Cell-type specific EWAS identifies genes involved in HIV pathogenesis and oncogenesis among people with HIV infection

Xinyu Zhang[1,2], Ying Hu[3], Ral E. Vandenhoudt[1,2], Chunhua Yan[3], Vincent C Marconi[4,5]

Mardge H. Cohen[6], Amy C Justice [2,7], Bradley E Aouizerat*[8,9], Ke Xu*[1,2]

1. Department of Psychiatry, Yale School of Medicine, New Haven, CT, USA;

2. VA Connecticut Healthcare System, West Haven, CT, USA;

3. Center for Biomedical Information and Information Technology, National Cancer Institute, Rockville, MD, 20850, USA;

4. Division of Infectious Diseases, Emory University School of Medicine;

5. Atlanta Veterans Affairs Healthcare System, Decatur;

6. Department of Medicine, Stroger Hospital of Cook County, Chicago, IL, USA;

7. Department of Internal Medicine, Yale School of Medicine, New Haven, CT;

8. Translational Research Center, College of Dentistry, New York University, New York, NY, USA;

9. Department of Oral and Maxillofacial Surgery, College of Dentistry, New York University, New York, NY, USA.

Corresponding to

Ke Xu, MD, PhD Email: ke.xu@yale.edu

Bradley E Aouizerat, Ph.D. Email: bea4@nyu.edu

Abstract

Epigenome-wide association studies (EWAS) of heterogenous blood cells have identified CpG sites associated with chronic HIV infection, which offer limited knowledge of cell-type specific methylation patterns associated with HIV infection. Applying a computational deconvolution method validated by capture bisulfite DNA methylation sequencing, we conducted a cell type-based EWAS and identified differentially methylated CpG sites specific for chronic HIV infection among five immune cell types in blood: CD4+ T-cells, CD8+ T-cells, B cells, Natural Killer (NK) cells, and monocytes in two independent cohorts (N_{total}=1,134). Differentially methylated CpG sites for HIV-infection were highly concordant between the two cohorts. Cell-type level meta-EWAS revealed distinct patterns of HIV-associated differential CpG methylation, where 67% of CpG sites were unique to individual cell types (false discovery rate, FDR <0.05). CD4+ T-cells had the largest number of HIV-associated CpG sites (N=1,472) compared to any other cell type. Genes harboring statistically significant CpG sites are involved in immunity and HIV pathogenesis (e.g. *CX3CR1* in CD4+ T-cells, *CCR7* in B cells, *IL12R* in NK cells, *LCK* in monocytes). More importantly, HIV-associated CpG sites were overrepresented for hallmark genes involved in cancer pathology (*FDR<0.05*) (e.g. *BCL family, PRDM16, PDCD1LGD, ESR1, DNMT3A, NOTCH2*). HIV-associated CpG sites were enriched among genes involved in HIV pathogenesis and oncogenesis such as Kras-signaling, interferon-α and -γ, TNF-α, inflammatory, and apoptotic pathways. Our findings are novel, uncovering cell-type specific modifications in the host epigenome for people with HIV that

contribute to the growing body of evidence regarding pathogen-induced epigenetic oncogenicity, specifically on HIV and its comorbidity with cancers.

Key words: Deconvoluted cell-type EWAS, Immunity, Oncogenesis, HIV infection

Introduction

With successful antiretroviral therapy (ART), people with HIV (PWH) have a similar lifespan to the general population [1]. However, the healthspan of PWH remains 9 years shorter [1] because of a high burden of comorbid chronic diseases such as cardiovascular diseases [2], diabetes, and non-AIDS-related cancers [3]. The prevalence of non-AIDS-related cancers among PWH is significantly higher compared to that among the people without HIV (PWoH) [4, 5], especially in PWH who are not virally suppressed [6, 7]. Elevated cancer incidence may be due to the high prevalence of cancer risk factors such as substance use and other co-infections [8]. It is important to characterize the underlying mechanisms involved in HIV pathogenesis that may contribute to cancer emergence for PWH.

Host epigenetic modifications play critical roles in HIV-1 induced cellular reprogramming at different stages of HIV-1 pathogenesis, including viral integration, maintenance, activation, or silencing [9]. Upon the integration of HIV-1 into the host genome, chromatin in the infected cells undergoes profound reorganization to control the virus by affecting proviral long terminal repeat (LTR) promoter complex formation [10]. HIV-1 proteins (e.g. Tat) in turn change the cellular environment to facilitate virus survival and replication by

disrupting chromatin structure and altering gene expression in the host cell. For example, expression of histone methyltransferase (*DNMT3A*, *DNMT3B*) and histone deacetylase (*HDAC2* and *HDAC3*) genes are significantly upregulated in host cells infected with HIV-1 [11]. Additionally, although some cells infected with HIV have a normal expression of the BCL family of anti-apoptotic proteins permitting apoptosis and viral propagation, other infected cells overexpress these proteins [12], thereby increasing the risk of cancer and promoting persistence of latently infected cells [13], which poses a barrier to HIV eradication. Epigenetic reprogramming in immune cells persists in chronically infected cells. Such dynamic virus-host genomic interaction results in distinct epigenetic profiles among different immune cell types in response to environmental changes. A common hallmark of pathogen-induced accumulation of DNA methylation maladaptation in multiple genes is an increase in risk for oncogenesis and cancer, which is estimated to occur for as much as 20% of cancers [14]. One example is hepatitis B virus-induced DNA methylation alteration of tumor-suppressor genes *p16*, *p21*, *CDH1*, and *SOCS1* that contributes to hepatocellular carcinoma [15]. However, the role of HIV-associated epigenetic alterations in carcinogenesis has not been explored.

Epigenome-wide association studies (EWAS) of the host methylome have identified numerous significant CpG sites for different stages of HIV infection. During the acute stage of HIV infection, up to 22,697 methylation sites are altered by HIV-1 [16]. ART initiation reverses patterns of DNA methylation in less than 1% of the altered CpG sites, leaving the majority of CpG sites with methylation states persisting into the chronic stage even among virally suppressed individuals [16]. Independently replicated studies have identified several CpG sites and genes associated with HIV infection, including three

NLRC5 promoter CpG sites that are less methylated in PWH with or without ART compared to PWoH [17-20]. DNA methylation profiles are predictive of HIV progression, frailty, and mortality for PWH [21-23]. These findings demonstrate the importance of methylation mechanisms in HIV infection and HIV-related comorbidities.

While EWAS in whole blood have identified CpG sites such as *NLRC5* that have been replicated by different studies, they provide limited insight into the epigenetic modifications of specific immune cell populations that underlie the pathogenic effects of HIV. HIV-1-induced alterations to the epigenomes of specific immune cell types remain unknown. Immune cells that originate from different lineages show distinct DNA methylation patterns [24, 25]. Thus, conventional blood- or peripheral blood mononuclear cell (PBMC)-based EWAS are likely to confound cell-type specific CpG sites for HIV-1 infection. EWAS signals identified from heterogeneous cells are likely to result from consistent CpG methylation signals across cell types or a strong cell-type specific signal that exceeds noise from inconsistent methylation of the same CpG in other cell types. However, only a few studies have conducted EWAS in specific cell types to identify HIV-1 associated CpG sites, and then only in a few cell types and in limited sample sizes. For example, one study showed that the number of CpG sites changed by HIV-1 were approximately 100 times greater in monocytes than CD4+ T-cells in the acute stage of HIV infection [16]. Little is known about the impact of HIV-1 infection on the epigenome of other cell types such as B cells, CD8+ T-cells, natural killer (NK) cells, in addition to monocytes and CD4+ T-cells. DNA methylation in specific cell types plays an important role in responding to, but is ultimately altered by, HIV-1 infection [26, 27].

Cell-type specific DNA methylome profiling of large population samples is technically challenging and cost prohibitive, especially for less abundant cell types in blood. Recently, computational methods have been developed and applied to deconvolute cell-type specific methylation signals from bulk PBMCs or whole blood samples [28-31]. You et al. (2021) successfully dissected several well-known smoking-associated hypomethylation signatures that derived from myeloid lineage immune cells [32].

In this study, we hypothesized that alterations in CpG methylation in the genome of individuals with chronic HIV differ among immune cell types and that HIV-associated CpGs are enriched among genes involved in HIV pathogenesis, but potentially other comorbid conditions including cancer development and progression. We applied a tensor composition analysis (TCA) method to computationally deconvolute cell-type specific methylation data in PBMCs without sorting cells [33]. We first validated the TCA deconvoluted methylomes by direct bisulfite DNA sequencing of sorted CD4+ T-cells, CD8+ T-cells, and monocytes from the same PBMC specimen for a subset of the sample (N=29). We then deconvoluted methylation data from whole blood or PBMCs into CD4+ T-cells, CD8+ T-cells, B cells, NK cells, and monocytes in 1,134 samples. These cell-type specific EWAS for chronic HIV infection were conducted in two cohorts: the Veteran Aging Cohort Study (VACS)[34] for men (N=702) and the Women's Interagency HIV Study (WIHS)[35] for women (N=432) **(Supplemental sTable 1)**. To the best of our knowledge, this is the first and largest cell-type specific EWAS for chronically infected PWH in a predominantly African American population. A flowchart of analytical strategies is presented in **Figure 1**.

Results

Validation of deconvoluted DNA methylome data measured using capture sequencing in CD4+ T-cells, CD8+ T-cells, and monocytes.

We first sought to validate the performance of TCA by comparing the DNA methylomes of CD4+ T-cells, CD8+ T-cells, and monocytes isolated from PBMCs to TCA-deconvoluted DNA methylation data collected employing capture sequencing in a subset of WIHS samples. The methylation value of each CpG site from the bulk (PBMC) sequencing data was deconvoluted to CpG methylation originating from CD4+ T-cells, CD8+ T-cells, and monocytes using TCA. In tandem, a separate aliquot of the same PBMC sample was subjected to cell sorting using magnetic beads and the CD4+ T-cells, CD8+ T-cells, and monocytes obtained were individually subjected to methylation capture sequencing (MC-seq) (**Supplemental Material**).

We compared the methylation β-value for the top 10,000 most-variable CpG sites between the TCA-deconvoluted and the directly measured methylation β-value for CD4+ T-cells, CD8+ T-cells, and CD14+ monocytes. We found a high correlation of methylation values between the two approaches in each cell type. Correlation coefficients for each pair were 0.96 in CD4+ T-cells, 0.97 in CD8+ T-cells, and 0.96 in CD14+ monocytes (**Figure 2a**). Distribution of CpG methylation in each cell type derived by TCA and MC-seq methods were almost identical. Compared to directly measured methylation, a small proportion of hypermethylated CpG sites with β>0.95 trended slightly higher using the TCA method (**Figure 2b**). These results suggest that TCA is a robust and effective

deconvolution method. Accordingly, EWAS for HIV infection using TCA-deconvolution was performed.

Cell type-based EWAS identified differentially methylated positions (DMPs) for HIV infection in men with HIV: the Veteran Aging Cohort Study

In the VACS cohort, the EWAS of HIV-infection using whole blood was carried out by applying a two-step regression model adjusting for age, self-reported race, cigarette smoking, alcohol use, ART adherence, HIV viral load, and the top 30 principal components (PCs) of DNA methylation. We identified 496 epigenome-wide significant (EWS) DMPs associated with HIV infection in the bulk tissue blood DNA methylome (false discovery rate, FDR<0.05) (**Figure 3a**, **Supplemental Figure 1a, sTable 2**). The significant DMPs included those previously reported by us and other groups. Examples include two previously replicated associations on *NLRC5* (cg16411857, t=-9.42, FDR=3.58E-14 and cg07839457, t=-9.05, FDR=5.68E-10) [17, 18, 36]. Hypomethylation of *HCP5* that was previously linked to HIV infection also reached EWS in this study (cg18808777, t=5.61, FDR=7.81E-04) [18]. While replication of these well-validated CpG sites and genes is important, whether these DMPs originate from specific cell types or sub-groups of cell types is unknown.

