Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

_____ _____

Raven Shah Date

Multiplexed approaches to investigate cellular mechanisms underlying HIV-1 transcriptional competence and viral replication

By
Raven Shah
Doctor of Philosophy

Biochemistry, Cell and Developmental Biology
Graduate Division of Biological and Biomedical Sciences

_____
Stefan G. Sarafianos, Ph.D.
Advisor

_____
Michael Koval, Ph.D.
Committee Member

_____
Graeme L. Conn, Ph.D.
Committee Member

_____
Gregory Melikian, Ph.D.
Committee Member

_____
Bo Liang, Ph.D.
Committee Member

_____
Victor G. Corces, Ph.D.
Committee Member

Accepted:

_____
Kimberly J. Arriola, Ph.D., MPH
Dean of the James T. Laney School of Graduate Studies

_____
Date

Multiplexed approaches to investigate cellular mechanisms underlying HIV-1 transcriptional competence and viral replication

By

Raven Shah

B.S., University of Rochester, 2015

Advisor: Stefan G. Sarafianos, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Biochemistry, Cell, and Developmental Biology
2021

ABSTRACT

Multiplexed approaches to investigate cellular mechanisms underlying HIV-1 transcriptional competence and viral replication

By Raven Shah

Human immunodeficiency virus type 1 (HIV-1) remains a global public health burden with approximately 38 million individuals living with the disease in 2019 and 1.7 million new annual infections. There is currently no effective HIV-1 cure or vaccine. HIV is a retrovirus that inserts its genome into the target host chromatin, thus exploiting the cellular transcriptional machinery for replication. Investigating features of chromatin architecture at HIV sites of integration can provide insight into the molecular mechanisms governing HIV gene expression. Previous studies have demonstrated that HIV preferentially integrates into the introns of transcriptionally active genes, and productive HIV proviruses are associated with active epigenetic markers. The work presented in this thesis expands upon preliminary studies, undertaking an integrative approach to profile the relationship between HIV-1 transcription and chromatin state, map the local cellular and viral transcriptome(s) at HIV-host gene boundaries, and monitor viral transcriptional dynamics using multiplexed fluorescence imaging.

In **Chapter II**, we present an innovative and highly sensitive methodology that we "coin" Multiplexed immunofluorescent cell-based detection of DNA, RNA, and protein (MICDDRP), to simultaneously label viral nucleic acid(s) and protein(s) to visualize viral replication kinetics at single-cell resolution across a broad spectrum of viruses. **Chapters III-V** further demonstrate the power of our imaging modality, detailing how this technology can be used to study virus-host factor interactions **(Chapter III)**, visualize how small molecules can impact viral transcription **(Chapter IV)**, and measure virus replication kinetics with high spatiotemporal resolution **(Chapter V)**. In **Chapter VI**, we apply a multiomics sequencing approach to profile chromatin and gene expression at proviral sites of integration using HIV-inducible cellular models. We seek to understand how lentiviral integration and activation of HIV-1 transcription can alter cellular chromatin structure and the local transcriptional environment. Our study provides the first in-depth integrative investigation of HIV-1 chromatin ultrastructure and viral transcription at provirus-host gene boundaries using high-resolution chromatin mapping techniques (ATAC-seq and Hi-C) and long-read Nanopore RNA-sequencing. Our presented findings may have translational implications providing insight into mechanisms influencing HIV-1 transcriptional competency, as well as providing a platform for applying cutting-edge sequencing and *in-situ* imaging technologies to study viral replication.

Multiplexed approaches to investigate cellular mechanisms underlying HIV-1 transcriptional competence and viral replication

By

Raven Shah

B.S., University of Rochester, 2015

Advisor: Stefan G. Sarafianos, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Biochemistry, Cell, and Developmental Biology
2021

PUBLICATIONS

PUBLICATIONS RELATED TO THESIS WORK

1) **Shah, R.,** Gallardo, C.M., Dixon, J.R., Jung, Y.H., McFadden W., Torbett, B., Corces, V.G., Tedbury, P.R., Sarafianos, S.G. (2021) Multiomics profiling of HIV-infected cells reveals unique chromatin signature associated with active proviruses. (Planned submission to *PNAS*)

2) **Shah, R.,** Lan, S., Ong, Y.T., Boggs, E.A., Tedbury, P.R., Sarafianos, S.G. (2021) Multiplexed fluorescence-based single-cell imaging to capture early replication events of SARS-CoV-2 infection. (*In prep*)

3) Lan, S., Ong, Y.T., **Shah, R.**, Boggs, E.A., Tedbury, P.R., Sarafianos, S.G. (2021) Development and characterization of SARS-CoV-2 replicon system. (*In prep*)

4) Mahboubi, D., **Shah, R.,** Puray-Chavez, M., Poeschla, E.M., Engelman, A.N., Tedbury, P.R., Sarafianos, S.G. (2021) Interference with LEDGF/p75-directed Integration Enhances Transcription of HIV-1 Antisense RNA. (*In prep*)

5) Dupont, L., Bloor, S., Williamson, J.C., Cuesta, S.M., **Shah, R.**, Namaati, A., Greenwood, E.J.D., Balasubramanian, S., Sarafianos, S.G., Matheson, N.J., Lehner, P.J. (2021) Nuclear Immunosurveillance by the SMC5/6 Complex is Antagonized by HIV-1. *Cell Host & Microbe.* (PMID: 33811831)

6) **Shah, R.,** Lan, S., Puray-Chavez, M., Liu, D., Tedbury, P.R., Sarafianos, S.G. (2020) Single-cell multiplexed fluorescence imaging to visualize viral nucleic acid and proteins to monitor HIV, HTLV, HBV, HCV, ZKV, and IAV infection. *J Vis Exp.* (PMID: 33191939)

OTHER PUBLICATIONS

1) Ke, Z., Dillard, R.S., Chirkova, T., Leon, F., Stobart, C.C., Hampton, C.M., Strauss, J.D., Rajan, D., Rostad, C.A., Taylor, J.V., Yi, H., **Shah, R**., Jin, M., Hartert, T.V., Stokes Peebles Jr., R., Graham, B.S, Moore, M.L., Anderson, L.J., Wright, E.R. (2018) The Morphology and Assembly of Respiratory Syncytial Virus Revealed by Cryo-Electron Tomography. *Viruses.* 10(8), 446 (PMID: 30127286)

2) Williams, A.*, **Shah, R.***, Shayne, M., Huston, A., Krebs, M., Murray, N., Thompson, B., Doyle, K., Korotkin, J., Hyland, S., Moynihan, J., Cory-Slechta, D., Janelsins, M. (2017) Associations between Inflammatory Markers and Cognitive Function

in Breast Cancer Patients Receiving Chemotherapy. *J. Neuroimmunol.* 314: 17-23 (PMID:29128118)

3) Dayan, O., Nagarajan A., **Shah, R.**, Ben-Yona, A., Forrest, L.R., Kanner, B.I. (2017) An Extra Amino Acid Residue in Transmembrane Domain 10 of the γ-Aminobutryic Acid (GAB) Transporter GAT-1 Is Required for Efficient Ion-coupled Transport. *J Biol Chem.* 5418-5428 (PMID: 28213519)

CONFERENCE PRESENTATIONS

1) **Shah, R.,** Gallardo, C.M., Jung, Y.H., Corces, V.G., Torbett, B.E., Tedbury, P.R., Sarafianos, S.G. (2021) Multiomics profiling of HIV-infected cells reveals unique chromatin signature associated with active proviruses **[Poster presentation, Cold Spring Harbor Retroviruses (virtual)]**

2) **Shah, R.,** Jung, Y.H., Corces, V.G., Tedbury, P.R., Sarafianos, S.G. (2021) Applications of RNA- and ATAC-seq to study patterns of HIV latency reactivation **[Oral presentation, HIV Interactions in Viral Evolution (HIVE), virtual]**

3) **Shah, R.,** Jung, Y.H., Corces, V.G.., Tedbury, P.R., Sarafianos, S.G. (2020) Applications of RNA- and ATAC-seq to study patterns of HIV latency reactivation **[Oral presentation, Cold Spring Harbor Retroviruses (virtual)]**

4) **Shah, R.,** Jung, Y.H., Corces, V.G., Tedbury, P.R., Sarafianos, S.G. (2020) Applications of RNA- and ATAC- seq to study patterns of HIV latency reactivation **[Oral presentation, HIV Interactions in Viral Evolution (HIVE) (La Jolla, CA)]**

5) **Shah, R.,** Majumder, K., Bixler, B.J., Jung, Y.H., Corces, V.G., Pintel, D.J., Tedbury, P.R., Sarafianos, S.G. (2019) Long-range chromosome interactions between the HIV-1 provirus and host genome during latency **[Oral presentation, West Coast Retroviruses (Palm Springs, CA)]**

6) **Shah, R.,** Majumder, K., Ukah, O.B., Bixler, B.J., Jung, Y.H., Corces, V.G., Pintel, D.J., Tedbury, P.R., Sarafianos, S.G. (2019) Long-range chromosome interactions between the HIV-1 provirus and host genome during latency **[Oral presentation, HIV Interactions in Viral Evolution (HIVE) (Bethesda, MD)]**

ACKNOWLEDGEMENTS

I would like to first express my sincerest gratitude to my PI, Dr. Stefan Sarafianos, for being a phenomenal mentor and providing continuous support and encouragement during my tenure in his lab. Dr. Sarafianos' mentorship, guidance, and critical feedback have been invaluable to my training as a scientist, and I am very grateful for the opportunity to have conducted my doctoral research under his guidance.

I am also extremely grateful for the amazing people in the lab over the years. Dr. Phil Tedbury has been a second advisor to me, helping me closely with the conception and development of my projects and ideas. I appreciate his guidance and feedback but also enjoyed hearing his perspective on science and life. I also have to thank Dr. Karen Kirby, Dr. Ryan Slack, Dr. Shuiyun Lan and many more in the lab. It has been a pleasure learning from your expertise and getting to know you all on a personal level. The lab has grown considerably since I joined about three years ago, and I'm grateful to be in an environment with so many diverse personalities and down-to-earth people. It's also been a pleasure going through the process of graduate school with Maria and Darius who have been nothing but supportive and helpful.

I am also very grateful to my committee for their guidance and expertise over the years. Thank you, Drs. Koval, Conn, Melikian, Liang, and Corces for your time, encouragement, and invaluable feedback. I am very fortunate and humbled to be part of the BCDB program and Laney Graduate School at Emory University, which has provided me with countless opportunities throughout my graduate career.

I am very fortunate to have made so many incredible friends over the past 5 years in Atlanta who have provided me with encouragement and support. I could not have asked for better friends in my cohort (Zane, Raven, Stephanie, & Emma). It has been an honor making it through Foundations with you all and growing as people and scientists. I need to give a shout-out to Ed, Fadi, and Alex who were third-year graduate students when I started the program. They, therefore, had a few years of wisdom to impart on me, as I worked through the process. Erika and Sam have been amazing housemates and friends, also going through the rigors of graduate school, understanding the common struggle. It's been awesome having one of my oldest friends, Eric, move down to Atlanta and getting me out of the Emory bubble and helping me experience and embrace our great city of ATL. I need to give a very special shout-out to Emily L. I definitely could not have made it to this point without you. Your countless hours of feedback on my project/presentations to being my closest friend have been instrumental in keeping me motivated and helping me be the best person I could possibly be. Thank you to my friends from the UR (Sol, Hogan, Niyi, Tim, Steve, Brian, Emily S, and many more) for all of your support over the years and always having my back.

Finally, I am beyond fortunate to have such an amazing family who have always been there for me, serving as incredible role models and providing me with endless opportunities to pursue my passions. Thank you, Mom, Dad, Ryan, Grandma, and Grandpa, for pushing me to keep my eye on the prize and making this all a reality. None of this would have been possible without you.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter I Figures:

Chapter II Figures:

Chapter III Figures:

Chapter IV Figures:

Chapter V Figures:

Chapter VI Figures:

Chapter VII Figures:

LIST OF TABLES

LIST OF ABBREVIATIONS

Acquired Immunodeficiency Syndrome	AIDS
Antisense RNA	(-) RNA
Assay for transposase-accessible chromatin with sequencing	ATAC-seq
Base pair	bp
Branched DNA	bDNA
C-C chemokine receptor type 5	CCR5
Carboxy-terminal domain	CTD
Catalytically dead Cas protein	dCas
CCCTC-binding factor	CTCF
Chromatin immunoprecipitation using sequencing	ChIP-seq
Chromosome conformation capture	3C
Circulating recombinant forms	CRFs
Cleavage and polyadenylation specificity factor 6	CPSF6
Co-immunoprecipitation	Co-IP
Combination antiretroviral therapy	cART
Complementary DNA	cDNA
Cytotoxic T-lymphocytes	CTLs
Deoxynucleoside triphosphate	dNTP
Double-stranded DNA	dsDNA
Downstream-of-gene	DoGs
DRB sensitivity-inducing factor	DSIF
Effector-memory-transition	EMT
Endoplasmic reticulum	ER
Facilitates chromatin transcription	FACT
Fetal bovine serum	FBS
Fluorescence in-situ hybridization	FISH
Fluorescence-activated cell sorting	FACS
Heat shock factor 1	HSF1
Hepatitis B virus	HBV
Hepatitis C virus	HCV
Herpes simplex virus 1	HSV-1
Histone deacetylases	HDACs
Histone methyltransferases	HMTs
HIV-1 Antisense protein	ASP
HIV-1 Capsid	CA
HIV-1 Envelope glycoprotein	Env
HIV-1 Integrase	IN
HIV-1 Matrix	MA
HIV-1 Negative regulatory factor	Nef

HIV-1 Nucleocapsid	NC
HIV-1 packaging signal	Psi (φ)
HIV-1 Protease	PR
HIV-1 Reverse Transcriptase	RT
HIV-1 Surface (gp120)	SU
HIV-1 Trans-activator protein	Tat
HIV-1 Transmembrane (gp41)	TM
HIV-1 Viral infectivity factor	Vif
HIV-1 Viral protein R	Vpr
HIV-1 Viral protein U	Vpu
Hours post-infection	Hpi
Human genome assembly	hg38
Human immunodeficiency virus type 1	HIV-1
Human leukocyte antigen	HLA
Human T lymphotropic virus 1	HTLV-1
Influenza A virus	IAV
Kaposi sarcoma	KS
Kilobases	kb
Kilodalton	kDa
Knock-down	KD
Knock-out	KO
Lamina-associated genes	LADs
Latency-reversing agents	LRAs
Lens epithelium-derived growth factor	LEDGF/p75
Long terminal repeat	LTR
Major histocompatibility complex	MHC
Messenger RNA	mRNA
Minute(s)	Min
Multiplexed immunofluorescent cell-based detection of DNA, RNA, and protein	MICDDRP
Multiplicity of infection	MOI
N-terminus domain	NTD
Negative elongation factor	NELF
Next-generation sequencing	NGS
NF-kappa-B inhibitor beta	IkBa
Nuclear export signal	NES
Nuclear factor kappa-light-chain-enhancer of activated B	NFkB
Nuclear factor of activated T-cells	NFAT
Nuclear pore complex	NPC
Nucleosome	Nuc
Open-reading frame	ORF

Peripheral blood mononuclear cells	PBMCs
Phosphate-buffer saline with 0.1% Tween-20	PBST
Poly-D-lysine	PDL
Polyadenylated	poly(A)
Polycomb repressive complex-2	PRC2
Positive transcription elongation factor	P-TEFb
Positive-sense RNA	(+) RNA
Pre-integration complex	PIC
Protein kinase C	PKC
Quantitative PCR	qPCR
Quantitative reverse transcription PCR	qRT-PCR
Quantitative viral outgrowth assay	QVOA
Recurrent integration genes	RIGs
Regulator of viral protein	Rev
Rev response element	RRE
Reverse transcription complex	RTC
RNA polymerase II	RNAPII
Room Temperature	RT
Second(s)	Sec
Severe acute respiratory syndrome coronavirus 2	SARS-CoV-2
Single-cell RNA-seq	scRNA-seq
Single-stranded RNA	ssRNA
Sodium dodecyl sulfate	SDS
Speckle-associated genes	SPADs
Super-elongation complex	SEC
Super-enhancer	SE
T-cell receptor	TCR
Targeted chromatin capture	T2C
Topologically Associating Domains	TADs
Trans-activation response	TAR
Transcriptional start sites	TSS
Untranslated region	UTR
Wild-type	WT
Zika virus	ZKV

CHAPTER I: INTRODUCTION TO HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (HIV-1), HIV-1 TRANSCRIPTIONAL REGULATION, & NUCLEAR ORGANIZATION

1-A: HIV-1 Disease Pathogenesis & Viral Replication

Since the early 1980s, human immunodeficiency virus type 1 (HIV-1) has been a major global public health burden, infecting more than 75 million individuals worldwide. There is currently no effective cure or preventative vaccine against HIV-1 infection [1]. HIV targets immune cells including T cells [2, 3], macrophages [4], and dendritic cells [5], establishing permanent infection by integrating a double-stranded complementary DNA (cDNA) copy of its RNA genome into the targeted cellular chromatin [6]. Integrated viral DNA or the "provirus" persists for the lifetime of the infected cell, allowing transcriptionally active and replication-competent proviruses to propagate virion progeny. About 10 days following infection, HIV becomes detectable in the blood and then viremia grows exponentially over the next few weeks, when HIV antibodies become detectable [1]. Untreated infection results in the targeted depletion of CD4+ T cells, a cell-type important for innate and adaptative immunity. When T cell levels drop below 200 cells/mL in the blood, infection leads to the onset of Acquired Immunodeficiency Syndrome (AIDS) [7-9]. When infection progresses to AIDS, the body experiences immunological abnormalities, increasing the risk for oncological complications and opportunistic infections [10]. HIV-associated immunodeficiency increases the risk of developing Kaposi sarcoma (KS), certain lymphomas, and cervical cancer [1].

There are currently two types of HIV: type 1 (HIV-1) and type 2 (HIV-2). HIV-1 accounts for over 95% of the infections globally, whereas HIV-2 is endemic to West Africa [1, 10, 11]. HIV-2 is associated with lower pathogenicity, lower levels of replication in the host, and

reduced transmission rates [1, 12]. HIV emerged from zoonotic transmissions of simian immunodeficiency virus (SIV) from monkeys to great apes to humans. SIV from chimpanzees (CIVcpz) or gorillas (SIVgor) to humans gave rise to the four HIV-1 groups: M and N from SIVcpz, and O and P from SIVgor. The group M strains can be further classified into nine subtypes (A, B, C, D, F, G, H, J, and K), with each subtype phylogenetically equidistant from one another. Genetic recombination between these subtypes can generate further strains known as circulating recombinant forms (CRFs) [13, 14]. The HIV-1 subtype B is the most dominant circulating strain in the Americas and accounts for around 10% of global infections. Subtype C is the most prevalent (~50% total global infections). Subtype A is ~25% of total infections [1, 11].

HIV transmission occurs through contact between infected bodily fluids and abrasions or ruptures in mucosal tissue. Upon exposure to the virus, HIV virions use the host receptor C-C chemokine receptor type 5 (CCR5) or to a significantly lesser extent, the C-X-C chemokine receptor type 4 (CXCR4), for entry into the host cell [15, 16]. Once HIV enters the cell, the virus will unleash its viral core into the cytoplasm, which consists of two viral single-stranded RNA (ssRNA) genomes encapsidated by HIV structural proteins. HIV will make its way to the nucleus of the cell for integration of reverse-transcribed viral cDNA into the cellular chromatin. Successful integration of the HIV provirus can then subsequently lead to viral replication and spread. The first detection of viral RNA in the blood marks the end of the eclipse phase of HIV infection (**See Figure I.A.1**). During the first few weeks of infection, HIV begins to actively replicate at the site of initial infection and begins to spread to distant susceptible tissues and organs, with no detectable viremia or symptoms at this point. Primary or acute infection begins when virus is typically at the peak of viral RNA levels (~10^6-10^7 copies/mL) in the blood, and viremia is rampant (large populations of HIV-infected CD4+ T cells). Acute infection may be

associated with a short symptomatic phase (fever, rashes, myalgias, and lymphadenopathy), however infected individuals may also be asymptomatic at this point in the course of HIV infection [1, 11]. The immune response begins to mount a defense, dropping viral loads to steady-state levels that are referred to as the viral set point. Set point directly correlates to patient clinical outcome. Both humoral (antibody-mediated response) and cell-mediated responses (CD8+ cytotoxic T cells) target infected cells. Most of the productively infected CD4+ T cells (actively replicating HIV) are eradicated *via* activation-induced cell death (AICD), cytopathic effects (CPE), or cytotoxic T-lymphocytes- (CTL)-mediated cell killing [11]. CTLs eliminate infected cells by recognizing specific viral peptides presented by human leukocyte antigen (HLA) class I molecules on the cellular surface and then elicit effector mechanisms that cause cell killing [17]. The decline in CD4+ T cell numbers, resulting from the host immune response, mark the end of the acute phase and beginning of the chronic infection phase. During this phase, viremia is stabilized to a set point. CD4+ T cell levels continue to gradually decrease due to the death of HIV-infected cells, leading to chronic immune modulation and inflammation. If left untreated, HIV infection advances to the onset of AIDS. Fortunately, decades of scientific advancements in the field of HIV virology and drug discovery have culminated in the development of effective combination antiretroviral therapies (cART), which involve a cocktail of antiretroviral drugs that manage and minimize HIV viral loads to near undetectable levels. Since the introduction of cART in 1996, treatment has been highly effective, suppressing HIV-1 replication in infected individuals and allowing successful management of disease [10].

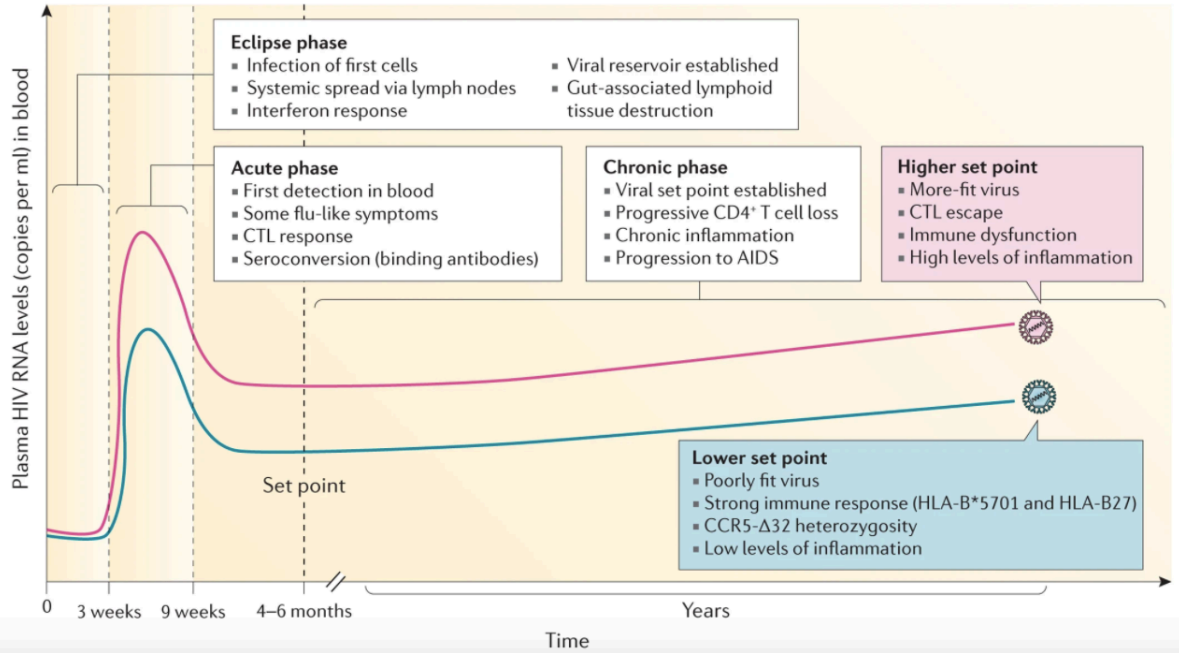


Figure I.A.1. Course of HIV Infection. HIV first infects target cells in mucosal tissues and then begins to spread through the lymphoid system during the eclipse phase of infection. The viral reservoir is established during this phase. During the acute phase, viral RNA levels become detectable after several days and then increase exponentially, reaching a set point a few weeks later. The adaptive immune response then elicits partial control. HIV infection leads to elimination of CD4+ T cells, culminating in immunodeficiency and chronic inflammation. The typical untreated HIV-infected individual progresses to death within 10 years, others astray from the mean, succumbing to viral infection more rapidly (higher set point) or display slower progression of infection (lower set point).The original copyrighted material was published by Deeks *et al.* 2015 [1]. Reprinted with permission from Springer Nature and Copyright Clearance Center. License number: 5095490407457.

HIV-1 Replication Cycle

HIV-1 replication begins when the GP120 subunit of the HIV-1 Envelope (Env) glycoprotein recognizes and binds to a CD4 molecule on a CD4+ cell, which causes a rearrangement of the variable loops (V1-V3) within GP120 [18-20]. This conformational reconfiguration of the Env glycoprotein facilities binding to either the host CCR5 or CXCR4 co-receptor, which is mediated by the V3 loop [21]. Co-receptor binding occurs following exposure and insertion of the GP41 subunit into the host cell membrane, which triggers membrane fusion

via formation of a six-helix bundle that brings the viral and host membranes into close proximity [22, 23].

Upon fusion and viral entry, the HIV-1 virion conical core is then unleashed into the cellular cytoplasm. The process of reverse transcription of the encapsulated viral genomic RNA likely begins in the cytoplasm [24, 25]. During this process, the viral core travels along microtubule tracks towards the host nucleus [26]. The resulting double-stranded DNA (dsDNA) forms a pre-integration complex (PIC) with the viral enzyme, integrase (IN), and other cellular host factors [27-29]. The PIC likely within an intact or partially intact capsid (CA) core traverses the nuclear pore complex (NPC) [30, 31] to target the introns of chromatin-accessible and transcriptionally active genes for proviral integration [6, 32, 33]. dsDNA can remain in a linear state that can be degraded [34], circularize to contain one or two long terminal repeats (LTRs) [35-37] (flanking 5' and 3' regions of the HIV-1 dsDNA genome), or integrate into cellular chromatin to continue the replication cycle. Integration is the hallmark of retroviral infection, permanently inserting the viral genome into the targeted host chromatin.

For initiation of proviral integration, IN removes two nucleotides from the 3'-ends of the blunt-ended viral dsDNA (3'-end processing), creating 5'-end overhangs [6, 38-40]. The cellular factor lens epithelium-derived growth factor (LEDFG/p75) mediates anchoring of the IN complex to the cellular chromatin [6, 41]. This facilitates a concerted strand transfer reaction, a transesterification that cuts both strands of target DNA, simultaneously joining them to the 3'-ends of the viral DNA molecule. This concerted strand transfer event targets a pair of phosphodiester bonds on the opposing strands of the target DNA across the major groove, approximately five base pairs (bp) apart in the host DNA [42]. The 5'-end overhangs are removed,

and the single-stranded gaps between the 5'-ends of the HIV-1 DNA are repaired *via* the cellular DNA repair machinery [36, 43].

Following integration of replication-competent HIV-1 dsDNA, the integrated genome (provirus) can then undergo productive proviral transcription or establish a latent or "quiescent" non-replicative state [44-46]. Mechanisms underlying HIV-1 transcriptional control and latency establishment will be further discussed in **Chapter I.E**. The cellular polymerase, RNA polymerase II (RNAPII), transcribes HIV-1 RNAs that will serve as either full-length viral genomic RNA or messenger RNAs (mRNA). Viral RNA (vRNA) can be packaged into the assembling virion or 5' capped and poly-adenylated (polyA) viral mRNA can be translated into viral proteins *via* the cellular ribosomal machinery.

The viral accessory protein regulator of viral protein expression (Rev) mediates the nuclear export of vRNA oligomerizing at the rev response element (RRE), a higher-order RNA domain found in canonical HIV-1 transcripts [47-49]. Rev bound to the RRE of viral transcripts interacts with Crm1 (Exportin 1) *via* a nuclear export signal (NES) within the C-terminus, mediating nuclear export of viral mRNA through the NPC [50]. This complex is destabilized within the cytosol by hydrolysis of Crm1-associated GTP, freeing Rev to renter the nucleus by binding to Importin-β [51].

The exported transcripts encoding Gag, Gag/Pol, Vif, Vpr, and Nef are translated in cytosolic polysomes, while Env and Vpu are translated at the rough endoplasmic reticulum (ER) [52]. Gag and Gag/Pol proteins traffic to the cellular plasma membrane *via* intracellular vesicular pathways along microtubule tracks [53, 54]. Gag monomers and dimers then assemble into detergent-resistant membrane microdomains [55, 56]. The Gag proteins undergo conformational changes that promote Gag-Gag, Matrix (MA)-membrane, and Nucleocapsid (NC)-RNA

interactions [57, 58]. Two unspliced Gag/Pol RNA transcripts are recognized by NC proteins at a packaging signal (φ) located in the 5'-end of the viral genomes and packaged within the assembling virion [59-61].

The Env protein traffics to the host membrane independently of Gag [52]. Upon arrival at the cellular membrane, Env interacts with the MA domain of Gag to promote Env incorporation into virions [62, 63]. The accessory protein, viral protein R (Vpr), viral infectivity factor (Vif), and negative regulatory factor (Nef) are also packaged [52]. As Gag continues to accumulate at assembly sites, immature virions bud from the host membrane using host ESCRT machinery. The ESCRT machinery plays vital roles in endosomal sorting, cellular abscission, and viral budding [64, 65]. HIV-1 protease (PR), which has also been incorporated into the immature virion, proteolyzes the Gag and Gag/Pol proteins at up to ten different positions to produce fully processed matrix (MA), capsid (CA), NC, p6, PR, Reverse Transcriptase (RT), and IN proteins [66, 67]. After PR cleavage, the virion is rearranged to create a mature infectious virion with two HIV-1 ssRNA genomes encapsidated by a fullerene ring of CA protein and enveloped by a host-derived lipid bilayer membrane [52].

Visual depiction of the HIV-1 life cycle is shown in **Figure I.A.2.** In addition, the Figure highlights critical host cellular factors that antagonize viral proteins, inhibiting viral replication (red boxes), as well as pharmacological inhibitors and their respective HIV-1 target proteins (green boxes). In the work presented in this dissertation, I am most focused on the steps of HIV-1 infection involved in retroviral integration (further discussed in **Chapter I.4**) and proviral transcriptional control (further discussed in **Chapter I.5**). I seek to improve our understanding of how HIV-1 integration site selection influences proviral transcriptional competence and the relationship between nuclear organization and HIV-1 transcription. In addition, I am interested in

understanding to what extent HIV-1 integration can alter chromatin architecture and how HIV-1 may perturb local cellular transcriptional patterns. These studies are presented in more detail in **Chapter VI**.

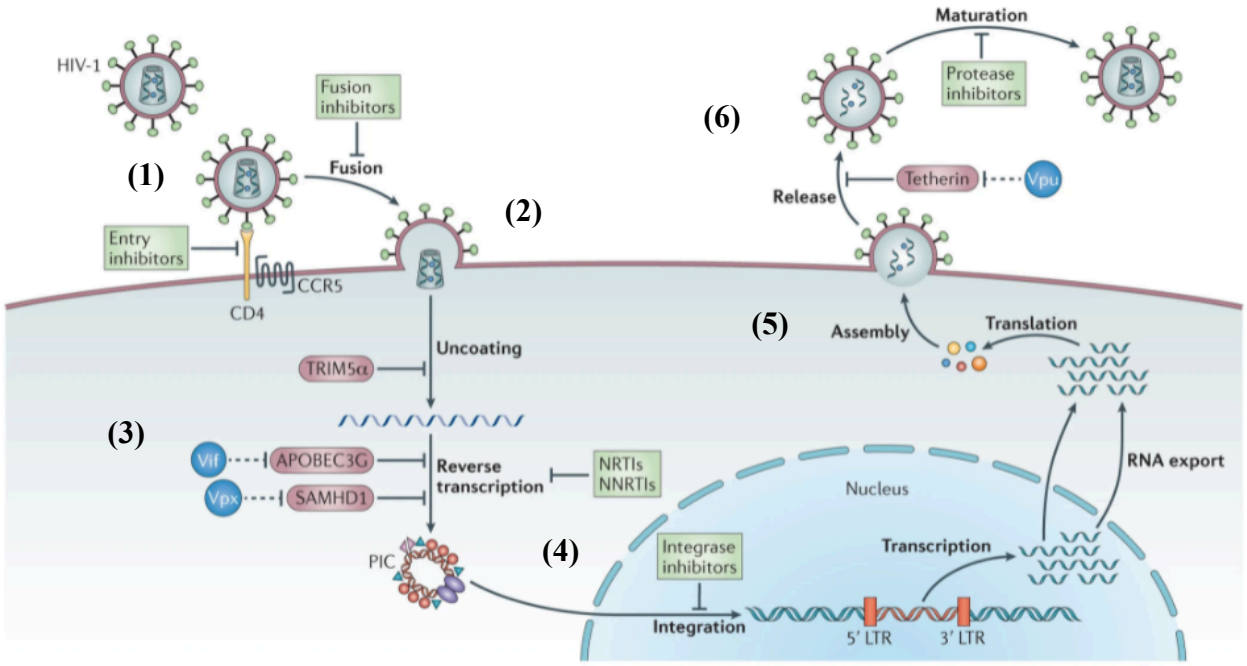


Figure I.A.2. Schematic of HIV-1 life cycle highlighting critical host restriction factors and classes of antiretroviral small molecules. This figure highlights the main steps in the HIV-1 replication cycle. 1) Binding of the HIV-1 virion to the CD4 receptor and co-receptors. 2) Fusion with the host membrane. 3) Viral capsid (CA) encapsulating the vRNA genome may begin to uncoat in the cytoplasm. Recent evidence suggests uncoating may not take place until nuclear entry of the viral core [30, 31]. Reverse transcription, however, likely initiates in the cytoplasm, synthesizing double-stranded viral cDNA from the HIV-1 ssRNA template. 4) Formation of the pre-integration complex (PIC), consisting of HIV-1 integrase, viral dsDNA, and cellular factors that mediate translocation of the complex into the nucleus. Following nuclear entry, intact viral DNA can be integrated into the cellular chromatin, subsequently leading to proviral transcription and translation of new viral proteins. 5) Viral proteins traffic to the cell surface to assemble into immature viral particles. 6) The newly assembled virions bud off and are released. The viral particle undergoes maturation, as protease cleaves the structural polyprotein to form mature Gag proteins, resulting in the formation of infectious virions. The major families of antiretroviral compounds and the viral processes that they inhibit are shown in green (further discussed in **Chapter I.4**). Also depicted are critical HIV restriction factors (tripartite motif-containing 5α (TRIM5α), APOBEC3G, SAMHD1, and tetherin, which are shown in red. The corresponding viral antagonists of the

cellular restriction factors (Vif, Vpx, and Vpu) are shown in blue. Vpx is an accessory protein found in HIV-2, which plays a role in increasing the rate of reverse transcription in macrophages [68]. The original copyrighted material was reproduced from Barré-Sinoussi *et al.* 2013 [10]. Reprinted with permission from Springer Nature and Copyright Clearance Center. License: 5097830461838.

1-B: HIV-1 Genome, Virion Organization, & Viral Proteins

HIV is grouped to the genus *Lentivirus* within the Retroviridae family [69] of viruses. HIV-1 virions are about 100 nanometers (nm) in diameter with a lipid envelope embedded with trimeric transmembrane glycoproteins [52]. The HIV-1 conical capsid (CA) core is composed of 250 CA hexamers and 12 CA pentamers[70-72]. The CA core is further enveloped within the lipid bilayer. Viral dsDNA is approximately 9.7 kilobases (kb) in length. A diagram of an assembled and mature HIV-1 virion is shown in **Figure I.B.1.**

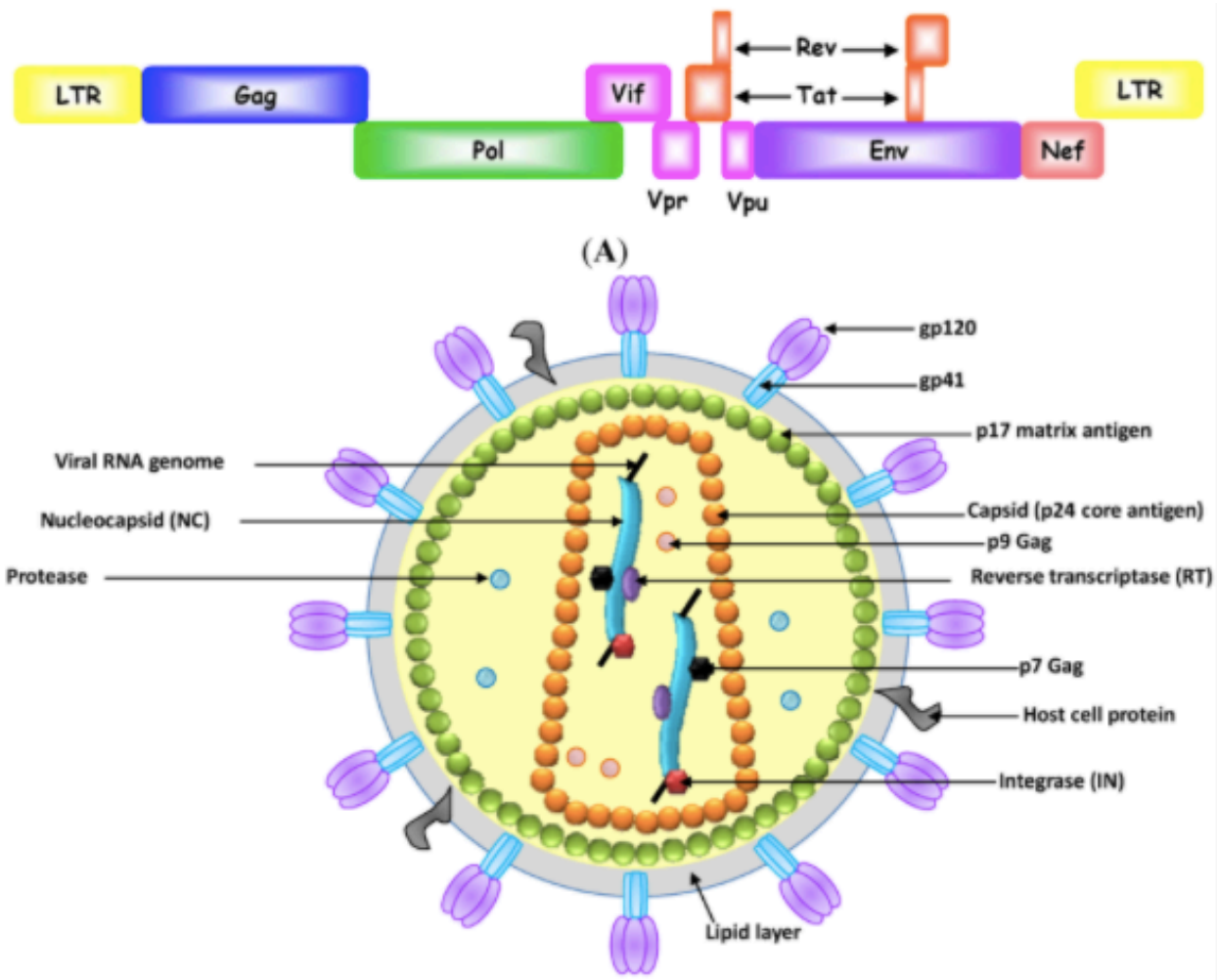


Figure I.B.1 HIV-1 Virion. The viral genome is depicted above with an assembled and mature HIV-1 particle with associated structural proteins. The original copyrighted material was published by Shum *et*

al. 2013 [73] in *Pharmaceuticals.* This is an open access article distributed under the terms of the Creative Commons CC BY license.

