medRxiv preprint doi: https://doi.org/10.1101/2020.05.11.20027458; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

DNA methylation mediates the effect of cocaine use on HIV severity

Chang Shu[1,2], Amy C. Justice[2,3], Xinyu Zhang[1,2], Zuoheng Wang[4], Dana B. Hancock[5], Eric O. Johnson[5,6], Ke Xu[1,2]

1. Department of Psychiatry, Yale School of Medicine, New Haven, CT, USA
2. Connecticut Veteran Healthcare System, West Haven, CT, USA
3. Department of Internal Medicine, Yale School of Medicine, New Haven, CT, USA
4. Department of Biostatistics, Yale School of Public Health, New Haven, CT, USA
5. Center for Omics Discovery and Epidemiology, Behavioral Health Research Division, RTI International, Research Triangle Park, NC, USA
6. Fellow Program, RTI International, Research Triangle Park, NC, USA.

Keywords: Cocaine use, HIV severity, mortality, DNA methylation, mediation effect

Running title: DNA methylation mediates cocaine exposure on HIV

Word count:

Abstract: 237; Text: 2989

Figure: 4

Table: 5

Corresponding Author

Ke Xu, MD., Ph.D.

Associate Professor of Psychiatry

Yale School of Medicine

950 Campbell Ave, West Haven, CT

Email: ke.xu@yale.edu

Tel: 203-932-5711 x7430

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.11.20027458; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Abstract

Background: Cocaine use accelerates HIV progression and worsens HIV outcomes. We assessed whether DNA methylation (DNAm) in blood mediates the association between cocaine use and HIV severity in a veteran population.

Methods: We analyzed 1,435 HIV-positive participants from the Veterans Aging Cohort Study Biomarker Cohort. HIV severity was measured by the Veteran Aging Cohort Study index (VACS index). We assessed the effect of cocaine use on VACS index and mortality among the HIV-positive participants. We selected candidate mediators from 408,583 CpGs that were both significantly associated with persistent cocaine use and VACS index by epigenome-wide association (EWA) scans. Mediation analysis was conducted on the selected CpGs and joint mediation effect of multiple CpGs were estimated.

Results: More frequent cocaine use was significantly associated with higher VACS index (p=2.7E-04) and cocaine use increased the risk of 10-year mortality (hazard ratio=1.10, p=0.011) with adjustment of confounding factors. Based on the EWA scan, 15 candidate mediator CpGs were selected and 12 CpGs showed significant mediation effect with each explained 11.3%-29.5% of the variation. The joint mediation effect of these 12 CpGs accounted for 47.2% of cocaine's effect on HIV severity. Genes harboring these 12 CpGs are involved in antiviral response (*IFIT3*, *IFITM1*, *NLRC5*, *PLSCR1*, *PARP9*) and HIV progression (*CX3CR1*, *MX1*). Furthermore, these genes were enriched in pathways involving in viral process (p=7.76E-08) and response to cytokine (p= 9.88E-07).

Conclusions: DNAm played a mediation role between cocaine use and HIV severity.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.11.20027458; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Introduction

Cocaine use is highly prevalent among persons with chronic HIV infection [1, 2]. Previous studies have shown that cocaine use accelerated HIV progression [3-6]. However, the biological mechanism of cocaine's effect on HIV outcomes remains largely unknown. Some studies have suggested that cocaine use may worsen HIV outcomes due to poor adherence to antiretroviral therapy (ART) among HIV-positive participants [7, 8]. Other studies have demonstrated that cocaine's adverse effect on HIV outcomes is independent of ART [5, 6, 9-11], supporting the hypothesis that cocaine exposure may lead to long-lasting pathophysiological changes in the immune system that worsen HIV outcomes.

DNA methylation (DNAm) is an important mechanisms that haven been shown to associated with many environmental exposures such as smoking and alcohol[12-21], and diseases such as cancer, diabetes and cardiovascular diseases [22-27]. Our previous study showed that two DNAm sites in *NLRC5* was differentially methylated between HIV-positive and HIV-negative participants in peripheral blood [28]. More importantly, DNAm may play a mediation role linking environmental exposure and disease outcomes [29-35]. A recent study reported that DNAm sites in *PIM3* (energy metabolism) and *ABCG1* (lipid metabolism) mediated the association between prenatal famine exposure and long-term metabolic outcomes [33]. Another study reported the mediation effect of cg05575921 (*AHRR*) between smoking and the risk of bladder cancer among postmenopausal women [36].

Previous studies have shown that the use of cocaine enhances HIV-1 replication and undermines immune function by dysregulating gene expression on HIV-1 entry co-receptors, enhancing HIV-1 cellular toxicity and dysregulating interleukins in the host [37, 38]. However, no studies have examined the role of DNAm on how cocaine exposure worsens HIV outcomes. We hypothesized that DNAm may mediate the effect of cocaine exposure on HIV severity.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.11.20027458; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

In this study, we first validated previous findings by examining cocaine's adverse effect on HIV severity and mortality. We further conducted a mediation analyses to assess the mediation role of DNAm sites (or CpGs) on cocaine's effect on HIV severity using the Veteran Aging Cohort Study – Biomarker Cohort (VACS-BC). Lastly, we examined genes and biological pathways related with these CpGs. Our results provide new insights for the role of DNAm on how cocaine affects HIV severity.

Methods

Study samples: The Veteran Aging Cohort Study (VACS) is a prospective cohort study of veterans designed to study substance use and HIV related outcomes with patient surveys, electronic medical records and biospecimen data [39]. Baseline survey was conducted at the enrollment[39]. The follow-up survey of 5 visits occurred in approximately one-year interval [39]. Blood samples were collected in the middle of follow-up for a subset of participants in the cohort (VACS-BC) [40]. A total of 1,435 HIV-positive participants from the VACS-BC were used to examine cocaine's effect on mortality and HIV severity, and a subset of samples (N=875) with DNAm data available were used for mediation analyses (**Figure 1**). Demographic and clinical information of baseline samples and a subset of the samples at the time of blood collection are summarized in **Table 1**.

Assessment of cocaine use: The timeline of cocaine use assessment for each analysis is illustrated in **Figure 1**. Information on cocaine use status was self-reported through telephone interview for a total of 5 visits. We defined the "persistent cocaine use" group as self-reported cocaine use across all 5 visits, and "no cocaine use" group with self-reported no cocaine use across all 5 visits. This definition led to a subset of samples with 265 persistent cocaine users

medRxiv preprint doi: https://doi.org/10.1101/2020.05.11.20027458; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

and 202 non-users for the mediation analyses to eliminate the inconsistent response across 5 visits and examine the effect of long-term cocaine exposure on DNAm and HIV severity. The frequency of cocaine use was also assessed at baseline (**Figure 1**). Each participant was asked "how often in the past year have you used cocaine or crack?", from which cocaine frequency of use was coded as an ordinal variable: 0=never tried; 1= no use in the last year; 2=less than once a month; 3=1-3 times a month; 4=1-3 times a week, and 5=4 or more times a week.

Assessment of mortality and HIV severity: The timeline of HIV severity measurement and survival information for each analysis is shown in **Figure 1**. Mortality and survival year information were based on medical records. VACS index was used as a measure of HIV severity [41-44] and was obtained at each visit and at the time of blood collection (**Figure 1**). The VACS index was calculated by summing pre-assigned points for age, routinely monitored indicators of HIV disease (CD4 count and HIV-1 RNA) and other general indicators of organ system injury [41]. High VACS index corresponds to worsened HIV outcomes and VACS index is positively associated with increased mortality [45]. VACS index at same time with DNAm profiling was used for the selection of candidate mediator CpGs, and the average VACS index after the blood collection was used for mediation analyses (**Figure 1**).