At the cell-type level, we identified considerably more EWS DMPs across the five cell types than in whole blood: 2,208 in CD4+ T-cells, 106 in CD8+ T-cells, 8 in B cells, 317 in NK cells, and 21 in monocytes (**Figure 3a**, **Supplemental Figure 1b**, **Supplemental sTable 3**-**7**). The majority of DMPs in each cell type differed among the five cell types; a small number of DMPs were common in more than one cell type. For example, one DMP,

LPCAT1 cg16272981*,* was significant in four cell types (CD4+ T-cells: FDR=1.15E-07; B cells: FDR=0.0016; NK cells: FDR=0.005 and monocytes: FDR=3.29E-06). In total, 765 DMPs share more than one cell type in the same direction.

In CD4+ T-cells, differentially methylated CpGs were located in 1,407 genes. We found more hypomethylated (N=1,336) than hypermethylated (N=827) CpG sites in samples from PWH relative to uninfected controls. The HIV-associated loci previously reported in CD4+ T-cells were replicated in this cohort. *LPCAT1* cg16272981 showed the largest effect (16.1% less methylated in samples from PWH relative to PWoH). The top 30 DMPs were located in 15 genes (i.e. *LPCAT1, SLC17A9, RUNX3, KLF7, SEPT9, PEX14, NLRC5, SPOCK2, SPATAS, MYT1L, CAPN11, SEMA3G, BCL9, XYLT1*) (FDR=8.49-09~ 5.21E-06). Some genes harbored multiple DMPs. For example, 4 EWS DMPs were located in *RUNX3*, a well-recognized tumor suppressor of gastric, colon and many other forms of solid tumors [37]. The majority of HIV-associated DMPs in CD8+ T-cells were hypomethylated. Six out of 8 DMPs in B cells were hypomethylated. *LPCAT1* cg16272981, also significantly enriched in CD4+ T-cells, showed the strongest EWS association in B cells (t=-5.97, FDR=0.003). Other significant DMPs were located on *TULP4, ETS1, KCNK9, STAT3, HDAC4*, and *GPC1*. Of note, 6 out of 8 DMPs were common between CD4+ T-cells and B cells except for *ETS1* and *GPC1*. In NK cells, 159 DMP sites overlapped with other cell types. *TULP4* cg02571055 was hypomethylated in B cells and the strongest EWS association in NK cells (t=-6.40, FDR=0.0001). Among 21 significant CpG sites in monocytes, the most significant DMP was *LPCAT1* cg16272981 (t=-6.97, FDR=3.29E-06), followed by *TULP4* cg02571055 (t=-5.99, FDR= 4.94E-04),

which was also identified in B cells and NK cells. Thirteen DMPs were only observed in monocytes, including DMPs located in in *NOTCH4* and *IGSF9.*

Cell-type EWAS for HIV infection in women with HIV: the Women's Interagency HIV Study

The EWAS of women with HIV using PBMCs was carried out by applying the same regression model and adjusting for the same covariates as for the EWAS in the VACS. We identified 13 EWS DMPs associated with HIV infection (**Figure 3b**, **Supplemental Figure 1c**, **Supplemental sTable 8**). Consistent with the VACS sample, *NLRC5* cg07839457 was one of the most significant CpG sites (t=-5.76, p=1.85E-08). *NLRC5* cg16411857 showed near epigenome-wide significance (t=-4.51, p=8.95=-06). Other EWS DMPs were located in *C12orf32*, *CD80*, *GADD45G*, *TXNIP*, *TMEM49*, *SGK269*, *DUSP16*, *RAC2*, *TNIP3*, and *GLB1L2.*

For the cell-type level EWAS, we identified 153 significant DMPs among the 5 cell types: 20 for CD4+ T-cells, 10 for CD8+ T-cells, 22 for B cells, 1 for NK cells, and 100 for monocytes (all FDR<0.05) (**Figure 3b, Supplemental Figure 1d, Supplemental sTable 9-13**). Several DMPs are worthy of mention. *C12orf32* cg12051710 displayed the strongest association in four out of five cell types: CD4+ T-cells (t=-6.85; FDR=1.35E-05), CD8+ T-cells (t=-7.45, FDR= 2.95E-07), B cells (t= -6.28, FDR=2.02E-04), and monocytes (t=-6.94 p= 7.43E-06). *CD80* cg13458803 was hypomethylated in multiple cell types: CD8+ T-cells (t=-5.86, FDR=0.0001), B cells (t=-6.41, FDR=0.0002), and NK cells (t=-5.86, FDR=0.004).Of note, , *NLRC5* cg07839457 was one of the top ranked DMPs in

CD4+ T-cells in the WIHS (t=-5.78, FDR= 0.00537), consistent with the EWAS in the VACS.

Concordance of DMPs for chronic HIV infection in the VACS and WIHS cohorts

The distinct demographic and clinical characteristics of the VACS and the WIHS may undermine whether the HIV-associated DMPs identified in these two cohorts are comparable and if the findings can be generalized to other studies. We conducted a correlation analysis of effect sizes for each DMP across data from bulk samples (whole blood, PBMC) and each cell type between the two cohorts (p<0.001). We found that effect sizes of the same CpG site between the two cohorts were highly correlated in bulk cells (r=0.784, p=5.86E-30) and in four out of five cell types (**Figure 3c**). At the cell type level, the correlation coefficient of DMPs from the VACS and WIHS were strongest in CD4+ T-cells (r=0.635, p=2.14E-30), followed by CD8+ T-cells (r=0.601, p=4.13E-05), B cells (r=0.396, p=1.89E-05), and monocytes (r=0.365, p=2.12E-06). The correlation of effect size in NK cells between the two cohorts was suggestive, but not significant (r=0.151, p=0.099). The directions of the correlations for the majority of DMPs (85.1%) were concordant between the two cohorts. The correlation analysis results show that DMPs for HIV infection in bulk and in individual cell types were largely consistent between the two cohorts; the results also underscored the potential value of EWAS meta-analysis of the two cohorts.

EWAS meta-analysis by cell type identified common and specific DMPs for chronic HIV infection

Cell-type based EWAS meta-analysis (meta-EWAS) identified EWS DMPs in each of five cell types. Prior to performing the meta-EWAS, DMPs with significant heterogeneity (heterogeneity p<0.05) between the two cohorts were removed.

Meta-EWAS in bulk cells revealed 453 DMPs including top significant genes *NLRC5*, *LPCAT1*, *HCP5*, and *PSMB8* (**Figure 4a, Supplemental sTable 14**). Meta-EWAS by cell type identified 1,472 epigenome-wide DMPs in CD4+ T-cells, no DMPs in CD8+ T-cells, 159 DMPs in B cells, 198 DMPs in NK cells, and 422 DMPs in monocytes (**Figure 4b and 4c**, **Supplemental sTable 15-18**). In CD8+ T-cells, the DMPs uniquely identified in the VACS cohort showed opposite direction in the WIHS cohort, with the exception of the DMP in *C12orf32* and *MYL3*. Because the CpG sites with opposite effect were removed from the meta-analysis and no significant DMPs in CD8+ T-cells were identified due to high heterogeneity. Several genomic regions harbored DMPs that were common to more than one cell type **(Figure 5a)**. Multiple loci on chromosome 3 (*ARHGEF3, DNAJB8, CCRL1, AHSG, CD80*), 5 (*LPCAT1*), chromosome 11 (*SHANK2*), chromosome 10 (*RUNX2*), chromosome 16 (*NLRC5*), and chromosome 20 (*SLC17A9*) were common in more than one cell type. Several top ranked CpG sites were located in genes encoding for transcription factors. For example, *ZNF326* (t=-5.92, FDR=0.002), *ZNF714* (t=-4.83; FDR=0.016), *ZNF90* (t=-4.45, FDR=0.042), and *ZNF76* (t=-4.44, FDR=0.044). Other DMPs were located in genes relevant to cancer such as *STRN3* cg18451035 (t=-5.56, FDR=0.004) and genes involved in innate immunity and metabolism such as *ENPP4* cg25606773 (t=-5.48, FDR=0.004) [38, 39].

Overall, the majority of DMPs (67%) from the meta-EWAS were unique to each cell type (**Figure 5b**). Cell-type specific DMPs accounted for 67.6% of DMPs in CD4+ T-cells,

20.8% in B cells, 42.4% in NK cells, and 39.3% in monocytes. Among lymphocytes, only 1.4% of DMPs overlapped between CD4+ T-cells and B cells, and 0.7% of DMPs between CD4+ T-cells and NK cells. The small number of overlapping DMPs may be because fewer DMPs were identified in B cells and NK cells. The results suggest that meta-EWAS revealed distinct DNA methylation modifications for HIV infection between cell types. Annotation of significant CpGs showed that the majority (40-76%) of CpG sites were located in gene bodies in each cell type. The proportion of DMPs located in promoter regions was greater in CD4+ T-cells (10%) than in other cell types, followed by monocytes (8%) (**Figure 6a**). The proportion of DMPs in CpG islands was also greater in CD4+ T-cells (**Figure 6b**).

Cell-type Specific DMPs from meta-EWAS are overrepresented in hallmark genes for cancer

Previous studies have demonstrated that pathogen-induced epigenetic alterations cumulatively contribute to cancer development [40]. Using hallmark genes from the COSMIC Cancer Gene Census (databasehttps://cancer.sanger.ac.uk/census), we found significant overrepresentation of differential methylation of hallmark genes for cancer in CD4+ T-cells (FDR=1.01E-05), in B cells (FDR=0.008), in NK cells (FDR=0.003), and in monocytes (FDR=0.02). Several hallmark genes were EWS for HIV infection from the meta-EWAS in each cell type except CD8+ T-cells (**Table 1**).

In CD4+ T-cells, 41 hallmark genes were differentially methylated for HIV infection (e.g. *BCL9* for B-ALL [41]*, GAS7* and *PRDM16 for* AML [42, 43]*, ESR1* for breast cancer [44]*, and GRIN2A* for colorectal, lung, and gastric carcinoma [45]) (**Figure 7**). Five hallmark

genes were differentially methylated for HIV in B cells (*PRDM16* and *DNMT3A* for AML*,* *PDCD1LG2* for Hodgkin's lymphoma*, LMNA* for spritzed tumor*,* and *BCL11B* for T-ALL). Three hallmark genes were differentially methylated for HIV in NK cells (*BCL11B* for T-ALL*, BCL10* for MALT*, and MAP3K7IP* for prostate cancer), and six in monocytes (*LCK* and *BCL11B* for T-ALL*, PDCD1LGD* for Hodgkin's lymphoma*, PRDM16* for AML*, CACNA1D* for prostate cancer*,* and *EBF1* for B-ALL).

