



Where Viruses Land: Correlations Between R-Loops and Retroviral Integration

Citation

Singh, Parmit and Alan Engelman. "Where Viruses Land: Correlations Between R-Loops and Retroviral Integration". Conference Talk, 2025

Link

<https://dash.harvard.edu/handle/1/42730337>

Terms of use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material (LAA), as set forth at

<https://harvardwiki.atlassian.net/wiki/external/NGY5NDE4ZjgzNTc5NDQzMGIzZWZhMGFIOWI2M2EwYTg>

Accessibility

<https://accessibility.huit.harvard.edu/digital-accessibility-policy>

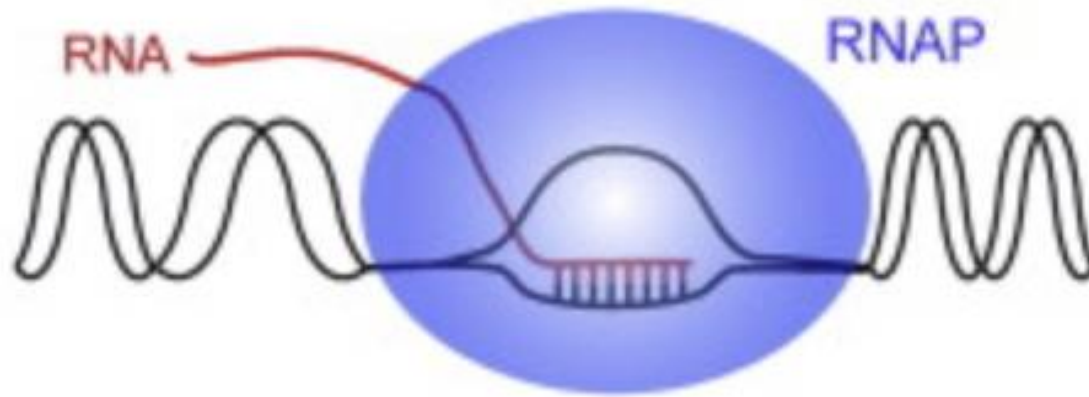
Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#)

Where Viruses Land: Correlations Between R-Loops and Retroviral Integration

Parmit Singh

What are R-loops?



Aguilera and Garcia-Muse, Molecular Cell, 2012

Tri-strand structure (DNA: RNA hybrid + displaced ssDNA)

Occurs co-transcriptionally, especially at gene ends and regulatory regions.

Retroviral integration is not random

Retroviruses HIV-1 prefer integration into active genes.

This preferences is dependent in viral – host proteins interactions.

LEDGF/p75 and integrase
Viral capsid and CPSF6

Speckle associated domains (SPADs)
Gene-dense chromosomal regions
Transcription/splicing

Key Findings from Recent Studies

Aquarius helicase: Promotes HIV-1 integration into **R-loop–enriched regions**. Penzo et al., *Nat Microbiol*, 2025

HIV-1 induces R-loops and preferentially integrates nearby. Park et al., *eLife*, 2024

PSIP1/LEDGF: Reduces R-loops at active genes to preserve genome integrity. Jayakumar et al., *Nature Communications*, 2024

Method

20 different published R-loops coordinates.

Detection is based on the use of S9.6 monoclonal antibody-based R-loop pulldown and followed by deep sequencing and mapping to the genome.

DRIP= DNA-RNA immunoprecipitation sequencing.

Published HIV-1 integration sites:

