Results Summary

In this study, we analyzed HIV integration sites within the human genome, focusing on their relationship with chromatin accessibility in primary CD4+ T cells. Our analysis revealed that 44.52% of the identified HIV integration sites were located within regions of open chromatin (See supplementary material). Open chromatin regions were identified using data from ENCODE (60 projects), derived from ATAC-seq, ChIP-seq, and DNase-seq assays, specifically from biospecimens labeled as “CD4-positive, alpha-beta T cells.” These assays are widely recognized for their ability to measure chromatin accessibility: ATAC-seq captures regions of accessible DNA by identifying transposase-accessible sites(1), ChIP-seq provides insights into histone modifications associated with active or repressed chromatin states(2), and DNase-seq identifies DNase I hypersensitive sites, which are markers of open chromatin(3).

Gene set enrichment analysis further revealed that several pathways were significantly enriched at HIV integration sites, including those involved in chromatin-modifying enzymes (adjusted p-value = 0.05), apoptotic processes, and vesicle-mediated transport. These pathways suggest a potential mechanistic link between chromatin accessibility, cellular stress responses, and the HIV life cycle. Moreover, the most frequently observed gene was NPLOC4 (See Supplementary Table 1), identified in 17 patient samples, indicating that this gene may play an important role in HIV integration which agree with previous studies (4,5). Similar support was found for the second most frequent gene, DNMT1(6).

We selected these criteria for defining open chromatin regions based on their broad use and validation in genomic studies, particularly within the ENCODE project(7). ATAC-seq, ChIP-seq, and DNase-seq provide complementary information on chromatin states: ATAC-seq and DNase-seq both measure regions of chromatin accessibility, while ChIP-seq identifies histone marks that are indicative of either active or repressed chromatin(1–3). This combination of data ensures a comprehensive view of chromatin structure, making it ideal for investigating HIV's integration preferences. Our findings support the hypothesis that HIV preferentially integrates into open chromatin regions (44.52% HIV integration sites were located within regions of open chromatin), which may facilitate viral replication by leveraging accessible host genomic region(8,9).

References:

1. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods. 2013 Dec;10(12):1213–8.
2. Park PJ. ChIP-seq: Advantages and challenges of a maturing technology. Vol. 10, Nature Reviews Genetics. 2009. p. 669–80.
3. Boyle AP, Davis S, Shulha HP, Meltzer P, Margulies EH, Weng Z, et al. High-Resolution Mapping and Characterization of Open Chromatin across the Genome. Cell. 2008 Jan 25;132(2):311–22.
4. Francis AC, Marin M, Singh PK, Achuthan V, Prellberg MJ, Palermino-Rowland K, et al. HIV-1 replication complexes accumulate in nuclear speckles and integrate into speckle-associated genomic domains. Nat Commun. 2020 Dec 1;11(1).
5. Lucic B, Chen HC, Kuzman M, Zorita E, Wegner J, Minneker V, et al. Spatially clustered loci with multiple enhancers are frequent targets of HIV-1 integration. Nat Commun. 2019 Dec 1;10(1).
6. Youngblood B, Reich NO. The early expressed HIV-1 genes regulate DNMT1 expression. Epigenetics. 2008;3(3):149–56.
7. Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F, et al. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012 Sep 6;489(7414):57–74.
8. Lewinski MK, Bisgrove D, Shinn P, Chen H, Hoffmann C, Hannenhalli S, et al. Genome-Wide Analysis of Chromosomal Features Repressing Human Immunodeficiency Virus Transcription. J Virol. 2005 Jun;79(11):6610–9.
9. Han Y, Lassen K, Monie D, Sedaghat AR, Shimoji S, Liu X, et al. Resting CD4 + T Cells from Human Immunodeficiency Virus Type 1 (HIV-1)-Infected Individuals Carry Integrated HIV-1 Genomes within Actively Transcribed Host Genes . J Virol. 2004 Jun 15;78(12):6122–33.

Supplementary Materials:

Supplementary table 1. Top 20 most frequently observed genes in the highest number of patients. Additional genes are available upon request.

