Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Raina Mathur Date

DNA Methylation Markers of Type 2 Diabetes among HIV-infected and Uninfected Individuals

By

Raina Mathur
Master of Public Health

Global Epidemiology

_____

Yan V. Sun, Ph.D.
Committee Chair

_____

Vincent C. Marconi, M.D.
Committee Member

DNA Methylation Markers of Type 2 Diabetes among HIV-infected and Uninfected Individuals

By

Raina Mathur

B.A., University of California, Berkeley, 2014

Thesis Committee Chair: Yan V. Sun Ph.D.

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Global Epidemiology
2018

Abstract

DNA Methylation Markers of Type 2 Diabetes among HIV-infected and Uninfected Individuals

By Raina Mathur

Type 2 Diabetes Mellitus (T2DM) has been increasingly diagnosed in HIV/AIDS patients in the United States, resulting in an approximately 3.8% difference in prevalence between HIV positive and HIV negative individuals in 2009-2010. HIV infection and/or HIV treatment could result in epigenetic modification of the T2DM-related loci in the human genome, affecting the risk of developing this disease. In this study, we investigated differential DNA methylation associated with T2DM in both HIV positive and negative populations using data from the Veterans Aging Cohort Study (VACS). We conducted association analyses to replicate five previously reported CpG sites associated with T2DM and to discover novel loci associated with T2DM in HIV positive individuals. DNA methylation level was modeled as a function of T2DM status, controlling for HIV-infection status, current smoking status, chronological age, BMI, and cell type proportions in all study participants and with stratification by HIV-infection status. Interaction effects were assessed between T2DM status and HIV-infection status. Among reported T2DM-associated CpG sites, cg1963031 in *TXNIP*, cg18181703 in *SOCS3*, and cg09152259 in *PROC* were negatively associated with T2DM in the HIV positive individuals, suggesting T2DM status is associated with hypomethylation of these CpG sites. Three novel sites, cg1231141 in *ADAMTS2* (p-value = 6.76×10^{-7}), cg19534769 in *HGFAC* (p-value = 2.09×10^{-6}), and cg13163919 in *TLE3* (p-value = 4.48×10^{-6}) showed suggestive statistical significance in the epigenome-wide association analysis among HIV positive individuals. The T2DM-HIV interaction was associated with cg17862404 in *TSC22D1* with suggestive statistical significance (interaction p-value = 9.87×10^{-7}). Though further validation is warranted, our study identified novel T2DM-associated CpG sites in people living with HIV and suggested modification by HIV-infection. The identified epigenetic associations with T2DM have implications in inflammation, pancreatic β-cell function, and T2DM pathogenesis.

Key words: epigenome; DNA methylation; HIV; Type 2 Diabetes

DNA Methylation Markers of Type 2 Diabetes among HIV-infected and Uninfected Individuals

By

Raina Mathur

B.A., University of California, Berkeley, 2014

Thesis Committee Chair: Yan V. Sun Ph.D.

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Global Epidemiology
2018

Table of Contents

CHAPTER I: BACKGROUND AND SIGNIFICANCE

Approximately 30 million people in the United States are afflicted with all types of diabetes, including Type 1, Type 2, and gestational diabetes (1). Specifically, type 2 diabetes (T2DM) is prevalent among 90-95% of these individuals (1). T2DM is primarily caused by insulin resistance within the body. This occurs when the cells of the muscle, fat, and liver do not properly respond to insulin and therefore do not absorb much glucose from the bloodstream. This ultimately leads to high blood sugar (1). In the early stages of the disease, the beta cells of the pancreas produce extra insulin to overcome the insulin resistance and ensure adequate glucose uptake. With time, however, the beta cells become exhausted and cannot sustain the insulin requirement, resulting in T2DM (2). Common risk factors for T2DM include presence of prediabetes, being overweight, age 45 years or older, having a family member with T2DM, remaining physically inactive, a previous diagnosis of gestational diabetes, and/or African American, Hispanic/Latino American, American Indian, Alaska Native, or Asian American ethnicity (1). Additionally, rising childhood obesity rates are contributing to an increased incidence of T2DM in children and young adults (1). This trend, combined with the fact that T2DM is the leading contributor to retinopathy and blindness, chronic kidney disease, and amputation, has led to a heightened concern over the increasing prevalence of the disease in the United States (1).

Notably, T2DM has been increasingly diagnosed among individuals living with HIV, and researchers have proposed several hypotheses to explain this disposition. One such theory is that effective antiretroviral therapy (ART) has increased the longevity for HIV positive individuals and may be increasing the likelihood of developing age-related

diseases such as T2DM. Another is that various antiretroviral medications, including protease inhibitors and nucleoside reverse transcriptase inhibitors, have been shown to negatively affect glucose metabolism, potentially also increasing the risk for T2DM. A third possible explanation is that HIV and/or HIV treatment could cause epigenetic modification of the T2DM related loci in the human genome thereby increasing the risk of developing this metabolic disease. It likely that one or more of these hypotheses contributes towards increased T2DM risk in HIV infected individuals, however further epidemiologic studies are needed to elucidate the pathogenesis of this metabolic disease.

Genetic Research on Type 2 Diabetes

Genetic research is one method by which researchers have sought to understand the pathophysiology of T2DM. Genome-wide association studies (GWAS) have helped researchers pinpoint novel genetic loci that are associated with increased risk for T2DM. By 2016, there were at least 75 independent genetic loci associated with T2DM found through these types of studies (3). Some of the initial GWAS for T2DM showed that loci *HHEX/IDE, SLC30A8, TCF7L2, PPARG, KCNJ11, FTO, CDKN2A/2B, CDKAL1*, and *IGF2BP2* were all significantly associated with risk of T2DM, with modest odds ratios ranging from 1.10 to 1.40 (3).

To continue the discovery of novel genetic loci associated with T2DM, large scale initiatives such as the DIAbetes Genetics Replication and Meta-analysis study (DIAGRAM) and DIAGRAM + were organized to increase overall sample size in GWA analyses. These studies were primarily composed of individuals of European descent and collectively recruited over 101,000 individuals in sample size (3). Some of the novel associated loci found using these large cohorts included *CDKALI, ADAMTS9, JAFZI,*

CDC123-CAMKID, TSPAN8-LGR5, THADA, and *NOTCH2*. These, and other identified loci, are found throughout the genome, including introns, intergenic regions, missense mutations, and non-coding RNAs (3).

Subsequently, GWAS were conducted in more diverse populations to investigate race specific loci associated with T2DM. The first GWAS results in East Asians showed variants in SNPs of the *KCNQ1* loci that had significant associations with T2DM in Japanese, Korean, and Chinese people. From these two studies, Yasuda et. al and Unoki et. al, identified odds ratios of 1.26 and 1.49 in their study populations (3-5). In 2011, the Asian Genetic Epidemiology Network (AGEN) meta-analysis study published results for eight novel loci, *GLIS3, PEPD, FITM2-R3HDML-HNF4A, KCNK16, MAEA, GCC1-PAX4, PSMD6, and ZFAND* that were associated with T2DM (3, 6). Other studies examining associations in Mexicans and other Latin American ethnicities found that sequence variants in *SLC16A11* and *SLC16A13* were significantly associated with T2DM and that the association was strongest in younger, and leaner individuals (3, 7). In African Americans specifically, GWAS studies using cases with T2DM associated end-stage kidney disease and population-based controls found 37 significant SNPs across eight different loci in total. Among these, variants in *MYH9, APOL1, SFI1*, and *LIMK2* were found to be most significant (8).

Additional GWAS studies were multi-ethnic analyses, using cohorts composed of East Asians, Europeans, South Asians, and Mexicans/Mexican Americans (3, 9). Seven novel loci associated with T2DM risk were discovered through these studies, including *TMEM154, SSR1/RREB1, FAF1, POU5F1/TCF19, LPP, ARL15,* and *MPHOSPH9* (9). Additionally, three variants identified in *KLF14, PEPD,* and *TCG7L2* were found to have

significant heterogeneity in risk allele frequency and odds ratios between the four ethic groups (9).

The majority of the common variants identified through the aforementioned GWAS studies have only modest effects upon T2DM and do not fully explain the effect of T2DM heritability. Consequently, rare or low-frequency variants with large functional effects have been investigated via parallel sequencing, variant calling, and association testing (3). Variants in the *MTNR1B, PPARG, PDX1*, and *PAM* genes resulted in loss of function mutations or frameshift mutations and were found to be significantly associated with increased T2DM risk (10-12). A loss of function variant of *SLC30A8*, a gene involved in zinc transport, was found significantly associated with decreased risk of T2DM (13). Similarly, an intronic low-frequency variant in *CCND2* was shown to decrease the risk of T2DM by half due to increased CCND2 expression (12). Another avenue of analysis to identify novel loci associated with risk of T2DM is glycemic trait analysis, which seeks to find genetic variants associated with quantitative T2DM markers such as fasting glucose, fasting insulin, and HbA1c in those with prediabetes and T2DM (3). Using this method, the Meta-Analysis of Glucose and Insulin-related traits Consortium (MAGIC) study found 53 glycemic trait associated loci, including *MTNR1B, ADCY5, MADD, ADRA2A, CRY2, FADS1, GLIS3, and SLC2A2* genes, all of which are associated with fasting glucose (14). Similarly, the AGEN study identified three variants in or near *PDK1-RAPGEF4, KANK1*, and *IGF1R* that were associated with fasting glucose in East Asians only (15). Studies to identify low or rare frequency variants in the context of the glycemic traits analyses were also performed, yielding variants in *GLP1R,*

URB2, and *G6PC2* which are associated with fasting glucose and fasting insulin levels (16, 17).

Epigenetic research on Type 2 Diabetes

A less well-studied area related to T2DM is one that relates epigenomic profiling to risk for developing this disease. Epigenetic modifications refer to processes that alter gene activity and expression without changing the underlying DNA sequence. These processes include methylation, acetylation, phosphorylation, and ubiquitylation among others. DNA methylation is the most widely studied strategy and is a process by which a methyl group is either added or removed to a G-C dinucleotide (CpG). Methylation of a CpG site downregulates transcription of the associated gene, ultimately decreasing protein expression. Epigenetic studies are key to understanding how environmental factors and various comorbidities may contribute to T2DM susceptibility. Several diabetes-related risk-factors have been shown to cause epigenetic changes in the genome such as age, obesity, physical activity, and diet (3). One of the first epigenome-wide association studies (EWAS) by Toperoff et. al found that known GWAS loci associated with T2DM had differential epigenetic changes via methylation (3, 18). The authors reported that for every 1% decrease in DNA methylation, there was a 6.01% increase in the odds of having T2DM (18). Supporting this hypothesis, Karachanak-Yankova et. al found that there was a 10.4 times increase on average in methyl-CpG-binding domain protein 2 (MBD2) in T2DM afflicted individuals compared to controls. MBD2 is known to mediate the effects of DNA methylation, leading to gene silencing of methylated CpG sites (19). Results from a prospective cohort study showed that individuals with hypomethylation at specific sites were more likely to develop T2DM in the future, and

that differential methylation at a CpG site in *FTO* gene contributed towards higher T2DM risk (18). Gu et al found that *IGFBP7* DNA methylation levels were increased in Swedish men with newly diagnosed T2DM and that *IGFBP7* was associated with development of insulin resistance (20). Interestingly, several other studies have also reported significant results for differential methylation at CpG site cg19693031 in the *TXNIP* locus that is involved in human skeletal muscle glucose uptake (21-24). *TXNIP* encodes thioredoxin interacting protein, which is involved in energy metabolism and regulation of cellular redox balance (24). Chambers et. al not only reported differential methylation at this same site in *TXNIP*, but also found CpG sites cg06500161, cg02650017, cg1818703, and cg11024682 within *ABCG1, PHOSPHO1, SOC3, and SREBF1*, respectively, to be significantly associated with development of T2DM (23). All five of these genes are involved in pathways underlying type 2 diabetes and related metabolic diseases. Studies aiming to replicate results from the Chambers et. al paper published results that were similar. In 2016, Dayeh et al used epigenetic data collected from the Botnia prospective study and replicated results for differential methylation within the same CpG sites of *ABCG1* and *PHOSPHO1* (25). The authors hypothesized that the lack of significant associations between T2DM and *TXNIP, SOC3*, and *SREBF1* was due to lack of a large sample size (25).

