Study Group reviewed the manuscript. RN is the guarantor of this work and, as such, had full access to the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

We sincerely thank Dr. David Nathan (Harvard Medical School, Boston, MA) and the EDIC publications committee for many helpful discussions and critical comments, JDRF for their invaluable support, and all the study participants for their generosity and interest in the study. We also thank Dr. Margaret Morgan (City of Hope National Medical Center, Duarte, CA) for her help with this manuscript. We are grateful to Dr. Rose Gubitosi-Klug (Case Western University, Cleveland, OH), all the EDIC coordinators and Stephan Villavicencio (The George Washington University, Washington, DC) for their valuable assistance throughout the project.

The authors declare no conflict of interests.
References

1. Pezzolesi MG, Katavetin P, Kure M, Poznik GD, Skupien J, Mychaleckyj JC, Rich SS, Warram JH, Krolewski AS: Confirmation of genetic associations at ELMO1 in the GoKinD collection supports its role as a susceptibility gene in diabetic nephropathy. Diabetes 2009;58:2698-2702
2. Pezzolesi MG, Poznik GD, Mychaleckyj JC, Paterson AD, Barati MT, Klein JB, Ng DP, Placha G, Canani LH, Bochenski J, Waggott D, Merchant ML, Krolewski B, Mirea L, Wanic K, Katavetin P, Kure M, Wolkow P, Dunn JS, Smiles A, Walker WH, Boright AP, Bull SB, Doria A, Rogus JJ, Rich SS, Warram JH, Krolewski AS: Genome-wide association scan for diabetic nephropathy susceptibility genes in type 1 diabetes. Diabetes 2009;58:1403-1410
3. Sandholm N, Salem RM, McKnight AJ, Brennan EP, Forsblom C, Isakova T, McKay GI, Williams WW, Sadlier DM, Makinen VP, Swan EJ, Palmer C, Boright AP, Ahlqvist E, Deshmukh HA, Keller BJ, Huang H, Ahola AJ, Fagerholm E, Gordin D, Harjutsalo V, He B, Heikkila O, Hietala K, Kyto J, Lahermo P, Lehto M, Lithovius R, Osterholm AM, Parkkonen M, Pitkaniemi J, Rosengard-Barlund M, Saraheimo M, Sarti C, Soderlund J, Soro-Paavonen A, Syreeni A, Thorn LM, Tikkanen H, Tolonen N, Tryggvason K, Tuomilehto J, Waden J, Gill GV, Prior S, Guiducci C, Mirel DB, Taylor A, Hosseini SM, Parving HH, Rossing P, Tarnow L, Ladenvall C, Alhenc-Gelas F, Lefebvre P, Rigalleau V, Roussel R, Tregouet DA, Maestroni R, Maestroni A,Falhammar H, Gu T, Mollsten A, Cimponeriu D, Ioana M, Mata M, Mota E, Serafinceanu C, Stavarachi M, Hanson RL, Nelson RG, Kretzler M, Colhoun HM, Panduru NM, Gu HF, Brismar K, Zerbini G, Hadjadj S, Marre M, Groop L, Lajer B, Bull SB, Waggott D, Paterson AD, Savage DA, Bain SC, Martin F, Hirschhorn JN, Godson C, Florez JC, Group PH, Maxwell AP: New susceptibility loci associated with kidney disease in type 1 diabetes. PLoS Genet 2012;8:e1002921
4. Leung A, Schones DE, Natarajan R: Using epigenetic mechanisms to understand the impact of common disease causing alleles. Curr Opin Immunol 2012;24:558-563
5. Villeneuve LM, Reddy MA, Natarajan R: Epigenetics: deciphering its role in diabetes and its chronic complications. Clinical and experimental pharmacology & physiology 2011;38:451-459
6. Reddy MA, Natarajan R: Epigenetic mechanisms in diabetic vascular complications. Cardiovascular research 2011;90:421-429
7. Cooper ME, El-Osta A: Epigenetics: mechanisms and implications for diabetic complications. Circulation research 2010;107:1403-1413
8. The DCCT Research Group: The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. The Diabetes Control and Complications Trial Research Group. N Engl J Med 1993;329:977-986
9. DCCT/EDIC Research Group: Retinopathy and nephropathy in patients with type 1 diabetes four years after a trial of intensive therapy. The Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications Research Group. N Engl J Med 2000;342:2563-2569
10. Genuth S: Insights from the diabetes control and complications trial/epidemiology of diabetes interventions and complications study on the use of intensive glycemic treatment to reduce the risk of complications of type 1 diabetes. Endocr Pract 2006;12 Suppl 1:34-41
11. Cleary PA, Orchard TJ, Genuth S, Wong ND, Detrano R, Backlund JY, Zinman B, Jacobson A, Sun W, Lachin JM, Nathan DM: The effect of intensive glycemic treatment on coronary artery calcification in type 1 diabetic participants of the Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications (DCCT/EDIC) Study. Diabetes 2006;55:3556-3565
12. DCCT/EDIC Writing Team: Effect of intensive therapy on the microvascular complications of type 1 diabetes mellitus. Jama 2002;287:2563-2569
13. DCCT/EDIC Writing Team: Sustained effect of intensive treatment of type 1 diabetes mellitus on development and progression of diabetic nephropathy: the Epidemiology of Diabetes Interventions and Complications (EDIC) study. Jama 2003;290:2159-2167
14. Nathan DM, Cleary PA, Backlund JY, Genuith SM, Lachin JM, Orchard TJ, Raskin P, Zinman B: Intensive diabetes treatment and cardiovascular disease in patients with type 1 diabetes. N Engl J Med 2005;353:2643-2653
15. Holman RR, Paul SK, Bethel MA, Matthews DR, Neil HA: 10-year follow-up of intensive glucose control in type 2 diabetes. N Engl J Med 2008;359:1577-1589
16. Chalmers J, Cooper ME: UKPDS and the legacy effect. N Engl J Med 2008;359:1618-1620
17. Miao F, Smith DD, Zhang L, Min A, Feng W, Natarajan R: Lymphocytes from patients with type 1 diabetes display a distinct profile of chromatin histone H3 lysine 9 dimethylation: an epigenetic study in diabetes. Diabetes 2008;57:3189-3198
18. Miao F, Wu X, Zhang L, Yuan YC, Riggs AD, Natarajan R: Genome-wide analysis of histone lysine methylation variations caused by diabetic conditions in human monocytes. J Biol Chem 2007;282:13854-13863
19. Pirola L, Balcerczyk A, Tothill RW, Haviv I, Kaspi A, Lunke S, Ziemann M, Karagiannis T, Tonna S, Kowalczuk A, Beresford-Smith B, Macintyre G, Kelong M, Hongyu Z, Zhu J, El-Osta A: Genome-wide analysis distinguishes hyperglycemia regulated epigenetic signatures of primary vascular cells. Genome research 2011;21:1601-1615
20. El-Osta A, Brasacchio D, Yao D, Pocai A, Jones PL, Roeder RG, Cooper ME, Brownlee M: Transient high glucose causes persistent epigenetic changes and altered gene expression during subsequent normoglycemia. The Journal of experimental medicine 2008;205:2409-2417
21. Miao F, Chen Z, Zhang L, Liu Z, Wu X, Yuan YC, Natarajan R: Profiles of epigenetic histone post-translational modifications at type 1 diabetes susceptible genes. J Biol Chem 2012;287:16335-16345
22. Sun G, Reddy MA, Yuan H, Lanting L, Kato M, Natarajan R: Epigenetic histone methylation modulates fibrotic gene expression. J Am Soc Nephrol 2010;21:2069-2080
23. Zhong Q, Kowluru RA: Role of histone acetylation in the development of diabetic retinopathy and the metabolic memory phenomenon. J Cell Biochem 2010;110:1306-1313
24. Villeneuve LM, Reddy MA, Lanting LL, Wang M, Meng L, Natarajan R: Epigenetic histone H3 lysine 9 methylation in metabolic memory and inflammatory phenotype of vascular smooth muscle cells in diabetes. Proc Natl Acad Sci U S A 2008;105:9047-9052
25. Chan PS, Kanwar M, Kowluru RA: Resistance of retinal inflammatory mediators to suppress after reinstitution of good glycemic control: novel mechanism for metabolic memory. J Diabetes Complications 2010;24:55-63
26. Bird A: Perceptions of epigenetics. Nature 2007;447:396-398
27. Bonasio R, Tu S, Reinberg D: Molecular signals of epigenetic states. Science 2010;330:612-616
28. Jirtle RL, Skinner MK: Environmental epigenomics and disease susceptibility. Nat Rev Genet 2007;8:253-262
29. Jenuwein T, Allis CD: Translating the histone code. Science 2001;293:1074-1080
30. Kouzarides T: Chromatin modifications and their function. Cell 2007;128:693-705
31. Razin A, Riggs AD: DNA methylation and gene function. Science 1980;210:604-610
32. Martin C, Zhang Y: The diverse functions of histone lysine methylation. Nature reviews Molecular cell biology 2005;6:838-849
33. Li Y, Reddy MA, Miao F, Shanmugam N, Yee JK, Hawkins D, Ren B, Natarajan R: Role of the histone H3 lysine 4 methyltransferase, SET7/9, in the regulation of NF-kappaB-dependent inflammatory genes. Relevance to diabetes and inflammation. J Biol Chem 2008;283:26771-26781
34. Miao F, Gonzalo IG, Lanting L, Natarajan R: In vivo chromatin remodeling events leading to inflammatory gene transcription under diabetic conditions. J Biol Chem 2004;279:18091-18097
35. Brasacchio D, Okabe J, Tikellis C, Balcerczyk A, George P, Baker EK, Calkin AC, Brownlee M, Cooper ME, El-Osta A: Hyperglycemia induces a dynamic cooperativity of histone methylase and demethylase enzymes associated with gene-activating epigenetic marks that coexist on the lysine tail. Diabetes 2009;58:1229-1236
36. Yuan H, Reddy MA, Sun G, Lanting L, Wang M, Kato M, Natarajan R: Involvement of p300/CBP and epigenetic histone acetylation in TGF-beta1-mediated gene transcription in mesangial cells. Am J Physiol Renal Physiol 2013;304:F601-613