DNA methylation profiling and quality control: DNA samples were extracted from blood for a subset of 875 HIV-positive participants (**Figure 1**). DNAm was profiled using two different methylation arrays, with 475 samples profiled by the Infinium Human Methylation 450K BeadChip (HM450K) and 400 samples profiled by the Infinium Human Methylation EPIC BeadChip (EPIC) [40]. DNA samples were randomly selected for each methylation array regardless of cocaine use status or other clinical demographic variables.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.11.20027458; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Quality control (QC) for samples measured by each array was conducted separately using the same pipeline as previously described [46] by the R package *minfi* [47]. After QC, a total of 408,583 CpGs were measured by both the HM450k and EPIC array remained for analysis. Six cell type proportions (CD4+ T cells, CD8+ T cells, NK T cells, B cells, monocytes and granulocytes) were estimated for each participant using the established method [48]. Negative control principal components were extracted by *minfi* to control for background noise[47]. Batch effect removal was conducted by *combat* after QC [49].

Statistical analysis

Cocaine survival analysis among HIV-positive participants at baseline. Survival analysis was conducted using baseline information among 1,435 HIV-positive participants with cocaine use frequency (0-5) and other covariates (**Figure 1**). Kaplan Meier analyses on 10-year follow-up among HIV-positive and HIV-negative participants by cocaine use frequency (0-5) at baseline was conducted, and the Kaplan Meier curves were plotted by using the R package *survminer* [50]. A test on ordered differences of Kaplan Meier curves by cocaine use frequency was conducted by *survminer* [50].

To adjust for confounding factors, Cox proportional-hazards model was used to assess the hazard ratio of baseline cocaine use frequency (0-5) on mortality during the follow-up using the R package *survival* [51]. The following model was used to calculate the adjusted hazard ratio among HIV-positive participants.

$$h(t) = h_0(t)\exp\,(\beta_1 cocaine\ use\ frequency\ +\ \beta_2\ sex +\ \beta_3\ baseline\ age\ +\ \beta_4\ race$$
$$+\ \beta_5 \log_{10}(viral\ load) + \beta_6\ CD4\ count\ +\ \beta_7\ antiviral\ medication\ adherence)$$

We then added VACS index to the above model as covariate to assess whether the association of cocaine use frequency with mortality remained significant after adjusting for HIV severity.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.11.20027458; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Association between the cocaine use frequency and HIV severity among HIV-positive participants at baseline. This analysis was conducted using baseline information on cocaine use frequency (0□5), VACS index and other covariates (**Figure 1**). The following linear regression model was performed to test the association of the cocaine use frequency and HIV severity, adjusting for confounders shown in the following model.

$$\begin{aligned} HIV\ disease\ severity \\ &= \beta_1 cocaine\ use\ frequency\ + \beta_2\, sex + \beta_3\, age\ + \beta_4\, race \\ &+ \beta_5 \log_{10}(viral\ load) + \beta_6\ CD4\ count\ + \beta_7\ antiviral\ medication\ adherence \end{aligned}$$

Selection of candidate CpGs by epigenome-wide association (EWA) on persistent cocaine use and HIV severity. To select candidate CpG for mediation analysis, we conducted two separate EWAs on persistent cocaine use and HIV severity (**Figure 1**). Each EWA model adjusted for sex, baseline age, race, smoking, self-reported antiviral medication adherence, white blood cell count, estimated cell type proportions, negative control principal components and residual principal components, and the model details have been described previously [28, 46, 52]. Since CD4+ T cell count is one component of VACS index, to avoid overrepresented CpGs associated with CD4+ T cell type in EWA results, we extracted top 1,000 CD4+ T cell type relevant CpGs based on data from *FlowSorted.Blood.450k* [53]. Top 2 principal components that in total account for > 80% variation of the 1,000 CD4+ T cell CpGs were used as covariates in the VACS index EWA model. The significance threshold was arbitrarily set at a more liberal threshed (p<0.001). CpGs with p<0.001 in both EWAs on the persistent cocaine use and HIV severity were selected as candidate CpGs for mediation analyses.

Single-site mediation analysis and joint mediation analysis. The selected candidate CpGs were assessed as potential mediators on the association between persistent cocaine use and HIV severity among HIV-positive participants (N=467). We performed single-site mediation analysis using the R package *mediation*[54].

medRxiv preprint doi: https://doi.org/10.1101/2020.05.11.20027458; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

We used M to represent the candidate CpGs(mediator), X to represent persistent cocaine use status(exposure), Y to represent the average VACS index after blood collection(outcome), and C_i to represent k confounding variables (sex, age, race, smoking, self-report antiviral medication adherence, white blood cell count, estimated CD8 T cells, granulocyte, NK cells, B cells, monocytes). The mediator model $f(M|X, C)$ examined the association between persistent cocaine use and CpGs:

$$f(M|X, C) = \beta_0 X + \sum_{i=1}^{k} \beta_i C_i$$

The outcome model $f(Y|X, M, C)$ examined both the direct effect of persistent cocaine use on VACS index and the mediation effect by CpG:

$$f(Y|X, M, C) = \alpha_0 X + \alpha_1 M + \sum_{i=1}^{k} \alpha_{i+1} C_i$$

Thus, the mediation effect of CpG M is $\alpha_1\beta_0$, the total effect is $\alpha_0 + \alpha_1\beta_0$ and the proportion mediated is $\alpha_1\beta_0/(\alpha_0 + \alpha_1\beta_0)$. The confidence interval and p-value were estimated by bootstrapping of 1,000,000 iterations.

The joint mediation analysis on all significant mediator CpGs were conducted as previously described [55]. Assuming the independence between multiple mediators $M_1, M_2, \ldots, M_n$, the mediator model $f(M_j|X, C)$ for M_j is:

$$f(M_j|X, C) = \beta_{0j} X + \sum_{i=1}^{k} \beta_{ij} C_i$$

The outcome model $f(Y|X, M_1, \ldots, M_n, C)$ is:

$$f(Y|X, M_1, \ldots, M_n, C) = \alpha_0 X + \sum_{j=1}^{n} \alpha_j M_j + \sum_{i=1}^{k} \alpha_{i+n} C_i$$

medRxiv preprint doi: https://doi.org/10.1101/2020.05.11.20027458; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

The joint mediation effect of CpGs $M_1, \ldots, M_j$ is $\sum_{j=1}^{n} \alpha_j \beta_{0j}$, the total effect is $\alpha_0 + \sum_{j=1}^{n} \alpha_j \beta_{0j}$, and the proportion mediated is $\sum_{j=1}^{n} \alpha_j \beta_{0j} / (\alpha_0 + \sum_{j=1}^{n} \alpha_j \beta_{0j})$. The confidence interval and p-value were estimated by bootstrapping of 1,000,000 iterations.

Gene ontology (GO) enrichment analysis. We performed GO enrichment analysis for genes located near significant mediator CpGs using Database for Annotation, Visualization and Integrated Discovery (DAVID) [56]. To avoid redundancy in pathway names, we only used level 4 GO terms defined in DAVID in the enrichment analysis. We considered biological pathways with false discovery rate (FDR) less than 0.05 as statistically significant pathways.

Results

Cocaine use affects HIV severity and mortality among HIV-positive participants. We found that among HIV-positive participants, higher frequency of cocaine use at baseline was significantly associated with higher VACS index (i.e., higher HIV severity, p= 0.00027), after adjusting for sex, age, race, viral load, CD4 count and antiviral medication adherence (**Table 2**). Higher cocaine use frequency was also associated with increased mortality (p=0.008, **Figure 2a**). Such difference was not found among HIV-negative participants (p= 0.180, **Figure 2b**). Using Cox proportional-hazards model, this trend remained significant with a hazard ratio (HR) of 1.10 (95%CI: 1.02-1.19, p=0.011), controlling for sex, baseline age, race, viral load, CD4 count and antiviral medication adherence (**Table 2**). Interestingly, when VACS index was added to the Cox model, frequency of cocaine use was no longer significantly associated with mortality (HR= 1.076, p= 0.058), indicating that cocaine's effect on increased mortality may through HIV severity. Our results were consistent with previous findings that cocaine use accelerated HIV progression and increased mortality independent of antiviral medication adherence [5, 6, 9-11].

medRxiv preprint doi: https://doi.org/10.1101/2020.05.11.20027458; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Selection of candidate DNAm sites for mediation analysis by EWA scan

The EWA scan on persistent cocaine use showed good control of inflation (λ=1.034, **Figure S1**). A total of 497 CpGs met our candidate selection threshold (p<0.001). Interestingly, top ranked CpG site cg22917487 was close to epigenome-wide significance threshold (p<1.2E-07) with p-value of 1.69E-07. This CpG site is located on *CX3CR1*, a gene that encodes a coreceptor for HIV-1 and leads to rapid HIV progression (**Table S1**).