In addition to the enrichment of hallmark genes for cancer, many of the HIV-associated DMPs in individual cell types are also reportedly involved in cancer. Some genes related to cancer harbored multiple CpG sites that all were hypomethylated in samples from PWH compared to PWoH Six DMPs on *SLC17A9* (cg14686919, cg01817521, cg00199007, cg04478428, cg26329715, cg00727912) were hypo-methylated in CD4+ T-cells, B cells, NK cells, and monocytes (**Table 1**). Notably, *SLC17A9* expression is associated with colorectal cancer [46]. We found 4 DMP sites on *RUNX3* that were hypomethylated (cg11585280, cg07236781, cg15498134, cg00147638). *RUNX3* is a tumor suppressor frequently deleted or transcriptionally silenced in cancers that encodes for one of the RUNX family proteins that are critical transcriptional regulators combining with transcription factor CBF-β in CD4+ T-cells [47]. Two *HCP5* DMPs were hypomethylated (cg18808777, cg25843003); *HCP5* is an oncogene associated with multiple cancers. DMPs in several genes from the *BCL* family were found in CD4+ T-cells, B cells, NK cells, and monocytes [48]. We found 5 EWS DMP sites on KLF7 that were hypermethylated. KLF7 affects cell proliferation and has been implicated in ovarian cancer progression [49]. Differential methylation of *KLF7* for HIV was previously linked to pancreatic ductal adenocarcinoma [5]. HIV-1 induced DNA methylation changes to CpGs in these genes,

which have been reported in the literature to play a role in cancer development, may explain the increased prevalence of cancer in PWH.

HIV-associated DMPs are enriched in gene sets for immunity and cancer biology

To better interpret the biological significance of HIV-associated DMPs in the aggregate, we carried out a gene set enrichment analysis. Among the set of hallmark genes, we found 20 significant pathways in CD4+ T-cells, 8 in B cells, 3 for NK cells, and 3 for monocytes (q<0.05). Multiple pathways were identified that were involved in immune evasion in multiple cancers (**Supplemental Table 19**). For example, the Kras pathway harbors a set of one of the most common oncogenic-driven mutations and genes that are targets of cancer therapeutics. In CD4+ T-cells, the Kras pathway contains HIV-associated hypomethylated DMPs in *IRF8, PSMBP8, PDCD1LG2, MYCN, and SNAP91* and hypermethylated DMPs in *BTBD3, CSF2*, and *NIN*. Interestingly, hypomethylation of *IRF8* was enriched in not only the Kras pathway, but also the allograft rejection pathway and the interferon γ response pathway. Multiple genes contained HIV associated DMPs in the allograft rejection pathway, including hypomethylated DMPs in *CD80, CD8A, CRTAM, HLA-DOA, PTPRC, STAT1, STAT4, and CD96* and hypermethylated DMPs in *MBL2* and *CD4*. HIV-associated DMPs in *STAT1, STAT4, HLA-A* and *TAP1* were common to both the allograft rejection pathway and the interferon γ response pathway (**Figure 8a**). The allograft rejection pathway and the Kras pathway are connected by HIV-associated DMPs located in the following hallmark genes for cancer: *HIF1A, CD79A, PDCD1LG2*, and *FLT4* (**Figure 8b**). The three above-mentioned pathways are involved in cancer development and progression.

Gene set enrichment analysis for pathways in B cells mostly overlapped that in CD4+ T-cells. One significant pathway unique to B cells was the IL2-STAT5_SIGNALING pathway, in which 7 of 194 genes harbored HIV-associated DMPs. Gene set enrichment analysis for pathways in NK cells largely overlapped with those identified in B cells. One significant pathway unique in the NK cell was the HALLMARK_PI3K_AKT_MTOR_SIGNALING pathway, in which 7 of 104 genes harbored HIV-associated DMPs. Monocytes had one significant pathway: HALLMARK_PANCREAS_BETA_CELLS. Enrichment analysis from the KEGG database showed overall distinct patterns of significant biological pathways in each cell type **(Supplemental Figure 2)**. The results further underscore the striking enrichment of genes in cell type that feature DMPs for HIV-pathogenesis that are enriched for cancer development among PWH.

Discussion

We identified DMPs for chronic HIV infection in five major immune cell types: CD4+ T-cells, CD8+ T-cells, B cells, NK cells, and monocytes. These include a number of previously reported DMPs associated with HIV infection. Despite differences in demographic and clinical characteristics between the two cohorts studied, we found a number of overlapping and highly concordant DMPs. The majority of DMPs identified in individual cell type meta-EWAS were unique to each cell type (67%). The occurrence of distinct profiles of HIV-associated DMPs among immune cell types highlights the importance of examining differences in DNA methylation profiles between individual cell

types. Among the five cell types, the number of DMPs identified in CD4+ T-cells were three-to-ten-fold greater than the other four cell types, suggesting that epigenetic alteration in CD4+ T-cells plays a major role in chronic HIV-infection. More importantly, we found that genes that harbored HIV-associated DMPs are overrepresented in cancer biology. The identified genes were enriched among hallmark pathways of HIV pathogenesis and cancer. The results provide new insights into the epigenetic mechanisms of HIV that may underlie the increased risk for cancer in PWH.

As expected, we identified the largest number of HIV-associated CpG sites in CD4+ T-cells. Several previously reported genes involved in chronic HIV infection from CD4+ T-cells and from PBMC samples were replicated in this study. Of note, we observed differentially methylated CpG sites harbored in the genes involved in the Th1 signaling process (e.g. *RUNX3, STAT4*) in CD4+ T-cells. Several *TNF* CpG sites were reported to be hypermethylated in samples from PWH [50]. These CpG sites were also hypermethylated in the present study. A noteworthy hypomethylated DMP identified in the present study is *PEX14* cg25310676. PEX14 is involved in the control of oxidative stress and is targeted by HIV Env-mediated autophagy [51]. Expression of *PEX14* was decreased in HIV-infected CD4+ T-cells and contributed to CD4+ T-cell apoptosis [7]. Compared to CD4+ T-cells, a much smaller number of DMPs were identified in the other four cell types, including monocytes. This observation is in line with previous reports showing that different numbers of DMPs between acute and chronic HIV infection are observed in CD4+ T-cells and monocytes. During acute HIV infection, the number of DMPs are 10-fold greater in monocytes than in CD4+ T-cells [16]. In SIV-infected macaques and African green monkeys, only 0.5% of DMPs overlapped in CD4+ T-cells

between acute and chronic HIV infection stages. This evidence suggests that distinct profiles of DNA methylation modification may occur in the different stages of HIV infection and in different cell types.

The fact that many HIV-associated DMPs and associated genes are also involved in cancer among the four cell types is intriguing. One possibility is that HIV-1 directly induces maladaptive changes in epigenetic regulation of oncogenes. For example, several BCL family genes were significantly associated with HIV infection, *BCL9* in CD4+ T-cells, and both *BCL11B and BCL2L2* in CD4+ T-cells, B cells, NK cells, and monocytes. The BCL family plays a crucial role in the development, proliferation, differentiation, and subsequent survival of T cells and is associated with multiple cancers. *BCL9* functions in cell-cell communication in colorectal cancer [52]. *BCL11B* encodes for a protein that is a transcriptional repressor and is regulated by the NURD nucleosome remodeling and histone deacetylase complex. *BCL11B* is a hallmark of B-cell CLL/Lymphoma. *BLC2L2* acts as an apoptotic regulator and is linked to multiple cancers including liver cancer, lung cancer, and breast cancer [53]. On the other hand, evidence shows that chronic inflammation is involved in pathogen-induced DNA methylation changes resulting in an "epigenetic field defect" for oncogenesis. For example, our results show that several proinflammatory genes (e.g. *TNF, IGFBPL1*) were differentially methylated in PWH compared to PWoH. Increased inflammation is a hallmark of chronic HIV infection. Whether chronic HIV-1 results in DNA methylation of inflammatory genes contributing to cancer warrants further study. The overrepresentation of HIV-1 integration in cancer genes has been reported previously [54]. In PWH on suppressive ART, a large proportion of persisting proviruses are found in proliferating cells. One possible mechanism is to

promote the proliferation and survival of latently HIV-infected cells, which in turn benefits HIV-1 persistence and reservoir expansion thereby frustrating attempts to eradicate the virus. Such interactions between HIV-1 and the host epigenome may point to underlying mechanisms of cancer development in PWH.

Overall, we found highly concordant effect sizes among the shared DMPs in four of five cell types between two distinct cohorts. The observation highlights the common biological pathways in ancestrally heterogenous populations among men and women with HIV. On the other hand, we observed a subset of DMPs that differed between the two cohorts. We speculate that the DMPs identified that differed between the two cohorts are likely due to different statical power and different rates of HIV viral suppression between the two cohorts. The sample size of the VACS cohort is over twice that of the WIHS cohort. More importantly, participants in the VACS cohort had higher HIV viral loads than participants in the WIHS cohort. Although HIV viral load was adjusted for in each EWAS, residual effects on DNA methylation cannot be ruled out. Another possibility is biological differences in HIV infection between men and women. Sex differences in HIV infection are observed in clinical settings. Women appear better able to control HIV-1 replication compared to men, typically having lower HIV viral loads, and higher CD8+ and CD4+ T-cell counts [55]. However, the rate of progression to AIDS between men and women are similar, suggesting that immune and inflammatory activation are higher among women [55]. The underlying reasons for the observed differences in CD8+ T-cell epigenome are unclear. One study showed that an increased CD8+ T-cell count may be due to an enhanced capacity to respond to the IL12 cytokine in women compared to men, which leads to more effector cell differentiation [56]. Treatment naïve women with HIV-1 had

significantly higher CD8+ T-cell activation than men, which appears mediated by interferon α in response to Toll-like receptor 7 [57]. Future studies including both sexes and addressing confounding factors are warranted to investigate potential sex differences in the host genome among PWH.

We acknowledge several limitations to this study. Cell type proportion was estimated based on DNA methylation, not cell count, which could result in inaccurate deconvolution of DNA methylation for individual cell types. While the results suggest the effect would be modest, TCA-deconvoluted DNA methylation profiles in each cell type may differ between the two cohorts due to differences in biospecimen collection. For the VACS cohort, cell-type DNAm was deconvoluted from whole blood that included granulocytes while cell-type DNAm from the WIHS cohort was deconvoluted from PBMCs, which excludes granulocytes. Computationally identified significant DMPs warrant confirmation in sorted cell types. Finally, only a small proportion of CpG sites in the methylome were investigated in this study (i.e., 450K and EPIC commercial arrays). Future studies to expand the number of CpG sites using a sequencing platform to comprehensively profile the methylome for chronic HIV infection are warranted.

In summary, leveraging a computational deconvolution approach, we identified cell-type level DPMs associated with HIV infection. The findings were enriched for genes involved in both HIV pathogenesis and cancer pathology, which underscore the important mechanisms of HIV persistence and comorbid cancers. The significant genes may be therapeutic targets for HIV disease and other comorbid medical diseases.

Methods

Sample characteristics and DNA methylation profiling

The VACS is a nationwide longitudinal veteran cohort including PWH and PWoH to study HIV infection and disease progression. A total of 702 samples from the VACS Biomarker Cohort, a subset of the entire VACS, were included in the analysis (**Supplemental Material**). The majority (86%) of the VACS sample were of African (African American/Black; AA) ancestry and all samples were collected from male participants. Clinical data and specimens used in this manuscript were collected by the Women's Interagency HIV Study (WIHS), now the Multicenter AIDS Cohort Study (MACS)/WIHS Combined Cohort Study (MWCCS) [58] **(Supplemental Material)**. WIHS included 245 samples from PWH and 187 uninfected controls from diverse ancestral populations and all samples were collected from female participants. Demographic and clinical characteristics are presented in **Supplemental sTable 1**.

Methylation of bulk DNA samples extracted from whole blood in the VACS was profiled using Illumina HumanMethylation 450K Beadchip. Methylation of bulk DNA samples extracted from PBMCs in the WIHS was profiled using Illumina HumanMethylation EPIC Beadchip. A total of 408,366 CpG sites were common to both the 450K and EPIC arrays, which were used to deconvolute bulk DNA methylation data to five cell types by TCA. More information about DNA methylation quality control and deconvolution are presented below in the Methods and in **Supplementary Material**.