The above studies were performed using peripheral blood samples, however epigenetic changes to the human genome have been shown to be tissue specific. Therefore, additional studies have demonstrated significant methylation patterns using other homogenous tissues from T2DM specific organs (3). Pancreatic islet cells are one such source that have been heavily studied. Stitzel et. al reported 34,000 distal regulatory

elements and determined that 47% of these were specific to the pancreatic islet cells being studied (26). Another study by Dayeh et. al published in 2014 showed 1649 differentially methylated CpG sites in 853 genes, including known T2DM associated loci *TCF7L2*, *FTO*, and *KCNQ1* in tissue samples of pancreatic islet cells (27). Furthermore, 102 of these genes that showed differential methylation patterns also had differential gene expression in the T2DM islets, affecting insulin and glucagon secretion functions in the α-cells and β-cells (27). Another type of homogenous tissue source that has been investigated is human skeletal muscle. Individuals with T2DM have shown hypermethylation at CpG sites in *PGC-1α,* and those who regularly exercised showed hypomethylation in promoter regions of *PGC-1α, PDK4,* and *PPAR-δ* (28, 29).

Epigenetic research on HIV

Individuals with HIV infection have increasingly been diagnosed with T2DM. At the end of 2016, the WHO and UNAIDS estimated that 36.7 million people globally were living with HIV, of which 1.8 million people were newly infected (30). Clinical variables such as CD4 counts and viral load quantitation have been monitored to determine the rate of disease progression and effectiveness of ART. Several virologic and host immune factors on a genetic and molecular level have been shown to interact to produce highly variable disease progression dynamics in HIV infected individuals. HIV has a high mutation rate that prevents perfect immunologic control, but for many years the immune system is able to maintain a steady-state viral load set-point (31). Furthermore, viral mutations may result in increased, decreased, or stagnant viral fitness depending upon the state of the infected individual's immune system (31). Immune system exhaustion due to chronic immune activation, dysfunction of effector T cells, and response to bacterial

pathogens that cross into systemic circulation via a compromised gut mucosal barrier are also critical factors influencing progression of the disease to AIDS (31, 32).

Several studies have examined the link between HIV-1 infection and the epigenome to characterize the overall HIV-infected individual's genome. Histone deacetylation and DNA hypermethylation prompted by HIV infection has been shown in several studies to result in inactivity of gene expression in infected cells. When examining the DNA methylation of the host genome, previous studies have found that CpG sites in the promoter region of *FOXP3* were hypermethylated in HIV-infected individuals due to increased expression of DNMT3b in Treg cells (33). The decreased expression of FOXP3 led to a loss of suppressive capacity in the Treg cells and alterations in cytokine secretion, including changes to TGF-β and IL-4 (33). Youngblood and Reich showed explicitly that infection with HIV can induce *DNMT1* promoter activity and expression, ultimately resulting in greater global methylation of the human genome regardless of cell type (34). Similarly, Maricato et. al determined that, following HIV-1 infection, one of the genes encoding the histone methyltransferase, *SETDB2*, was significantly upregulated (35).

Several epigenetic studies in HIV research have focused on genes aside from those involved in methylation of histone acetylation. In other studies, the promoter region of *IL2* was found to be demethylated in memory CD4+ T cells in cells of individuals with chronic HIV infection (35). Using data from the Veterans Aging Cohort Study (VACS), Zhang et. al determined 20 epigenome-wide significant CpG sites from HIV-1 infection, two of which were hypomethylated in the promoter region of the *NLRC5* which regulates major histocompatibility complex class I gene expression (36). Another VACS analysis

by Nelson et al showed that DNAm age is, on average, 11.2 years higher in patients infected with HIV at baseline, indicating that HIV infection may manipulate DNA methylation patterns in certain age-related regions of the human genome (37). Once ART treatment was initiated, however, these methylation patterns were more varied and less predictable (37).

Addressing the Gap: The intersection between Type 2 Diabetes and HIV

Though epigenetic research has been conducted separately for T2DM and HIV alike, little research has been done to investigate the potential connection between epigenetic changes to T2DM and HIV. This gap is especially large when examining the potential epigenetic effect of HIV on T2DM related loci. A few preliminary studies in this area include Butt et. al who used VACS data to investigate the connection between HIV infection and the risk of T2DM (38). The study results showed that HIV infected individuals in the study population had a lower prevalence of diabetes at baseline and that they had a lower risk of T2DM, with an OR of 0.84 (38). The odds ratio of T2DM was also found to be associated with increasing age, male gender, minority race, and high BMI. Additionally, HCV coinfection and nucleoside and non-nucleoside reverse transcriptase inhibitor therapies were associated with higher risk of T2DM (38). Another study published in 2010 showed that up to 80% of the HIV infected individuals treated with protease inhibitors, nucleoside reverse transcriptase inhibitors, and nonnucleoside reverse transcriptase inhibitors developed insulin resistance, as compared to approximately only 2% before administration of any antiretroviral therapy (39).

Though collectively these studies demonstrate the vast array of genetic and epigenetic changes to the human genome that can influence development of T2DM, there

is a clear gap in the scientific literature regarding T2DM EWAS in HIV-infected cohorts. The increased incidence of T2DM in HIV infected populations demonstrates the need to elucidate the genetic and metabolic mechanisms behind T2DM pathogenesis and make a comparison between HIV-infected and uninfected populations. Research contributing to this area will play a critical role in shaping recommendations regarding management of the HIV infection in the long term to avoid complications from age-related comorbidities.

CHAPTER II: MANUSCRIPT

Abstract

DNA Methylation Markers of Type 2 Diabetes among HIV-infected and Uninfected Individuals

By Raina Mathur

Type 2 Diabetes Mellitus (T2DM) has been increasingly diagnosed in HIV/AIDS patients in the United States, resulting in an approximately 3.8% difference in prevalence between HIV positive and HIV negative individuals in 2009-2010. HIV infection and/or HIV treatment could result in epigenetic modification of the T2DM-related loci in the human genome, affecting the risk of developing this disease. In this study, we investigated differential DNA methylation associated with T2DM in both HIV positive and negative populations using data from the Veterans Aging Cohort Study (VACS). We conducted association analyses to replicate five previously reported CpG sites associated with T2DM and to discover novel loci associated with T2DM in HIV positive individuals. DNA methylation level was modeled as a function of T2DM status, controlling for HIV-infection status, current smoking status, chronological age, BMI, and cell type proportions in all study participants and with stratification by HIV-infection status. Interaction effects were assessed between T2DM status and HIV-infection status. Among reported T2DM-associated CpG sites, cg1963031 in *TXNIP*, cg18181703 in *SOCS3*, and cg09152259 in *PROC* were negatively associated with T2DM in the HIV positive individuals, suggesting T2DM status is associated with hypomethylation of these CpG sites. Three novel sites, cg1231141 in *ADAMTS2* (p-value = 6.76×10^{-7}), cg19534769 in *HGFAC* (p-value = 2.09×10^{-6}), and cg13163919 in *TLE3* (p-value = 4.48×10^{-6}) showed suggestive statistical significance in the epigenome-wide association analysis among HIV positive individuals. The T2DM-HIV interaction was associated with cg17862404 in *TSC22D1* with suggestive statistical significance (interaction p-value = 9.87×10^{-7}). Though further validation is warranted, our study identified novel T2DM-associated CpG sites in people living with HIV and suggested modification by HIV-infection. The identified epigenetic associations with T2DM have implications in inflammation, pancreatic β-cell function, and T2DM pathogenesis.

Key words: epigenome; DNA methylation; HIV; Type 2 Diabetes

INTRODUCTION

At the end of 2016, the Centers for Disease Control and Prevention estimated that 1,122,900 individuals in the United States were living with HIV (40). Approximately 1 in 7 did not know they had the disease, and 38,782 individuals were newly diagnosed (40). Clinical parameters such as CD4 counts and HIV viral load are monitored to determine the rate of HIV disease progression and effectiveness of antiretroviral therapy (ART). However, several virologic and host immune factors on a genetic and molecular level have been shown to interact to produce highly variable disease progression dynamics in these HIV-infected individuals. As ART has become more effective at lengthening the lives of HIV-infected individuals, they are at greater risk for age-related diseases such as Type 2 Diabetes (T2DM), cardiovascular disease, kidney disease, and various cancers. Notably, T2DM has been increasingly diagnosed in HIV positive patients. Data from the annual Health and Nutrition Survey (NHANES) from 2009-2010 showed that the adjusted prevalence of T2DM amongst HIV infected individuals was 11.8% compared to an adjusted prevalence of 8% in the general population (41). HIV-infected individuals with various metabolic syndromes have notable disturbances in inflammation and adipokines that may contribute towards the pathogenesis of T2DM, including higher levels of C-reactive protein (CRP), leptin, and lower level of adiponectin (42). In a study of prevalent T2DM in HIV-infected and uninfected veterans from the Veterans Aging Cohort Study (VACS) study, Butt et. al found that HIV infection itself was not associated with increased risk of diabetes, however, increasing age, HCV coinfection, and high BMI have a larger effect upon the risk of T2DM among HIV positive veterans (38). Furthermore, long-term ART was reported to increase risk for T2DM (38). Although

previous studies elucidated key risk factors for T2DM among people living with HIV (PLWH), the genetic and molecular mechanisms related to T2DM risk have not been thoroughly investigated to fully understand the development of T2DM among PLHW.

DNA methylation is the most widely studied epigenetic modification of the human genome and is a process by which a methyl group is either added to or removed from a C-G dinucleotide (CpG). Epigenetic studies are key to understanding how environmental factors and comorbidities may contribute to disease susceptibility, including T2DM and HIV infection. Several diabetes-related risk factors were shown to be associated with epigenetic changes in the genome such as age, obesity, physical activity, and diet (3). For example, several studies have reported significant results for differential methylation at CpG site cg19693031 in the *TXNIP* locus that is involved in human skeletal muscle glucose uptake (21-24). Chambers et. al not only reported differential methylation at this same site in *TXNIP*, but also found CpG sites within *ABCG1*, *PHOSPHO1*, *SOC3*, and *SREBF1* to be significantly associated with development of T2DM in a large epigenome-wide association study (EWAS) (23). All of these CpG sites are on or in close proximity to genes that regulate pathways underlying T2DM pathogenesis and related metabolic disorders (23).

HIV has also been shown to cause epigenetic modification in various loci of the human genome. For example, using data from the VACS, Zhang et. al found 20 epigenome-wide significant CpG sites associated with HIV-1 infection, two of which were hypomethylated in the promoter region of *NLRC5* which regulates major histocompatibility complex class I gene expression (36). Another VACS analysis by Nelson et al showed that calculated DNA methylation related age is on average 11.2

years higher than chronological age in patients infected with HIV but untreated at baseline, indicating that HIV infection may manipulate DNA methylation patterns in certain age-related regions of the human genome (37). Based on these data, HIV infection and/or HIV treatment could result in epigenetic modification of the T2DM-related loci in the human genome thereby change the risk of developing this metabolic disease. Therefore, in this study we aimed to investigate differential DNA methylation associated with T2DM in both HIV-infected and uninfected populations.

METHODS

VACS Dataset

Data was obtained from the VACS, a prospective and observational study of HIV-infected and uninfected veterans matched on age, race, ethnicity, sex enrolled from eight Veterans Affairs (VA) facilities around the United States (43). Upon enrollment, study participants completed a demographic survey and a questionnaire detailing comorbidities, tobacco and drug use, height and weight. Laboratory results and information were collected from each patient's electronic medical record. Total white blood cell counts and CD4+ and CD8+ T-cell subsets were analyzed at the time of peripheral blood sample collection. The dataset used in this analysis contains both phenotype and DNA methylation data from 564 HIV-infected and 117 HIV-uninfected veterans. The study was approved by the Veteran's Administration Research and Development Committee and the Institutional Review Board of Atlanta Veteran's Administration. All study participants signed an informed consent.

Phenotypes Data and Analysis

Descriptive statistics of key phenotype variables were obtained while stratifying for HIV status (Table 1). The VACS defined the following criteria as confirmation of Type 2 diabetes status: 1) Glucose ≥ 200 mg/dl on two separate occasions or 2) Glucose ≥ 200 mg/dl on one occasion plus treatment with an oral hypoglycemic or insulin for ≥ 30 days (38). Differences in continuous phenotype variables between HIV infected and uninfected groups were evaluated using Student's t-test, and differences in categorical variables between HIV-infected and uninfected groups were evaluated using a Chi-squared test.