The HIV-1 genome is flanked by two identical long terminal repeats (LTRs). Regulatory elements within the LTRs interact with *cis*-acting inducible transcription factors to direct assembly of stable transcription complexes to drive processive proviral transcription *via* RNAPII [74]. The HIV-1 genome encodes nine open reading frames (ORFs) in the positive-sense (+) orientation (coding strand of viral genome). Three of these ORFs encode the Gag, Pol, and Env polyproteins, which are proteolyzed into individual proteins. The four Gag proteins, MA, CA, NC, and p6, and the two Env proteins, surface or gp120 (SU) and transmembrane or gp41 (TM), are the core structural components of the HIV-1 virion. The three Pol proteins, PR, RT, and IN are the essential viral enzymes that are also encased within the viral particle. The Pol protein is generated from the Gag-Pol polyprotein that is produced *via* a ribosomal frameshift that occurs approximately 5% of the time during translation [75]. The HIV-1 genome encodes two regulatory proteins (Tat & Rev) and four accessory proteins. Three of these accessory proteins are Vif, Vpr, Nef, which are also found within the viral particle. An additional protein found in HIV-1 but not HIV-2 is Viral protein U (Vpu), which indirectly plays a role in virion assembly Rev and *trans*-activator protein (Tat) provide important gene regulatory functions [76, 77]. A detailed schematic of the HIV-1 genome can be found in **Figure I.B.2.**

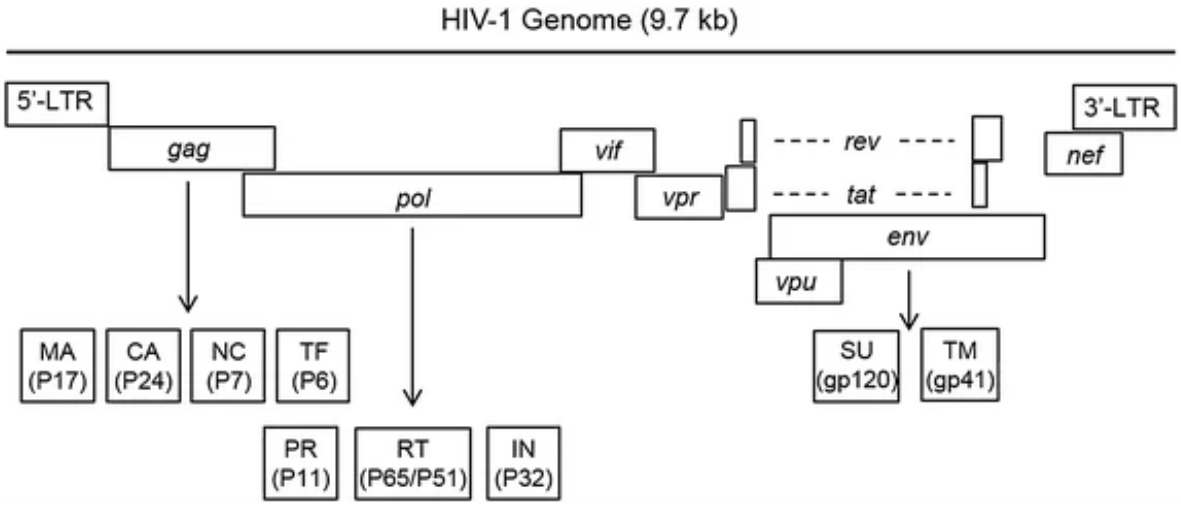


Figure I.B.2. Schematic of HIV-1 genome. The HIV-1 genome is about 9.7 kb in length. Each of the viral genes are depicted based on their relative orientation in the HIV-1 genome. Arrows indicate cleaved protein products, generated *via* polyprotein processing. Dashed lines represent RNA splicing. The number in parentheses is the molecular weight of each protein. The genomic elements and genes consist of LTR (long terminal repeats) flanking the 5' & 3' ends of the viral genome, *Gag, MA* (matrix), *CA* (capsid domain), *NC* (nucleocapsid), *TF* (trans-frame protein), *Pol* (polymerase), *PR* (protease), *RT* (reverse transcriptase), *IN* (integrae), *Env* (envelope glycoprotein), *SU* (surface membrane protein), *TM* (transmembrane protein), *Vif, Vpr, Vpu, Nef, Rev,* and *Tat*. This Figure is reproduced with permission from Nkeze *et al.* 2015 [78]. This is an open access article distributed under the terms of the Creative Commons CC BY license. This work was originally published in *Cell Bioscience.*

The HIV-1 genome also has an antisense (-) ORF (template strand of HIV-1 genome) encoding the putative "antisense protein" (ASP). The ASP gene overlaps the RRE and the Env glycoprotein. The ASP protein is a highly hydrophobic protein of about 190 amino acids in length. To date, the biological function(s) of HIV-1 ASP has not been established [79]. In **Chapter IV**, we demonstrate applications of single-molecule RNA FISH to track expression of HIV-1 (-) RNA during viral infection. We discuss studies to characterize the composition of these HIV-1 (-) transcripts and mechanistically define the functional role of HIV-1 (-) transcription in influencing viral replication and pathogenesis.

HIV-1 Structural Proteins:

Matrix

MA is the N-terminal protein of the Gag polyprotein and is important for trafficking Gag and Gag-Pol precursor polyproteins to the plasma membrane for virus assembly [80]. In the mature virion particle, the 132 amino acid residue MA protein lines the inner surface of the viral membrane, stabilizing the formation of the assembling virion. MA consists of an N-terminal myristate group and basic residues located within the first 50 amino acids of the protein that allow efficient membrane-targeting. Trimerization of MA allows the three myristate groups to embed themselves into the lipid bilayer [76]. MA may also play a role in facilitating incorporation of the glycoproteins into the assembling virion *via* interactions with the cytoplasmic tails of Env, however, this function of MA is not conclusive in the field [80, 81].

Capsid

CA is the second protein in the Gag polyprotein and forms the core of the virus particle, with ~2000 molecules per virion [76]. The C-terminal domain primarily functions in assembly and is important for CA dimerization and Gag oligomerization [82]. CA relies on many virus-host factor interactions throughout the course of infection. The HIV-1 CA core consists of ~250 hexamers and ~12 pentamers assembled from CA monomers, encapsulating the viral genome. Once the viral core is present in the cytoplasm following viral entry, it traffics along the microtubule network to the nuclear envelope prior to nuclear translocation through the NPC [83]. The interplay between CA and host co-factors such as inositol hexakisphosphate (IP6) [84, 85], Cyclophilin A (CypA) [86-88], restriction factors such as TRIM5α [89], and small molecules such as PF74 [90-92]

impact reverse transcription kinetics and transport of the PIC to the nucleus for HIV-1 integration [93]. CA is therefore a major structural protein involved in several aspects of the viral life cycle. Key questions remain regarding the molecular mechanisms regulating CA disassembly and CA-mediated nuclear transport of the HIV-1 PIC, however, cutting-edge innovations in live-cell microscopy and cryo-electron tomography (cryo-ET) are unraveling critical insight into CA's diverse role in HIV-1 replication [30, 31, 94].

Nucleocapsid

NC is the third protein within the Gag polyprotein and coats the vRNA genome within the CA conical core. The primary function of NC is to bind to the packaging signal within the vRNA genome, recruiting full-length vRNAs into the assembling viral particle. The core encapsulation or packaging signal within the RNA genome, Ψ, is composed of RNA hairpins located around the major splice donor site and translation initiation site [60, 61, 95]. NC is a basic protein that binds ssRNA, leading to encapsidation of the vRNA and mitigating nuclease-mediated degradation [76].

p6

p6 is comprised of 51 amino acid residues of the C-terminal of Gag and is important for incorporating Vpr into the virion during assembly. p6 is also implicated in helping mediate efficient particle release from the cellular membrane by recruiting ALIX and TSG101 prior to viral budding mediated by the ESCRT machinery [96].

Reverse Transcriptase

HIV-1 RT is comprised of two subunits: a 560-residue p66 subunit with polymerase and RNase H activity and a 440-amino acid p51 subunit that helps stabilize the reverse transcription

complex (RTC) [97-98]. Reverse transcription initiates from the 3' end of a $tRNA_3^{Lys}$ primer annealed to the primer binding site (PBS) near the 5' end of the genomic RNA [99]. The 3'-end of the primer is positioned at the primer-binding site (P-site), and the next deoxyribonucleotide (dNTP) to be incorporated into the nascent RT product binds at the polymerase active site (nucleotide-binding site (N-site)) [100-102]. At the polymerase active site, the α-phosphate group at the 5'-position of the incoming dNTP forms a phosphodiester bond with the hydroxyl group located at the 3'-position of the ribonucleotide terminating the 3'-end of the primer [101]. $tRNA^{Lys3}$ packages inside viral particles during assembly [99]. The kinetics of reverse transcription differ during the initiation and deoxynucleoside triphosphate (dNTP) elongation phases, becoming highly processive during elongation. Following tRNA-primed initiation, reverse transcription involves two DNA strand transfer reactions that are catalyzed by RT, leading to the synthesis of double-stranded cDNA from the ssRNA template [103, 104 101, 102]. For several years, the general consensus in the field of retrovirology was that reverse transcription was completed in the cytoplasm prior to nuclear entry, as the CA core begins to disassemble [76]. Recent live-cell imaging studies and high-resolution cryo-ET suggest that reverse transcription may actually complete in the nucleus within an intact or partially intact viral core [30, 31]. Integrative approaches utilizing live-cell imaging, structural biology, and computational modeling are improving elucidation of the process of reverse transcription during the viral life cycle with high spatiotemporal resolution, answering critical questions related to RT kinetics, RTC cellular compartmentalization, and the physiological molecular context (intact CA *vs.* disassembled CA) of the reaction in the cell [105, 106].

Integrase

Following reverse transcription of HIV-1 cDNA, IN catalyzes a series of reactions to integrate the viral genome in the cellular chromatin [29]. IN first removes two 3' nucleotides from each strand of the linear viral cDNA, leaving overhanging CA_{OH} ends. The CA dinucleotide is conserved at the ends of multiple retrotransposons. In the second step of the reaction, the processed 3' ends are covalently joined to the 5' ends of the target DNA. Retroviral IN comprises three domains: 1) An N-terminal domain (NTD) containing the Zn^{2+}-binding HHCC motif, 2) a catalytic core domain (CCD) containing the active site, and 3) a positively charged C-terminal domain (CTD). All three IN domains have been implicated in multimerization and DNA-binding [42, 107]. IN forms a tetrameric complex in its functionally relevant state [108]. Lentiviral INs have a tight interaction with the ubiquitous chromatin-associated protein, lens epithelium-derived growth factor (LEDGF/p75). The cellular functions of LEDGF are not well characterized, however, LEDGF is an important cellular factor involved in retroviral integration [42]. In **Chapter I.4,** we further delve into host factor-IN interactions that influence HIV-1 integration selection that affect downstream proviral transcriptional competency.

Protease

HIV virions budding from the cellular membrane release immature viral particles that are noninfectious. HIV-1 maturation requires cleavage of Gag and Gag-Pol polyproteins by PR. Cleavage triggers conformational rearrangements, leading to generation of the conical CA core encapsulating the viral ssRNA genomes. PR cleaves at several polyprotein sites to produce the final MA, CA, NC, and p6 proteins from Gag, as well as PR, RT, and IN proteins from Pol [76, 109].

Surface Env glycoproteins

HIV-1 Env (gp160) is a heavily glycosylated, type I membrane protein that trimerizes in the ER prior to being cleaved by cellular furin-like proteases to generate surface (SU, gp120) and transmembrane (TM, gp41) proteins [110]. Viral entry begins with binding of the SU glycoprotein, located at the viral membrane surface, to specific cell surface receptors. The major receptor for HIV-1 is CD4, an immunoglobulin (Ig)-like protein expressed on the surface of a subset of T cells and primary macrophages. Critical co-receptors involved in entry include a group of chemokine receptors (family of seven transmembrane G protein-coupled receptors). CXCR4/fusion was the first co-receptor identified, which permits entry of T-cell tropic viruses. CCR5 is a major co-receptor for macrophage-tropic viruses [76, 111]. Binding of CD4 to SU causes conformational changes in Env that facilitate co-receptor binding and subsequent viral fusion and entry. The variable V3 loop of SU is an important determinant of viral tropism [19, 20].

The primary function of the 345 amino acid residue viral transmembrane protein, TM, is to mediate fusion between the viral and cellular membranes following receptor binding. A hydrophobic N-terminal glycine-rich fusion peptide initiates fusion. Env is the major HIV-1 target for neutralizing antibodies and thus an important candidate for targeted vaccine development [112].

HIV-1 Regulatory & Accessory Proteins:

Trans-activator protein (Tat)

Tat is an essential HIV-1 regulatory protein that is expressed early in infection from multiply spliced viral mRNAs. Tat binds to the TAR RNA element, significantly increasing processive

transcription required for synthesis of full-length HIV-1 RNA transcripts for packaging and replication [113]. The role of Tat in transcriptional elongation at the HIV-1 5' LTR promoter may play an important role in transitioning the provirus from latency or to an activated state [114]. Tat function appears to be similar to cellular activator proteins, except Tat acts directly on vRNA rather than directly on a DNA domain [115]. The interactions at the Tat/TAR axis will be discussed in more detail in **Chapter I-E-B.** Proviral transcriptional activation leads to an increased number of accessible TAR RNA molecules. This results in an accumulation of Tat at the HIV-1 5' LTR promoter, perhaps, triggering a positive feedback loop that facilitates maximal HIV-1 transcriptional output.

Rev

The first HIV-1 mRNAs that are produced are mostly doubly spliced and encode for the Tat, Nef, and Rev proteins following proviral transcriptional activation. When other structural factors are needed for the assembly of infectious virions, singly spliced and unspliced transcripts are transported to the cytoplasm for translation and packaging into the viral particle. The viral regulatory factor, Rev, exports HIV-1 mRNA into the cytoplasm. Rev contains a leucine-rich nuclear export signal (NES) that allows it to shuttle between the nucleus and cytoplasm, interacting with nucleoporins at the nuclear pore [76, 116, 117]. HIV-1 mRNA also consists of another highly structured, *cis*-acting domain known as the Rev response element (RRE). The RRE is ~350 nucleotide sequence, located in the *Env* coding region of the viral genome, present in all canonically spliced and unspliced HIV-1 mRNA transcripts. The RRE serves as a scaffold for Rev oligomerization, which forms a complex that mediates export of HIV-1 mRNA from the nucleus to the cytoplasm, facilitating translation of viral proteins. Rev may also play an

important role in the splicing pathways of viral transcripts. Rev can directly inhibit splicing by preventing entry of additional small nuclear ribonucleoproteins (snRNPs), RNA-protein complexes that form the spliceosome machinery [118].

Vpr

The viral accessory protein, Vpr, is 96-amino acids in length. In collaboration with the Lehner group at the University of Cambridge, we recently showed that Vpr plays a role in counteracting host suppression of transcription from unintegrated lentiviral cDNA by mediating proteasomal degradation of the Smc5/6 protein complex [119]. We further detail this finding in **Chapter III**. Vpr has also previously been shown to interact with the cellular transcription factors, TFIIB and Sp1 [120, 121] and interacts with the coactivator CREB binding protein (CBP). This directly up-regulates proviral transcription [122]. For a small protein, there are numerous biological functions attributed to Vpr other than enhancement of viral gene expression [119, 123] including manipulation of the DNA-damage response [124, 125], induction of G2/M cell cycle arrest [126], and facilitating nuclear import of the HIV-1 genome [127].

Vpu

Vpu is an accessory protein that plays a role in mediating Env trafficking to the cellular surface during assembly of viral particles. Newly synthesized Env glycoproteins (gp160), which are cleaved into SU (gp120) and TM (gp41), may be sequestered in the ER through interactions with CD4 molecules. Vpu promotes the degradation of these CD4 molecules, freeing Env protein. Vpu is 81-residues and an integral membrane protein with a hydrophobic membrane-spanning domain at the N-terminus and a C-terminal cytoplasmic tail [128, 129]. Vpu has also been shown to stimulate virion release [76].

Nef

Nef is a 206 amino acid protein with an NTD. Similar to Vpu, Nef reduces cellular levels of CD4 *via* lysosome-mediated degradation, preventing sequestration of Env. Nef may therefore enhance Env incorporation into virions, promoting particle assembly. Nef can also downregulate expression of major histocompatibility complex (MHC) class I molecules, providing an immunomodulatory effect that protects infected cells from CTL-mediated elimination. About 70 Nef molecules incorporate into a virion [130, 131]. HIV-1 Nef also promotes infection by excluding the host factor SERINC5 from virion incorporation, redirecting SERINC5 to a Rab7-positive endosomal compartment [132]

Vif

Vif is a 192-residue protein that has multiple functions in the replication of productive infectious mature virions. Vif has been shown to interact with cellular co-transcription factor (CBF- β) to recruit CRL5, a ubiquitin ligase complex, to bind and degrade the host restriction enzymes APOBEC3D (A3D), APOBEC3F (A3F), APOBEC3G (A3G), and APOBEC3H (A3H). The APOBEC family of cellular polynucleotide cytidine deaminases are potent inhibitors of HIV-1 replication, causing excessive cytidine (C) to uridine (U) editing of negative sense reverse transcripts in newly infected cells. APOBEC enzymes can also inhibit viral DNA synthesis by impeding the translocation of RT along template RNA [133, 134]. Vif has also been shown to induce G2/M cell cycle arrest through degradation of PPP2R5s [135, 136].

1-C: Current Antiviral Therapies

Reverse Transcriptase (RT) Inhibitors

The remarkable progress in cART has transformed HIV infection from a fatal to a manageable chronic disease with marginal to no impedance to life expectancy. The first major advancement in HIV therapy was introduced in 1987, when a clinical trial showed that azidothymidine (AZT; also known as zidovudine) decreased mortality and onset of opportunistic infections in patients with AIDS [10, 137]. AZT inhibited the reverse transcription step of the HIV life cycle[138]. Virus resistance to AZT treatment quickly emerged, making it imperative to develop more compounds to target and inhibit HIV-1 replication. AZT is a nucleoside/nucleotide reverse transcriptase inhibitor (NRTI). NRTIs are administered as prodrugs in the patient, which require host cellular entry and phosphorylation by cellular kinases before they can function as inhibitors of HIV-1 reverse transcription [138, 139]. Their structure mimics canonical nucleosides incorporated by HIV-1 RT. The lack of 3'-5'phosphodiester bond between NRTIs and the incoming 5'-nucleoside triphosphates culminates in early termination of the nascent viral cDNA chain. This chain termination can occur during RNA-dependent DNA or DNA-dependent DNA synthesis, effectively halting synthesis of viral cDNA [140, 141]. FDA-approved NRTIs including AZT, Tenofovir disoprovil fumarate (TDF), didanosine (ddI), emtricitabine (FTC), lamivudine (3TC), stavudine (d4T), and zalcitabine (ddC) have proved to be effective treatment options against HIV-1 infection. However, viral evolution culminating in mutations (examples of early RT mutations include K65R, Y115F, L74V, Q151M) within RT catalytic domains conferred antiviral resistance. This thus allowed HIV-1 strains to emerge with low susceptibility to once potent NRTI treatments [142]. HIV-1 resistance to NRTIs can be attributed to NRTI discrimination (worsening NRTI incorporation efficiency relative to the natural dNTP pool) and NRTI excision

(nucleoside analog gets excised from nascent vRNA chain allowing reverse transcription to continue) [142].

Novel second-class RT inhibitors were more potent and long-acting small molecules that were developed with hopes of reducing the escape potential of viruses. 4'-Ethynyl-2-fluuoro-2'-deoxyadeonosine (EFdA, MK-8591, or islatravir) is a very promising RT inhibitor in phase III clinical trials [143-145]. EFdA has an extremely impressive biostability conferred by a 2-fluoro group, which prevents metabolism and degradation by adenosine deaminase, which targets adenosine-based nucleoside analogs, as well as adenosines. In addition, EFdA also contains a 4'-ethynyl (4'E), and 3'-OH [146, 147]. The retention of the 3'-OH is novel amongst current HIV-1 RT inhibitors. The 4'-E group stabilizes the molecule into a conserved hydrophobic pocket at the polymerization active site [147]. EFdA inhibits reverse transcription primarily by blocking RT translocation. EFdA is therefore known as a nucleoside reverse transcriptase translocation inhibitor (NRTTI). EFdA blocks translocation through immediate chain termination (ICT) and delayed chain termination (DCT). During ICT, RT is arrested immediately following incorporation of EFDA-monophosphate (EFdA-MP), and the 3'-end of the primer remains bound and locked at the pre-translocation site (N-site) [144, 145]. During DCT, EFdA-MP is incorporated into the 3'-end of the primer and translocates to the P-site, thus allowing incorporation of an additional single nucleotide before chain termination. The Sarafianos group has contributed extensive biochemical, structural, and cell-based characterization of EFdA to define these mechanisms of action [144-147]. Current FDA-approved NNRTIs include etravirine, delavirdine, efavirenz, and nevirapine [141, 142].

Integrase Inhibitors

A class of compounds "coined" integrase strand transfer inhibitors (InSTIs) target the strand transfer reaction of HIV-1 integration [38, 40]. Functionally, InSTIs interfere with binding of IN to viral cDNA ends, interfere with IN oligomerization, affect 3'-P activity, and can inhibit protein-protein interactions between IN and cellular cofactors [29]. Raltegravir (RAL), MK-0518, was FDA-approved in 2007, and a current InSTI commonly prescribed, robustly inhibiting HIV-1 proviral integration [148]. The common mechanism of action of InSTIs is binding to a specific complex between integrase and the viral DNA and chelation of two essential Mg^{2+} ion cofactors in the integrase catalytic site [39], thus abrogating integration. The emergence of resistance mutations drove the development of second-generation of InSTIs, including elvitegravir (EVG) [148] and dolutegravir (DTG) [149], which are both FDA-approved for treatment against HIV/AIDS.

Another class of integrase interacting compounds includes allosteric INIs (ALLINIs). ALLINIs bind to a site distinct from the active site but still manage to inhibit critical protein-protein interactions of HIV-1 IN and cellular cofactors [150]. When ALLINIs bind to an allosteric site, they induce major structural changes to the catalytic site [151]. Since ALLINIs target a distinct site than InSTIs, the two classes of IN inhibitors experience a different resistance profile [152]. The majority of ALLINIs target the IN-LEDGF interaction [153].

Protease Inhibitors

The HIV-1 protease is an essential viral factor involved in virion maturation, cleaving the viral Gag and Gag-pol polyproteins to generate the mature proteins and CA conical core [109]. Due to the relatively small size of HIV-1 PR (11 kilodalton (kDa)), it was initially expected that resistance to PR inhibitors would be rare. However, resistance quickly emerged with

polymorphisms found in 49 of the 99 codons of the PR gene [154]. FDA-approved PR inhibitors include amprenavir (APV, Agenerase), atazanavir (ATZ, Reyataz), darunavir (TMC114, Prezista), fosamprenavir (Lexiva), indinavir (IDV, Crixivan), lopinavir (LPV), nelfinavir (NFV, Viracept), ritonavir (RTV, Norvir), saquinavir (SQV, Fortovase/Invirase), and tipranavir (TPV, Aptivus) [141].

Capsid-Targeting Compounds

In the Sarafianos group, we have considerable interest in the development of potent CA inhibitors that block viral replication. CA inhibitors can inhibit viral replication *via* two distinct mechanisms of action: 1) disruption of CA-CA interactions, altering overall core stability and 2) compete against the nucleoporin, Nup153, and cleavage and polyadenylation specificity factor 6 (CPSF6), both cellular host factors that are required for the nuclear import of the HIV-1 PICs. CPSF6 also plays a role in integration site selection [91], which will be further discussed in **Chapter I.4**. A promising CA-targeting molecule, GS-6207, inhibits HIV-1 by stabilizing and preventing functional disassembly of the CA shell in infected cells. GS-6207 tightly binds two CA subunits and promotes distal intra- and inter-hexamer interactions that stabilize the curved CA lattice. GS-6207 also interferes with CA-host factor interactions (Nup153 & CPSF6) [155].

HIV-1 Vaccine Development

One of the many challenges in developing a preventive HIV-1 vaccine is the lack of a known correlate of protection. Despite a strong immune response against HIV infection, infected individuals do not eliminate the virus or the determinantal effects that are associated, as in other infectious diseases [1]. The development of protective and neutralizing antibodies through

vaccination is challenging, attributed to the significant sequence diversity in Env proteins and differential glycosylation patterns seen across HIV-1 glycoproteins. The RV144 HIV vaccine efficacy trial showed evidence of protection. This vaccine regimen showed a 31% reduction in HIV acquisition among 16,000 Thai men and women over a 3.5-year period. Interestingly, analysis of this trial suggested that non-neutralizing antibodies were playing a significant role in protection with high titers of IgA antibodies correlating with reduced vaccine efficacy. Perhaps, infection was inhibited through antibody-dependent cell-mediated cytotoxicity, rather than direct neutralization [1, 156].

1-D: HIV-1 Proviral Integration

1-D-a: Cellular & Viral Factors that Compose the Pre-Integration Complex

The intasome, a higher-order nucleoprotein complex composed of IN and the ends of viral dsDNA, mediates proviral integration [6, 108, 157]. The interactions between the HIV-1 intasome and cellular host proteins, as well as other viral factors are critical for efficient nuclear entry of the PIC and integration site selection. The CA conical core, encapsulating the PIC, interacts with several nucleoporins, including Nup358 and Nup153, and other cellular factors such as cleavage and polyadenylation specificity factor 6 (CPSF6). These CA-host interactions facilitate nuclear entry of the viral core [33, 158]. CPSF6 frees the core from the NPC and mediates progression of the PIC beyond the nuclear periphery for proviral integration [159].

The HIV-1 PIC also interacts with LEDGF/p75, which directs integration into the interior of gene bodies, favoring HIV-1 integration into the introns of transcriptionally active and chromatin-accessible genes [160, 161]. LEDGF/p75, a chromatin-associated cellular protein, was first identified as a tight interactor of lentiviral INs, playing an important role in HIV-1 replication and stimulating IN catalytic efficiency *in vitro* [162, 163]. LEDGF tethers HIV-1 intasomes to chromatin, protecting the viral genome from degradation and thus also strongly influencing the genome-wide pattern of HIV integration [29, 160]. Depletion of LEDGF from cells impairs viral infectivity [161]. A schematic of the interactions between the HIV-1 PIC, LEDGF, and cellular chromatin are depicted in **Figure I.D.1.** The HIV-1 PIC interacts with the integrase-binding domain (IBD) of LEDGF, which remains tethered to cellular chromatin. While the association between IN-LEDGF interactions mediating integration efficiency and integration site selection have been established, the role nuclear architecture and other chromatin factors play in influencing integration site selection and proviral transcriptional capacity have not been well

elucidated. Questions remain regarding what DNA domains or other chromatin-associated proteins are perhaps binding at the PWWP domain of LEDGF, mediating the anchoring of the intasome complex to cellular chromatin.

A putative important interactor with the IN-LEDGF/p75 complex influencing integration efficiency and site selection is the protein complex Facilitates Chromatin Transcription (FACT), a heterodimer composed of SSRP1 and Spt16 subunits [41, 164]. FACT has been biochemically shown to interact with the PWWP domain of LEDGF and the LEDGF-IN complex *via* co-immunoprecipitation (co-IP) [164]. A proteomics screen of the HIV-1 IN interactome also identified FACT as a strong interactor with the intasome complex *in cellulo* following *de novo* infection [41]. Further studies using *in-vitro* concerted integration assays, demonstrated that the FACT complex improves integration efficiency at H3/H4 histone tetramers of reconstituted chromatin by increasing chromatin accessibility by promoting nucleosome disassembly [41]. Recent atomic cryo-EM structures of FACT complexed with subnucleosomes revealed that FACT engages in extensive interactions with nucleosomal DNA [165], promoting nucleosome disassembly during gene transcription and reassembly during DNA repair [166]. Pharmacological inhibition of FACT-mediated nucleosome assembly improved HIV-1 integration efficiency *in cellulo*, further demonstrating that chromatin structure plays a critical role in proviral integration and the HIV-1 life cycle [41]. Early studies highlighting this potential important HIV-host interaction warrants further mechanistic investigation of FACT-mediated chromatin remodeling and HIV-1 integration. Understanding the interaction network between HIV-1 IN and nuclear proteins can provide insight into cellular factors influencing HIV-1 transcriptional control and latency establishment.

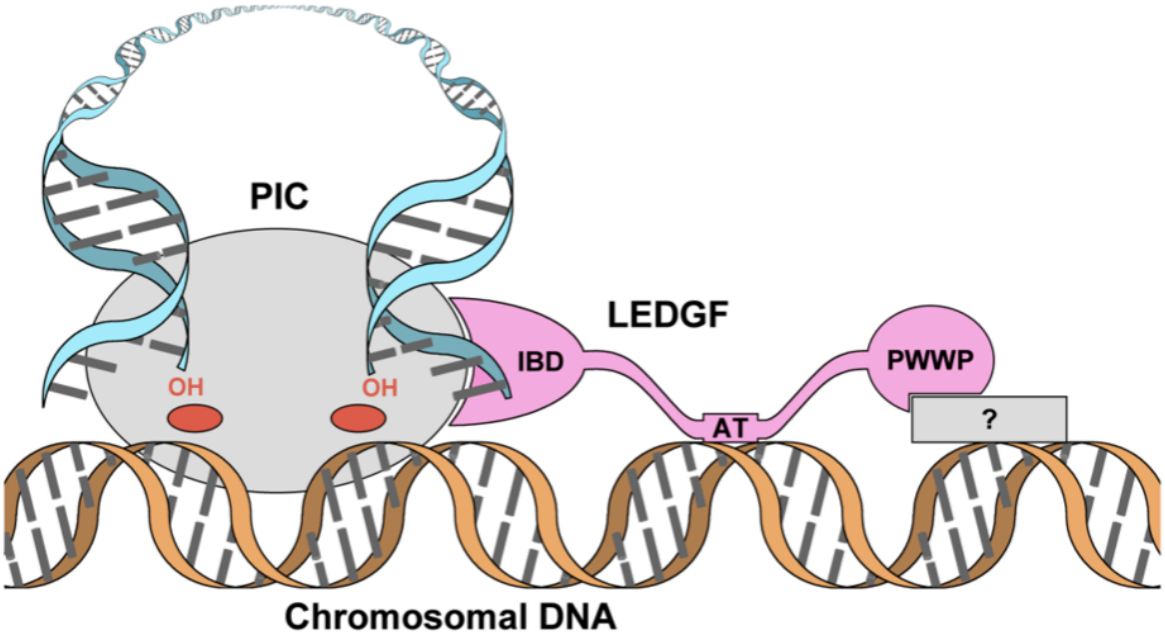


Figure I.D.1. Model of pre-integration complex (PIC) directed to chromosomal DNA for integration mediated *via* direct interaction with the host factor, LEDGF. LEDGF (pink) interacts with the PIC at the C-terminal integrase-binding domain (IBD) and with an unidentified chromatin factor at the N-terminal PWWP domain (gray box). Interactions at the PWWP interface with putative chromatin factors may influence HIV-1 integration site selection and local chromatin structure in the vicinity of integrated vDNA [41, 164]. Figure is reused from Hare *et al.* 2009 [42] with permission under an open access Creative Commons CC BY license.

1-D-b: Determinants of Integration Site Selection

The first genome-wide studies of HIV-1 integration site selection revealed that the retrovirus has a preference for integrating into gene-dense regions with highly expressed genes [167]. The advent of epigenomic profiling further revealed that HIV-1 integrates into regions with active chromatin marks such as H3K4me1, H3K36me3, H4K15a, H3K27ac and disfavors heterochromatin markers such as H3Kme3, H3K27me3, and lamina-associated domains (LADs) [168, 169]. Comparisons of HIV-1 integration frequency into specific genes across multiple studies revealed the presence of recurrent integration genes (RIGs) or integration "hotspots" [170].

Fluorescence-based cytological imaging of RIGs and HIV-1 proviruses in activated CD4+ T cells revealed that HIV may be targeting genes at the nuclear periphery for integration, suggesting HIV-1 may be targeting specific nuclear structures [169-171]. The idea of HIV targeting the nuclear periphery is contentious with multiple studies also demonstrating that PICs accumulate throughout the nucleus (pan-nuclear distribution) prior to integration [159, 172-174]. Activation of CD4+ T cells leads to significant rearrangements in nuclear architecture and cellular cytoskeleton networks [175]. RIG distribution in resting T cells was disperse throughout the nucleus, however, activation resulted in RIG relocalization closer to the nuclear periphery [170]. The implications of this nuclear reorganization on HIV integration site targeting are not clear. Integration occurred at a similar frequency in resting *vs.* active T cells, however, integration in resting T cells occurred in less gene-dense regions to a marginal extent relative to integration in active T cells [176].

More recent studies utilizing *in-situ* imaging of HIV-1 nucleic acid and integration site analysis have demonstrated that HIV-1 PICs accumulate at nuclear speckles and have a preference for integrating into speckle-associated domains (SPADs), regions enriched with transcriptional machinery [94]. Integration site analysis in Lucic *et al.* 2019 suggests that super-enhancers (SEs), genomic regions enriched in enhancers, active epigenetic marks such as H3K27ac and H3K4me1, and transcription factor binding sites, are enriched for RIGs [170]. The precise mechanism of PIC-targeting to these specific nuclear features are not well understood; however, these studies highlight the importance in unraveling the relationship between nuclear organization, HIV-1 integration site selection, and proviral transcriptional competence.

Ablation of IN-LEDGF/p75 interactions resulted in HIV-1 integration favoring interior regions of gene bodies towards the gene 5’ end rather than towards more promoter-distal introns

[177, 178]. LEDGF/p75 interacts with numerous mRNA splicing factors, perhaps mediating HIV-1 integration site targeting *via* interactions with cellular mRNA splicing machineries [178]. Knock-down (KD) of LEDGF on the cell appears to have less of an effect on integration frequency compared to cellular depletion of NPC-related factors [179]. Results from fluorescence imaging of HIV-infected cells have provided great insight into the role of CA protein in HIV-1 integration site targeting. Depletion of CPSF6 or infection with CA mutants (N74D and A77V) that attenuate interactions with cellular factors resulted in PIC and viral dsDNA accumulation in the peripheral region of the nucleus, often in the transcriptionally repressed LADs [94, 158, 159, 180]. These studies suggested that perhaps ablation of critical HIV-1 CA/IN-host factor interactions could influence aberrant integration site selection, which may lead to the establishment of latent infection. Barcoded viruses that facilitate multiplexed identification of HIV-1 integration sites and gene expression levels at these respective integration sites have shown that latent HIV-1 proviruses are more distal from active epigenetic marks compared to productive integrated viral genomes [181]. Battivelli *et al.* surveyed integration sites of productive (transcriptionally active), latent but reactivatable, and latent but not reactivatable HIV-1 and found that productive and reactivatable proviruses integrate into indistinguishable chromatin environments, enriched in active epigenetic markers and "open" chromatin [32]. This result suggested that, perhaps, superficially, integration site selection of productive and reactivatable proviruses are similar; however, subtle differences in higher-order chromatin architecture may be influencing viral latency and HIV-1 transcriptional control. In **Chapter VI**, we aim to identify critical higher-order chromatin and transcriptional signatures associated with active HIV-1 proviruses, assessing chromatin reorganization in response to proviral transcriptional activation at defined sites of integration.

1-E: HIV-1 Latency & Proviral Gene Expression

1-E-a: Establishment of the Latent Reservoir

Complete eradication of HIV-1 disease in the infected host is hindered by the establishment of latently infected reservoirs of resting CD4+ T cells. Within these latent cell populations, intact viral cDNA has been integrated into the host genome, yet HIV remains transcriptionally inactive. Consequently, these infected cells evade immune responses and treatment, enabling HIV to persist within the infected host [182-185]. Latent proviruses may be inactive but can undergo reactivation, leading to proviral transcriptional activation and subsequent viral replication. HIV latency is therefore a major obstacle in uncovering a functional cure.

Viral loads in infected patients on effective cART may reach below the threshold required for detection by PCR (<50 copies/mL), however, the virus can quickly rebound in patients that cease treatment [186, 187]. Kinetically, the latent reservoir is established within days after infection. The estimated stability of latently infected resting memory CD4+ T cells in patients receiving suppressive therapy has a half-life of ~44 months [188]. Homeostatic proliferation of memory CD4+ T cells makes them a good source for latent infection, and reactivation of latently infected cells allows for rapid and stealthy expansion of the viral pool. Latent reservoirs are found in lymphoid tissues, bone marrow, genital tract, the brain, and circulating blood [189]. It is estimated that while on continued cART, a patient would not achieve eradication of HIV-1 infection until about 70 years of treatment [188, 190, 191]. Earlier attempts to measure the latent reservoir used PCR-based methods that overestimated the size of the reservoir, as the majority of latently infected cells *in vivo* harbor replication-incompetent or "defective" proviruses [192-195]. Defective genomes have deletions or ABOBEC-mediated

hypermutations, thus making them incapable of supporting virus replication. Defective viral genomes are estimated make up $\geq 98\%$ of total integrated proviruses in patient cells [192, 196]. While these defective proviruses are not contributing to viral rebound, transcriptional activity from defective proviruses contributes to sustained immune activation and inflammation [194, 197]. The frequency of a latently infected cell in a patient is approximately one per 10^6 cells [198]. Although this appears to be a rare incidence, the persistence of latent infection remains the major barrier to eradication of infection. A significant fraction of persistently infected cells in patients have been shown to be clonally expand, allowing a small cell population of inducible, replication-competent proviruses to potentially reseed the active viral reservoir if cART is interrupted [196, 199, 200]. Studies have demonstrated that sufficient intact proviruses exist in clonally expanded populations of cells to sustain this viral rebound [201-203].

The precise mechanism(s) responsible for the establishment of latent infection is not clear. HIV-1 primarily infects activated CD4+ T cells but also infects resting T cells *in-vitro*. The latent reservoir may be established by infection of activated effector CD4+ T cells that then transition to a resting state. This transition to a resting memory state is called the effector-memory-transition (EMT) [46, 204, 205]. As discussed in the previous chapter (**Chapter I.4**), integration site selection influences the status of latency, highlighting the importance in determining the role nuclear ultrastructure plays in mediating HIV-1 site selection and transcription. Multiple studies have shown that HIV RIGs are found within SEs, SPADs, or other highly active genomic compartments that are dynamic and prone to major structural reorganization, suggesting that remodeling of chromatin architecture around proviral sites of integration may be important for driving or sustaining HIV-1 transcriptional activation [94, 170]. In addition to integration site selection and local proviral nuclear environment, several other

cellular factors within the infected host cell have been shown to suppress HIV-1 replication post-integration.