Capture Methylation Sequencing

Three cell types, CD4+ T-cells, CD8+ T-cells, and monocytes, were isolated from 4 PBMC samples using a magnetic bead-based method [59]. DNA was extracted from each isolated cell type. Methylation sequencing target enrichment library preparation was performed per manufacturer protocol (Agilent). Samples were sequenced using 100bp paired-end sequencing on an Illumina HiSeq NovaSeq according to Illumina standard protocol. Detailed quality control and data processes are presented in **Supplementary Material**. CpG sites were annotated using Homer annotatePeaks.pl, including intergenic, 5'UTR, promoter, exon, intron, 3'UTR, transcription start site (TTS), and non-coding categories. CpG island, shore, shelf, and open sea annotation was defined by locally developed bash and R scripts based on genomic coordinates (hg19) of CpG islands from the UCSC genome browser. CpG shore was defined as up to 2 kb from CpG islands and CpG shelf was defined as up to 2 kb from a CpG shore. Methylation CpG sites on the X and Y chromosomes were removed for subsequential analyses.

Deconvolution of DNA methylation from bulk cells to five cell types

To deconvolute bulk methylation of each CpG to specific cell types, TCA requires a DNA methylation data matrix in heterogeneous cells and cell type proportions for each sample in the cohort. We first estimated the proportion of six cell types from the methylation of whole blood in the VACS cohort (CD4+ T-cells, CD8+ T-cells, B cells, NK cells, monocytes, granulocytes) and from the PBMCs in the WIHS cohort (CD4+ T-cells, CD8+ T-cells, B cells, NK cells, monocytes). Because the approach to generating PBMC in the WIHS cohort results in the near-total depletion of granulocytes, we excluded granulocytes from the analyses to provide a consistent set of cell-type specific epigenome profiles shared between the two cohorts. The estimated proportions of each cell type in each

cohort were similar except for CD4+ T-cells, for which the proportion in the VACS cohort was greater than CD4+ T-cells in the WIHS cohort (**Supplemental Figure 3**). We used TCA in the R environment to estimate methylation beta values at each CpG site for each cell type. A total of 408,366 CpG sites were deconvoluted to five cell types using either a whole blood methylation matrix (VACS) or from PBMC methylation matrix (WIHS). Our results showed that TCA deconvoluted methylation in individual cell types robustly removed cell type confronting effects in (**Supplemental Figure 4**).

Comparison between TCA-deconvoluted and methylation capture sequencing-based methylation beta values in three cell types

We validated the accuracy of the TCA-derived estimates of DNA methylation at each CpG site by comparing the TCA-derived beta value with DNA methylation beta estimates measured directly using capture sequencing. We selected the top 10,000 most variable CpG sites among the samples for this comparison, which was performed using Pearson correlation analysis with significance set at p<0.05.

Cell-type based epigenome-wide association analysis

We performed a cell-type based EWAS for HIV-infection using TCA-deconvoluted methylation beta values for each cell type. In each cell type, we conducted a two-step regression analysis using the strategy proposed by Lehne *et al [60]*. The first regression model addressed global covariates that may confound the association of methylation with HIV infection. We first estimated the residual β using regression model ([1]):

$$\beta\ value\ (quantile\ normalized, QN) \sim age + race + smoking\ status + White\ Blood\ Cell\ (WBC)$$
$$+\ 6\ cell\ type\ proportions + PCs\ 1\text{-}30\ on\ intensities\ of\ control\ probes$$

We then performed a second PCA on the resulting regression residual β values and regressed out the first 5 PCs to further control for unmeasured confounders in the regression model ([2]).

$$\beta\ value\ (QN) \sim age + race + smoking\ status + WBC + 6\ cell\ type\ proportions$$
$$+ PCs\ 1\text{-}30\ on\ intensities\ of\ control\ probes$$
$$+ PCs\ 1\text{-}5\ on\ residuals\ from\ model\ (1)$$

Significance was set at a false discovery rate (FDR)<0.05. To further confirm the correction of global confounders, we performed Pearson correlation analysis between the first 30 PCs on residual methylation from model (1) and batch, demographic, clinical, and 6 cell type confounders (**Supplemental Figure 4**). Cutoff of correlation analysis was set at p<0.05

Correlation analysis of HIV-associated CpG sites between the two cohorts

In each cell type, we selected CpG sites with FDR<0.05 in the VACS and with a nominal p<0.05 in the WIHS for correlation analysis. The rationale for the significance cut off chosen for the WIHS cohort is to treat it as a replication cohort. Pearson correlation of the effect sizes at each resulting CpG between the two cohorts was performed. The significance threshold was set at p<0.05. We also compared the direction of effect of each CpG between two cohorts.

Cell type-based Meta-EWAS

We conducted an EWAS meta-analysis for each cell type by combining the data from the VACS and WIHS samples. Effect sizes and p-values for each probe were obtained from

analyses in the VACS and WIHS samples, respectively. We performed inverse-variance weighted meta-analysis, with scheme parameters of sample size and standard error as implemented in the METAL program, combining summary statistics from the two sample sets. We investigated heterogeneity between the two samples using the I^2-statistic. CpG sites with I^2>50% and heterogeneity p<0.05 were excluded from subsequent analysis.

Enrichment of hallmark genes for cancer

We performed an enrichment analysis of HIV-associated genes among 736 cancer hallmark genes from the Cancer Gene Census database. CpG sites with FDR<0.1 from the cell type based meta-EWAS were mapped to the nearest gene. The set of genes that met the above criteria were used to test whether the gene set was significantly overrepresented among the hallmark genes for each cell type at FDR<0.05.

Gene set enrichment analysis

Genes adjacent to CpG sites with FDR<0.1 in meta-EWAS for each cell type were selected for gene set enrichment analysis. We focused on the hallmark gene sets from the Molecular Signature Database (https://www.gsea-msigdb.org/gsea/msigdb/). Enrichment analysis using GO and KEGG annotations were also performed.

Declarations

Ethics approval and consent to participate

The study was approved by the committee of the Human Research Subject Protection at Yale University and the Institutional Research Board Committee of the Connecticut Veteran Healthcare System, and by the participating clinical sites. All subjects provided

written consent. The views and opinions expressed in this manuscript are those of the authors and do not necessarily represent those of the Department of Veterans Affairs or the United States government.

Availability of data and materials

Demographic and clinical variables and DNAm data for the VACS samples were submitted to GEO dataset (GSE117861) and are publicly available. All codes for analysis are also available upon a request to the corresponding author.

Competing interests

VCM has received investigator-initiated research grants (to the institution) and consultation fees (both unrelated to the current work) from Eli Lilly, Bayer, Gilead Sciences, and ViiV. The remaining authors declare that they have no competing interests.

Funding

Authors' contributions

XZ was responsible for data analysis and manuscript preparation. YH and CY were involved analytical strategy and supervised data analysis. RV contributed to manuscript preparation. ACJ provided DNA samples, clinical data, and contributed to interpretation results and manuscript preparation. MHC contributed to data collection. VM involved in interpretation results and manuscript preparation. BA and KX contributed equally to study design, interpretation findings, and manuscript preparation. All co-authors reviewed and approved the manuscript.

Acknowledgements

Legends

Figure 1. Flowchart of analytical strategies. VACS: Veteran Aging Cohort Study; WIHS: Women's Interagency HIV Study; TCA: Tensor Component Analysis; EWAS: Epigenome-wide Association Study.

Figure 2. Benchmarking TCA-deconvoluted cell-type specific DNA methylation. (a) Comparison of methylation β-values for the top 10,000 most variable CpG sites between the deconvoluted and the directly measured methylation. β-values for each cell type were compared between three cell types [CD4+ T cells, CD8+ T cells, and monocytes (CD14+)]; (b) Distribution of genome-wide DNA methylome by the TCA-deconvoluted and MC-seq methods. MC: methylation capture sequencing; TCA: Tensor Composition Analysis.

Figure 3. Summary of cell-type level EWAS in the VACS and WIHS cohorts in five cell types (CD4+ T, CD8+ T, B, Natural Killer, Monocyte). (a) Volcano plots for the VACS cohorts with top common and unique hyper- and hypomethylated gene-associated sites annotated, where PBMCs derive from whole blood samples. (b) Volcano plots for the WIHS cohort with similar annotations. (c) Correlation of significant DMPs between VACS and WIHS cohorts among the shared DMPs between the two cohorts. PBMC: peripheral blood mononuclear cell; EWAS: Epigenome-wide Association Study; VACS: Veteran Aging Cohort Study; WIHS: Women's Interagency HIV Study; DMPs: Differential Methylation Positions. *Significant genes shared between at least two cell types.

Figure 4. Summary of cell type level epigenome-wide meta-analysis of the combined Veteran Aging Cohort Study (VACS) and Women's Interagency HIV study (WIHS) data. (a) Manhattan plot of epigenome-wide significant CpG sites prior to computational deconvolution of data into cell-type-specific methylation. (b) Manhattan plot of epigenome-wide significant CpG sites after computational deconvolution into cell-type-specific signals. (c) Volcano plots of hyper- and hypomethylated DMPs for HIV infection in each cell type following Meta-EWAS (Epigenome-wide Association Study). DMP: Differential Methylation Position. * Significant genes shared between at least two cell types.

Figure 5. Common and distinct DMP profiles among cell types following cell type level meta-EWAS. (a) Stacked Manhattan plots displaying -log10(p) for CpG sites of each cell type. (b) Unique and common DMPs among five cell types and peripheral blood mononuclear cells. Overlap size represents the number of shared DMPs between the designated cell types. DMP: Differential Methylation Position.

Figure 6. Characterization of epigenome-wide significant DMP from cell-type level meta-epigenome-wide association analysis. DMP: Differential Methylation Position.

Figure 7. Gene enrichment analysis of significant DMPs for HIV infection in CD4+ T cells with Catalogue Of Somatic Mutations in Cancer (COSMIC) Gene Census Tier 1 database. CGC Tier 1 denotes the gene has cancer gene mutation patterns and evidence of functional impact reported in the literature. DMP: Differential Methylation Position.

Figure 8. Gene enrichment analysis of DMPs in the genes enriched on hallmark gene pathways in CD4+ T cells. Notably, the allograft rejection, early estrogen response, interferon alpha response, interferon gamma response, and kras signaling pathways were significantly enriched. (a) Circos plot showing DMPs in four pathways; (b) Relationships of four significant pathways. DMP: Differential Methylation Position.