The EWA scan on VACS index also showed good control of inflation (λ=1.116, **Figure S1**). There were 876 CpGs that reached candidate selection threshold (p<0.001) (**Table S2**). Of note, 6 CpGs reached epigenome-wide significance threshold (p<1.2E-07). These CpGs were located near the genes involved in viral and immune response (*PARP9*, *IFITM1, CD247, IFIT3, VASN*, *RUNX1*).

We selected candidate CpGs that were both significantly associated with cocaine use and HIV severity (p<0.001) by two separate EWA scans of 408,583 CpGs for mediation analysis. Fourteen CpGs that both met our candidate selection threshold. Additionally, cg22917487 in *CX3CR1* showed strong association with cocaine (p=1.69E-07) and marginal association with VACS index (p=1.73E-03). Given its biological plausibility, this CpG was also included as a candidate mediator for mediation analysis. Overall, a total of 15 CpGs were selected as candidates to assess their potential mediation roles on the association between persistent cocaine use and HIV severity.

Mediation analysis of candidate CpGs between persistent cocaine use and HIV severity

We examined the mediation role of DNAm between persistent cocaine use and HIV severity (**Figure 1**). Average VACS index after blood collection to ensure that DNAm profiling preceded of HIV severity. Twelve out of the 15 candidate CpGs showed significant mediation effects on

medRxiv preprint doi: https://doi.org/10.1101/2020.05.11.20027458; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

the association between persistent cocaine use and VACS index with p-values ranging from 1.00E-06 to 0.00307 (**Table 4**). These results remained significant after Bonferroni correction (p<0.0033). Each CpG mediator explained between 11.3% to 29.5% of persistent cocaine use affecting HIV severity. These 12 CpGs collectively mediated 47.2% of the cocaine's effects on the HIV severity.

Significant mediator CpGs are located near 11 viral and immune response genes: *MX1*, *PARP9*, *IFIT3*, *IFITM1*, *NLRC5*, *EPSTI1*, *PLSCR1*, *TAP2*, *TAP1*, *CX3CR1* and *RIN2*. Five CpGs are located on 5' gene regulatory region, 4 CpGs on gene body, 2 CpGs on transcription start sites and 1 CpG on 3' gene regulatory region. Notably, these 12 CpGs were mostly less methylated in persistent cocaine use group compared to never cocaine use group (**Table 3**, **Figure 3**). **Figure 4** illustrates the mediation effect of cg26312951 (*MX1*), cg08122652 (*PARP9*), cg07839457 (*NLRC5*) and cg22917487 (*CX3CR1*) on persistent cocaine use affecting HIV severity.

Biological pathways

Twelve significant CpG mediators were annotated to 11 genes. These genes were significantly enriched in 6 biological pathways (**Table 5**). These significant pathways were viral process (p=7.76E-08), response to cytokine (p=9.88E-07), defense response to virus (p=4.40E-06), regulation of viral process (p=7.50E-06), response to type I interferon (p=8.49E-06) and response to virus (p=1.48E-05).

Discussion

Our findings provide evidence that cocaine use worsens HIV severity and increases mortality among HIV-1 positive participants, and cocaine's adverse effect is mediated by DNAm in blood. We identified 12 CpGs that collectively accounted for a total of 47.2% cocaine use's adverse

medRxiv preprint doi: https://doi.org/10.1101/2020.05.11.20027458; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

effect on HIV severity. These CpGs, which are enriched in viral process and inflammatory response pathways, offer new insights into the mechanisms of how cocaine use may affect HIV outcomes through DNAm.

These 12 CpGs are located in or near 11 biologically meaningful genes that were previously reported to be involved in inflammation, HIV-1 viral replication and other pathways that play critical roles in HIV progression. Specifically, cg06188083 on *IFIT3* mediated 28.8% of the variation, and *IFIT3* encodes an IFN-induced antiviral protein which acts as an inhibitor of viral processes and viral replication[57]. Another significant mediator CpG site, cg06188083, is located near interferon gene *IFITM1*. We previously reported hypomethylation of cg07839457 due to HIV infection, which is located in the promoter region of *NLRC5* [28]. This CpG site was also a significant mediator between cocaine and HIV severity in this study. *NLRC5* plays an important role in cytokine response and antiviral immunity through its inhibition of NF-kappa-B activation and negative regulation of type I interferon signaling pathways[58]. The converging evidence on cg07839457 (*NLRC5*) warrants further investigation of its role in HIV infection and progression. Another interesting CpG site, cg22917487 on *CX3CR1*, showed both strong association with persistent cocaine use and significant mediation effect of cocaine affecting HIV severity. *CX3CR1* is involved in leukocyte adhesion and migration and was recently identified as an HIV-1 coreceptor [59]. Some studies also showed that genetic variants on *CX3CR1* were associated with HIV susceptibility and rapid HIV progression to AIDS[60]. cg25114611, located in the promoter region of *FKBP5*, is also biologically plausible, given the implication for chronic cocaine administration upregulating *FKBP5* expression in rats [61].

The joint mediation effect of the 12 CpGs was less than the total sum of the mediation effects across the individual CpG. This result suggests that the mediation effect for 12 CpGs were not

medRxiv preprint doi: https://doi.org/10.1101/2020.05.11.20027458; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

entirely independent. CpGs located in the same gene or the biological pathways may collectively mediate cocaine use on HIV severity.

There are several strengths of this study. First, instead of selecting candidate mediator CpGs based on literature or hypotheses, we applied an unbiased epigenome-wide screening to select CpGs associated with both cocaine use and HIV severity. Second, to limit self-reporting bias of cocaine use, we leveraged longitudinal data in defining persistent cocaine use and never cocaine use. We included only those participants who consistently reported cocaine use or no cocaine use across all 5 visits for the selection of candidate CpGs and the mediation analyses. Lastly, we used the average VACS index after blood collection so that DNAm measurements(mediator) preceded HIV severity(outcome) for the mediation analyses.

One limitation of the study is that our sample size on mediation analyses is small. However, the strict definition of cocaine use helped reduced self-reporting bias and can potentially increase power by comparing extreme groups. Additionally, our samples consisted of mostly male veterans, which may limit the generalizability of our findings.