The host transcription factor, nuclear factor kappa-light-chain-enhancer of activated B cells (NFκB), can be sequestered in the cytoplasm by NF-kappa-B inhibitor beta (IκBα), preventing localization and occupancy at the HIV-1 LTR promoter [206, 207]. Phorbol myristate acetate (PMA) and ionomycin (Io) in combination robustly activate protein kinase C (PKC) and downstream effectors that lead to the release of NFκB from IκB, which facilitates NFκB translocation into the nucleus [208, 209]. Prostatin (Pro) and byrostatin (Byro) are latency-reversing agents (LRAs) that act as agonists of PKC by directly binding to IκBα, blocking NFκB cytoplasmic sequestration [210, 211]. The provirus may also be silenced by repressive epigenetic marks at the Nuc-0 and Nuc-1 nucleosomes flanking the HIV-1 5' LTR, causing the proviral promoter to become tightly wound and inaccessible from host transcription factors [212]. Repressive chromatin marks such as H3K9me3, H3K27me3, and CpG methylation may be enriched at latent proviruses [181, 213-216]. *Trans*-acting chromatin factors have been implicated in facilitating the recruitment of histone deacetylases (HDACs), polycomb repressive complex-2 (PRC2), and other repressor complexes to the core HIV-1 5' LTR promoter [217, 218]. Bromodomain protein BRD4 occupancy is normally associated with processive transcription; however, BRD4 occupancy proximal to the HIV-1 5' LTR promotes latency by engaging repressive SWI/SNF chromatin remodeling complexes, leading to a reduction in proviral chromatin accessibility [219].

Interference of HIV-1 transcription caused by active host gene expression and provirus orientation relative to the flanking host gene have also been proposed to influence latency [220-222]. It has been proposed that if the provirus and host gene are in opposite transcriptional

orientations, convergent transcription causes proviral gene expression to be prematurely terminated due to the collision of RNAPII transcription complexes originating from both the oppositely-oriented HIV-1 5' LTR and flanking host gene promoters [222]. In **Chapter VI**, we interrogate HIV-host transcriptional landscapes at defined sites of integration to understand how proviral orientation and host-driven transcriptional read-through are influencing HIV-1 transcriptional states. Factors of the mTOR complex and related downstream effectors have also been shown to influence latency, possibly *via* TCR/CD28 signaling or NFκB activation [223]. Interestingly, a recent study has shown that proviruses from elite controllers (patients with low viral setpoint), a small subset of infected patients that can control their viremia in the absence of cART, have integrated into heterochromatin such as centromeric satellite DNA [224]. Understanding the molecular mechanisms that confer increased differential integration into heterochromatin for the elite controller populations may provide insight into factors that influence proviral transcriptional competency and the overall relationship between integration site selection and viral transcriptional output.

Transcriptional heterogeneity has been observed in both latently infected clonal CD4+ T cells [185] and primary resting CD4+ T cells [192]. In a population of clonally expanded primary T cells, reactivation or latency reversal occurs in only a small fraction of the cells. Single-cell RNA-seq (scRNA-seq) has been applied to reveal the heterogeneity present even within a clonal population of cells, potentially suggesting that state of the cell influences proviral gene expression over HIV-centric mechanisms [225]. Other models suggest the HIV-1 transcriptional program is a Tat feedback loop, where stochastic Tat fluctuations drive entry into latency [114, 226, 227].

Specific integration sites have been linked to persistence and clonal expansion of infected cells in patients on long-term cART [199, 200, 228]. Integration into specific regions of *MKL2*, *BACH2*, and *STAT5B* are highly enriched in primary patient samples [196, 199, 200]. These genes play a role in tumorigenesis, T cell homeostasis, immune signaling, and B cell development [229-231].

1-E-b: HIV-1 Transcriptional Regulation

The HIV-1 LTR promoter (636 base-pairs (bp)) can be subdivided into four functional domains: 1) a base core promoter with a canonical TATA element composed of three tandem Sp1 binding sites along and an enhancer element containing two binding sites for the transcriptional nuclear factor kappa B (NF-κB) , 2) an upstream regulatory region containing elements necessary for cell type-specific expression, and 3) a downstream regulator region containing secondary enhancer elements [74, 232]. The U3 and R regions of the proviral 5' LTR act as the viral promoter. Although the 5' and 3' LTRs have identical sequences, simultaneous transcription from both LTRs is highly inefficient due to the interference of the U3 region by (+) transcriptional complexes that terminate RNA synthesis at the R/U5 junction [233]. The 5' LTR also encodes a novel RNA, the *trans*-activation response (TAR) element, located directly downstream the transcription initiation site. The TAR sequence consists of a 59-nucleotide RNA stem-loop structure that forms a trans-activator protein- (Tat)-binding molecular scaffold at the 5' end of all canonical HIV-1 transcripts [232, 234]. Interaction between Tat protein and TAR RNA elements promote formation of super-elongation complexes (SEC) that drive processive HIV-1 transcription [218].

In the absence of Tat, proviral transcription can still initiate with the formation of stable transcription complexes, however, transcription prematurely terminates, generating only short

abortive transcripts. The majority of mammalian protein-encoding genes have an occupied RNAPII at the promoter proximal region, poised for transcription [235], yet promoter proximal pausing leads to the formation of short transcripts that accumulate following abortive RNAPII escape. In HIV-1 transcription, short transcripts are generated by processing of TAR RNA by termination factors Setx and Xrn2 and the 3' to 5' exoribonuclease, Rrp6 [236].

The negative elongation factor (NELF) and the DRB sensitivity-inducing factor (DSIF) block productive HIV-1 transcriptional elongation in the absence of Tat. DSIF is a two-subunit complex comprised of Spt4 and Spt5, which are recruited to newly initiated transcription factors and promote 5' capping [237]. Spt5 directly interacts with nascent RNA and facilitates the recruitment of NELF [238]. NELF forces premature termination, and thus, shRNA KD of NELF leads to enhanced basal HIV transcriptional activity [239, 240]. The switch from promoter proximal pausing to processive proviral transcription is mediated by Tat and the positive transcription elongation factor (P-TEFb). P-TEFb is comprised of the CDK9 serine/threonine kinase and cyclin T [241, 242]. Structural studies have shown that the binding of Tat to CDK9/cyclin T1 (P-TEFb) allows the viral factor to establish partial helical secondary structure and induces conformational changes in P-TEFb which allows stable binding of Tat and the cyclin T1 subunit to TAR [242, 243]. Recruitment of P-TEFb at TAR enables CDK9 to phosphorylate NELF, leading to NELF disengagement from TAR RNA and the paused polymerase complex [239]. CDK9 also hyper-phosphorylates the carboxy terminal domain (CTD) of RNAPII at Ser2 and Ser5 residues of the heptad repeats and homologous heptapeptide repeats of Spt5 at Thr4 residues to convert Spt5 into a stimulatory elongation factor [244-246]. The effect of Tat and P-TEFb is thus to remove the "roadblocks" to elongation imposed by NELF and DSIF, promoting processive proviral transcription.

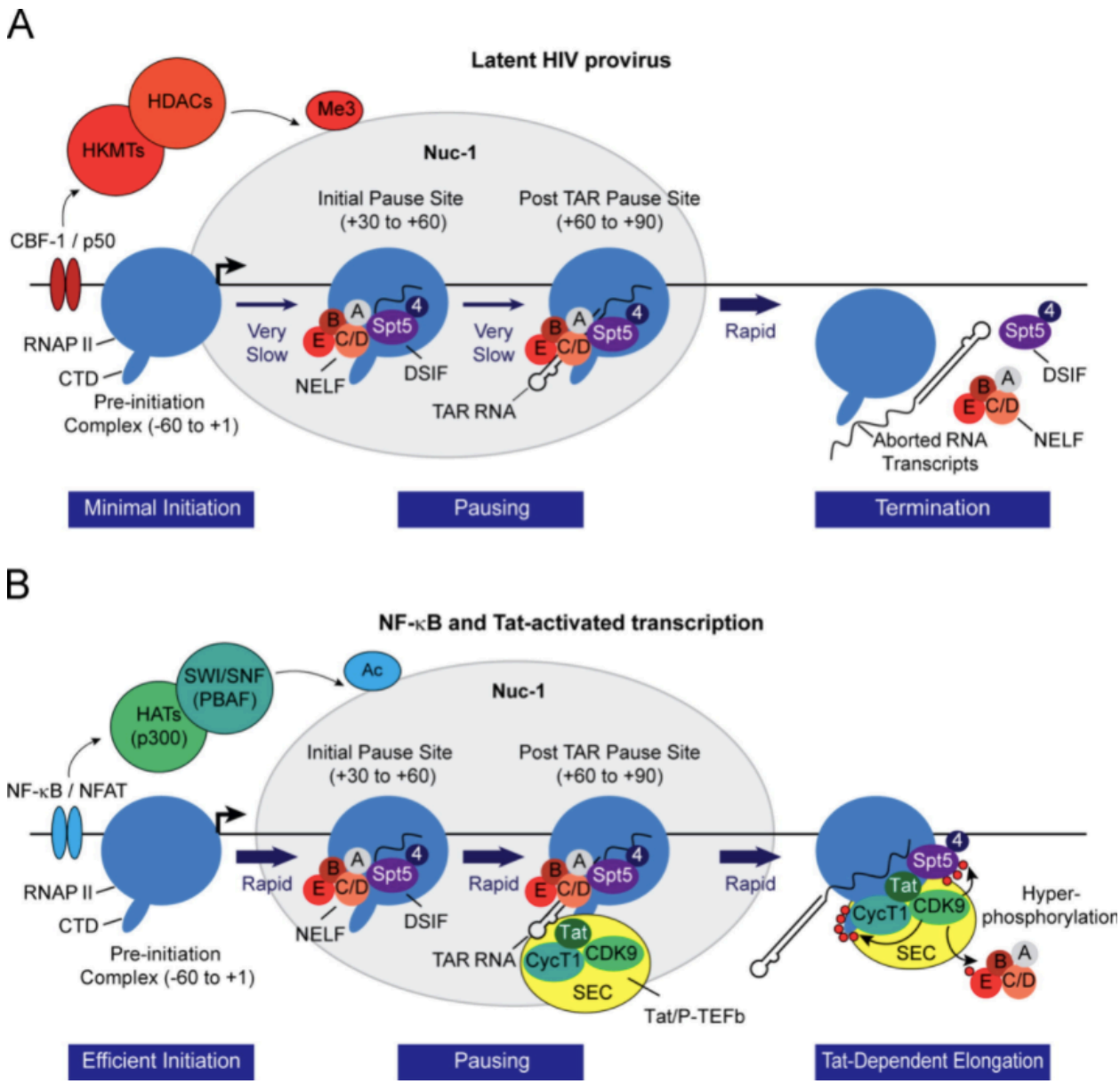


Figure I.E.1. Regulation of productive HIV-1 transcription. Transcriptional control at latent HIV-1 proviruses. Latent proviruses have higher enrichment of histone deacetylases (HDACs), histone methyltransferases (HMTs), and associated repressive modifications at Nuc-1 at the HIV-1 5' transcriptional start site (TSS). The deacetylated HIV-1 chromatin becomes a target for additional epigenetic silencing *via* recruitment of polycomb repressive complex-2 (PRC2), which induces histone methylation. Viral transcripts are prematurely terminated, and there is an accumulation of transcriptionally paused polymerase complexes due to the absence of the viral protein Tat. Escaping RNAPII complexes induce abortive transcription due to NELF **(A)**. Transcriptional activators at

productive HIV-1 proviruses. Proviral transcriptional initiation is driven by NF-κB/NFAT activity and associated co-activators such as p300. The transitioning of the stalled polymerase into processive elongation requires recruitment and activation of P-TEFb kinase activity. Tat is a *trans*-acting viral factor that helps elicit HIV transcriptional elongation by recruiting P-TEFb to the paused complex at the HIV-1 TAR RNA hairpin. The assembly of the SEC leads to hyper-phosphorylation of the CTD of RNAPII, facilitating processive transcription at the HIV-1 provirus **(B)**. The original copyrighted material was reproduced from Mbonye *et al.* 2014 [10]. Reprinted with permission from Elsevier and Copyright Clearance Center. License: 5098551131161.

HIV-1 exhibits a diverse transcriptome, utilizing the cellular spliceosome machinery. Alternative splicing of HIV-1 mRNAs increases viral protein coding potential and controls the dynamics of viral replication [247, 248]. Full-length HIV-1 serves as the genome or as an mRNA that can undergo splicing using four donors and 10 acceptors to create over 50 physiologically relevant transcripts in two size classes (1.8 kb and 4 kb) [249]. A schematic depicting HIV-1 splicing diversity can be seen in **Figure I.E.2.**

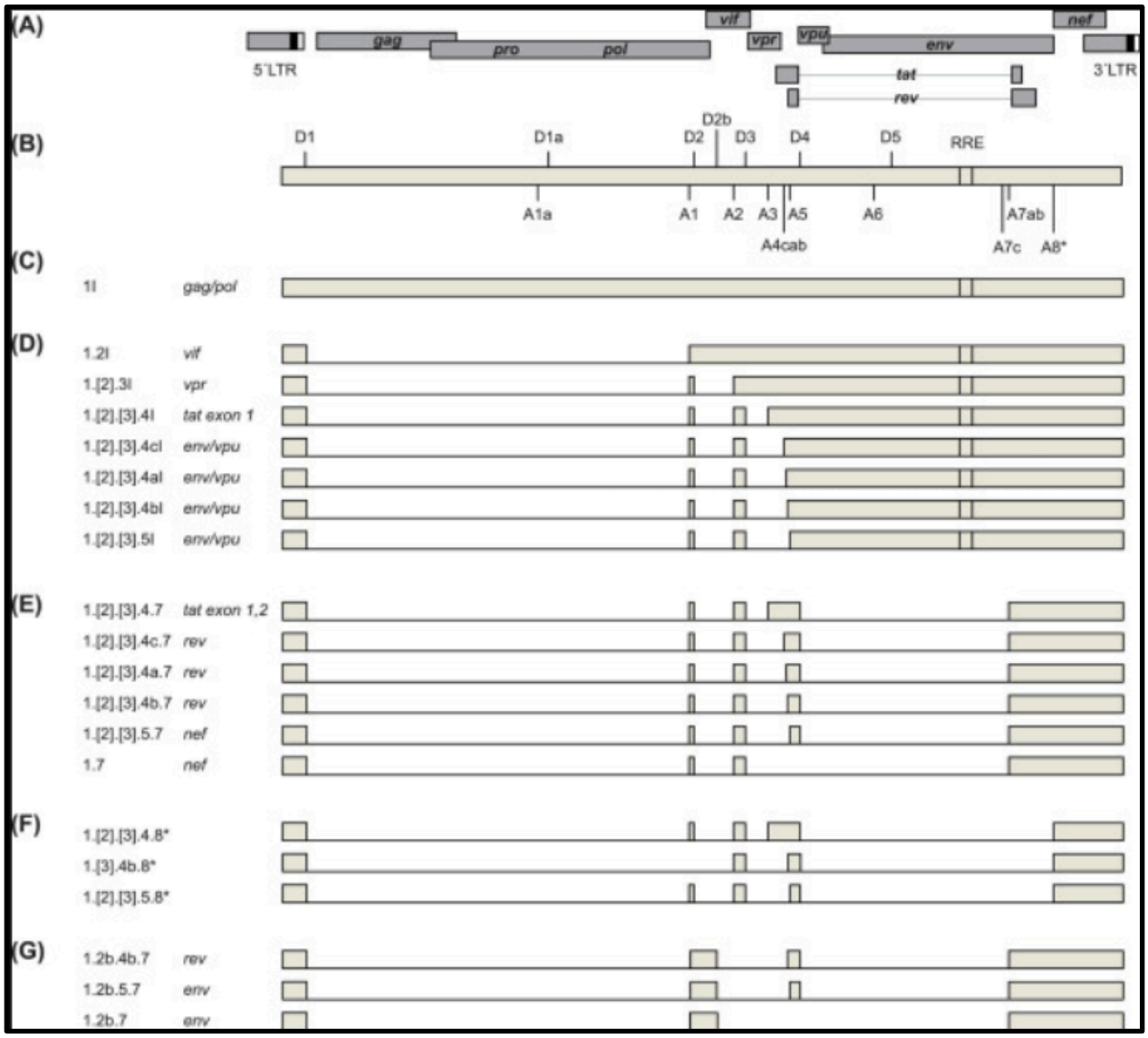


Figure I.E.2. Schematic of HIV-1 splicing and transcript diversity. The HIV-1 ORFs are shown as grey boxes. The LTRs are located at both ends of the HIV-1 genome and are depicted in a three-color code: U3 (grey), R (black), and U5 (white) **(A)**. Positions of 5' and 3' splice sites (ss). The RRE corresponds to the binding site of Rev, which mediates HIV-1 mRNA export of unspliced and intron-containing transcripts to the cytoplasm for translation **(B)**. The unspliced 9 kb mRNA is expressed from the 5' LTR promoter, serving as genomic RNA or cods for Gag and Pol. Leader exon 1 is non-coding but is present in all transcript variants. Alternative splicing of non-coding leader exons 2 and 3 and subsequent insertion lead to alternatively spliced mRNAs. Intron-containing transcripts are marked with an "I" **(C-F).** Additional splicing occurs for the 2 kb mRNA class **(E)** between splice sites D4 and A7, which leads to the absence of the RRE in these mRNAs. Recognition of the alternative splice acceptor A8, the 1 kb mRNA is formed **(F).** Other alternative transcripts encoding for protein isoforms of Rev4b and Env-gp41 C-terminus originating from splicing at the D2b site **(G)**. The original copyrighted material was reproduced from Sertznig *et al.* 2018 [250] and originally published in *Virology*. Reprinted with permission from Elsevier and Copyright Clearance Center. License: 5102820643273.

1-E-c: Therapeutic Strategies to Target the Latent Reservoir

Current strategies for HIV-1 eradication include induction of HIV-1 gene expression to expose latently infected cells for immune clearance, which is known as the "shock-and-kill" approach [251]. Small molecule compound library screens have identified drugs that can reactivate latent HIV-1 proviruses, which have been "coined" latency-reversing agents (LRAs) [251]. However, none of these LRAs have reached a durable HIV-1 remission in clinical trials. HIV-1 gene expression is not deterministic. Maximal T cell activation only reactivates a small percentage of latently infected cells, due to the stochastic fluctuation of Tat expression [114, 227]. The stochastic nature of HIV-1 transcriptional activation makes the design of effective LRAs very challenging, and there is therefore a need to identify pharmacological compounds that can modulate the fluctuations of proviral gene expression.

T cell activation remains the most robust way to reactivate HIV-1 proviruses, however the systemic side effect and potential of increasing the proliferative capacity of HIV-infected cells makes global T cell activation not amenable for clinical use [252]. Drug screens testing HIV-1 latency reversal potential identified multiple viral and cellular pathways mediating proviral gene expression. HIV-1 can be reactivated by increasing Tat-dependent transcriptional elongation by inhibiting BRD4 interaction with the P-TEFb. JQ1, a bromodomain and extra-terminal (BET) inhibitor, blocks the interaction between BRD4 and P-TEFb, thus increasing P-TEFb occupancy at the HIV-1 5' LTR promoter [219, 253]. Other LRAs modulate the epigenetic environment around the provirus such as HDAC inhibitors (vorinostat [254], suberoylanilide hydroxamic acid (SAHA) [255], panobinostat [256], and romidepsin [257]). Cellular pathways that modulate latency reversal to an extent include NF-κB activation through Toll-like receptor (TLR) activation (TLR1/2 agonist Pam3CSK4 [258]), noncanonical NF-κB activation (SMAC mimetics [259, 260]), PKC activation

(bryostatin [261]), and other pathways such as reactive oxidative stress inducer (juglone [262]). Other recent studies have shown a positive association between administration of stress-inducing compounds and latency reversal in primary T cell models [263, 264]. In **Chapter VII.5**, we begin to investigate how the cellular heat shock response increases the frequency of latency reactivation.

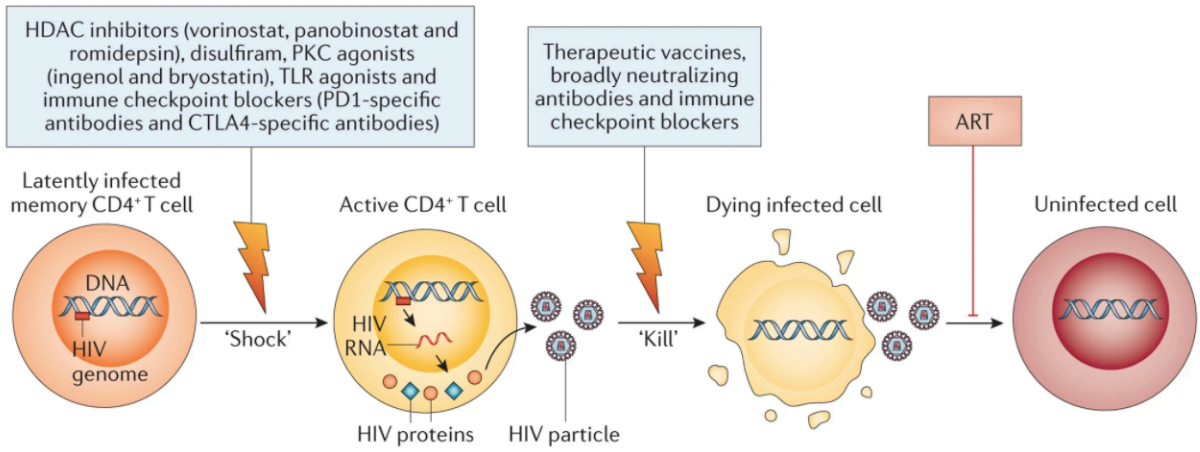


Figure I.E.3. 'Shock & kill' strategy to purge the latent reservoir using latency-reversing agents (LRAs) that activate processive proviral transcription. 'Shock-and-kill' reverses latency by inducing proviral gene expression, resulting in the production of viral proteins and eventual host-mediated clearance of the virus-infected cell. CTLA4, cytotoxic T lymphocyte antigen 4; HDAC, histone deacetylase; PD1; programmed cell death protein 1; PKC, protein kinase C; TLR, Toll-like receptor. The original copyrighted material was published by Deeks *et al.* 2015 [1]. Reprinted with permission from Springer Nature and Copyright Clearance Center. License number: 5095490407457.

Pharmacological compounds that also suppress HIV-1 gene expression and replication have been characterized for their abilities to block latency reversal. These compounds include CDK9 inhibitors (flavopiridol [265]), mTOR inhibitors such as rapamycin [223] that inhibit NF-κB activation, JAK inhibitors (filgotinib [265] and ruxolitinib [266]), and RNA splicing inhibitors (filgotinib [265]). This strategy of blocking latency reversal is called "block-and-lock", which aims

to permanently lock the virus into a transcriptionally silent or deep latent state, ablating viral rebound upon cessation of cART [267]. Perhaps, compounds that can influence the integration of HIV-1 into transcriptionally repressed chromatin environments (e.g., centromeres, LADs) can ensure permanent silencing of any replication-competent provirus. Compounds that interfere with IN and LEDGF/p75 interactions ("coined" LEDGINs) [268] shifts integration preference toward the inner nucleus compared to wild-type (WT) integration. Small molecules that also interfere with CPSF6 and viral CA interactions may modulate integration site selection [159]. Other compounds that interfere with chromatin remodeling and viral transcriptional activators such as Tat are also potentially promising therapeutic targets for inducing long-term proviral suppression [269].

1-E-d: Molecular & Cellular Toolbox for Mechanistic Latency Studies

Latency Cell Lines

HIV-1 latency studies have been hindered by the fact that only a small percentage of cells become latently infected *in vivo* [270]. *In-vitro* latency cell models have been developed to circumvent this issue. However, notable differences exist among different latency cell model systems, including differential reactivation profiles, disparities in T-cell subsets being represented, genetic composition of viruses employed (WT to functional deletions), and differences in experimental approaches taken to generate latent infection in these models [270-272]. The variability across these cell models make it difficult to reliably predict LRA efficacy that is consistent across the different latency models, as well as efficacy within infected patient cells tested *ex vivo*. The model developed by Bosque *et al.* [272], derived from central memory CD4+ T cells (T_{CM}), are highly responsive to stimuli that activate the nuclear factor of activated T-cells

(NFAT). J-Lat clones developed in the Verdin lab [185] tend to be highly responsive to stimuli that activate NF-κB (PKC agonists & TNF-α). While these models may not be ideal for testing LRA efficacy relative to primary infected cells taken *ex vivo*, the signaling pathways that activate these latency cell lines are capable of reactivating latent primary CD4+ T cells, suggesting they may be a useful tool for mechanistic latency studies. Multiple characterized cell models from the Greene [273], Lewin, [274] Planelles [272], Siliciano [262], and Verdin [185] groups have been developed and applied to latency studies. The Verdin group has generated a number of Jurkat T cell line-derived clones of J-Lat cells, harboring latent HIV-1 in single and defined integration sites. The provirus was engineered to express GFP in lieu of Nef, therefore, enabling facile identification of transcriptionally activated 5' LTR promoters [185]. The J-Lat cell model contains WT Tat and TAR. For our studies presented in **Chapter VI**, we use different inducible J-Lat clones to study chromatin organization and transcriptional patterns at defined sites of proviral integration. The benefit of having an inducible clonal system with a unique and mapped integration site allows us to assess how local nuclear environment is influencing HIV-1 transcriptional state. In **Figure I.E.4,** we show the reactivation profiles of four J-Lat clones (10.6, 8.4, 9.2, 15.4), all with distinct sites of integration (highlighted in **Table I.1).** These clonal J-Lat cells are treated with a diverse panel of LRAs.

TABLE I.1. Integration site and transcriptional orientation of provirus and respective host gene. Table adapted from Symons *et al.* [275].

Cell Line	Chromosome	Position	Gene	Gene Orientation	HIV Orientation
J-Lat 10.6	9	136,468,579	SEC16A	-	+
J-Lat 8.4	1	7,946,384	FUBP1	-	-
J-Lat 9.2	19	4,381,104	PPP5C	+	+
J-Lat 15.4	19	34,441,293	UBA2	+	+

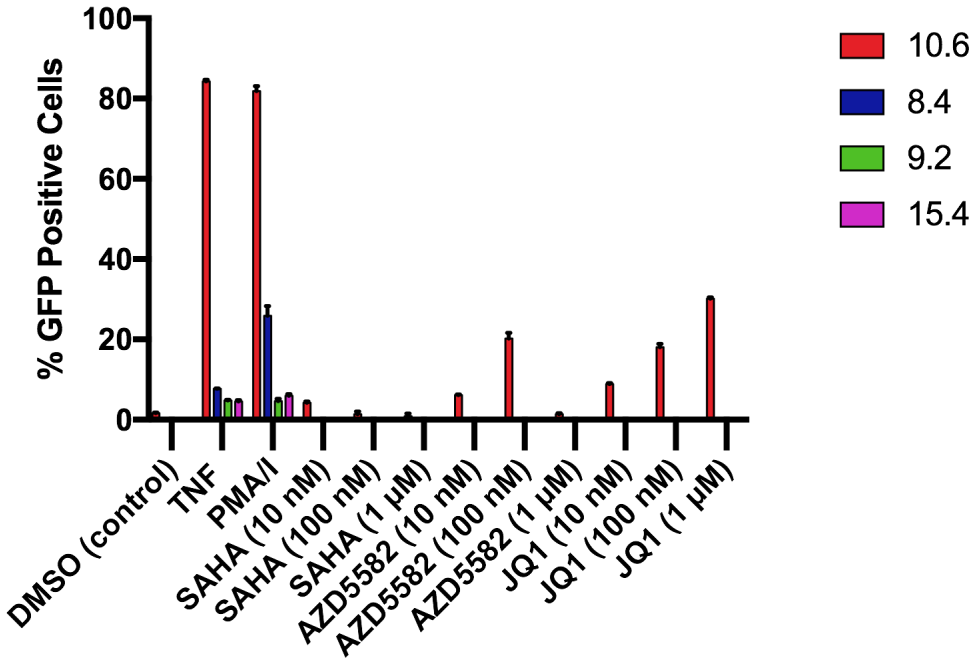


Figure I.E.4. Reactivation profiles of four distinct J-Lat clonal models treated with a diverse panel of LRAs. Reactivated cells were GFP+. Reactivation frequency was quantified *via* flow cytometric analysis.

Mechanistic latency studies in primary cell models

With the variability across latency cell models and potential for a specific cell line to not reflect relevant disease physiology, studies in primary cell models or in infected cells taken *ex*

vivo are preferrable, if possible. An optimized HIV-1 dual reporter virus, enabling facile isolation of productive and latent proviruses, is the HIV_{GKO} dual reporter [32, 276]. Within this HIV-1 reporter system, pseduovirus particles can be generated with a proviral genome that allows identification of latently infected cells *via* EF1a-driven mKO2 expression and lack of LTR-driven csGFP. Productive virus can be identified by infected cells expressing both mKO2 and csGFP. Following infection of primary T cell models, latent and active cell populations can be sorted *via* fluorescence-activated cell sorting (FACS). Isolation of these different infected cell populations can then enable mechanistic studies probing for differences in proviral integration sites, gene expression networks, and chromatin environments across productive and latent cells.

Latency studies in primary cell models following *de novo* infection is amenable for single-cell and mechanistic studies, as various laboratories have developed well-established protocols for latent cell isolation. The Karn group has demonstrated that a high number of latently infected cells can be obtained by co-culturing HIV-infected CD4+ T cells and H80 feeder cells [277]. A method developed by the Bosque and Planelles labs uses T-cell receptor (TCR) stimulation in the presence of TGF-β and anti-IL4/anti-IL2 to differentiate CD4+ T cells into non-polarized subsets (central memory) [272, 278]. The Siliciano group developed a protocol to first synchronize cells to a resting state in the presence of anti-apoptotic factor Bcl2 to prolong the culture. Cells are then infected with HIV-1, and latently infected cells are then isolated [262]. These models can be used following infection with the dual reporter virus, $HIV_{GKO,}$ to preferentially isolate latently infected cells with relative ease.

While studies using primary cells derived from patients is the ultimate goal, the small fraction of latently infected cells makes functional latency studies difficult. Nonetheless, robust assays have been developed to estimate the latent viral reservoir in patients, as well as animal

models, for *in vivo* studies. A commonly used methodology to measure replication-competent HIV-1 is the quantitative viral outgrowth assay (QVOA). In this assay, stimulated peripheral blood mononuclear cells (PMBCs) from HIV-positive individuals on cART are co-cultured with CD4+ T cells purified from HIV-negative individuals. Replication-competent HIV from the infected population is amplified by infecting the HIV-negative CD4+ T cells over time. Viral replication can then be measured *via* p24 (CA protein) ELISA or reverse transcription assays [279, 280]. QVOA is considered by many to be the "gold standard" for estimation of the replication-competent latent reservoir [281]. However, there are several limitations to the assay including: 1) large blood volume is required from multiple donors, 2) long and laborious, 3) sample-to-sample variation across patient-derived PBMCs, and 4) low viral spread and copy number of intact latent proviruses leads to issues in sensitivity and detection [282]. The limitations of QVOA and other quantitative PCR-based assays to estimate latent viral reservoirs presents a commodity for an assay that is sensitive enough for simultaneous detection of total viral DNA, RNA, and protein molecules at the single-cell level. Through our multiplexed fluorescence imaging approach (MICDDRP), presented in **Chapter II**, we provide a detailed step-by-step protocol that allows for simultaneous detection of HIV-1 viral nucleic acid (DNA & RNA) and protein to monitor viral replication. We present several examples where we applied this technology to visualize replication across a broad spectrum of viruses, highlighting the potential to implement our fluorescence-based approach to measure the percentage of cells that have nuclear viral dsDNA and the fraction of those cells that are producing vRNA and protein. This can potentially provide a more reliable metric of the "true" number of cells that have defective proviral DNA *vs.* infected cells with intact HIV-1 DNA that can reactivate from a latent state (reactivated cells will produce vRNA and protein).

1-F: Nuclear Chromatin Organization & Relevance to Viral Transcription

1-F-a: Current "Dogma" of Nuclear Organization

The spatial organization of the nuclear genome in 3D space is nonrandom, playing a critical role in establishing and maintaining cellular identify [283-285]. Higher-order organization of eukaryotic chromosomes regulate gene expression through long-range chromatin interactions that help define the cellular transcriptome [286]. Cis-regulatory sequences may be located long distances from the target gene but the formation of chromatin loops enables close physical contact between distal enhancer and promoter regions [287]. Chromosomes are composed of Topologically Associating Domains (TADs), structural genomic units that vary in size from ~40 kb to 1 Mb [288-290] and are characterized by sharp boundaries that promote long-range interactions [291, 292] (**Figure I.F.1**). TADs are local epigenomic compartments that form intrachromosomal contacts and are found in functionally distinct A or B compartments, which are composed of largely active (euchromatin) or inactive (heterochromatin) chromatin, respectively [291-293]. Interphase chromatin is highly dynamic, and as the transcriptional state of a TAD changes, chromatin contacts within the domain are altered, remodeling chromatin architecture [293].

The dynamic nature of TADs can be attributed to chromatin looping, mediated by the CCCTC-binding factor (CTCF). CTCF is an 82-kDa highly conserved zinc finger protein that functions as a transcriptional activator, repressor, or an insulator protein, blocking enhancer and promoter interactions by forming distinct directional chromatin loops[294, 295]. CTCF forms loop domains likely through an extrusion mechanism involving cohesin [290, 295]. TADs are separated by boundaries often enriched in CTCF-binding sites and highly transcribed genes. These boundaries help delineate genomic domains, which are dynamic and susceptible to major structural

reorganization, contributing to the complexity of the hierarchical organization of cellular chromatin [290, 295-298].

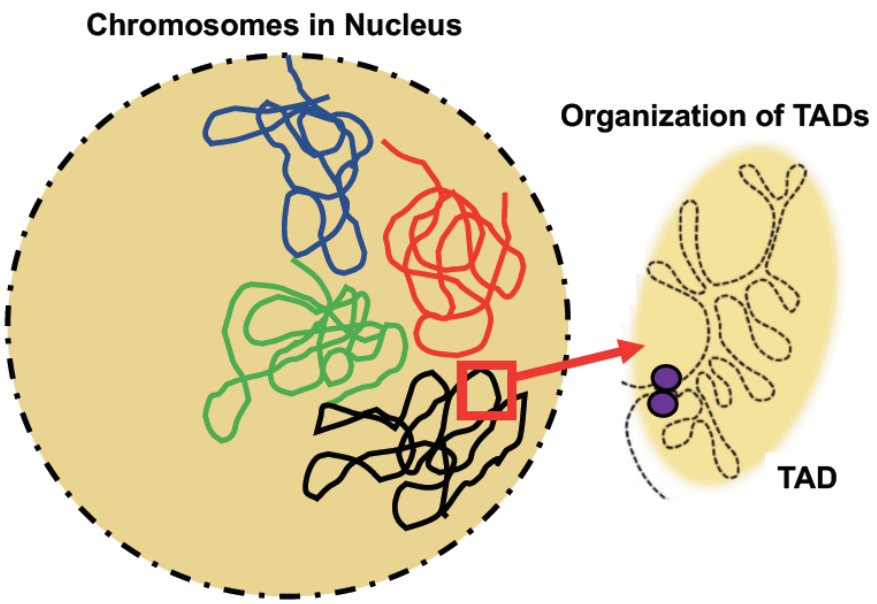


Figure I.F.1. Chromatin organization and topological domains. Chromosomes occupy distinct territories in the nucleus (indicated by different colors on left), and long-range looping or chromosomal interactions mediated by CCCTC-Binding Factor (CTCF; purple spheres) organize chromatin into smaller Topologically Associating Domains (TADs) (on right).

Early cytological microscopy studies of eukaryotic nuclei observed that euchromatin preferentially localizes towards the nuclear interior and heterochromatin at the nuclear periphery [299, 300]. Heterochromatin contains densely packed nucleosomes with repressive epigenetic modifications and is often transcriptionally silent. Conversely, euchromatin is more highly accessible and packed with histones marked with active epigenetic modifications, promoting transcriptional activation [301]. Spatial partitioning of these functionally distinct chromatin states can be deconvolved *via* chromosome conformation capture (3C) technologies such as Hi-C [292, 293]. In addition to A/B functional compartmentalization, the genome is organized by anchoring to

subcellular nuclear structures such as the nuclear lamina. The lamina is found towards the nuclear periphery. Here heterochromatin forms Lamina-Associated Domains (LADs), genomic regions in contact with nuclear lamina proteins such as Lamin B1 [284, 302]. The dynamics of chromatin localization have been linked to gene expression. Multiple studies report chromatin-lamina contacts correlating with gene repression [284, 303, 304]. Similarly, a study reprogramming spatial chromatin organization *via* a novel methodology they "coined" CRISPR-GO, demonstrated that nuclear positioning directly impacts the transcriptional state of a gene [305]. In our presented work, we seek to understand how spatial organization and local chromatin environments shape the viral transcriptional landscape. We apply 3C-based methodologies to probe the long-range chromatin interaction profiles between virus-host DNA (**Chapter VI & VII.1)** and visualize viral gene expression patterns with high spatiotemporal resolution *via* single-molecule fluorescence *in-situ* hybridization (FISH), presented in **Chapter II.**

1-F-b: Molecular Toolbox to Investigate Chromatin Organization & Dynamics

Chromosome conformation capture (3C) technologies to probe 3D chromatin organization

Chromosome conformation capture (3C) technology and 3C-based derivative methods are important tools utilized for studying the 3D spatial organization of chromatin architecture. The strategy of these methods is to quantify the frequencies of contacts between distal DNA segments in a cellular population [306]. The principal steps of 3C and 3C-based experiments are very similar across the different platforms. They all involve 1) crosslinking chromatin using a chemical fixative agent in solution (most often formaldehyde) to create covalent bonds between DNA fragments bridged by proteins, 2) isolating and digesting chromatin using a restriction enzyme to create pairs of crosslinked DNA fragments that are distant in linear distance but close

in 3D space, 3) ligate the sticky digested ends of crosslinked DNA fragments to bring together spatially proximal DNA fragments, 4) reverse the crosslinks to obtain 3C templates, and 5) interrogate the ligated DNA fragments by quantitative PCR (qPCR) or deep sequencing to determine chromatin interaction profiles at genetic *loci* of interest [307-309].

3C technology detects frequency of ligation junctions, which is a proxy for chromatin-chromatin interactions or close spatial proximity, by quantitative PCR (qPCR). 3C provides a method for visualizing the genome at high resolution, however, this methodology is low throughput, requiring PCR primers designed to amplify regions of interest. For this reason, 3C can only detect spatial relationships between targeted DNA sequences ("one *vs.* one") [310]. To overcome this limitation, several 3C-based methods have been developed to generate higher throughput data.