37. Kadiyala CS, Zheng L, Du Y, Yohannes E, Kao HY, Miyagi M, Kern TS: Acetylation of retinal histones in diabetes increases inflammatory proteins: effects of minocycline and manipulation of histone acetyltransferase (HAT) and histone deacetylase (HDAC). J Biol Chem 2012;287:25869-25880

38. Pirola L, Balcerczyk A, Okabe J, El-Osta A: Epigenetic phenomena linked to diabetic complications. Nat Rev Endocrinol 2010;6:665-675

39. Zhong Q, Kowluru RA: Epigenetic changes in mitochondrial superoxide dismutase in the retina and the development of diabetic retinopathy. Diabetes 2011;60:1304-1313

40. Caramori ML, Kim Y, Moore JH, Rich SS, Mychaleckyj JC, Kikyo N, Mauer M: Gene expression differences in skin fibroblasts in identical twins discordant for type 1 diabetes. Diabetes 2012;61:739-744

41. Ko YA, Mohtat D, Suzuki M, Park AS, Izquierdo MC, Han SY, Kang HM, Si H, Hostetter T, Pullman JM, Fazzari M, Verma A, Zheng D, Greally JM, Susztak K: Cytosine methylation changes in enhancer regions of core pro-fibrotic genes characterize kidney fibrosis development. Genome Biol 2013;14:R108

42. Bieda M, Xu X, Singer MA, Green R, Farnham PJ: Unbiased location analysis of E2F1-binding sites suggests a widespread role for E2F1 in the human genome. Genome research 2006;16:595-605

43. Thomas-Chollier M, Hufton A, Heinig M, O'Keeffe S, Masri NE, Roider HG, Manke T, Vingron M: Transcription factor binding predictions using TRAP for the analysis of ChIP-seq data and regulatory SNPs. Nature protocols 2011;6:1860-1869

44. Yun JM, Jialal I, Devaraj S: Epigenetic regulation of high glucose-induced proinflammatory cytokine production in monocytes by curcumin. J Nutr Biochem 2011;22:450-458

