

Epigenome-wide characterization reveals aberrant DNA methylation of host genes regulating CD4+ T cell HIV-1 reservoir size in women with HIV

Ke Xu, MD, PhD^{1,2}, Xinyu Zhang, PhD^{1,2}, Kesava Asam^{3,4}, MS, Bryan C. Quach, PhD⁵, Grier P. Page, PhD^{5,6}
Deborah Konkle-Parker, PhD⁷, Claudia Martinez, MD⁸, Cecile D. Lahiri, MD⁹, Elizabeth T. Topper, PhD¹⁰, Mardge H. Cohen, MD¹¹, Seble G. Kassaye, MD¹², Jack DeHovitz, MD¹³, Mark H. Kuniholm, PhD¹⁴, Nancie M. Archin, PhD^{15,16}, Amir Valizadeh, MD^{1,2}, Phyllis C. Tien, MD^{17,18}, Vincent C. Marconi¹⁹, Dana B. Hancock, PhD⁵, Eric Otto Johnson, PhD^{5,6}, Bradley E. Aouizerat, PhD^{3,4}

1. Department of Psychiatry, School of Medicine, Yale University, New Haven, CT, United States.
2. Veteran's Affairs Connecticut Healthcare System, West Haven, CT, United States.
3. Department of Oral and Maxillofacial Surgery, New York University, New York, NY, United States.
4. Translational Research Center, New York University, New York, NY, United States.
5. GenOmics, Bioinformatics, and Translational Research Center, Biostatistics and Epidemiology Division, RTI International, Research Triangle Park, NC, United States.
6. Fellow Program, RTI International, Research Triangle Park, NC, United States.
7. Schools of Nursing, Medicine, and Population Health, University of Mississippi Medical Center, Jackson, MS, United States.
8. Miller School of Medicine, Division of Cardiovascular Medicine, University of Miami, Miami, FL, United States.
9. Department of Medicine, Division of Infectious Diseases, Emory University School of Medicine, Atlanta, GA, United States.
10. Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, United States.
11. Department of Medicine, Stroger Hospital, Cook County Health System, Chicago IL, United States.
12. Department of Medicine, Division of Infectious Diseases, Georgetown University, Washington, D.C., United States.
13. Department of Medicine, Division of Infectious Diseases, Downstate Health Sciences University, Brooklyn, NY, United States.
14. Department of Epidemiology and Biostatistics, University at Albany, State University of New York, Rensselaer, NY, United States.
15. UNC HIV Cure Center, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, NC, United States.

Xu et al., Epigenome-wide characterization reveals aberrant DNA methylation of host genes regulating CD4+ T cell HIV-1 reservoir size in women with HIV

16. Department of Medicine, Division of Infectious Diseases, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, NC, United States.

17. Department of Medicine, University of California, San Francisco, San Francisco, CA, United States.

18. Department of Veterans Affairs Medical Center, San Francisco, CA, United States.

19. Emory University School of Medicine and Rollins School of Public Health; the Atlanta Veterans Affairs Medical Center, Atlanta, GA, United States.

Corresponding author:

Ke Xu, MD, PhD

Department of Psychiatry, Yale School of Medicine, New Haven, CT, United States.

Email: ke.xu@yale.edu

Xu et al., Epigenome-wide characterization reveals aberrant DNA methylation of host genes regulating CD4⁺ T cell HIV-1 reservoir size in women with HIV

Abstract

The underlying mechanism of the HIV-1 reservoir, a major barrier to an HIV cure, is largely unknown. The integration of HIV-1 DNA and immune defense mechanisms can disrupt the host epigenetic landscape, potentially silencing HIV-1 replication. Using bisulfite capture DNA methylation sequencing, we profiled approximately 3.2 million CpG sites in CD4⁺ T cells isolated from the blood of 427 virally suppressed women with HIV. The average total CD4⁺ T cell HIV-1 Reservoir (HR_{CD4}) size was 1,409 copies per million cells. Most proviruses were defective with only a small proportion being intact. We found 245 differentially methylated positions (CpG sites) and 85 methylated regions associated with the total HR_{CD4} size. Notably, 52% of significant methylation sites were in intronic regions. HR_{CD4}-associated genes were involved in viral replication (e.g., *ISG15*), HIV-1 latency (e.g., *MBD2*), and cell growth and apoptosis (e.g., *IRF9*). A subset of the identified genes with aberrant methylation was an established target of HIV-1 integration (e.g., *NFIA*, *SPPL3*, *DLEU2*, *ELMSAN1*). Overall, HR_{CD4} size was inversely associated with DNA methylation of interferon signaling genes and positively associated with methylation at established HIV-1 integration sites. HR_{CD4}-associated genes were enriched in pathways including immune defense against the virus (i.e., interferon- α response and interferon- γ response), DNA binding transcription repression, and host-virus interaction such as Tau protein binding. Together, our results show that epigenomic alterations in CD4⁺ T cells are associated with total HIV-1 reservoir size, offering new insights into HIV-1 latency and may provide potential molecular targets for future HIV-1 eradication strategies.

Xu et al., Epigenome-wide characterization reveals aberrant DNA methylation of host genes regulating CD4⁺ T cell HIV-1 reservoir size in women with HIV

Introduction

The HIV-1 reservoir (HR) has posed a major barrier to an HIV cure. HIV latency is established early during infection, with the provirus persisting primarily in resting or memory CD4⁺ T cells^{1,2}. The HR evades eradication by the immune system and is largely suppressed by antiretroviral therapy (ART). However, cessation of ART leads to HIV-1 rebound within approximately 2 weeks³. The HR size decays slowly with an estimated half-life of 44 months for the total HR⁴ and 59 months for the intact provirus⁵. At this decay rate, eradication would take over 60 years, necessitating lifelong ART for people with HIV (PWH).

In recent years, efforts to cure HIV have focused on latency reversal followed by immune-mediated clearance of infected cells. However, latency reversal agents have failed to effectively reduce HR size, presenting a major challenge^{6,7}. Another challenge is the limited understanding of how HR establishes, maintains, and rebounds in CD4⁺ T cells before, during, and after ART discontinuation.

Emerging evidence shows that epigenetic regulation of both the viral and host genomes is crucial to establishing and maintaining HIV-1 latency. Transcription of provirus depends upon the dynamic interaction between genomic alteration of the provirus and host immune activity⁸. The HIV-1 provirus genome reduces chromatin accessibility in latently infected cells⁹. Mechanistically, two CpG sites in the promoter region of long terminal repeats (LTR) of the provirus are hypermethylated, mediated by the host methyl-CpG binding domain protein 2 (MBD2) in infected CD4⁺ T cells¹⁰. In contrast, some studies reported that methylation of LTR was not associated with HIV-1 latency^{11,12}. Research on the host genome's DNA methylation and HR is still in its

Xu et al., Epigenome-wide characterization reveals aberrant DNA methylation of host genes regulating CD4⁺ T cell HIV-1 reservoir size in women with HIV

infancy. Some studies have reported that hyper- and hypo-methylated regions are associated with HR size and positively correlate with disease progression assessed by viral response to ART¹³, but these findings are based on small sample sizes or *ex vivo* experiments, making it challenging to draw strong conclusions about *in vivo* host DNA methylation dynamics associated with HR.

Genome-wide DNA methylation analysis with high-density CpG profiling is a useful approach to comprehensively investigate the role of the host DNA methylome in HR size. Among the total reservoir in infected CD4⁺ T cells, only a small proportion contains intact provirus (~1 per million cells) capable of full transcription activation. Most proviruses are defective because of apolipoprotein B mRNA editing enzyme catalytic polypeptide-like 3G (*APOBEC3C*)-mediated hypermutation¹⁴, large internal deletions, packaging signal (Ψ) deletions, major splice donor mutations, and inactivating point mutations¹⁵. Most defective proviruses accumulate swiftly during early HIV-1 infection¹⁴. While previous studies have focused on intact reservoirs, recent research indicates that defective provirus can be activated to produce HIV-1 transcripts when integrated at active transcription sites¹⁵, and it can be recognized by cytotoxic T cells¹⁶. Proviruses escape mutations and remain stable during ART¹⁷, leading to chronic immune activation and inflammation, which are associated with the progression of HIV-related comorbidities¹⁸⁻²⁰. Thus, it is plausible that host epigenome alterations occur in response to both intact and defective reservoirs after HIV-1 integration. Characterization of the perturbed host epigenome by intact and defective proviruses may help us discover new targets to eradicate the reservoir.

Xu et al., Epigenome-wide characterization reveals aberrant DNA methylation of host genes regulating CD4⁺ T cell HIV-1 reservoir size in women with HIV

In this study, we examined the association of DNA methylation features with HR size in CD4⁺ T cells (HR_{CD4}) from a cohort of women with HIV (WWH) who were virally suppressed on ART (N=427). DNA samples were extracted from isolated CD4⁺ T cells. Using a methylation capture sequencing method, we profiled DNA methylation sites across the epigenome and found 245 methylation sites associated with total HR_{CD4}. The genes annotated for significant methylated sites were involved in HIV-1 latency, inflammation, and cell differentiation and harbored HIV-1 integration sites. These findings provide new evidence on the potential roles of host epigenetic factors in HR size through type I interferon genes, immune activation, and chromosome stability.

Results

Participant characteristics and HIV-1 latent reservoir detection

The study population was derived from the Women's Interagency HIV Study (WIHS) a multi-center interval cohort study of women with HIV (WWH) and a comparison group of women without HIV infection in the United States^{21,22}. In 2019, WIHS merged with the Multicenter AIDS Cohort Study (MACS) to form the MACS/WIHS Combined Cohort Study (MWCCS)²³. Participants underwent follow-up visits every six months which included the collection of peripheral blood mononuclear cells (PBMC) and demographic, clinical, and ART treatment data. The 427 participants who met the inclusion and exclusion criteria were selected for the current study (See **Methods**). The average age was 47.3 ± 8.2 years; 76% of the participants self-identified as African American non-Hispanic, 12% as White non-Hispanic, and 11% as other races. Most participants (80%) currently smoked tobacco. All participants were on ART and had

Xu et al., Epigenome-wide characterization reveals aberrant DNA methylation of host genes regulating CD4+ T cell HIV-1 reservoir size in women with HIV

undetectable viral loads (below the assay lower limit of quantitation) for a minimum of six months, with an average duration of 2.1 years, up to the visit at which CD4+ T cells were isolated from PBMCs and were processed to estimate the HR_{CD4}. The genomic DNA extracted from CD4+ T cells was subjected to bisulfite capture sequencing for DNA methylation profiling.

The HR_{CD4} quantification method for this cohort was detailed previously by Aouizerat et al.²⁴. Briefly, HR_{CD4} was detected using droplet digital polymerase chain reaction (ddPCR). Intact proviral HIV DNA was measured using a modified intact proviral DNA assay (IPDA). The average HR_{CD4} size was 1,409 ± 1,639 copies per million CD4⁺ T cells. As expected, defective HIV-1 proviruses were more common than intact HR_{CD4}. The average intact HR_{CD4} was 279 ± 463 copies per million; defective HR_{CD4} was 531 ± 886 copies per million for APOBEC3G-mediated mutations (3'-defective HR_{CD4}) and 600 ± 748 copies per million for large internal deletions (5'-defective HR_{CD4}). Demographic and clinical characteristics are presented in **Table 1**.

Epigenome-wide methylation characterization of HR_{CD4}

CD4⁺ T cells epigenome was profiled using bisulfite methylation capture sequencing, yielding data on methylation of 3.2 million CpG sites. Quality control (QC) was conducted according to our previous report²⁵ (see **Methods**). Only CpG sites with a coverage greater than 10x depth were included. Approximately 2 million CpG sites were analyzed after QC. Ancestry and chronological age were included in generalized linear regression models.

Xu et al., Epigenome-wide characterization reveals aberrant DNA methylation of host genes regulating CD4+ T cell HIV-1 reservoir size in women with HIV

We found 245 differentially methylated sites associated with total HR_{CD4} size (False Discovery Rate, FDR<0.05) (**Figure 1A, Supplementary Table**). Among them, 67% (n=165) were positively associated with HR_{CD4} size while 33% (n=80) were negatively associated with it (**Figure 1B**). A substantial proportion of these methylation sites (89%) were annotated within or near genes, with the rest (11%) in intergenic regions. Within gene regions, 52% of CpG sites were in introns, 20% in promoter regions, and 11% in exons (**Figure 1C**). Additionally, 26% of the associated sites mapped to CpG islands, with the rest in non-island regions (open sea 34%, shore 30%, and shelf 9%). Notably, although most of the observed associations were positive, 50% of the methylation sites in the promoter regions were negatively associated with total HR_{CD4}.