Sowd et al., PNAS, 2016

Li et al, mBio, 2020

Ferris et al., PLoS Pathog, 2019

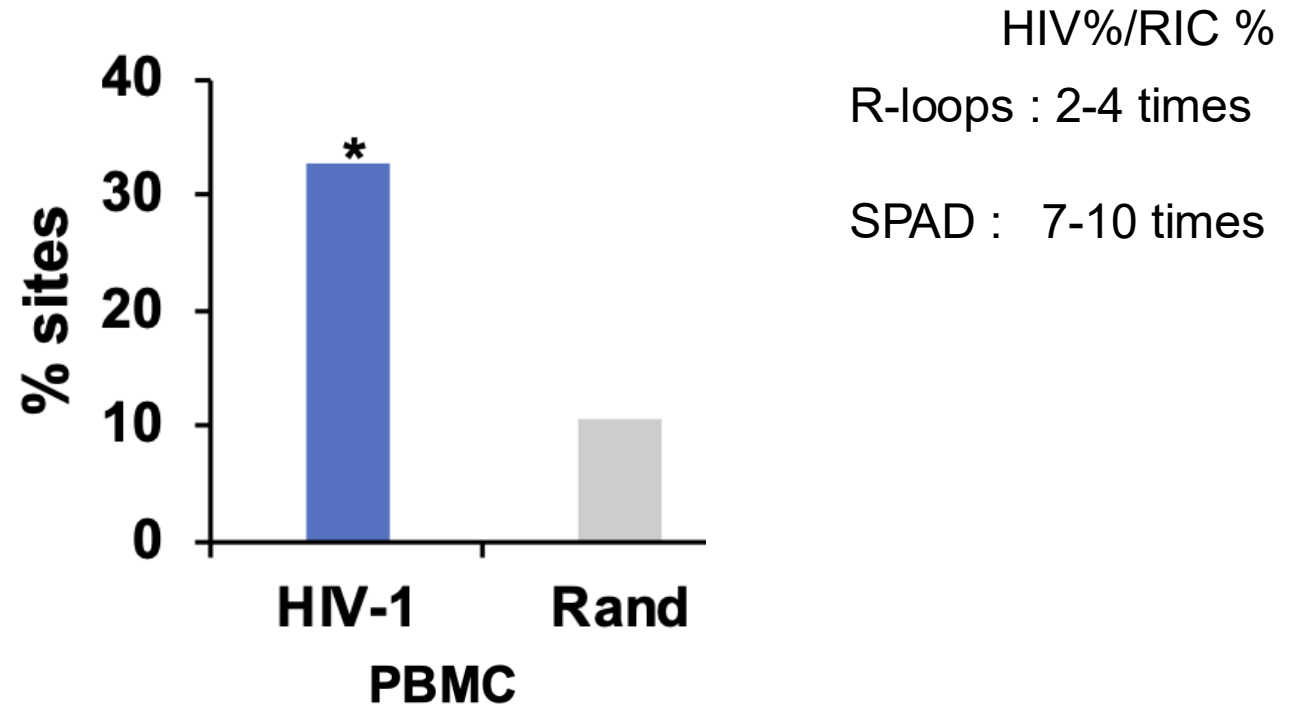
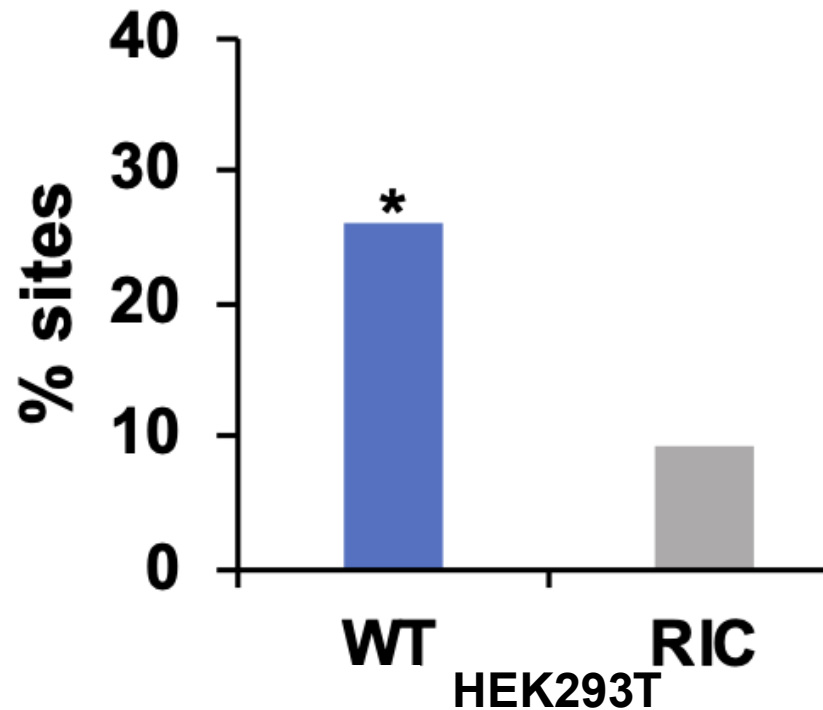


hg19 human genome

Elife Datasets

Calculated the amount of integration sites +/- 2.5 kb of R-loops.