DNA Methylation Assessment and Statistical Analysis

Epigenome-wide DNA methylation were measured on samples of peripheral blood mononuclear cells (PBMCs) using the Illumina 450K platform. Using detection p-value and missing rate >5%, no individuals were removed. However, 927 CpG sites showed a call rate of less than 0.95 and were removed from the final methylation dataset. The data were normalized using subset-quantile within array normalization in the minfi package in R. A total of 35,605 CpG sites were subsequently removed as their probes contained single nucleotide polymorphisms (SNPs) within 10 base pairs of the CpG site. Two diabetes-associated CpG sites, cg06500161 and cg11024682, were reported by Chambers et al, but were excluded from this study because of the SNP overlapping with CpG probes (23). An additional 24,729 CpG sites were removed as their probes mapped to multiple locations in the genome. After the quality control procedure, 412,583 autosomal CpG sites and 11,232 X chromosome CpG sites from 648 individuals remained. Heterogeneity of cell type proportions in blood and other tissues is a well-

established confounder in epigenetic epidemiological studies (44). Using probes from the 450K array that were highly correlated with six cell types ($CD4^+$ T cells, $CD8^+$ T cells, NK T cells, B cells, monocytes, and granulocytes) in the blood and estimated the cell type proportions, and an algorithm developed by Houseman, et al (implemented in the R minfi package), we calculated cell type proportions for each participant (45). These estimated cell type proportions were subsequently adjusted in the epigenetic association analyses.

Several multiple linear regressions were performed to model DNA methylation as a function of diabetes status. First, DNA methylation intensity was modeled as a function of diabetes status, controlling for HIV status, current smoking status, chronological age, BMI, and cell type proportions by pooling all HIV-infected and uninfected individuals. This model also included a random effect for the chip ID and was implemented using the nlme package in R. The same association analysis was also performed by stratifying HIV infection status and controlling for all other confounders previously described. Lastly, another pooled multivariate linear regression was run to assess potential interaction effects between diabetes status and HIV infection status. These three EWAS analyses were run for the X chromosome data as well as previously described (46). For the replication analyses of published T2DM-associated sites, statistical significance was evaluated using a threshold of 0.05 and a multiple-testing corrected p-value was used for the EWAS.

The DNAm age was calculated using the algorithm developed by Horvath *et. al*, using age-related CpG sites from the Illumina Infinium platform (450K or 27K) can be uploaded to the web-based DNAm age calculator (47). The age difference between

chronological age and DNAm age was calculated to represent accelerated epigenetic aging (37).

A multivariate logistical regression analysis was performed to assess whether DNAm age was independently associated with diabetes status. For this analysis diabetes status was modeled as a function of the age difference between chronological and DNAm age. An adjusted model additionally controlling for chronological age was also run.

RESULTS

A total of 681 male participants had both phenotype and DNA methylation data collected and were included for analysis, including 564 in the HIV positive group and 117 in the HIV negative group. The median HIV viral load for the HIV-infected individuals was 75 copies/μL, and the majority of them were on ART at baseline (83.2% and 77.9% respectively). On average, there was a higher proportion of study participants in the HIV negative group with T2DM than in the HIV positive group (35% vs. 20%, χ^2 = 11.1, p-value = 9.0×10^{-4}).

A smoking EWAS was conducted as a positive control analysis, stratified by HIV status and controlling for chronological age. In the HIV positive group, CpG sites cg0557592, cg23576855, cg21161138, and cg26703534 were among the top associations. All of these sites are associated with the *AHRR* gene, a gene which has been repeatedly demonstrated to be significantly associated with smoking status in many EWAS to date. These CpG sites were also statistically significant in the HIV negative strata.

To replicate previously reported T2DM-associated CpG sites, we conducted an association analysis with T2DM, controlling for HIV status, current smoking status,

chronological age, BMI, and cell type proportions. As summarized in Table 2 we observed significant association of cg1963031 (*TXNIP* gene) with a p-value of 5.0×10^{-7}. CpG site cg18181703 (*SOCS3* gene) was also replicated with a p-value of 2.5×10^{-3} Table 2). The regression coefficients for cg1963031 and cg18181703 showed different directionality (β coefficient of -0.023 and 0.013). None of the other three CpG sites under consideration were found to be statistically significant (p-value > 0.05).

In the association analysis stratified by HIV status, the epigenetic association of cg19693031 was replicated in both the HIV-positive and HIV-negative groups (Table 2, p-value = 9.0×10^{-5}, p-value = 3.0×10^{-4} among HIV positive and HIV negative participants, respectively). In the HIV positive group, cg18181703 and cg09152259, associated with *SOCS3* and *PROC* genes, were significantly associated with T2DM (Table 2, p-value = 9.7×10^{-3} and p-value = 3.0×10^{-2}, respectively), however, this association was not significant for the HIV negative group. For CpG sites cg04999691 and cg02650017, the association with T2DM was not significant in either the HIV positive or HIV negative groups. For the stratified analyses of cg1963031, we observed approximately two-fold difference in the β coefficients in the HIV negative comparing to the HIV positive groups (β = -0.042 and β = -0.020), which indicated potential interaction effect of HIV infection. However, a t-test comparing the differences of these two regression coefficients was not statistically significant (t = 1.78, p-value=0.08).

For cg19693031, the interaction analysis showed no statistically significant interaction (p-value=0.41), despite the difference in magnitude of the β coefficients seen in the HIV positive and HIV negative groups. Additionally, none of the other previously reported CpG sites showed statistically significant interaction.

The CpG regional plot for chromosome 1 shown in figure 1 shows that CpG site cg1963031 is located in the middle of the *TXNIP* gene. There were 17 CpG sites tested within the *TXNIP* gene and 136 CpG sites tested within ±200 kb of the flanking region. None of these CpG sites except cg1963031 in the surrounding region were statistically significant in the association analysis.

To discover novel loci associated with T2DM in HIV positive and HIV negative individuals, we conducted an EWAS analysis on all HIV positive and negative individuals in the data set, controlling for chronological age, smoking status, BMI, and HIV infection status, and similar EWAS stratified by HIV infection status. While no CpG sites were epigenome-wide significant in the pooled analysis, five total CpG sites were suggestive of significance using a p-value threshold of 10^{-5} (Table 3). Notably, CpG sites cg01920980 (*ST18* gene) and cg03575666 (*SFRP1* gene) were suggestive of statistical significance (p-value = 9.89 x10^{-6} and p-value = 1.1 x 10^{-5}). An EWAS analysis performed on just the group of HIV positive individuals also showed no epigenome-wide significant CpG sites, when controlling for chronological age, current smoking status, and BMI. Nine CpG sites total passed the aforementioned suggestive threshold, and notable loci include cg1231141 (*ADAMTS2* gene, p-value = 6.76 x 10^{-7}), cg19534769 (*HGFAC* gene, p-value = 2.09 x 10^{-6}), and cg13163919 (*TLE3* gene, p-value = 4.48 x 10^{-6}). Table 4 summarizes these results.

We also conducted an EWAS on all HIV positive and negative individuals controlled for chronological age, current smoking status, BMI, HIV status, and accounted for interaction between T2DM status and HIV status. Again, no CpG sites passed the epigenome-wide significant threshold, however a notable CpG site from this analysis

nearing a suggestive cut-off of 1×10^{-5} was cg17862404 (*TSC22D1* gene), T2DM p-value = 1.18×10^{-5}, T2DM-HIV interaction p-value = 9.87×10^{-7}). When comparing the results from the stratified HIV positive and HIV negative EWAS (Table 5), the regression coefficients for diabetes status were different in both magnitude and direction of effect on DNA methylation (HIV positive $\beta = -1.72 \times 10^{-3}$, p-value = 0.038; HIV negative $\beta = 8.7 \times 10^{-3}$, p-value = 1.61×10^{-4}).

The Manhattan and Q-Q plots for all three described EWAS analyses supported the lack of epigenome-wide significance of any CpG sites. The Q-Q plots for the EWAS analyses without interaction performed on the pooled group and HIV-infected group showed moderate inflation (IF = 1.03, and IF = 1.12, respectively), which did not require further correction for global inflation. The Q-Q plot for the EWAS evaluating interaction effects between T2DM status and HIV status shows deflation, with an inflation factor of 0.87.

The initial positive control analysis on the data showed a strong correlation between chronological age and epigenetic age (r = 0.85). The median chronological age did not differ between the HIV positive and HIV negative groups. However, the median DNAm age in the HIV positive group was 52.9 years and 51.8 years in the HIV negative group. This difference in epigenetic age was statistically significant (p-value = 4.0×10^{-3}). The association analysis showed that the DNA methylation age was not independently associated with T2DM status (p-value = 0.14). This was true in the adjusted model that controlled for chronological age as well.

DISCUSSION

The association study of DNA methylation at the candidate sites and T2DM among both HIV-infected and uninfected participants replicated previously reported T2DM-associated CpG site cg19693031 located in the *TXNIP* gene. The negative regression coefficient related to *TXNIP* indicates that positive T2DM status is associated with hypomethylation at the *TXNIP* CpG site. Furthermore, the difference in magnitude of the regression coefficients in the replication analysis stratified by HIV status indicates that the association of T2DM on DNA methylation at cg1963031 is larger in HIV negative individuals. *TXNIP* encodes thioredoxin-interacting protein that has several different physiological functions (48). This protein is a member of the alpha arrestin protein family that is a regulator of cellular redox signaling that protects cells from oxidative stress, thereby functioning as a tumor suppressor in several cell types (48). Expression of *TXNIP* is inhibited by insulin and stimulated by hyperglycemic conditions within the body (24). In prediabetes and diabetes, TXNIP expression is consistently elevated in the muscle, and expression of the gene is inversely associated with glucose uptake in this cell type (24) Mechanistic studies in animal models have also found that pancreatic beta cell death was vastly decreased in *Txnip* knockout mice (49).

We also replicated the CpG site cg18181703 located in the *SOCS3* gene in the pooled EWAS analysis, controlling for chronological age, BMI, HIV status, and cell types. The effect estimate on DNA methylation indicates that those with T2DM had increased methylation at this particular CpG site. In the stratified EWAS analysis the significance of *SOCS3* was replicated in the HIV positive group, but not the HIV negative group. It is possible that the non-significance of *SOCS3* in the HIV negative

group could be attributed to a lack of statistical power in the strata of the analysis. This gene encodes a member of the suppressor of cytokine signaling family that regulates cytokine signal transduction (50). *SOCS3* inhibits this signaling pathway by binding to tyrosine kinase receptors, notably including insulin receptors. In animal models, *Socs3* knockout mice had comparatively more normal adiposity and energy expenditure and were protected against the development of hyperinsulinemia and insulin resistance (51). Interestingly, *SOCS3* expression in pancreatic beta-cells has been shown to increase the risk for Type 1 Diabetes by interfering with cytokine signaling and inducing beta-cell death (52).

The *PROC* gene was found to be statistically significant in the HIV positive group of VACS participants. The magnitude and direction of the regression coefficient indicates T2DM status was associated with hypomethylation of cg09152259. This gene provides instructions for protein C, an important factor for controlling blood clotting and inflammation (53). Activated protein C, a derivative of protein C, functions in reducing inflammation and apoptosis and prevents pancreatic β-cell death (54).

The pooled EWAS analyses combining both HIV positive and HIV negative VACS participants showed significance of cg01920980, and cg03575666. These CpG sites map to the *ST18* gene, and *SFRP1* gene, respectively. Both of these genes have functions that may play a role in the pathophysiology of T2DM. The *ST18* gene encodes a tumor-suppressor protein associated with various cancers, including breast and liver carcinogenesis (55-57). *ST18* was previously found to be increased in the pancreatic islet cells of obese animals, which stimulates pancreatic β-cell apoptosis (56). Furthermore, increased expression of ST18 was found to decrease insulin secretion in the pancreas

(56). The protein encoded by the *SFRP1* gene is a critical component of the Wnt-Frizzled signaling pathway(58). Changes in regulation of this pathway has been associated with various types of cancers(58, 59). *SFRP1* expression is also associated with determining retina cell differentiation which has implications in progression of diabetic retinopathy (59).

The EWAS in the HIV positive group of veterans only showed suggestive significance of cg1231141 (*ADAMTS2* gene), cg19534769 (*HGFAC* gene), and cg13163919 (*TLE3* gene). The *ADAMTS2* gene encodes enzymes that modify procollagen molecules, precursor molecules to collagens, and ensures that the collagen molecules can assemble in the proper structure (60). While *ADAMTS2* has not specifically been associated with T2DM, other members of the ADAMTS protein family, such as *ADAMTS9*, are found to be associated with T2DM and insulin resistance (61). The *HGFAC* gene encodes the hepatocyte growth factor activator and regulates pancreatic β-cell mass and β-cell function. *HGFAC* expression is regulated by HFN1α, a transcription factor that has been implicated in forms of monogenic diabetes (62). *TLE3* is a transcriptional regulator in the Wnt and PPARγ pathways, and functions in opposing ways to effect adipogenesis. TLE3 interacts with PPARγ to stimulate adipogenesis, while on the other hand it interacts with the Wnt signaling pathway to repress adipocyte proliferation (63). Because of its various effects on adiposity, *TLE3* may be associated with development of T2DM.