We further examined whether the total HR_{CD4}-associated CpG sites were also significantly associated with intact and defective HR_{CD4}. Among the 245 significant CpG sites for total HR_{CD4}, 38 were nominally associated with intact or defective HR_{CD4} (p<0.05). For example, *SMC4* on chromosome 3 at position 160119770 was negatively associated with total (t=-5.79, p=1.64e-08), intact (t=-4.0, p=7.63e-05), and 5'-defective (t=-4.78, p=2.64e-06) HR_{CD4}. *PARP9* on chromosome 3 at position 122281940 was also negatively associated with total (t=-5.76, p=1.92e-08), intact (t=-3.19, p=1.57e-03), 5'-defective (t=-3.66, p=2.93e-04), and 3'-defective (t=-4.28, p=2.39e-05) HR_{CD4}. *ISG15* on chromosome 1 at position 949432 was negatively associated with total (t=-5.14, p=4.66e-07), intact (t=-3.39, p=7.90e-04), and 5'-defective (t=-4.53, p=8.24e-06) HR_{CD4}. These genes were previously reported to be involved in interaction with HIV-1 tau

Xu et al., Epigenome-wide characterization reveals aberrant DNA methylation of host genes regulating CD4+ T cell HIV-1 reservoir size in women with HIV

protein (*SMC4*)²⁶, HIV-1 induced apoptosis (*PARP*)²⁷, and HIV-1 replication (*ISG15*)²⁸, which may contribute to HIV-1 latency.

Differentially methylated regions were significantly associated with total HR_{CD4}

To find significant differentially methylated regions (MR), we conducted an analysis using the “bump hunting” approach (see **Methods**). This analysis found MRs containing at least 2 CpG sites associated with total HR_{CD4}. We found 85 significant MRs across 22 chromosomes (**Table 2**). Among them, 28 MRs were negatively associated with HR_{CD4} while 57 MRs were positively associated.

Notably, 4 MRs on *ISG15*, spanning 21 CpG sites, were significant, including one region harboring 12 CpG sites. All 4 regions on *ISG15* were inversely associated with HR_{CD4} (**Figure 2A**). These CpG sites were either located in the TSS region or exons, suggesting that the aberrant methylation of *ISG15* may regulate *ISG15* expression, which is associated with HR. Another large significant MR was located on chromosome 3 at *SMC4*, spanning 2,872 base pairs and harboring 12 CpG sites (**Figure 2B**). Similarly, this MR was also inversely associated with HR_{CD4}. Other negatively HR_{CD4}-associated MR were in genes such as *PARP9*, *MX1*, *EPSTI1*, *USP18*, *IRF9*, *IFI44L*, and *IRF7* (**Figure 2C-D**).

Total HIV-1 CD4+ T cell reservoir-associated DNA methylation is related to HIV-1 integration

Because half of the HR_{CD4}-associated differentially methylated sites were located in intronic regions, we explored whether our 245 methylation sites were enriched in

Xu et al., Epigenome-wide characterization reveals aberrant DNA methylation of host genes regulating CD4+ T cell HIV-1 reservoir size in women with HIV

known HIV-1 genic integration sites by compiling HIV-1 integration sites reported across three recent studies²⁹⁻³¹. We found that a subset of the genes harboring significant CpG sites were targets of HIV-1 integration. Methylation at most of these integration sites was positively associated with HR_{CD4} size.

One gene, *IMPDH2*, was a common HIV-1 integration site between our study and the three previous studies²⁹⁻³¹. A CpG site on chromosome 3 at position 49066890, which was in the promoter region of *IMPDH2*, was positively associated with HR_{CD4} size. We also found that 23 HR_{CD4}-associated genes overlapped with HIV-1 integration sites reported by Kok et al.²⁹. Also, a few other genes were shared between our study and the other two previous studies^{29,30}, including *NFIA*, *TMTC3*, *SPPL3*, *DLEU2*, *ELMSAN1*, and *ACSF3*. A total of 53 genes overlapped with HIV-1 integration sites reported by Maldarelli et al.³¹ (e.g., *ISG15*, *TTC19*, *KCNG1*, *EIF3J*, *COX7A2L*, *TTC19*, *LAIR1*, *PARP9*, *JAK1*, and *CCDC97*), and 17 with sites reported by Wagner et al.³⁰ (e.g., *TTBK2*, *CDK16*, *ATP5PO*, and *TFDP1*) (**Figure 3**).

The most frequently reported HIV-1 integration site, *BACH2*^{30,31}, was nominally associated with HR_{CD4} ($p=0.001$) in our study. Additionally, several other previously reported integration sites displayed consistent patterns of differential methylation, with higher methylation levels in the top 25th percentile of total HR_{CD4} size compared to the lowest 25th percentile. For example, a CpG site at *JAK1* had 15% higher methylation in the top 25th percentile for HR_{CD4} size, compared to the lowest 25th percentile (*chr1* 65503557, $p=5.74e-07$) (**Figure 4A**). Similarly, two CpG sites at *NFIA* showed 7% and 10% higher methylation levels in the top 25th percentile (*chr1* 61542330, $p=0.00004$; *chr1* 61545122, $p=5.49e-10$) (**Figure 4B**). Other examples include 3 CpG sites for

Xu et al., Epigenome-wide characterization reveals aberrant DNA methylation of host genes regulating CD4⁺ T cell HIV-1 reservoir size in women with HIV

ELMSAN1 (*chr14* 74244367, $p=6.62e-05$; *chr14* 74250897, $p=1.43e-07$; *chr14* 74250926, $p=0.00004$) (**Figure 4C**), 2 CpG sites for *DLEU2* (*chr13* 50653658, $p=0.00001$; *chr13* 506554306, $p=1.88e-06$) (**Figure 4D**), and 1 CpG site for *SMAD2* (*chr18* 45458622: $p=1.28e-05$) (**Figure 4E**).

Total HR_{CD4}-associated CpG sites are enriched in pathways relevant to HIV-1 latency

We performed gene enrichment analysis using the human MSigDB database (<https://www.gsea-msigdb.org/gsea/msigdb/>), which includes 33,591 genes from nine major collections. In the hallmark gene set, we found significant enrichment for HR_{CD4}-associated genes in the interferon α response (gene ratio: 16/79; $p_{\text{adj}}=1.50e-10$) and interferon γ response (gene ratio: 16/79; $p_{\text{adj}}=3.45e-06$) (**Figure 5A-B**) (**Table 3**). In the interferon α gene set, 10 genes were associated with HR_{CD4}, with 8 of these overlapping between the “interferon α response” and “interferon γ response” gene sets. These shared genes were *USP18*, *MX1*, *ISG15*, *IRF7*, *IFIT3*, *IFI44L*, *IFI27*, and *EPST1*, all of which reached epigenome-wide significant association with total HR_{CD4}. Four of these genes are involved in the regulation of inflammatory function (i.e., *IRF7*, *IFIT3*, *IFI44L*, *IFI27*).

Gene set enrichment analysis employing Gene Ontology (GO) molecular function terms resulted in 34 HR_{CD4}-associated gene sets that were nominally significant but did not remain significant after multiple-testing correction. Nevertheless, these gene sets were highly relevant to HIV-1 pathogenesis (**Figure 6A, B**) (**Table 3**). The top-ranked gene sets included “tau protein binding” ($p=0.001$), “vinculin binding” ($p=0.001$),

Xu et al., Epigenome-wide characterization reveals aberrant DNA methylation of host genes regulating CD4+ T cell HIV-1 reservoir size in women with HIV

“transcription factor binding”, “DNA binding transcription repressor activity” ($p=0.006$), “DNA binding transcription factor binding” ($p=0.01$), and “interferon binding” ($p=0.02$). Nine genes were enriched in more than one gene set (i.e., *SMAD2*, *FOXK2*, *FOXP1*, *IRX3*, *PURB*, *SK1*, *ISL1*, *NFIA*, *TFDP1*).

Interestingly, *SMAD2* connects the “tau-protein binding” pathway to the “DNA binding transcription factor” and “DNA binding transcription activator activity” pathways (**Figure 6C**). A methylation site at *SMAD2* was positively associated with HR_{CD4} ($t=4.43$, $p=1.28e-05$). *SMAD2* is responsible for transmitting the signal of transforming growth factor (TGF)-beta³², which plays a crucial role in regulating various cellular processes like cell proliferation, differentiation, and apoptosis.

Discussion

Here, we report the first epigenome-wide association study on total HR_{CD4}, highlighting significant new insights into DNA methylation mechanisms involved in the HIV-1 reservoir. Using bisulfite capture sequencing on samples collected from 427 women with virally suppressed HIV-1 infection, we comprehensively characterized the epigenomic landscape of HR_{CD4} over 2 million CpG sites. We found 245 DNA methylation sites and 87 regions associated with total HR_{CD4} size. Notably, genes and pathways related to significant CpG sites were involved in the host immune responses, particularly type I interferon genes and JAK1, which regulate HIV-1 replication and integration. We also observed HR_{CD4}-associated aberrant methylation on the genes involved in chromosome stability, which may contribute to HIV-1 integration. Our results, derived from the largest sample size reported to date, reveal specific methylation

Xu et al., Epigenome-wide characterization reveals aberrant DNA methylation of host genes regulating CD4+ T cell HIV-1 reservoir size in women with HIV

mechanisms for HIV-1 persistence that may inform the potential target eradication strategies for HIV care.

Consistent with previous findings¹⁵, we found that the majority of the integrated DNA proviruses were defective. Recent studies have reported that both intact and defective integrated HIV-1 DNA proviruses remain transcriptionally active under suppressive ART, contributing to prolonged immune activation³³⁻³⁵ and greater comorbidity among PWH. Epigenetic regulation of host factors plays a critical role in HIV-1 transcriptome activation and silencing. In line with our current understanding of the mechanisms of HIV-1 pathogenesis, we found epigenetic alterations in genes involved in immune defense, interactions with viral proteins, and HIV-1 integration, suggesting that aberrant DNA methylation of these genes is linked to HIV-1 persistence.

Notably, we found aberrant DNA methylation of multiple type I interferon genes associated with total HR_{CD4} size. For example, four significant CpG sites for the interferon-induced protein *ISG15* were inversely associated with HR_{CD4}. A 445 base pairs region in *ISG15* containing multiple CpG sites was significantly less methylated in the highest total HR_{CD4} size quartile compared to the lowest quartile. Similarly, three regions in *interferon regulatory transcription factor 7 (IRF7)* encompassed 20 CpG sites that were less methylated in the highest compared to the lowest HR_{CD4} quartile. Additionally, a region spanning 4 CpG sites on *IRF9* was significantly less methylated in the highest compared to the lowest HR_{CD4} quartile. Other interferon-induced genes such as *IFI44L*, *MX1*, *PARP9*, and *USP18* consistently showed an inverse relationship between CpG methylation and HR_{CD4} size. Interferon genes are known to function as antiviral signaling molecules, regulate immune function, and induce apoptosis. As a

Xu et al., Epigenome-wide characterization reveals aberrant DNA methylation of host genes regulating CD4+ T cell HIV-1 reservoir size in women with HIV

result, it is presumed that increased expression of interferon genes due to hypomethylation may enhance restriction factor activity, resulting in limited HIV-1 replication and maintaining HR_{CD4} in a latent state. Furthermore, interferons also promote increased CD4+ T cell localization and HIV coreceptor expression, potentially enlarging the reservoir through seeding other cells. Our results support this complex and critical role of interferons in HIV-1 replication and latency³⁶.

Another notable finding is that a hypermethylated CpG site at *JAK1* was associated with larger HR_{CD4} size. *JAK1* is a critical mediator of immune activation and is involved HIV-1 latency. Previous studies have shown that *JAK1* interacts with the HIV-1 transcriptome and is essential for its life cycle. HIV-1 envelop surface glycoprotein gp120^{37,38} has been found to induce the expression of *JAK1* through the Tat-mediated activity of PIK3R1³⁹. Additionally, other studies have found that reduced expression levels of *JAK1* are associated with the inhibition of HIV LTR-beta-gal activity⁴⁰. A clinical trial showed that Ruxolitinib, a JAK1/2 inhibitor, significantly decreased immune activation markers in PWH and reduced HIV-1 reservoirs⁴¹. Thus, the aberrant methylation of *JAK1* in relation to HR_{CD4} size in this study further supports the significant role of *JAK1* in HIV-1 latency.