45. Miao F, Chen Z, Zhang L, Wang J, Gao H, Wu X, Natarajan R: RNA-Sequencing Analysis of High Glucose Treated Monocytes Reveals Novel Transcriptome Signatures and associated Epigenetic Profiles. Physiological genomics 2013;45:287-299

46. Gohda T, Niewczas MA, Ficociello LH, Walker WH, Skupien J, Rosetti F, Cullere X, Johnson AC, Crabtree G, Smiles AM, Mayadas TN, Warram JH, Krolewski AS: Circulating TNF receptors 1 and 2 predict stage 3 CKD in type 1 diabetes. J Am Soc Nephrol 2012;23:516-524

47. Lopes-Virella MF, Baker NL, Hunt KJ, Cleary PA, Klein R, Virella G, Group DER: Baseline markers of inflammation are associated with progression to macroalbuminuria in type 1 diabetic subjects. Diabetes Care 2013;36:2317-2323

48. The DCCT Research Group: The absence of a glycemic threshold for the development of long-term complications: the perspective of the Diabetes Control and Complications Trial. Diabetes 1996;45:1289-1298

49. The DCCT Research Group: The relationship of glycemic exposure (HbA1c) to the risk of development and progression of retinopathy in the diabetes control and complications trial. Diabetes 1995;44:968-983

50. Petruk S, Sedkov Y, Johnston DM, Hodgson JW, Black KL, Kovermann SK, Beck S, Canaani E, Brock HW, Mazo A: TrxG and PcG proteins but not methylated histones remain associated with DNA through replication. Cell 2012;150:922-933
Table 1. Clinical Characteristics of Cases and Controls at DCCT baseline, DCCT Closeout (end of randomized therapy), and at EDIC year 10, and HbA1c at different times.

|                      | DCCT baseline |                          | DCCT Closeout |                          | EDIC Year 10 |
|----------------------|---------------|---------------------------|---------------|---------------------------|--------------|
|                      | Control*      | Case*                     | P †           | Control*                  | Case*        | P †          |
| Number               | 30            | 30                        |               | 37 (6.1)                  | 34 (6.7)     | 0.18         |
| Intensive (%)        | 100           | 0                         | n/a           | 13 (4.8)                  | 13 (4.3)     | 0.86         |
| Primary Cohort (%)   | 43            | 20                        | 0.05          | 25.1 (40.6)               | 0.0007       |
| Female (%)           | 43            | 57                        | 0.3           | 0                         | 10           | 0.076        |
| Age                  | 30 (6.3)      | 28 (6.8)                  | 0.27          | 37 (6.1)                  | 34 (6.7)     | 0.18         |
| Diabetes Duration (yr)| 6.6 (4.6)    | 6.8 (3.8)                 | 0.65          | 13 (4.8)                  | 13 (4.3)     | 0.86         |
| AER (mg/24hr)        | 9.3 (5.2)     | 16.8 (15.3)               | 0.017         | 7.1 (3.9)                 | 25.1 (40.6)  | 0.0007       |
| AER≥300mg/24hr (%)   | 0             | 0                         | n/a           | 0                         | 0            | n/a          |
| GFR (mL/min/1.73m²)  | 116 (13.7)    | 142 (18.8)                | 0.045         | 114 (11.3)                | 118 (11.6)   | 0.11         |
| PDR† (%)             | 0             | 0                         | -             | 0                         | 17           | 0.019        |
| Clinical Neuropathy‡ (%) | 0             | 3                         | 0.31          | 3                         | 30           | 0.0056       |
| Mean BP § (mm Hg)    | 87 (9.5)      | 86 (7.6)                  | 0.56          | 87 (6.8)                  | 89 (10.4)    | 0.27         |
| HDL (mg/dL)          | 51 (12.7)     | 53 (12.5)                 | 0.61          | 52 (13.9)                 | 52 (11.9)    | 0.69         |
| LDL (mg/dL)          | 104 (28.7)    | 118 (24.4)                | 0.08          | 108 (26.9)                | 125 (27.6)   | 0.01         |
| Triglyceride (mg/dL) | 67 (21.6)     | 90 (41.7)                 | 0.041         | 61 (19.4)                 | 95 (43.9)    | 0.0027       |
| Hyperlipidemia (%)   | 0             | 0                         | -             | 23                        | 40           | 0.17         |
| HbA1c                |               |                           |               |                           |              |
| HbA1c at DCCT Eligibility (%) | 8.1 (1.3) | 10.1 (1.9) | <0.0001 |

(Continued on next page)
| Mean HbA1c over the DCCT (%) |  6.4 (0.4) | 10.2 (0.8) | <0.0001 |
|-----------------------------|-----------|-----------|--------|
| (mmol/mol)                  | 46 (4.4)  | 88 (8.7)  |        |
| EDIC mean to EDIC Year 10 (%) | 7.1 (0.8) | 8.7 (1.2) | <0.0001 |
| (mmol/mol)                  | 54 (8.7)  | 72 (13.1) |        |
| EDIC mean to blood draw ‡ (%) | 7.1 (0.8) | 8.5 (1.1) | <0.0001 |
| (mmol/mol)                  | 54 (8.7)  | 69 (12.0) |        |
| HbA1c at EDIC Year 10 (%)   | 7.2 (0.9) | 8.1 (1.3) | 0.0038  |
| (mmol/mol)                  | 55 (9.8)  | 65 (14.2) |        |
| HbA1c at blood draw ¶ (%)   | 7.4 (1.0) | 8.0 (1.0) | 0.024   |
| (mmol/mol)                  | 57 (10.9) | 64 (10.9) |        |
| DCCT/EDIC weighted mean to EDIC Year 10 (%) | 6.8 (0.6) | 9.3 (0.9) | <0.0001 |
| (mmol/mol)                  | 51 (6.6)  | 78 (9.8)  |        |
| DCCT/EDIC weighted mean to blood draw ¶ (%) | 6.9 (0.6) | 9.0 (0.9) | <0.0001 |
| (mmol/mol)                  | 52 (6.6)  | 75 (9.8)  |        |
| pre-DCCT/DCCT/EDIC weighted mean to EDIC Year 10 (%) | 7.2 (0.6) | 9.6 (0.9) | <0.0001 |
| (mmol/mol)                  | 55 (6.6)  | 81 (9.8)  |        |
| pre-DCCT/DCCT/EDIC weighted mean to blood draw ¶ (%) | 7.2 (0.6) | 9.3 (0.9) | <0.0001 |
| (mmol/mol)                  | 55 (6.6)  | 78 (9.8)  |        |