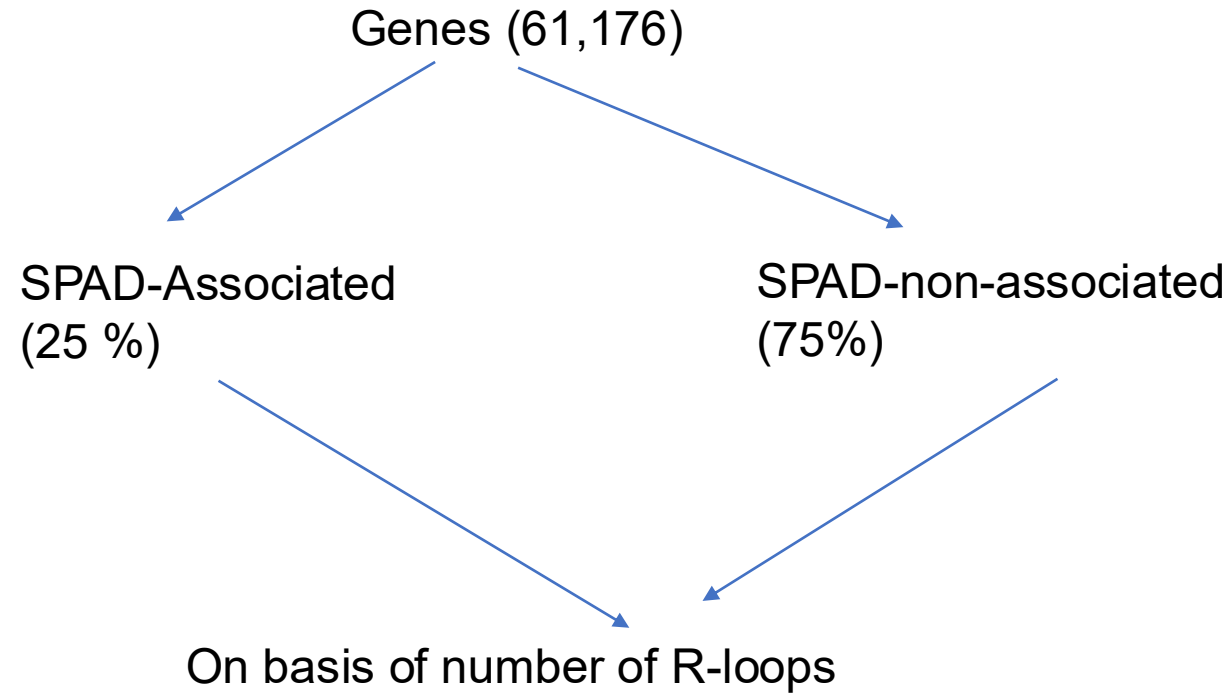
HIV-1 targets R-loops-proximal regions (+/-2.5kb)



HIV-2, Equine infectious anemia virus (EIAV), Feline immunodeficiency virus (FIV), Bovine immunodeficiency virus (BIV), Porcine endogenous retrovirus (PERV), Xenotropic murine leukemia virus-related virus (XMRV), Murine Leukemia Virus (MLV), Human T-cell leukemia virus (HTLV)