HIV-1 genome tends to integrate into highly expressed genes and genes involved in clonal expansion and cell cycle control²⁹. In this study, several genes involved in HIV-1 integration and latency showed significant association with HR_{CD4}. For example, the methylation of *NFIA*, a transcription factor highly expressed in HIV-1 latently infected CD4+ T cells⁴², was positively associated with HR_{CD4} size, with the methylation being more pronounced in the highest HR_{CD4} size quartile.

Xu et al., Epigenome-wide characterization reveals aberrant DNA methylation of host genes regulating CD4+ T cell HIV-1 reservoir size in women with HIV

Aberrant DNA methylation associated with HR_{CD4} was also discovered in genes responsible for chromosome stability, such as *SMC4*. *SMC4* is part of a family of genes that encode the structural maintenance of chromosome (SMC) proteins, which are parts of the cohesion and condensin complexes essential for chromatid pairing and chromosome segregation during mitosis⁴³. Dysfunction of SMC proteins may increase vulnerability to HIV-1 integration into the host genome. They also play critical roles in regulating transcription and ATPase domains and are crucial for maintaining chromosome structure and the repair and stability of DNA. *SMC4* is also a positive regulator of innate inflammatory immune responses and cell proliferation⁴⁴, which are key host factors for HIV-1 persistence. It is also believed that in T cells, *SMC4* interacts with HIV-1 Tat protein, which controls HIV-1 DNA replication and nuclear architecture²⁶. The observed aberrant DNA methylation associated with HR_{CD4} size in *SMC4*, further supports its involvement in the establishment and maintenance of HR.

We acknowledge some limitations in our study. First, we were unable to find epigenome-wide significant methylation sites specific for intact or defective reservoirs, necessitating more sensitive detection methods. Additionally, sample input requirements and technical challenges of studying HR_{CD4} in a large sample of PWH precluded the measurement of the functional HR_{CD4} (e.g., QVOA) in conjunction with IPDA. A precise measurement of HIV-1 latency requires quantifying cellular HIV-1 RNA transcription. As a result, confirming the identified methylation sites for HR_{CD4} using single-cell methylation profiles is needed. In addition, we are unable to rule out the potential confounding effect of cigarette smoking on DNA methylation marks, as most of the participants were smokers. Lastly, our cross-sectional study design limits the ability

Xu et al., Epigenome-wide characterization reveals aberrant DNA methylation of host genes regulating CD4+ T cell HIV-1 reservoir size in women with HIV

to assess the causal relationship between altered DNA methylation and HR size. Both longitudinal and *ex vivo* studies are needed to figure out the causality between DNA methylation and reservoir size. Future research is also needed to replicate these findings in men living with HIV and explore the dynamic interaction between the genes found in this study and the HIV-1 genome in reservoir establishment and maintenance.

In summary, this comprehensive characterization of the epigenome concerning HR_{CD4} sheds light on the epigenetic mechanisms underlying HIV-1 persistence. Most importantly, these findings offer new gene targets such as *ISG15*, *PARP*, *NFIA*, and *SMC4* for an in-depth exploration of host responses to the HIV-1 reservoir. These targets could potentially serve as molecular points of interest for the development of novel treatments.

Methods

Ethics approval and consent to participate

The study was approved by the Committee of Human Research Subject Protection at Yale University and the Institutional Research Board Committee of the Connecticut Veteran Healthcare System. Informed consent was provided by all MWCCS participants via protocols approved by institutional review committees at each affiliated institution.

Participants and HR detection

Clinical data and specimens used in this study were collected by MWCCS²³. MWCCS is the largest observational cohort of HIV infection in the United States. The

Xu et al., Epigenome-wide characterization reveals aberrant DNA methylation of host genes regulating CD4⁺ T cell HIV-1 reservoir size in women with HIV

inclusion criteria were Women with HIV who had undetectable HIV-1 viral load (below the limit of quantitation for the assay) on ART treatment for at least 6 months⁴⁵. A total of 427 WWH who met the criteria were selected for the study. The average duration of ART treatment was ~2.1 years among the participants. CD4⁺ T cells were isolated by negative bead selection using the DynaBeads Untouched Human CD4 kit. Genomic DNA was then extracted from CD4⁺ T cells using magnetic bead-based nucleic acid isolation. ddPCR was employed to measure the intact proviral HIV DNA using the modified IPDA assay as described previously by Bruner *et al.*⁴⁶, providing estimates of intact provirus per one million CD4⁺ T cells. Additional details are discussed elsewhere²⁴.

Genome-wide DNA methylation capture sequence (MC-seq)

Genome-wide DNA methylation was conducted at the Yale Center for Genome Analysis according to the following steps:

Methyl-Seq target enrichment library prep. Genomic DNA quality was determined by estimating the A260/A280 and A260/A230 ratios using spectrophotometry and the concentration from fluorometry. DNA integrity and fragment size were confirmed using a microfluidic chip run on an Agilent Bioanalyzer. Indexed paired-end whole genome sequencing libraries were prepared using the SureSelect XT Methyl-Seq kit (Agilent, part#G9651B). The library preparation was conducted following the manufacturer's protocol. Briefly, samples yielding >350 ng were enriched for targeted methylation sites using the custom SureSelect Methyl-Seq Capture Library. The SureSelect enriched, bisulfite-converted libraries were PCR amplified using custom-made indexed primers

Xu et al., Epigenome-wide characterization reveals aberrant DNA methylation of host genes regulating CD4+ T cell HIV-1 reservoir size in women with HIV

(IDT, Coralville, Iowa). Dual-indexed libraries were quantified by quantitative polymerase chain reaction (qPCR) using the Library Quantification Kit (KAPA Biosystems, Part#KK4854). Insert size distribution was assessed using the Caliper LabChip GX system. Samples with a yield of ≥ 2 ng/ul proceeded to sequencing.

Flow cell preparation and sequencing. Sample concentrations were normalized to 10 nM and loaded onto an Illumina NovaSeq flow cell at a concentration that yielded 40 million passing filter clusters per sample. Samples were sequenced using 100bp paired-end sequencing on an Illumina HiSeq NovaSeq, according to Illumina's standard protocol. The 10bp dual index was read during additional sequencing reads that automatically followed the completion of the first read. Data generated during sequencing runs were transferred to the Yale Center for Genome Analysis high-performance computing cluster. A positive control (prepared bacteriophage Phi X library) provided by Illumina was spiked into every lane at a concentration of 0.3% to monitor sequencing quality in real-time.

Preprocessing and quality control. Signal intensities were converted to individual base calls during a run using the system's Real Time Analysis (RTA) software. Sample demultiplexing and alignment to the human genome were performed using Illumina's CASAVA 1.8.2 software suite. The sample error rate was required to be less than 1% and the distribution of reads per sample in a lane was required to be within reasonable tolerance. Quality control (QC) of sequencing reads was conducted following the standard procedure as previously described⁴⁷. The quality of sequence data was

Xu et al., Epigenome-wide characterization reveals aberrant DNA methylation of host genes regulating CD4+ T cell HIV-1 reservoir size in women with HIV

examined using *FastQC* (ver. 0.11.8). Adapter sequences and fragments at 5' and 3' (phred score <30) with inadequate quality were removed by *Trim_galore* (ver. 0.6.3_dev). We used Bismark pipelines (ver. v0.22.1_dev) to align the reads to the bisulfite human genome (hg19) with default parameters⁴⁸. An average of 87% of the sequence per sample was mapped to the genome. Quality-trimmed paired-end reads were transformed into a bisulfite-converted forward strand version (C->T conversion) or a bisulfite-treated reverse strand (G->A conversion of the forward strand). Only CpG sites with coverage >10 depth were used for final analysis to ensure the quality of CpG sites. A total of 427 samples and 2,078,054 CpG sites remained for epigenome-wide association analysis. Genes 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 shores were defined as up to 2 kb from CpG islands and CpG shelves were defined as up to 2 kb from a CpG shore.

Identification of CpG sites associated with HR

To ensure no other cell types were represented in the genomic DNA isolated from CD4⁺ T cells obtained from magnetic bead separation, we first estimated the PBMC cell type proportion using the Houseman method⁴⁹. The distribution of the HIV-1 proviral DNA count (i.e., HR size) was normalized using a log₁₀ transformation. Epigenome-wide association analysis was performed using generalized linear models, in which methylation beta value at each CpG site was modeled as the dependent

Xu et al., Epigenome-wide characterization reveals aberrant DNA methylation of host genes regulating CD4+ T cell HIV-1 reservoir size in women with HIV

variable, and total HR size was modeled as the independent variable, while also adjusting for age, self-reported race, cell type proportion, and smoking status. We also adjusted the model for the lower detection limit of each HIV-1 RNA viral load assay. The significance level was set at a false discovery rate (FDR) of 0.05. CpG sites that were evaluated for significant association with total HR were then re-evaluated for association with intact HR_{CD4}, 5'-defective HR_{CD4}, and 3'-defective HR_{CD4} separately.

Identification of DNA methylation regions associated with total HR

Leveraging the bump-hunting framework⁵⁰, we analyzed DNA methylation regions (as opposed to individual CpG sites) that were associated with total HR_{CD4}. Bump hunting provides the advantage of effectively modeling measurement errors, removing batch effects, and detecting regions of interest for continuous measures such as HR_{CD4}. We modified the bump hunting pipeline of Jaffe AE et al. to increase the sensitivity for detecting methylation regions. In the regression model, instead of using methylation β value at each CpG site, we used the *t*-statistic value that was derived from EWAS regressed on the total HR_{CD4}. The *p*-value of each region was then estimated following the method proposed by Lui et al.⁵¹. A significant region was defined as having at least two consecutive CpG sites and an adjusted $p < 0.05$.

Gene set enrichment analysis

Genes adjacent to CpG sites were selected for gene set enrichment analysis. A cutoff of $p < 0.05$ was used to ensure enough genes were selected for enrichment analysis. We used the Hallmark database from the Molecular Signatures Database

Xu et al., Epigenome-wide characterization reveals aberrant DNA methylation of host genes regulating CD4+ T cell HIV-1 reservoir size in women with HIV

(<https://www.gsea-msigdb.org/gsea/msigdb/>) and the Gene Ontology (GO) database for the enrichment analyses.

Administrative information

Author contributions

KX oversaw the study, including DNA methylation sequencing data collection, data processing, statistical analyses, and manuscript preparation. XZ conducted data processing and analysis and contributed to manuscript preparation. BEA and KA handled CD4⁺ T cell isolation, genomic DNA isolation, and HIV-1 reservoir detection. BCQ, GPP, and DBH participated in analytic strategy and manuscript preparation. DKP, HHB, CDL, ETG, MHC, NMA, MHK, and PCT participated in the study's conduct, including protocol development, participant recruitment, follow-up, and biospecimen collection. AV contributed to manuscript preparation. VCM contributed to the interpretation of the findings and preparation of the manuscript. EOJ and BEA contributed to the study design, interpretation of findings, and manuscript preparation. All authors read and approved the final manuscript.

Acknowledgments

The authors appreciate the support of the Yale Center of Genomic Analysis. Data for this study were collected by the Women's Interagency HIV Study (WIHS), now the MACS/WIHS Combined Cohort Study (MWCCS). The authors gratefully acknowledge

Xu et al., Epigenome-wide characterization reveals aberrant DNA methylation of host genes regulating CD4+ T cell HIV-1 reservoir size in women with HIV

the contributions of the study participants and the dedication of the staff at the MWCCS sites.

Conflicts of interest statement

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.

Data availability statement

Summary statistics of the epigenome-wide analysis are presented in the Supplementary Table. Access to individual-level DNA methylation data from the MACS/WIHS Combined Cohort Study Data (MWCCS) may be obtained upon review and approval of a MWCCS concept sheet. □ Links and instructions for online concept sheet submission are on the study website (<https://statepi.jhsph.edu/mwccs/>).