* Cases and Controls are defined in Methods section under “Subjects and Experimental Design”. † Proliferative diabetic retinopathy. ‡ Clinical Neuropathy is defined as 'Definite Clinical Neuropathy' and was obtained at DCCT baseline, DCCT closeout, and EDIC Year 13/14 (ref: The DCCT Research Group: The effect of intensive diabetes therapy on the development and progression of neuropathy. Ann Intern Med. 122:561-568, 1995). § Blood Pressure. ¶ All participants for this study had blood draw in EDIC Year 16 or Year 17. † Clinical characteristics were compared using the Wilcoxon rank-sum test and chi-square test for quantitative and categorical variables, respectively.
Table 2. Case or Control H3K9 Hyperacetylated Regions in Monocytes.

| ID | Chrom | Region Start | Region End | annotated | Fold change | Age + Gender adjusted log2 Fold-change | Age + Gender adjusted one-sided p-value |
|---|-------|--------------|------------|-----------|-------------|---------------------------------------|----------------------------------------|
| 1 | chr1  | 111962206    | 111962875  | RAP1A     | 0.15        | 0.006                                 |                                        |
| 2 | chr1  | 151785673    | 151786123  | S100A4    | 0.143       | 0.007                                 |                                        |
| 3 | chr1  | 152187791    | 152188172  | DENND4B   | 0.145       | 0.038                                 |                                        |
| 4 | chr1  | 152200405    | 152200880  | CRT2      | 0.173       | < 1E-3                                |                                        |
| 5 | chr2  | 68549953     | 68550333   | FBXO48    | 0.179       | <1E-3                                 |                                        |
| 6 | chr2  | 87534775     | 87535339   | NCRNA00152| 0.152       | 0.049                                 |                                        |
| 7 | chr2  | 111970045    | 111970416  | LOC541471 | 0.166       | 0.041                                 |                                        |
| 8 | chr2  | 111971091    | 111971556  | LOC541471 | 0.199       | 0.032                                 |                                        |
| 9 | chr2  | 113259272    | 113260041  | IL1A      | 0.138       | 0.034                                 |                                        |
| 10| chr2  | 191589560    | 191589923  | STAT1     | 0.181       | 0.02                                  |                                        |
| 11| chr2  | 241900740    | 241901302  | SEPT2     | 0.152       | 0.01                                  |                                        |
| 12| chr3  | 33812515     | 33812866   | PDCD6IP   | 0.197       | 0.009                                 |                                        |
| 13| chr3  | 103029890    | 103030241  | NFKBIZ    | 0.158       | <1E-3                                 |                                        |
| 14| chr4  | 25927998     | 25928776   | RBPJ      | 0.175       | 0.027                                 |                                        |
| 15| chr4  | 103639447    | 103639809  | NFKB1     | 0.203       | 0.002                                 |                                        |
| 16| chr5  | 169688141    | 169688720  | LOC257358 | 0.15        | 0.047                                 |                                        |
| 17| chr6  | 30821389     | 30822036   | FLOT1, IER3| 0.153      | 0.014                                 |                                        |
| 18| chr6  | 30823196     | 30823551   | IER3      | 0.173       | 0.039                                 |                                        |
| 19| chr6  | 31647450     | 31647891   | LTA       | 0.195       | 0.002                                 |                                        |
| 20| chr6  | 31650396     | 31650771   | TNF       | 0.165       | 0.035                                 |                                        |
| 21| chr6  | 159388691    | 159389164  | TAGAP     | 0.19        | 0.001                                 |                                        |
| 22| chr7  | 95790825     | 95791685   | SLC25A13  | 0.16        | 0.016                                 |                                        |
| 23| chr7  | 122729570    | 122730540  | TRAF1     | 0.148       | 0.001                                 |                                        |
| 24| chr10 | 45541225     | 45541586   | FAM21C    | 0.17        | 0.001                                 |                                        |
| 25| chr10 | 104209340    | 104210129  | TMEM180   | 0.164       | 0.003                                 |                                        |
| 26| chr11 | 47357165     | 47357764   | SPI1      | 0.142       | 0.002                                 |                                        |
| 27| chr11 | 56952484     | 56953709   | SLC43A3   | 0.145       | 0.044                                 |                                        |

(Continued on the next page)
|   | chr   | Start   | End     | Gene   | p-value | q-value |
|---|-------|---------|---------|--------|---------|---------|
| 28 | chr11 | 56953884| 56954534| SLC43A3| 0.172   | 0.021   |
| 29 | chr11 | 61202834| 61203277| DAGLA  | 0.161   | 0.002   |
| 30 | chr12 | 24964370| 24964817| DAD1L  | 0.141   | 0.034   |
| 31 | chr12 | 49707385| 49707850| SLC11A2| 0.153   | 0.006   |
| 32 | chr14 | 74150441| 74151092| LTBP2  | 0.147   | 0.019   |
| 33 | chr14 | 74151271| 74152036| LTBP2  | 0.165   | 0.028   |
| 34 | chr14 | 102126331| 102126800| RCOR1  | 0.142   | 0.003   |
| 35 | chr16 | 49331047| 49331408| CYLD   | 0.172   | 0.005   |
| 36 | chr17 | 4795432 | 4795889 | PFN1, ENO3 | 0.147 | 0.038   |
| 37 | chr17 | 6865049 | 6865522 | MIR497, BCL6B | 0.15 | 0.011   |
| 38 | chr17 | 35512673| 35513129| NR1D1  | 0.214   | 0.002   |
| 39 | chr19 | 53446461| 53446916| CARD8  | 0.164   | 0.017   |