In summary, we validated previous reports that use of cocaine worsened HIV severity and increased risk of all-cause mortality among HIV positive participants. More importantly, this study, for the first time, found that several biologically meaningful DNAm sites mediated the adverse effect of cocaine use on HIV severity. These results merit future studies to further explore the biological mechanisms revealed by these DNAm sites on how cocaine affects HIV disease outcomes.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.11.20027458; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Tables

Table 1: Sample characteristics in HIV-positive participants

	Baseline*	Follow-up (time of blood collection)†
Sample size	1435	875
Age [mean (sd)]	48.8 (8.1)	51.5 (7.7)
Sex		
Male	1399 (97.5%)	861 (98.4%)
Female	36 (2.5%)	14 (1.6%)
Race		
Caucasian	290 (20.2%)	84 (9.6%)
African American	964 (67.2%)	726 (83.0%)
Other	181 (12.6%)	65 (7.4%)
Persistent Cocaine use status		
Persistent cocaine use		265 (30.3%)
Non-cocaine use		202 (23.1%)
Cocaine use at baseline (%)		
have never tried	555 (38.7%)	
no use in the last year	540 (37.6%)	
less than once a month	122 (8.5%)	
1-3 times a month	103 (7.2%)	
1-3 times a week	39 (2.7%)	
≥4 times a week	76 (5.3%)	
Cigarette smoking	762 (53.1%)	521 (59.5%)
VACS index‡	30.0 (19.1)	
Average VACS index after blood collection[mean(sd)]		39.20 (22.4)
CD4+ count [mean (sd)]	418.7 (273.0)	440.02 (289.1)
log10 viral load [mean (sd)]	3.1 (1.2)	2.74 (1.2)
Antiviral medication adherence (%)	1117 (77.8%)	666 (76.1%)

*Samples at baseline for the survival analysis

†A subset of samples at the time of blood collection with DNA methylation measurements

‡VACS index: Veteran Aging Cohort Study index

medRxiv preprint doi: https://doi.org/10.1101/2020.05.11.20027458; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Table 2: Association between cocaine use frequency and HIV severity and survival analysis among HIV-positive participants

	HIV severity*				Mortality†			
	Estimate	**SE**	**T**	**P value**	**Hazard ratio**	**95% CI**	**Z**	**P value**
Cocaine use frequency‡	1.00	0.28	3.65	2.70E-04	1.10	(1.02,1.19)	2.54	1.10E-02
Sex (reference: male)	5.78	2.33	2.48	1.34E-02	0.27	(0.07,1.08)	-1.86	6.33E-02
Age‡	1.09	0.05	23.58	< 2E-16	1.05	(1.04,1.07)	7.35	1.90E-13
Race (reference: Caucasian)								
African American	7.30	0.94	7.73	2.00E-14	1.05	(0.80,1.39)	0.37	7.15E-01
Other	5.33	1.31	4.08	4.80E-05	0.71	(0.47,1.09)	-1.57	1.17E-01
log 10 viral load‡	4.72	0.36	13.11	< 2eE-16	1.07	2	1.36	1.73E-01
CD4 count‡	-0.03	0.00	-19.05	< 2E-16	1.00	(1.00,1.00)	-3.74	1.80E-04
Antiviral medication adherence‡	2.59	0.99	2.62	8.87E-03	1.02	(0.76,1.37)	0.13	8.95E-01

*Summary statistics from linear regression model

†Summary statistics from Cox proportional hazards model

‡Measured at baseline

Table 3: Epigenome-wide association(EWA) scan to select candidate CpGs

CpG	Chr	Position	Nearest gene	Cocaine EWA effect size	Cocaine EWA p-value	VACS index* EWA effect size	VACS index* EWA p-value	Reference Gene Group	Relations to CpG Islands
cg22917487	3	39322103	*CX3CR1*	1.71E-02	1.69E-07	1.96E-04	1.73E-03	Body;TSS200;5UTR;TSS1500	
cg07839457	16	57023022	*NLRC5*	-5.18E-02	3.16E-05	-7.08E-04	4.02E-04	TSS1500	N_Shore
cg26312951	21	42797847	*MX1*	-3.93E-02	3.24E-05	-7.80E-04	2.50E-07	TSS200;5UTR	N_Shore
cg06188083	10	91093005	*IFIT3*	-5.06E-02	3.34E-05	-9.96E-04	4.76E-08	Body	
cg08122652	3	1.22E+08	*PARP9*	-6.13E-02	1.10E-04	-1.26E-03	2.30E-10	5UTR;TSS1500	N_Shore
cg22385827	2	2.11E+08	*C2orf67*	8.10E-03	2.01E-04	1.20E-04	7.52E-04	TSS1500	Island
cg22930808	3	1.22E+08	*PARP9*	-6.52E-02	2.44E-04	-1.17E-03	1.52E-06	5UTR;TSS1500	N_Shore
cg06981309	3	1.46E+08	*PLSCR1*	-4.29E-02	2.63E-04	-5.72E-04	4.21E-04	5UTR	N_Shore
cg03753191	13	43566902	*EPSTI1*	-1.36E-02	2.81E-04	-2.44E-04	5.50E-05	TSS1500	S_Shore
cg00096307	2	1.07E+08	*UXS1*	-1.12E-02	3.59E-04	1.74E-04	5.73E-04	Body	
cg26396492	20	19915762	*RIN2*	-1.26E-02	4.76E-04	-2.00E-04	9.15E-04	Body	
cg22940798	6	32805554	*TAP2*	-1.93E-02	4.82E-04	-2.89E-04	4.72E-04	Body	N_Shore
cg08623256	5	3858275		-1.73E-02	6.96E-04	3.17E-04	7.18E-04		S_Shelf
cg08818207	6	32820355	*TAP1*	-2.67E-02	8.18E-04	-6.18E-04	2.11E-07	Body	N_Shore
cg03038262	11	315262	*IFITM1*	-2.98E-02	9.59E-04	-6.46E-04	7.65E-09	3UTR	N_Shore

*VACS index: Veteran Aging Cohort Study index

medRxiv preprint doi: https://doi.org/10.1101/2020.05.11.20027458; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

Table 4: Mediation analyses on candidate CpGs between cocaine use and VACS index*

						Average Causal Mediation Effect			Proportion Mediated				
CpG	**Chr**	**Position**	**Nearest gene**	**Reference Gene Group**	**Relations to CpG Islands**	**Effect estimate**	**95% CI**	**p value**	**Proportion Mediated**	**95% CI**	**p value**	**Average Direct Effect**	**Total Effect**
cg26312951	21	42797847	*MX1*	TSS200;5UTR	N_Shore	0.097	(0.051,0.153)	1.00E-06	29.5%	(15.9%,53.0%)	1.00E-06	0.231	0.329
cg08122652	3	122281939	*PARP9*	5UTR;TSS1500	N_Shore	0.092	(0.049,0.142)	1.00E-06	28.2%	(15.7%,49.2%)	1.00E-06	0.236	0.329
cg06188083	10	91093005	*IFIT3*	Body		0.095	(0.052,0.145)	2.00E-06	28.8%	(16.2%,50.7%)	2.00E-06	0.234	0.329
cg03038262	11	315262	*IFITM1*	3UTR	N_Shore	0.078	(0.040,0.123)	4.00E-06	23.7%	(12.4%,42.8%)	4.00E-06	0.251	0.329
cg07839457	16	57023022	*NLRC5*	TSS1500	N_Shore	0.079	(0.038,0.127)	8.00E-06	24.0%	(11.0%,47.5%)	8.00E-06	0.250	0.329
cg22930808	3	122281881	*PARP9*	5UTR;TSS1500	N_Shore	0.091	(0.047,0.141)	8.00E-06	27.7%	(15.1%,48.9%)	8.00E-06	0.237	0.329
cg03753191	13	43566902	*EPSTI1*	TSS1500	S_Shore	0.078	(0.036,0.128)	3.90E-05	23.7%	(10.6%,46.8%)	3.90E-05	0.251	0.329
cg06981309	3	146260954	*PLSCR1*	5UTR	N_Shore	0.074	(0.033,0.123)	4.70E-05	22.6%	(10.7%,41.8%)	4.70E-05	0.254	0.329
cg22940798	6	32805554	*TAP2*	Body	N_Shore	0.058	(0.025,0.100)	5.70E-05	17.6%	(7.6%,35.6%)	5.70E-05	0.271	0.329
cg08818207	6	32820355	*TAP1*	Body	N_Shore	0.081	(0.037,0.130)	5.80E-05	24.5%	(11.9%,44.7%)	5.90E-05	0.248	0.329
cg22917487	3	39322103	*CX3CR1*	Body;TSS200;5UTR;TSS1500		0.054	(0.015,0.104)	1.81E-03	16.3%	(4.6%,34.4%)	1.81E-03	0.276	0.329
cg26396492	20	19915762	*RIN2*	Body		0.037	(0.009,0.073)	3.07E-03	11.3%	(2.7%,24.9%)	3.07E-03	0.291	0.329