Funding statement

The contents of this publication are solely the responsibility of the authors and do not represent the official views of the National Institutes of Health (NIH). The project was supported by the National Institute on Drug Abuse [R61DA047011 (Johnson and Aouizerat), R33DA047011 (Johnson and Aouizerat), R01DA052846 (Xu), R01DA047063 (Xu and Aouizerat), R01DA047820 (Xu and Aouizerat)], the Emory Center for AIDS Research [P30-AI050409 (Marconi)] and the MWCCS (Principal

Xu et al., Epigenome-wide characterization reveals aberrant DNA methylation of host genes regulating CD4+ T cell HIV-1 reservoir size in women with HIV

Investigators): Atlanta CRS (Ighovwerha Ofotokun, Anandi Sheth, and Gina Wingood), U01-HL146241; Bronx CRS (Kathryn Anastos, David Hanna, and Anjali Sharma), U01-HL146204; Brooklyn CRS (Deborah Gustafson and Tracey Wilson), U01-HL146202; Data Analysis and Coordination Center (Gypsyamber D'Souza, Stephen Gange and Elizabeth Topper), U01-HL146193; Chicago-Cook County CRS (Mardge Cohen, Audrey French, and Ryan Ross), U01-HL146245; Northern California CRS (Bradley Aouizerat, Jennifer Price, and Phyllis Tien), U01-HL146242; Metropolitan Washington CRS (Seble Kassaye and Daniel Merenstein), U01-HL146205; Miami CRS (Maria Alcaide, Margaret Fischl, and Deborah Jones), U01-HL146203; UAB-MS CRS (Mirjam-Colette Kempf, James B. Brock, Emily Levitan, and Deborah Konkle-Parker), U01-HL146192; UNC CRS (M. Bradley Drummond and Michelle Floris-Moore), U01-HL146194. The MWCCS is funded primarily by the National Heart, Lung, and Blood Institute (NHLBI), with additional co-funding from the Eunice Kennedy Shriver National Institute Of Child Health & Human Development (NICHD), National Institute On Aging (NIA), National Institute Of Dental & Craniofacial Research (NIDCR), National Institute Of Allergy And Infectious Diseases (NIAID), National Institute Of Neurological Disorders And Stroke (NINDS), National Institute Of Mental Health (NIMH), National Institute On Drug Abuse (NIDA), National Institute Of Nursing Research (NINR), National Cancer Institute (NCI), National Institute on Alcohol Abuse and Alcoholism (NIAAA), National Institute on Deafness and Other Communication Disorders (NIDCD), National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institute on Minority Health and Health Disparities (NIMHD), and in coordination and alignment with the research priorities of the National Institutes of Health, Office of AIDS Research (OAR). MWCCS data

Xu *et al.*, Epigenome-wide characterization reveals aberrant DNA methylation of host genes regulating CD4+ T cell HIV-1 reservoir size in women with HIV

collection is also supported by UL1-TR000004 (UCSF CTSA), P30-AI-050409 (Atlanta CFAR), P30-AI-073961 (Miami CFAR), P30-AI-050410 (UNC CFAR), P30-AI-027767 (UAB CFAR), P30-MH-116867 (Miami CHARM), UL1-TR001409 (DC CTSA), KL2-TR001432 (DC CTSA), and TL1-TR001431 (DC CTSA).

References

- 1 Bruner, K. M. & Cohn, L. B. HIV-1 reservoir dynamics in CD4+ T cells. *Curr Opin HIV AIDS* **14**, 108-114 (2019). <https://doi.org:10.1097/COH.0000000000000521>
- 2 Board, N. L., Moskovljevic, M., Wu, F., Siliciano, R. F. & Siliciano, J. D. Engaging innate immunity in HIV-1 cure strategies. *Nat Rev Immunol* **22**, 499-512 (2022). <https://doi.org:10.1038/s41577-021-00649-1>
- 3 Li, J. Z. *et al.* The size of the expressed HIV reservoir predicts timing of viral rebound after treatment interruption. *AIDS* **30**, 343-353 (2016). <https://doi.org:10.1097/QAD.0000000000000953>
- 4 Siliciano, J. D. *et al.* Long-term follow-up studies confirm the stability of the latent reservoir for HIV-1 in resting CD4+ T cells. *Nat Med* **9**, 727-728 (2003). <https://doi.org:10.1038/nm880>
- 5 Cho, A. *et al.* Longitudinal clonal dynamics of HIV-1 latent reservoirs measured by combination quadruplex polymerase chain reaction and sequencing. *Proc Natl Acad Sci U S A* **119** (2022). <https://doi.org:10.1073/pnas.2117630119>
- 6 Bullen, C. K., Laird, G. M., Durand, C. M., Siliciano, J. D. & Siliciano, R. F. New ex vivo approaches distinguish effective and ineffective single agents for reversing HIV-1 latency in vivo. *Nat Med* **20**, 425-429 (2014). <https://doi.org:10.1038/nm.3489>
- 7 Laird, G. M. *et al.* Ex vivo analysis identifies effective HIV-1 latency-reversing drug combinations. *J Clin Invest* **125**, 1901-1912 (2015). <https://doi.org:10.1172/JCI80142>
- 8 Einkauf, K. B. *et al.* Parallel analysis of transcription, integration, and sequence of single HIV-1 proviruses. *Cell* **185**, 266-282 e215 (2022). <https://doi.org:10.1016/j.cell.2021.12.011>
- 9 Jefferys, S. R. *et al.* Epigenomic characterization of latent HIV infection identifies latency regulating transcription factors. *PLoS Pathog* **17**, e1009346 (2021). <https://doi.org:10.1371/journal.ppat.1009346>
- 10 Kauder, S. E., Bosque, A., Lindqvist, A., Planelles, V. & Verdin, E. Epigenetic regulation of HIV-1 latency by cytosine methylation. *PLoS Pathog* **5**, e1000495 (2009). <https://doi.org:10.1371/journal.ppat.1000495>
- 11 Boltz, V. F. *et al.* CpG Methylation Profiles of HIV-1 Pro-Viral DNA in Individuals on ART. *Viruses* **13** (2021). <https://doi.org:10.3390/v13050799>
- 12 Blazkova, J. *et al.* Paucity of HIV DNA methylation in latently infected, resting CD4+ T cells from infected individuals receiving antiretroviral therapy. *J Virol* **86**, 5390-5392 (2012). <https://doi.org:10.1128/JVI.00040-12>

Xu *et al.*, Epigenome-wide characterization reveals aberrant DNA methylation of host genes regulating CD4+ T cell HIV-1 reservoir size in women with HIV

- 13 Mantovani, N. *et al.* Latency-associated DNA methylation patterns among HIV-1 infected individuals with distinct disease progression courses or antiretroviral virologic response. *Sci Rep* **11**, 22993 (2021). <https://doi.org/10.1038/s41598-021-02463-0>
- 14 Bruner, K. M. *et al.* Defective proviruses rapidly accumulate during acute HIV-1 infection. *Nat Med* **22**, 1043-1049 (2016). <https://doi.org/10.1038/nm.4156>
- 15 Ho, Y. C. *et al.* Replication-competent noninduced proviruses in the latent reservoir increase barrier to HIV-1 cure. *Cell* **155**, 540-551 (2013). <https://doi.org/10.1016/j.cell.2013.09.020>
- 16 Pollack, R. A. *et al.* Defective HIV-1 Proviruses Are Expressed and Can Be Recognized by Cytotoxic T Lymphocytes, which Shape the Proviral Landscape. *Cell Host Microbe* **21**, 494-506 e494 (2017). <https://doi.org/10.1016/j.chom.2017.03.008>
- 17 Duette, G. *et al.* The HIV-1 proviral landscape reveals that Nef contributes to HIV-1 persistence in effector memory CD4+ T cells. *J Clin Invest* **132** (2022). <https://doi.org/10.1172/JCI154422>
- 18 Sieg, S. F. Persistent Inflammation in Treated HIV Disease. *J Infect Dis* **214 Suppl 2**, S43 (2016). <https://doi.org/10.1093/infdis/jiw185>
- 19 Dirajlal-Fargo, S. & Funderburg, N. HIV and cardiovascular disease: the role of inflammation. *Curr Opin HIV AIDS* **17**, 286-292 (2022). <https://doi.org/10.1097/COH.0000000000000755>
- 20 Snell, L. M., McGaha, T. L. & Brooks, D. G. Type I Interferon in Chronic Virus Infection and Cancer. *Trends Immunol* **38**, 542-557 (2017). <https://doi.org/10.1016/j.it.2017.05.005>
- 21 Barkan, S. E. *et al.* The Women's Interagency HIV Study. WIHS Collaborative Study Group. *Epidemiology* **9**, 117-125 (1998).
- 22 Adimora, A. A. *et al.* Cohort Profile: The Women's Interagency HIV Study (WIHS). *Int J Epidemiol* **47**, 393-394i (2018). <https://doi.org/10.1093/ije/dyy021>
- 23 D'Souza, G. *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* **190**, 1457-1475 (2021). <https://doi.org/10.1093/aje/kwab050>
- 24 Aouizerat, B. E. *et al.* Frequent Cocaine Use is Associated with Larger HIV Latent Reservoir Size. *JAIDS Journal of Acquired Immune Deficiency Syndromes* (2024). <https://doi.org/10.1097/qai.00000000000003472>
- 25 Shu, C., Zhang, X., Aouizerat, B. E. & Xu, K. Comparison of methylation capture sequencing and Infinium MethylationEPIC array in peripheral blood mononuclear cells. *Epigenetics Chromatin* **13**, 51 (2020). <https://doi.org/10.1186/s13072-020-00372-6>
- 26 Gautier, V. W. *et al.* In vitro nuclear interactome of the HIV-1 Tat protein. *Retrovirology* **6**, 47 (2009). <https://doi.org/10.1186/1742-4690-6-47>
- 27 Rumlova, M. *et al.* HIV-1 protease-induced apoptosis. *Retrovirology* **11**, 37 (2014). <https://doi.org/10.1186/1742-4690-11-37>
- 28 Jurczynszak, D. *et al.* ISG15 deficiency restricts HIV-1 infection. *PLoS Pathog* **18**, e1010405 (2022). <https://doi.org/10.1371/journal.ppat.1010405>
- 29 Kok, Y. L. *et al.* HIV-1 integration sites in CD4+ T cells during primary, chronic, and late presentation of HIV-1 infection. *JCI Insight* **6** (2021). <https://doi.org/10.1172/jci.insight.143940>

Xu *et al.*, Epigenome-wide characterization reveals aberrant DNA methylation of host genes regulating CD4+ T cell HIV-1 reservoir size in women with HIV

- 30 Wagner, T. A. *et al.* HIV latency. Proliferation of cells with HIV integrated into cancer genes contributes to persistent infection. *Science* **345**, 570-573 (2014). <https://doi.org/10.1126/science.1256304>
- 31 Maldarelli, F. *et al.* HIV latency. Specific HIV integration sites are linked to clonal expansion and persistence of infected cells. *Science* **345**, 179-183 (2014). <https://doi.org/10.1126/science.1254194>
- 32 Bergstresser, S. & Kulpa, D. A. TGF-beta Signaling Supports HIV Latency in a Memory CD4+ T Cell Based In Vitro Model. *Methods Mol Biol* **2407**, 69-79 (2022). https://doi.org/10.1007/978-1-0716-1871-4_6
- 33 Ruggiero, A. *et al.* During Stably Suppressive Antiretroviral Therapy Integrated HIV-1 DNA Load in Peripheral Blood is Associated with the Frequency of CD8 Cells Expressing HLA-DR/DP/DQ. *EBioMedicine* **2**, 1153-1159 (2015). <https://doi.org/10.1016/j.ebiom.2015.07.025>
- 34 Vallejo, A. HIV-1 Reservoir Association with Immune Activation. *EBioMedicine* **2**, 1020-1021 (2015). <https://doi.org/10.1016/j.ebiom.2015.08.025>
- 35 Huang, Y. *et al.* Immune activation and exhaustion marker expression on T-cell subsets in ART-treated adolescents and young adults with perinatal HIV-1 infection as correlates of viral persistence. *Front Immunol* **14**, 1007626 (2023). <https://doi.org/10.3389/fimmu.2023.1007626>
- 36 Utay, N. S. & Douek, D. C. Interferons and HIV Infection: The Good, the Bad, and the Ugly. *Pathog Immun* **1**, 107-116 (2016). <https://doi.org/10.20411/pai.v1i1.125>
- 37 Cicala, C. *et al.* HIV envelope induces a cascade of cell signals in non-proliferating target cells that favor virus replication. *Proc Natl Acad Sci U S A* **99**, 9380-9385 (2002). <https://doi.org/10.1073/pnas.142287999>
- 38 Galey, D. *et al.* Differential transcriptional regulation by human immunodeficiency virus type 1 and gp120 in human astrocytes. *J Neurovirol* **9**, 358-371 (2003). <https://doi.org/10.1080/13550280390201119>
- 39 Brass, A. L. *et al.* Identification of host proteins required for HIV infection through a functional genomic screen. *Science* **319**, 921-926 (2008). <https://doi.org/10.1126/science.1152725>
- 40 Zhou, H. *et al.* Genome-scale RNAi screen for host factors required for HIV replication. *Cell Host Microbe* **4**, 495-504 (2008). <https://doi.org/10.1016/j.chom.2008.10.004>
- 41 de Armas, L. R. *et al.* The Effect of JAK1/2 Inhibitors on HIV Reservoir Using Primary Lymphoid Cell Model of HIV Latency. *Front Immunol* **12**, 720697 (2021). <https://doi.org/10.3389/fimmu.2021.720697>
- 42 Kim, K. C. *et al.* Identification of novel genes associated with HIV-1 latency by analysis of histone modifications. *Hum Genomics* **11**, 9 (2017). <https://doi.org/10.1186/s40246-017-0105-7>
- 43 Wang, J. *et al.* Mutation of Arabidopsis SMC4 identifies condensin as a corepressor of pericentromeric transposons and conditionally expressed genes. *Genes Dev* **31**, 1601-1614 (2017). <https://doi.org/10.1101/gad.301499.117>
- 44 Wei-Shan, H., Amit, V. C. & Clarke, D. J. Cell cycle regulation of condensin Smc4. *Oncotarget* **10**, 263-276 (2019). <https://doi.org/10.18632/oncotarget.26467>