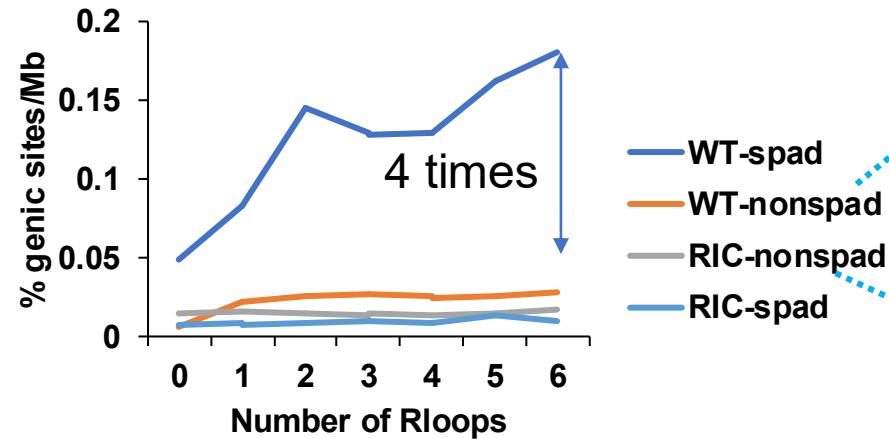
Prototype Foamy virus (PFV) and Mouse mammary tumor virus (MMTV) did not show such preference compared to RIC.

Gene Grouping by SPADs and R-Loops

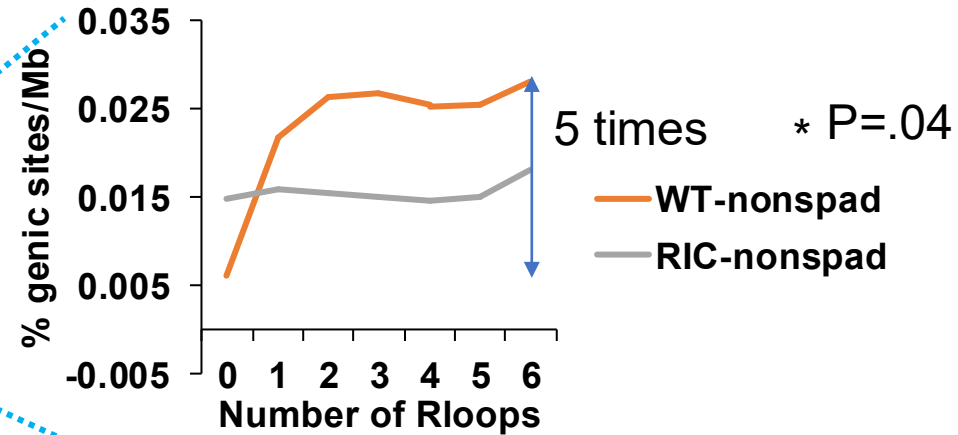


Genes with More R-Loops Are Preferentially Targeted Regardless of SPAD Association