*VACS index: Veteran Aging Cohort Study index

medRxiv preprint doi: https://doi.org/10.1101/2020.05.11.20027458; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

Table 5: Gene ontology (GO) pathway enrichment analysis and joint mediation effect

Term	n	%	P value	Genes	CpG	Fold Enrichment	FDR*
GO:0016032~viral process	8	73	7.76E-08	*PLSCR1, CX3CR1, IFITM1, IFIT3, MX1, TAP2, NLRC5, TAP1*	cg06981309,cg22917487,cg03038262,cg06188083,cg26312951,cg22940798,cg07839457,cg08818207	15.1	9.89E-05
GO:0034097~response to cytokine	7	64	9.88E-07	*PARP9, PLSCR1, CX3CR1, IFITM1, IFIT3, MX1, NLRC5*	cg08122652,cg06981309,cg22917487,cg03038262,cg06188083,cg26312951,cg07839457	15.9	1.26E-03
GO:0051607~defense response to virus	5	45	4.40E-06	*PLSCR1, IFITM1, IFIT3, MX1, NLRC5*	cg06981309,cg03038262,cg06188083,cg26312951,cg07839457	39.8	5.61E-03
GO:0050792~regulation of viral process	5	45	7.50E-06	*PLSCR1, IFITM1, MX1, TAP2, TAP1*	cg06981309,cg03038262,cg26312951,cg22940798,cg08818207	34.8	9.56E-03
GO:0034340~response to type I interferon	4	36	8.49E-06	*IFITM1, IFIT3, MX1, NLRC5*	cg03038262,cg06188083,cg26312951,cg07839457	93.5	1.08E-02
GO:0009615~response to virus	5	45	1.48E-05	*PLSCR1, IFITM1, IFIT3, MX1, NLRC5*	cg06981309,cg03038262,cg06188083,cg26312951,cg07839457	29.3	1.88E-02

*FDR: False discovery rate

medRxiv preprint doi: https://doi.org/10.1101/2020.05.11.20027458; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.11.20027458; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Figure Legends

Figure 1. Timeline of data and blood sample collection for each analysis

Figure 2. Kaplan Meier curves by cocaine use frequency at baseline among HIV-positive (n=1,463) and HIV-negative (n=795) participants. The higher frequency of cocaine use is associated with lower survival probability among HIV-positive participants, but not among HIV-negative participants.

Figure 3. DNA methylation level of the selected CpG mediators by persistent cocaine use status

Figure 4. Significant mediation effect of cg26312951 (*MX1*),cg08122652(*PARP9*), cg07839457 (*NLRC5*) and cg22917487 (*CX3CR1*) between persistent cocaine use and HIV severity ($p<0.0033$)

Figure S1: Manhattan and quantile-quantile (QQ) plot of persistent cocaine use epigenome-wide association (EWA) ($\lambda=1.034$) and HIV severity EWA ($\lambda=1.116$)

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. All participants provided written consents.

Availability of data and materials

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

Conflict of interest

All authors declare that they have no conflict of interest.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.11.20027458; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Funding

The work was supported by the National Institute on Drug Abuse (R03DA039745, R01DA038632, R01DA047063, R01DA047820).

Authors' contributions

CS was responsible for data analysis and manuscript preparation. ACJ provided DNA samples, clinical data, and contributed to manuscript preparation. XZ was responsible for the bioinformatics data processing. ZW contributed to mediation analysis and manuscript preparation. DH and EJ contributed to analytical approach and the manuscript preparation. KX was responsible for the study design, study protocol, sample preparation, data analysis, interpretation of findings, and manuscript preparation.

Acknowledgements The authors appreciate the support of the Veteran Aging Study Cohort Biomarker Core and Yale Center of Genomic Analysis

medRxiv preprint doi: https://doi.org/10.1101/2020.05.11.20027458; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Reference

1. Chaisson RE, Bacchetti P, Osmond D, Brodie B, Sande MA, Moss AR. Cocaine use and HIV infection in intravenous drug users in San Francisco. Jama **1989**; 261(4): 561-5.
2. Tyndall MW, Currie S, Spittal P, et al. Intensive injection cocaine use as the primary risk factor in the Vancouver HIV-1 epidemic. Aids **2003**; 17(6): 887-93.
3. Webber MP, Schoenbaum EE, Gourevitch MN, Buono D, Klein RS. A prospective study of HIV disease progression in female and male drug users. Aids **1999**; 13(2): 257-62.
4. Vittinghoff E, Hessol NA, Bacchetti P, Fusaro RE, Holmberg SD, Buchbinder SP. Cofactors for HIV disease progression in a cohort of homosexual and bisexual men. Journal of acquired immune deficiency syndromes (1999) **2001**; 27(3): 308-14.
5. Cook JA, Burke-Miller JK, Cohen MH, et al. Crack cocaine, disease progression, and mortality in a multi-center cohort of HIV-1 positive women. AIDS (London, England) **2008**; 22(11): 1355.
6. Baum MK, Rafie C, Lai S, Sales S, Page B, Campa A. Crack-cocaine use accelerates HIV disease progression in a cohort of HIV-positive drug users. JAIDS Journal of Acquired Immune Deficiency Syndromes **2009**; 50(1): 93-9.
7. Sharpe TT, Lee LM, Nakashima AK, Elam-Evans LD, Fleming PL. Crack cocaine use and adherence to antiretroviral treatment among HIV-infected black women. J Community Health **2004**; 29(2): 117-27.
8. Cofrancesco Jr J, Scherzer R, Tien PC, et al. Illicit drug use and HIV treatment outcomes in a US cohort. AIDS (London, England) **2008**; 22(3): 357.
9. Carrico AW, Johnson MO, Morin SF, et al. Stimulant use is associated with immune activation and depleted tryptophan among HIV-positive persons on anti-retroviral therapy. Brain, behavior, and immunity **2008**; 22(8): 1257-62.
10. Dash S, Balasubramaniam M, Villalta F, Dash C, Pandhare J. Impact of cocaine abuse on HIV pathogenesis. Front Microbiol **2015**; 6: 1111-.
11. Rasbach DA, Desruisseau AJ, Kipp AM, et al. Active cocaine use is associated with lack of HIV-1 virologic suppression independent of nonadherence to antiretroviral therapy: use of a rapid screening tool during routine clinic visits. AIDS care **2013**; 25(1): 109-17.
12. Breitling LP, Yang R, Korn B, Burwinkel B, Brenner H. Tobacco-smoking-related differential DNA methylation: 27K discovery and replication. The American Journal of Human Genetics **2011**; 88(4): 450-7.
13. Joubert BR, Håberg SE, Nilsen RM, et al. 450K epigenome-wide scan identifies differential DNA methylation in newborns related to maternal smoking during pregnancy. Environmental health perspectives **2012**; 120(10): 1425-31.
14. Lee KW, Pausova Z. Cigarette smoking and DNA methylation. Frontiers in genetics **2013**; 4: 132.
15. Tsaprouni LG, Yang T-P, Bell J, et al. Cigarette smoking reduces DNA methylation levels at multiple genomic loci but the effect is partially reversible upon cessation. Epigenetics **2014**; 9(10): 1382-96.
16. Gao X, Zhang Y, Saum K-U, Schöttker B, Breitling LP, Brenner H. Tobacco smoking and smoking-related DNA methylation are associated with the development of frailty among older adults. Epigenetics **2017**; 12(2): 149-56.
17. Zhang X, Hu Y, Aouizerat BE, et al. Machine learning selected smoking-associated DNA methylation signatures that predict HIV prognosis and mortality. **2018**; 10(1): 155.
18. Zhang R, Miao Q, Wang C, et al. Genome-wide DNA methylation analysis in alcohol dependence. Addiction biology **2013**; 18(2): 392-403.
19. Zhang H, Gelernter J. DNA methylation and alcohol use disorders: Progress and challenges. The American journal on addictions **2017**; 26(5): 502-15.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.11.20027458; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