Xu *et al.*, *Epigenome-wide characterization reveals aberrant DNA methylation of host genes regulating CD4+ T cell HIV-1 reservoir size in women with HIV*

- 45 Blankson, J. N. *et al.* Biphasic decay of latently infected CD4+ T cells in acute human immunodeficiency virus type 1 infection. *J Infect Dis* **182**, 1636-1642 (2000). [https://doi.org:10.1086/317615](https://doi.org/10.1086/317615)
- 46 Bruner, K. M. *et al.* A quantitative approach for measuring the reservoir of latent HIV-1 proviruses. *Nature* **566**, 120-125 (2019). [https://doi.org:10.1038/s41586-019-0898-8](https://doi.org/10.1038/s41586-019-0898-8)
- 47 Wreczycka, K. *et al.* Strategies for analyzing bisulfite sequencing data. *J Biotechnol* **261**, 105-115 (2017). [https://doi.org:10.1016/j.jbiotec.2017.08.007](https://doi.org/10.1016/j.jbiotec.2017.08.007)
- 48 Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571-1572 (2011). [https://doi.org:10.1093/bioinformatics/btr167](https://doi.org/10.1093/bioinformatics/btr167)
- 49 Houseman, E. A. *et al.* DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* **13**, 86 (2012). [https://doi.org:10.1186/1471-2105-13-86](https://doi.org/10.1186/1471-2105-13-86)
- 50 Jaffe, A. E. *et al.* Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies. *Int J Epidemiol* **41**, 200-209 (2012). [https://doi.org:10.1093/ije/dyr238](https://doi.org/10.1093/ije/dyr238)
- 51 Liu, Y. *et al.* ACAT: A Fast and Powerful p Value Combination Method for Rare-Variant Analysis in Sequencing Studies. *Am J Hum Genet* **104**, 410-421 (2019). [https://doi.org:10.1016/j.ajhg.2019.01.002](https://doi.org/10.1016/j.ajhg.2019.01.002)

Xu et al., Epigenome-wide characterization reveals aberrant DNA methylation of host genes regulating CD4+ T cell HIV-1 reservoir size in women with HIV

Tables

Table 1 | Demographic and clinical characteristics of participants.

Variable		Participants (N=427)
Age (Mean \pm SD)		47.29 \pm 8.22
Sex (F/M)		427/0
Ancestry (%)	African American	76.3
	European	12.4
	Other	11
Smoking Status (%)	Current	79.9
	Former	8.9
	Never	11
Alcohol Drinks per Week (Mean \pm SD)		9.3 \pm 26.5
On ART + Undetected VL (%)		100
HLR (Mean \pm SD)	Total	1,409 \pm 1,639
	Intact	279 \pm 463.4
	5'-depletion	600 \pm 747.7
	3'-depletion	530 \pm 885.7

ART: Antiretroviral Therapy; HR: HIV-reservoir; VL: Viral load

Xu et al., Epigenome-wide characterization reveals aberrant DNA methylation of host genes regulating CD4⁺ T cell HIV-1 reservoir size in women with HIV

Table 2 | Differentially methylated regions associated with total HIV-1 reservoir in CD4⁺ T cells.

Chr	Start	End	Gene	# CpGs	t	p	FDR
1	61545077	61545123	<i>NFIA</i>	2	5.4	1.1e-09	9.2e-08
3	160119633	160122505	<i>MIR16-2, MIR15B, SMC4</i>	12	-5.0	4.7e-09	2.0e-07
3	122281883	122281941	<i>DTX3L, PARP9</i>	2	-5.4	3.8e-08	1.1e-06
13	43565289	43565401	<i>EPSTI1</i>	3	-4.7	5.8e-08	1.2e-06
14	74250898	74250927	<i>ELMSAN1</i>	2	4.8	2.9e-07	4.1e-06
3	54888035	54888057	<i>CACNA2D3</i>	6	-4.4	2.9e-07	4.1e-06
2	43398273	43398277	-	2	5.1	5.4e-07	6.5e-06
7	44924186	44924194	<i>PURB</i>	2	-4.5	6.7e-07	7.1e-06
1	12676155	12676183	<i>DHRS3</i>	2	5.0	9.4e-07	7.7e-06
19	41815093	41815101	<i>CCDC97</i>	3	4.5	1.0e-06	7.7e-06
10	88852547	88852560	<i>GLUD1</i>	2	4.6	1.1e-06	7.7e-06
1	119522623	119522631	<i>TBX15</i>	2	4.5	1.1e-06	7.7e-06
1	949394	949433	<i>ISG15</i>	3	-4.6	1.2e-06	7.7e-06
11	118790174	118790223	-	3	4.6	1.4e-06	7.7e-06
1	207996299	207996461	-	3	4.8	1.4e-06	7.7e-06
17	7381527	7381577	<i>ZBTB4</i>	2	4.5	2.4e-06	1.3e-05
21	42798945	42799143	<i>MX1</i>	7	-4.4	2.7e-06	1.3e-05
15	44829519	44829525	<i>EIF3J, EIF3J-AS1</i>	2	-4.5	3.0e-06	1.4e-05
1	207996636	207996676	-	2	4.7	3.3e-06	1.4e-05
9	134139856	134139880	<i>FAM78A</i>	3	4.2	3.3e-06	1.4e-05
10	22612799	22613027	<i>COMMD3-BMI1, BMI1</i>	2	4.5	3.4e-06	1.4e-05
22	18635396	18635462	<i>USP18</i>	3	-4.4	3.8e-06	1.4e-05
16	4714768	4714796	<i>MGRN1</i>	3	4.5	4.1e-06	1.4e-05
2	37551058	37551083	-	5	4.6	4.1e-06	1.4e-05
11	75062095	75062115	<i>ARRB1</i>	2	4.4	6.3e-06	2.1e-05
21	42797710	42797717	<i>MX1</i>	2	-4.5	7.6e-06	2.4e-05
14	50159543	50159550	<i>KLHDC1</i>	2	4.4	8.7e-06	2.6e-05
11	121324233	121324511	<i>SORL1</i>	3	4.3	8.7e-06	2.6e-05
19	54876395	54876448	<i>LAIR1</i>	2	4.2	9.3e-06	2.6e-05
1	12538489	12538543	<i>VPS13D</i>	4	4.4	9.3e-06	2.6e-05
1	949698	949878	<i>ISG15</i>	12	-4.2	1.1e-05	2.8e-05
1	949563	949599	<i>ISG15</i>	4	-4.3	1.1e-05	2.8e-05
15	69707926	69707938	<i>KIF23</i>	2	-4.4	1.2e-05	3.1e-05
17	47437978	47438153	<i>ZNF652, LOC102724596</i>	3	4.2	1.4e-05	3.4e-05
15	55562701	55562788	<i>RAB27A</i>	4	-4.1	1.7e-05	4.0e-05
22	27068060	27068092	<i>MIAT, MIATNB</i>	4	4.2	1.8e-05	4.2e-05
4	38666920	38667432	<i>KLF3-AS1, KLF3</i>	2	4.3	1.8e-05	4.2e-05
12	120700363	120700438	<i>PXN</i>	3	4.0	2.4e-05	5.2e-05
9	6716463	6716482	-	4	-4.2	2.5e-05	5.2e-05
17	79265630	79265645	<i>SLC38A10</i>	2	4.1	2.5e-05	5.2e-05
4	103682822	103682824	<i>MANBA</i>	2	4.2	2.5e-05	5.2e-05
21	45509115	45509120	<i>TRAPPC10</i>	2	4.3	2.7e-05	5.3e-05

Xu et al., Epigenome-wide characterization reveals aberrant DNA methylation of host genes regulating CD4+ T cell HIV-1 reservoir size in women with HIV

11	614844	614857	<i>IRF7</i>	3	-4.1	2.8e-05	5.4e-05
1	949479	949508	<i>ISG15</i>	3	-4.2	2.8e-05	5.4e-05
10	90641813	90642005	<i>STAMBPL1</i>	3	-4.2	2.9e-05	5.5e-05
14	105751631	105751638	<i>BRF1</i>	2	4.1	3.1e-05	5.6e-05
22	39637150	39637172	<i>PDGFB</i>	4	4.1	3.2e-05	5.6e-05
1	226069169	226069191	<i>TMEM63A</i>	2	4.1	3.2e-05	5.6e-05
19	4380572	4380611	<i>SH3GL1</i>	2	4.1	3.3e-05	5.6e-05
3	43385089	43385113	<i>SNRK</i>	2	4.2	3.4e-05	5.6e-05
14	24631374	24631383	<i>IRF9</i>	2	-4.2	3.4e-05	5.6e-05
6	137630457	137630478	-	2	4.2	3.4e-05	5.6e-05
11	614567	614763	<i>IRF7</i>	2	-4.2	3.7e-05	5.9e-05
9	6716396	6716425	-	5	-4.1	3.9e-05	6.0e-05
17	42462262	42462268	<i>ITGA2B</i>	2	4.1	4.4e-05	6.7e-05
20	62530409	62530421	<i>DNAJC5</i>	2	4.1	4.7e-05	6.8e-05
2	43398215	43398245	-	2	4.1	4.7e-05	6.8e-05
2	173294414	173294494	<i>ITGA6</i>	2	4.0	4.8e-05	6.8e-05
21	45509391	45509401	<i>TRAPPC10</i>	3	4.1	4.8e-05	6.8e-05
22	38610341	38610350	<i>MAFF</i>	2	4.1	4.9e-05	6.8e-05
9	6716209	6716228	-	2	-4.1	4.9e-05	6.8e-05
11	61213606	61213621	<i>SDHAF2</i>	2	4.1	5.2e-05	7.0e-05
11	74975246	74975248	<i>ARRB1</i>	2	4.0	5.3e-05	7.1e-05
17	7381831	7381857	<i>ZBTB4</i>	2	4.1	5.6e-05	7.3e-05
2	71681539	71681542	<i>DYSF</i>	2	4.0	6.1e-05	7.9e-05
19	19281257	19281272	<i>MEF2BNB-MEF2B, MEF2B</i>	2	4.0	6.3e-05	8.0e-05
11	109962729	109962762	<i>ZC3H12C</i>	2	4.0	6.8e-05	8.5e-05
21	45509454	45509457	<i>TRAPPC10</i>	2	4.0	6.8e-05	8.5e-05
14	61109826	61109857	-	2	4.0	7.1e-05	8.7e-05
4	85416937	85416953	<i>NKX6-1</i>	2	3.9	7.9e-05	9.5e-05
16	53535169	53535242	<i>AKTIP</i>	2	3.9	9.0e-05	1.1e-04
1	154318280	154318316	<i>ATP8B2</i>	2	3.9	9.7e-05	1.1e-04
8	133098472	133098504	<i>HHLA1</i>	2	-3.9	1.1e-04	1.2e-04
17	37308830	37308847	<i>PLXDC1</i>	2	3.9	1.1e-04	1.3e-04
9	6716240	6716271	-	2	-3.9	1.2e-04	1.3e-04
6	168714672	168714675	<i>DACT2</i>	2	3.9	1.3e-04	1.4e-04
9	6716324	6716351	-	3	-3.9	1.3e-04	1.4e-04
9	79791171	79791179	<i>VPS13A</i>	2	3.9	1.4e-04	1.5e-04
3	122401131	122401302	<i>PARP14</i>	3	-3.8	1.4e-04	1.5e-04
5	142435222	142435443	<i>ARHGAP26</i>	2	-3.8	1.5e-04	1.5e-04
15	85177989	85178063	<i>SCAND2P</i>	2	3.8	1.5e-04	1.5e-04
1	79090602	79092806	<i>IFI44L</i>	2	-3.8	1.5e-04	1.5e-04
21	38065618	38065639	-	2	3.8	1.9e-04	1.9e-04
5	17219328	17219331	<i>BASP1</i>	2	3.8	1.9e-04	1.9e-04

Xu et al., Epigenome-wide characterization reveals aberrant DNA methylation of host genes regulating CD4+ T cell HIV-1 reservoir size in women with HIV

Table 3 | Significantly perturbed pathways associated with CD4+ T-cell total HIV-1 reservoir size.