The interaction analysis showed suggestive statistical significance of cg17862404 (*TSC22D1* gene). The *TSC22D1* gene encodes a protein that is stimulated by transforming growth factor beta and plays a critical role in pathways involving cancer

cell apoptosis (64) . This gene has also been implicated in protein-protein interactions associated with T2DM and diabetic retinopathy (65).

Several of the CpG sites identified in the pooled EWAS analyses are associated with genes that have implications not only in T2DM but various types of cancers as well. HIV positive individuals have a substantially higher risk for some types of cancers, including AIDS-defining cancers and non-AIDS-defining cancers (66). These notable genes include *TSCD221, ST18, SFRP1,* and *TXNIP*. T2DM is associated with increased risk of various cancers including, liver, colon, breast, bladder, and pancreas (67). Several risk factors for T2DM are also risk factors for cancer, including aging, obesity, diet, and physical activity (67). Other factors such as hyperinsulinemia, hyperglycemia, and inflammation are thought to play a role in development of both T2DM and these types of cancers (67). These CpG sites and genes indicate possible novel connections between HIV and chronic comorbidities candidates for further investigation in HIV positive populations.

To our knowledge, this study is the first to examine epigenetic association with T2DM in HIV infected populations. This analysis has suggested several CpG sites associated with T2DM status that can be further investigated in studies in HIV infected populations. We observed several epigenetic associations with T2DM different between HIV-infected and uninfected participants, which suggested a possible epigenetic mechanism linking to the development of T2DM among PLWH. Additionally, phenotypic data had been collected on most known confounders of the relationship between T2DM status and DNA methylation, including BMI, chronological age, and current smoking status. These confounders were controlled for in the analysis, enhancing

the validity of the results. One limitation of this study is that two of the candidate CpG sites from the study done by Chambers et al were removed from the dataset in data cleaning. Therefore, no conclusions regarding the association of these CpG sites with T2DM in an HIV infected population could be made. Additionally, the data were not collected prospectively, therefore we could not examine the temporal direction of the association between T2DM and DNA methylation. Due to the gender and ethnic composition of the cohort included in this analysis we should be cautious when generalizing the results to the entire population. The study participants in this analysis were all male individuals and, the racial composition of the study participants is heavily weighted in favor of African American participants. Future directions for this study could include female study volunteers and include an even distribution of races/ethnicities. Despite these limitations, this study provides evidence for novel CpG candidate sites that may contribute to differential T2DM risk in HIV infected populations. Future studies in this population can help quantify the associated increased T2DM risk and whether these sites may become useful biomarkers for T2DM risk and progression.

Tables

Table 1. Characteristics of a Cohort of Male Veterans Enrolled in Veterans Aging Cohort Study by HIV Status

	HIV Positive (n=567)	HIV Negative (n=117)	
	Median (IQR) or n (%)	**Median (IQR) or n (%)**	***P*-Value**
Age (years)	52 (10)	52 (11)	0.1
DNA Methylation Age	52.9 (11.5)	51.8 (11.8)	4.0×10^{-3}
Race Group			3.6×10^{-7}
White	58 (10.2%)	30 (25.6%)	
Black	484 (84.9%)	73 (62.4%)	
Hispanic	15 (2.6%)	9 (7.7%)	
Other	13 (2.3%)	5 (4.3%)	
VACS Index	29 (28)		
White Blood Cell Count (no. x 10^9 cell/L)	4.9 (2.35)	6.9 (3.5%)	$<2.2 \times 10^{-16}$
CD4 Count (cells/μL)	389 (335)	1,462 (1077)	3.6×10^{-7}
Proportion of CD4T	0.06 (0.08)	0.15 (0.06)	$<2.2 \times 10^{-16}$
Proportion of CD8T	0.14 (0.1)	0.03 (0.02)	$<2.2 \times 10^{-16}$
Proportion of Natural Killer Cells	0.09 (0.07)	0.06 (0.05)	1.1×10^{-4}
Proportion of B-cells	0.09 (0.06)	0.08 (0.06)	2.2×10^{-4}
Proportion of Monocytes	0.11 (0.05)	0.09 (0.02)	2.4×10^{-5}
Proportion of Granulocytes	0.51 (0.17)	0.58 (0.13)	3.1×10^{-9}
Type 2 Diabetes Status			9.0×10^{-4}
Yes	116 (20.4%)	41 (35.0%)	
No	454 (79.6%)	76 (65.0%)	
Body Mass Index (BMI)	25.1 (5.3)	29.8 (8.6)	2.9×10^{-12}
Glucose (mg/dL)	96 (23)	98 (36)	0.4
A1c	5.7 (1.1)	6 (1.2)	0.9
Smoking Status			0.02
Current Smoker	324 (57.1%)	52 (44.4%)	
Ever Smoker	243 (42.9%)	65 (55.6%)	
HIV Viral Load	75 (3,879)		
On ART at baseline	474 (83.2%)		
On HAART at baseline			
Yes	444 (77.9%)		
No	126 (22.1%)		

Table 2. Replication of Candidate CpG sites in association analysis of Type 2 Diabetes from models including all study participants, stratified by HIV status, and including interaction

CpG Site	Chr.	Gene Name	MAPINFO	Regression Coefficient (SE), Pooled	P-value, Pooled	Regression Coefficient (SE), HIV Positive	P-value, HIV Positive	Regression Coefficient (SE), HIV Negative	P-value, HIV Negative
cg18181703	17	*SOCS3*	76354621	0.013 (0.004)	2.5×10^{-3}	-0.013 (0.005)	9.7×10^{-3}	-0.011 (0.010)	0.26
cg19693031	1	*TXNIP*	145441552	-0.023 (0.005)	5.0×10^{-7}	-0.020 (0.005)	9.0×10^{-5}	-0.042 (0.011)	3.0×10^{-4}
cg02650017	17	*PHOSPHO1*	47301614	-0.003 (0.002)	0.23	-0.002 (0.003)	0.97	-1.4×10^{-4} (0.004)	0.97
cg04999691	7	*C7orf29*	150027050	-0.002 (0.002)	0.16	-0.003 (0.002)	0.15	-4.4×10^{-4} (0.004)	0.91
cg09152259	2	*PROC*	128156114	-0.011 (0.005)	0.19	-0.011 (0.005)	3.1×10^{-2}	-4.0×10^{-3} (0.013)	0.74

Table 3. CpG Sites Passing Suggestive Significance Threshold in EWAS in HIV positive and HIV Negative VACS participants

CpG Site	Chromosome	Gene Name	Regression Coefficient (SE)	P-Value (Unadjusted)
cg19693031	1	*TXNIP*	-0.023 (0.005)	5.4×10^{-7}
cg14721531	6	*CDK19*	-0.007 (0.001)	1.89×10^{-6}
cg08069675	12	*NOC4L*	-0.009 (0.002)	2.40×10^{-6}
cg01920980	8	*ST18*	-0.006 (0.001)	9.89×10^{-6}
cg03575666	8	*SFRP1*	0.020 (0.005)	1.07×10^{-5}

Table 4. CpG Sites Passing Suggestive Significance Threshold in EWAS of T2DM among HIV positive VACS participants

CpG Site	Chromosome	Gene Name	Regression Coefficient (SE)	P-Value (Unadjusted)
cg01231141	5	*ADAMTS2*	0.054 (0.011)	6.76×10^{-7}
cg19534769	4	*HGFAC*	0.010 (0.002)	2.09×10^{-6}
cg13163919	15	*TLE3*	-0.015 (0.003)	4.48×10^{-6}
cg18595258	11	*TMEM138; CYBASC3*	-0.015 (0.003)	4.51×10^{-6}
cg14836313	11	*MRPL23*	0.008 (0.002)	4.55×10^{-6}
cg01505421	22	*CHKB-CPT1B*	-0.004 (0.001)	7.74×10^{-6}
cg20480097	13	*MBNL2*	-0.004 (0.001)	8.77×10^{-6}
cg02246876	14	*CEBPE*	0.009 (0.002)	9.34×10^{-6}
cg01439753	16	*CMIP*	-0.004 (0.001)	1.03×10^{-5}

Table 5. Top five significant CpG sites in the HIV-Interaction EWAS of T2DM and their association results in HIV positive and HIV Negative VACS participants

CpG Site	Chromosome	Gene Name	Diabetes, Regression Coefficient (SE)	P-Value (Unadjusted)	Interaction, Regression Coefficient (SE)	P-Value (Unadjusted)	Diabetes Regression Coefficient (SE), HIV Positive	Diabetes Regression Coefficient (SE), HIV Negative
cg17862404	13	*TSC22D1*	0.007 (0.002)	1.18×10^{-5}	-0.009 (0.002)	9.87×10^{-7}	-0.002 (0.001)	0.009 (0.002)
cg26799209	8	*MSC*	0.022 (0.005)	9.26×10^{-7}	-0.024 (0.002)	2.00×10^{-6}	-0.002 (0.002)	0.027 (0.003)
cg05111645	6	*OR2H1*	-0.114 (0.027)	2.00×10^{-5}	0.140 (0.0050	2.63×10^{-6}	0.024 (0.013)	-0.142 (0.041)
cg11046602	17	*SPOP*	-0.02 (0.005)	4.9×10^{-5}	0.025 (0.030)	4.28×10^{-6}	0.006 (0.003)	-0.026 (0.005)
cg05124190	6		-0.013 (0.003)	7.66×10^{-5}	0.016 (0.0050)	9.81×10^{-6}	0.003 (0.002)	-0.013 (0.003)

Figures

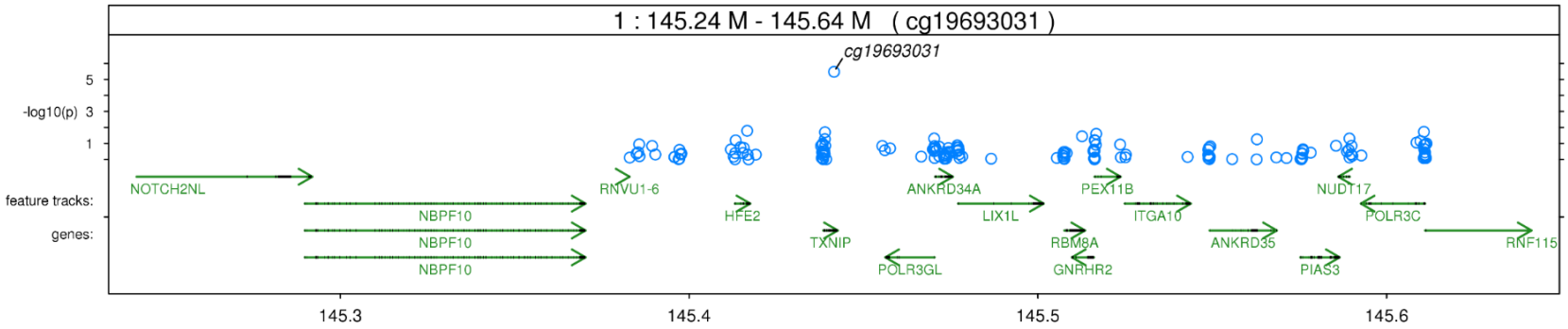

Figure 1. Regional plot of CpG site cg19693031 (*TXNIP* gene) on chromosome 1 in HIV positive VACS participants