References

1. Marcus JL, Leyden WA, Alexeeff SE, Anderson AN, Hechter RC, Hu H, Lam JO, Towner WJ, Yuan Q, Horberg MA *et al*: **Comparison of Overall and Comorbidity-Free Life Expectancy Between Insured Adults With and Without HIV Infection, 2000-2016**. *JAMA Netw Open* 2020, **3**(6):e207954.
2. Smidt U, Muysers K, Pichotka J: **Alveolar-arterial O 2 and CO 2 gradients in normals during hypoxia**. *Bull Physiopathol Respir (Nancy)* 1970, **6**(2):501-504.
3. Silverberg MJ, Chao C, Leyden WA, Xu L, Horberg MA, Klein D, Towner WJ, Dubrow R, Quesenberry CP, Jr., Neugebauer RS *et al*: **HIV infection, immunodeficiency, viral replication, and the risk of cancer**. *Cancer Epidemiol Biomarkers Prev* 2011, **20**(12):2551-2559.
4. Gupta R, Malvi P, Parajuli KR, Janostiak R, Bugide S, Cai G, Zhu LJ, Green MR, Wajapeyee N: **KLF7 promotes pancreatic cancer growth and metastasis by up-regulating ISG expression and maintaining Golgi complex integrity**. *Proceedings of the National Academy of Sciences* 2020, **117**(22):12341-12351.
5. Sigel K, Park L, Justice A: **HIV and cancer in the Veterans Health Administration System**. *Semin Oncol* 2019, **46**(4-5):334-340.
6. Silverberg MJ, Lau B, Achenbach CJ, Jing Y, Althoff KN, D'Souza G, Engels EA, Hessol NA, Brooks JT, Burchell AN *et al*: **Cumulative Incidence of Cancer Among Persons With HIV in North America: A Cohort Study**. *Ann Intern Med* 2015, **163**(7):507-518.
7. Park LS, Tate JP, Sigel K, Brown ST, Crothers K, Gibert C, Goetz MB, Rimland D, Rodriguez-Barradas MC, Bedimo RJ *et al*: **Association of Viral Suppression With Lower AIDS-Defining and Non-AIDS-Defining Cancer Incidence in HIV-Infected Veterans: A Prospective Cohort Study**. *Ann Intern Med* 2018, **169**(2):87-96.
8. Park LS, Hernandez-Ramirez RU, Silverberg MJ, Crothers K, Dubrow R: **Prevalence of non-HIV cancer risk factors in persons living with HIV/AIDS: a meta-analysis**. *AIDS* 2016, **30**(2):273-291.
9. Abdel-Hameed EA, Ji H, Shata MT: **HIV-Induced Epigenetic Alterations in Host Cells**. *Adv Exp Med Biol* 2016, **879**:27-38.
10. Roebuck KA, Saifuddin M: **Regulation of HIV-1 transcription**. *Gene Expr* 1999, **8**(2):67-84.
11. Bogoi RN, de Pablo A, Valencia E, Martín-Carbonero L, Moreno V, Vilchez-Rueda HH, Asensi V, Rodriguez R, Toledano V, Rodés B: **Expression profiling of chromatin-modifying enzymes and global DNA methylation in CD4+ T cells from patients with chronic HIV infection at different HIV control and progression states**. *Clin Epigenetics* 2018, **10**:20.
12. Zheng L, Yang Y, Guocai L, Pauza CD, Salvato MS: **HIV Tat protein increases Bcl-2 expression in monocytes which inhibits monocyte apoptosis induced by tumor necrosis factor-alpha-related apoptosis-induced ligand**. *Intervirology* 2007, **50**(3):224-228.
13. French AJ, Natesampillai S, Krogman A, Correia C, Peterson KL, Alto A, Chandrasekar AP, Misra A, Li Y, Kaufmann SH *et al*: **Reactivating latent HIV with PKC agonists induces resistance to apoptosis and is associated with phosphorylation and activation of BCL2**. *PLoS Pathog* 2020, **16**(10):e1008906.
14. Rajagopalan D, Jha S: **An epi(c)genetic war: Pathogens, cancer and human genome**. *Biochim Biophys Acta Rev Cancer* 2018, **1869**(2):333-345.
15. Tian Y, Yang W, Song J, Wu Y, Ni B: **Hepatitis B virus X protein-induced aberrant epigenetic modifications contributing to human hepatocellular carcinoma pathogenesis**. *Mol Cell Biol* 2013, **33**(15):2810-2816.

16.   Corley MJ, Sacdalan C, Pang APS, Chomchey N, Ratnaratorn N, Valcour V, Kroon E, Cho KS, Belden AC, Colby D *et al*: **Abrupt and altered cell-type specific DNA methylation profiles in blood during acute HIV infection persists despite prompt initiation of ART**. *PLoS Pathog* 2021, **17**(8):e1009785.

17.   Zhang X, Justice AC, Hu Y, Wang Z, Zhao H, Wang G, Johnson EO, Emu B, Sutton RE, Krystal JH *et al*: **Epigenome-wide differential DNA methylation between HIV-infected and uninfected individuals**. *Epigenetics* 2016, **11**(10):750-760.

18.   Yang CX, Schon E, Obeidat M, Kobor MS, McEwen L, MacIsaac J, Lin D, Novak RM, Hudson F, Klinker H *et al*: **Occurrence of Accelerated Epigenetic Aging and Methylation Disruptions in Human Immunodeficiency Virus Infection Before Antiretroviral Therapy**. *J Infect Dis* 2021, **223**(10):1681-1689.

19.   Nelson KN, Hui Q, Rimland D, Xu K, Freiberg MS, Justice AC, Marconi VC, Sun YV: **Identification of HIV infection-related DNA methylation sites and advanced epigenetic aging in HIV-positive, treatment-naive U.S. veterans**. *AIDS* 2017, **31**(4):571-575.

20.   Shu C, Jaffe AE, Sabunciyan S, Ji H, Astemborski J, Sun J, Bakulski KM, Mehta SH, Kirk GD, Maher BS: **Epigenome-wide association scan identifies methylation sites associated with HIV infection**. *Epigenomics* 2020, **12**(21):1917-1927.

21.   Shu C, Justice AC, Zhang X, Marconi VC, Hancock DB, Johnson EO, Xu K: **DNA methylation biomarker selected by an ensemble machine learning approach predicts mortality risk in an HIV-positive veteran population**. *Epigenetics* 2021, **16**(7):741-753.

22.   Zhang X, Hu Y, Aouizerat BE, Peng G, Marconi VC, Corley MJ, Hulgan T, Bryant KJ, Zhao H, Krystal JH *et al*: **Machine learning selected smoking-associated DNA methylation signatures that predict HIV prognosis and mortality**. *Clin Epigenetics* 2018, **10**(1):155.

23.   Zhang X, Hu Y, Justice AC, Li B, Wang Z, Zhao H, Krystal JH, Xu K: **DNA methylation signatures of illicit drug injection and hepatitis C are associated with HIV frailty**. *Nat Commun* 2017, **8**(1):2243.

24.   Schuyler RP, Merkel A, Raineri E, Altucci L, Vellenga E, Martens JHA, Pourfarzad F, Kuijpers TW, Burden F, Farrow S *et al*: **Distinct Trends of DNA Methylation Patterning in the Innate and Adaptive Immune Systems**. *Cell Rep* 2016, **17**(8):2101-2111.

25.   Loyfer N, Magenheim J, Peretz A, Cann G, Bredno J, Klochendler A, Fox-Fisher I, Shabi-Porat S, Hecht M, Pelet T *et al*: **A DNA methylation atlas of normal human cell types**. *Nature* 2023, **613**(7943):355-364.

26.   Moir S, Fauci AS: **B-cell responses to HIV infection**. *Immunol Rev* 2017, **275**(1):33-48.

27.   McMichael AJ, Rowland-Jones SL: **Cellular immune responses to HIV**. *Nature* 2001, **410**(6831):980-987.

28.   Titus AJ, Gallimore RM, Salas LA, Christensen BC: **Cell-type deconvolution from DNA methylation: a review of recent applications**. *Hum Mol Genet* 2017, **26**(R2):R216-R224.

29.   Adams C, Nair N, Plant D, Verstappen SMM, Quach HL, Quach DL, Carvidi A, Nititham J, Nakamura M, Graf J *et al*: **Identification of cell-specific differential DNA methylation associated with methotrexate treatment response in rheumatoid arthritis**. *Arthritis Rheumatol* 2023.

30.   Jeong Y, de Andrade ESLB, Thalmeier D, Toth R, Ganslmeier M, Breuer K, Plass C, Lutsik P: **Systematic evaluation of cell-type deconvolution pipelines for sequencing-based bulk DNA methylomes**. *Brief Bioinform* 2022, **23**(4).

31.   Moss J, Magenheim J, Neiman D, Zemmour H, Loyfer N, Korach A, Samet Y, Maoz M, Druid H, Arner P *et al*: **Comprehensive human cell-type methylation atlas reveals**

origins of circulating cell-free DNA in health and disease. *Nat Commun* 2018, **9**(1):5068.

32.   You C, Wu S, Zheng SC, Zhu T, Jing H, Flagg K, Wang G, Jin L, Wang S, Teschendorff AE: **A cell-type deconvolution meta-analysis of whole blood EWAS reveals lineage-specific smoking-associated DNA methylation changes**. *Nat Commun* 2020, **11**(1):4779.

33.   Rahmani E, Schweiger R, Rhead B, Criswell LA, Barcellos LF, Eskin E, Rosset S, Sankararaman S, Halperin E: **Cell-type-specific resolution epigenetics without the need for cell sorting or single-cell biology**. *Nat Commun* 2019, **10**(1):3417.

34.   Justice AC, Dombrowski E, Conigliaro J, Fultz SL, Gibson D, Madenwald T, Goulet J, Simberkoff M, Butt AA, Rimland D *et al*: **Veterans Aging Cohort Study (VACS): Overview and description**. *Med Care* 2006, **44**(8 Suppl 2):S13-24.

35.   Adimora AA, Ramirez C, Benning L, Greenblatt RM, Kempf MC, Tien PC, Kassaye SG, Anastos K, Cohen M, Minkoff H *et al*: **Cohort Profile: The Women's Interagency HIV Study (WIHS)**. *Int J Epidemiol* 2018, **47**(2):393-394i.

36.   Shiau S, Strehlau R, Wang S, Violari A, Do C, Patel F, Liberty A, Krupska I, Arpadi SM, Foca M *et al*: **Distinct epigenetic profiles in children with perinatally-acquired HIV on antiretroviral therapy**. *Scientific Reports* 2019, **9**(1):10495.

37.   Manandhar S, Lee YM: **Emerging role of RUNX3 in the regulation of tumor microenvironment**. *BMB Rep* 2018, **51**(4):174-181.

38.   Zhang Z, Li J, He T, Ouyang Y, Huang Y, Liu Q, Wang P, Ding J: **The competitive endogenous RNA regulatory network reveals potential prognostic biomarkers for overall survival in hepatocellular carcinoma**. *Cancer Sci* 2019, **110**(9):2905-2923.

39.   Tang Y, Fang G, Guo F, Zhang H, Chen X, An L, Chen M, Zhou L, Wang W, Ye T *et al*: **Selective Inhibition of STRN3-Containing PP2A Phosphatase Restores Hippo Tumor-Suppressor Activity in Gastric Cancer**. *Cancer Cell* 2020, **38**(1):115-128.e119.

40.   Djomkam Zune AL, Olwal CO, Tapela K, Owoicho O, Nganyewo NN, Lyko F, Paemka L: **Pathogen-Induced Epigenetic Modifications in Cancers: Implications for Prevention, Detection and Treatment of Cancers in Africa**. *Cancers (Basel)* 2021, **13**(23).

41.   Suzuki K, Okuno Y, Kawashima N, Muramatsu H, Okuno T, Wang X, Kataoka S, Sekiya Y, Hamada M, Murakami N *et al*: **MEF2D-BCL9 Fusion Gene Is Associated With High-Risk Acute B-Cell Precursor Lymphoblastic Leukemia in Adolescents**. *J Clin Oncol* 2016, **34**(28):3451-3459.

42.   Megonigal MD, Cheung NK, Rappaport EF, Nowell PC, Wilson RB, Jones DH, Addya K, Leonard DG, Kushner BH, Williams TM *et al*: **Detection of leukemia-associated MLL-GAS7 translocation early during chemotherapy with DNA topoisomerase II inhibitors**. *Proc Natl Acad Sci U S A* 2000, **97**(6):2814-2819.

43.   Yamato G, Yamaguchi H, Handa H, Shiba N, Kawamura M, Wakita S, Inokuchi K, Hara Y, Ohki K, Okubo J *et al*: **Clinical features and prognostic impact of PRDM16 expression in adult acute myeloid leukemia**. *Genes Chromosomes Cancer* 2017, **56**(11):800-809.

44.   Brett JO, Spring LM, Bardia A, Wander SA: **ESR1 mutation as an emerging clinical biomarker in metastatic hormone receptor-positive breast cancer**. *Breast Cancer Research* 2021, **23**(1):85.