Pathway	Gene Ratio	p	q
Hallmark Pathways			
Interferon-α Response	10/97	2.2e-05	1.0e-03
Interferon-γ Response	8/200	4.8e-02	7.7e-01
GO Molecular Function Pathways			
Tau Protein Binding	5/272	1.1e-03	3.0e-01
Vinculin Binding	3/272	1.2e-03	3.0e-01
DNA Binding Transcription Repressor Activity	13/272	6.1e-03	6.7e-01
DNA binding Transcription Activator Activity	16/272	1.1e-02	6.7e-01
DNA Binding Transcription Factor Binding	16/272	1.2e-02	6.7e-01
Aromatic Amino Acid Transmembrane Transporter Activity	2/272	1.7e-02	6.7e-01
Oligopeptide Transmembrane Transporter Activity	2/272	1.7e-02	6.7e-01
Intermediate Filament Binding	2/272	2.0e-02	6.7e-01
Peptide Transmembrane Transporter Activity	2/272	2.0e-02	6.7e-01
Amino Acid Transmembrane Transporter Activity	5/272	2.0e-02	6.7e-01
Transcription Factor Binding	18/272	2.2e-02	6.7e-01
Cell Adhesion Molecule Binding	17/272	2.2e-02	6.7e-01
Organic Anion Transmembrane Transporter Activity	8/272	2.3e-02	6.7e-01
2-Oxoglutarate Dependent Dioxygenase Activity	4/272	2.3e-02	6.7e-01
Glycosyltransferase Activity	10/272	2.3e-02	6.7e-01
B-1-3-Galactosyltransferase Activity	2/272	2.6e-02	6.7e-01
Proton Transmembrane Transporter Activity	6/272	2.7e-02	6.7e-01
Glycosaminoglycan Binding	9/272	2.9e-02	6.7e-01
Mechanosensitive Ion Channel Activity	2/272	3.0e-02	6.7e-01
L-Amino Acid Transmembrane Transporter Activity	4/272	3.2e-02	6.7e-01
Demethylase Activity	3/272	3.2e-02	6.7e-01
Organic Acid Transmembrane Transporter Activity	7/272	3.4e-02	6.7e-01
Cadherin Binding	11/272	3.4e-02	6.7e-01
Insulin-Like Growth Factor Receptor Binding	2/272	3.4e-02	6.7e-01
Heparin Binding	7/272	3.6e-02	6.7e-01
Core Promoter Sequence-Specific DNA Binding	3/272	3.7e-02	6.7e-01
G-Protein Activity	3/272	3.7e-02	6.7e-01
Monosaccharide Binding	4/272	3.8e-02	6.7e-01
Magnesium Ion Transmembrane Transporter Activity	2/272	3.8e-02	6.7e-01
Protein Self-Association	4/272	4.2e-02	6.7e-01
NAD+ Binding	2/272	4.2e-02	6.7e-01
Opsonin Binding	2/272	4.2e-02	6.7e-01
Mannose Binding	2/272	4.7e-02	6.9e-01
Protein Serine Threonine Tyrosine Kinase Activity	3/272	4.9e-02	6.9e-01

Xu et al., Epigenome-wide characterization reveals aberrant DNA methylation of host genes regulating CD4⁺ T cell HIV-1 reservoir size in women with HIV

Figure Legends

Figure 1. Epigenome-wide association analysis of CD4⁺ T cell total HIV-1 reservoir size: (A) Manhattan plot; (B) Volcano plot showing positive and negative association of CpG sites with total HR; (C) Proportion of CpG site annotations.

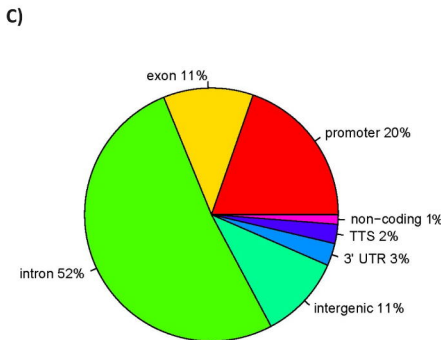
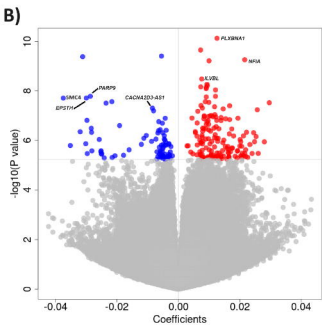
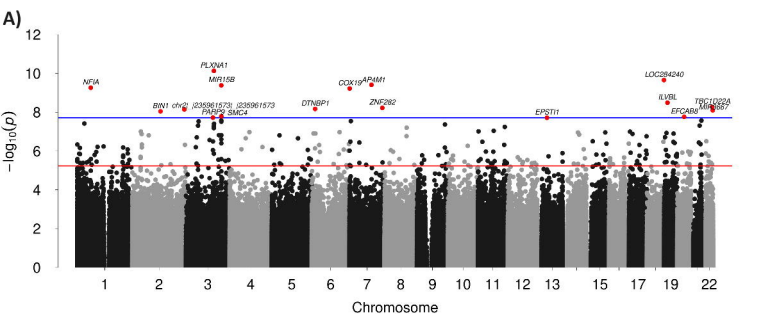
Figure 2. Differential methylation regions on four genes associated with HIV-1 reservoir size. (A) Interferon-stimulated gene 15 (*ISG15*); (B) Structural maintenance of chromosomes 4 (*SMC4*); (C) Interferon regulatory factor 9 (*IRF9*); (D) Interferon-induced protein 44 like (*IFI44L*). ddPCR-Total-low: bottom 25th percentile HIV-1 reservoir size; ddPCR-Total-medium: medium HIV-1 reservoir size; ddPCR-Total-high: top 25th percentile HIV-1 reservoir size.

Figure 3. The number of significant genes associated with the HIV-1 reservoir found in this study (Xu) and its overlap with the previously identified HIV-1 integration sites in three other studies by Wagner et al, Maldarelli et al, and Kok et al.

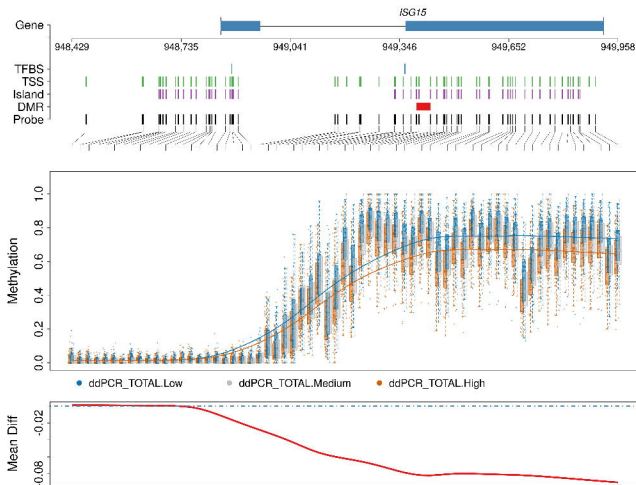
Figure 4. CpG sites differentially methylated between the high and low HIV-1 reservoir sizes on genes involved in HIV-1 integration sites. (A) *JAK1_chr1_6550357* $p=5.74e-07$; (B) two CpG sites on *NFIA* (*chr1_6145122*, $p=5.49e-10$ and *chr1_61542330*, $p=0.00004$); (C) three CpG sites on *ELMSAN1*(*chr14_74244367*, $p=6.62e-05$; *chr14_74250897*, $p=1.43e-07$; *chr14_74250926*: $p=0.00004$), and (D) two CpG sites on *DLEU2* (*chr13_50653658*: $p=0.00001$ and *chr13_506554306*: $p=1.88e-06$); (E) *SMAD2 chr18_45458622*: $p=0.0000128$

Figure 5. Significant pathways enriched on Hallmark gene set. (A) IFN- α and IFN- γ pathways; (B) HIV-1 reservoir-associated genes on those pathways.

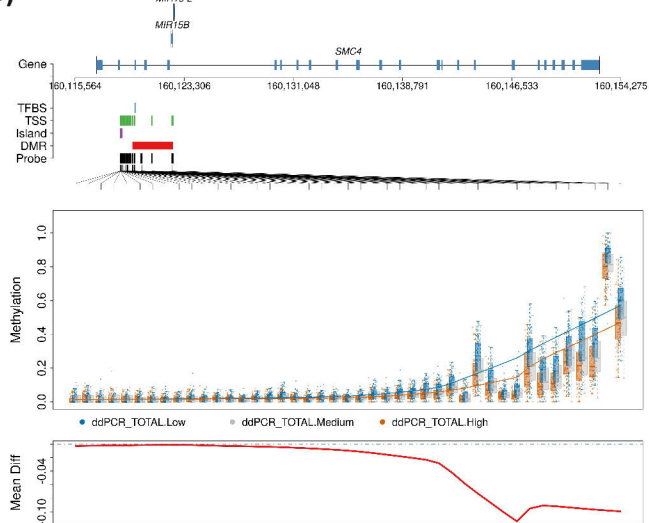
Figure 6. Gene set enrichment analysis results using GO Molecular function terms. (A) significant pathways; (B) HIV-1 reservoir-associated genes enriched on eight significant pathways; (C) the relationship between the top five pathways.



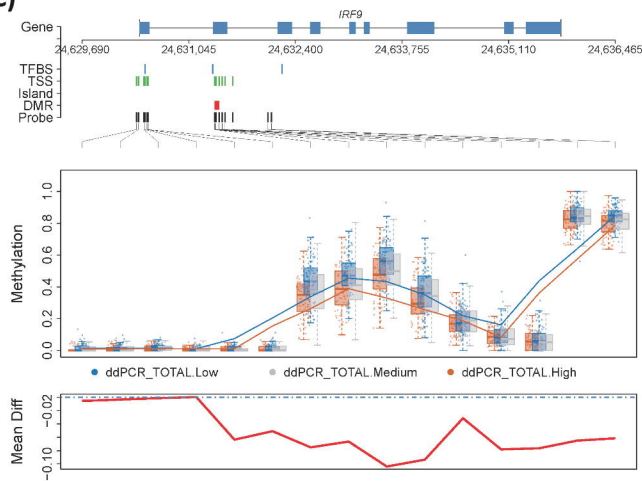
A)



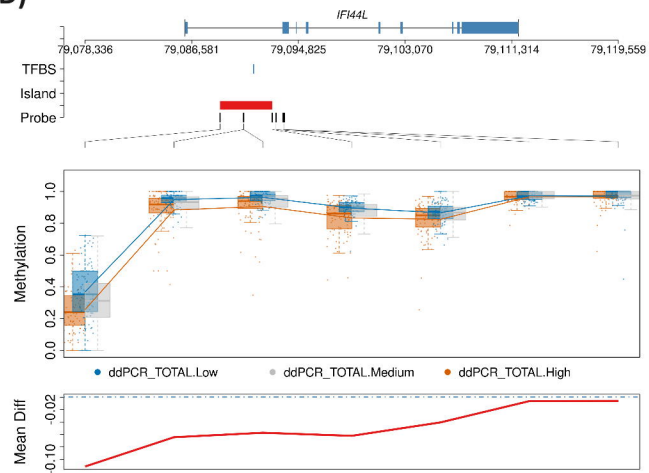
B)



C)

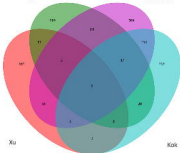


D)