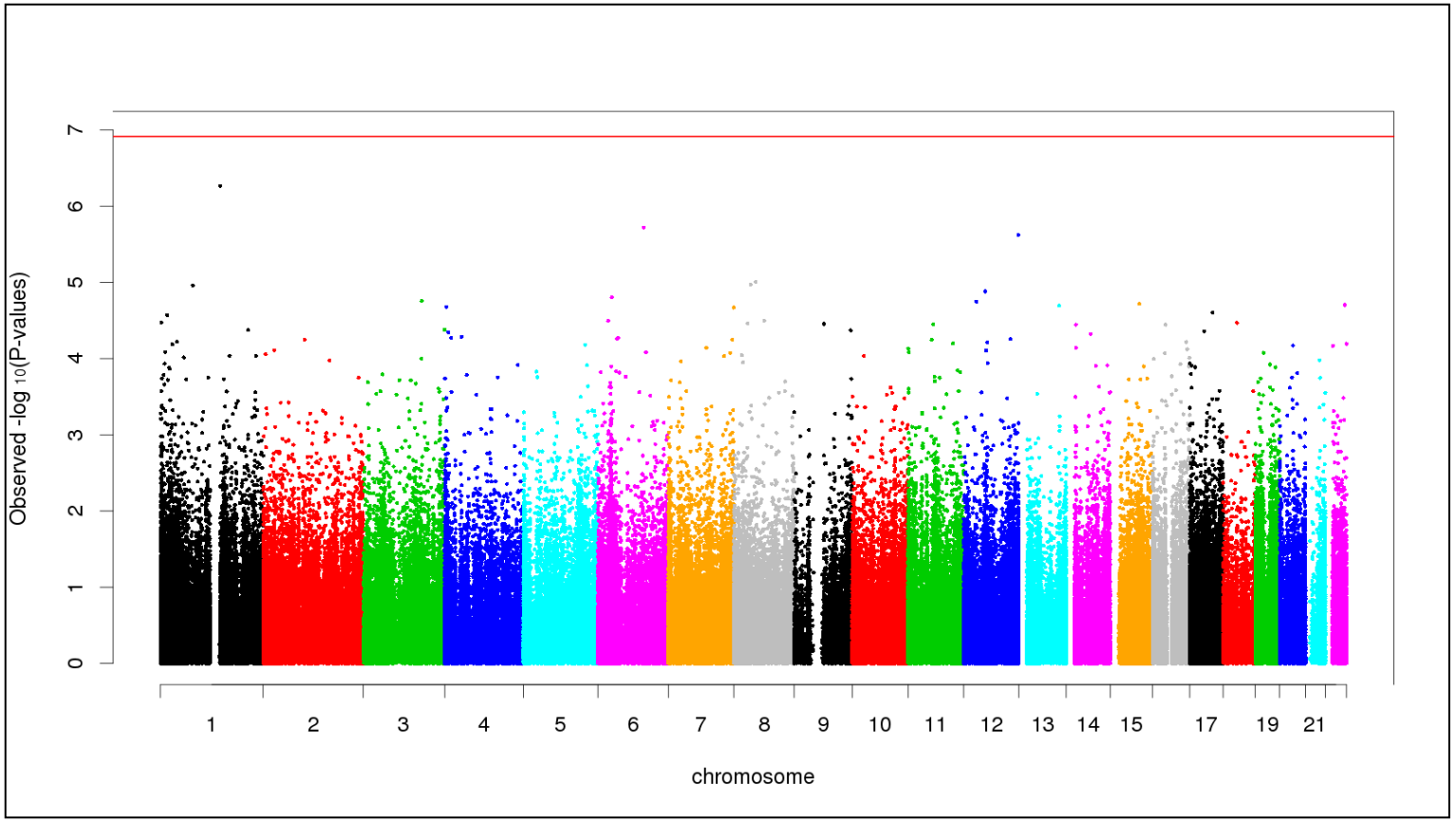

Figure 2. Linear mixed regression p-values of type 2 diabetes term, adjusted for chronological age, current smoking status, HIV status, BMI, and PBL subtypes by chromosome in HIV positive and HIV negative VACS participants

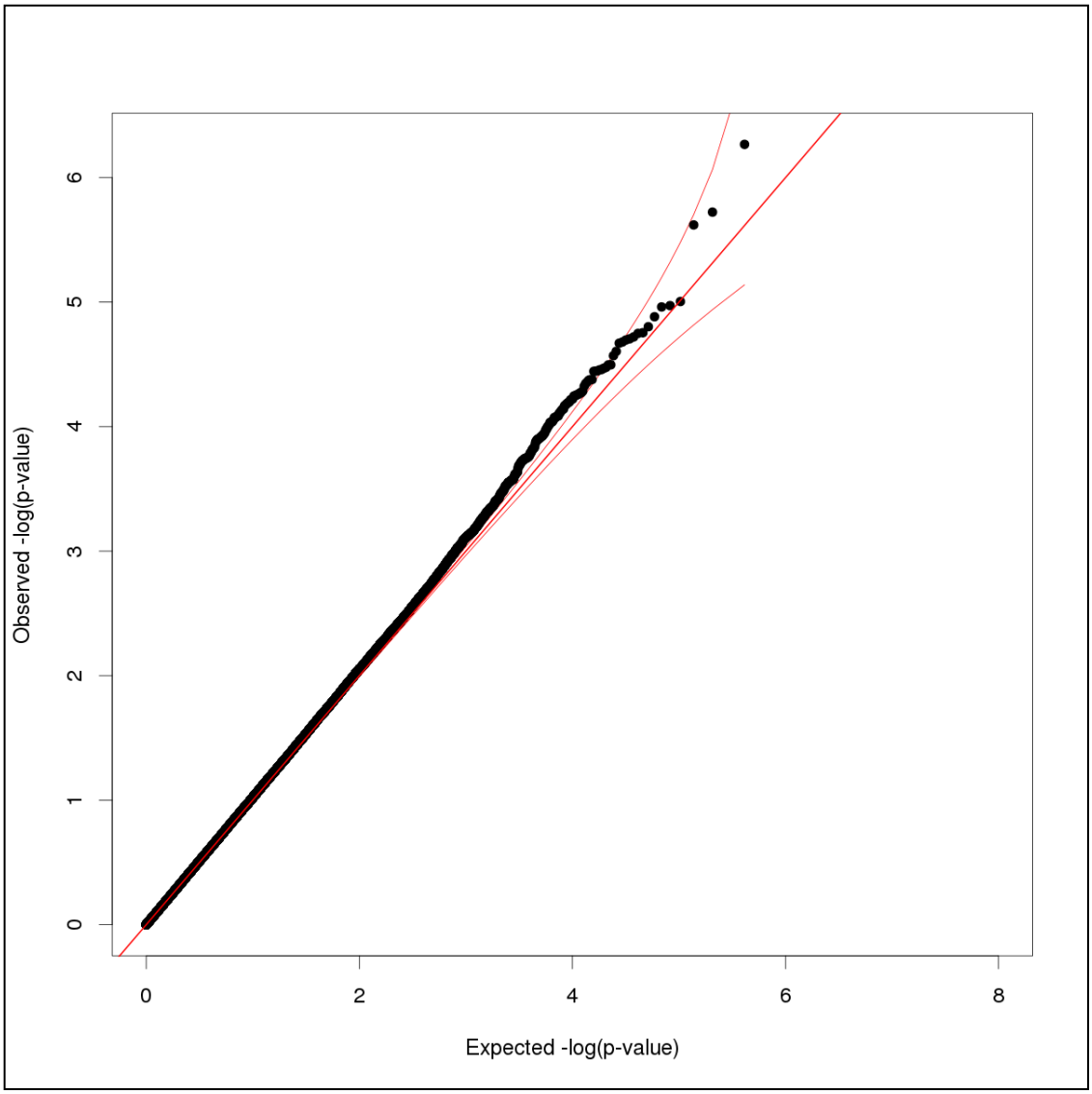

Figure 3. Linear mixed regression p-values of type 2 diabetes term, adjusted for chronological age, current smoking status, HIV status, BMI, and PBL subtypes in HIV positive and HIV negative VACS participants, Inflation factor = 1.03

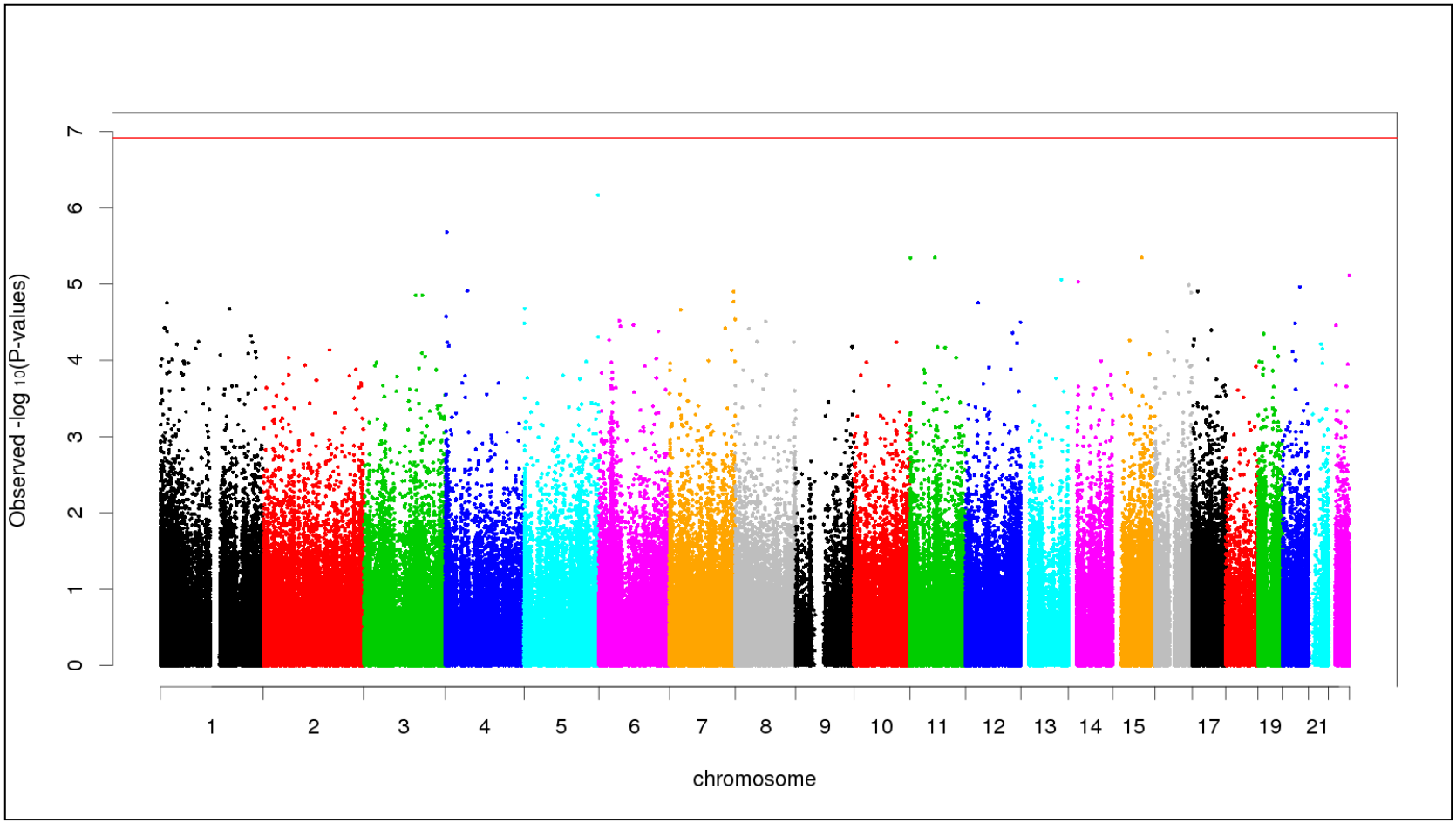

Figure 4. Linear mixed regression p-values of type 2 diabetes term, adjusted for chronological age, current smoking status, BMI, and PBL subtypes by chromosome in HIV positive VACS participants