20. Sharp GC, Arathimos R, Reese SE, et al. Maternal alcohol consumption and offspring DNA methylation: findings from six general population-based birth cohorts. Epigenomics **2018**; 10(1): 27-42.
21. Liu C, Marioni RE, Hedman ÅK, et al. A DNA methylation biomarker of alcohol consumption. Molecular psychiatry **2018**; 23(2): 422.
22. Feinberg AP, Koldobskiy MA, Göndör A. Epigenetic modulators, modifiers and mediators in cancer aetiology and progression. Nature Reviews Genetics **2016**; 17: 284.
23. Hao X, Luo H, Krawczyk M, et al. DNA methylation markers for diagnosis and prognosis of common cancers. Proceedings of the National Academy of Sciences **2017**; 114(28): 7414.
24. Michalak EM, Burr ML, Bannister AJ, Dawson MA. The roles of DNA, RNA and histone methylation in ageing and cancer. Nature Reviews Molecular Cell Biology **2019**.
25. Davegårdh C, García-Calzón S, Bacos K, Ling C. DNA methylation in the pathogenesis of type 2 diabetes in humans. Mol Metab **2018**; 14: 12-25.
26. Zhong J, Agha G, Baccarelli AA. The role of DNA methylation in cardiovascular risk and disease: methodological aspects, study design, and data analysis for epidemiological studies. Circulation research **2016**; 118(1): 119-31.
27. Nakatochi M, Ichihara S, Yamamoto K, et al. Epigenome-wide association of myocardial infarction with DNA methylation sites at loci related to cardiovascular disease. Clinical epigenetics **2017**; 9(1): 54.
28. 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.
29. Liu Y, Aryee MJ, Padyukov L, et al. Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. Nature biotechnology **2013**; 31(2): 142.
30. Bind M-A, Lepeule J, Zanobetti A, et al. Air pollution and gene-specific methylation in the Normative Aging Study: association, effect modification, and mediation analysis. Epigenetics **2014**; 9(3): 448-58.
31. Cao-Lei L, Dancause KN, Elgbeili G, et al. DNA methylation mediates the impact of exposure to prenatal maternal stress on BMI and central adiposity in children at age 13½ years: Project Ice Storm. Epigenetics **2015**; 10(8): 749-61.
32. Timms JA, Relton CL, Rankin J, Strathdee G, McKay JA. DNA methylation as a potential mediator of environmental risks in the development of childhood acute lymphoblastic leukemia. Epigenomics **2016**; 8(4): 519-36.
33. Tobi EW, Slieker RC, Luijk R, et al. DNA methylation as a mediator of the association between prenatal adversity and risk factors for metabolic disease in adulthood. Science advances **2018**; 4(1): eaao4364.
34. Barker ED, Walton E, Cecil CA. Annual Research Review: DNA methylation as a mediator in the association between risk exposure and child and adolescent psychopathology. Journal of Child Psychology and Psychiatry **2018**; 59(4): 303-22.
35. Rutten BPF, Mill J. Epigenetic Mediation of Environmental Influences in Major Psychotic Disorders. Schizophrenia Bulletin **2009**; 35(6): 1045-56.
36. Jordahl KM, Phipps AI, Randolph TW, et al. Differential DNA methylation in blood as a mediator of the association between cigarette smoking and bladder cancer risk among postmenopausal women. Epigenetics **2019**: 1-9.
37. Shirazi J, Shah S, Sagar D, et al. Epigenetics, drugs of abuse, and the retroviral promoter. J Neuroimmune Pharmacol **2013**; 8(5): 1181-96.
38. Dhillon NK, Williams R, Peng F, et al. Cocaine-mediated enhancement of virus replication in macrophages: implications for human immunodeficiency virus-associated dementia. J Neurovirol **2007**; 13(6): 483-95.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.11.20027458; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

39. Justice AC, Dombrowski E, Conigliaro J, et al. Veterans aging cohort study (VACS): overview and description. **2006**; 44(8 Suppl 2): S13.
40. Veterans Aging Cohort Study. VACS Biomarker Cohort Description. **2016**.
41. Tate JP, Justice AC, Hughes MD, et al. An internationally generalizable risk index for mortality after one year of antiretroviral therapy. AIDS (London, England) **2013**; 27(4): 563-72.
42. Justice AC, Modur SP, Tate JP, et al. Predictive accuracy of the Veterans Aging Cohort Study index for mortality with HIV infection: a North American cross cohort analysis. Journal of acquired immune deficiency syndromes (1999) **2013**; 62(2): 149-63.
43. Bebu I, Tate J, Rimland D, et al. The VACS Index Predicts Mortality in a Young, Healthy HIV Population Starting Highly Active Antiretroviral Therapy. J Acquir Immune Defic Syndr **2014**; 65(2): 226-30.
44. Brown ST, Tate JP, Kyriakides TC, et al. The VACS index accurately predicts mortality and treatment response among multi-drug resistant HIV infected patients participating in the options in management with antiretrovirals (OPTIMA) study. PLoS One **2014**; 9(3): e92606.
45. Justice AC, Modur S, Tate JP, et al. Predictive accuracy of the Veterans Aging Cohort Study (VACS) index for mortality with HIV infection: a north American cross cohort analysis. Journal of acquired immune deficiency syndromes (1999) **2013**; 62(2): 149.
46. Lehne B, Drong AW, Loh M, et al. A coherent approach for analysis of the Illumina HumanMethylation450 BeadChip improves data quality and performance in epigenome-wide association studies. **2015**; 16(1): 37.
47. Aryee MJ, Jaffe AE, Corrada-Bravo H, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. **2014**; 30(10): 1363-9.
48. Houseman EA, Accomando WP, Koestler DC, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics **2012**; 13(1): 86.
49. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics **2007**; 8(1): 118-27.
50. Kassambara A, Kosinski M, Biecek P. survminer: Drawing Survival Curves using'ggplot2'. R package version 03 **2017**; 1.
51. Therneau T. A Package for Survival Analysis in S. version 2.38. **2015**.
52. Zhang X, Hu Y, Justice AC, et al. DNA methylation signatures of illicit drug injection and hepatitis C are associated with HIV frailty. Nature Communications **2017**; 8(1): 2243.
53. Jaffe A. Flowsorted. blood. 450k: illumina humanmethylation data on sorted blood cell populations. R Package Version **2015**; 1(0).
54. Tingley D, Yamamoto T, Hirose K, Keele L, Imai K. Mediation: R package for causal mediation analysis. **2014**.
55. VanderWeele T, Vansteelandt S. Mediation analysis with multiple mediators. Epidemiologic methods **2014**; 2(1): 95-115.
56. Fresno C, Fernández EAJB. RDAVIDWebService: a versatile R interface to DAVID. **2013**; 29(21): 2810-1.
57. Schmeisser H, Mejido J, Balinsky CA, et al. Identification of alpha interferon-induced genes associated with antiviral activity in Daudi cells and characterization of IFIT3 as a novel antiviral gene. J Virol **2010**; 84(20): 10671-80.
58. Cui J, Zhu L, Xia X, et al. NLRC5 negatively regulates the NF-kappaB and type I interferon signaling pathways. Cell **2010**; 141(3): 483-96.
59. Garin A, Tarantino N, Faure S, et al. Two novel fully functional isoforms of CX3CR1 are potent HIV coreceptors. J Immunol **2003**; 171(10): 5305-12.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.11.20027458; this version posted May 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

60. Faure S, Meyer L, Costagliola D, et al. Rapid progression to AIDS in HIV+ individuals with a structural variant of the chemokine receptor CX3CR1. Science **2000**; 287(5461): 2274-7.

61. Connelly KL, Unterwald EM. Chronic cocaine administration upregulates FKBP5 in the extended amygdala of male and female rats. Drug Alcohol Depend **2019**; 199: 101-5.