$R^2 \text{ SPAD} = 0.83$

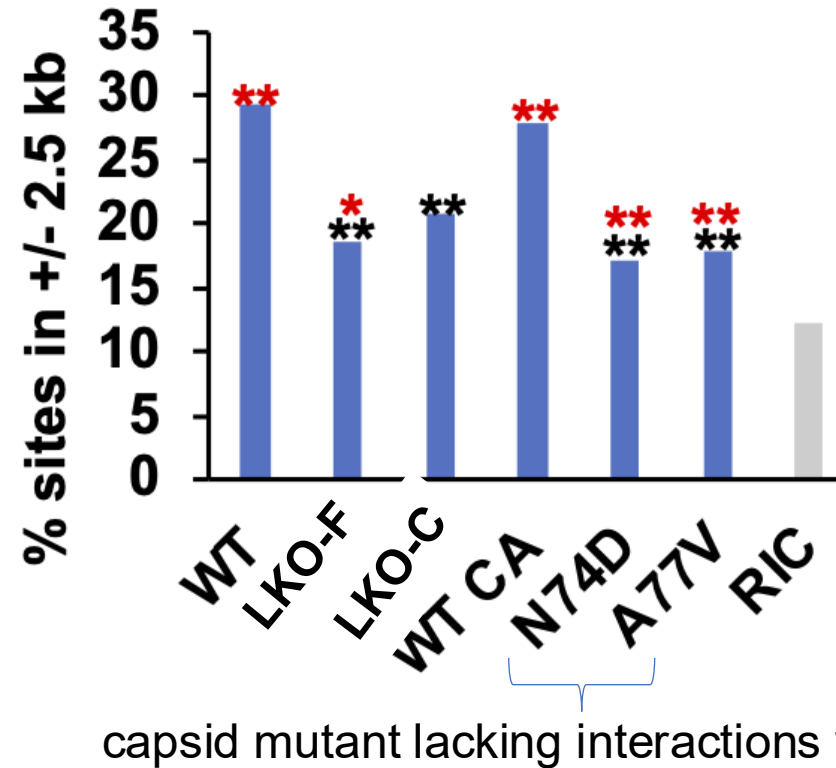
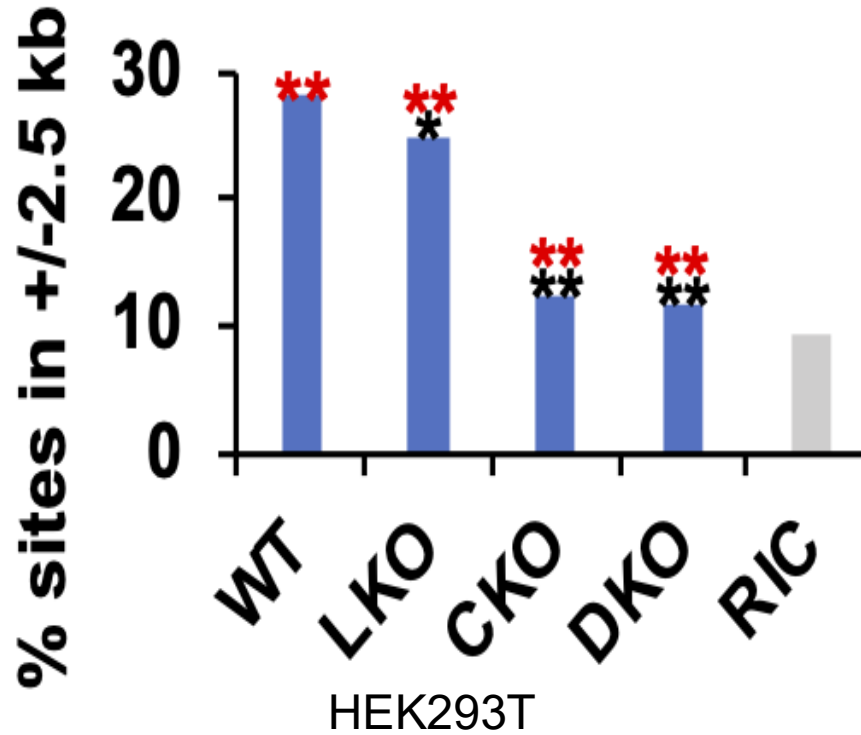