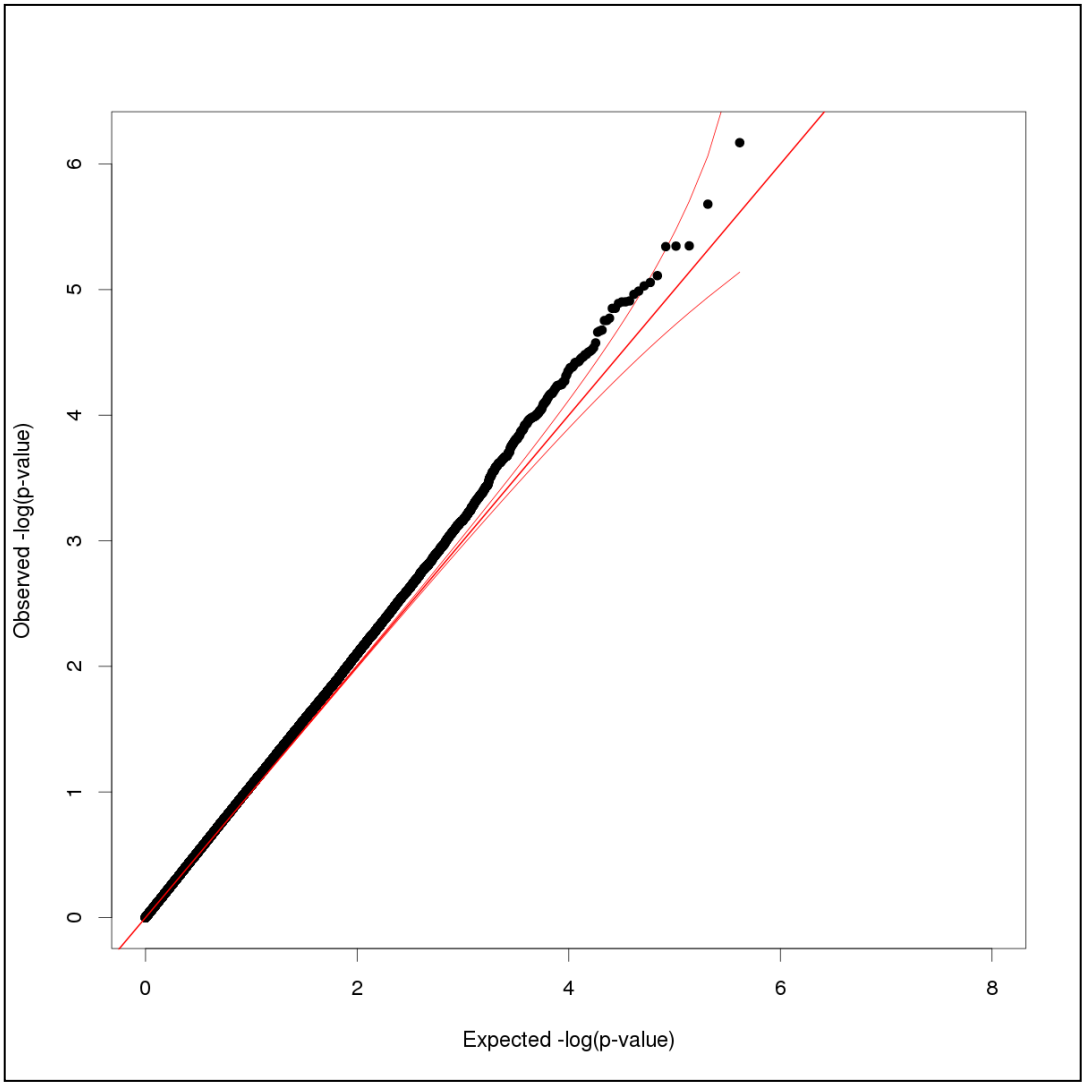

Figure 5. Linear mixed regression p-values of type 2 diabetes term, adjusted for chronological age, current smoking status, BMI, and PBL subtypes in HIV positive VACS participants, Inflation factor = 1.08

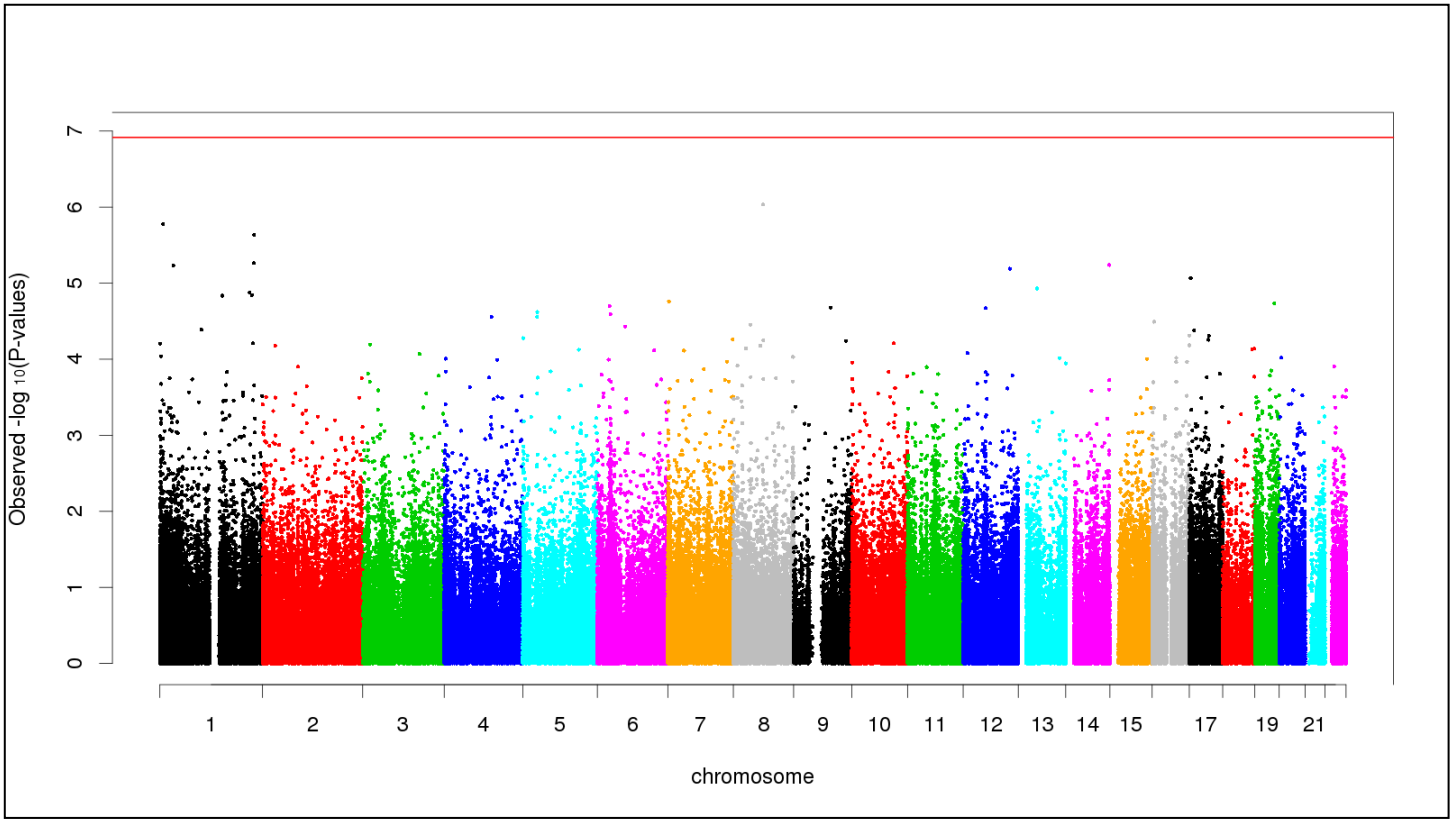

Figure 6. Linear mixed regression p-values of type 2 diabetes term evaluating interaction between HIV status and T2DM status, adjusted for chronological age, current smoking status, BMI, and PBL subtypes by chromosome in HIV positive and HIV negative VACS participants

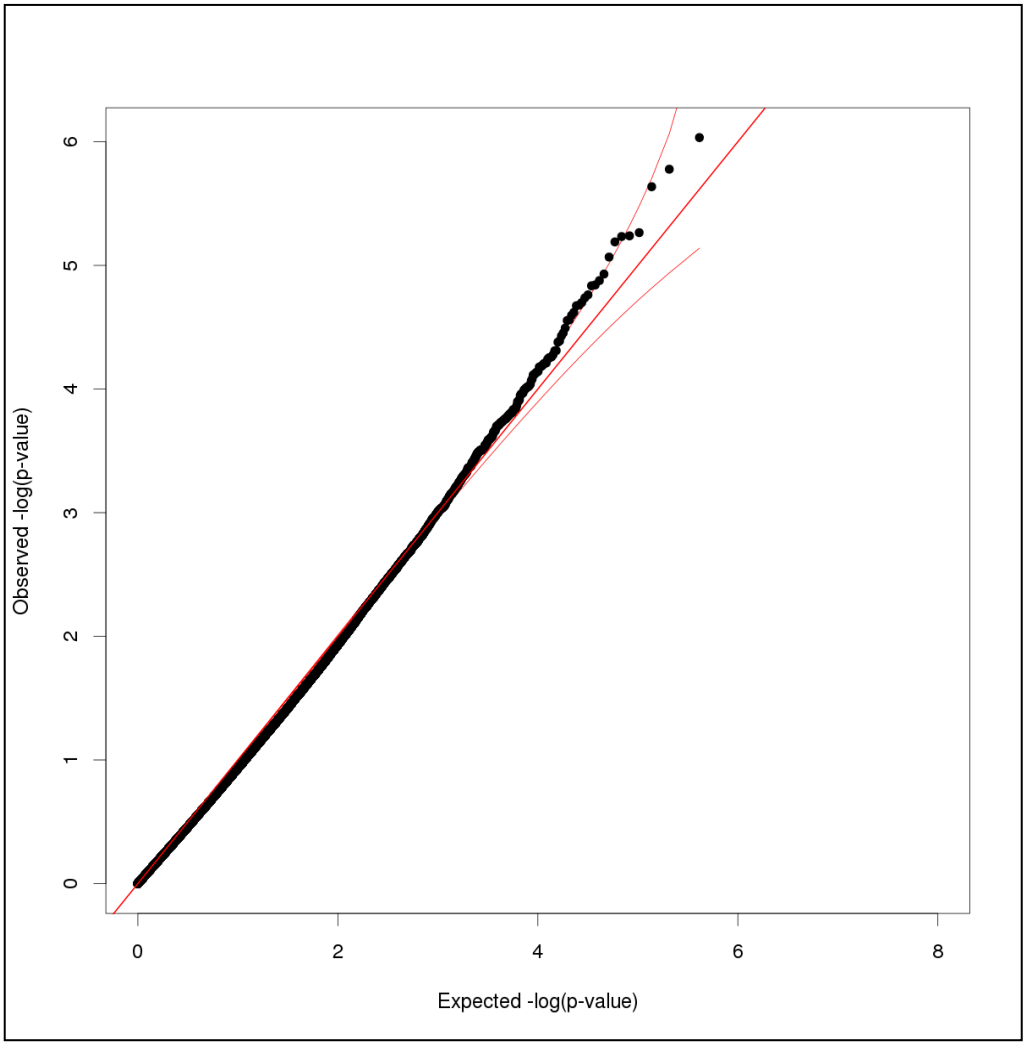

Figure 7. Linear mixed regression p-values of type 2 diabetes term evaluating interaction between HIV status and T2DM status, adjusted for chronological age, current smoking status, BMI, and PBL subtypes in HIV positive and HIV negative VACS participants, Inflation factor = 0.868

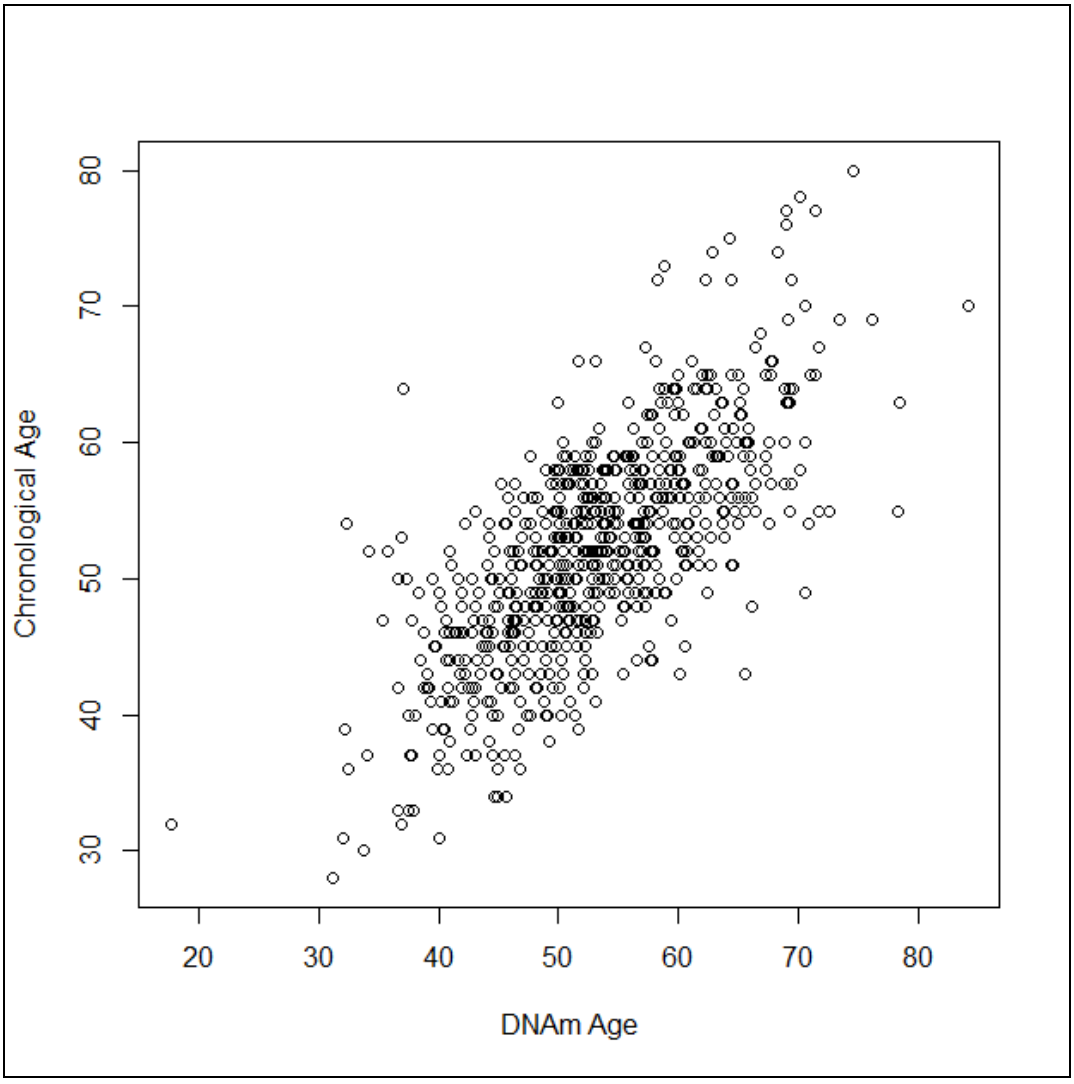

Figure 8. Scatterplot of chronological age against epigenetic age in HIV positive and HIV negative VACS participants, calculated with algorithm developed by Horvath et. al

CHAPTER III: SUMMARY, PUBLIC HEALTH IMPLICATIONS, AND POSSIBLE FUTURE DIRECTIONS

Summary of Results

Previous EWAS studies have identified differential methylation in five CpG sites, cg19693031, cg09152259, cg04999691, cg02650017, and cg18181703 as significantly associated with increased T2DM risk. Data from VACS was used to examine replication results in a cohort of HIV positive individuals and discover novel CpG sites associated with T2DM in HIV infected and uninfected groups.

