0358

Chimeric Virus-host Transcripts Reveal Integration Sites in Macaque SIV+ CD4 T Cells

Mykhaylo Artamonov,[1,2*] Ales Varabyou,[3,4*] Kevin Tracy,[1,2] Sean C. Patro[5], Wenjie Wang,[5] Xiaolin Wu,[5] Sophia Cheng,[1,2] Thuy Nguyen,[5] Frank Maldarelli,[5] **Diane L. Bolton**[1,2]

[1]U.S. Military HIV Research Program, CIDR, Walter Reed Army Institute of Research, Silver Spring, MD; [2]Henry M. Jackson Foundation for the Advancement of Military Medicine, Bethesda, MD; [3]Center for Computational Biology, Johns Hopkins University, Baltimore, MD; [4]Department of Computer Science, Johns Hopkins University, Baltimore, MD; [5]Cancer Research Technology Program, Leidos Biomedical Research, Inc., Frederick National Laboratory for Cancer Research, Frederick, MD

BACKGROUND

- We hypothesized that transcriptionally active proviruses would exhibit integration site bias towards highly expressed host genes.

Elucidating mechanisms governing proviral integration and transcription may aid in identifying strategies to limit HIV-1 reservoirs. Chimeric host-viral transcripts are a mechanism by which HIV-1 modulates host gene expression to promote reservoir persistence, though they are poor characterized *in vivo*. We analyzed SIV RNA+ CD4 T cells for viral, host, and chimeric transcripts during acute infection in rhesus macaques by scRNAseq, identifying integration sites (IS) and their expression.

METHODS

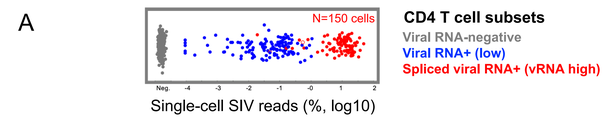


Figure 1 . SIV RNA+ CD4 T cell isolation for host-virus chimeric RNA detection and integration site mapping. A. SIV RNA content of memory CD4 T cells isolated by single-cell FACS sorting. **B.** Analysis pipeline for identification of host-virus chimeric RNA from spliced viral RNA+ cells and identification of retroviral integration site (RIS).

Rhesus macaques (n=2) experimentally infected with SIVmac251 or 239X were sampled for peripheral blood and lymph node 10-21 days post-infection. Memory CD4 T cells were isolated by flow cytometry for single-cell RNA sequencing of cDNA generated by SMART-Seq technology followed by Illumina NovaSeq sequencing. Cells expressing spliced viral transcripts (svRNA; n=150)[1] were analyzed for chimeric host-viral mRNA (**Fig.1A**). A custom Python script detecting chimeric transcripts identified sites of host-virus DNA fusion,[2] differentiating between spliced and readthrough chimeric transcription events **(Fig. 1B).** Separate cell aliquots were DNA sequenced in bulk using linker-mediated PCR to both 5' and 3' SIV LTR and mapping to RheMac10 (UCSC).

Integration site genes harboring active proviruses exhibit diverse transcriptional activity, including evidence that host gene expression is not required for integration or proviral transcription.

RESULTS

Chimeric mRNA constitutes ~0.1% of vRNA

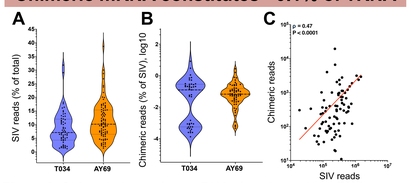


Figure 2. Chimeric host-virus mRNA abundance in svRNA+ CD4+ T cells. A. Proportion of scRNAseq reads aligning to SIV. **B.** Proportion of SIV reads that are rhesus-SIV chimeras. **C.** Spearman rank correlation between SIV and chimeric SIV-rhesus reads.

Multiple mechanisms contribute to chimeric host-virus mRNA

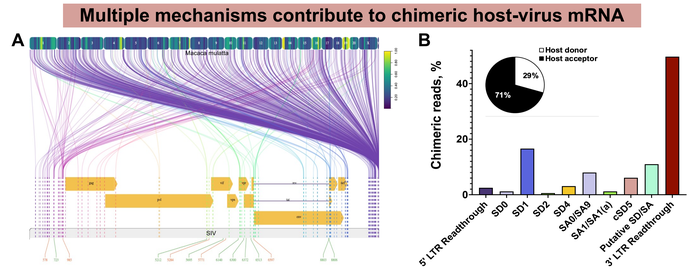


Figure 3. Chimeric read junction mapping to host and viral genome. A. Chimeric read break points depicted as lines connecting alignment position in the host and viral genomes. **B.** Proportion of chimeric reads formed through the indicated splicing or readthrough mechanism.

Frequent CD4+ T cell superinfection

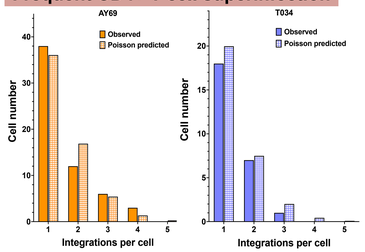


Figure 4. CD4+ T cell superinfection during acute SIV infection. Number of unique integration sites identified per cell by scRNAseq chimeric SIV-host mRNA analysis. Poisson distribution probability predicted values are overlaid.