|   | chr   | Start   | End     | Gene   | p-value | q-value |
|---|-------|---------|---------|--------|---------|---------|
| 1  | chr11 | 11332779| 11333154| CSNK2A1P | 0.194  | 0.001   |
| 2  | chr17 | 31547484| 31547949| CCL3L1, CCL3L3 | 0.210 | 0.046   |
| 3  | chr19 | 48583397| 48583760| TEX101  | 0.151   | 0.004   |
| 4  | chr8  | 7308555 | 7308917 | SPAG11B | 0.197   | 0.013   |
| 5  | chr8  | 7742450 | 7742897 | SPAG11A | 0.198   | 0.016   |
| 6  | chr8  | 7743064 | 7743420 | SPAG11A, SPAG11B | 0.175 | 0.035   |
| 7  | chr9  | 137592388| 137592759| PAEP    | 0.142   | 0.005   |
| 8  | chrX  | 23672897| 23673297| ACOT9   | 0.186   | 0.024   |
| 9  | chrX  | 52548220| 52548671|          | 0.154   | 0.002   |
| 10 | chrX  | 52561114| 52561565|          | 0.145   | 0.003   |

* The March 2006 human genome assembly (NCBI Build 36.1/hg18) was used for annotation. † Promoter is defined as -3200bp to 800bp relative to transcript start site.
LEGENDS

**Figure 1. Distribution of the number of regions enriched with the histone PTMs in the case and control groups.**

For each histone modification mark studied, the promoter regions of genes modified by the specific chromatin marks were identified in each sample using TAMAL. Modified regions were defined as regions that had at least four consecutive probes with modification levels in the form of log2 ratios greater than 85th percentile of all the probes on the Nimblegen promoter array. The distribution of the number of histone-modified regions across the case and control groups are shown by boxplots for H3K9Ac (A) and H3K4Me3 (B) in monocytes; and H3K9Ac (C), H3K4Me3 (D) and H3K9Me2 (E) in lymphocytes. Boxplot function in R was used to generate the plot where the middle lines within the boxes represent the median value, and the top and bottom whiskers are the maximum and minimum values excluding outliers that are over 1.5× interquantile range. Outliers are represented by empty circles. Two-sided Wilcoxon Rank-Sum Exact tests were used to compare the number of modified regions between the controls and cases in each dataset. The increased number of H3K9 acetylated regions in monocytes from cases versus controls was statistically significant (* p = 0.0096).

**Figure 2. Association of monocyte H3K9 acetylation levels with HbA1c at different time periods.**

A. The five time periods from which documented HbA1c values for the case and control groups were used. Each EDIC participant’s history of type 1 diabetes is divided into three phases, namely Pre-D CCT, DCCT and EDIC. Pre-D CCT phase is from the date of type 1 diabetes diagnosis to the date of entry into DCCT in 1983. DCCT is from entry into DCCT to the end of
DCCT in 1993 within which HbA1c was measured quarterly. EDIC is from 1993 (start of EDIC) to current, during which HbA1c was documented annually. All our study participants had the blood drawn in EDIC year 16/17 (2009-2010), and HbA1c level measured at this time is termed Blood draw. HbA1c levels in the other four periods (DCCT, EDIC, DCCT/EDIC and Pre-DCCT/DCCT/EDIC) are represented by mean values across specified periods. B. Association of acetylation levels with HbA1c at H3K9 commonly-acetylated regions. Commonly-acetylated regions across all 60 samples were defined as regions enriched with H3K9Ac in at least 15 samples (25%) that were at least 350-bp in length. The average acetylation level at each of the resulting 6248 commonly acetylated regions was calculated in each sample and its relationship with HbA1c at different time periods (color coded) is represented by Pearson’s correlation coefficient. Sample permutation (N=100) was performed to generate the correlation coefficients under the null condition. The cumulative densities of the correlation coefficients are plotted (color solid lines, original data; black dashed line, permuted data) and a shift to the right indicates a higher positive association. The statistical significances of distances between two empirical distributions were tested by K-S tests. The time periods are color coded as indicated.

Figure 3. Identification of group-specific promoter regions depicting the largest differences in the levels of histone H3K9Ac in monocytes.

The H3K9 commonly-acetylated regions in cases and controls were first identified. The average log2 ratio of at least four probes within these regions in cases and controls were compared, and the hyperacetylated regions in cases or controls were identified using multiple linear regression with one-sided test before and after adjusting for age and gender. The analyses were performed using custom R codes built on top of the functions implemented in “preprocessing Core” and
“GenomicRanges” bioconductor packages, as well ”stats” package. A. The volcano plot of 4644 commonly acetylated regions found in cases. B. The volcano plot of 4286 commonly acetylated regions found in controls. The y-axis indicates the –log_{10} (Pvalue) and the x-axis indicates the log_{2} ratio differences between the two groups. Each dot represents one group-specific commonly-acetylated region. Regions considered to be hyper-modified (fold change ≥ 1.1 and nominal p < 0.05) in cases or controls are highlighted as small empty circles. C. Venn Diagrams showing the comparison of the hyperacetylated promoters in the case and control groups, with and without adjustment for age and gender.

Figure 4. Regions with variations in histone H3K9Ac in monocytes of the cases and control groups, and the biological functions related to the annotated promoters in these regions.

A. Hierarchical clustering of hyperacetylated regions in cases and controls. The H3K9 hyperacetylated regions (age and gender adjusted nominal p < 0.05, fold change ≥ 1.1) in the case and control groups were pooled and the acetylation levels were calculated for each region by averaging the log_{2} ratios of probes falling into these regions. The acetylation levels were mean centered across the 60 samples, and hierarchical clustering of these regions was generated using average linkage method with Pearson correlation as the similarity metric and visualized by Java Treeview v2.1. Blue represents acetylation levels above the mean, and yellow represents acetylation levels below the mean. The annotated genes containing a hyperacetylated region in their promoters are shown on the right. For regions located in the promoters of more than one gene, all the corresponding gene symbols are listed (separated by commas). For multiple regions located within the same gene promoter, the same gene symbol is listed for each region. Thus 39 case hyperacetylated regions (spanning the promoters of 38 genes) and 8 control hyperacetylated
regions (spanning the promoters of 8 genes) are depicted. B. Ingenuity Pathway Analysis (IPA) of H3K9 hyperacetylated promoters in cases (in silico analyses). The 38 genes identified in case samples that have H3K9 hyperacetylated regions in their promoters (-3200bp to +800bp of TSS) were imported into IPA for identification of over-represented canonical pathways. The 162 pathways in IPA knowledge-based database were sorted by P values (Fisher’s exact test p values adjusted by Bonferroni-Holm [B-H] method) and the top 10 pathways are shown. Blue bars represent $-\log_{10}(P$-value) of pathways. The B-H adjusted p-values of the enrichments of all the 10 pathways are less than 0.0025.