Wagner

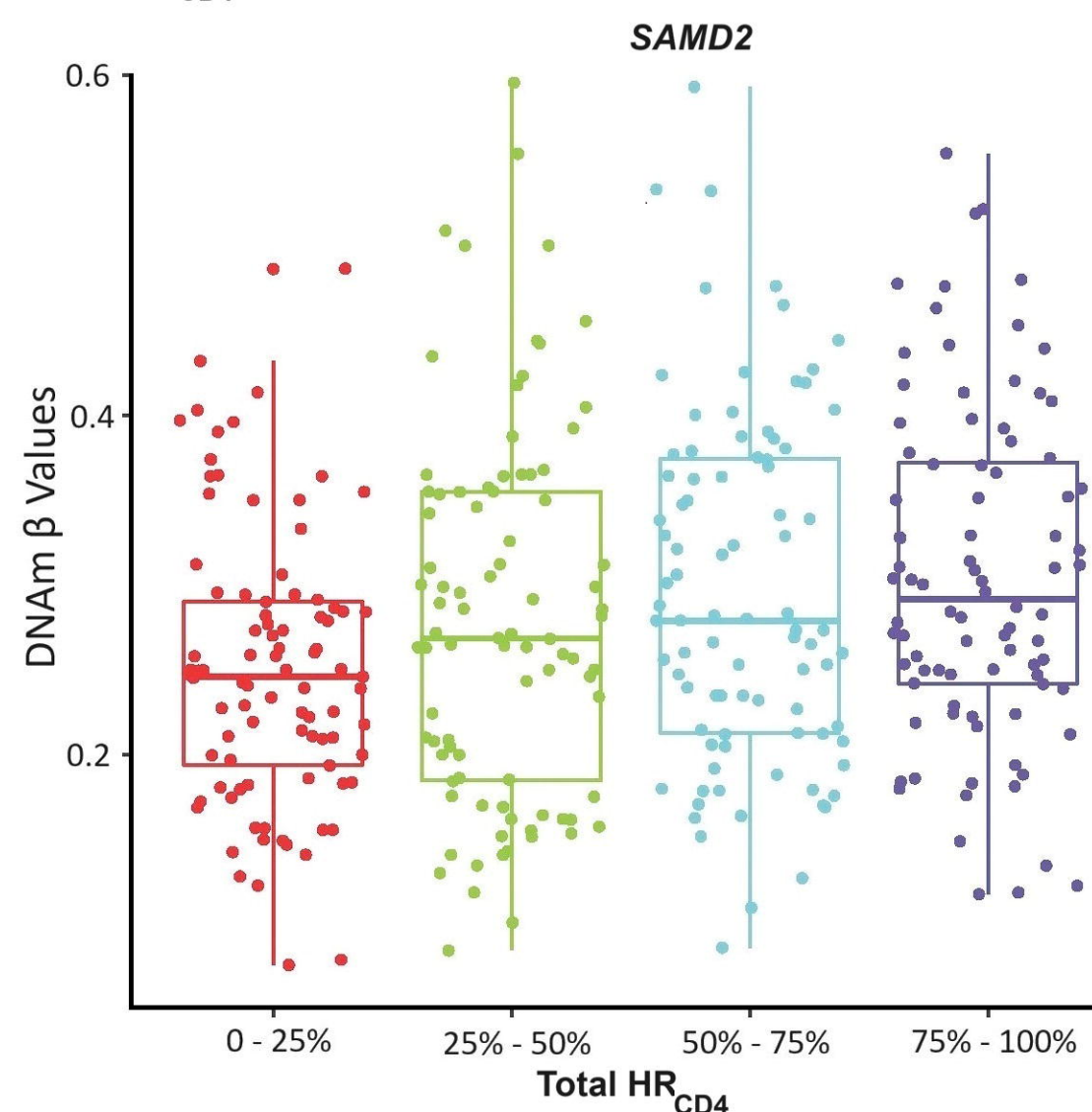
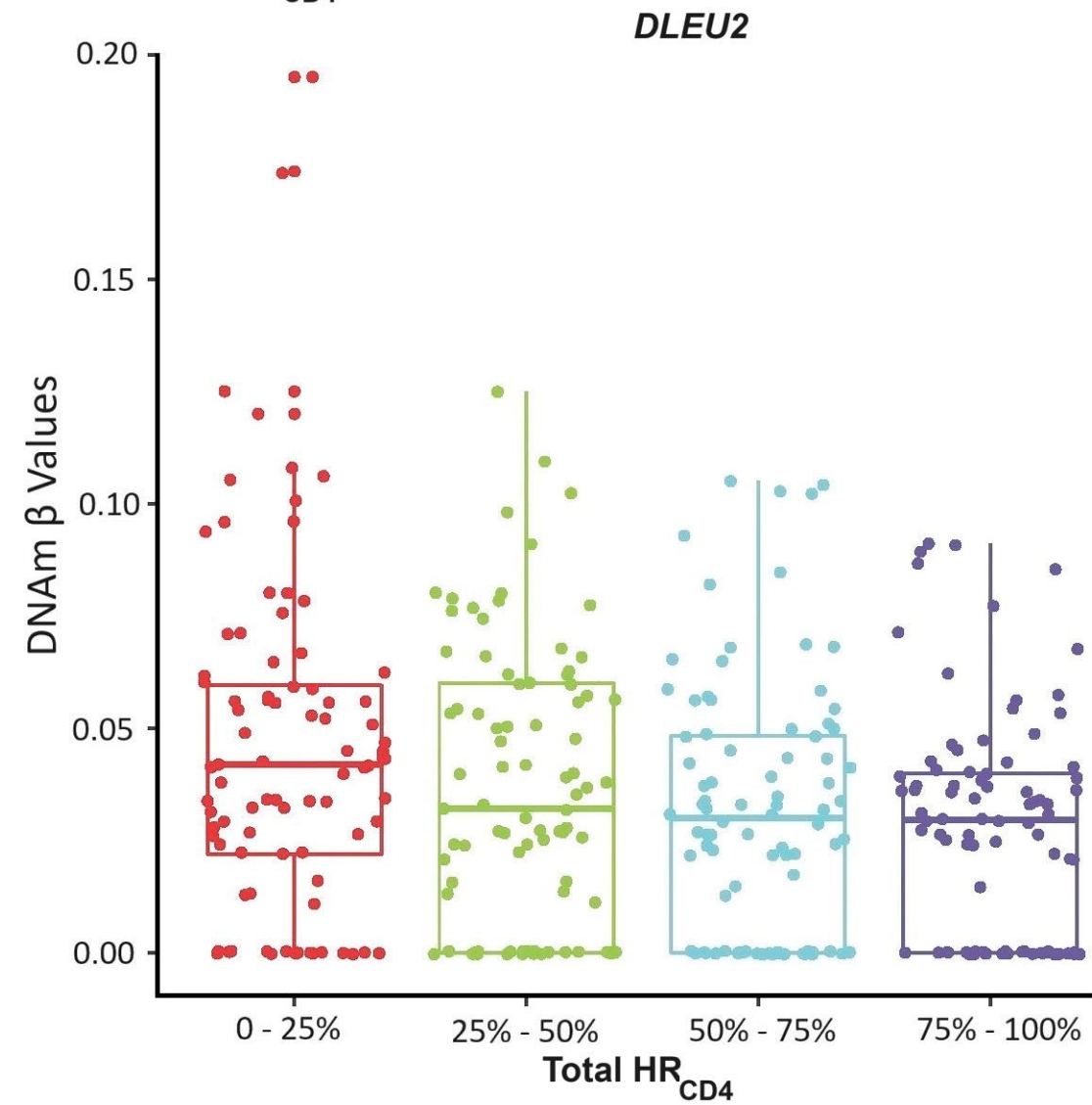
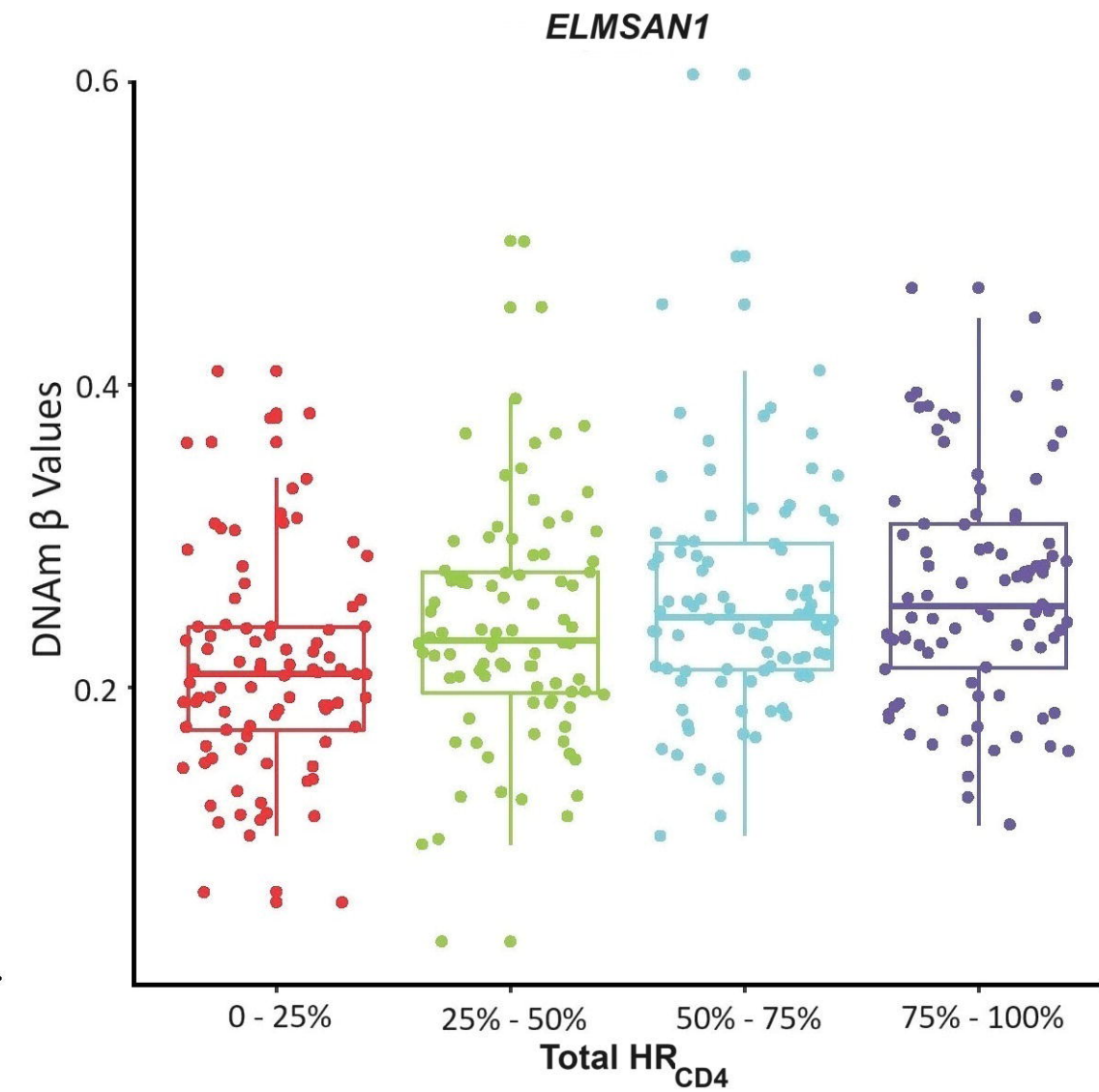
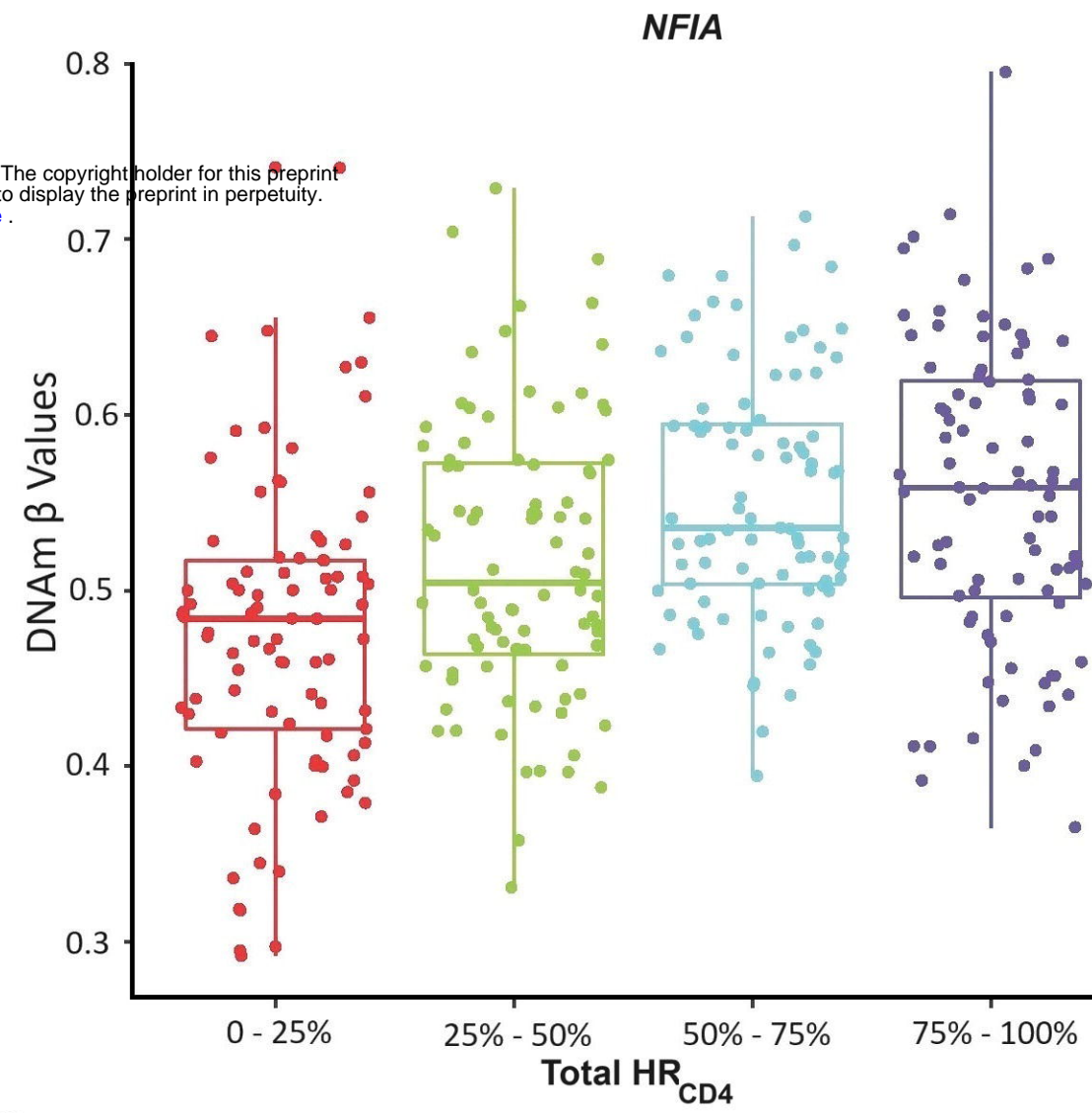
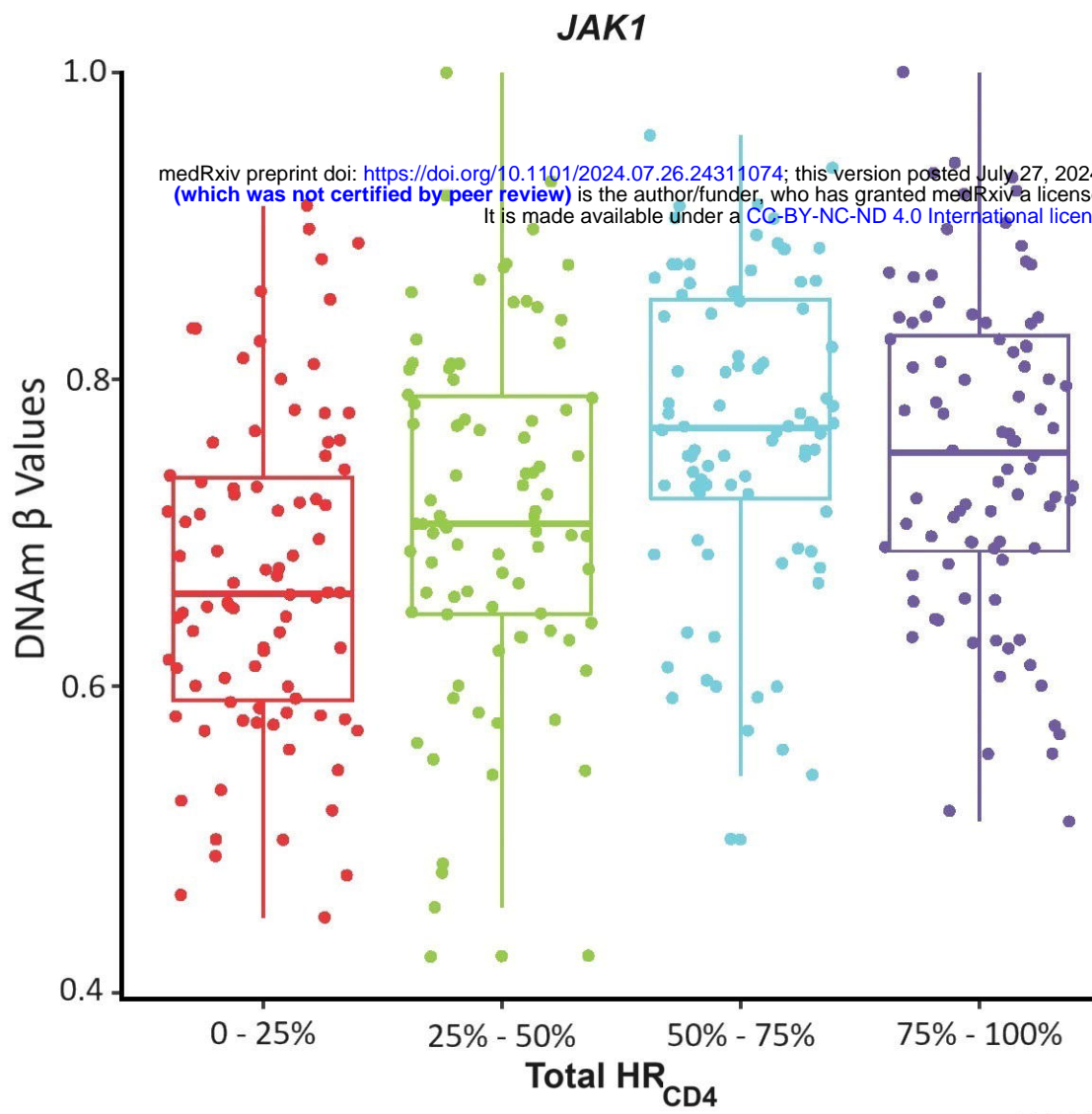
Makareli