On average, there was a higher proportion of study participants in the HIV negative group with T2DM than in the HIV positive group (35% vs. 20%, $\chi^2 = 11.1$, p-value = 9.0×10^{-4}). In an association analysis replicating previously reported T2DM-associated CpG sites, DNA methylation was modeled as a function of T2DM status, controlling for HIV status, current smoking status, chronological age, BMI, and cell type proportions. We observed significant association of cg19693031 (*TXNIP* gene) with a p-value of 5.0×10^{-7}. CpG site cg18181703 (*SOCS3* gene) was also replicated with a p-value of 2.5×10^{-3}. The regression coefficients for cg1963031 and cg18181703 showed different directionality (β coefficient of -0.023 and 0.013). None of the other three CpG sites under consideration were found to be statistically significant (p-value > 0.05).

In the association analysis stratified by HIV status, the epigenetic association of cg19693031 was replicated in both the HIV positive and negative groups (p-value = 9.0×10^{-5} and p-value = 3.0×10^{-4} in HIV positive and HIV negative participants, respectively). In the HIV positive group, cg18181703 and cg09152259 (*SOCS3* and *PROC* genes) were significantly associated with T2DM (p-value =9.7×10^{-3} and 3.0×10^{-2} respectively),

however this did not hold true for the HIV negative group, perhaps due to lack of statistical power in the HIV negative study participants. Cg19693031 showed potential interaction effect of HIV infection due to a two-fold difference in the β coefficients between the HIV positive and HIV negative groups. However, a t-test comparing the differences of these two regression coefficients was not statistically significant (t = 1.78, p-value = 0.08).

For cg1963031, the interaction analysis showed no statistically significant interaction (p-value = 0.41), despite the difference in magnitude of the β coefficients seen in the HIV positive and HIV negative groups. Additionally, none of the other previously reported CpG sites showed significant interaction.

To discover novel loci associated with T2DM in HIV positive and HIV negative individuals, we conducted an EWAS analysis on all HIV positive and negative individuals in the data set, controlling for chronological age, smoking status, BMI, and HIV infection status, and similar EWAS stratified by HIV infection status. While no CpG sites were epigenome-wide significant in the pooled analysis, five total CpG sites were suggestive of significance using a p-value threshold of 10^{-5} (Table 3). Notably, CpG sites cg01920980 (*ST18* gene) and cg03575666 (*SFRP1* gene) were suggestive of statistical significance (p-value = 9.89 $x10^{-6}$ and p-value = 1.1 x 10^{-5}). An EWAS analysis performed on just the group of HIV positive individuals also showed no epigenome-wide significant CpG sites, when controlling for chronological age, current smoking status, and BMI. Nine CpG sites total passed the aforementioned suggestive threshold, and notable loci include cg1231141 (*ADAMTS2* gene, p-value = 6.76 x 10^{-7}), cg19534769 (*HGFAC*

gene, p-value = 2.09 x 10^{-6}), and cg13163919 (*TLE3* gene, p-value = 4.48 x 10^{-6}). Table 4 summarizes these results.

We also conducted an EWAS on all HIV positive and negative individuals controlled for chronological age, current smoking status, BMI, HIV status, and accounted for interaction between T2DM status and HIV status. Again, no CpG sites passed the epigenome-wide significant threshold, however a notable CpG site from this analysis nearing a suggestive cut-off of 1 x 10^{-5} was cg17862404 (*TSC22D1* gene), T2DM p-value = 1.18 x 10^{-5}, T2DM-HIV interaction p-value = 9.87 x 10^{-7}). When comparing the results from the stratified HIV positive and HIV negative EWAS (Table 5), the regression coefficients for diabetes status were different in both magnitude and direction of effect on DNA methylation (HIV positive β = -1.72 x 10^{-3}, p-value = 0.038; HIV negative β = 8.7 x 10^{-3}, p-value = 1.61 x 10^{-4}).

The Manhattan and Q-Q plots for all three described EWAS analyses supported the lack of epigenome-wide significance of any CpG sites. The Q-Q plots for the EWAS analyses without interaction performed on the pooled group and HIV-infected group showed moderate inflation (IF = 1.03, and IF = 1.12, respectively), which did not require further correction for global inflation. The Q-Q plot for the EWAS evaluating interaction effects between T2DM status and HIV status shows deflation, with an inflation factor of 0.87.

The initial positive control analysis on the data showed a strong correlation between chronological age and epigenetic age (r = 0.85). The association analysis showed that the DNA methylation age was not independently associated with T2DM status (p-value = 0.14). This was true in the adjusted model that controlled for chronological age as well.

Public Health Implications and Possible Future Directions

All of the CpG sites from the replication and discovery EWAS analyses are associated with genes that have known implications in T2DM development. *TXNIP* encodes thioredoxin-interacting protein that has several different physiological functions (48). This protein is a member of the alpha arrestin protein family that is a regulator of cellular redox signaling that protects cells from oxidative stress, thereby functioning as a tumor suppressor in several cell types (48). The *SOCS3* gene encodes a member of the suppressor of cytokine signaling family that regulates cytokine signal transduction (50). *SOCS3* inhibits this signaling pathway by binding to tyrosine kinase receptors, notably including insulin receptors. The *PROC* gene provides instructions for protein C, an important factor for controlling blood clotting and inflammation (53). Activated protein C, a derivative of protein C, functions in reducing inflammation and apoptosis and prevents pancreatic β-cell death (54). The *TSC22D1* gene encodes a protein that is stimulated by transforming growth factor beta and plays a critical role in pathways involving cancer cell apoptosis (64) . This gene has also been implicated in protein-protein interactions associated with T2DM and diabetic retinopathy (65). The *ST18* gene encodes a tumor-suppressor protein associated with various cancers, including breast and liver carcinogenesis (55-57). ST18 was previously found to be increased in the pancreatic islet cells of obese animals, which stimulates pancreatic β-cell apoptosis (56). Furthermore, increased expression of ST18 was found to decrease insulin secretion in the pancreas (56). The protein encoded by the *SFRP1* gene is a critical component of the Wnt-Frizzled signaling pathway(58). Changes in regulation of this pathway has been associated with various types of cancers(58, 59). SFRP1 expression is also associated

with determining retina cell differentiation which has implications in progression of diabetic retinopathy (59). The *ADAMTS2* gene encodes enzymes that modify procollagen molecules, precursor molecules to collagens, and ensures that the collagen molecules can assemble in the proper structure (60). While *ADAMTS2* has not specifically been associated with T2DM, other members of the ADAMTS protein family, such as *ADAMTS9*, are found to be associated with T2DM and insulin resistance (61). The *HGFAC* gene encodes the hepatocyte growth factor activator and regulates pancreatic β-cell mass and β-cell function. HGFAC expression is regulated by HFN1α, a transcription factor that has been implicated in forms of monogenic diabetes (62). TLE3 is a transcriptional regulator in the Wnt and PPARγ pathways, and functions in opposing ways to effect adipogenesis. TLE3 interacts with PPARγ to stimulate adipogenesis, while on the other hand it interacts with the Wnt signaling pathway to repress adipocyte proliferation (63). Because of its various effects on adiposity, TLE3 may be associated with development of T2DM. Differential methylation of CpG sites that effect expression and functions of the proteins encoded by these genes in HIV infected cohorts indicate that these genes and CpG sites could be candidates for replication in subsequent population-level studies.

Several of the CpG sites and associated genes identified in the pooled EWAS have implications not only in T2DM but various types of cancers as well. HIV positive individuals have a substantially higher risk for some types of cancers, including AIDS-defining cancers and non-AIDS-defining cancers (66). These notable genes include *TSCD221, ST18, SFRP1,* and *TXNIP*. Additionally, T2DM is associated with increased risk of various cancers including, liver, colon, breast, bladder, and pancreas (67). Several

risk factors for T2DM are also risk factors for cancer, including aging, obesity, diet, and physical activity (67). Other factors such as hyperinsulinemia, hyperglycemia, and inflammation are thought to play a role in development of both T2DM and these types of cancers (67). These CpG sites and genes indicate possible novel connections between HIV and chronic comorbidities and are candidates for further investigation in HIV infected populations.

This study is one of the first to examine methylation markers associated with T2DM in HIV infected populations. This analysis has identified several novel CpG sites associated with T2DM status that can be candidates for further replication. Additionally, phenotypic data had been collected on most known confounders of the relationship between T2DM status and DNA methylation, including BMI, chronological age, and current smoking status. These confounders were controlled for in the analysis, enhancing the validity of the results.

One limitation of this study is that two of the candidate CpG sites from the study done by Chambers et al were removed from the dataset in data cleaning. Therefore, no conclusions regarding the association of these CpG sites with T2DM in an HIV infected population could be made. The data used in this analysis were also not collected prospectively, therefore nothing can be said about the temporal direction of the association between T2DM and DNA methylation. Additionally, the study participants in this analysis were all male individuals and, the racial composition of the study participants is heavily weighted in favor of African American participants. Due to this gender and ethnic composition of the cohort we should be cautious when generalizing the results to the entire population.

Future directions for this study could target inclusion of female study volunteers and achievement of even distribution of races/ethnicities. Replication of the new CpG sites found from these EWAS analyses could confirm their significance and association with T2DM and could be examined in both HIV infected and HIV uninfected populations separately. Additionally, prospective studies in this HIV infected population can help quantify the associated increased T2DM risk and whether these sites may become useful biomarkers for T2DM risk and progression. The CpG sites associated with genes implicated in oncology studies could be candidates for replication in other cancer cohorts to clearly elucidate the genetic epidemiologic connection. Association of these sites with AIDS defining and non-AIDS defining cancers could also be looked at in HIV infected cohorts.

This study takes a needed step towards clarifying the nature of risk for Type 2 in the HIV infected population. The replication of known CpG sites that increase risk for Type 2 Diabetes in the HIV infected VACS participants indicate that DNA methylation assessment could be used to identify HIV infected individuals who could benefit from early interventions and life style changes that control and monitor blood glucose, BMI, and other phenotypic risk factors for type 2 diabetes. This study also shows that DNA methylation quantitation can provide information beyond the current set of differential clinical biomarkers to quantify T2DM risk in cohorts with special characteristics. As medications for HIV positive individuals evolve and lengthen life spans, genetic studies in this area will help determine the chronic comorbidities for which these individuals are at highest risk.