**Figure 5. Correlation of HbA1c with acetylation levels at case/control H3K9 hyperacetylated regions.**

Pearson’s correlation coefficients of acetylation levels with HbA1c at regions depicting hyperacetylation in cases and controls (same regions shown in Figure 4A) at the different time periods across 60 samples are shown as heatmap. The rows represent the case/control hyperacetylated regions, and the columns represent different time periods (labels on top of the heatmap).

**Figure 6. Validation of ChIP-chip data of histone H3K9 hyperacetylated regions in cases versus controls.**

A. Probe level H3K9Ac levels at the promoter region (-3200bp to +800bp of TSS) of the *STAT1* gene (from ChIP-chip data). This region contained 38 probes. For each probe, each dot represents one sample, with red dots representing the cases and blue dots representing the controls. The red line represents the average probe log2 ratio of samples in all the cases, and the
blue line represents the average probe log2 ratio of all the controls. The dotted rectangle highlights the hyperacetylated region in the cases compared to controls. B-E: Results of ChIP-QPCR validation of the case hyperacetylated region located in the promoter/enhancer regions of the indicated genes *STAT1*, *TNFα* *IL1A* and *SPI1*. The histone H3K9Ac ChIPs were prepared from monocytes of each individual participant (30 cases and 30 controls). Standardized ChIP-QPCRs were then used to quantify the amount of specific modified ChIP-enriched DNA in monocyte samples from each patient. Results shown were obtained using PCR primers designed to amplify the promoter/enhancer regions of the indicated four genes selected from Table 2. Data shown are the average of samples run in duplicate. One-sided p-values (cases vs. controls) were calculated by Wilcoxon Rank-Sum exact test: *STAT1* p=0.018; *TNFα* p<0.001; *IL1A* p=0.007 and *SPI1* p=0.002.
Figure 1. Distribution of the number of regions enriched with the histone PTMs in the case and control groups.

A
H3K9Ac in Monocytes

B
H3K4Me3 in Monocytes

C
H3K9Ac in Lymphocytes

D
H3K4Me3 in Lymphocytes

E
H3K9Me2 in Lymphocytes

* p = 0.0096

p = 0.56

p = 0.14

p = 0.81

p = 0.57

Diabetes
Figure 2. Association of monocyte H3K9 acetylation levels with HbA1c at different time periods.

A

| Year | 1983 | 1993 | 2009/2010 |
|------|------|------|-----------|
| Phase | Pre-DCCT | DCCT | EDIC |

- 1. DCCT Mean
- 2. EDIC Mean
- 3. DCCT/EDIC combined Mean
- 4. Pre-DCCT/DCCT/EDIC combined Mean
- 5. Blood draw

B

Empirical Cumulative Distribution

Pearson Correlation Coefficient
Figure 3. Identification of group-specific promoter regions depicting the largest differences in the levels of histone H3K9Ac in monocytes.
Figure 4. Regions with variations in histone H3K9Ac in monocytes of the cases and control groups, and the biological functions related to the annotated promoters in these regions.
Figure 5. Correlation of HbA1c with acetylation levels at case/control H3K9 hyper-acetylated regions.
Figure 6. Validation of ChIP-chip data of histone H3K9 hyperacetylated regions in cases versus controls.
Online Supplementary Table

Table S1. Summary of the number of regions/genes with the most histone PTM variations between Controls and Cases. *

| Cell Type | PTM   | No Adjustment |             |             | Adjusted for Age+Gender |             |             |
|-----------|-------|---------------|-------------|-------------|-------------------------|-------------|-------------|
|           |       | Control-hyper | Case-hyper  | Control-hyper | Case-hyper           | Control-hyper | Case-hyper |
|           |       | Region # | Gene # | Region # | Gene # | Region # | Gene # | Region # | Gene # | Region # | Gene # |
| Monocyte  | H3K9Ac | 6       | 7     | 35     | 35     | 10      | 8     | 39      | 38     |
| Monocyte  | H3K4Me3| 30      | 34    | 10     | 20     | 32      | 36    | 19      | 28     |
| Lymphocyte| H3K9Ac | 13      | 14    | 0      | 0      | 23      | 25    | 0       | 0      |
| Lymphocyte| H3K4Me3| 4       | 3     | 0      | 0      | 5       | 4     | 0       | 0      |
| Lymphocyte| H3K9Me2| 36      | 38    | 0      | 0      | 24      | 25    | 1       | 1      |

* The top histone-modified regions were identified using linear regression with and without adjusting for age and gender. The regions were considered case or control hyper-modified if the log2 ratio difference was ≥ 1.1 and linear regression one-sided p < 0.05. These regions were annotated to the promoter regions of Refseq genes, which are defined as -3200bp to +800bp of the TSS. Note that one peak can also be annotated to multiple genes.
Legends for the Online Supplemental Figures:

Figure S1. The Study Scheme.
A total of 60 subjects with T1D (30 cases and 30 controls) were selected from the EDIC study based on criteria described in the Methods. Peripheral blood lymphocytes and monocytes were isolated from each individual. Each sample was processed for ChIPs with antibodies to histone modification marks (histone H3K9Ac, H3K4Me3, H3K9Me2). ChIP enriched DNA and no-antibody control samples were amplified and labeled with Cy5 and Cy3 respectively. The Cy5 and Cy3 labeled samples were hybridized to human promoter tiling arrays (Roche Nimblegen), then washed and dried. Hybridizations of ChIP enriched samples with same antibody and from same cell type were all performed at one place, either at the Clinical Microarray Core at University of California, Los Angeles or at the Functional Genomics Core at the Beckman Research Institute of City of Hope. Within each dataset, 60 samples were separated into 5 groups by the EDIC coordination center, each containing 12 samples (balanced with 6 cases and 6 controls) without releasing the details of group IDs and patient information when the samples were processed. Thus each group of samples including equal numbers of case and control samples labeled with masked IDs were hybridized and processed at same time in the same core as 5 batches, to prevent possible bias being introduced when comparing 30 cases and 30 controls. The image of each array was then scanned and processed to generate data profiles in the form of log2 Cy5/Cy3 for each probe. Quality controls and data normalization were performed before subsequent analyses that included identifying case/control hyper-modified regions and analyses of associations between H3K9Ac and HbA1c. The regions identified were further annotated to Refseq genes’ promoters and explored for potential biological functions. Detailed data analysis pipeline is shown in the Online Supplementary Figure S2.