The development of 4C coupled to next-generation sequencing (NGS) technologies provided a high-throughput 3C-based method to probe the chromatin interaction network [311, 312]. 4C-seq is able to assess chromatin interactions between one genomic locus of interest (referred to as "bait" or viewpoint) and all other genomic *loci* ("one *vs.* all"). In 4C library preparation, small DNA circles are created by cleaving DNA with a second restriction enzyme and re-ligating the 3C DNA templates. Inverse PCR, using bait-specific primers, is then applied to amplify any interacting fragments. The genome-wide interactions with the "bait" sequence are then analyzed following NGS. **Figure I.F.2** presents a workflow we have developed for 3C-based assays to specifically target the HIV-1 genome. Designating a sequence within the *Vpr* gene of the HIV-1 genome, we probe direct interactions between cellular chromatin and integrated HIV-1. We apply 4C-seq to identify genetic *loci* interacting with integrated HIV-1 and then use 3C-qPCR to validate our top "hits" (cellular chromatin interactors with HIV-1 chromatin). Our protocol is

under optimization, and we present preliminary work measuring differential chromatin interaction profiles between HIV-inactive and -active J-Lat 10.6 cells (4C-seq & 3C-qPCR), as well as early 4C-seq studies on Hepatitis B virus (HBV). The current status of this work can be found in **Chapter VII.1.**

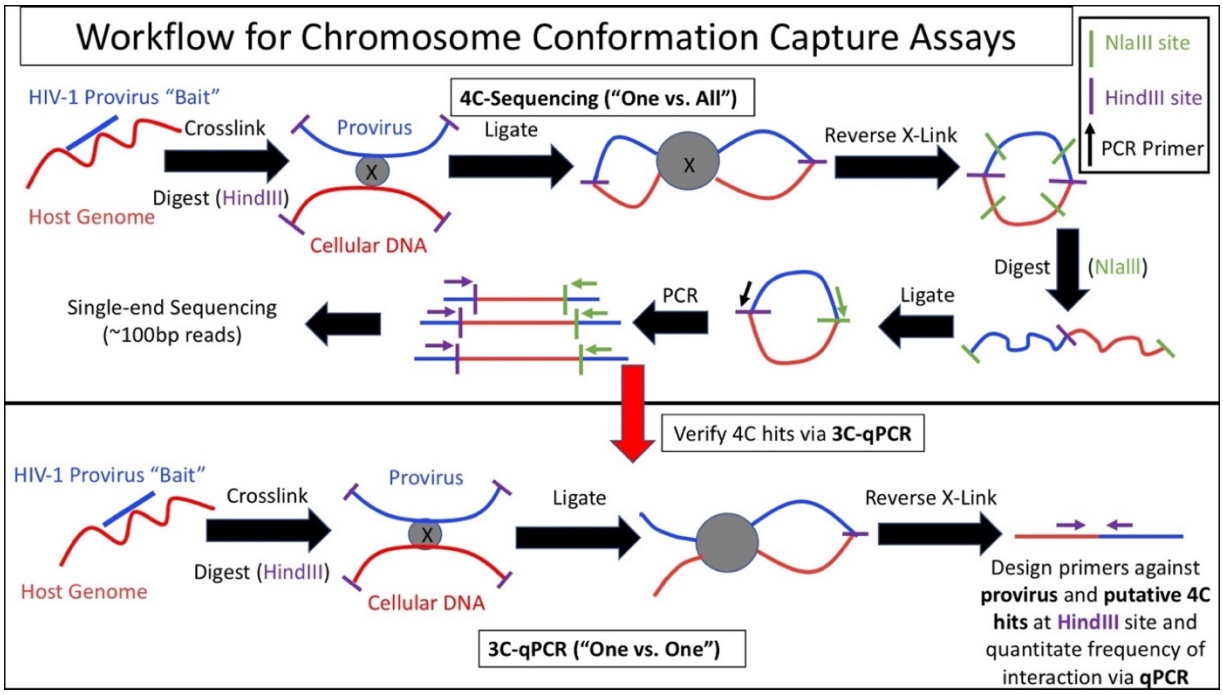


Figure I.F.2. Chromosome conformation capture (3C) pipeline for targeted interrogation of HIV-host chromatin interactions. 4C-seq and 3C-qPCR assays to assess viral-host chromatin interactions were performed as in Majumder *et al.* [309]. Provirus and host chromatin interactions are captured *via* chemical crosslinking, followed by digestion (HindIII) and intramolecular ligation to generate provirus-host DNA hybrids. In 4C library preparation, a second round of digestion (NlaIII) is performed before circularizing and generating a sequencing library of all hybrid fragments (~100 bp).

The improvements in high-throughput sequencing promoted the development of 3C-based methods that can interrogate genome-wide chromatin interactions in an unbiased manner ("all *vs.* all"). Hi-C is one of the first of these higher throughput methods to be developed and does not depend on specific primers to generate genome-wide contact maps [293]. In Hi-C

experiments, the first step is to generate contact segments similarly to 3C. However, after restriction digestion, the sticky ends are filled in with biotin-labeled nucleotides followed by blunt-end ligation. The expected interacting DNAs are sheared and then purified *via* a biotin pull-down using streptavidin beads to ensure only biotinylated junctions are selected for deep sequencing and downstream analyses.

High-resolution Hi-C datasets (> 10^9 paired reads/library), which enable bin reads at up to 1-5 kb resolution, enable the identification of contact domains, smaller than TADs. Point-to-point chromatin contacts between sequences bound by CTCF now yield a strong punctate signal in the contact density matrix, allowing accurate detection of chromatin looping between two distinct genetic *loci* [292]. Hi-C data binned at 10-50 kb suggests that compartments can be rather small, perhaps, consisting of a single active or inactive locus [290]. This degree of resolution makes Hi-C a powerful methodology in not only studying 3D chromatin organization at a genome-wide level but also enables targeted interrogation of genetic *loci* of interest. In **Chapter VI**, we apply Hi-C to map changes in chromatin organization at defined sites of HIV-1 proviral integration, mapping how chromatin restructures as a function of HIV-1 integration and proviral transcriptional activation. We specifically probe the functional compartmentalization of the HIV-1 provirus in differential transcriptional states and assess the role of chromatin looping in modulating proviral activity. Other 3C-based strategies such as Targeted Chromatin Capture (T2C) now offer a high-throughput method to interrogate large selected regions of the genome with an unbiased view of the spatial organization of a genetic locus of interest. T2C offers improvements in resolution compared to Hi-C and requires significantly lower sequencing depth [313, 314]. **Table I.2** describes the advantages/limitations of various 3C-based methods below.

TABLE I.2. Chromosome conformation capture (3C) technologies. The advantages and limitations of 3C-qPCR, 4C-seq, 5C, Hi-C, & T2C are presented. This table is reused from Kolovos *et al.* [314]. This is an open access article distributed under the terms of the Creative Commons CC BY license. This work was originally published in *Epigenetics Chromatin.*

Method	Applications	Advantages	Limitations
3C-qPCR	One-to-one	Simple analysis	Laborious, requires knowledge of the locus and proper controls
3C-seq/4C-seq	One-to-all	Good resolution, good signal-to-noise ratio	Restricted to single viewpoint per experiment when multiplexing several viewpoints, analysis requires extra bioinformatics expertise, not an all-to-all genome-wide method
3C-on-chip (4C)	One-to-all	Relatively simple data analysis	Poor signal-to-noise ratio, difficult to obtain genome-wide coverage
5C	Many-to-many	Identifies interactions between many individual fragments	Very laborious, no genome-wide coverage, primer design can be challenging. Analysis requires advanced bioinformatics expertise
Hi-C	All-to-all	Explores the genome-wide interactions between all individual fragments	Very expensive, requires a large sequence effort to obtain sufficient coverage, approximately 10 to 40 kbp resolution, requires advanced bioinformatics expertise
T2C	Many-to-all	Explores the interactome of a selected region in cis but also in trans, high (restriction fragment) resolution, cheaper than Hi-C and 5C, requiring only half a lane of Illumina HiSeq2000	Is restricted to the selected regions of the genome, requires advanced bioinformatics expertise

In-situ hybridization (ISH) of nucleic acid(s) of interest to follow transcriptional dynamics

Advancements in single-molecule DNA/RNA fluorescence *in-situ* hybridization (smFISH) have improved visualization of chromosome spatial organization, chromatin domains, and individual genes at the single-cell level [315]. Specific improvements in target hybridization strategies and oligonucleotide probe design have enabled facile fluorescence-based detection of scarce nucleic acid(s) of interest [316-319]. In **Chapter II**, we provide a detailed procedure of our innovative multiplexed fluorescence imaging methodology that allows us to simultaneously label DNA, RNA, and proteins in cell samples to microscopically visualize and monitor viral gene expression and replication across a broad spectrum of viruses. We demonstrate several examples of how we can label both DNA and RNA molecules and nucleic acid(s) of different strandedness (positive-sense (+) or antisense (-)). We have previously also shown that we can label spliced *vs.* unspliced RNA,

demonstrating how elegantly designing hybridization probes across splicing junctions can allow us to differentiate between different spliced variants [119, 317-319]. This *in-situ* imaging method is therefore a powerful tool to interrogate the transcriptional landscapes of genetic *loci* of interest.

Live-cell nucleic acid imaging to track gene expression real-time

While our presented *in-situ* hybridization protocol in **Chapter II** is broadly applicable, enabling us to investigate a diverse set of questions related to virus replication, our method relies on fixation of our cell or tissue samples. This thus allows us to take a molecular "snapshot" of a biological process in motion. However, improvements in fluorescent labeling techniques and probe sensitivity have opened the door for live-cell imaging of nucleic acid(s) [30, 94, 320-323]. Catalytically inactive Cas (dCas) protein is fused to fluorescent protein(s) and retains the ability to bind nucleic acid of interest in an RNA-guided fashion. dCas imaging systems in combination with fluorescence microscopy can be a powerful modality for tracking transcriptional dynamics, chromatin organization, and direct chromatin-chromatin interactions real-time [321-323].

Chromatin profiling methods to study epigenetic landscape

Chromatin regulatory landscapes play a critical role in influencing the cellular transcriptome and driving many important biological processes. Epigenome sequencing technologies such as chromatin immunoprecipitation with sequencing (ChIP-seq) and assay for transposase-accessible chromatin with sequencing (ATAC-seq) have enabled researchers to interrogate the chromatin landscape of cellular populations of interest on a genome-wide scale, enabling assessment of chromatin functional state and identifying transcription factors bound at key regulatory elements [324].

In ChIP-seq experiments, DNAs bound to specific proteins of interest are enriched and deep sequenced. The main steps include: 1) crosslinking DNA and proteins *in situ* with formaldehyde, 2) sonication of DNA into small (200-600 bp) fragments, 3) IP of the DNA-protein complexes of interest, and 4) reversing crosslinking, which frees DNA for sequencing adapter ligation prior to sequencing. A major limitation of conventional ChIP-seq is the high cell input for high-quality datasets (10^5-10^7 cells) [324]. Other methods such as cleavage under targets and release using nuclease (CUT&RUN) have been developed for detection of DNA-protein interactions. CUT&RUN, however, does not require formaldehyde crosslinking and sonication-based fragmentation, but instead uses MNase fused to Protein A/G to cut and release target DNA fragments *in situ*. This minimizes background and can be applied to as low as 100-1000 cells [325]. While CUT&RUN offers an alternative to ChIP-seq, transient DNA-protein interactions may be difficult to capture [325].

The precursor to ATAC-seq was DNase I hypersensitive site sequencing (DNase-seq), a method to investigate chromatin accessibility [326]. However, DNase-seq datasets often had high background noise and did not have the resolution to determine accurate transcription factor and nucleosome occupancy at regulatory regions. MNase digestion with deep sequencing (MNase-seq) was developed to probe nucleosome positioning [327]. MNase digests DNA until it encounters nucleic acid binding proteins such as transcription factors or nucleosomes. A major limitation of MNase-seq is the large amount of input material needed for library preparation, resulting in a need for a nucleosome mapping technique that is amenable for more rare and precious samples like embryotic tissue. The development of ATAC-seq successfully filled this niche, enabling higher-resolution mapping of chromatin accessibility, nucleosome positioning, and transcription factor occupancy and requiring relatively low input material (500-5000 cells) [328]. In ATAC-seq

experiments, the nuclei of cells are harvested following cellular lysis, and nuclear chromatin is fragmented by a hyperactive Tn5 transposase that ligates sequencing adapters to "open" regions of chromatin. Condensed chromatin is not accessible by Tn5. DNA is fragmented, and transposase-accessible chromatin is PCR-amplified prior to NGS preparation. Our ATAC-seq studies presented in **Chapter VI** apply the refined Omni-ATAC method to interrogate chromatin accessibility at defined sites of HIV-1 integration [329].

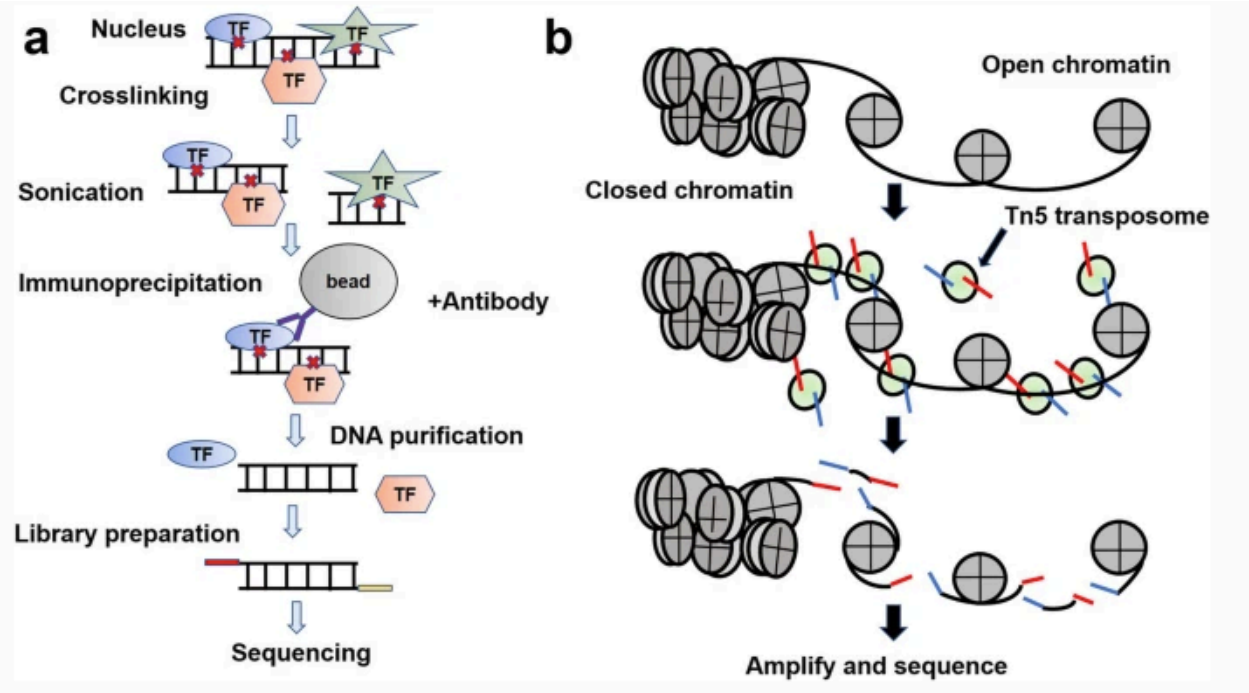


Figure I.F.3. Workflows of ChIP- & ATAC-seq. In ChIP-seq, chromatin is chemically crosslinked using formaldehyde and sonicated to obtain DNA fragments between 200-600 bp. The DNA protein complex of interest can then be immunoprecipitated. Final libraries are then deep sequenced **(A).** ATAC-seq workflow utilizes a hyperactive TN5 transposase, which preferentially integrates into 'open' chromatin and inserts sequencing adapters. Final library preparation concludes with PCR amplification of transposase-accessible sites and sequencing (**B).** This Figure is reused from Ma *et al.* [324]. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International License. This work was originally published in *Molecular Biomedicine.*

RNA-sequencing assays to capture complex transcriptomes

Compared to previous Sanger sequencing- and microarray-based methods, RNA-seq is a significantly higher throughput methodology, providing improved coverage and greater resolution of the dynamic nature of the cellular and viral transcriptomes. RNA-seq allows for quantification of gene expression, identification of novel transcripts, isoform analysis of spliced variants, and detection of allele-specific expression [330]. Advances in RNA-seq library preparation have enabled researchers to further delve into the complexity of the transcriptome. In addition to capturing and sequencing poly(A) mRNA, RNA-seq can also be applied to investigate expression of different RNA populations including total RNA, pre-mRNA, and noncoding RNA (microRNA and long ncRNA) [330]. RNA-seq has become a relatively streamlined methodology, generating very large and informative datasets and facilitating hypothesis-driven research. Integrative transcriptomic analysis with other chromatin profiling methods, cytological *in-situ* imaging, and functional studies have provided invaluable insight into important biological questions. However, a technical limitation of conventional Illumina-based RNA-seq is the acquisition of short sequence read length, which makes it difficult to accurately map and identify long and complex RNA species.

To circumvent this limitation of conventional RNA-seq, long-read sequencing platforms have been developed. Oxford Nanopore Technologies have launched a portable sequencer which offers the ability to sequence very long reads. Nanopore sequencers function by monitoring changes to an electrical current as nucleic acids are passed through a protein nanopore. The resulting signal is deconvolved to produce a DNA or RNA chromatogram [247, 331-335]. A recent study collected vRNA from HIV-infected cells at multiple time-points and applied long-read Nanopore RNA-seq to demonstrate that this technology can be used to robustly track HIV-1

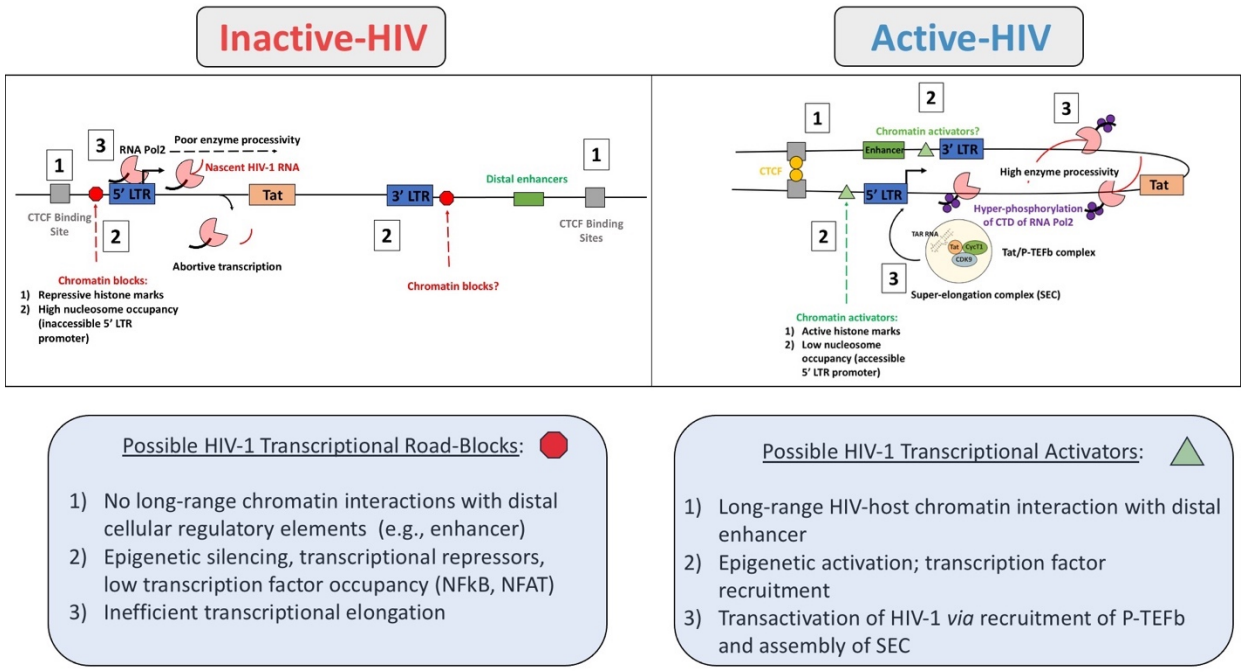
splicing dynamics [247]. In **Chapter VI**, we apply Nanopore sequencing to characterize chimeric HIV-host transcripts that result from transcriptional read-through events and assess how HIV-1 integration and proviral activation affect local cellular splicing patterns.

1-F-c: Association between Chromatin Organization & Virus Replication

The association between nuclear organization and viral gene expression has been demonstrated for several viruses. Nuclear positioning, host regulatory elements, and higher-order nuclear spatial compartmentalization directly influence viral replication efficacy for viruses that require nuclear entry for their life cycles. The Pintel group has shown that the genome of parvovirus Minute Virus of Mice (MVM) associates significantly with previously identified cellular TADs, suggesting that some level of nuclear hierarchical organization may be influencing MVM transcription[309]. Virus-host chromatin interactions have also been shown to be important for regulating viral transcription in Kaposi's Sarcoma-Associated Herpesvirus (KSHV) [336] and Human T lymphotropic virus 1 (HTLV-1) [337], with the presence of viral DNA altering local cellular chromatin structure and host transcriptional patterns. The Bangham group has shown that HTLV-1 has an inherent CTCF-binding site that can mediate differential chromatin looping patterns with the host genome around HTLV-1 retroviral sites of integration [338]. CTCF-mediated looping at HTLV-1 sites of integration can induce dysregulation of local RNA landscape [337-339].

Several important questions remain regarding how HIV can alter host chromatin organization. Using a targeted 3D chromatin mapping technique, 4C-sequencing, Dieudonne *et al.* demonstrated that long-range chromatin interactions between integrated HIV and the host genome change in response to proviral transcriptional activation in the J-Lat A1 cell model [171]. This result suggested higher-order chromatin reorganization, leading to large scale remodeling of the HIV

chromatin contact network, may be important for proviral transcriptional activation. Other more recent studies have demonstrated that assessing the local functional chromatin state surrounding the provirus may be a predictor of latency establishment and proviral transcriptional competence [32, 219]. Host factors (LEDGF/p75, CPSF6) that affect integration site selection and penetration of the PIC into the nucleus have also been implicated in influencing the status of latency and proviral gene expression [172]. HIV PICs preferentially target higher-order nuclear structures such as SPADs [94] and SEs [170], linking specific chromatin features to integration site selection. We, however, still do not have a clear grasp of how HIV-1, once integrated, can alter the local chromatin environment. Further elucidation of the determinants of HIV-1 transcriptional activity can be invaluable for HIV latency research and may be useful for implementing more effective lentiviral-based gene therapies. **Figure I.F.4** depicts possible HIV-1 transcriptional "road-blocks", as well as possible transcriptional activators that are explored in in this dissertation. Critical *cis*- and *trans*-acting factors regulating HIV-1 transcription have been identified. However, many questions remain regarding how long-range chromatin remodeling may influence assembly of stable transcription complexes at the HIV-1 5' promoter. **Figure I.F.4** presents three possible HIV-1 transcriptional "road-blocks", as well as possible activators of proviral transcription explored in more detail in **Chapter VI**.



Possible HIV-1 Transcriptional Road-Blocks:

1) No long-range chromatin interactions with distal cellular regulatory elements (e.g., enhancer)
2) Epigenetic silencing, transcriptional repressors, low transcription factor occupancy (NFkB, NFAT)
3) Inefficient transcriptional elongation

Possible HIV-1 Transcriptional Activators:

1) Long-range HIV-host chromatin interaction with distal enhancer
2) Epigenetic activation; transcription factor recruitment
3) Transactivation of HIV-1 *via* recruitment of P-TEFb and assembly of SEC

Figure I.F.4. HIV-1 transcriptional regulation and Tat transactivation of proviral transcription. Transcriptionally repression of HIV-1 (inactive) may be influenced by a lack of 1) long-range chromatin interactions that bring an enhancer close to the HIV-1 5' LTR promoter, 2) chromatin "blocks" such as inaccessible chromatin or repressive histone modifications at the 5' LTR, and 3) inefficient transcriptional elongation. HIV-1 may be transcriptionally activated due to 1) long-range HIV0host chromatin interactions with a cellular enhancer, 2) active epigenetic modifications or accessible HIV-1 chromatin, and 3) transactivation of HIV-1 *via* recruitment of P-TEFb.

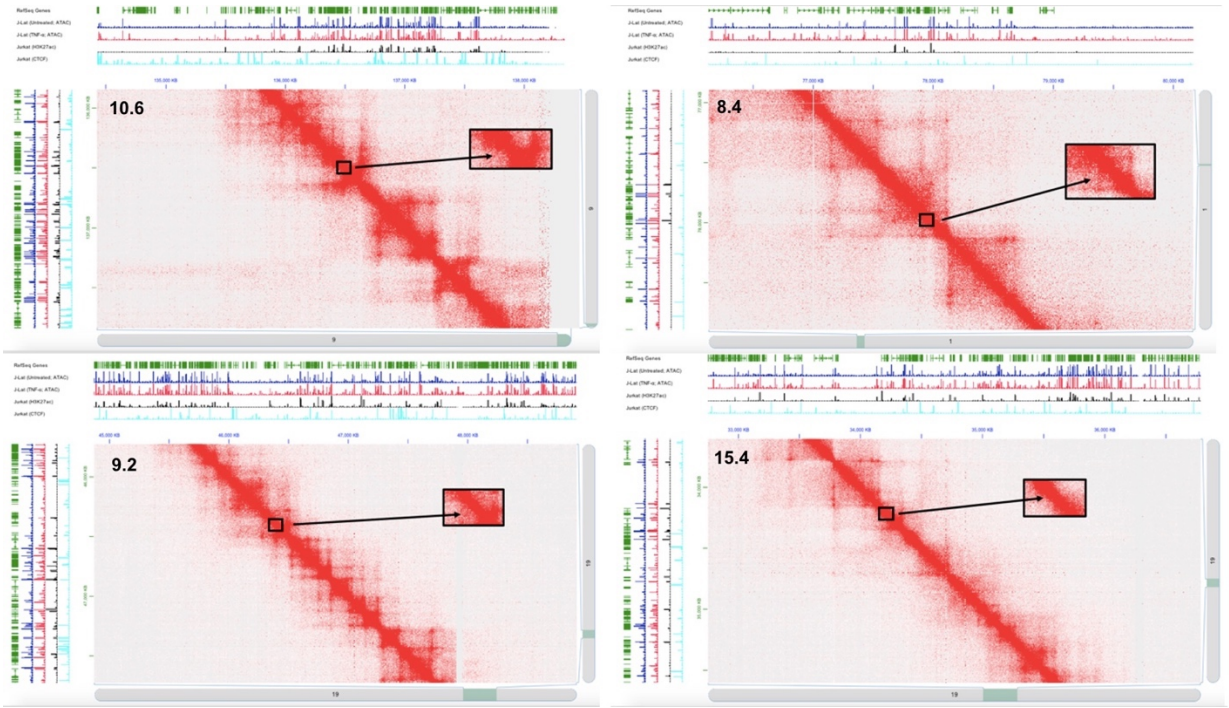


Figure I.F.5. Jurkat T cell Hi-C contact map at site of integration across J-Lat models (10.6, 8.4, 9.2, & 15.4). Chromatin contact network at the *SEC16A* gene within the 10.6 clone **(A)** 8.4 contact network **(B).** 9.2 **(C)** 15.4 **(D)**. All proviral integration sites are within TADs, proximal to CTCF-BS. Tracks in green above Hi-C maps are the reference genome (Refseq), blue and red are ATAC-seq tracks of Jurkats +/- TNF treatment. The black colored track is H3K27ac enrichment, and CTCF occupancy is shown in cyan. All sequencing maps are from Jurkat T cells. Hi-C map was processed using Juicebox and visualized on Juicer [340]. Hi-C data downloaded from the Gene Expression Omnibus (GEO): GSE122958.

1-G: Dissertation Direction

Multiplexed profiling of the chromatin and transcriptional landscape of virus-infected cells using deep sequencing and *in-situ* imaging approaches can provide insight into critical virus-host factor interactions that are regulating viral transcriptional control and replication. Within this dissertation, I present studies utilizing complementary NGS-based technologies and FISH-based imaging to study a diverse set of questions pertinent to viral gene regulation.

Chapter I presents a thorough background on HIV-1 replication and the current status of HIV antiviral therapies, focusing on factors that influence HIV-1 transcriptional competency (integration site selection and nuclear organization). In addition, **Chapter I** provides background on our current understanding of eukaryotic chromatin organization and methods to profile chromatin structure and RNA transcription. Collectively, this chapter aims to demonstrate the interplay between nuclear organization, critical virus-host factor interactions, and HIV-1 gene expression.

Chapter II focuses on the development and optimization of a single-cell multiplexed fluorescence imaging approach to simultaneously visualize viral nucleic acid(s) and protein(s) during the course of viral infection. Applications of fluorescence-based methods to visualize viral replication have been hindered by the fact that conventional DNA and RNA FISH have not been compatible with immunostaining, prohibiting multiplexed labeling of viral nucleic acid(s) and protein(s). Tracking synthesis of vRNA and protein is important for monitoring critical stages of virus replication and understanding what factors influence viral transcription. Microscopy-based approaches have the added advantage of interrogating single cells, allowing researchers to observe the variability inherently present within an infected cellular population, as well enabling visualization of the spatial cellular compartmentalization of critical virus-host

factor interactions. Work in the Sarafianos group has led to the development of an innovative platform for multiplexed fluorescence imaging, which we have "coined" **m**ultiplexed **i**mmunofluorescent **c**ell-based **d**etection of **D**NA, **R**NA, and **p**rotein (MICDDRP). We provide a streamlined step-by-step protocol for performing MICDDRP to simultaneously fluorescently label viral nucleic acid (DNA & RNA of different strandedness) and protein across a broad spectrum of viruses. In **Chapters III-V,** we provide additional examples of applications of MICDDRP to investigate HIV-host factor interactions regulating transcription from unintegrated viral DNA templates (**Chapter III**), effects of small molecules on HIV-1 transcription (**Chapter IV**), and visualization of SARS-CoV-2 early replication kinetics (**Chapter V**). All of these applications collectively highlight the robustness and broad applicability of our multiplexed fluorescence imaging modality.

In **Chapter VI,** we apply chromatin profiling methods such as ATAC-seq and Hi-C to investigate differential chromatin accessibility and HIV chromatin 3D organization, respectively, around integrated proviruses in HIV-inducible cellular models. Using these inducible cell models provides us control over HIV-1 transcription (inactive *vs.* active). In addition, we apply RNA-seq to interrogate gene expression at HIV-host gene boundaries. We implement an innovative long-read Nanopore sequencing approach to improve our mapping of the HIV transcriptome and understand how HIV-1 integration can affect the local RNA landscape. We aim to understand how lentiviral integration, in general, may alter chromatin structure and dysregulate cellular transcription.

To summarize, we present future directions where this work serves as a foundation for multiplexed sequencing and imaging studies in primary cell models using single-cell sequencing approaches. Our key contributions to the field include improving our understanding

of how lentiviral integration can alter chromatin organization and methodological advancements to probe the HIV-host transcriptomes. Our applications of various NGS-based assays also provide a pipeline for further in-depth interrogation of the association between chromatin architecture and viral gene expression. Our work can be expanded to a broad spectrum of viruses that exploit the nuclear machinery and are interacting with cellular chromatin to influence virus-host gene expression patterns. In addition, our findings implicate putative biological pathways involved in HIV-1 transcriptional control. Future mechanistic studies will seek to further understand how virus-host interactions influence HIV-host functional chromatin state.

CHAPTER II: SINGLE-CELL MULTIPLEXED FLUORESCENCE IMAGING TO VISUALIZE VIRAL NUCLEIC ACID & PROTEINS TO MONITOR VIRAL REPLICATION

This research and presented protocol for multiplexed imaging is published in *Journal of Visualized Experiments.*

Chapter II is adapted from:

Shah, R., Lan, S., Puray-Chavez, M. N., Liu, D., Tedbury, P. R., Sarafianos, S. G. Single-cell Multiplexed Fluorescence Imaging to Visualize Viral Nucleic Acids and Proteins and Monitor HIV, HTLV, HBV, HCV, Zika Virus, and Influenza Infection. *J. Vis. Exp.* (164), e61843, doi:10.3791/61843 (2020)."

Copyright © MyJove Corp.

SUMMARY

Presented here is a protocol for a fluorescence imaging approach, **m**ultiplex **i**mmunofluorescent **c**ell-based **d**etection of **D**NA, **R**NA, and **p**rotein (MICDDRP), a method capable of simultaneous fluorescence single-cell visualization of viral protein and nucleic acids of different type and strandedness. This approach can be applied to a diverse range of systems.

ABSTRACT

Capturing the dynamic replication and assembly processes of viruses has been hindered by the lack of robust *in situ* hybridization (ISH) technologies that enable sensitive and simultaneous labeling of viral nucleic acid and protein. Conventional DNA fluorescence *in situ* hybridization (FISH) methods are often not compatible with immunostaining. We have therefore developed an imaging approach, **MICDDRP** (**m**ultiplex **i**mmunofluorescent **c**ell-based **d**etection of **D**NA, **R**NA and **p**rotein), which enables simultaneous single-cell visualization of DNA, RNA, and protein. Compared to conventional DNA FISH, MICDDRP utilizes branched DNA (bDNA) ISH technology, which dramatically improves oligonucleotide probe sensitivity and detection. Small modifications of MICDDRP enable imaging of viral proteins concomitantly with nucleic acids (RNA or DNA) of different strandedness. We have applied these protocols to study the life cycles of multiple viral pathogens, including human immunodeficiency virus (HIV)-1, human T-lymphotropic virus (HTLV)-1, hepatitis B virus (HBV), hepatitis C virus (HCV), Zika virus (ZKV), and influenza A virus (IAV). We demonstrated that we can efficiently label viral nucleic acids and proteins across a diverse range of viruses. These studies can provide us with improved mechanistic understanding of multiple viral systems, and in addition, serve as a template for

application of multiplexed fluorescence imaging of DNA, RNA, and protein across a broad spectrum of cellular systems.

INTRODUCTION

While thousands of commercial antibodies are available to specifically label proteins *via* conventional immunostaining approaches, and while fusion proteins can be engineered with photo-optimized fluorescent tags for tracking multiple proteins in a sample [341], microscopic visualization of protein is often not compatible with conventional DNA fluorescence *in situ* hybridization (FISH) [169]. Technical limitations in simultaneous visualization of DNA, RNA, and protein using fluorescence-based approaches have hindered in-depth understanding of virus replication. Tracking both viral nucleic acid and protein during the course of infection allows virologists to visualize fundamental processes that underly virus replication and assembly [158, 318, 319, 342].

We have developed an imaging approach, **m**ultiplex **i**mmunofluorescent **c**ell-based **d**etection of **D**NA, **R**NA, and **P**rotein (MICDDRP) [318], which utilizes branched DNA (bDNA) *in situ* technology to improve the sensitivity of nucleic acid detection [316, 343, 344]. In addition, this method utilizes paired probes for enhanced specificity. bDNA sequence-specific probes use branching preamplifier and amplifier DNAs to produce an intense and localized signal, improving upon previous hybridization methods that relied on targeting repeated regions in the DNA [344]. Infected cells in a clinical context often do not contain abundant viral genetic material, providing a commodity for a sensitive method for fluorescent nucleic acid detection in diagnostic settings. The commercialization of bDNA technology through approaches such as RNAscope [316] and

ViewRNA [345] have filled this niche. The sensitivity of bDNA fluorescence imaging also has important utility in cell biology, allowing detection of scarce nucleic acid species in cell culture models. The vast improvement of sensitivity makes bDNA-based imaging methods suitable for studying viruses. A potential shortcoming, however, is that these methods focus on visualizing RNA or RNA and protein. All replicating cells and many viruses have DNA genomes or form DNA during their replication cycle, making methods capable of imaging both RNA and DNA, as well as protein, highly desirable.

In our MICDDRP protocol, we perform bDNA FISH for detection of viral nucleic acid using the RNAscope method, with modifications [316]. One of our major modifications to this protocol is optimization of protease treatment following chemical fixation. Protease treatment facilitates removal of proteins bound to nucleic acid to improve probe hybridization efficiency. Protease treatment is followed by incubation with branched oligonucleotide probe(s). After application of bDNA probe(s), samples are washed and subsequently incubated with signal pre-amplifier and amplifier DNAs. Multiplexed *in-situ* hybridization (ISH), labeling of multiple gene targets, requires target probes with different color channels for spectral differentiation [316]. Incubation with DNA amplifiers is followed by immunofluorescence (IF). Finding the optimal protease conditions are necessary to ensure efficient probe hybridization, while still preserving target epitopes for immunostaining.

bDNA ISH imparts improvements in signal-to-noise by amplification of target-specific signals, with a reduction to background noise from non-specific hybridization events [316, 346]. Target probes are designed using software programs publicly available that predict the probability of

non-specific hybridization events, as well as calculate melting temperature (T_m) of the probe-target hybrid [316, 346]. Target probes contain an 18- to 25-base region complementary to the target DNA/RNA sequence, a spacer sequence, and a 14-base tail sequence. A pair of target probes, each with a distinct tail sequence, hybridize to a target region (spanning ~50 bases). The two tail sequences form a hybridization site for the pre-amplifier probes, which contain 20 binding sites for the amplifier probes, which, in addition contain 20 binding sites for the label probe. As an example, a one kilobase (kb) region on the nucleic acid molecule is targeted by 20 probe pairs, creating a molecular scaffold for sequential hybridization with the preamplifier, amplifier, and label probe. This can thus lead to a theoretical yield of 8000 fluorescent labels per nucleic acid molecule, enabling detection of single molecules and vast improvements over conventional FISH approaches [316] (See **Figure II.1** for schematic of bDNA signal amplification). To set up probes for multiplexed ISH, each target probe must be in a different color channel (C1, C2, or C3). These target probes with different color channels possess distinct 14-base tail sequences. These tail sequences will bind distinct signal amplifiers with different fluorescent probes, thus enabling facile spectral differentiation across multiple targets. In our presented **Protocol**, **Table II.4** in **Step 9,** provides further information on fluorescently labeling target probes. In addition, **Figures 2-3** provide examples of how we chose the appropriate Amplifier 4-FL (A, B, or C) (fluorescent probe & the final hybridization step) to achieve specific fluorescent labeling of multiple viral nucleic acid targets following HIV-1 and HTLV-1 infections.

We have demonstrated several applications of simultaneous fluorescence visualization of RNA, DNA, and proteins, observing critical stages of virus replication with high spatiotemporal resolution [318, 319, 342]. For example, simultaneous single-cell visualization of vRNA, cytoplasmic

and nuclear DNA, and protein have allowed us to visualize key events during HIV-1 infection, including following RNA containing cores in the cytoplasm prior to nuclear entry and integration of proviral DNA [318]. In addition, we have applied MICDDRP to characterize the effects of host factors and drug treatment on viral infection and replication [319, 342]. In Ukah *et al.* 2018, we tracked reactivation of HIV-1 transcription in latency cell models treated with different latency-reversing agents to visualize HIV transcription and latency reversal [319]. In addition, MICDDRP can allow us to visualize phenotypic changes associated with antiviral inhibition attributed to small molecule treatment or host factor restriction. As a proof of concept to the robustness and broad applicability of our approach, we have demonstrated that we can use modifications of our protocol to efficiently label viral nucleic acid to follow infection not only in human immunodeficiency virus (HIV)-1, but also human T-lymphotropic virus (HTLV)-1, hepatitis B virus (HBV), hepatitis C virus (HCV), Zika virus (ZIKV), and influenza A virus (IAV). As the HIV-1 life cycle consists of both viral DNA and RNA species, we have performed the majority of our optimization of MICDDRP following HIV-1 replication kinetics. However, in addition, we have demonstrated that we can track synthesis of different vRNA transcripts of either or both sense (+) and antisense (-) strandedness in viruses such as ZIKV, IAV, HBV, and HCV to monitor viral transcription and replication [3-6]. Our studies aim to improve our mechanistic understanding of several viral processes and serve as a guideline to implement this fluorescence imaging technology to a broad range of cellular models.