45.   Wei X, Walia V, Lin JC, Teer JK, Prickett TD, Gartner J, Davis S, Stemke-Hale K, Davies MA, Gershenwald JE *et al*: **Exome sequencing identifies GRIN2A as frequently mutated in melanoma**. *Nat Genet* 2011, **43**(5):442-446.

46.   Yang L, Chen Z, Xiong W, Ren H, Zhai E, Xu K, Yang H, Zhang Z, Ding L, He Y *et al*: **High expression of SLC17A9 correlates with poor prognosis in colorectal cancer**. *Hum Pathol* 2019, **84**:62-70.

47.   Lau QC, Raja E, Salto-Tellez M, Liu Q, Ito K, Inoue M, Putti TC, Loh M, Ko TK, Huang C *et al*: **RUNX3 is frequently inactivated by dual mechanisms of protein mislocalization and promoter hypermethylation in breast cancer**. *Cancer Res* 2006, **66**(13):6512-6520.

48.   Zou Y, Chen B: **Long non-coding RNA HCP5 in cancer**. *Clinica Chimica Acta* 2021, **512**:33-39.

49.   De Donato M, Babini G, Mozzetti S, Buttarelli M, Ciucci A, Arduini G, De Rosa MC, Scambia G, Gallo D: **KLF7: a new candidate biomarker and therapeutic target for high-grade serous ovarian cancer**. *Journal of Experimental & Clinical Cancer Research* 2020, **39**(1):265.

50.   Moron-Lopez S, Urrea V, Dalmau J, Lopez M, Puertas MC, Ouchi D, Gómez A, Passaes C, Mothe B, Brander C *et al*: **The Genome-wide Methylation Profile of CD4+ T Cells From Individuals With Human Immunodeficiency Virus (HIV) Identifies Distinct Patterns Associated With Disease Progression**. *Clin Infect Dis* 2021, **72**(9):e256-e264.

51.   Daussy CF, Galais M, Pradel B, Robert-Hebmann V, Sagnier S, Pattingre S, Biard-Piechaczyk M, Espert L: **HIV-1 Env induces pexophagy and an oxidative stress leading to uninfected CD4(+) T cell death**. *Autophagy* 2021, **17**(9):2465-2474.

52.   Jiang M, Kang Y, Sewastianik T, Wang J, Tanton H, Alder K, Dennis P, Xin Y, Wang Z, Liu R *et al*: **BCL9 provides multi-cellular communication properties in colorectal cancer by interacting with paraspeckle proteins**. *Nature Communications* 2020, **11**(1):19.

53.   **BCL2L2**

54.   Wagner TA, McLaughlin S, Garg K, Cheung CYK, Larsen BB, Styrchak S, Huang HC, Edlefsen PT, Mullins JI, Frenkel LM: **Proliferation of cells with HIV integrated into cancer genes contributes to persistent infection**. *Science* 2014, **345**(6196):570-573.

55.   Sterling TR, Vlahov D, Astemborski J, Hoover DR, Margolick JB, Quinn TC: **Initial plasma HIV-1 RNA levels and progression to AIDS in women and men**. *N Engl J Med* 2001, **344**(10):720-725.

56.   Yee Mon KJ, Goldsmith E, Watson NB, Wang J, Smith NL, Rudd BD: **Differential Sensitivity to IL-12 Drives Sex-Specific Differences in the CD8+ T Cell Response to Infection**. *Immunohorizons* 2019, **3**(4):121-132.

57.   Meier A, Chang JJ, Chan ES, Pollard RB, Sidhu HK, Kulkarni S, Wen TF, Lindsay RJ, Orellana L, Mildvan D *et al*: **Sex differences in the Toll-like receptor-mediated response of plasmacytoid dendritic cells to HIV-1**. *Nat Med* 2009, **15**(8):955-959.

58.   D'Souza G, Bhondoekhan F, Benning L, Margolick JB, Adedimeji AA, Adimora AA, Alcaide ML, Cohen MH, Detels R, Friedman MR *et al*: **Characteristics of the MACS/WIHS Combined Cohort Study: Opportunities for Research on Aging With HIV in the Longest US Observational Study of HIV**. *Am J Epidemiol* 2021, **190**(8):1457-1475.

59.   Schuler PJ, Harasymczuk M, Schilling B, Lang S, Whiteside TL: **Separation of human CD4+CD39+ T cells by magnetic beads reveals two phenotypically and functionally different subsets**. *J Immunol Methods* 2011, **369**(1-2):59-68.

60.   Lehne B, Drong AW, Loh M, Zhang W, Scott WR, Tan ST, Afzal U, Scott J, Jarvelin MR, Elliott P *et al*: **A coherent approach for analysis of the Illumina HumanMethylation450 BeadChip improves data quality and performance in epigenome-wide association studies**. *Genome Biol* 2015, **16**(1):37.

a Data collection from two study cohorts. The Veterans Aging Cohort Study (VACS) provided whole blood samples where the Women's Interagency HIV Study (WIHS) did PBMCs.

b Validation of TCA computational deconvolution model by three sorted cell types.

Comparison of cell-type specific methylation data validates TCA as a computational alternative to cell sorting

Methylation sequencing vs. TCA

Methylation sequencing

Cell-type specific TCA methylation

Bead-based cell sorting

Computational deconvolution

c The TCA method yields cell-type specific CpG data following array-based DNA methylation in bulk cells (450K and 850K).

d Cell type-level EWAS of each cohort studying the methylation landscape associated with HIV in five cell types.

e Cell type-based EWAS meta-analysis combining data from both cohorts.

f Overrepresentation of COSMIC Cancer Gene Census (CGC) hallmark gene differential methylation. Differentially methylated hallmark genes (CGC Tier 1) reached epigenome-wide significance for HIV infection in each cell type.

g Gene set enrichment analysis with hallmark gene sets from the Molecular Signature Database. Significant pathways, many involved in cancer development and progression, emerged from four cell types.
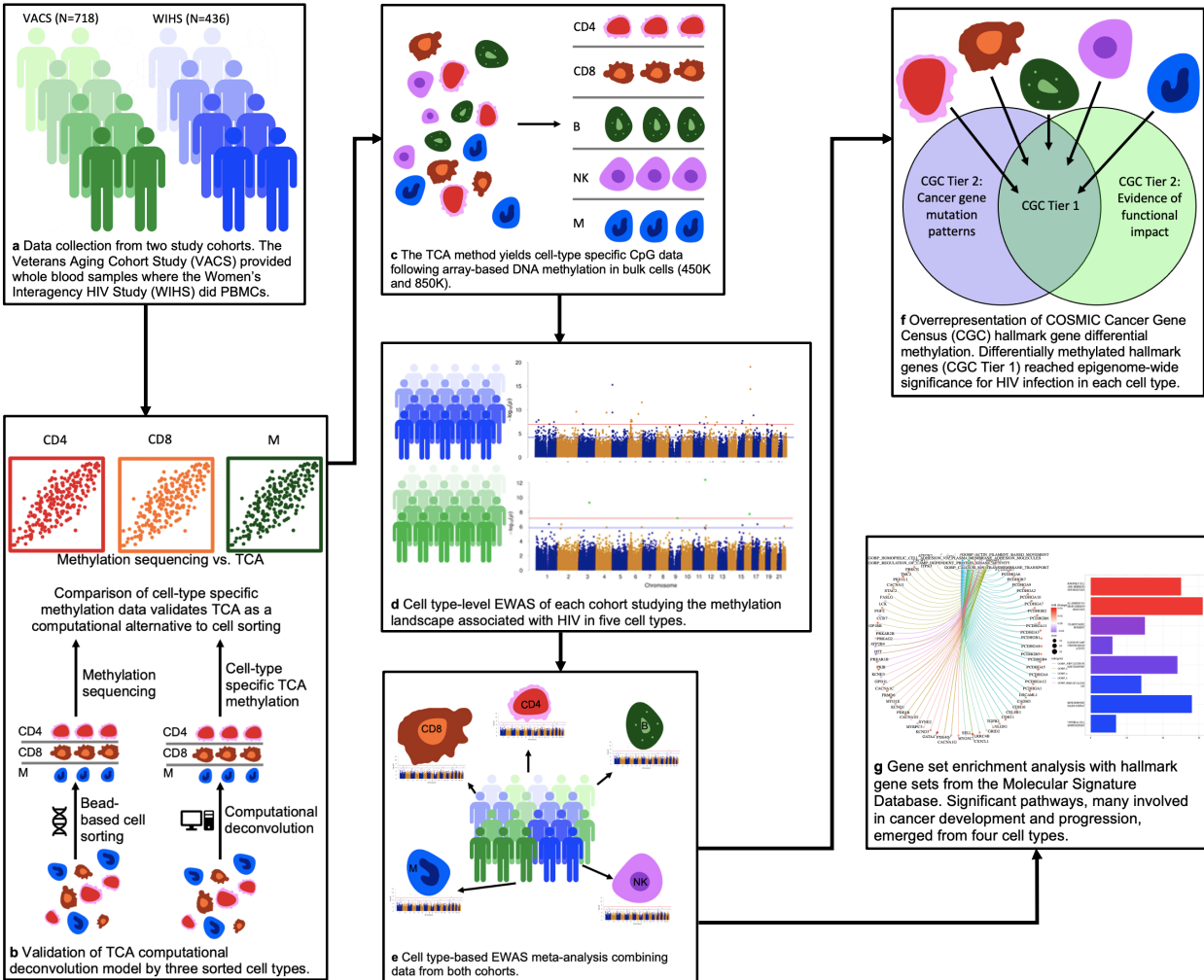
Figure 1