Figure S2. Pipeline used for ChIP-chip data quality controls, data normalization and analysis.

Figure S3. Quality control (QC) assessment protocols for the ChIP-chip data. QC methods applied were 1) reviewing of NimbleGen’s experimental metrics reports to identify potential problems during the array processing; 2) generating MA plots and histogram plots to detect possible dye-bias and evaluate sample enrichment; 3) plotting profile patterns of each histone
PTM at gene promoter regions on gene sets with different expression levels to confirm its relationship with gene expression; and 4) performing cluster analysis of samples to identify outliers based on summarized signal (mean log2 ratio) on defined RefSeq gene promoters. Example plots except experimental metrics report are shown in panel A-D. A. Smoothed MA plot. This is used to detect possible dye bias and to confirm the presence of enrichment using raw pair data. M represents Cy5/Cy3 intensity ratio and A represents the average intensity (‘A’). The dots, each representing one probe, are smoothed so that the higher density regions are darker, whereas lower density regions are lighter. Usually the majority of the probes have no enrichment, hence most of the dots are located closer to M = 0 (i.e., the baseline). Enrichment by antibodies is represented by the probes above the baseline. A lack in this population indicates no enrichment. B. Histogram of Cy5/Cy3 log2 ratios. A heavy right tail indicates good sample enrichment whereas the lack of a heavy right tail or the presence of an obvious left tail suggests poor enrichment. C. Composite profiles of histone modification levels in the form of log2 ratios around the promoter regions. Four gene sets representing different expression were initially generated based on each gene’s average expression level across 12 EDIC samples profiled by Affymetrix Human Gene 1.0ST arrays as described earlier (Miao et. al. J. Biol. Chem.2012;287:16335-16345.). Specifically, transcript clusters were ordered based on their average expression level and stratified into 10 groups whose expression levels range from low to high. The 1<sup>st</sup>, 4<sup>th</sup>, 7<sup>th</sup> and 10<sup>th</sup> groups, each consisting of 1803 genes, were chosen to represent genes that showed none, low, moderate and high expression, respectively. The promoter region of each transcript in a gene set was divided into 100-bp windows. The H3K9 acetylation level within each window was calculated as the average log2 enrichment ratios of the probes falling within it. These calculations resulted in a data matrix where each row represented one transcript and there were 41 columns corresponding to the 41 windows. The average H3K9Ac level of all the transcripts within the gene set on each column was calculated and plotted across the promoter. The plot was used to examine if the correlation of gene expression and histone marks follows known patterns. The anticipated positive correlation between H3K9Ac level and gene expression by composite profile was used to confirm data’s good quality. D. Hierarchical clustering of samples based on Pearson correlation coefficient (R) of histone modification levels at the promoters of all the genes. The average Cy5/Cy3 log2 ratios of all the probes at the promoter regions were used to perform hierarchical clustering of all the samples. This was performed to detect technical outliers by examining the presence of samples in isolated clusters.
Figure S4. Technical reproducibility by cluster analysis. ChIP enriched DNA from lymphocyte samples derived from five subjects were randomly selected. H3K4me3 levels in each one were measured twice using Roche Nimblegen promoter tiling arrays. All the resulting 10 datasets were pooled together and quantile-normalized. H3K4me3 modification levels at 22007 non-redundant RefSeq promoters (-1000bp of TSS to TSS) were subsequently calculated and clustered based on Pearson’s correlations. Rep 1 and Rep2 indicate technical duplicate data from the same sample. Good reproducibility is seen.

Figure S5. Correlations of HbA1c with monocyte H3K9Ac at the 47 differentially hyperacetylated regions separately within the 30 cases (left panel) and 30 controls (right panel). The 47 differentially acetylated regions include 39 case-specific hyperacetylated and 8 control specific hyperacetylated regions. Pearson’s correlation coefficients of acetylation levels at these regions with HbA1c at the different time periods are shown as heatmaps using blue as positive correlation and yellow as negative correlation. The rows represent the case/control hyper-acetylated regions, and the columns represent different time periods (labels on top of the heatmaps). In the main paper we examined the correlations of the HbA1c levels over different periods and at different times with the level of acetylation over all 6248 commonly-acetylated regions among the 60 subjects in the case and control groups combined (Figure 2). The heatmap (Figure 5) then showed the correlations of HbA1c in the combined groups with H3K9Ac at the 47 specific regions that were differentially hyperacetylated between the case and control groups. In this Supplemental Figure S5, the degrees of correlation of HbA1c with H3K9Ac at these 47 differentially hyperacetylated regions separately within the case and control groups were analyzed and presented. Among cases, the levels of acetylation at the 39 annotated case-specific hyperacetylated regions (spanning 38 gene promoters) were positively correlated with the measures of HbA1c, and the correlations among the 8 annotated regions identified in controls tended to be negatively correlated with HbA1c. Among the controls, there was little correlation of the HbA1c levels with the acetylation levels in these 47 regions.