$R^2 \text{ nonSPAD} = 0.53$



SPAD=speckles associated domains

Role of host factors

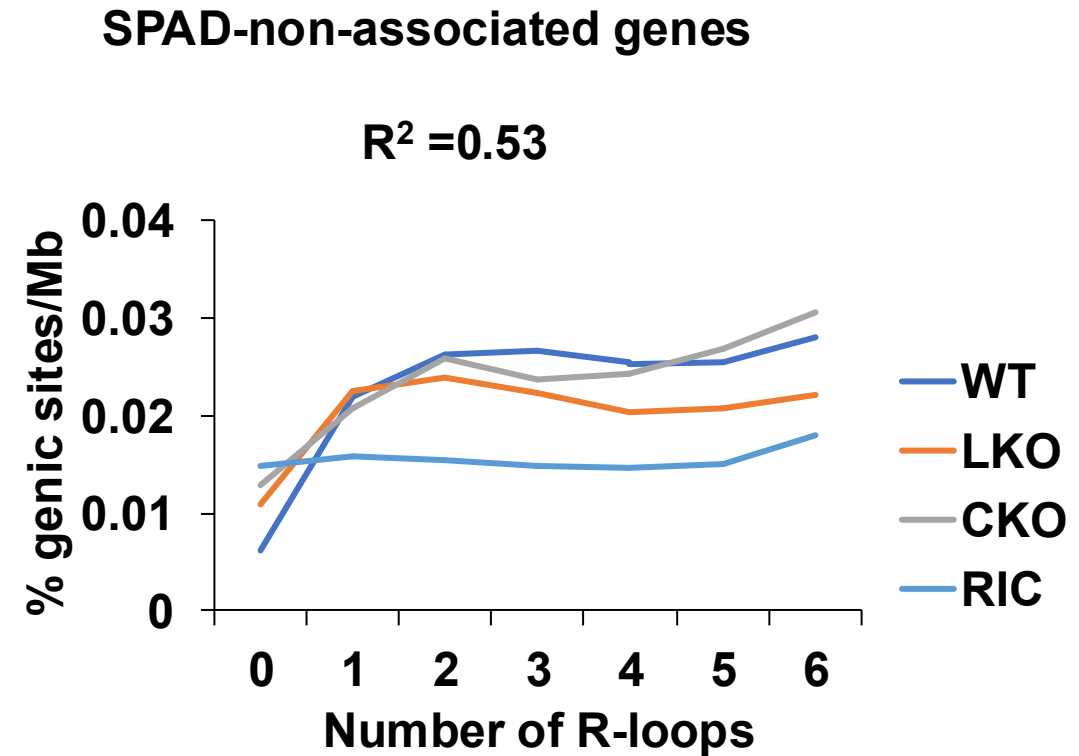
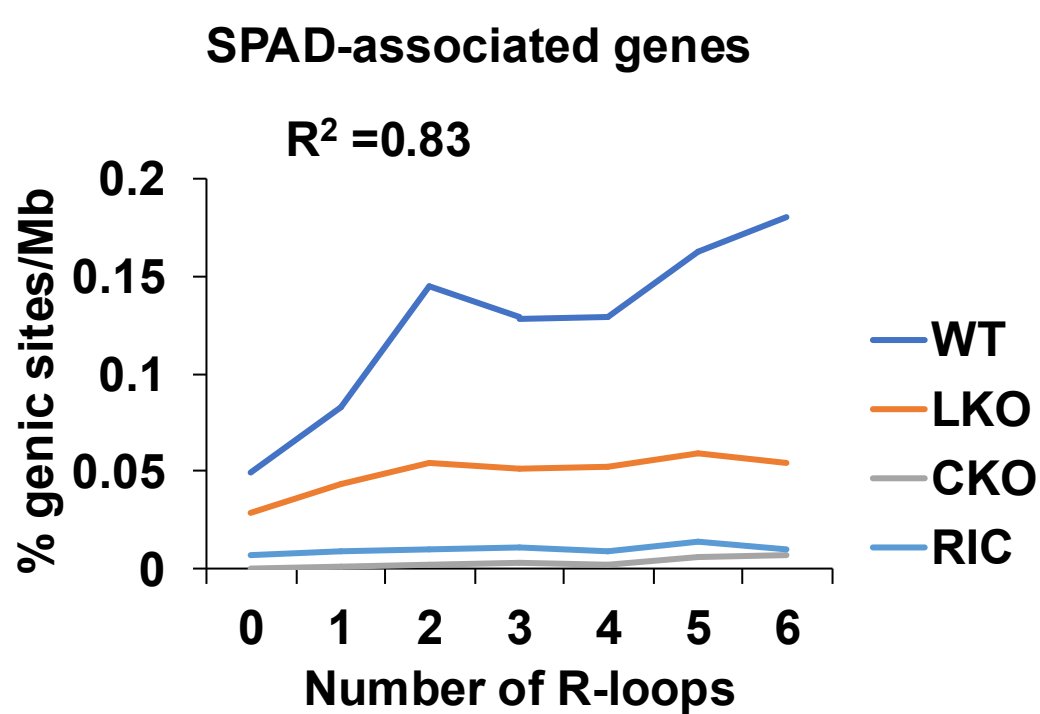