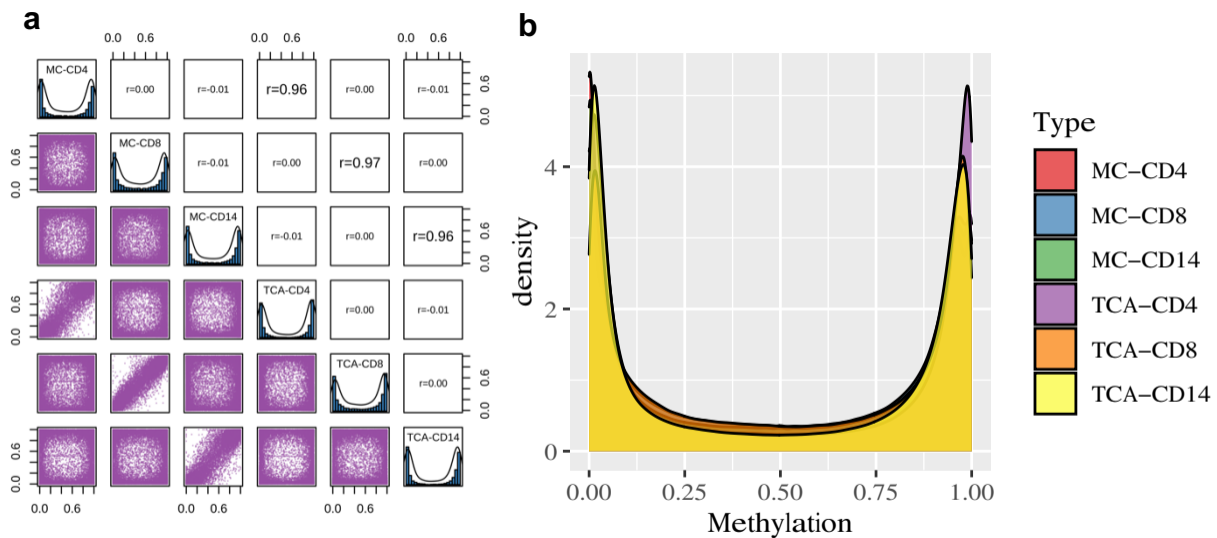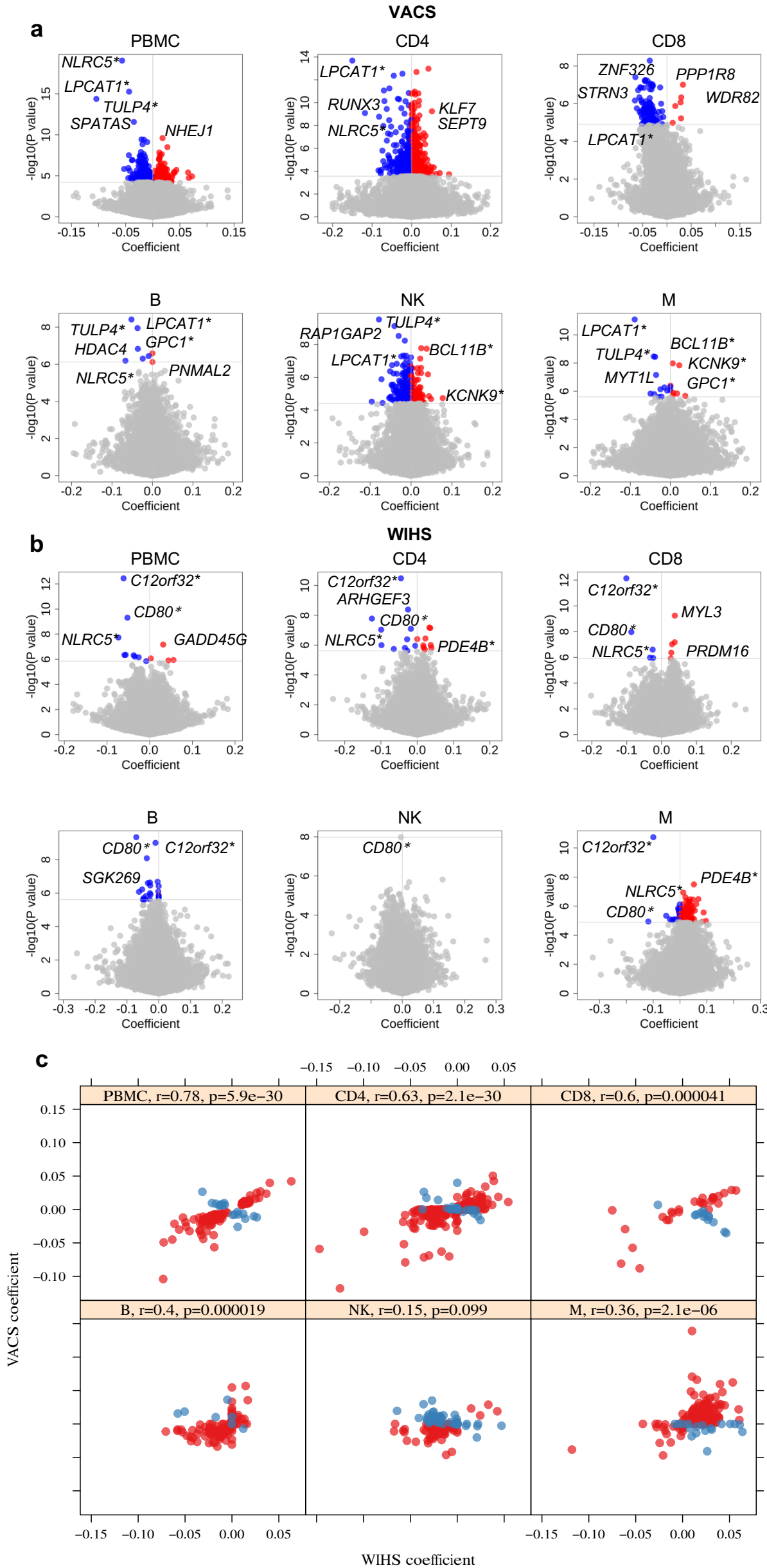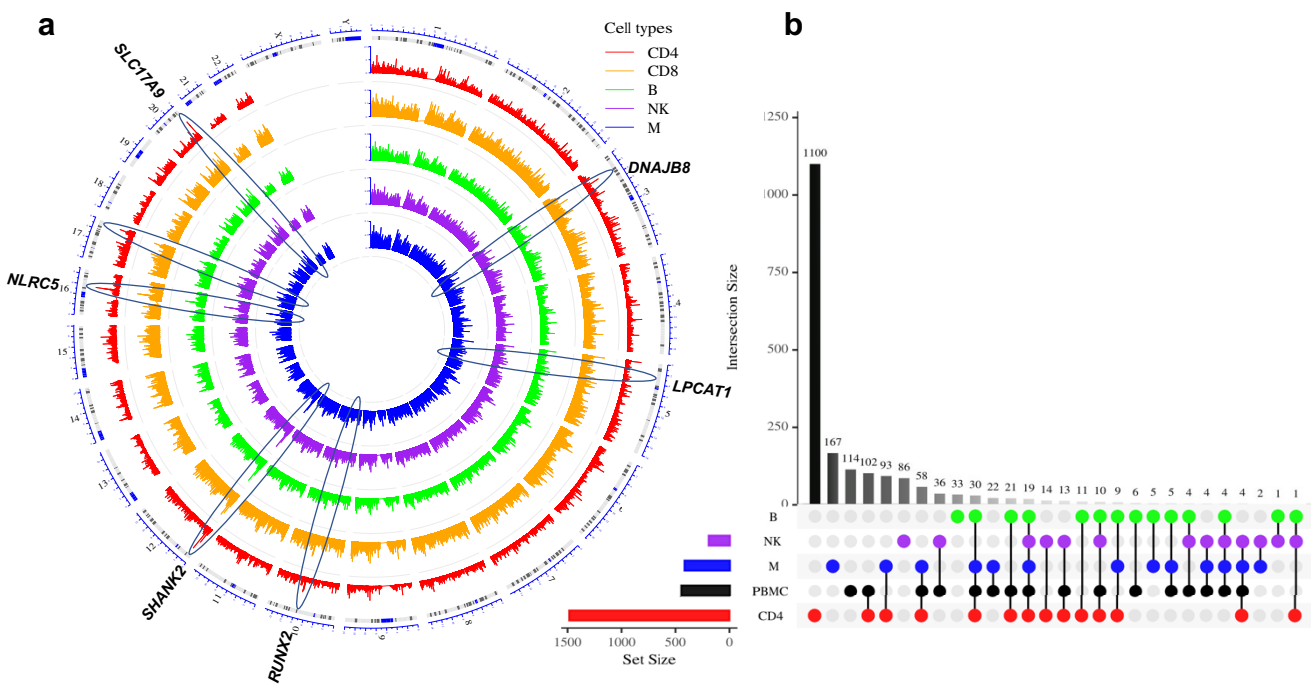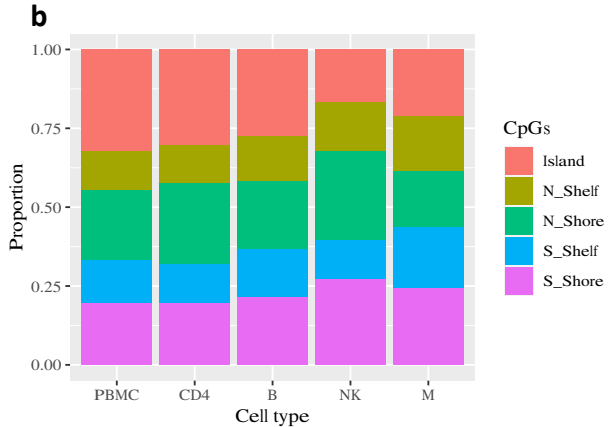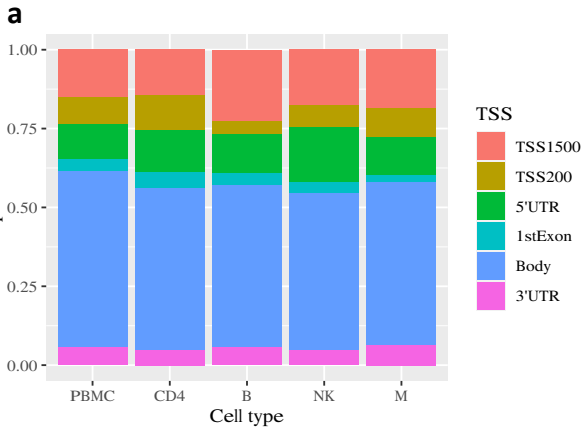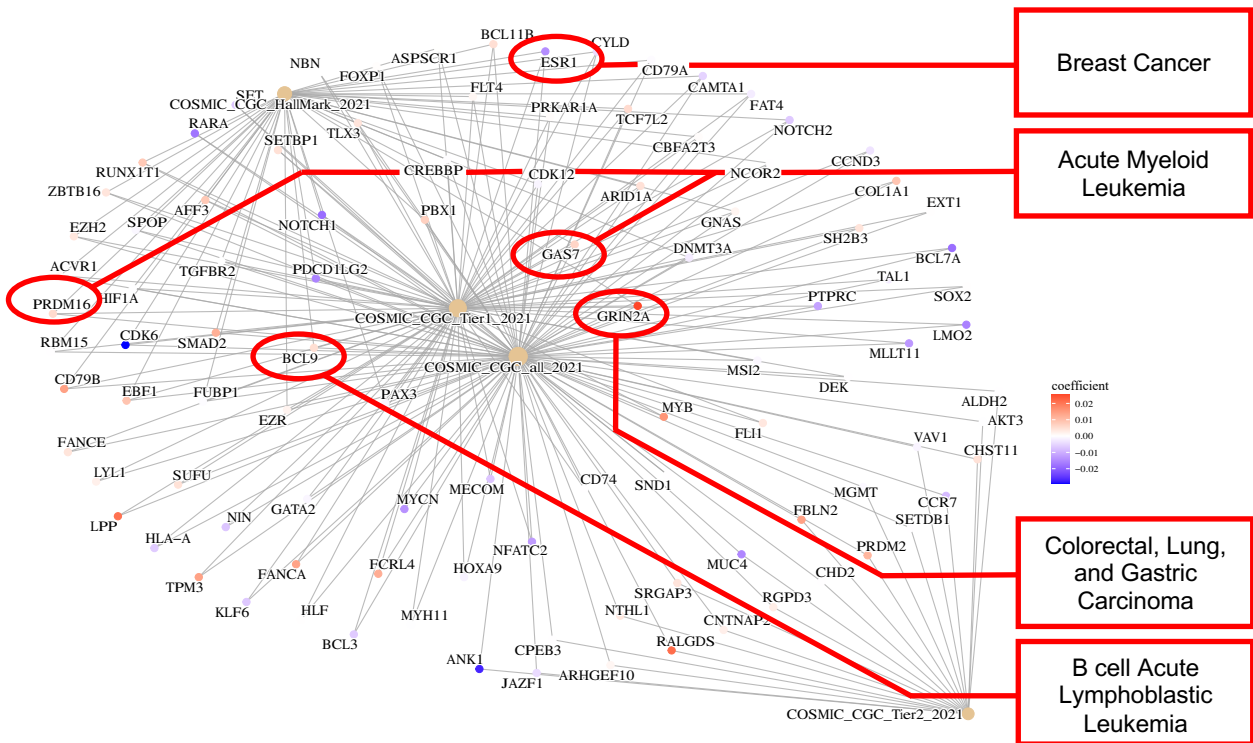Figure 2