STEP-BY-STEP PROTOCOL FOR MULTIPLEXED IMAGING

1. Seed cells (Suspension *vs.* Adherent Cells) on coverslips or chamber slides (protocol presented uses coverslips).

1. Seeding of suspension cells
    1. Prepare poly-D-lysine (PDL) coated coverslips (facilitates adherence of suspension cells to coverslip) by first incubating coverslips in ethanol (EtOH) for 5 minutes (sterilizes coverslips and removes any residue). Then wash 2x in phosphate buffered saline (PBS) before incubating coverslips in PDL (20 µg/mL) for 30 minutes (min) at room temperature (RT).
    2. Remove PDL and wash 2x in PBS. Pellet cells (previously infected/treated based on desired imaging conditions) and resuspend 10^6cells in 50 µL of PBS. Spot 50 µL of cells on glass (PDL-coated) and incubate at RT for 30 min.

2. Seeding of adherent cells
    1. Culture cells on sterile coverslip placed in 6-well dish and infect cells with viral particles or treat with compound of interest. Allow cells to reach 50-70% confluency prior to sample preparation for imaging experiments.

2. Cellular fixation: To preserve cellular morphology for fluorescence imaging studies

NOTE: Keep 4% PFA and PBS at RT for at least 30 minutes before cellular fixation.

1. Aspirate the cellular media, wash cells on coverslips 3x in PBS, and fix cells in 4% paraformaldehyde (PFA) for 30 min. Aspirate PFA and wash cells in PBS 2x.
   NOTE: Cells can now be dehydrated and stored, if the experimenter wishes to resume the experiment on another date. Fixed cells can be dehydrated in EtOH prior to storage.

    1. Remove PBS following wash after cellular fixation and replace with 500 µL of 50% EtOH (v/v in water). Incubate at RT for 5 min.

    2. Remove 50% EtOH and replace with 500 µL of 70% EtOH. Incubate at RT for 5 min.

    3. Remove 70% EtOH and replace with 500 µL of 100% EtOH. Incubate at RT for 5 min.
    4. Remove 100% EtOH and replace with fresh 100% EtOH.

NOTE: Dehydrated cells can be stored at -20 °C for 6 months. Seal plates with tape or parafilm prior to storage to prevent evaporation of EtOH.

5. Rehydrate cells to move onto multiplexed fluorescence labeling of cells/virus.

6. Remove 100% EtOH and replace with 500 µL of 70% EtOH. Incubate at RT for 2 min.

NOTE: Do not let cells dry out at any time. Always use enough solution to submerge all the cells.

7. Remove 70% EtOH and replace with 500 µL of 50% EtOH. Incubate at RT for 2 min.

8. Remove 50% EtOH and replace with 1x PBS.

NOTE: Prepare protease dilution, hybridization probes, and wash buffer in advance of protease treatment (**Step 4)** and target probe application (**Step 5**). Specific instructions on reagent preparation are presented at the beginning of each respective step.

TABLE II.1. Key reagents in MICDDRP protocol.

Reagents	Other Notes
Protease solution	Prepare in 1X PBS
Target oligonucleotide probe(s)	Dilute in hybridization buffer
Wash Buffer	For hybridization steps
Hybridization buffer	Recipe presented in **Table II.3**

3. Cell permeabilization: Increases access into the cell and cellular organelles for entry of large molecules (antibodies, nucleic acid hybridization probes)

1. Remove 1x PBS following cellular fixation or rehydration and replace with 500 µL of 0.1% (v/v) Tween-20 in 1x PBS. Incubate at RT for 10 min. Replace one by one. Wash once with 500 µL of 1x PBS and add fresh 1x PBS.

4. *Coverslip immobilization* on glass slide and *protease treatment* to remove nucleic acid binding proteins from fixed viral/cellular nucleic acid to improve hybridization efficiency

NOTE: Prepare the diluted Protease III (see specifics of reagent in **Table of Materials**) solution during cellular permeabilization. Let protease reach RT for 10 min before permeabilization.

1. Place a small drop of nail polish on a sterilized glass slide. Dry back (side with no cell layer) and place the edge of coverslip on the nail polish drop, with the side with the adhered cells facing upwards. Add few drops of PBS on the immobilized coverslip to prevent drying.

1. Draw a circle (about 3 mm away from the coverslip) around the perimeter of the coverslip now adhered to the slide using hydrophobic barrier pen (water-repellant pen that keeps reagents localized on cells).

2. Dilute Protease III in 1x PBS (100 µL/coverslip).

   NOTE: Protease concentration may need to be adjusted depending on differences in cell types, probes, or target nucleic acid(s) through empirical optimization (See **Discussion**). For most efficient RNA/DNA labeling across different viral systems, we had success with the following dilutions presented in **Table II.2** below with dilutions ranging from (1 to 2)-(1 to 15) (protease to 1x PBS).

TABLE II.2. Protease III dilutions in PBS for viral nucleic acid hybridization.

Probe Targets	Protease Dilution in 1X PBS (Protease III to 1X PBS)
HIV-1 DNA/RNA	1 to 5
HTLV-1 DNA/RNA	1 to 5
HBV pgRNA and total HBV RNA	1 to 15
IAV RNA	1 to 15
ZIKV RNA	1 to 2

3. Decant the 1x PBS on the coverslip following immobilization and apply the diluted Protease III.

4. Incubate in a humidified oven at 40 °C for 15 min.

5. Decant protease solution and submerge slides in 1x PBS. Agitate with a rocking dish for 2 min at RT. Repeat wash with new 1x PBS.

    1. For **only** DNA detection, wash samples three times with nuclease-free water for 2 min each, followed by incubation with 5 mg/mL RNase A diluted in PBS for 30 min at 37 °C.

    2. Decant RNase A solution, and wash 3x for 2 min with ultrapure water. Continue with ISH of target probe(s).
    NOTE: Hybridization buffer improves vDNA detection without affecting vRNA staining efficiency for the results presented (**Representative Results**). Dilute DNA Channel 1 (C1) probes 1:1 with hybridization probe. Dilute RNA C2 and C3 probes in hybridization buffer. The C2 and C3 probes used in our imaging studies are in 50X solutions (1:50; target probe **to** hybridization buffer).

6. Prepare hybridization buffer in nuclease-free water following the step-by-step procedure below:

    1. In a 15 mL tube, add 700 µL of nuclease free-water, 300 µL of 50% (weight/volume (w/v)) dextran sulfate, 300 µL of 5 M NaCl, 125 µL of 200 mM sodium citrate (pH 6.2), and 375 mg (powder) of ethylene carbonate.

    2. Mix well using a vortex to dissolve ethylene carbonate (ensure all powder is dissolved and clear solution).

    3. Add 25 µL of 10% (volume/volume (v/v)) Tween-20 and enough nuclease-free water to complete 2.5 mL (2x solution).
    NOTE: The recipe for the hybridization buffer presented can be found in **Table 3** below. Solution is stable for a week. Ensure sufficient mixing of Tween-20 detergent, while being careful to prevent bubbling of solution.

TABLE II.3. Hybridization buffer list of reagents.

Reagent	Stock Concentration	Additional Notes
Nuclease-free water	NA	
Dextran sulfate	50% (w/v)	Very viscous
Sodium chloride	5 M	
Sodium citrate (pH 6.2)	200 mM	Store at 4°C
Ethylene carbonate	NA	Powder
Tween-20	10% (v/v)	

5. Incubation with DNA/RNA target hybridization probes: Target oligonucleotide probes bind to region(s) of interest, creating a molecular scaffold for pre-amplifiers, amplifiers, and fluorescent probes to bind.

NOTE: Warm DNA/RNA probes at 40 °C for 10 min (during Cell Permeabilization) and cool down to RT for at least 10 min, if no RNase-treatment is included. If RNase-treatment or further sample treatment is needed prior to target probe hybridization, warm probes accordingly. Spin down C2 and C3 probes after warming and dilute in hybridization buffer. After dilution, C2 and C3 probes can be briefly warmed. Warm 50x wash buffer (See **Table of Materials** for more detail) at 40 °C for 10-20 min and dilute to 1x in molecular biology grade water.

1. Incubate 200 µL of the hybridization buffer at 67 °C for 10 min prior to addition of hybridization probe(s). Dilute probe in hybridization buffer (recipe listed in **Table II.2**). Add 50 µL/coverslip.

2. Incubate in humidified oven at 40 °C for 2 hours (h). Decant probes and submerge slides in 1x wash buffer. Agitate by rocking dish for 2 min at RT. Repeat wash with new 1x wash buffer.

6. Amplifier (Amp) 1-FL Hybridization: Addition of pre-amplifier that is complementary to the tail sequence of the target DNA/RNA probes (Step 5)

NOTE: Amplifiers should be at RT before use. Get each individual amplifier out of the fridge 30 min before use and leave on the bench at RT.

1. Remove slides from 1x wash buffer and tap/absorb to remove excess liquid.

2. Add 1 drop of Amp 1-FL on the coverslip. Incubate in humidified oven at 40 °C for 30 min.
3. Decant Amp 1-FL and submerge in 1x wash buffer. Agitate by rocking dish 2 min at RT. Repeat wash with new 1x wash buffer.

7. Amp 2-FL Hybridization: Incubation with signal amplifier with cognate recognition sequence to pre-amplifiers (Amp 1-FL)

1. Remove slides from 1x wash buffer and tap/absorb to remove excess liquid.

2. Add 1 drop of Amp 2-FL on the coverslip. Incubate in humidified oven at 40 °C for 15 min.

3. Decant Amp 2-FL and submerge in 1x wash buffer. Agitate by rocking dish 2 min at RT. Repeat wash with new 1x wash buffer.

8. Amp 3-FL Hybridization: Incubation with second signal amplifier

1. Remove slides from 1x wash buffer and tap/absorb to remove excess liquid.

2. Add 1 drop of Amp 3-FL on the coverslip. Incubate in humidified oven at 40 °C for 30 min.

3. Decant Amp 3-FL and submerge in 1x wash buffer. Agitate by rocking dish 2 min at RT. Repeat wash with new 1x wash buffer.

9. Amp 4-FL Hybridization: Fluorescent label and final hybridization step

NOTE: First, see **Table II.4** to choose the suitable Amp 4 (A,B, or C)- FL based on the channels of the target probe(s). Assess what Amp 4-FL is needed to label DNA/RNA of interest. An example is provided by the table below:

TABLE II.4. Selection of Amp 4 (A, B, or C)-FL fluorescent probe for multiplexed ISH.

	A	**B**	**C**
DNA Channel 1 (C1)	488	550	550
DNA/RNA Channel 2 (C2)	550	488	647
RNA Channel 3 (C3)	647	647	488

NOTE: For multiplexed FISH (multiple targets), choosing the correct Amp 4 (A, B, or C)- FL is critical for properly labeling your target of interest(s). Target probes (Step 5) have different color channels (C1, C2, or C3), which dictate their respective fluorescent label (Alexa 488, Atto 550, or Alexa 647), based on the Amp 4-FL chosen. Examples of choosing fluorophore combinations for multiplexed imaging are provided in the legends of **Figure II.2** and **Figure II.3**. As an additional example, selection of Amp 4B-FL will selectively label DNA C1 probes with Atto 550 and RNA C3 probes with Alexa 647.

1. Remove slides from 1x wash buffer and tap/absorb to remove excess liquid.

2. Add 1 drop of Amp 4-FL on the coverslip. Incubate in humidified oven at 40 °C for 15 min.

   NOTE: Following step 9.2, keep samples covered, protected from the light.
3. Decant Amp 4-FL and submerge in 1x wash buffer. Agitate by rocking dish for 2 min at RT. Repeat wash with new 1x wash buffer. Wash with 1x PBS (2 min) and store in PBS.

10. Protein immunostaining: To label protein(s) of interest

    1. Decant PBS and add 200 µL of blocking buffer (1% w/v BSA, 10% v/v FBS in PBS with 0.1% v/v Tween-20 (PBST)) to the coverslip. Incubate 1 h at RT.

    2. Decant blocking buffer and apply 200 µL of primary antibody diluted in PBST + 1% w/v BSA. Incubate 1 h at RT.

    3. Wash the slide twice with PBST for 10 min at RT with shaking.

    4. Apply secondary antibody of choice for 1 h at RT in PBST + 1% w/v BSA.

    5. Wash the slide with PBST for 10 min at RT with shaking.

11. Nuclear staining: Counter-stain nuclei following immunostaining

    1. Decant PBST and apply DAPI or nuclear stain of choice for 1 min at RT.

    2. Wash the slide twice with PBS for 10 min at RT with shaking.

12. Mounting

    1. Place 1 drop of antifade solution (e.g., Prolong Gold) on new sterile glass slide (First, clean slide with EtOH and let dry to ensure no residues are on glass). With the same tip, spread the antifade solution drop to cover an area approximately the size of the coverslip.

NOTE: The antifade solution is very viscous and may be difficult to pipette. Cutting the tip off a 200 µL tip prior to pipetting may mitigate these issues.

2. Remove coverslip with cell sample from the slide and submerge in PBS to remove residual nail polish at the back using the forceps and PBS. Dry forceps and back of the coverslip using a Kimwipe.

3. Gently imbed coverslip in the drop of antifade solution, placing sample side (side with cell layer) of coverslip on drop).

4. Let samples dry overnight.

13. Imaging

1. Image with an epifluorescent microscope.

REPRESENTATIVE RESULTS

A schematic of MICDDRP is depicted in **Figure II.1**. Labeling of DNA and RNA is followed by immunostaining. The use of branching amplifiers increases signal, allowing detection of single nucleic acid molecules.

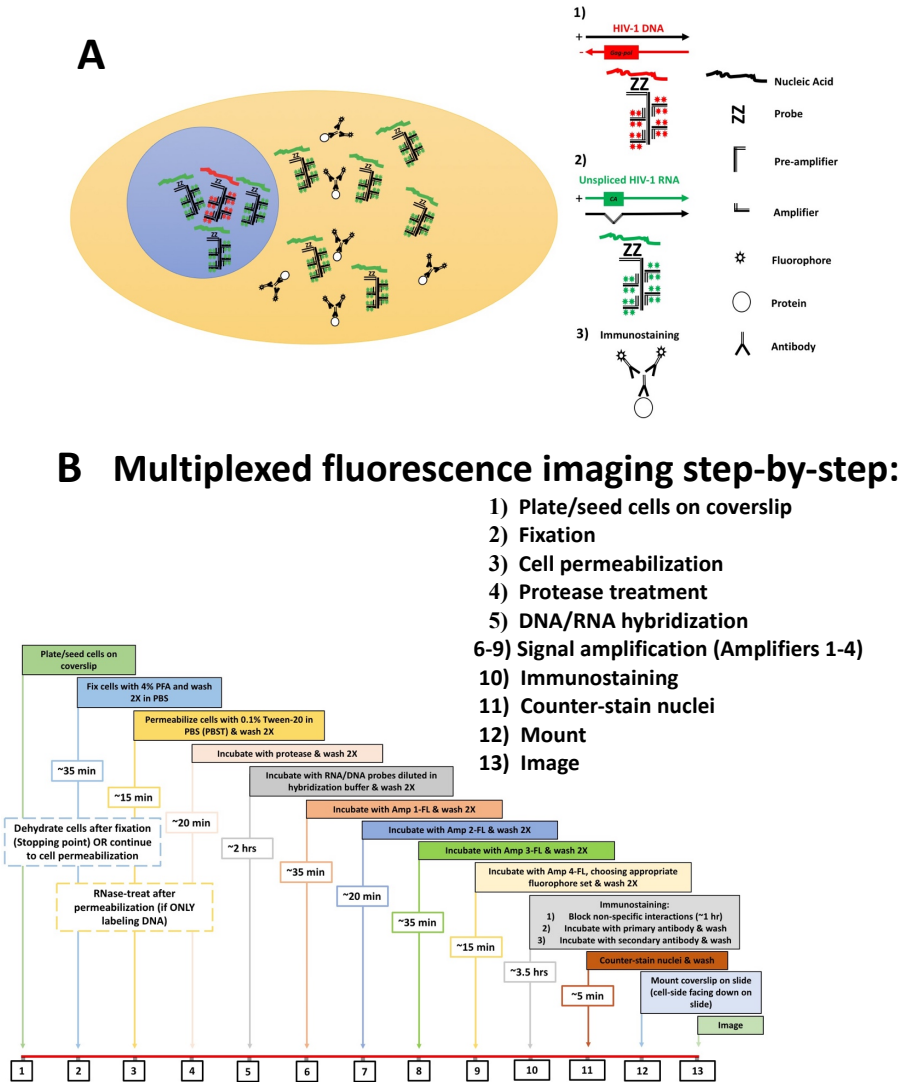


Figure II.1. Schematic of MICDDRP and step-by-step workflow. (A) bDNA signal is amplified *via* branching preamplifier and amplifier DNAs to enhance detection of viral DNA (1) and RNA species (2). Oligonucleotide probes are hybridized in pairs (ZZ in schematic) to target(s) of interest, creating a scaffold for pre-amplifier (Amp 1-FL, **Step 6** in **Protocol**), amplifier (Amp 2- & 3-FL), and fluorescent probes (Amp 4, **Step 9** in **Protocol**). Consult **Table II.4** for choosing appropriate Amp 4 (A,B, or C)-FL for multiplexed ISH. Labeling of nucleic acid is followed by immunostaining proteins of interest (3). (B) Thirteen main steps in MICDDRP protocol with estimation of time duration for each respective step.

Application of MICDDRP to study the course of HIV-1 infection has been a useful tool in tracking viral replication kinetics in primary cells. As a proof of concept of this procedure, HIV-1 DNA, RNA, and protein are simultaneously labeled and visualized microscopically at the single-cell level (**Figure II.2**). Two HIV-1 DNA genomes are visualized in a single cell, as they are actively transcribing viral RNA (vRNA). vRNA has been exported through the nuclear pore complex and viral protein is synthesized in the cytoplasm.

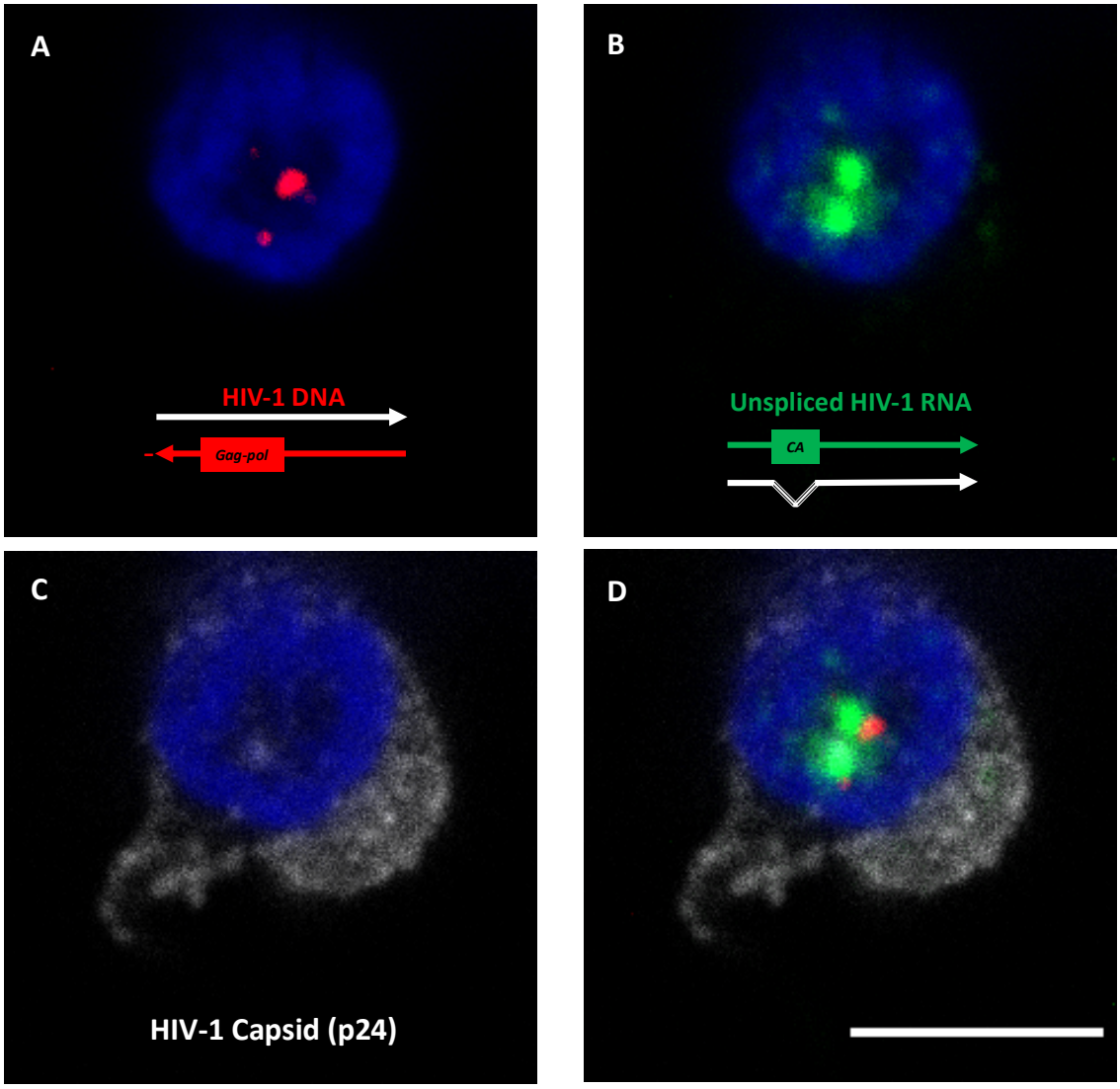


Figure II.2. MICDDRP of primary blood mononuclear cells (PMBCs) infected with HIV-1. PMBCs infected with HIV-1 (NL4.3) at a multiplicity of infection (MOI) of 2. Cells were fixed 48 hours post infection (hpi). (A) HIV-1 bDNA FISH probe (labeled with ATTO 550; red in Figure). This probe hybridizes to the template (3'<—5') vDNA strand to prevent cross-talk with sense (+) vRNA. (B) Unspliced HIV-1 RNA (labeled with Alexa 647; green in Figure). The HIV-1 vRNA probe hybridizes to viral transcripts transcribed in the 5'—>3' orientation. (C) Immunostaining of HIV-1 capsid (p24) protein (secondary antibody conjugated to Alexa 488; white in figure). (D) Merged image. Scale bar represents 5 µm. Nuclei were stained with DAPI (blue). To label HIV-1 DNA (DNA C1 probe) with ATTO 550 and HIV-1 RNA (RNA C3 probe) with Alexa 647, respectively, samples were incubated with Amp 4-FL '**B**' (Consult **Table II.4** in **Protocol** to choose appropriate channel colors for hybridization probes).

In addition, we have performed dual viral DNA (vDNA) and vRNA staining to follow HTLV-1 infection. For optimization of our nucleic acid labeling, we adhered closely to our vDNA/VRNA staining procedure developed for multiplexed fluorescence imaging of HIV infection. We have demonstrated that we can specifically label HTLV-1 DNA and RNA simultaneously.

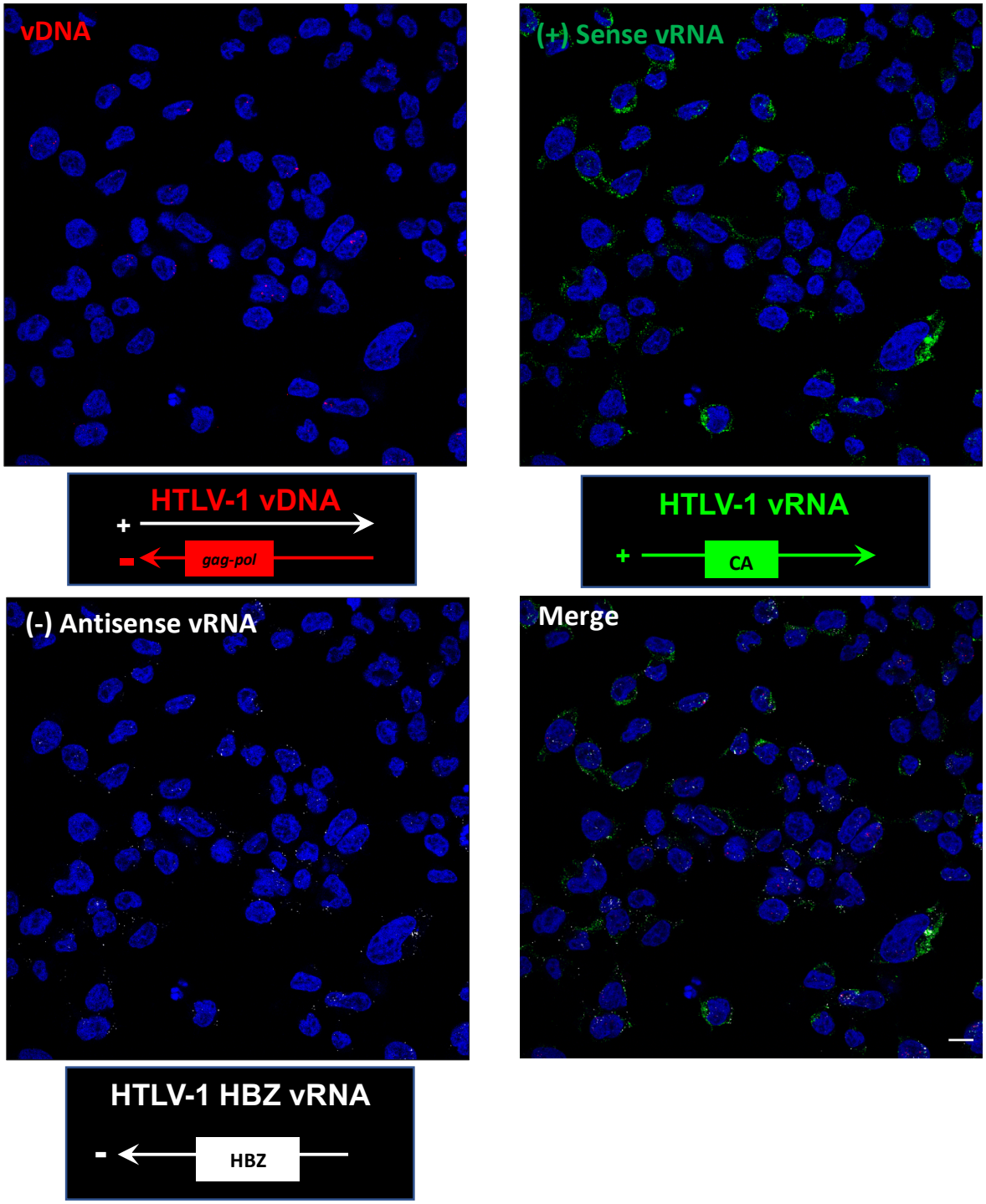


Figure II.3. Simultaneous labeling of HTLV-1 vDNA and vRNA in MT-2 cells. (A) HTLV-1 vDNA (labeled with ATTO 550; red in Figure) (B) Unspliced HTLV-1 sense (+) RNA (RNA C2 probe & labeled with Alexa 488; green in Figure)). (C) HTLV-1 HBZ antisense (-) vRNA (RNA C3 probe & (labeled with Alexa 647; white in Figure)). (D) Merged image. Scale bar represents 20 µm. Nuclei were

stained with DAPI (blue). Amp 4-FL '**B**' was used (Consult **Table II.4** in **Protocol**) to achieve multiplexed labeling of HTLV-1 ('+' & '-') RNA**.**

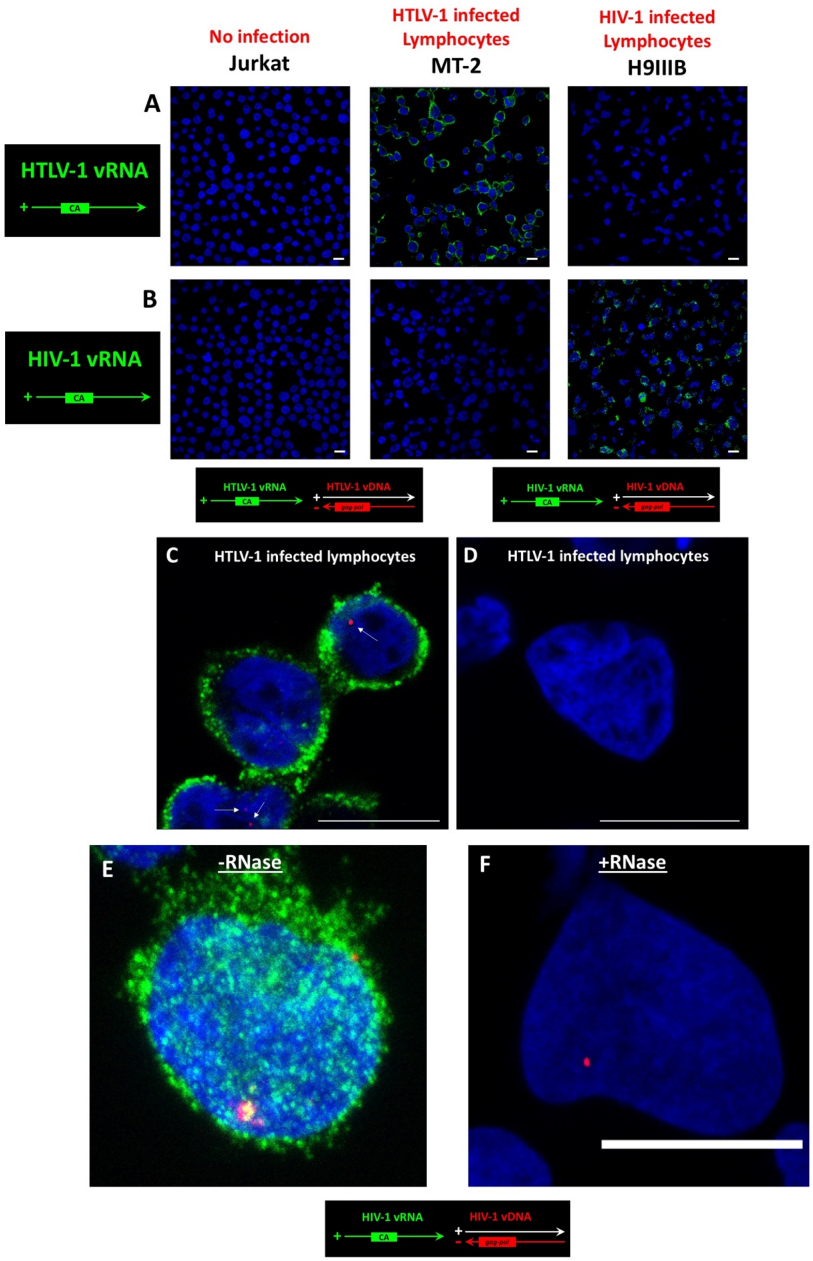


Figure II.4. Control experiments to assess target probe specificity and background. Specificity of HIV-1 and HTLV-1 target probes were assessed following infection in lymphocytic cell lines (Uninfected Jurkat T cells, HTLV-1 infected cells (MT2), and HIV-1 infected cells (H9111B)). Scale bars represent 10 µm. (A) Cells were treated with HTLV-1 RNA probes. (B) Cells were treated with HIV-1 (+) RNA probe (C-D). Further demonstration of the specificity of HTLV-1 probes with no cross-reactivity with HIV-1. White arrows in Figure C denote HTLV-1 DNA (red). (E-F). vRNA staining (green) +/- RNase-treatment.

To verify the specificity of our hybridization probes and to assess how our ISH labeling method impacts protein staining efficiency and overall background, we perform critical controls to ensure the highest level of rigor and reproducibility for our experiments. As an example, in **Figure II.4A-D**, we verify the specificity of our HIV-1 and HTLV-1 vDNA and vRNA probes, as we show very little to no cross-reactivity between the probe sets across the two viruses. Despite labeling of two retroviruses with the potential for probe cross-reactivity, the HIV-1 probes are *only* specific to HIV-1 genetic material and not HTLV-1. The same trend is true for the HTLV-1 probes. In addition, in **Figure II.4F**, we show that we can eliminate vRNA staining if we RNase-treat our cells during sample preparation. To assess the possibility of attenuation of protein staining efficiency due to ISH (protease treatment (**Step 4 in Protocol & Figure II.1B)** and hybridization conditions, which can lead to ablation of epitope recognition or increased background signal, we demonstrate that protein staining efficiency for the nuclear speckle marker, sc-35, is comparable across conventional IF approaches and immunostaining during our MICDDRP protocol. In the conventional IF protocol, cells were permeabilized with 0.1% Triton X-100 for 15 minutes, rather than permeabilization with 0.1% Tween-20 for 10 minutes, which is used in MICDDRP protocol (**Step 3** in **Protocol** & workflow in **Figure II.1B**). Protein staining across both conditions (IF *vs.* MICDDRP) produced a signal several orders of magnitude greater than the control (MICDDRP with no primary antibody), further demonstrating the low background generated following this protocol and preservation of protein epitopes for efficient immunostaining. Image quantification of mean integrated fluorescence intensity of sc-35 signal per cell **(Figure II.5E**) was performed as previously described [318]. For all imaging results shown, we ensured probe and antibody specificity, as well as assessed any possible perturbations or higher than normal background noise attributed to our ISH approach.

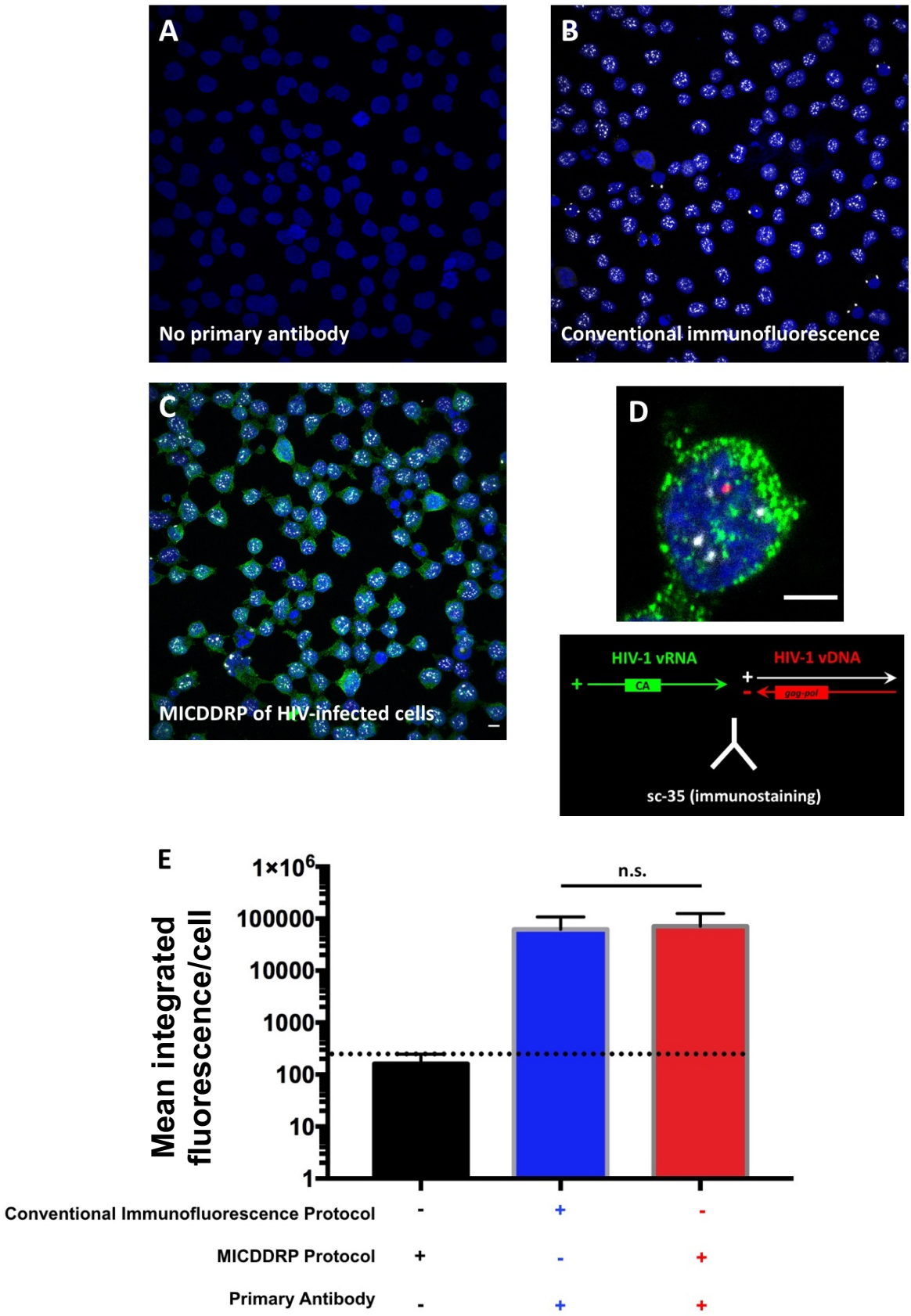

A
No primary antibody
B
Conventional immunofluorescence
C
MICDDRP of HIV-infected cells
D
HIV-1 vRNA
HIV-1 vDNA
CA
gag-pol
sc-35 (immunostaining)
E
Mean integrated fluorescence/cell
1×10^6
100000
10000
1000
100
10
1
n.s.
Conventional Immunofluorescence Protocol
MICDDRP Protocol
Primary Antibody

Figure II.5. Comparison of protein staining efficiency following conventional IF *vs.* MICDDRP. The cellular protein, sc-35, a biomarker for nuclear speckles, was immunostained in Jurkat T cells. All scale bars represent 10 µm. **(**A) Uninfected Jurkat T cells underwent the MICDDRP protocol and were treated with HIV-1 DNA/RNA hybridization probes used in **Figures II.2** and **II.4** above. During immunostaining, no primary antibody was added. (B). Uninfected Jurkat T cells underwent conventional IF, labeling sc-35 (white). Cells were permeabilized with 0.1% Triton X-100 for 15 minutes. (C). MICDDRP was performed on HIV-infected Jurkat T cells. vRNA is labeled green, vDNA is red, and sc-35 is white. (D). Close-up of vRNA, vDNA, and protein labeling (sc-35) in HIV-infected cell. (E) Quantification of mean integrated fluorescence signal per cell of sc-35 across different immunostaining conditions. The y-axis is on a logarithmic scale. The dotted line represents the background signal from uninfected cells that underwent the MICDDRP protocol where no primary antibody was added. No significant difference in signal following MICDDRP *vs.* conventional IF. Over 500 cells were sampled for quantification for reach respective condition.

This method can also be applied to study RNA viruses that may or may not include DNA templates for viral replication. For instance, strand-specific bDNA probes can be designed to monitor expression of sense (+) or antisense (-) strand RNA and different vRNA species in viruses such as HBV, HCV, IAV, and ZIKV. The visualization of different RNA species during the course of infection can provide insight into the replication kinetics of various viral systems.