Jurkat

* RIC
* WT

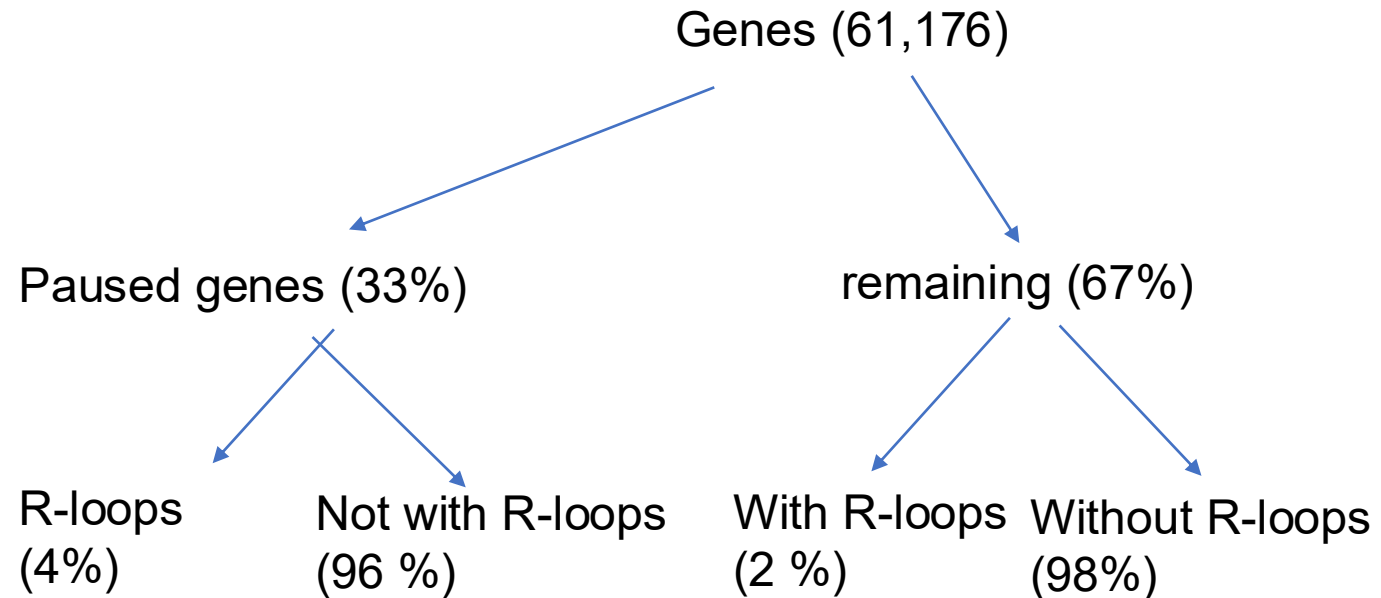
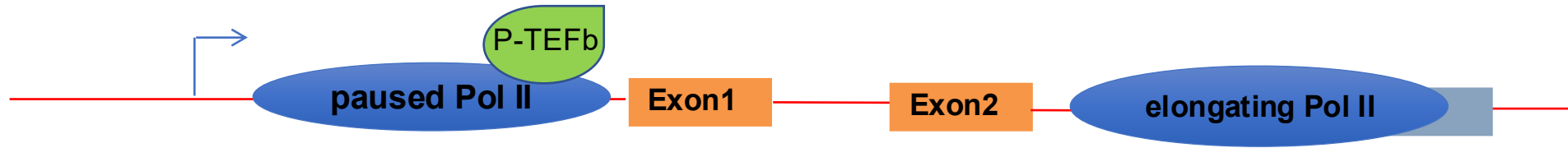
LKO cells lack LEDGF/p75. CKO cells lack CPSF6 and DKO cells lack both factors.