Bibliography

1. Diabetes Home. Centers for Disease Control and Prevention. (https://www.cdc.gov/diabetes/basics/type2.html). (Accessed April 12 2018).
2. Prediabetes & Insulin Resistance. National Institute of Diabetes and Digestive and Kidney Diseases. (https://www.niddk.nih.gov/health-information/diabetes/overview/what-is-diabetes/prediabetes-insulin-resistance). (Accessed April 12 2018).
3. Kwak SH, Park KS. Recent progress in genetic and epigenetic research on type 2 diabetes. *Experimental & Molecular Medicine* 2016;48(3):e220.
4. Liu Y, Zhou DZ, Zhang D, et al. Variants in KCNQ1 are associated with susceptibility to type 2 diabetes in the population of mainland China. *Diabetologia* 2009;52(7):1315-21.
5. Unoki H, Takahashi A, Kawaguchi T, et al. SNPs in KCNQ1 are associated with susceptibility to type 2 diabetes in East Asian and European populations. *Nature Genetics* 2008;40:1098.
6. Cho YS, Chen C-H, Hu C, et al. Meta-analysis of genome-wide association studies identifies eight new loci for type 2 diabetes in east Asians. *Nature Genetics* 2011;44:67.
7. The STDC. Sequence variants in SLC16A11 are a common risk factor for type 2 diabetes in Mexico. *Nature* 2013;506:97.

8. Palmer ND, Ng MCY, Hicks PJ, et al. Evaluation of Candidate Nephropathy Susceptibility Genes in a Genome-Wide Association Study of African American Diabetic Kidney Disease. *PLOS ONE* 2014;9(2):e88273.

9. Replication DIG, Meta-analysis C, Asian Genetic Epidemiology Network Type 2 Diabetes C, et al. Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nature Genetics* 2014;46:234.

10. Bonnefond A, Clément N, Fawcett K, et al. Rare MTNR1B variants impairing melatonin receptor 1B function contribute to type 2 diabetes. *Nature Genetics* 2012;44:297.

11. Majithia AR, Flannick J, Shahinian P, et al. Rare variants in PPARG with decreased activity in adipocyte differentiation are associated with increased risk of type 2 diabetes. *Proceedings of the National Academy of Sciences* 2014;111(36):13127.

12. Steinthorsdottir V, Thorleifsson G, Sulem P, et al. Identification of low-frequency and rare sequence variants associated with elevated or reduced risk of type 2 diabetes. *Nature Genetics* 2014;46:294.

13. Flannick J, Thorleifsson G, Beer NL, et al. Loss-of-function mutations in SLC30A8 protect against type 2 diabetes. *Nature Genetics* 2014;46:357.

14. Scott RA, Lagou V, Welch RP, et al. Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nature Genetics* 2012;44:991.

15. Hwang J-Y, Sim X, Wu Y, et al. Genome-Wide Association Meta-analysis Identifies Novel Variants Associated With Fasting Plasma Glucose in East Asians. *Diabetes* 2015;64(1):291-8.

16. Mahajan A, Sim X, Ng HJ, et al. Identification and Functional Characterization of G6PC2 Coding Variants Influencing Glycemic Traits Define an Effector Transcript at the G6PC2-ABCB11 Locus. *PLOS Genetics* 2015;11(1):e1004876.

17. Wessel J, Chu AY, Willems SM, et al. Low-frequency and rare exome chip variants associate with fasting glucose and type 2 diabetes susceptibility. *Nature Communications* 2015;6:5897.

18. Toperoff G, Aran D, Kark JD, et al. Genome-wide survey reveals predisposing diabetes type 2-related DNA methylation variations in human peripheral blood. *Human Molecular Genetics* 2012;21(2):371-83.

19. Karachanak-Yankova S, Dimova R, Nikolova D, et al. Epigenetic alterations in patients with type 2 diabetes mellitus. *Balkan Journal of Medical Genetics : BJMG* 2015;18(2):15-24.

20. Gu HF, Gu T, Hilding A, et al. Evaluation of IGFBP-7 DNA methylation changes and serum protein variation in Swedish subjects with and without type 2 diabetes. *Clinical Epigenetics* 2013;5(1):20-.

21. Florath I, Butterbach K, Heiss J, et al. Type 2 diabetes and leucocyte DNA methylation: an epigenome-wide association study in over 1,500 older adults. *Diabetologia* 2016;59(1):130-8.

22. Kulkarni H, Kos MZ, Neary J, et al. Novel epigenetic determinants of type 2 diabetes in Mexican-American families. *Human Molecular Genetics* 2015;24(18):5330-44.

23. Chambers JC, Loh M, Lehne B, et al. Epigenome-wide association of DNA methylation markers in peripheral blood from Indian Asians and Europeans with incident type 2 diabetes: a nested case-control study. *The lancet Diabetes & endocrinology* 2015;3(7):526-34.

24. Parikh H, Carlsson E, Chutkow WA, et al. TXNIP Regulates Peripheral Glucose Metabolism in Humans. *PLOS Medicine* 2007;4(5):e158.

25. Dayeh T, Tuomi T, Almgren P, et al. DNA methylation of loci within ABCG1 and PHOSPHO1 in blood DNA is associated with future type 2 diabetes risk. *Epigenetics* 2016;11(7):482-8.

26. Stitzel ML, Sethupathy P, Pearson DS, et al. Global epigenomic analysis of primary human pancreatic islets provides insights into type 2 diabetes susceptibility loci. *Cell metabolism* 2010;12(5):443-55.

27. Dayeh T, Volkov P, Salö S, et al. Genome-Wide DNA Methylation Analysis of Human Pancreatic Islets from Type 2 Diabetic and Non-Diabetic Donors Identifies Candidate Genes That Influence Insulin Secretion. *PLOS Genetics* 2014;10(3):e1004160.

28. Barrès R, Osler ME, Yan J, et al. Non-CpG Methylation of the PGC-1α Promoter through DNMT3B Controls Mitochondrial Density. *Cell Metabolism*;10(3):189-98.

29. Barrès R, Yan J, Egan B, et al. Acute Exercise Remodels Promoter Methylation in Human Skeletal Muscle. *Cell Metabolism*;15(3):405-11.

30. HIV/AIDS. (http://www.who.int/mediacentre/factsheets/fs360/en/). (Accessed April 12 2018).

31. Khaitan A, Unutmaz D. Revisiting Immune Exhaustion During HIV Infection. *Current HIV/AIDS reports* 2011;8(1):4-11.

32. Troyer RM, Collins KR, Abraha A, et al. Changes in Human Immunodeficiency Virus Type 1 Fitness and Genetic Diversity during Disease Progression. *Journal of Virology* 2005;79(14):9006-18.

33. Pion M, Jaramillo-Ruiz D, Martínez A, et al. HIV infection of human regulatory T cells downregulates Foxp3 expression by increasing DNMT3b levels and DNA methylation in the FOXP3 gene. *AIDS* 2013;27(13):2019-29.

34. Youngblood B, Reich NO. The early expressed HIV-1 genes regulate DNMT1 expression. *Epigenetics* 2008;3(3):149-56.

35. Maricato JT, Furtado MN, Takenaka MC, et al. Epigenetic Modulations in Activated Cells Early after HIV-1 Infection and Their Possible Functional Consequences. *PLOS ONE* 2015;10(4):e0119234.

36. Zhang X, Justice AC, Hu Y, et al. Epigenome-wide differential DNA methylation between HIV-infected and uninfected individuals. *Epigenetics* 2016;11(10):750-60.

37. Nelson KN, Hui Q, Rimland D, et al. Identification of HIV infection-related DNA methylation sites and advanced epigenetic aging in HIV+, treatment-naïve U.S. veterans. *AIDS (London, England)* 2017;31(4):571-5.

38. Butt AA, McGinnis K, Rodriguez-Barradas MC, et al. HIV Infection and the Risk of Diabetes Mellitus. *AIDS (London, England)* 2009;23(10):1227-34.

39. Ryan JG. Increased risk for type 2 diabetes mellitus with HIV-1 infection. *Insulin* 2010;5(1):37-45.

40. HIV in the United States: At A Glance. Division of HIV/AIDS Prevention, National Center for HIV/ADS, Viral Hepatits, STD, and TB Prevention, Centers for Disease Control and Prevention, Centers for Disease Control and Prevention. (Accessed).

41. Harrison P. Diabetes Prevalence High in HIV-Infected Adults in US.

42. Kalra S, Kalra B, Agrawal N, et al. Understanding diabetes in patients with HIV/AIDS. *Diabetology & Metabolic Syndrome* 2011;3:2-.

43. Justice AC, Dombrowski E, Conigliaro J, et al. Veterans Aging Cohort Study (VACS): Overview and Description. *Medical care* 2006;44(8 Suppl 2):S13-S24.

44. Adalsteinsson BT, Gudnason H, Aspelund T, et al. Heterogeneity in White Blood Cells Has Potential to Confound DNA Methylation Measurements. *PLOS ONE* 2012;7(10):e46705.

45. Houseman EA, Accomando WP, Koestler DC, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* 2012;13(1):86.

46. Klebaner D, Huang Y, Hui Q, et al. X chromosome-wide analysis identifies DNA methylation sites influenced by cigarette smoking. *Clinical Epigenetics* 2016;8(1):20.

47. Zhao Y. DNA Methylation Age Calculator. UCLA Department of Human Genetics; 2018. (Accessed 4/8/2018 2018).

48. *TXNIP* Gene. Weizmann Institute of Science. (http://www.genecards.org/cgi-bin/carddisp.pl?gene=TXNIP). (Accessed April 9, 2018).

49. Singh LP. Thioredoxin Interacting Protein (TXNIP) and Pathogenesis of Diabetic Retinopathy. *Journal of clinical & experimental ophthalmology* 2013;4:10.4172/2155-9570.1000287.

50. *SOCS3* Gene.

51. Jorgensen SB, O'Neill HM, Sylow L, et al. Deletion of Skeletal Muscle SOCS3 Prevents Insulin Resistance in Obesity. *Diabetes* 2013;62(1):56-64.

52. Mori H, Shichita T, Yu Q, et al. Suppression of SOCS3 expression in the pancreatic β-cell leads to resistance to type 1 diabetes. *Biochemical and Biophysical Research Communications* 2007;359(4):952-8.

53. PROC gene. NIH U.S. National Library of Medicine. (https://ghr.nlm.nih.gov/gene/PROC). (Accessed April 9, 2018).

54. Xue M, Jackson CJ. Chapter Thirteen - Activated Protein C and Its Potential Applications in Prevention of Islet β-Cell Damage and Diabetes. In: Litwack G, ed. *Vitamins & Hormones*: Academic Press, 2014:323-63.

55. ST18 Gene. Weizmann Institute of Science. (http://www.genecards.org/cgi-bin/carddisp.pl?gene=ST18). (Accessed April 9 2018).

56. Henry C, Close A-F, Buteau J. A Critical Role for the Neural Zinc Factor ST18 in Pancreatic β-Cell Apoptosis. *The Journal of Biological Chemistry* 2014;289(12):8413-9.

57. Ravà M, D'Andrea A, Doni M, et al. Mutual epithelium-macrophage dependency in liver carcinogenesis mediated by ST18. *Hepatology (Baltimore, Md)* 2017;65(5):1708-19.

58. SFRP1 gene. Weizmann Institute of Science. (http://www.genecards.org/cgi-bin/carddisp.pl?gene=SFRP1). (Accessed April 9 2018).

59. Esteve P, Trousse F, Rodríguez J, et al. SFRP1 modulates retina cell differentiation through aβ -catenin-independent mechanism. *Journal of Cell Science* 2003;116(12):2471.

60. ADAMTS2 gene. NIH U.S. National Library of Medicine. (https://ghr.nlm.nih.gov/gene/ADAMTS2). (Accessed April 9 2018).

61. Boesgaard TW, Gjesing AP, Grarup N, et al. Variant near ADAMTS9 Known to Associate with Type 2 Diabetes Is Related to Insulin Resistance in Offspring of Type 2 Diabetes Patients—EUGENE2 Study. *PLOS ONE* 2009;4(9):e7236.

62. Ohki T, Sato Y, Yoshizawa T, et al. Identification of hepatocyte growth factor activator (Hgfac) gene as a target of HNF1α in mouse β-cells. *Biochemical and Biophysical Research Communications* 2012;425(3):619-24.

63. Villanueva CJ, Waki H, Godio C, et al. TLE3 is a dual function transcriptional coregulator of adipogenesis. *Cell metabolism* 2011;13(4):413-27.

64. TSC22D1 Gene. Weizmann Institute of Science. (http://www.genecards.org/cgi-bin/carddisp.pl?gene=TSC22D1). (Accessed April 9, 2018).

65. Rasche A, Al-Hasani H, Herwig R. Meta-Analysis Approach identifies Candidate Genes and associated Molecular Networks for Type-2 Diabetes Mellitus. *BMC Genomics* 2008;9:310-.

66. HIV Infection and Cancer Risk. NIH National Cancer Institute. (https://www.cancer.gov/about-cancer/causes-prevention/risk/infectious-agents/hiv-fact-sheet). (Accessed April 9 2018).

67. Giovannucci E, Harlan DM, Archer MC, et al. Diabetes and Cancer: A consensus report. *Diabetes Care* 2010;33(7):1674-85.