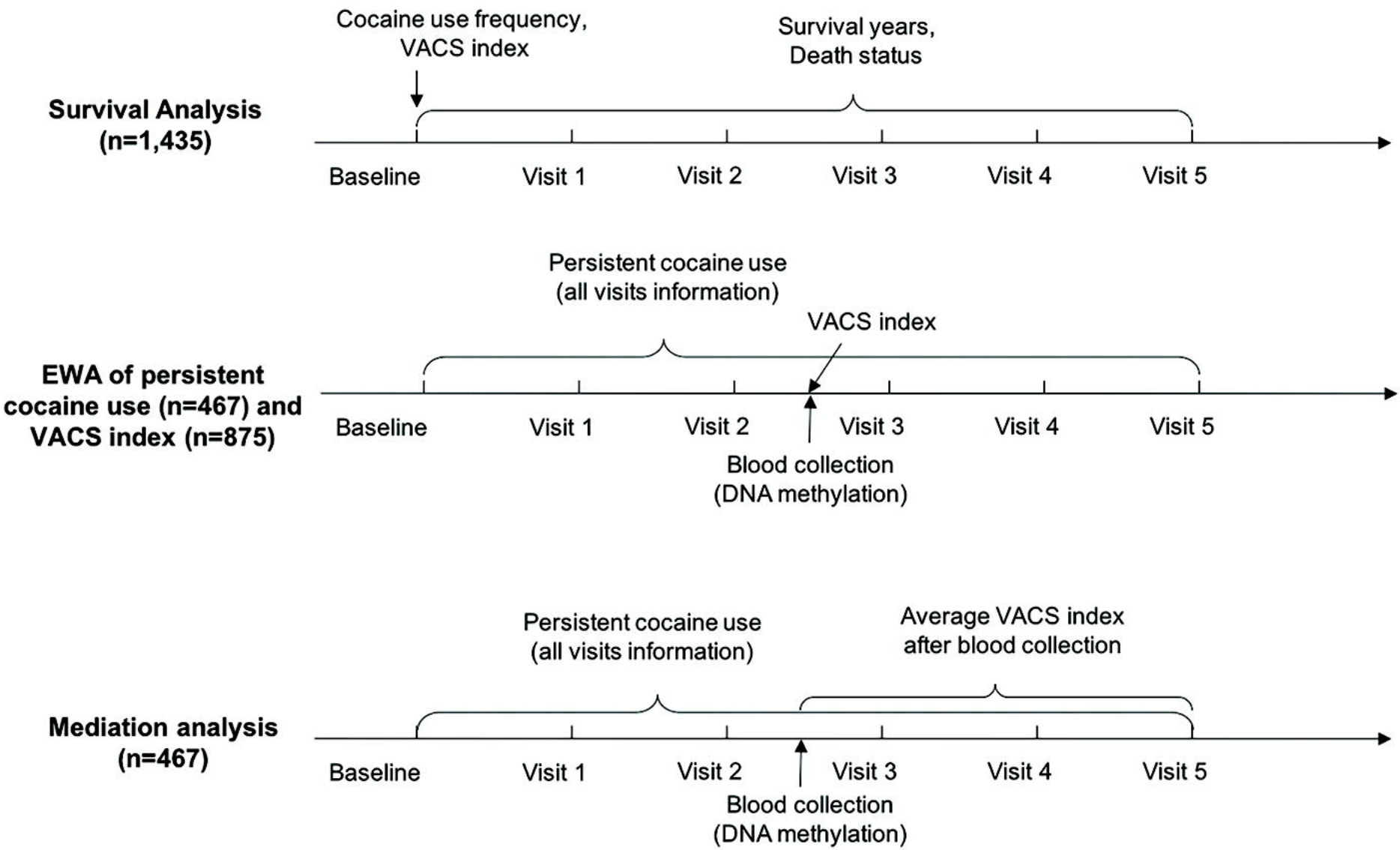
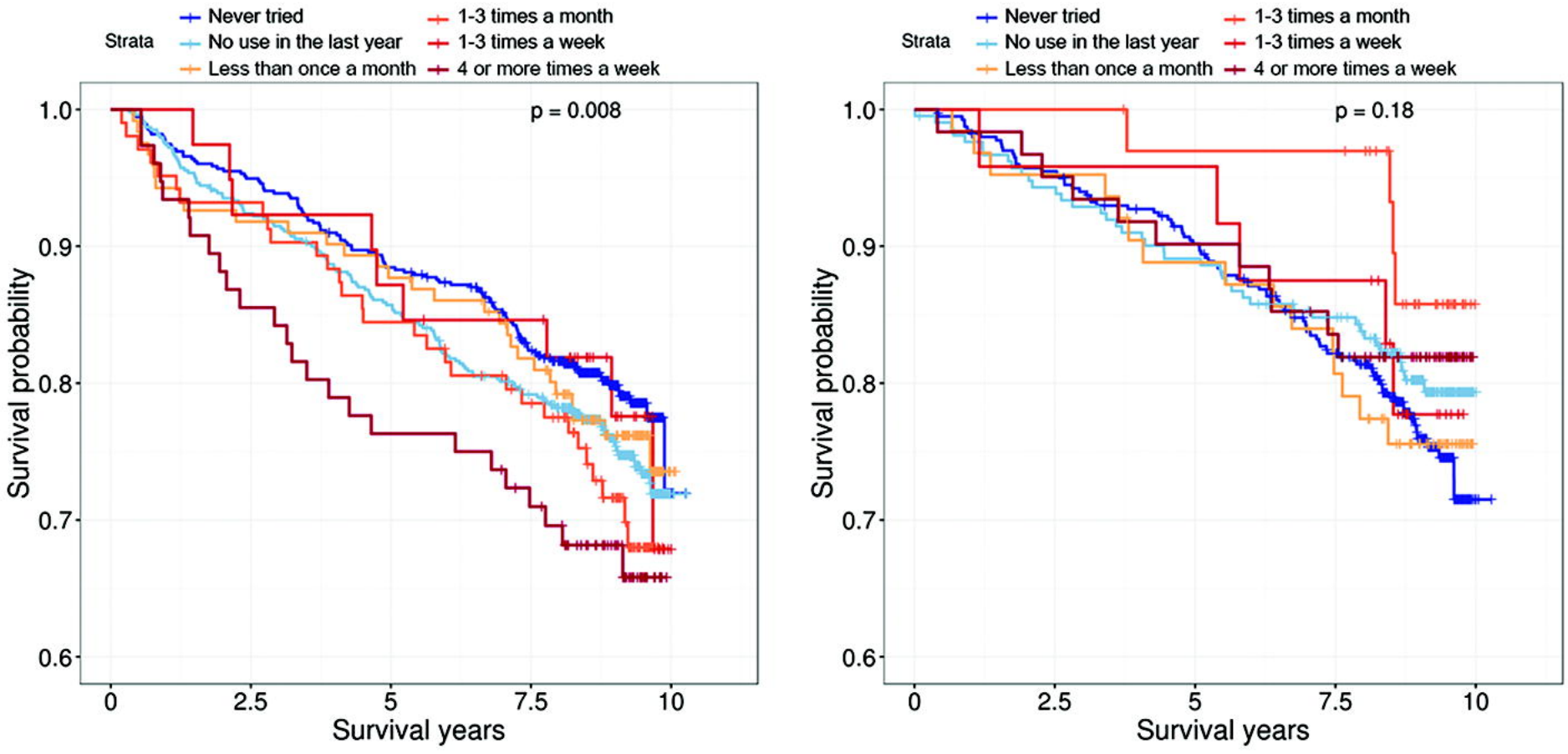


Figure 1. Timeline of data and blood sample collection for each analysis



a. HIV-Positive

b. HIV-Negative

Figure 2. Kaplan Meier curves by cocaine use frequency at baseline among HIV positive (n=1435) and HIV negative(n=795) participants