Genes with More R-Loops Are Preferentially Targeted Regardless of SPAD Association



SPAD=speckles associated domains

Gene Grouping by Pol II Pausing and R-Loops



Elife datasets

R-loops

After 0, 3, 6 and 12 hrs of infection

HeLa, Jurkat

CD4

Integration sites

After 5 days of infection

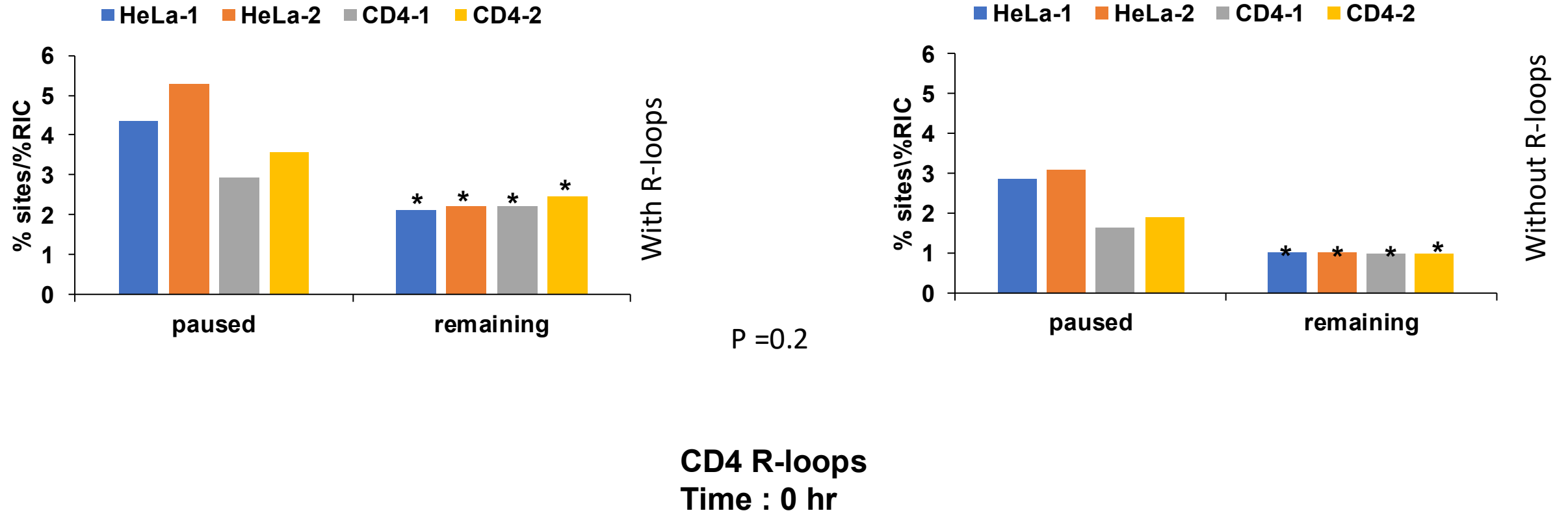
HeLa-1

HeLa-2

CD4-1

CD4-2

Paused Genes Are Preferentially Targeted Regardless of R-Loop Presence



Preference for paused genes persists with or without R-loops.

Summary

HIV-1 highly targets the proximal regions of R-loops.

The preference of integration in R-loop regions is reduced in LKO and CKO cells.

Lentiviruses, gamma-retroviruses, and delta-retroviruses also show a preference to integrate near R-loops.

Spumaviruses and betaretroviruses do not show such preference.

Genes with a higher number of R-loops are more frequently targeted, independent of SPAD status.

Paused genes are more frequently targeted than non-paused genes, independent of R-loop levels.

Challenges in Interpreting the Aquarius Helicase Findings

- R-loops are enriched in recurrent integration genes (RIGs), but also present in 38% of recurrently avoided genes (RAGs) and 55% of genes with only 1 integration → presence alone does not determine HIV-1 targeting.
- R-loops enriched in SPADs, but a major fraction of RIGs occur outside SPADs (Fig. 1g) → unclear whether SPADs or R-loops drive integration.
- Authors state integration bias is strong for HIV-1/HIV-2 but *markedly lower for MLV*, yet their own KD data show MLV behavior changes in the opposite direction — implies inconsistent interpretation.

Acknowledgments

Dr. Alan Engelman (mentor)

Lab members

Gregory Bedwell

Soojin Jang

Rajalingam Radhakrishnan

Bjarki Arnarson

Zhen Li

Zach Taylor