Diverse transcriptional states of integration site genes

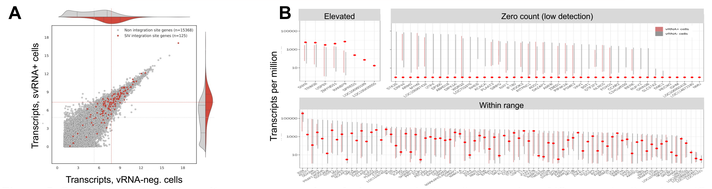


Figure 5. Single-cell integration site gene expression. A. Average single-cell expression of IS genes (red) and non-IS genes (gray) among svRNA+ (y-axis) and viral RNA-negative (x-axis) memory CD4+ T cells. Dotted lines depict medians. Gene expression is displayed as log2 normalized counts. **B.** Comparison of IS gene expression for the infected cell with SIV integrated at the indicated gene (red circle) to the range of expression observed for vRNA-negative cells (black bar) and svRNA+ cells with integration elsewhere (pink bar; 95% prediction intervals).

Strong biases and non-randomness of integration sites identified by DNA sequencing

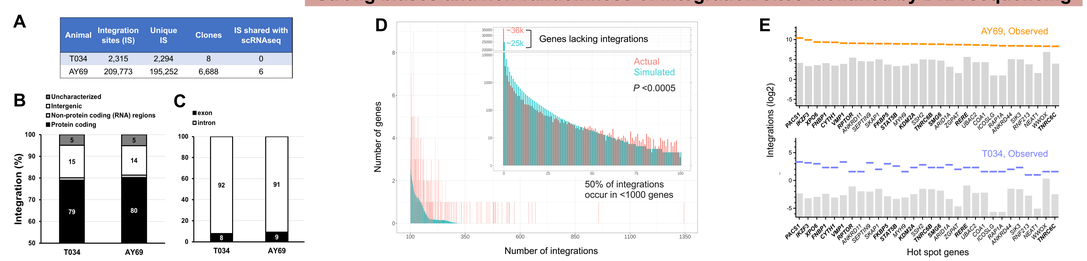


Figure 6. SIV integration sites identified by DNA sequencing during acute SIV infection. Mononuclear cells from two acutely SIV-infected macaques were DNA sequenced for proviral integrations. **A.** Number of IS mapped for each animal. **B.** Proportion of SIV integrations located in genic/unannotated versus non-genic/pseudogenic/RNA regions. **C**. Proportion of genic integrations within extronic and intronic regions (limited to IS mapped with nucleotide precision by 3' readthrough transcription). **D.** Cumulative distribution of integrations per gene is shown (actual) compared to predicted values (simulated) in animal AY69. Inset depicts genes with 0-100 integrations. P value depicts Chi-squared with Monte Carlo simulation for non-randomness. **E.** Comparison of the expected (gray bars) versus observed (colored lines) number of integrations within the 30 genes with the most integrations.

CONCLUSIONS

- Virus-host chimeric transcription common among replication active infected cells.
- Transcriptionally active proviruses preferentially reside in genes with elevated expression, although a substantial fraction (~30%) target genes with low or undetectable expression, indicating that host gene activity is not required for viral integration or transcription.
- Host–virus chimeric transcripts revealed integration sites concordant with DNA-based mapping, with 3' LTR readthrough as a major source of chimeric mRNA.
- 30% of replication-active infected cells harbor multiple proviruses during acute infection, highlighting a potential mechanism for retroviral recombination and increased viral diversity.

IMPORTANCE

SIV transcription independent of integration site gene expression *in vivo* suggests latency reversal strategies may have limited impact on inducing HIV-1. Insertional activation of integration site genes and potential novel proteins formed via chimeric host-virus RNA may contribute to proliferation and survival of infected CD4⁺ T cells, posing further barriers to HIV-1 cure.

ADDITIONAL KEY INFORMATION

AUTHOR CONTACT INFORMATION

DBOLTON@hivresearch.org

REFERENCES

[1]Cobos-Jiménez V et al. *iScience*. 2023.
[2]Liu R et al. *Sci Transl Med*. 2020.

ACKNOWLEDGEMENTS

We thank Drs. Michael Corley and Howard Fox for sharing long-read single-cell RNA sequencing data from an SIV-infected macaque used here for aligning SIV transcripts. Dr. Rasmi Thomas performed Illumina scRNAseq.[1] This work was supported by a cooperative agreement (W81XWH-18-2-0040) between the Henry M. Jackson Foundation for the Advancement of Military Medicine, Inc., and the U.S. Department of Defense (DOD).

PLAIN LANGUAGE SUMMARY

The location of HIV-1 in the host cell genome is believed to influence whether the virus is latent or active. Here, we find that active viruses frequently reside in inactive host genes, suggesting that HIV-1 eradication strategies may be most effective if they target the viral promoter.

Material has been reviewed by the Walter Reed Army Institute of Research. There is no objection to its presentation and/or publication. The opinions or assertions contained herein are the private views of the author, and are not to be construed as official, or as reflecting true views of the Department of the Army or the Department of Defense. Research was conducted under an IACUC-approved animal use protocol in an AAALAC International - accredited facility with a Public Health Services Animal Welfare Assurance and in compliance with the Animal Welfare Act and other federal statutes and regulations relating to laboratory animals.