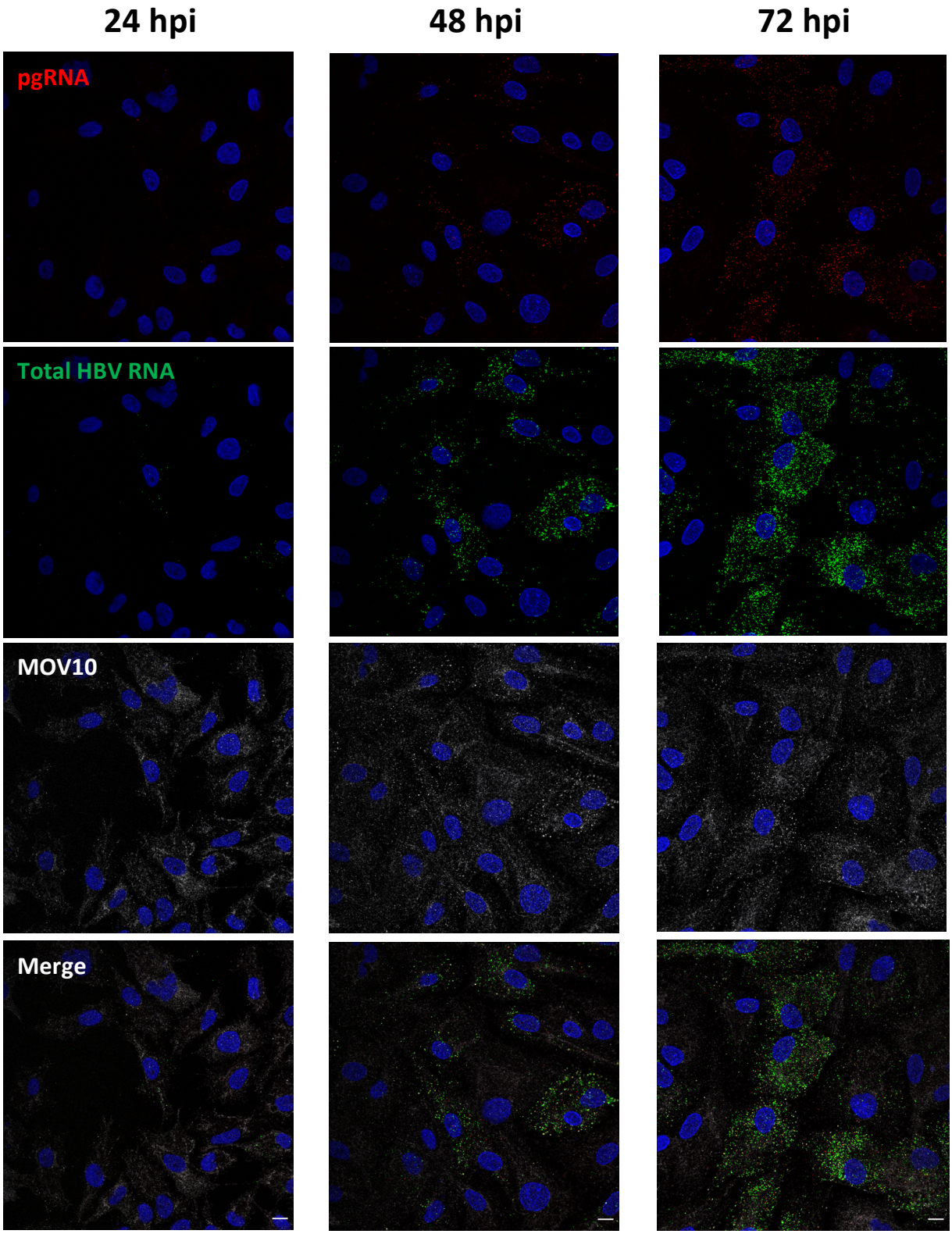

24 hpi
48 hpi
72 hpi
pgRNA
Total HBV RNA
MOV10
Merge

Figure II.6. Time-course of HBV infection of 3E8 cells. Cells were infected with 300 HBV genomes per cell. Viral replication is shown at three time points (24, 48, and 72 hpi). pgRNA (labeled with ATTO 550; red in Figure), total HBV RNA (labeled with Alexa 647; green in Figure), and MOV10 (secondary antibody conjugated to Alexa 488; white in Figure). Nuclei are stained with DAPI in blue. Scale bar on merged images represent 10 µm. hpi, hours post-infection.

Following the time course of HBV infection, we can see that the amount of HBV pre-genomic RNA (pgRNA) and total HBV RNA increases as a function of time. In addition, we simultaneously immunostained a cellular host factor, MOV10.

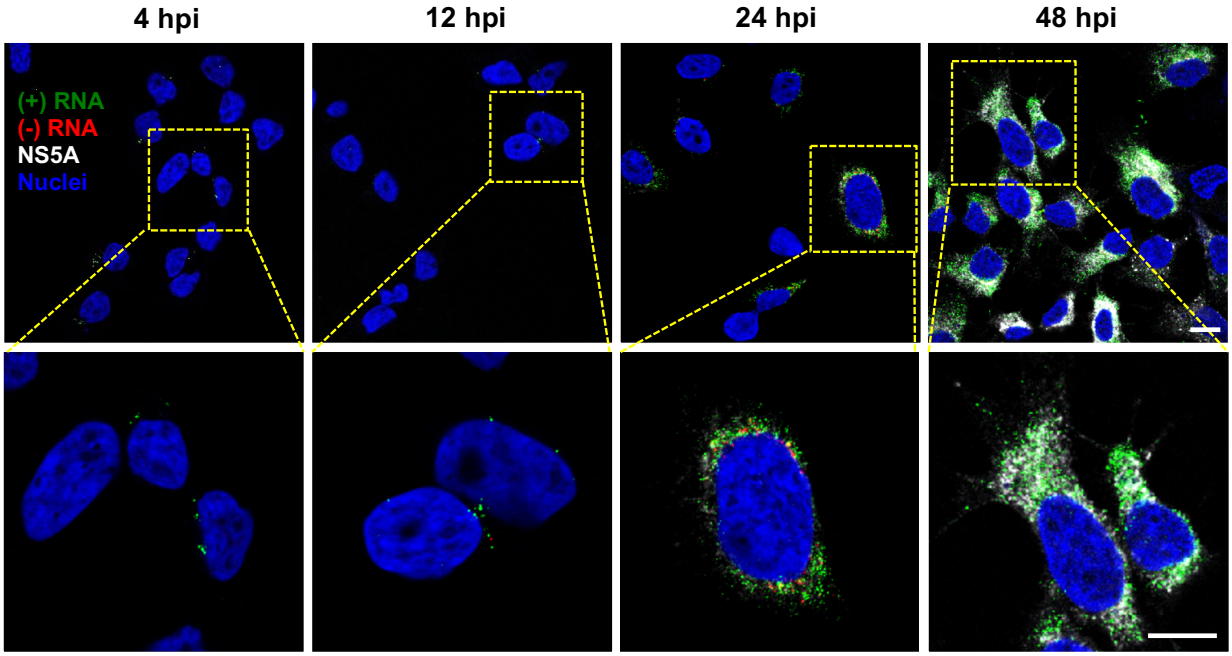


Figure II.7. Time-course of HCV infection of Huh-7.5.1 cells. Huh-7.5.1 cells were infected with hepatitis C virus (HCV) Jc1/Gluc2A at an MOI of 0.5. At the time intervals indicated, the cells were fixed and probed sequentially for sense (+) vRNA, antisense (-) vRNA and NS5A (HCV protein). Nuclei were stained with DAPI. Representative merged images from each time-point, showing (+) RNA in green (labeled with Alexa 647), (-) RNA in red (labeled with Atto 555), NS5A in white (secondary goat anti-mouse conjugated to Alexa 488), and nuclei in blue. Scale bars represent 10 µm. The lower images are enlarged cut outs from the corresponding time-point. hpi, hours post-infection.

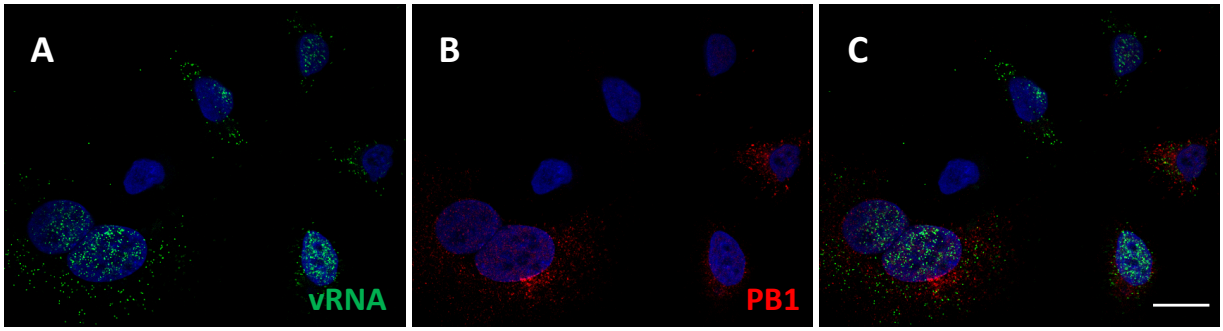


Figure II.8. Strand-specific bDNA FISH and immunostaining of A549 cells infected with influenza A virus. A549 cells infected with PR8 Flu A virus were fixed and probed for (A) IAV nucleoprotein (NP) RNA (labeled with Alexa 488; green in Figure), and (B) IAV polymerase protein (PB1) (secondary goat anti-mouse Atto 550; red in Figure) and nuclei were stained with DAPI (blue). (C) Merged image. Scale bar represents 10 µm.

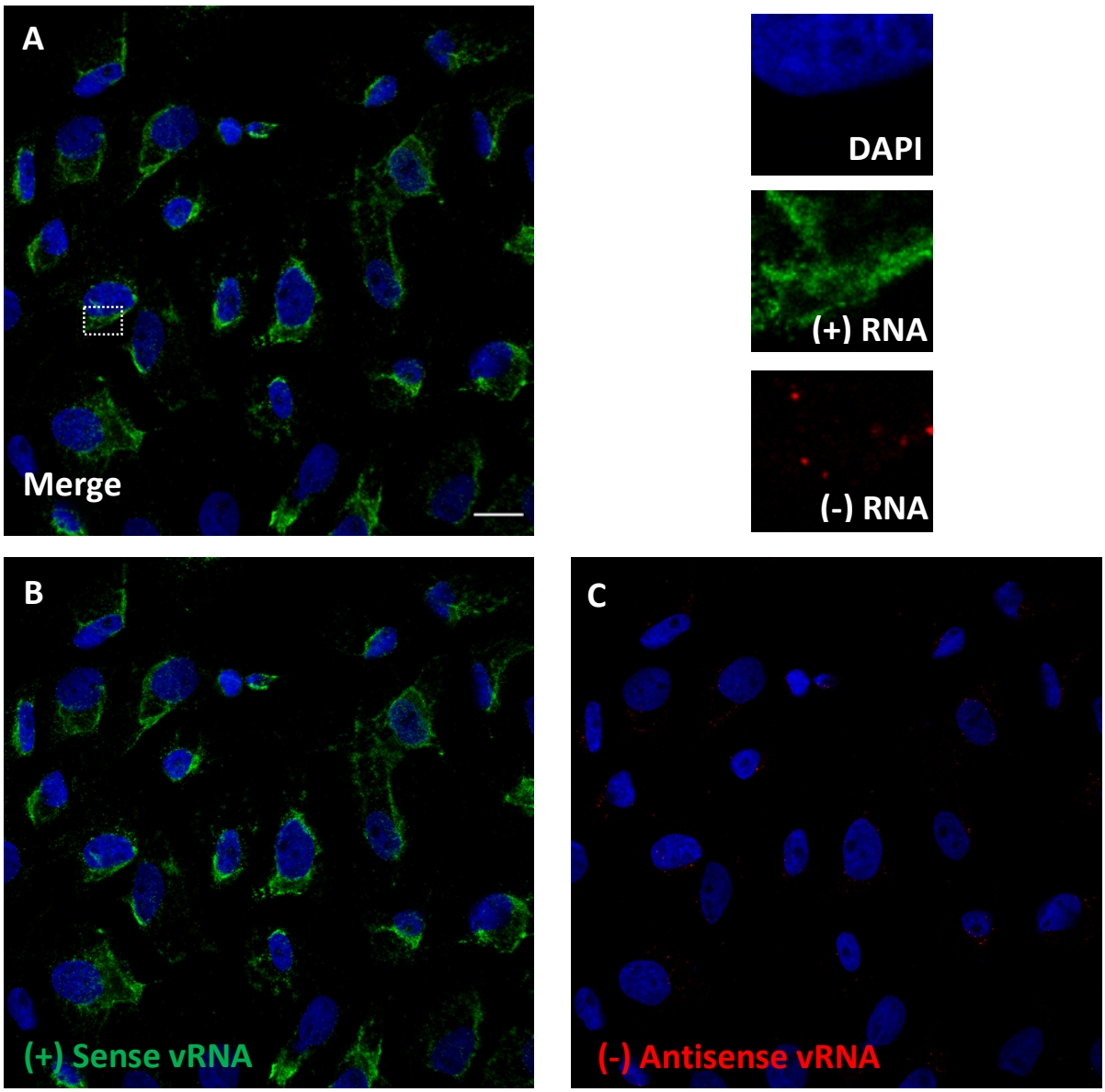


Figure II.9. Strand-specific bDNA FISH in Zika virus (ZIKV)-infected cells. Vero cells were infected with ZIKV at a MOI of 0.1. Cells were fixed at 48 hpi. (A) Cells were simultaneously stained for sense (+) vRNA (labeled with Alexa 488; green in Figure) and antisense (-) vRNA (labeled with Atto 550; red in Figure). Nuclei were stained with DAPI. In (A), the white box denotes a region with both (+) and (-) vRNA. The insets present a close-up of (+) vRNA (green) and the scarcer (-) vRNA species (red). (B) sense (+) vRNA (green). (C) antisense (-) vRNA (red). Scale bar in (A) represents 10 µm.

Imaging was performed *via* confocal microscopy using a 60x oil-immersion objective. The excitation/emission bandpass wavelengths used to detect DAPI, Alexa 488, ATTO 550, and Alexa 647 were set to 405/420-480, 488/505-550, 550/560-610, and 647/655-705 nm, respectively.

TABLE II.5: All reagents and materials used for multiplexed fluorescence imaging of viral infection.

Name	Company	Catalog Number	Comments
4% PFA			
50% dextran sulfate	Amreso	198	For DNA hybridization buffer
50X wash buffer	ACD Bio	320058	
6-well plates			
Amplifier 1-FL	ACD Bio		
Amplifier 2-FL	ACD Bio		
Amplifier 3-FL	ACD Bio		
Amplifier 4-FL	ACD Bio		Consult Amp-4 table in protocol
Anti-HCV NS5a antibody	Abcam	ab13833	Mouse monoclonal; works with HCV genotypes 1a, 1b, 3, and 4
Anti-HIV-1 p24 monoclonal antibody	NIH AIDS Reagent Program	3537	
Anti-Mov10 antibody	Abcam	ab80613	Rabbit polyclonal
Anti-PB1 antibody	GeneTex	GTX125923	Antibody against flu protein
Bovine serum albumin			Blocking reagent for immunostaining
Cell media with supplements			Media appropriate for cell model
Coverslips			
DAPI	ACD Bio		Nuclear stain (RNAscope kit from ACD Bio)
Dulbecco's phosphate buffered saline (1X PBS)	Gibco	14190250	No calcium and magnesium
Ethylene carbonate	Sigma	E26258	
Fetal bovine serum (FBS)			Use specific FBS based on what serum secondary antibody was raised in (e.g goat FBS)

Table II.5 Continued:

Fisherbrand colorfrost plus microscope slides	Fisher Scientific	12-550-17/18/19	Precleaned
HCV-GT2a-sense-C2 probe	ACD Bio	441371	HCV(+) sense RNA probe
HIV-gagpol-C1	ACD Bio	317701	HIV-1 cDNA probe
HIV-nongagpol-C3	ACD Bio	317711-C	HIV-1 RNA probe
HybEZ hybridization oven	ACD Bio	321710/321720	
ImmEdge hydrophobic barrier pen	Vector Laboratories	H-4000	
Nail polish			For immobolizing coverslip to slide prior to protease treatment
Nuclease free water	Ambion	AM9937	
Poly-d-lysine (PDL)			Coat coverslips in 20 µg/mL of PDL for 30 minutes
Probe diluent	ACD Bio	300041	For diluting RNA C2 or C3 probes
Prolong gold antifade	Invitrogen	P36930	
Protease III	ACD Bio	322337	
RNAscope® Probe- V-Influenza-H1N1-H5N1-NP	ACD Bio	436221	
RNase A	Qiagen		
Secondary antibodies			
Slides			
Sodium chloride			For DNA hybridization buffer
Sodium citrate, pH 6.2			For DNA hybridization buffer
Tween-20			For DNA hybridization buffer and PBS-T
V-HBV-GTD	ACD Bio	441351	Total HBV RNA
V-HBV-GTD-01-C2	ACD Bio	465531-C2	HBV pgRNA probe
V-HCV-GT2a probe	ACD Bio	441361	HCV(-) sense RNA probe
V-HTLV-HBZ-sense-C3	ACD Bio	495071-C3	HTLV-1 (-) sense RNA probe targetting HBZ
V-HTLV1-GAG-C2	ACD Bio	495051-C2	HTLV-1 DNA probe
V-HTLV1-GAG-POL-sense	ACD Bio	495061	HTLV-1 (+) sense RNA probe
V-Influenza-H1N1-H5N1-NP	ACD Bio	436221	IAV RNA probe
V-ZIKA-pp-O2	ACD Bio	464531	Zika(+) sense RNA probe
V-ZIKA-pp-O2-sense-C2	ACD Bio	478731-C2	Zika(-) sense RNA probe

DISCUSSION

Simultaneous visualization of RNA, DNA, and protein often requires extensive optimization. Two commonly used methods are 5-ethynyl-2-deoxyuridine (EdU) labeling and DNA FISH. EdU labeling has been applied to visualize viral DNA and protein simultaneously, as EdU is incorporated in nascent DNA and subsequently labeled with azide-containing fluorescent dyes *via* click chemistry. EdU labeling can thus be used to monitor native virus replication kinetics of DNA viruses or viruses with DNA templates for replication [174]. A shortcoming of EdU labeling, however, is that in dividing cells, the replicating genome will incorporate EdU, generating high background and confounding image analysis. DNA FISH can circumvent these issues by directly hybridizing a nucleic acid probe to the respective target regardless of the cell cycle. However, conventional FISH often relied on high temperatures to achieve efficient probe hybridization, hindering immunostaining or even simultaneous RNA staining [347]. MICDDRP can potentially circumvent these issues providing robust simultaneous fluorescent labeling of DNA, RNA, and protein across a variety of cellular systems.

While we have demonstrated that we can label protein and nucleic acid simultaneously using our MICDDRP protocol, optimization was needed across different systems. The first major parameter that we had to optimize was protease treatment. We varied protease III concentration across our conditions. Optimization of protease treatment was empirical, as we used different dilutions to assess what yielded the greatest hybridization efficiency, without compromising immunostaining efficiency. Appropriate controls were performed side-by-side to assess probe specificity and changes to protein staining efficiency attributed to protease treatment. The next major parameters that needed optimization were probe design and probe hybridization.

Proper design of capture and amplifier probes are critical for achieving the sensitivity and specificity of bDNA technology. Software packages that predict the probability of non-specific hybridization events are available to improve probe design [316, 346]. bDNA probes with the accompanying pre-amplifier, amplifier, and fluorescent label probes can now be commercially purchased to ensure compatibility with bDNA imaging kits. Users can supply manufacturers with sequence information (~300-1000 base pairs) for the target region(s) in the form of a fasta file (text-based format for representing nucleotide sequence). Target probes are generated with > 90% sequence homology to the supplied sequence.

For DNA labeling, we have found that dilution of the probes in the hybridization buffer described in **Step 5** of the **Protocol** improves DNA hybridization. When labeling both DNA and RNA, the RNA probe can be diluted in the hybridization buffer. DNA labeling in the absence of RNase cannot exclude the possibility that the observed nucleic acid includes RNA of the targeted strandedness. Temperature may also have to be adjusted for improving hybridization efficiency. Increasing temperature may affect protein staining efficiency, as increased temperatures may promote protein denaturation, ablating epitope recognition of the primary antibody. In our presented representative data, we have performed ISH at 40 °C.

Compared to conventional DNA FISH, MICDDRP provides an improved procedure for simultaneously labeling DNA, RNA, and protein to visualize *via* fluorescence microscopy. A potential limitation is that the selection of probe may affect the efficiency of hybridization and ability to quantitatively compare data between probes. This protocol has been effective across

diverse cellular and viral systems in our hands with only minor optimization needed across varying conditions. Recent high-profile publications have utilized our approach to study HIV integration site selection [94] and HIV reverse transcription kinetics [348]. Future applications of MICDDRP could include visualization of viral nucleic acids concomitantly with nucleic acid sequences of specific cellular genes and cellular proteins.

ACKNOWLEDGMENTS:

This work was supported in whole or in part by the National Institutes of Health (R01 AI121315, R01 AI146017, R01 AI148382, R01 AI120860, R37 AI076119, and U54 AI150472). We thank Dr. Raymond F. Schinazi and Sadie Amichai for providing cells infected with influenza A Virus.

CHAPTER III: APPLICATIONS OF SINGLE-MOLECULE FISH TO STUDY VIRUS-HOST INTERACTIONS INFLUENCING TRANSCRIPTION FROM UNINTEGRATED HIV-1 DNA

The research and data presented in this chapter is included in a published manuscript in *Cell Host & Microbe.*

Chapter III contains data & analysis included in:

Dupont, L., Bloor, S., Williamson, J.C., Cuesta, S.M., **Shah, R.**, Namaati, A., Greenwood, E.J.D., Balasubramanian, S., Sarafianos, S.G., Matheson, N.J., Lehner, P.J. (2021) Nuclear Immunosurveillance by the SMC5/6 Complex is Antagonized by HIV-1. *Cell Host & Microbe.* (PMID: 33811831)

This is an open access article distributed under the terms of the Creative Commons CC-BY license.

Copyright © Cell Press

INTRODUCTION

The work presented in this Chapter utilizes the multiplexed fluorescence imaging approach, outlined in **Chapter II** to visualize host restriction of transcription from unintegrated HIV-1 cDNA. While integration of the HIV-1 reverse-transcribed genome into cellular chromatin is a major step in the retroviral life cycle, much of the nuclear viral cDNA ends up as extrachromosomal DNA species, consisting of linear unintegrated DNA, 1-long terminal repeat (LTR), and 2-LTR circles [349]. Integrated and unintegrated viral DNA rapidly become chromatinized within the host nucleus, however, transcription from unintegrated lentiviral cDNA is severely reduced relative to integrated proviral DNA [350]. Specifically, in this study, we visualize how knock-out (KO) of SLF2 affects transcription of unintegrated HIV-1 cDNA. SLF2 is a recruitment factor for the Smc5/6 complex. The Lehner group at the University of Cambridge have identified Smc5/6 as a cellular restriction factor that inhibits transcription of unintegrated HIV-1 DNA. Smc5/6 has multiple functions in the cell, including reorganization of eukaryotic chromosomes during cellular division, compaction of genomic DNA, and supports DNA repair and genome stability [351, 352]. Smc5/6 inhibits HIV-1 transcription by influencing the compaction and epigenetic silencing of unintegrated viral cDNA. Suppression of the Smc5/6 signaling pathway leads to an increase in chromatin accessibility of unintegrated HIV-1 DNA. In addition, the Lehner group found that the HIV-1 accessory protein, Vpr, counteracts this silencing mechanism by promoting proteasomal-mediated degradation of the Smc5/6 complex. We collaborated closely with the Lehner group to verify the role of Smc5/6 in HIV-1 transcriptional silencing and the counteraction of Vpr. We have applied our multiplexed *in-situ* imaging approach to visualize the HIV-1 transcriptional landscape of Jurkat T cells +/-SLF2 (Jurkat SLF2 knock-out (KO) cell line generated by Lehner lab) following infection with VSV-G

pseudotyped NL4-3-Δenv-GFP (+/-Vpr) (commonly used HIV-1 lab-adapted strain). In addition, we interrogated if PML bodies play a role in influencing HIV-1 transcription by assessing localization between PML bodies and unintegrated HIV-1 DNA.

MATERIALS/METHODS

Cell Culture

Jurkat T cells (WT & SLF2KO) were cultured in RPMI 1640 medium (Gibco, Waltham, MA, USA), supplemented with 10% heat-inactivated fetal bovine serum (FBS) and 2 mM L-glutamine (Gibco), in a humidified incubator at 37 °C with 5% CO_2.

Virus production and infection

Pseudoviruses were generated by transfection of HEK293T cells with NL4-3 (+/-vpr) constructs and VSV-G at a DNA ratio of 5:1 using X-tremeGENE (Roche) following the manufacturers recommendation. 48 hours post transfection (hpi), supernatants were collected, filtered (0.45 µm pore size), and stored at -80°C. All infections were performed in the presence of DEAE-Dextran (8 µg/mL) [353]. To block proviral integration, infected cells were treated with Raltegravir (RAL, Cayman Chemical; 1 µM). Cells were fixed for imaging 48 hpi.

Multiplexed in-situ viral RNA detection and immunostaining, imaging, and analysis of RNA fluorescence intensity

HIV-1 RNA detection was performed by branched DNA in situ hybridization (bDNA FISH) following a modified RNAscope protocol with RNAscope reagents from Advanced Cell Diagnostics (ACD) [316]. Briefly, cells were seeded on poly-d-lysine coated coverslips 48 h post infection, fixed in 4% PFA (30 min, RT), washed three times in PBS, incubated 10 minutes in

0.1% Tween-20-PBS (PBS-T), and washed twice in PBS. Cells were incubated with manufacturers protease treatment (Pretreat 3; 1:5 dilution in PBS) in a humidified ACD HybEZ oven at 40°C, 15 min. Protease solution was decanted, and samples were washed twice in PBS. A probe that recognizes HIV-1 RNA (HIV-nongagpol-C3; ACD 317711-C) was applied following manufacturers recommendations and samples incubated at 40 °C for 2 h in the HybEZ oven. Remaining wash steps, hybridization of preamplifiers, amplifiers, and fluorescent label were performed as previously described [119, 317, 318].

Following in-situ hybridization, immunostaining was performed. Cells were blocked for 1 hour at RT (blocking buffer; 1% w/v BSA, 10% v/v FBS in PBS with 0.1% v/v Tween-20 (PBST)). Following blocking, cells were incubated with primary antibody (anti-PML; 1:100 dilution (Abcam; ab96051)) diluted in PBST and 1% w/v BSA. Slides were washed twice witb PBST for 10 min at RT with shaking. Secondary antibody (Alexa Fluor 647 goat anti-mouse) was incubated for 1 hr on the cells in PBST and 1% w/v BSA. Slides were washed twice with PBST for 10 min at RT with shaking. PBST was decanted. Nuclei were next counter-stained with 4',6'-diamino-2-phenylinndole (DAPI) and mounted using Prolong Gold Antifade (Invitrogen).

Imaging was performed using a Nikon C2 confocal microscope using a 60x APO oil-immersion objective (numerical aperture 1.4). The excitation/emission bandpass wavelengths to detect DAPI (405 nm) and HIV-1 RNA (647 nm) were set to 420-480 nm and 655-705 nm, respectively. Images were quantified using Gen5 software (BioTek) to count individual cells and determine the integrated fluorescence intensity of HIV-1 RNA per cell. Background signal was determined using uninfected Jurkat T cells processed as described above.

Statistical Analysis

Statistical significance was calculated using ordinary one-way ANOVA with multiple comparisons correction performed in GraphPad Prism v8. Error bars denote standard deviation. ns, $P > 0.05$. *, $P < 0.05$. **, $P < 0.01$. ***, $P < 0.001$.

RESULTS

Figure III.1 provides a schematic of sample preparation for this assay. Jurkat T cells (WT & SLF2KO) are treated with RAL prior to infection with pseudovirus +/- *vpr.* Cells are fixed 48 h.p.i. Imaging results can be seen in **Figure III.2,** and the respective image quantification vRNA fluorescence intensity is shown in **Figure III.3.** Infected cells were filtered from the uninfected cells (three standard deviations > background fluorescence level). Mean integrated fluorescence intensity of each infected cell above the selected threshold was quantified and plotted for each condition. The association between unintegrated nuclear HIV-1 DNA and PML bodies was quantified *via* object-based localization analysis calculating the distance of the geometric means of PML bodies and viral DNA foci **(Figure III.4).** We do not see any differences in physical proximity between nuclear unintegrated HIV-1 DNA and PML bodies applying our cytological approach.

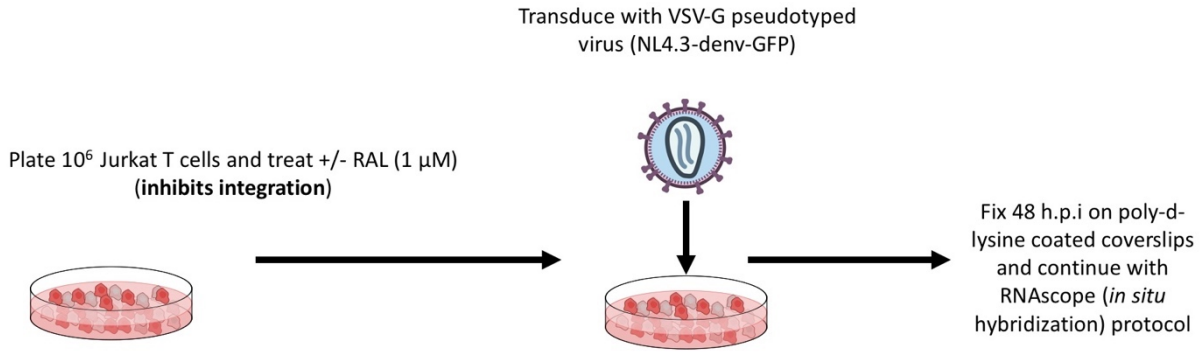


Figure III.1. Workflow for *in-situ* viral RNA detection and imaging. Suspension Jurkat T cells are treated with 1 µM RAL and infected with pseudovirus. 48 hpi, cells are harvested and plated on poly-d-lysine coated coverslips. *In-situ* hybridization or multiplexed imaging are then performed as described in **Chapter II**. Hpi = hours post-infection.

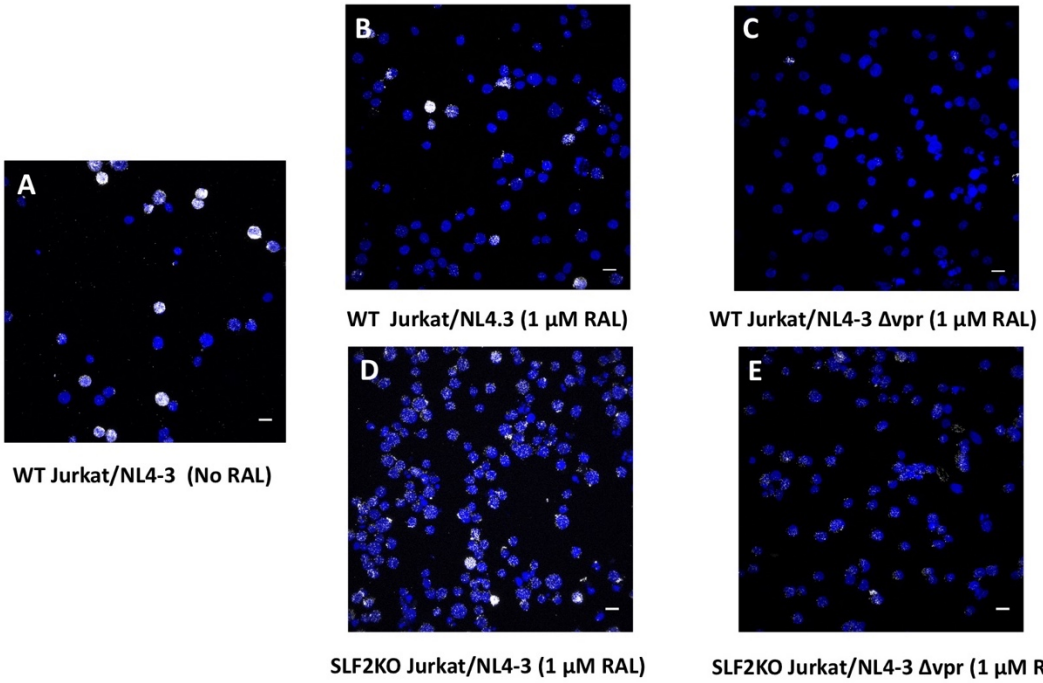


Figure III.2. Imaging of viral RNA in WT & SLF2KO Jurkats infected with NL4-3-Δ-GFP +/- vpr. *In situ* hybridization of vRNA in WT and SLF2KO Jurkat T cells infected with NL4-3^{GFP} +/- vpr reporter viruses in the absence/presence of 1 µM RAL. Cells were fixed 48 h.p.i. Scale bars are 10 µm **(A-E).** WT Jurkats infected with NL4-3^{GFP} in the absence of RAL treatment **(A)**. WT Jurkats infected with NL4-3^{GFP} in the presence of 1 µM RAL **(B)**. WT Jurkats infected with Δvpr NL4-3^{GFP} in the presence of 1 µM RAL **(C).** SLF2KO Jurkats infected with

NL4-3^{GFP} in the presence of 1 µM RAL **(D)**. SLF2KO Jurkats infected with Δvpr NL4-3^{GFP} in the presence of 1 µM RAL **(E).**

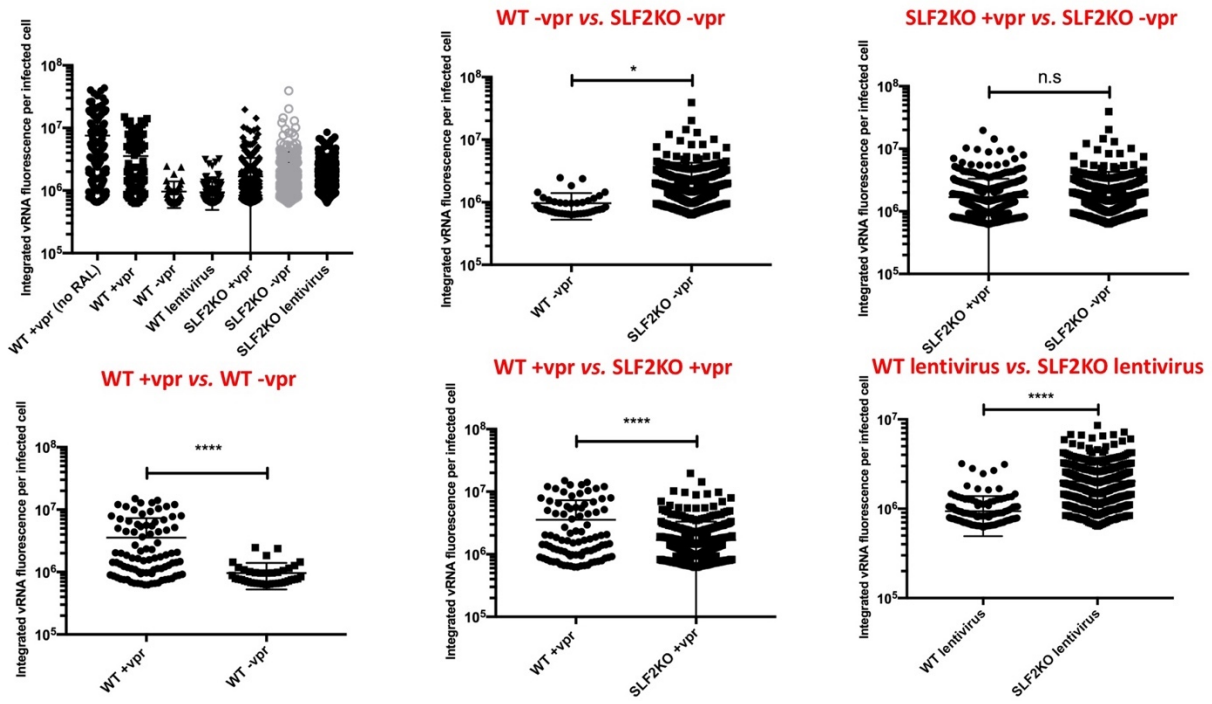


Figure III.3. Quantification of (+) HIV-1 RNA expression in the presence of RAL in WT & SLF2KO Jurkat T cells infected with pseudovirus +/- *vpr*. Infected cells were filtered from the uninfected cells (three standard deviations > background fluorescence level) and integrated HIV-1 RNA fluorescence intensity was plotted per cell for each condition. *$p_{value} < 0.05$; **** $p_{value} < 0.001$.

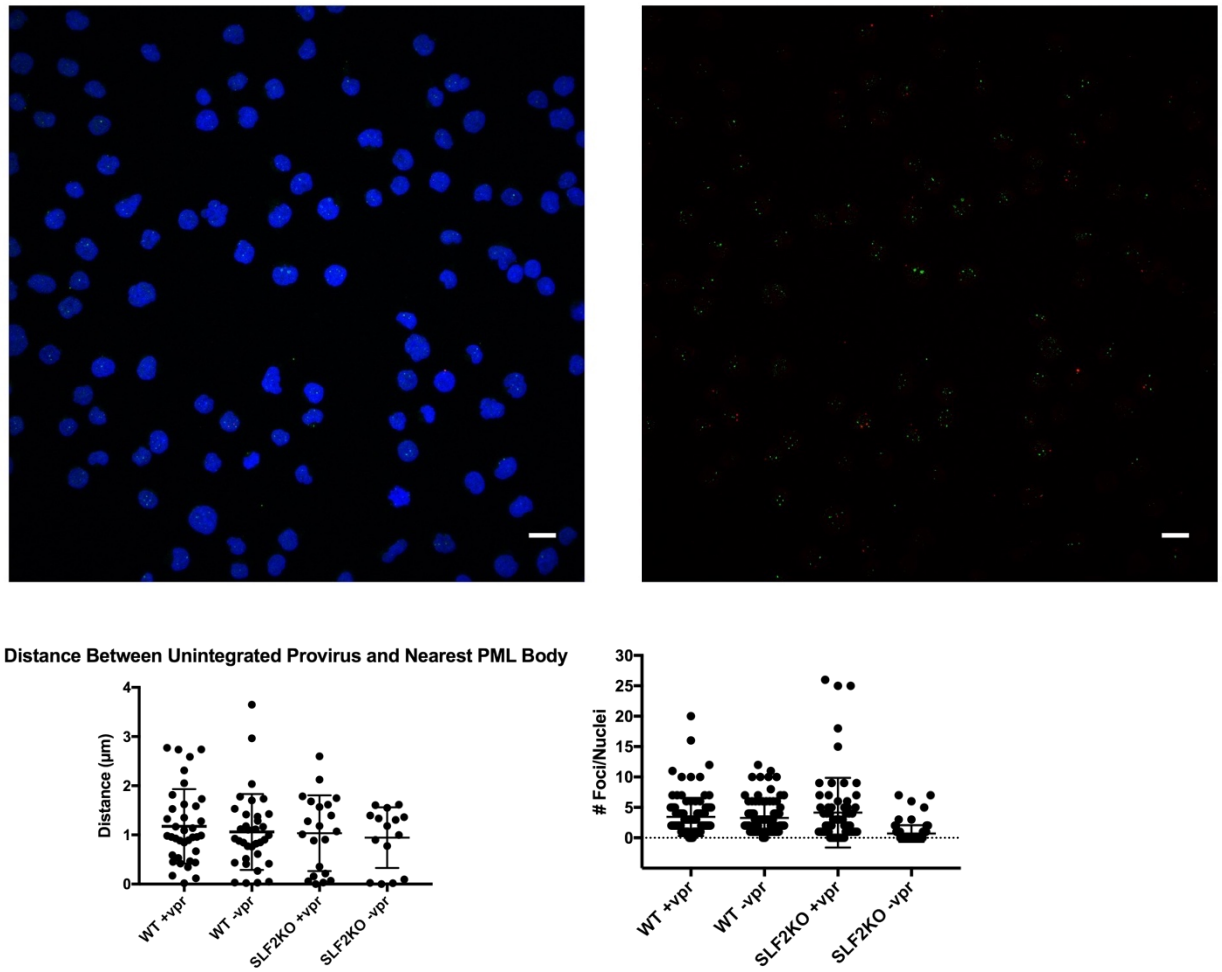


Figure III.4. Quantification of distance between unintegrated nuclear HIV-1 cDNA & PML bodies. Distance between geometric means of viral DNA foci & PML bodies and number of PML foci/nuclei per condition. No significance detected across samples.