Gene	Participants
NPLOC4	17
DNMT1	13
QRICH1	12
HSF1	12
PACS1	11
NSD1	11
NADK	10
SPTAN1	10
LUC7L	10
KDM2A	10
PLEC	10
FBXW5	10
MIR7112	10
ZNF276	10
HERC1	10
LIME1	10
FAM193A	10
SF1	9
IMPDH2	9
MROH1	9

Supplementary table 2. Table ranking the top 20 most frequently observed gene sets harboring HIV integration sites.

pathway	pval	padj	ES	NES	nMoreExtrem	size
REACTOME_CHROMATIN_MODIFYING_ENZYMES	1.00E-04	0.050995	0.480316	1.601326	0	121
REACTOME_APOPTOTIC_CLEAVAGE_OF_CELLULAR_PROTEINS	0.000213	0.050995	0.716253	1.913341	1	15
REACTOME_VESICLE_MEDIATED_TRANSPORT	0.0003	0.050995	0.393231	1.357061	2	266
REACTOME_MEMBRANE_TRAFFICKING	0.0004	0.050995	0.396889	1.369016	3	263
REACTOME_SUMOYLATION	0.0007	0.070251	0.473192	1.547853	6	88
REACTOME_RNA_POLYMERASE_II_TRANSCRIPTION	0.001	0.070251	0.355878	1.252331	9	497
REACTOME_SIGNALING_BY_WNT	0.0011	0.070251	0.444656	1.47374	10	110
REACTOME_CELLULAR_RESPONSE_TO_HEAT_STRESS	0.001102	0.070251	0.518046	1.618969	10	48
REACTOME_ACTIVATION_OF_ANTERIOR_HOX_GENES_IN_HINDBRAIN_DEVELOPMENT_DURING_EARLY_EMBRYOGENESIS	0.001614	0.091462	0.556482	1.676724	15	33
REACTOME_ANTIGEN_PROCESSING_UBIQUITINATION_PROTEASOME_DEGRADATION	0.0024	0.116321	0.416004	1.398953	23	148
REACTOME_RMTS_METHYLATE_HISTONE_ARGININES	0.002509	0.116321	0.640427	1.772564	23	18
REACTOME_APOPTOTIC_EXECUTION_PHASE	0.003241	0.134119	0.627592	1.73704	30	18
REACTOME_TRANSCRIPTIONAL_REGULATION_BY_RUNX1	0.0038	0.134119	0.446163	1.454263	37	83
REACTOME_RUNX1_INTERACTS_WITH_CO_FACTORS_WHOSE_PRECISE_EFFECT_ON_RUNX1_TARGETS_IS_NOT_KNOWN	0.004015	0.134119	0.593188	1.694558	38	22
REACTOME_SIGNALING_BY_NOTCH	0.004601	0.134119	0.458377	1.47325	45	69
REACTOME_TCF_DEPENDENT_SIGNALING_IN_RESPONSE_TO_WNT	0.0047	0.134119	0.452575	1.4619	46	74
REACTOME_PKMTS_METHYLATE_HISTONE_LYSINES	0.00488	0.134119	0.559185	1.645594	47	27
REACTOME_TRANSCRIPTIONAL_REGULATION_BY_TP53	0.005099	0.134119	0.40067	1.350652	50	155
REACTOME_CLASS_I_MHC_MEDIATED_ANTIGEN_PROCESSING_PRESENTATION	0.005499	0.134119	0.394819	1.335389	54	166
REACTOME_RAB_REGULATION_OF_TRAFFICKING	0.005507	0.134119	0.48264	1.51826	54	52

Errors: Here I mention a couple of findings that I think should be fixed.

1. Only 25 patients matched the MGS-PCR data. The guidelines says there is 26 participants. The missing files correspond to the label ID SCO2453.
2. I think the slide with the figures for Hit count is wrong.

Coding: I implemented a GitHub repository where you can find the codes, outputs, and inputs I used for this small project. In the input folder there is not data because it is patient info. So, I added a note saying were to find the data. Here is the link of the repository: https://github.com/bciezah1/HIV_Insertion_Test.git