Figure S6. Scatterplot of STAT1 H3K9 acetylation levels assessed by ChIP-chip and ChIP-QPCR in monocyte samples from 60 patients (30 cases and 30 controls). The acetylation level of each sample measured by ChIP-chip was defined as average quantile-normalized log2(ChIP/input) of all the probes located at 2702-2339bp upstream of the TSS of STAT1 and shown on the X-axis. As for the ChIP-QPCR method, standardized ChIP-QPCR were used to
quantify the amount of histone H3K9 acetylation in ChIP-enriched DNA in monocyte samples from individual patients around 2700bp upstream of the TSS. The quantities of ChIP enriched DNA and input DNA of each sample were measured by Qubit assay (Invitrogen) and the same amount of ChIPed DNA and input DNA were used in QPCR. Data in the form of log2(scaled ChIP/input) are shown on the Y-axis. Pearson’s r of the acetylation levels measured by the two methods was subsequently calculated.

Figure S7. Genome-wide correlation of promoter histone modification levels between different datasets in cells from the same individual. For each of the five datasets shown, average histone modification levels at all the 22121 promoters (-1000bp to TSS) were first calculated for each sample. The Pearson’s correlation coefficients of two specified datasets were then calculated for each patient. The resulting 60 (59 for monocyte H3K4Me3 dataset due to one missing sample) coefficients were plotted as X-axis in boxplots where the middle lines within the boxes represent the median value, and the top and bottom whiskers are the maximum and minimum values excluding outliers that are over 1.5× interquantile range. Outliers are represented by black dots. A total of 6 pairs of datasets (which shared the same modification mark or the same cell type) were compared, listed from top to bottom as A) H3K9Ac in monocytes vs. H3K9Ac in lymphocytes; B) H3K4Me3 in monocytes vs. H3K4Me3 in lymphocytes; C) H3K9Ac in monocytes vs. H3K4Me3 in monocytes; D) H3K9Ac in lymphocytes vs. H3K4Me3 in lymphocytes; E) H3K4Me2 in lymphocytes vs. H3K9Me2 in lymphocytes; and F) H3K4Me3 in lymphocytes vs. H3K9Me2 in lymphocytes. The correlations of H3K9Ac levels in monocytes and lymphocytes are the highest (median correlation coefficient of 0.85 across 60 patients), and the correlation of H3K4Me3 between the two cell types is the second highest (median Pearson’s r =0.77). H3K9Ac and H3K4Me3 levels at Refseq promoter regions in monocytes show moderately positive correlations (median Pearson’s r=0.47), while in lymphocytes, the two marks show greater positive correlation (median Pearson’s r = 0.71 across 60 lymphocyte samples). Finally, the correlations of H3K9Me2 with H3K9Ac or H3K4Me3 levels in lymphocytes were, as expected, both poor with median Pearson’s r as 0.26 and 0.39 respectively.

Figure S8. Overlaps of commonly-modified regions between different datasets. For each of the five datasets, namely H3K9Ac and H3K4Me3 in monocytes, and H3K9Ac, H3K4Me3 and H3K9Me2 in lymphocytes, commonly-modified regions were used to examine the correlations
between two datasets. These regions were defined as the genomic regions of at least 350bp in length which contain histone modified regions in at least 15 out of all the 60 (59 for monocyte H3K4Me3 datasets) samples. The total length (kb) of the overlapping regions between the commonly-modified regions of 2 datasets and the total length (kb) of the regions not showing overlap were calculated in kilobase (kb), and the data is shown as Venn diagrams. Six Venn Diagrams are shown to compare either the same modification mark between different cell types or two different marks from the same cell type, as labeled above each circle. The percentages of commonly-modified regions that overlap with the other dataset are shown in parenthesis. The comparisons are: A. monocyte H3K9Ac vs. lymphocyte H3K9Ac; B. monocyte H3K4Me3 vs. lymphocyte H3K4Me3; C. monocyte H3K9Ac vs. monocyte H3K4Me3; D. lymphocyte H3K9Ac vs. lymphocyte H3K4Me3; E. lymphocyte H3K9Ac vs. lymphocyte H3K9Me2; and F. lymphocyte H3K4Me3 and H3K9Me2. It is seen that around 65% H3K9-acetylated regions and 56% H3K4-trimethylated regions in monocytes overlap with those identified in lymphocytes (panels A and B). This suggests that the same histone modification mark (H3K9Ac or H3K4Me3) are likely to have similar genome wide modification status in monocytes and lymphocytes. But these similarities are relatively lesser when comparing the modification levels of two active marks, H3K9Ac and H3K4Me3, in same cell type (either monocyte or lymphocyte), as 28% and 36% of H3K9-acetylated regions were identified as H3K4-trimethylated regions in monocytes and lymphocytes respectively (panels C and D). This suggests that these active marks can also have distinct biological functions. As expected, poor similarity was observed between regions modified with active marks (H3K9Ac or H4K4me3) and the repressive mark (H3K9Me2), with only 8% of H3K9-acetylated regions and 11% H3K4-trimethylated regions overlapping with the H3K9-dimethylated regions in lymphocytes (panels E and F).
Figure S1. The Study Scheme.

1. Lymphocytes or Monocytes from EDIC Patients (Cases & Controls)
2. ChIPed DNA vs. no Antibody Control labeled with Cy5 or Cy3
3. Hybridization to promoter arrays
4. Data Acquisition
5. Quality Control
6. Normalization

- Genome-wide Identification of Histone Modification Enriched Regions
  - Identification of Differentially-modified Regions between Cases and Controls
  - Association of Histone Modification and HbA1c Levels at Commonly-modified Regions
  - Annotation to Gene Promoters (-3200bp to 800bp relative to TSS)
  - In silico analyses. Exploration of Biological Functions
Figure S2. Pipeline used for ChIP-chip data quality controls, normalization and analysis.
Figure S3. Quality control assessment protocols for the ChIP-chip data.

A. MA plot

B. Histogram

C. Composite Profile

D. Cluster Analysis
Figure S4. Technical reproducibility by cluster analysis.
Figure S5. Correlations of HbA1c with monocyte H3K9Ac at the 47 differentially hyperacetylated regions separately within the 30 cases (left panel) and 30 controls (right panel)
Figure S6. Scatterplot of STAT1 H3K9 acetylation levels assessed by ChIP-chip and ChIP-QPCR in monocyte samples from 60 patients (30 cases and 30 controls).
Figure S7. Genome-wide correlations of histone modification levels at promoters of same individuals between different datasets.
Figure S8. Overlaps of commonly-modified regions between different datasets.