DISCUSSION

Through application of our multiplexed *in-situ* imaging approach, we have clearly demonstrated that the cellular host factor, Smc5/6, reduces expression of unintegrated lentiviral cDNA and that the HIV-1 accessory protein, Vpr, can antagonize this host inhibitory effect. Our results directly corroborate what was shown biochemically by the Lehner group, contributing imaging analyses

to this publication in *Cell Host & Microbe*. In general, we see that inhibition of HIV-1 proviral integration leads to a severe attenuation of HIV-1 (+) expression, as expected. SLF2 KO leads to an increase in HIV-1 transcription from unintegrated viral cDNA templates.

CHAPTER IV: RALTEGRAVIR TREATMENT LEADS TO AN INCREASE IN HIV-1 ANTISENSE (-) RNA EXPRESSION

The research and data presented in this chapter will be included in a future manuscript.

Chapter IV contains data & analysis that will be included in:

Mahboubi, D., **Shah, R.,** Puray-Chavez, M., Poeschla, E.M., Engelman, A.N., Tedbury, P.R., Sarafianos, S.G. (2021) Interference with LEDGF/p75-directed Integration Enhances Transcription of HIV-1 Antisense RNA. *In preparation.*

D.M., R.S., & M.P. performed imaging experiments. P.R.T., D.M., R.S., & S.G.S. conceptualized experiments.

INTRODUCTION

Antisense (-) transcripts are RNA molecules expressed from the opposite strand of a protein-coding gene and may play a role in transcriptional regulation of the respective cognate gene [354, 355]. In human and *Saccharomyces cerevisiae* eukaryotic systems, (-) transcripts have been found to play a role in regulating epigenetic silencing [356], transcriptional interference [357], and RNA stability [358, 359]. The role of (-) expression of HIV-1 RNA remains poorly elucidated. Many open questions remain regarding the possible biological function(s) of HIV-1 (-) RNA in regulating viral replication, as well as how (-) vRNA is expressed. HIV-1 (-) RNA can be translated into the Antisense Protein (ASP), with the (-) ASP gene overlapping the HIV-1 Rev Response Element (RRE) and the envelope glycoprotein gene [79, 360]. (-) vRNA has also been found in samples of *ex vivo* donor CD4+ T cells at about 10-30 copies per million cells [361]. For this study, we are interested in defining how HIV-1 (-) RNA can be expressed, as well as identifying the transcriptional "start" and "end" sites of HIV-1 (-) transcripts. Recent work has suggested that retroviral (-) RNAs may be retained within the nucleus, lacking proper poly(A) expressed from the retroviral 3' LTR [362]. We seek to unravel to what extent HIV-1 (-) transcription is originating from proximal host promoters, LTRs of unintegrated HIV-1 dsDNA, and the 3' LTR of integrated proviral DNA.

MATERIALS/METHODS

Imaging conditions and sample workflow

Samples were prepared for multiplexed *in-situ* imaging analysis, as performed in **Chapter III.** Jurkat T cells were treated with 1 μM RAL (Cayman Chemicals) to block proviral integration.

vRNA of HIV-infected cells was quantified to assess how RAL treatment impacts the proviral transcriptional landscape.

RESULTS

Treatment with RAL prior to infection ablates HIV-1 integration allowing us to quantify the viral transcriptional output from unintegrated HIV-1 DNA. We tested the same conditions presented in **Chapter III**. In **Figure IV.1**, we show that RAL treatment reduces the number of HIV-infected cells that are expressing vRNA. KO of the Smc5/6 recruitment factor, SLF2, increases the number of cells expressing (+) and (-) vRNA. RAL treatment, however, increases the number of cells expressing (-) vRNA 48 h.p.i relative to untreated cells, as well as increases overall transcriptional output of (-) vRNA in cells expressing (-) vRNA relative to the untreated cells (**Figure IV.2**). Overall, these studies demonstrate that HIV-1 (-) RNA can be expressed from unintegrated vDNA templates.

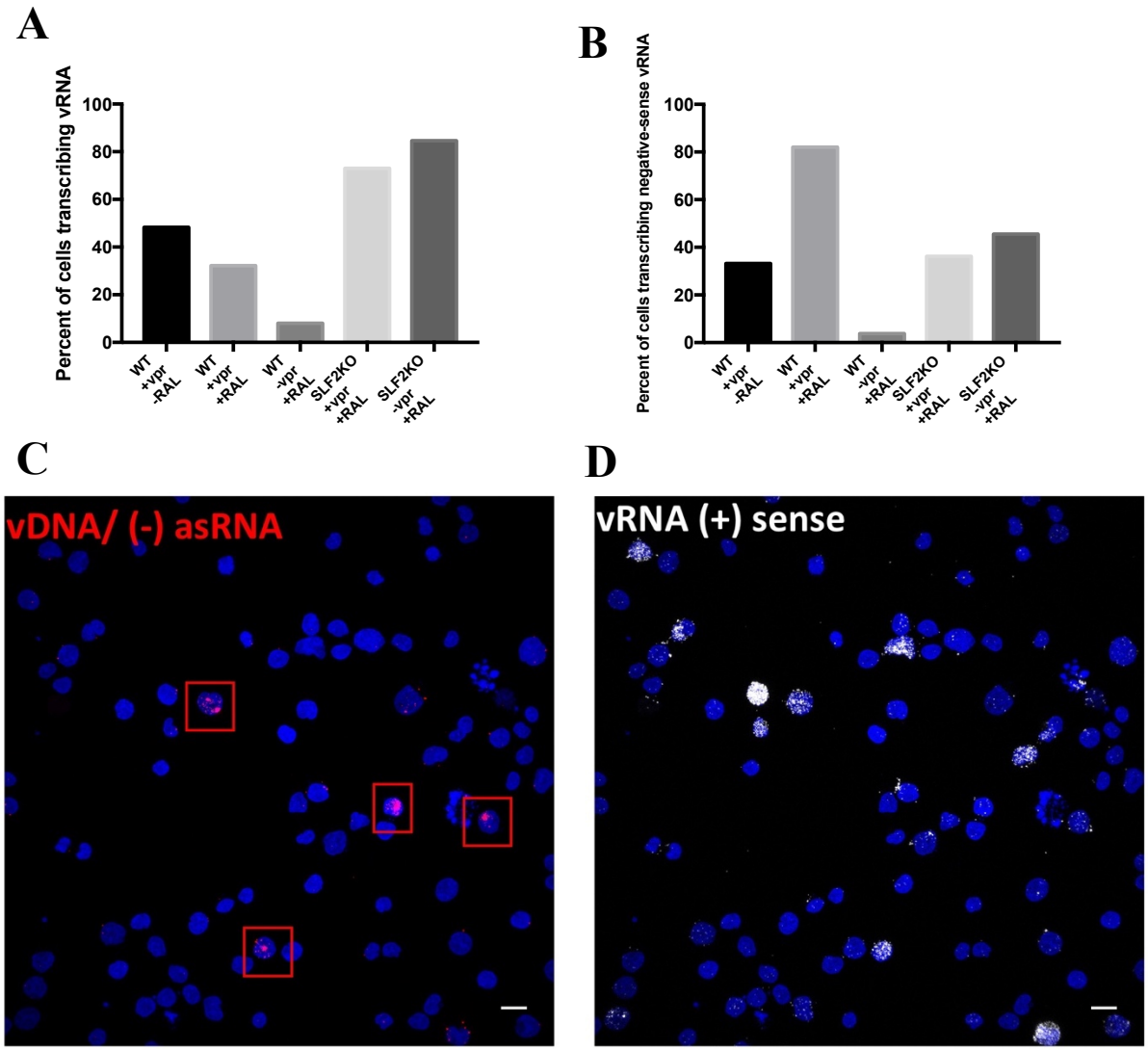


Figure IV.1. Increased expression of HIV-1 (-) RNA following RAL treatment. Jurkat T cells are infected with HIV-1 pseudovirus in the presence of 1 µM RAL. Cells are fixed 48 h.p.i. Quantification of cells expressing (+) vRNA **(A)** or (-) vRNA **(B)** across conditions analyzed in **Chapter III.** HIV-expressing cells were selected if integrated fluorescence intensity per cell was three standard deviations above background control. Representative cells expressing (-) vRNA are demarcated by red boxes **(C).** Corresponding (+) vRNA expression (white) **(D).** N (# of biological replicates) = 1. ~150-500 vRNA-producing cells quantified/condition.

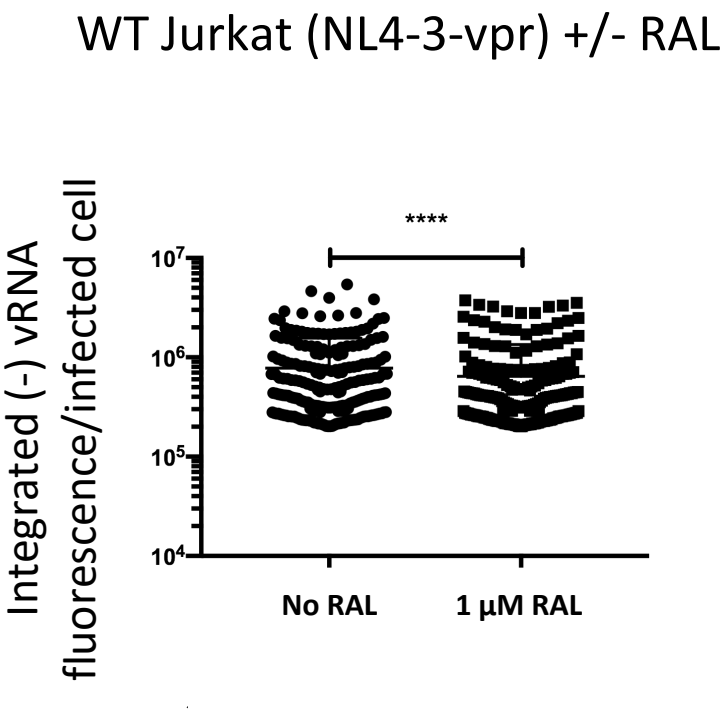


Figure IV.2. Comparison of RAL treatment on expression of (-) vRNA in HIV-infected Jurkat T cells. Infected cells expressing (-) vRNA three standard deviations above background control were selected and plotted. Cell number ~500 cells/condition. **** $p_{value} < 0.001$.

DISUCSSION

Our finding that RAL treatment leads to an increase in HIV-1 (-) RNA expression suggests (-) vRNA transcripts can originate from unintegrated vDNA. This demonstrates a possible route of expression of HIV-1 (-) transcripts. In **Chapter VI**, we also demonstrate that HIV-1 (-) transcription can be induced by proximal host promoters in the case of J-Lat 10.6 cells. In this cellular model, HIV-1 DNA has integrated into a single and defined locus of the human genome. Transcriptional activation of this respective provirus can be modulated *via* TNF-**α** treatment.

When the HIV-1 provirus is inactive in the absence of TNF-**α**, we microscopically visualize expression of HIV-1 (-) RNA. We attribute expression of this (-) vRNA to the proximal *SEC16A* host promoter, which is in the opposite transcriptional orientation of the J-Lat 10.6 provirus, thus leading to host-driven HIV-1 (-) RNA expression (**Figure VI.4A**). Further validation will be performed to quantify changes in (-) vRNA production following RAL treatment (stranded quantitative reverse transcription PCR (qRT-PCR) assays). In addition, the sequence identity and the definitive transcriptional "start" and "end" sites of the (-) transcripts will be determined *via* long-read Nanopore RNA-sequencing. Nanopore sequencing will enable improved mapping of long and complex vRNA species [247, 333-335], improving transcriptomic assemblies of these respective (-) vRNAs. We will also assess the extent ALLINIs and other small molecules that redirect canonical HIV-1 integration site selection contribute to HIV-1 (-) expression. Preliminary data suggests that ALLINI treatment increases expression of (-) vRNA (*unpublished data*). This investigation into the characterization of HIV-1 (-) transcription will improve our understanding of the formation of (-) vRNA species during infection, laying the foundation for further in-depth mechanistic studies defining possible biological function(s) of proviral antisense transcription.

CHAPTER V: MULTIPLEXED FLUORESCENCE-BASED SINGLE-CELL IMAGING TO CAPTURE EARLY REPLICATION EVENTS OF SARS-CoV-2 INFECTION.

The research and data presented in this chapter will be included in a future manuscript.

Chapter V contains data & analysis that will be included in:

Shah, R., Lan, S., Ong, Y.T., Boggs, E.A., Tedbury, P.R., Sarafianos, S.G. (2021) Multiplexed fluorescence-based single-cell imaging to capture early replication events of SARS-CoV-2. (*In preparation)*

R.S., E.A.B., & S.L. performed imaging experiments. Y.T.O. performed infections with SARS-CoV-2 in our BSL3 facility. R.S. performed stranded RT-qPCR assays. R.S., P.R.T., & S.G.S. conceptualized experiments. R.S. will write final manuscript.

INTRODUCTION

In **Chapters II-IV**, we demonstrated the broad applicability of our multiplexed imaging platform (MICDDRP). In **Chapter V**, we present ongoing work seeking to use *in-situ* imaging approaches to visualize severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) transcriptional dynamics and early replication events.

SARS-CoV-2 is an enveloped positive-sense (+) single-stranded RNA virus of the order *Nidovirales* with a genome of ~ 30 kb in length [363]. Nidoviruses share a common genome organization and coordinated viral RNA expression occurs at replication-transcription complexes (RTCs). Following cellular entry of the target host cell, SARS-CoV-2 transcription initiates with the synthesis of a complementary antisense (-) genome *via* the viral RNA-dependent RNA polymerase (RdRp), NSP12 [364]. The (-) genome serves as an intermediate template of (+) vRNA expression, as well as a template for transcription of viral (+) subgenomic RNAs (sgRNAs). sgRNAs have a common leader sequence joined to the gene sequences in the 3'-end of the viral genome, which occurs *via* discontinuous transcription. sgRNAs can then be fed into the cellular translational machinery for synthesis of viral proteins. (-) sgRNA expression is likely caused by paused (-) RNA synthesis leading to a nested set of (-) sgRNAs from the 3' end of the viral genome, which are joined to a common 5'-leader sequence [364-366]. (-) sgRNAs also serve as a template for synthesis of (+) sgRNAs.

We seek to better characterize SARS-CoV-2 early replication dynamics, tracking simultaneous expression of different viral nucleic acid species ((+) and (-) genomic RNAs (gRNAs) and sgRNAs)), as well as viral protein. We infected different epithelial-based cell lines for kinetic studies of SARS-CoV-2 replication. Using stranded qRT-PCR in conjunction with our smFISH protocol, we quantified changes in viral nucleic acid expression in the course of an

eight-hour infection period. Our imaging approach allows us to detect scarce viral nucleic acids within individual cells at early time points to sample the heterogeneity of infection across cell-to-cell. The work in this chapter serves as a guideline for further single-cell microscopy-based studies to assess SARS-CoV-2 infection, providing validation of highly sensitive target oligonucleotide probes that can specifically label vRNAs of particular strandedness ((+) or (-)) and sequence composition. Our imaging analysis is directly compared to our stranded RT-qPCR assay to validate the robustness and precision of our *in-situ* assay.

MATERIALS/METHODS

Cell Culture of Caco-2 and Vero E6 cells

Caco-2 and Calu3 cells were cultured using ATCC-formulated Eagle's Minimum Essential Medium (EMEM) supplemented with 10% fetal bovine serum (FBS) and 1% penicillin/streptomycin. Vero E6 cells were grown with DMEM supplemented with 10% fetal bovine serum (FBS) and 1% penicillin/streptomycin. All cells were incubated at 37°C with 5% CO_2. Confluent cells were detached from the cell culture flask with trypsin.

Virus preparation and titration

Virus stock of SARS-CoV-2 isolate USA-WA1/2020 (BEI resources) was prepared in Vero E6 cells for one passage and harvested 3 days post-infection. Virus stock and extracellular virus (collected from medium at the respective time points) were tittered on Vero E6 in 96-well focus-forming assay as previously described [367] with minor modifications. Cells (pre-seeded at 2 x 10^4 cells per well in 96-well plate) were inoculated with 50 µl of virus diluted by 5-fold serial

dilutions for about 1 hour at 37 ºC. Next, 100 µl of overlay medium (1% carboxymethylcellulose, 2% FBS, 1X MEM) was added to cells and allowed to incubate in 37 ºC humidified CO_2 incubator for ~ 20-24 hours. After removal of overlay medium and washing with 1X DPBS, cells were fixed with 4% paraformaldehyde (PFA) at room temperature for 30 min. Fixed cells were washed with 1X DPBS and permeabilized with 0.1% Triton X-100 for 10 min, and subsequently blocked with blocking buffer (PBS, 0.1% Tween 20, 5% FBS). Foci were stained with anti-SARS-CoV nucleocapsid rabbit monoclonal antibody (SinoBiological, cat # 40143-R001) at 1:5000 dilution and AlexaFluor 647 conjugated goat anti-rabbit secondary antibody (ThermoFisher Scientific; cat # A27040). Images were taken using Cytation 5 (BioTek) and foci were counted manually in ImageJ to determine virus titer (ffu/ml).

Time-course experiments following SARS-CoV-2 infection

Cells were pre-seeded in 24-well plates a day prior to infection. Cells for imaging were seeded on collagen coated coverslip (Neuvitro Corporation). Synchronous infection was performed by infecting cells (prechilled at 4 ºC for 30 min) with virus (diluted in ice-cold medium) at MOI ~ 1. After incubation at 4 ºC for 1-hour, unbound virus was removed by washing cells with 1X DPBS. Subsequently prewarmed medium was added to cells and plates were incubated in 37 ºC humidified CO_2 incubator to initiate 0 h time point. At various time points (2, 4, 6, 8 hour), medium was removed (or aliquots from duplicate wells were collected for extracellular virus titration) and cells were washed with 1X DPBS. Cells for imaging were fixed with 4% PFA, while cells for RNA extraction was lysed with PureLink RNA Lysis buffer.

Stranded RT-qPCR to quantify SARS-CoV-2 genomic & subgenomic (sg) RNAs (positive-sense (+) & antisense (-))

Total RNA was extracted using PureLink RNA Mini Kit (Thermofisher Scientific), and RNA concentration and purify was assessed using a NanoDrop spectrophotometer (Thermofisher Scientific). 50 ng of RNA was denatured in the presence of primer prior to cDNA amplification and 10 mM dNTPs at 65 °C for 5 min. This was followed by incubation at 4 °C for 5 min. To enrich for antisense (-) vRNA populations, a primer targeting the 5' UTR (CCCAGGTAACAAACCAACCAAC) of SARS-CoV-2 was administered. An oligo$(dT)_{20}$ was used to capture total positive (+)-sense RNAs during cDNA amplification. See **Figure V.1** for more information on cDNA amplification strategies for preferentially capturing SARS-CoV-2 (+ & -) RNAs. cDNA synthesis was performed using SuperScript III First-Strand Synthesis System (Invitrogen), following the manufacturer's instructions. qPCR amplification was conducted using PowerUp SYBR Green Master Mix (Applied Biosystems, Foster City, CA, USA) in a QuantStudio 3 PCR system (Applied Biosystems). The cycle conditions were uracil-DNA glycosylase activation at 50 °C for 2 min, dual-lock DNA polymerase at 95 °C for 2 min, followed by 40 cycles of denaturation at 95 °C for 15 sec, annealing at 55 °C for 15 sec, and extension at 72 °C for 1 min. qPCR primers were designed to target N or NSP12 RNA. N primers (Forward: ACCCCAAAATCAGCGAAATG); Reverse: AGTGAGAGCGGTGAACCAAG). NSP12 primers (Forward: TGTGTACCTTCCTTACCAG; Reverse: ATGAAAGACATCAGCATACTCC). Plasmids harboring an insert for N or NSP12 ORFs, respectively, were used to generate a standard curve and were subject to the same qPCR protocol.

Multiplexed fluorescence imaging of SARS-CoV-2 infection

Samples were prepared for multiplexed *in-situ* imaging as performed in **Chapter II** with modifications. Samples were protease treated at a 1:15 dilution (protease III to PBS) for labeling of sgRNAs and a 1:2 dilution for labeling of gRNAs. The probes used for labeling SARS-CoV-2 nucleic acids can be found in **Table V.3**. For immunostaining, SARS-CoV-2 Nucleocapsid protein (N) protein was labeled with a mouse monoclonal anti-N antibody raised against recombinant, full-length N protein (*Novus Biologicals,* 3865) at a 1:5,000 dilution. A goat anti-mouse secondary antibody conjugated to Atto 555 (Invitrogen) were used for fluorescent detection of N. Z-stacks were deconvolved using Microvoluton, using an experimentally derived point-spread-function (PSF) from our instrument (Nikon C2) at 60x magnification.

RESULTS

The primers and cDNA amplification strategy for our stranded RT-qPCR assay is shown in **Tables V.1-V.2** and **Figure V.1**, respectively. Our RT-qPCR expression assay allows us to track strand-specific synthesis of SARS-CoV-2 sgRNA and gRNA across different cell-types. In **Figure V.2,** we demonstrated the kinetics of bulk viral replication across Vero E6, Caco2, and Calu3 cells. Vero cells are most susceptible to SARS-CoV-2 infection relative to the Caco2 and Calu3 cell lines. vRNA production accelerated between two and four h.p.i in the Vero E6 cells, while replication began to exponentially increase in the Caco2 and Calu3 cells between four and six h.p.i. SARS-CoV-2 replication kinetics in the Caco2 and Calu3 cells are therefore slower relative to infection in the Vero E6 cells.

TABLE V.1. Primers for cDNA library preparation for RT-qPCR assays.

Primer	Target	Template for qPCR
SEM467	5' UTR	(-) vRNA
Oligo-$(dT)_{20}$	Total (+) RNA	(+) vRNA & (+) cellular RNA

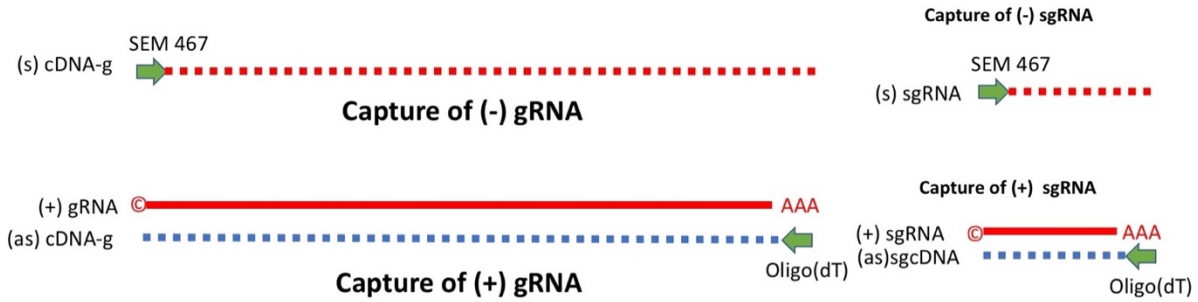


Figure V.1. cDNA synthesis strategy for stranded RT-qPCR assays to capture SARS-CoV-2 transcripts. The SEM 467 primer targets the 5' UTR of SARS-CoV-2 transcripts. This allows capture of antisense (-) vRNA. Oligo(dT) allows capture of (+) vRNA, as well as total poly(A) (+) transcripts present in the virus-host transcriptome. Stranded cDNA library generation of (+) & (-) vRNA species is then followed by qPCR.

TABLE V.2. Primers for qPCR of stranded cDNA libraries of (+) & (-) vRNAs.

Target	Primer		Sequence (5'-3')
nsp12	WA1-15918	forward	TGTGTACCTTCCTTACCCAG
	cWA1-16086	reverse	ATGAAAGACATCAGCATACTCC

Target	Primer		Sequence (5'-3')
N	WA1-28288	forward	ACCCCAAAATCAGCGAAATG
	cWA1-28444	reverse	AGTGAGAGCGGTGAACCAAG

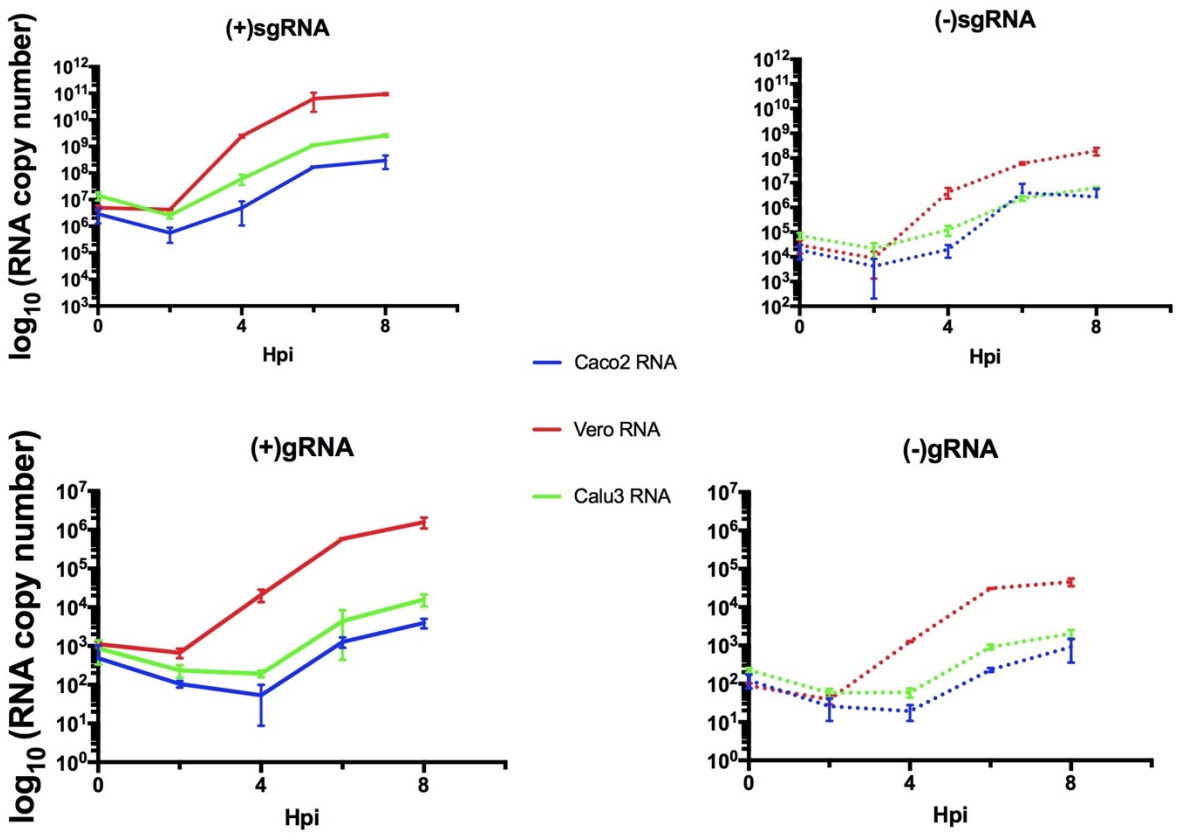


Figure V.2. Stranded RT-qPCR of different cell-types to track early SARS-CoV-2 replication kinetics. Expression of positive (+) and antisense (-) vRNA from 0-8 h.p.i. Top two panels show expression of subgenomic RNA (sgRNA), and the bottom two panels show expression of full-length gRNA. Experiments are performed in biological replicates for three cell types listed (Caco2, Vero E6, & Calu3). cDNA synthesis strategy can be seen in **Figure V.1** and qPCR targets are shown in **Table V.2.** Solid lines = (+) vRNA; dotted lines = (-) vRNA. RNA was collected from a total of 1 X 10^5 cells following infection (MOI ~ 1) for each replicate.

Oligonucleotide probes for SARS-CoV-2 smFISH experiments can be found in **Table V.3**. Imaging of Vero-infected cells allowed us to demonstrate the kinetics of early SARS-CoV-2 replication. At eight h.p.i, we observed viral spread and monitored viral replication at the single-cell level (**Figures V.3-V.4**). Patterns of expression observed *via* imaging coincide with vRNA copies quantified using our stranded RT-qPCR expression assay. Immunostaining of viral protein, NSP8 (red), is also performed concomitantly with RNA hybridization and presented in

Figures V.3-V.4. NSP8 is a catalytic subunit of the RTC [363, 364]. In **Figure V.3**, we visualized high expression of (+) sgRNA amongst infected Vero E6 cells by only four h.p.i, consistent with the number of (+) sgRNA copies quantified *via* RT-qPCR (**Figure V.2**). By six h.p.i, we observed high levels of NSP8 expression in infected cells. In **Figure V.4**, we can visualize the heterogeneity of vRNA and NSP8 expression across the cluster of infected Vero E6 cells.

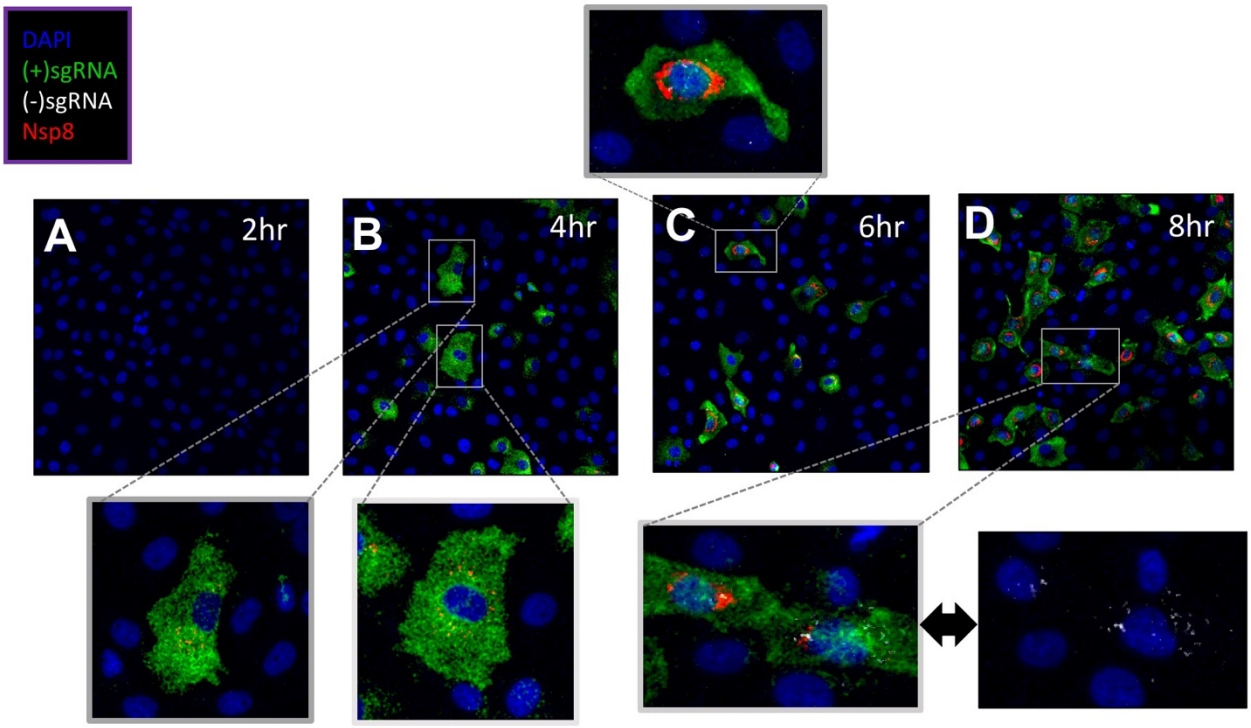


Figure V.3. Time-course of SARS-CoV-2 early replication kinetics in Vero E6 cells. Infected Vero cells were fixed and imaged at 2 (**A**), 4 (**B**), 6 (**C**), and 8 h.p.i (**D**). vRNA production can be visualized 4 h.p.i ((+) sgRNA (green)). In panel **B**, gray boxes denote two cells expressing an abundance of (+) sgRNA and beginning translation of viral protein (NSP8; red). In panel **D**, the two bottom-right images show a close-up of infected cells and their respective viral transcriptional states. The left image is the merge of the two cells that are demarcated by the gray box in the zoomed-out image at 8 h.p.i, and the image to the far-right is displaying only the (-) sgRNA channel (white). DAPI is pseudo-colored blue, (+) sgRNA is green, (-) sgRNA is white, and NSP8 protein is red.

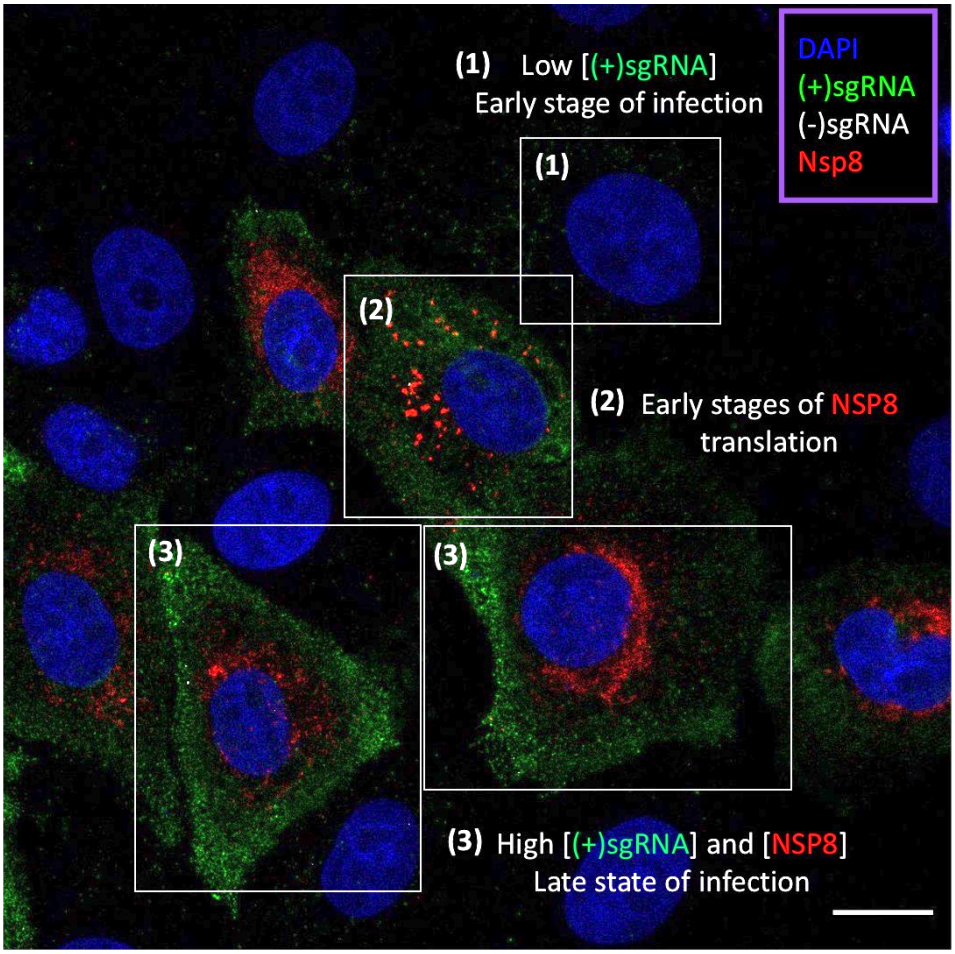


Figure V.4. Tracking cell-to-cell variability following SARS-CoV-2 replication. Split channels show DAPI (blue; nuclei), (+) sgRNA (green), (-) sgRNA (white), and NSP8 protein staining (red). Cluster of infected Vero E6 cells 8 h.p.i. Box **(1)** highlights a cell displaying early replication events (low level of (+) sgRNA staining), box **(2)** depicts early stages of NSP8 translation, and box **(3)** highlights two cells that are producing high amounts of vRNA and protein. Box **(3)** also indicates infected cells that are likely producing infectious virions by 8 h.p.i. Scale bars represent 10 µm. Images were taken at 60x magnification with the Nikon C2 confocal microscope. Images are shown as single slices.

TABLE V.3. Oligonucleotide hybridization probes for fluorescent detection of SARS-CoV-2 transcripts.

Oligonucleotide Probe ID	Viral Genomic Target	Channel	Special Notes
V-SARS-CoV-2-N-O1-C2	(+)sgRNA	C2	Dilute in RNAscope diluent
V-SARS-C0V-2-N-O2-sense-C3	(-)sgRNA	C3	Dilute in RNAscope diluent
V-SARS-CoV-2-orf1ab	(+)gRNA	C1	Dilute in DNA hybridization buffer **(Table II.3)**
V-SARS-CoV-2-orf1ab-O1-sense-C2	(-)gRNA	C2	Dilute in DNA hybridization buffer **(Table II.3)**

DISUCSSION

Our two orthogonal quantitative assays, stranded RT-qPCR and smFISH, provide robust and highly sensitive complementary methodologies for tracking SARS-CoV-2 replication across multiple cell-types. While RT-qPCR assays enable us to track bulk vRNA expression across the infected cellular population, our smFISH technique allows us to not only visualize the variability of replication across individual infected cells but also provides a method to monitor viral spread, measure viral transcriptional kinetics (**Figure V.3**), and observe the spatial compartmentalization of putative virus-host factor interactions. At eight h.p.i, we observe cells only expressing (+) sgRNA and no viral protein. Proximal to these cells, we observe infected cells that have already expressed an abundance of vRNA and viral protein, suggesting productive virions are budding from these respective cells, leading to very rapid SARS-CoV-2 cell-to-cell spread.

Ongoing work is looking to pinpoint incoming virions prior to viral entry, quantify viral transcriptional bursts and viral protein synthesis, and quantitatively track cell-to-cell viral spread.

Collectively, this work will allow us to improve our spatiotemporal understanding of critical SARS-CoV-2 replication events.

Further ongoing work consists of optimization of fluorescently labeling full-length viral gRNAs for our imaging studies. These vRNA species are scarce relative to sgRNAs (**Figure V.2**) and likely exhibit complex secondary structures, hindering efficient probe hybridization. We are introducing a pre-denaturation step prior to target oligonucleotide probe hybridization (incubate at 90 °C for 1 hr), where we also incubate our cell samples in denaturant or conditions that accelerate probe hybridization (90% DMSO, 70% formamide). These conditions may promote improved fluorescence-based detection of viral gRNAs. In addition, we are applying our strand-specific assays to quantify changes in vRNAs following treatment with a panel of small molecules. The FISH probes we are using for our studies are compatible with super-resolution stimulated emission depletion (STED) microscopy, which we are utilizing to visualize co-localization of viral and cellular factors localized at putative RTCs. Overall, this work provides an innovative platform for single-cell fluorescence-based simultaneous detection of SARS-CoV-2 RNAs and protein to track viral replication with improved spatiotemporal resolution.