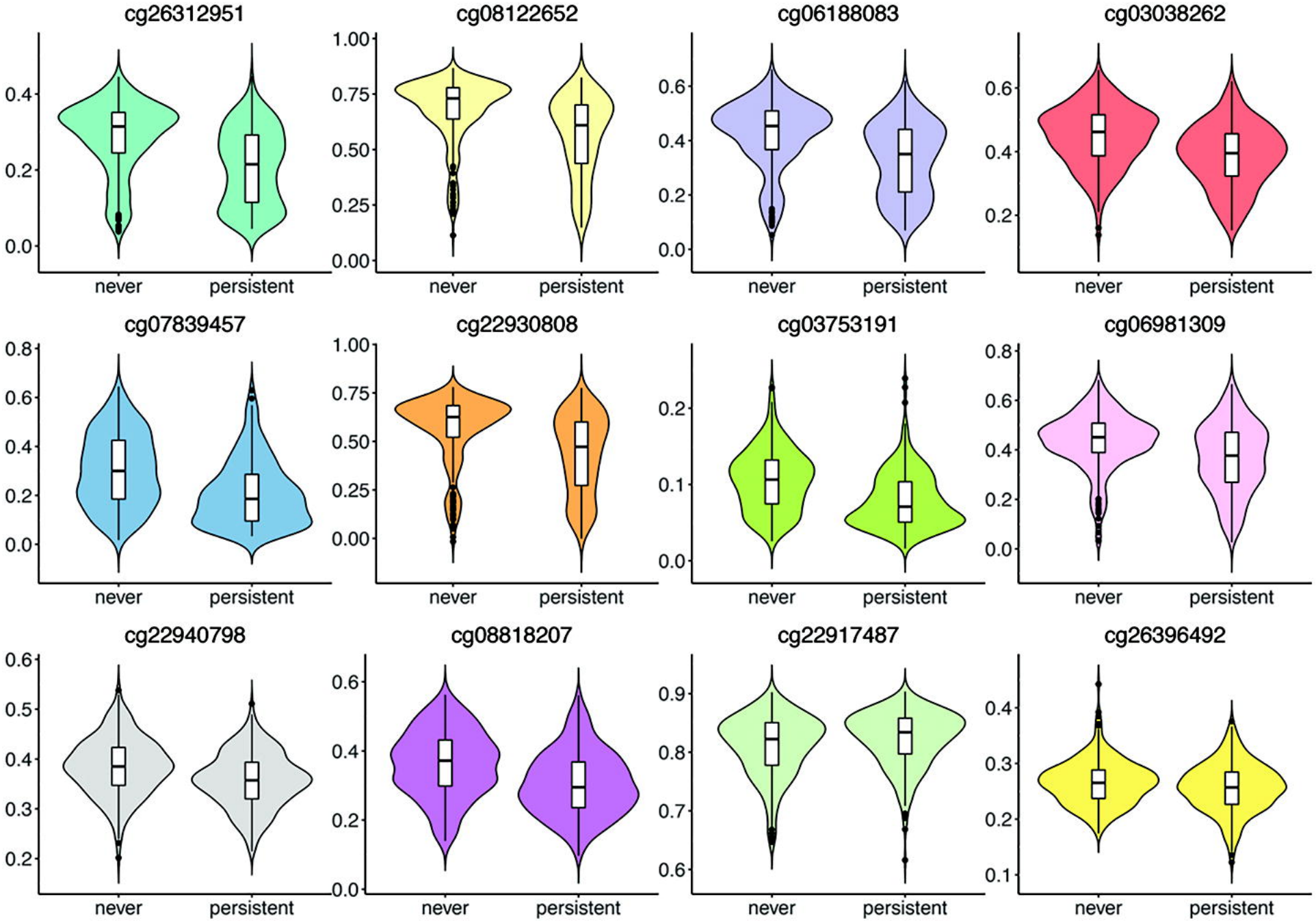


Figure 3. DNA methylation level of the selected CpG mediators by persistent cocaine use status

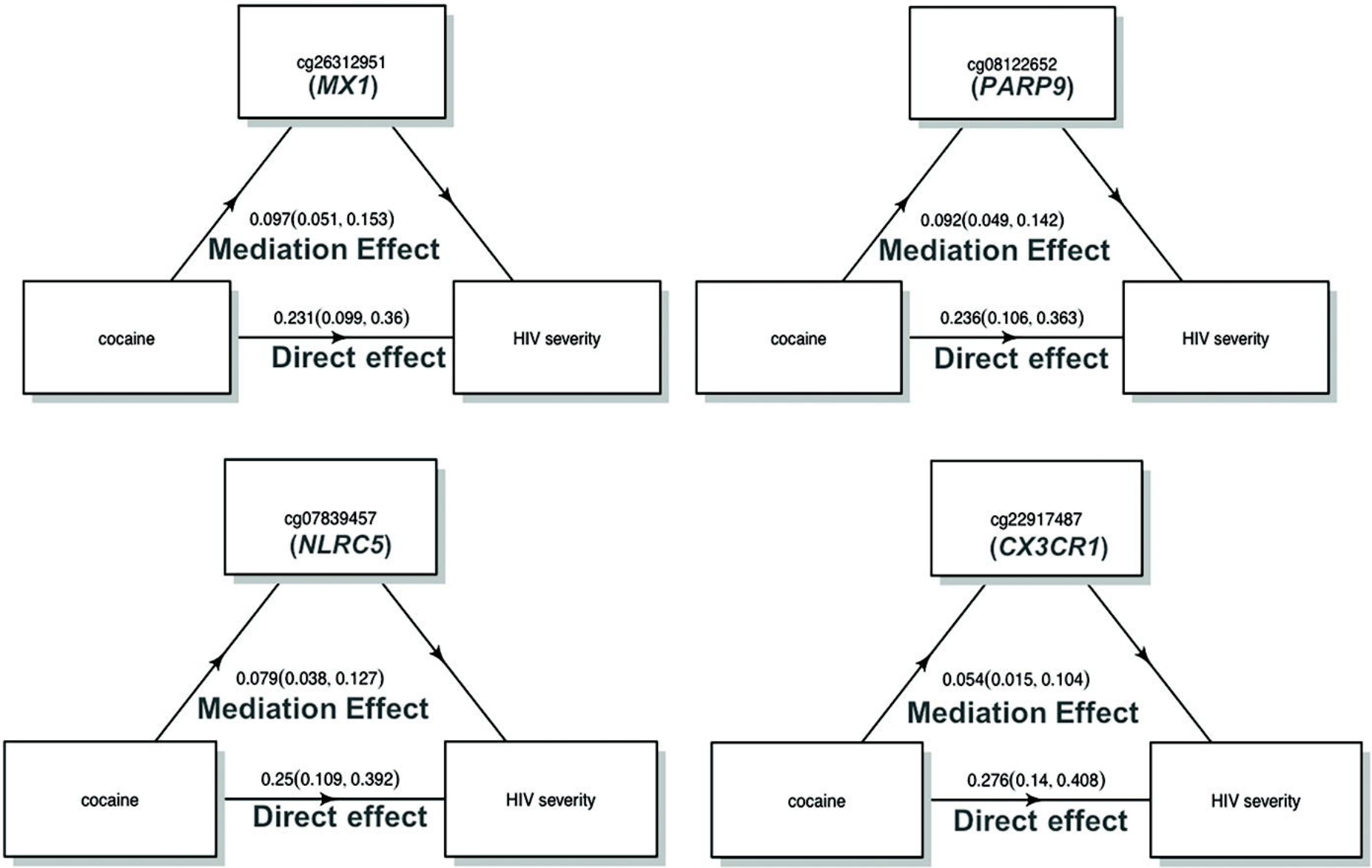


Figure 4. Significant mediation effect of cg26312951 (*MX1*),cg08122652(*PARP9*), cg07839457 (*NLRC5*) and cg22917487 (*CX3CR1*) between persistent cocaine use and HIV severity ($p<0.0033$)