Figure 3

Figure 4

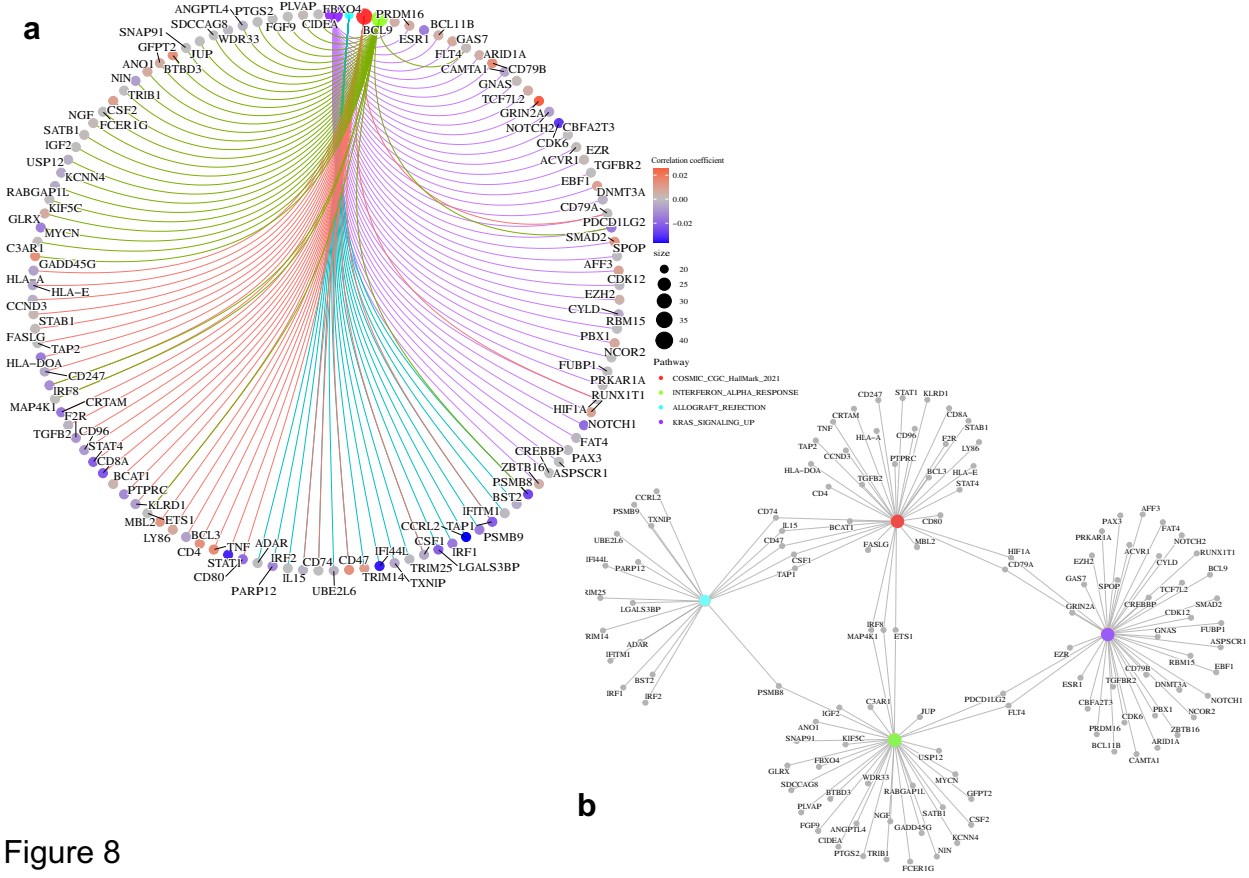Figure 5

Figure 6

Figure 7

Figure 8

Table 1 Overlap between hallmark genes for cancer and meta-EWAS identified significant genes for HIV infection

Cell Type	probe	Chromosome	Position	Gene Symbol	Name	Chr Band	Tumour Types(Somatic)	Role in Cancer	MetaEffect	Zscore	Meta P	Hat P	FDR
CD4+T cell	cg11754402	1	147012719	BCL9	B-cell CLL/lymphoma 9	21.2	B-ALL	oncogene, fusion	0.00486	6.788	1.14E-11	0.1161	3.42E-07
	cg05573412	1	3238152	PRDM16	PR domain containing 16	36.32	MDS, AML	oncogene, fusion	0.00544	5.649	1.61E-08	0.9565	9.60827E-05
	cg00168240	1	3347457						0.00604	4.892	9.98E-07	0.6543	0.001835991
	cg01062116	1	3352435						0.00567	4.281	0.00001861	0.1391	0.012478968
	cg24900983	6	152128528	ESR1	estrogen receptor 1	25.2	breast	oncogene, TSG, fusion	-0.01417	-5.535	3.12E-08	0.677	0.000151436
	cg23467008	6	152128537						-0.002735	-5.213	1.86E-07	0.5923	0.000557299
	cg21950534	6	152128483						-0.0088855	-4.928	8.32E-07	0.176	0.001611014
	cg24900983	6	152128528						-0.01417	-5.535	3.12E-08	0.677	0.000151436
	cg12833931	17	9862893	GAS7	growth arrest-specific 7	13.1	AML*	fusion	0.00629	5.067	4.04E-07	0.2609	0.000979112
	cg06470804	5	180071930	FLT4	fms-related tyrosine kinase 4	35.3	soft tissue sarcoma	oncogene	0.001158	4.866	1.14E-06	0.3619	0.001972615
	cg05307957	1	27028974	ARID1A	AT rich interactive domain 1A (SWI-like)	36.11	clear cell ovarian carcinoma, RCC, breast	TSG, fusion	0.00448	4.666	3.08E-06	0.3885	0.004052045
	cg08355301	17	62009651	CD79B	CD79b molecule, immunoglobulin-associated beta	23.3	DLBCL, WM	oncogene	0.013975	4.629	3.68E-06	0.8054	0.00447258
	cg21772773	1	6880294	CAMTA1	calmodulin binding transcription activator 1	36.31	epithelioid haemangioendothelioma	TSG, fusion	-0.005355	-4.621	3.81E-06	0.6247	0.00457825
	cg07344096	1	7724122						0.007465	4.184	0.00002868	0.8856	0.016949257
	cg03051946	1	6946515						0.000540046	3.768	0.0001645	0.1644	0.047070759
	cg10546626	20	57424521	GNAS	guanine nucleotide binding protein (G protein), alpha stimulating activity polypeptide 1	13.32	pituitary adenoma, pancreatic intraductal papillary mucinous neoplasm, fibrous dysplasia	oncogene	0.0015175	4.499	6.84E-06	0.8962	0.006643465
	cg03613625	20	57408174						0.003505	3.949	0.00007835	0.06884	0.030883664
	cg26380291	10	114787843	TCF7L2	transcription factor 7-like 2	25.3	colorectal	oncogene, fusion	0.005369	4.46	8.19E-06	0.8992	0.007316706
	cg03935060	16	10272021	GRIN2A	glutamate receptor, ionotropic, N-methyl D-aspartate 2A	13.2	melanoma, colorectal carcinoma, gastric carcinoma, lung carcinoma	TSG	0.025175	4.449	8.63E-06	0.2726	0.007566163
	cg26196087	1	120530251	NOTCH2	Notch homolog 2	12	marginal zone lymphoma, DLBCL, bladder	oncogene, TSG	-0.006609	-4.296	0.00001735	0.5559	0.01198841
	cg07266431	7	92460113	CDK6	cyclin-dependent kinase 6	21.2	ALL	oncogene, fusion	-0.0288	-4.139	0.00003482	0.1561	0.019241278
	cg00907204	7	92461971						-0.0143	-3.923	0.00008733	0.2113	0.033421433
	cg06982885	16	89029055	CBFA2T3	core-binding factor, runt domain, alpha subunit 2; translocated to, 3 (MTG-16)	24.3	AML	TSG, fusion	0.0011815	3.813	0.0001373	0.3417	0.042156881
	cg01517968	12	52348879	ACVR1	activin A receptor, type I	24.1	DIPG	oncogene	0.016675	4.765	1.89E-06	0.05167	0.002830141
	cg15130433	6	159240081	EZR	ezrin	25.3	NSCLC	fusion	0.001795	4.184	0.0001046	0.3215	0.036602471
	cg13859541	3	30647802	TGFBR2	transforming growth factor beta receptor II	24.1	head and neck, colorectal	TSG	-4.16225E-05	-3.874	0.0001069	0.4247	0.036901374
	cg25353990	5	158164199	EBF1	early B-cell factor 1	33.3	lipoma	TSG, fusion	0.00889	3.871	0.0001083	0.4636	0.037195995
	cg20669908	2	25498444	DNMT3A	DNA (cytosine-5-)-methyltransferase 3 alpha	23.3	AML	TSG	-0.002447	-3.823	0.0001317	0.6217	0.041370617
B Cell													
	cg26425711	1	3251680	PRDM16	PR domain containing 16	36.32	MDS, AML	oncogene, fusion	0.01325	4.281	0.00001858	0.8995	0.048768324
	cg00168240	1	3347457	PRDM16;PRDM16					0.00483	5.241	2.18e-08 (6.16e-09)	0.3239	0.005427865
	cg14133064	9	5530115	PDCD1LG2	programmed cell death 1 ligand 2	24.1	PMBL, Hodgkin lymphoma	oncogene, fusion	-0.00993	-4.937	7.915E-07	0.5993	0.011243238
	cg00856404	2	25498451	DNMT3A	DNA (cytosine-5-)-methyltransferase 3 alpha	23.3	AML	TSG	-0.0211	-4.731	0.000002235	0.07902	0.019841261
	cg20669908	2	25498444						-0.016585	-4.868	0.000001126	0.1058	0.012772781
	cg22820188	1	156086004	LMNA	lamin A/C	22	Spitzoid tumour	fusion	-0.0219	-4.703	0.000002563	0.05563	0.021112522
	cg07440398	14	99712966	BCL11B	B-cell CLL/lymphoma 11B (CTIP2)	32.2	T-ALL	oncogene, TSG, fusion	0.019905	4.34	0.00001428	0.2027	0.043196048
Natural Killer Cell													
	cg07440398	14	99712966	BCL11B	B-cell CLL/lymphoma 11B (CTIP2)	32.2	T-ALL	oncogene, TSG, fusion	0.0395	4.713	0.000002438	0.3084	0.016187883
	cg01636910	1	85740751	BCL10	B-cell CLL/lymphoma 10	22.3	MALT	TSG, fusion	-0.011200007	-4.484	0.000007312	0.07226	0.02843783
Monocyte													
	cg07440398	14	99712966	BCL11B	B-cell CLL/lymphoma 11B (CTIP2)	32.2	T-ALL	oncogene, TSG, fusion	0.008815	6.033	1.61E-09	0.1174	0.000329551
	cg05100282	14	99713430						2.52E-08	4.65	3.32E-06	0.9355	0.015425098
	cg27229529	1	3162676	PRDM16	PR domain containing 16	36.32	MDS, AML	oncogene, fusion	0.009845	4.414	0.00001017	0.3783	0.024287031
	cg00168240	1	3347457						0.0191	6.14	8.24E-10	0.104	0.000329551
	cg14133064	9	5530115	PDCD1LG2	programmed cell death 1 ligand 2	24.1	PMBL, Hodgkin lymphoma	oncogene, fusion	-0.008485	-5.19	2.10E-07	0.135	0.00339572
	cg00625963	1	32740034	LCK	lymphocyte-specific protein tyrosine kinase	35.2	T-ALL	oncogene, fusion	-0.00885	-4.336	0.00001453	0.1171	0.027574071
	cg25353990	5	158164199	EBF1	early B-cell factor 1	33.3	lipoma	TSG, fusion	0.0197	4.081	0.00004491	0.05484	0.046915091