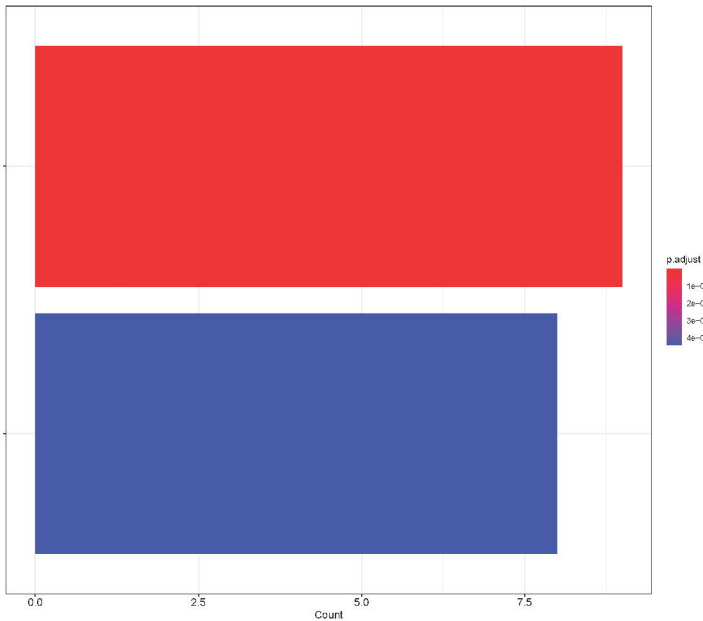
Xu

Kork

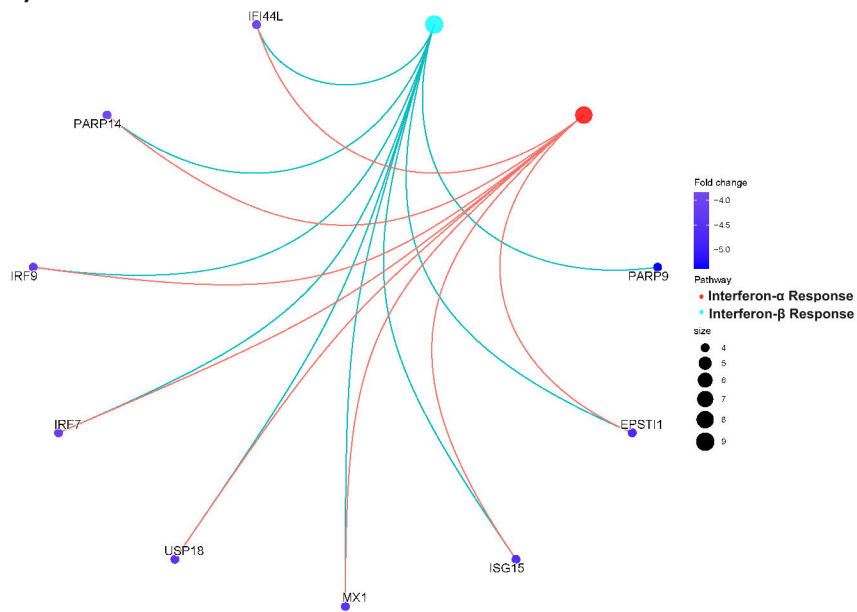
medRxiv preprint doi: <https://doi.org/10.1101/2024.07.26.24311074>; this version posted July 27, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY-NC-ND 4.0 International license](#).



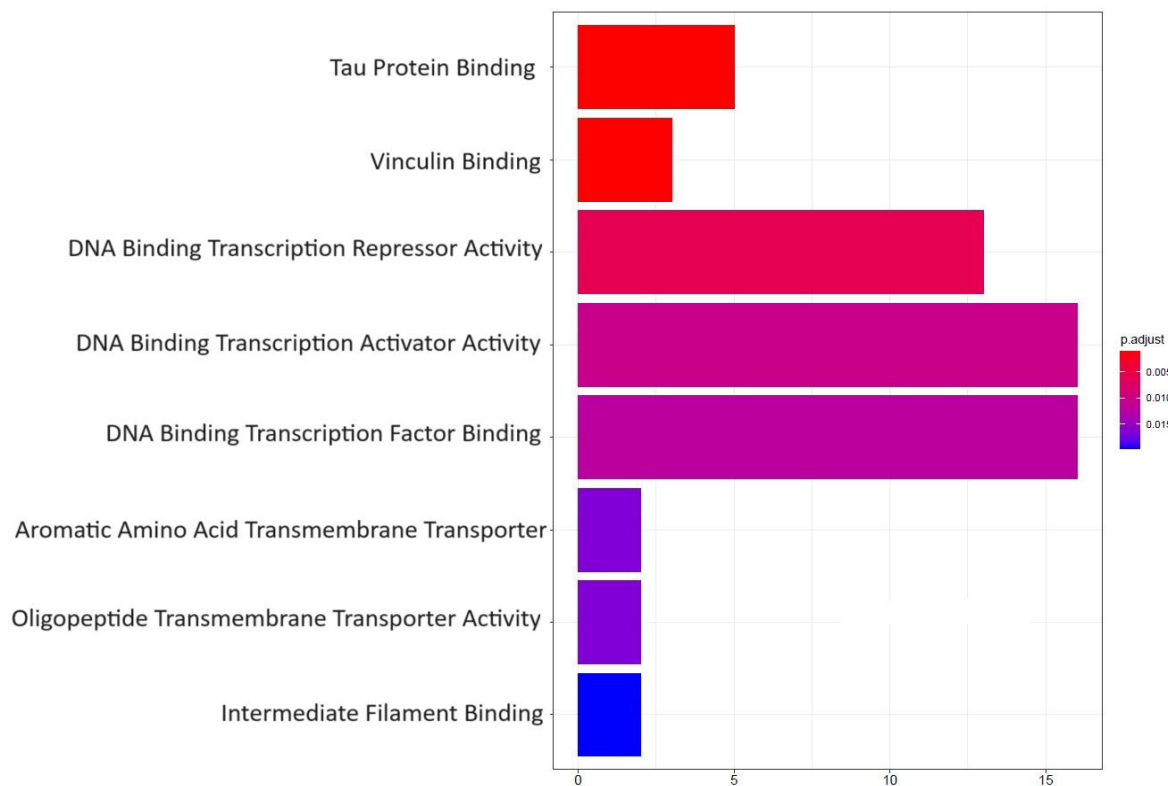
A)

Interferon- α ResponseInterferon- β Response

B)



A)



B)