CHAPTER VI: MULTIOMICS PROFILING REVEALS A UNIQUE CHROMATIN SIGNATURE ASSOCIATED WITH ACTIVE HIV-1 PROVIRUSES

The research and data presented in this chapter will be included in a future manuscript.

Chapter VI contains data & analysis that will be included in:

Shah, R., Gallardo, C.M., Dixon, J.R., Jung, Y.H., McFadden, W., Torbett, B., Corces, V.G., Tedbury, P.R., Sarafianos, S.G. (2021) Multiomics profiling reveals unique chromatin signature associated with active HIV-1 proviruses. (Planned submission to *PNAS*)

ABSTRACT

The possible effects of lentiviral integration on restructuring of chromatin architecture are not well understood. We interrogated genome-wide chromatin organization and the structure of chromatin around HIV-1 integration sites using Hi-C and ATAC-seq, respectively, and examined RNA transcription of the provirus and neighboring genes in HIV-inducible cellular models. We found chromatin interaction networks around integrated HIV-1 are predominantly preserved with respect to uninfected cells, consistent with a lack of association between HIV-1 integration and major chromatin remodeling. Instead, we find that induction of proviral transcription leads to stark local changes in chromatin accessibility downstream from the HIV-1 3' LTR, demonstrating how HIV-1 can directly alter local cellular chromatin structure post-integration. Using long-read Nanopore RNA-seq, we interrogated the local host and HIV transcriptomes, and observe that 1-5% of HIV-1 transcripts initiated at the HIV-1 5' LTR promoter, extend into the flanking cellular genome, generating long chimeric virus-host RNAs. Despite provirus-driven read-through, HIV-1 appears to have only a modest effect on local cellular splicing patterns; however, HIV-1 integration may attenuate expression of full-length cellular RNAs if HIV-1 and the respective host gene are in the same transcriptional orientation. HIV-driven read-through resembles cellular stress-induced transcriptional reprogramming. Our studies provide an in-depth investigation of the impact of HIV-1 integration and transcription on 3D chromatin organization, chromatin accessibility, and cellular transcriptional patterns. This work provides a mechanism of how integrated HIV-1 may perturb local cellular transcription, while still preserving overall cellular chromatin structure.

SIGNIFICANCE STATEMENT

Applications of joint chromatin and RNA profiling of infected cells can provide insight into important virus-host factor interactions that regulate viral gene expression and replication. Our studies provide an in-depth investigation of the impact of HIV-1 integration and transcription on 3D chromatin organization, chromatin accessibility, and the local cellular and HIV-1 transcriptomes. This work provides a mechanism of how integrated HIV-1 may perturb local cellular transcription. This work has broad relevance to HIV-1 latency research and may be important in evaluating potential off-target effects of lentiviral-based genetic therapies. Our work serves as a foundation for further in-depth multiomics studies of HIV-1 chromatin structure, splicing analysis of complex RNAs, and retroviral gene regulation.

INTRODUCTION

HIV-1 is a retrovirus that targets immune cells including T cells [2, 3], macrophages [4], and dendritic cells [5], establishing permanent infection by integrating a double-stranded complementary DNA (cDNA) copy of its RNA genome into the targeted cellular chromatin [6]. Prior to integration, HIV nuclear import is likely driven by interactions with capsid (CA) core [30, 31, 173] and host factors that facilitate nuclear entry of the pre-integration complex (PIC). The PIC, composed of HIV-1 integrase and the viral cDNA genome, interacts with nuclear factors such as the cellular coactivator lens epithelium-derived growth factor (LEDGF/p75), which help direct the PIC to intronic sites in chromatin-accessible and active genes for integration [6, 94, 159]. Post-

integration, the virus exploits the cellular transcriptional machinery to produce the genomic and subgenomic RNAs required for viral replication.

The transcription of integrated HIV-1 genomes (provirus) is regulated by viral and cellular factors [47, 218, 239, 277]. Factors that may influence HIV transcription include location of integration site and local proviral nuclear environment [368], transcriptional orientation of the provirus relative to the respective flanking host gene [222], and transcription factor occupancy at the HIV-1 5' LTR promoter [369]. Researchers have begun to investigate the association between integration site selection and higher-order nuclear organization [159, 167, 370]. A recent study demonstrated that HIV-1 reoccurring integration sites are enriched proximally to super-enhancers (SE), intricate chromatin structures that drive expression of critical regulatory gene networks [170]. Another study applied integration site analysis and live-cell imaging studies to show that HIV-1 PICs traffic and accumulate at nuclear speckles with a preference for proviral integration into speckle-associated domains (SPADs) [94]. SPADs and SE networks are generally found in highly structured genomic compartments with elevated transcriptional activity [371]. While there appears to be a direct relationship between HIV-1 integration preference and hierarchical nuclear compartmentalization, we do not have a clear understanding of how higher-order chromatin structure affects HIV-1 transcription or whether the HIV-1 provirus alters the nuclear environment and transcriptome, globally or in the vicinity of the integration site.

The spatial organization of the genome has been probed using chromosome conformation capture (3C) methods (based on chemical cross-linking of chromatin in cells) to identify regions of chromatin that are in close proximity to one another. The 3C-based method, Hi-C, has been used to capture the global chromatin contact network of cellular populations, revealing the intricate hierarchal organization of nuclear chromatin [283-286, 289, 372, 373]. Additionally, Assay for

Transposase-Accessible Chromatin using sequencing (ATAC-seq) has been developed as a tool for genome-wide interrogation of the functional state of chromatin, nucleosome positioning, degree of chromatin openness, and transcription factor occupancy [296, 328, 329]. Together, these NGS-based assays have revealed that the genome can be split into two self-associating compartments (compartment A and B) with differential chromatin accessibility states. Compartment A is mainly composed of euchromatin and localized towards the nuclear interior, while compartment B is mainly composed of heterochromatin [291-293] and enriched towards the nuclear periphery [299, 300]. Euchromatin is highly accessible to nuclear factors and packed with histones marked with active epigenetic modifications, promoting RNA polymerase occupancy and transcriptional activity, conversely, heterochromatin contains densely packed nucleosomes with repressive epigenetic modifications and is often transcriptionally silent [301]. The discovery of A and B compartments provided an early example of the role that the spatial position of a gene can play in regulation of its expression.

A further level of organization is provided by chromosome looping to form topologically associating domains (TADs), structural genomic units that vary in size from ~40 kb to 1 Mb [288-290] and are characterized by sharp boundaries that promote long-range interactions [291, 292]; for example, *cis*-regulatory sequences that are far apart on the chromosome may be brought close together by the formation of TADs [287]. Protein complexes at TAD boundaries insulate these domains from one another, permitting the formation of discrete epigenetic environments and providing another level of transcriptional control. Interphase chromatin is highly dynamic, and as the transcriptional state of a TAD changes, chromatin contacts within the domain are altered,

remodeling chromatin architecture [293]. The dynamic formation and dissolution of TADs is mediated by the action of CCCTC-binding factor (CTCF) in complex with cohesin [290, 295, 296, 298].

T cell activation induces large scale nuclear reorganization, repositioning SE networks proximal to the nuclear pore and dramatically altering cellular transcriptional patterns and cellular function [170, 371]. Viruses also exploit the regulatory effects of position with the nucleus. Integration into heterochromatin mediated by knock-down of the HUSH complex has been shown to repress retroviral transcription, demonstrating the position effect variegation of viral gene expression [374].Viruses have been shown to also interact with cellular chromatin structure to regulate viral transcription, as well alter local cellular transcriptional patterns *via* virus-induced chromatin remodeling. The Epstein-Barr virus (EBV) genome preferentially associates with repressive nuclear compartments during latency and with active compartments during viral transcriptional activation, with its own transcription regulated alongside that of the host [375]. Conversely, Human T-cell leukemia virus type 1 (HTLV-1) integration directly restructures cellular chromatin near viral genome upon integration. The HTLV-1 genome contains a CTCF binding site that drives formation of novel chromatin loops and dysregulation of local host gene transcription [337-339]. Through applications of a high-through viral chromosome conformation capture assay to measure virus-host chromatin interactions, the parvovirus Minute Virus of Mice (MVM) genome was shown to be associated with previously identified TADs. This study suggested that MVM was interacting with highly ordered regions of the cellular genome [309]. White spot syndrome virus (WSSV) globally influences cellular chromatin structure by decreasing compaction of chromatin through activity of viral protein 9 (VP9) [376]. Host chromatin organization is a key regulatory feature, essential to the correct function of the cell, and

vulnerable to exploitation and manipulation by a broad spectrum of viruses through multiple mechanisms.

Work performed with HIV-1 has examined the influence of integration site epigenetic environment and integration site interactions in regulating HIV-1 transcription. These studies suggest that chromatin functional states (active *vs.* repressed) and proviral nuclear positioning (compartmentalization) are important for HIV-1 gene regulation [32, 196, 374]. By contrast, little is known about the impact of HIV-1 on the host chromatin. Unlike HTLV-1, the HIV-1 genome does not appear to contain a CTCF binding-site [337]; however, the impact of HIV-1 integration and gene expression on chromatin organization has not been directly investigated.

In this work, we have applied high-resolution Hi-C to examine long range chromatin interactions throughout the genome in uninfected T cells, T cells harboring an HIV-1 provirus, and T cells harboring an actively transcribing HIV-1 provirus; we have considered the potential for both global changes in chromatin organization and changes local to the integration site. Additionally, we interrogated chromatin accessibility around sites of HIV-1 integration *via* ATAC-seq. Finally, by pairing high-resolution chromatin mapping methods to RNA-seq, we are able to associate changes in chromatin organization with changes in virus and host transcription. We have applied this multiomics approach to evaluate how HIV-1 affects chromatin architecture and host transcription in the vicinity of viral integration sites.

RESULTS

Characterization of inducible J-Lat cellular models for multiomics profiling studies

For studies of chromatin structure, we require clonal cell populations, as several of the methods cannot be reliably applied to single cells. In particular, to allow us to study the

chromatin local to the HIV-1 provirus, we require cells with known proviral integration sites; additionally, to determine whether any observed phenotype depends merely on the insertion of the HIV-1 genome or requires HIV-1 transcription, we required a cell system where HIV-1 transcription can be regulated. For these reasons, we selected the well-established J-Lat cell lines. Clonal J-Lat models are derived from Jurkat T cells latently infected with an HIV-1 provirus carrying an LTR-driven GFP reporter; HIV-1 transcription can be activated by treatment with TNF-**α** [185, 270]. Using HIV-inactive clonal J-Lat cell models that can be activated enabled us to investigate chromatin structure at HIV-1 sites of integration pre- and post-activation of transcription. Following treatment with TNF-**α**, J-Lat cells can be sorted into HIV-inactive (GFP-) and -active (GFP+) populations for our NGS-based studies (cell-sorting scheme in **Figure VI.S.1)**. To probe how HIV-1 transcriptional orientation ((5' → 3') or (3' ←5')) relative to the flanking host gene and how local nuclear environment impacts HIV-1 transcription, we used two J-Lat cell lines with distinct integration sites (J-Lat 10.6 and 8.4). Within the J-Lat 10.6 model, the provirus has integrated in the opposite orientation relative to the host gene (*SEC16A)*; in the J-Lat 8.4 model, the provirus and the host gene (*FUBP1*) are in the same transcriptional orientation (**Figure VI.1A and VI.1B)** [185, 275]. Assessment of occupancy of active epigenetic markers (H3K27ac, Pol2, BRD4) and higher-order chromatin structures (SPADs and super-enhancers) present proximally to the J-Lat 10.6 provirus suggests that within this model, HIV-1 has integrated into a transcriptionally primed environment. There is significantly less enrichment of these factors proximal to the J-Lat 8.4 provirus (**Figure VI.1C)**. However, within both J-Lat clonal systems, HIV-1 has integrated within what appears to be a canonical integration site (intron within transcriptionally active gene) [167, 169]. In the case of these two models, local proviral epigenetic landscape was a good predictor of HIV transcriptional

competency, as the J-Lat 10.6 model not only activates at a higher frequency following TNF-**α** stimulation but also activates at a higher transcriptional threshold compared to the J-Lat 8.4 model **(Figure VL.1D**). There is over a 10-fold greater proviral transcriptional output in the J-Lat 10.6 model *vs.* the 8.4. These two models provide two distinct chromatin environments for our downstream studies.

A

J-Lat clone	Chromosome[a]	Position (bp)[b]	Host Gene[c]	Host Gene Orientation[d]	HIV Orientation[e]
10.6	9	136,468,579	SEC16A	-	+
8.4	1	77,946,384	FUBP1	-	-

[a] Chromosome where HIV provirus integrated
[b] First human genome nucleotide following 3'-end LTR
[c] Cellular gene where HIV integrated
[d-e] Transcription direction of the host or HIV gene

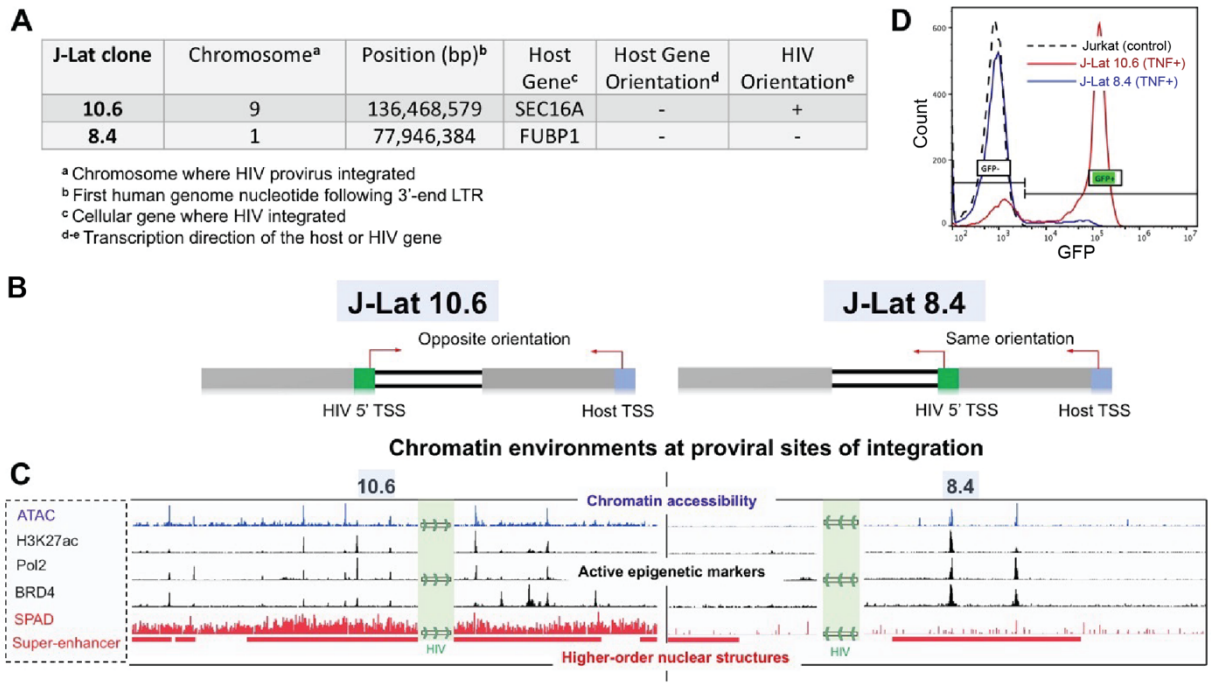


Figure VI.1. Variability in chromatin environments at HIV-1 sites of integration. Table adapted from Symons *et al.* 2018 [275], displaying HIV-1 site of integration in J-Lat 10.6 & 8.4 models. The two models were selected for differences in integration orientation **(A).** Schematic depicting orientation of HIV-1 provirus and respective host gene in J-Lat 10.6 & 8.4 models. In the 10.6 model, the provirus and host gene are in opposite transcriptional orientations, whereas in the 8.4 model, HIV-1 and *FUBP1* are in the same transcriptional orientation (3' ← 5') **(B).** Epigenomic profile of chromatin environment at representative sites of integration. ATAC-seq (blue) is an assay to quantify chromatin accessibility on a genome-wide scale, ChIP-seq against H3K27ac, Pol2, and BRD4 (black tracks) are active enhancer/promoter markers, and speckle-associated domains (SPADs) and super-enhancers (both red tracks) are higher-order nuclear structures that are putative targets for HIV-1 integration [94, 170]. HIV marker (green) denoting exact sites of integration is not to scale. Sequencing tracks are visualized using Integrated Genomics Viewer (IGV) [32] **(C).** Flow cytometric analysis of J-Lat activation potentials. Cells are treated with TNF-**α**. HIV-active cells are GFP+. Jurkat T cells (no reporter provirus) serve as a gating control **(D).**

Hi-C profiling to map the chromatin contact network of HIV-infected cells

In order to evaluate the impact of proviral transcriptional activation on higher-order chromatin structure, we performed Hi-C assays in wild-type (WT) Jurkat and J-Lat model T-cell lines. First, in order to characterize the native local 3D folding in the Jurkat cell line, we deeply sequenced Hi-C libraries and performed Hi-C analysis (**Figure VI.2A-B**). We obtained ~680 million contacts in the WT Jurkat cells, sufficient for analysis of the 3D chromatin folding at ~1 kb resolution. We then examined the chromatin interaction profiles in the vicinity of the two integration sites in J-Lat 10.6 and 8.4 cell lines (within *SEC16A* gene on chromosome 9 and *FUBP1* gene on chromosome 1, respectively). In both cases, in WT Jurkat T cells (absence of virus) these genes are relatively devoid of any clear features of higher-order chromatin structure. However, in both cases, there are strong looping events several kilobases downstream the respective HIV-1 integration sites. In the J-Lat 10.6 cells, this is a loop surrounding the *NOTCH1* gene, while in the 8.4 cells, this is a loop surrounding the *PTGFR* gene.

Next, to examine the impact of proviral integration and transcription on 3D chromatin folding in the nucleus overall and in the vicinity of the provirus, we performed Hi-C in the J-Lat 10.6 and 8.4 cell lines under untreated conditions or following activation of HIV-1 transcription with TNF-**α** (**Figure VI.2C**). For the J-Lat 10.6 cell line, we obtained 823 and 641 million contacts in the untreated and TNF-**α** treated conditions, respectively. In the J-Lat 8.4 cell line, we obtained 578 and 549 million contacts for the untreated and treated conditions, respectively. Mapped reads of all Hi-C libraries in this study are shown in **Figure VI.S.2**. Examining the patterns of chromosome folding at the sites of integration, we observe only mild differences in chromatin structure in the context of integration or proviral activation (**Figure VI.2D**). In both the 10.6 and 8.4 cell lines, there are no major changes in chromatin interaction frequency at

either the sites of integration, or in the downstream loops proximal to each gene. We do observe subtle difference at specific loci. For example, the loop downstream from the *SEC16A* integration site on chromosome 9 in the J-Lat 10.6 cell line shows modestly reduced chromatin interaction frequency, though the differences are not significant (**Figure VI.2E**). In addition, at the *FUBP1* site on chromosome 1, the *FUBP1* gene does show increased interactions with an upstream region near the *MIGA1* gene (**Figure VI.2F**), but these differences are modest and there are no other notable changes in chromatin interaction frequency at the *FUBP1* locus. Statistical significance is determined using a Wilcoxon signed-ranked test, assessing differences in direct chromatin interactions at the respective genetic locus. In both models, HIV-1 is also found in nuclear Compartment A in both transcriptional states. In summary, 3D genome organization is preserved following HIV-1 integration or proviral transcriptional activation, suggesting large scale chromatin remodeling of HIV-1 is not a pre-requisite of virus transcription and replication. In addition to the lack of change in the local 3D chromatin organization around HIV-1 proviruses, we do not observe any significant global changes in chromatin architecture of HIV-infected cells relative to the uninfected WT Jurkat T cells.

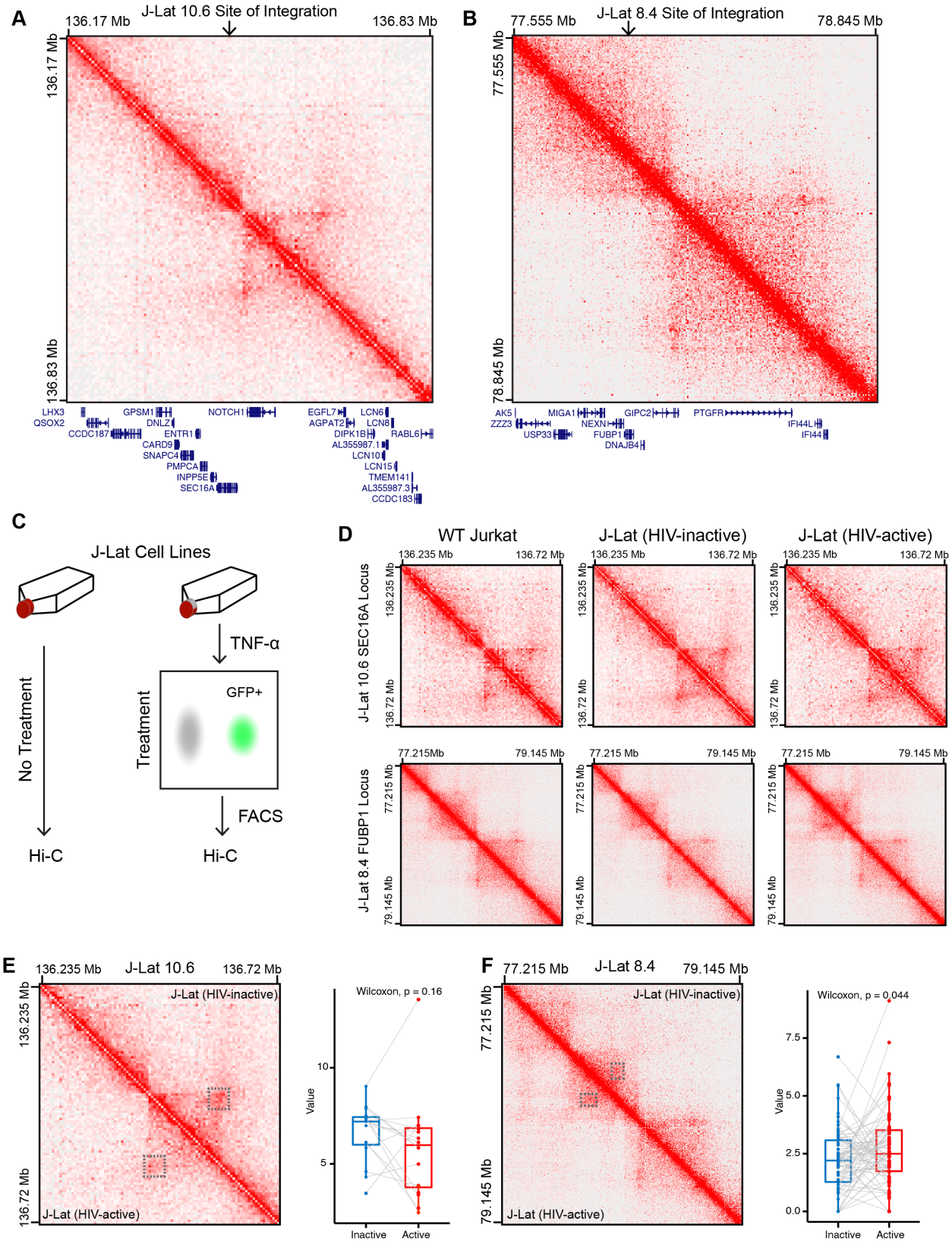

A
136.17 Mb
J-Lat 10.6 Site of Integration
136.83 Mb
136.17 Mb
136.83 Mb
LHX3
QSOX2
CCDC187
GPSM1
DNLZ
ENTR1
CARD9
SNAPC4
PMPCA
INPP5E
SEC16A
NOTCH1
EGFL7
AGPAT2
DIPK1B
AL355987.1
LCN10
LCN15
TMEM141
AL355987.3
CCDC183
LCN6
LCN8
RABL6
B
77.555 Mb
J-Lat 8.4 Site of Integration
78.845 Mb
77.555 Mb
78.845 Mb
AK5
ZZZ3
USP33
MIGA1
NEXN
FUBP1
DNAJB4
GIPC2
PTGFR
IFI44L
IFI44
C
J-Lat Cell Lines
TNF-α
No Treatment
Treatment
GFP+
FACS
Hi-C
Hi-C
D
WT Jurkat
J-Lat (HIV-inactive)
J-Lat (HIV-active)
J-Lat 10.6 SEC16A Locus
J-Lat 8.4 FUBP1 Locus
136.235 Mb
136.72 Mb
77.215Mb
77.215 Mb
79.145 Mb
E
136.235 Mb
J-Lat 10.6
136.72 Mb
J-Lat (HIV-inactive)
J-Lat (HIV-active)
Wilcoxon, p = 0.16
Value
Inactive
Active
F
77.215 Mb
J-Lat 8.4
79.145 Mb
J-Lat (HIV-inactive)
J-Lat (HIV-active)
Wilcoxon, p = 0.044
Value
Inactive
Active

Figure VI.2. Minimal changes in integration site higher-order chromatin structure upon HIV-1 transcriptional activation. Hi-C interaction map in wild-type Jurkat T-cells of the region on chr9 where integration site is located in the J-Lat 10.6 cell line. The arrow indicates the site of integration, within the *SEC16A* gene. Below the heat map are the location of genes at the locus, including *SEC16A* (**A**). Similar as in panel **(A)** but depicting the region on chr1 where the integration site is located in the J-Lat 8.4 model within the *FUBP1* gene (**B**). Schematic showing the experimental design of Hi-C in the context of HIV-1 transcriptional induction. TNF-**α** is added for 24 hours to the J-Lat model lines. GFP+ cells that have successfully activated and are then sorted and processed for Hi-C. Untreated J-Lat cells are used as controls **(C)**. Hi-C heat maps of the integration sites in WT Jurkat cells (left column), untreated J-Lat lines (middle column), and HIV-active (right column), for the J-Lat 10.6 sample (top row) and the J-Lat 8.4 sample (bottom row) **(D)**. Comparison of the *SEC16A* locus in the J-Lat 10.6 cell line. The heat map shows the untreated conditions in the upper right-hand half of the heat map, and the activated population in the lower left-hand. Dashed lines mark a loop downstream from the *SEC16A* locus that surrounds the *NOTCH1* gene. The plot on the right shows the quantification of the observed/expected Hi-C interaction frequencies for the HIV-inactive (blue) &-activated (yellow) conditions of pixels within the dashed line. The loop generally is weakened upon HIV activation, but the results do not reach a threshold for statistical significance (p=0.16, Wilcoxon) **(E)**. Comparison of the *FUBP1* locus in the J-Lat 8.4 cell line. The heat map shows the untreated conditions in the upper right-hand half of the heat map, and the activated cells in the lower left-hand. Dashed lines mark a region of increased interactions between the *FUBP1* gene and a region upstream. The plot on the right shows the quantification of the observed/expected Hi-C interaction frequencies for the HIV-inactive (blue) & -active (yellow) conditions of pixels within the dashed line (p=0.0.44, Wilcoxon) **(F)**.

Chromatin and transcriptional profiling at sites of HIV integration reveals a unique chromatin signature associated with active proviruses

To understand how HIV-1 integration and proviral transcriptional activation affect the local chromatin functional state, we applied ATAC-seq to our WT Jurkat and J-Lat T cell models. The distribution of ATAC reads used for analysis is shown in **Figure VI.S.3**. Analysis was performed filtering for reads between 1-115 base pairs (bp). In a novel result, we found that following HIV-1 transcriptional activation, chromatin accessibility increases downstream the HIV-1 3' LTR into the cellular chromatin within both J-Lat models **(Figure VI.3).** The effect was significantly more apparent within the J-Lat 10.6 model, where following proviral activation, we not only saw an increase in overall HIV-1 chromatin accessibility but the most prominent change in chromatin openness was downstream the 3' LTR. In the J-Lat 8.4 model,

following proviral activation, there is little to no change in ATAC density observed within the proviral chromatin; however, we observe a modest increase in accessibility downstream the 3' LTR (**Figures VI.3C and VI.S.4**). An increase in accessibility downstream from the 3' LTR into the flanking host chromatin is consistent across both models.

In addition to the use of ATAC-seq to reveal a change in chromatin accessibility at the downstream integrated HIV-1, we also applied Illumina-based RNA-seq to measure changes in transcription. Comparing activated and inactive J-Lat 10.6 and 8.4 cell lines, we observed an increase in read density downstream from the 3' LTR in both models following HIV-1 activation. The strandedness of the RNA fragments corresponding to this flanking region are in the same transcriptional orientation as the respective HIV-1 genome, supporting the observation that HIV-driven transcription can run into the cellular genome [220, 222, 377]. In the Jurkat T cells, we found that there was no difference in chromatin accessibility and gene expression +/- TNF-**α** at the SEC16A and FUBP1 loci, suggesting the observed changes in chromatin organization is an HIV-specific effect and not due to TNF-**α** stimulation. In addition, there is no change in chromatin accessibility within the regulatory regions of the proximal host genes, further suggesting changes in chromatin organization are localized to the provirus and the region directly downstream the 3' LTR following HIV-1 transcriptional activation. Assessment of chromatin accessibility of constitutively active house-keeping genes such as GAPDH reveals the most prominent chromatin accessibility towards the 5' gene promoter and not the 3' end of the gene, in contrast to what we have shown for active HIV-1 chromatin accessibility (**Figure VI.S.5)**.

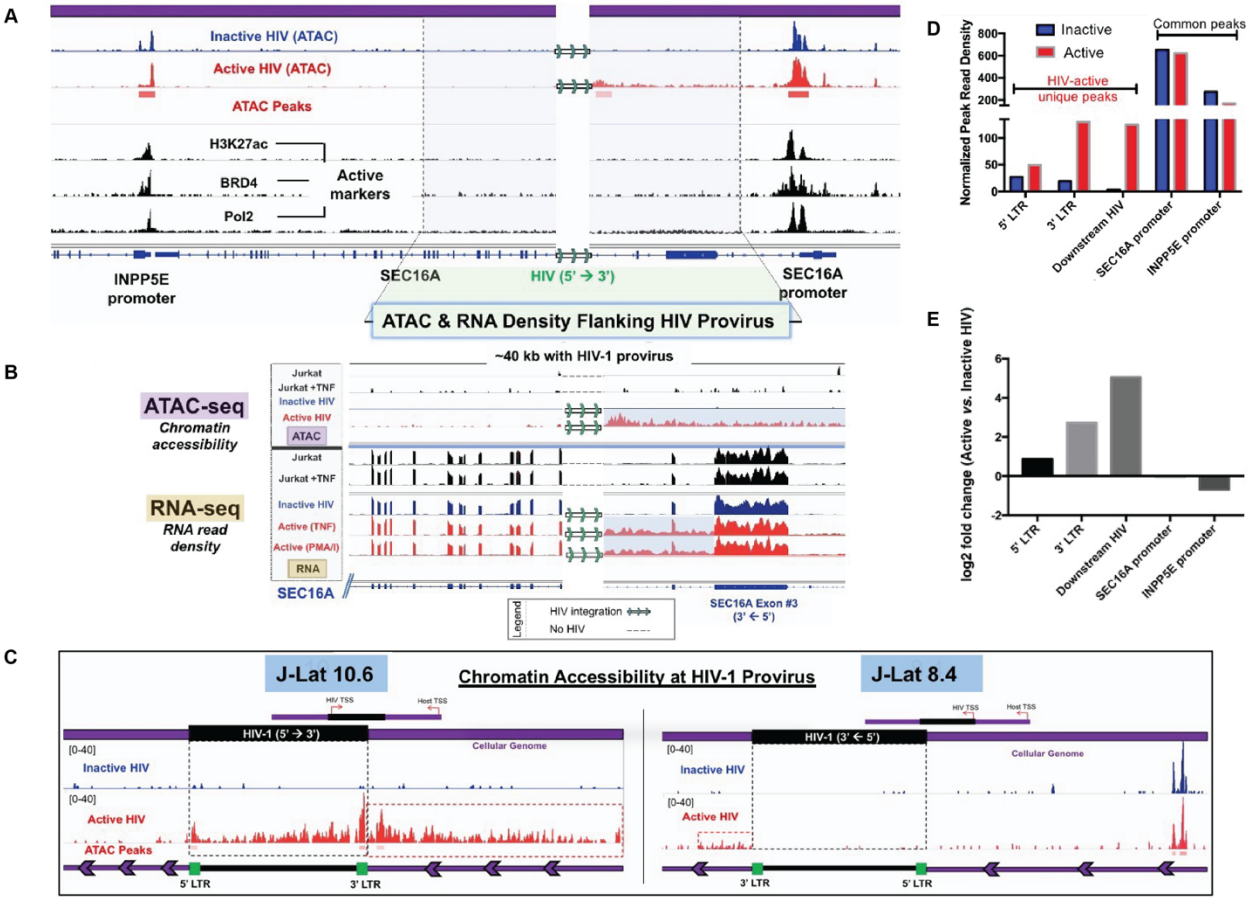


Figure VI.3. ATAC- & RNA-seq read density at HIV-1 integration sites to profile local HIV-host chromatin environment. Representative ATAC-seq tracks of inactive (blue) and active (red) populations of J-Lat 10.6 cells. ChIP-seq of Jurkat T cells (black) against active enhancers/promoters (H3K27ac, Pol2, & BRD4) are superimposed to demarcate host promoters (*INPP5E* & *SEC16A*). **(A)** Close-up of the highlighted region in panel **(A)**, visualizing ATAC & RNA (Illumina-based) density of the flanking host genome (reads aligned to the human reference genome (hg38)). Sequencing tracks to the left are upstream HIV-1, flanking the 5' LTR, and tracks to the right are downstream. The highlighted regions in ATAC and RNA density demarcate regions of increased read pileup directly downstream the activated provirus. The sequencing tracks in black are of uninfected Jurkat T cells +/- TNF-**α** to visualize what the native chromatin and transcriptional state is at this site of proviral integration. The host gene, *SEC16A,* has a 3' ← 5' orientation, whereas HIV-1 has the opposite 5' → 3' transcriptional orientation **(B).** ATAC read density at the proviral site of integration in the J-Lat 10.6 & 8.4 models. Reads were mapped to a concatenated hg38 & custom HIV-1 proviral genome. The red dotted boxes highlight regions of increased chromatin accessibility downstream the HIV 3' LTR following proviral transcriptional activation. HIV-1 and host transcriptional start sites (TSS) are shown to bring attention to transcriptional orientation of HIV-1 and the respective host genes in both J-Lat models. Chromatin accessibility (ATAC density) is observed downstream HIV-1 in both models, despite differences in this integration orientation **(C).** Quantification of differential ATAC read density at provirus, flanking cellular genome, and proximal host promoters in the J-Lat 10.6 cells (inactive *vs.* active states). Differential peak analysis performed using biological replicates for each condition. All ATAC- & RNA-seq experiments were performed with biological replicates **(D-E).**

Proviral transcriptional read-through observed regardless of the transcriptional orientation of HIV-1 relative to the flanking host gene

To visualize the dynamics of transcriptional read-through at HIV-host intergenic boundaries, we applied stranded single-molecule RNA FISH (smRNA FISH), monitoring expression of positive (+) and antisense (-) HIV-1 RNA [317, 318] in HIV-inactive and -active states. Our smFISH imaging reveals that in the J-Lat 10.6 model, there is expression of antisense HIV-1 (-) RNA in the inactive state **(Figure VI.4A)**. The HIV-1 (-) RNA remains within the nucleus, suggesting it is not exported to the cytoplasm for translation. Following proviral activation of the J-Lat 10.6 cells, we can observe expression of HIV-1 (+) RNA as expected. The viral RNA (vRNA) signal disappears following RNase treatment, leaving the single HIV-1 DNA foci (**Figure VI.4A)**. In the J-Lat 8.4 model, similarly, we observe expression of vRNA in the inactive cells. However, in this case, we are seeing expression of HIV-1 (+) RNA. While there are a few vRNA copies that were exported into the cytoplasm, there is no detectable GFP or HIV-1 protein expression in the inactive 8.4 cells [378]. Since we are observing HIV-1 (-) RNA in the inactive 10.6 cells, (provirus and SEC16A opposite transcriptional orientations) and are observing HIV-1 (+) RNA in the inactive 8.4 cells (provirus and FUBP1 are in the same transcriptional orientation), we suspect these are chimeric RNAs that are the by-product of host-driven transcriptional read-through, where transcription originates at the proximal host promoter and runs into the HIV genome. Maximal intensity projections of larger fields of view show the homogeneity of expression of these host-driven chimeric RNAs across the HIV-inactive clonal cellular populations (**Figure VI.S.6**). A similar observation was previously reported, where the authors demonstrate host-driven transcription or "transcriptional interference" is suppressing viral transcription *via* possible convergence by incoming Pol2 originating at the host promoter [222]. The authors used semi-quantitative RT-qPCR assays to quantify HIV-1 transcriptional read-

through products in these studies [222]. We utilized RNA-seq to determine the precise identify of the virus-host RNA chimeras and evaluate the role transcriptional interference is playing in regulating gene expression at HIV-host gene boundaries.

To capture long reads to identify virus-host chimeric RNAs, we utilized long-read Nanopore sequencing. Illumina-based pair end RNA-seq provided great sequencing depth for HIV-host gene profiling studies (>40 million reads/library), however, the shorter fragments captured via Illumina-based sequencing made it difficult to assemble long chimeric HIV-host RNA transcripts [334, 335] **(Figure VI.S.7)**. Nanopore RNA-seq has the advantage of sequencing very long stretches of RNA, enabling accurate mapping of complex transcripts for improved transcriptomic analyses [333, 379]. In the inactive J-Lat 8.4 cells, we observe host-driven chimeric transcripts to support our smFISH imaging. About 1-5% of the total HIV-1 transcripts captured within the J-Lat 10.6 and 8.4 models were chimeric HIV-host RNAs (**Figure VI.4C**). Our long-read Nanopore sequencing platform enabled characterization of the transcriptional start and end sites of the chimeric transcripts within the 8.4 model. In the active J-Lat 8.4 cells, we saw that transcription originated at the HIV-1 5' LTR and terminated at the 3' UTR of the *FUBP1* gene (**Figure VI.5**). We observe that in both J-Lat models, activation of proviral transcription does not seem to affect local host transcriptional output **(Figure VI.6C).** However, isoform analysis of *SEC16A* and *FUBP1* transcripts in the presence and absence of HIV-1 integration suggest that the mere presence of an integrated provirus may affect expression of full-length host transcripts if HIV-1 and the respective host gene are in the same transcriptional orientation. Premature termination can occur when both the provirus and host gene are in the same orientation, as the host gene can run into the termination site within the 5' or 3' LTRs [74, 380]. This phenomenon is observed in **Figures VI.5-VI.6**, within the J-Lat 8.4 cells. We also observe that at the *FUBP1*

locus in the 8.4 model, HIV-1 activation may alter host gene splicing patterns, leading to differential *FUBP1* expression. Transcription at the HIV-host boundary can lead to intergenic splicing when HIV and the host gene are in the same transcriptional orientation [377]. We observe several transcripts of *FUBP1* splicing into HIV-1 and then prematurely terminating (**Figure VI.5**). While our analyses support the notion of HIV- and host-driven transcriptional read-through at the HIV-host boundary, as previously reported [222, 377], our findings suggest that HIV-1 integration and proviral activation only modestly alter the local RNA landscape.

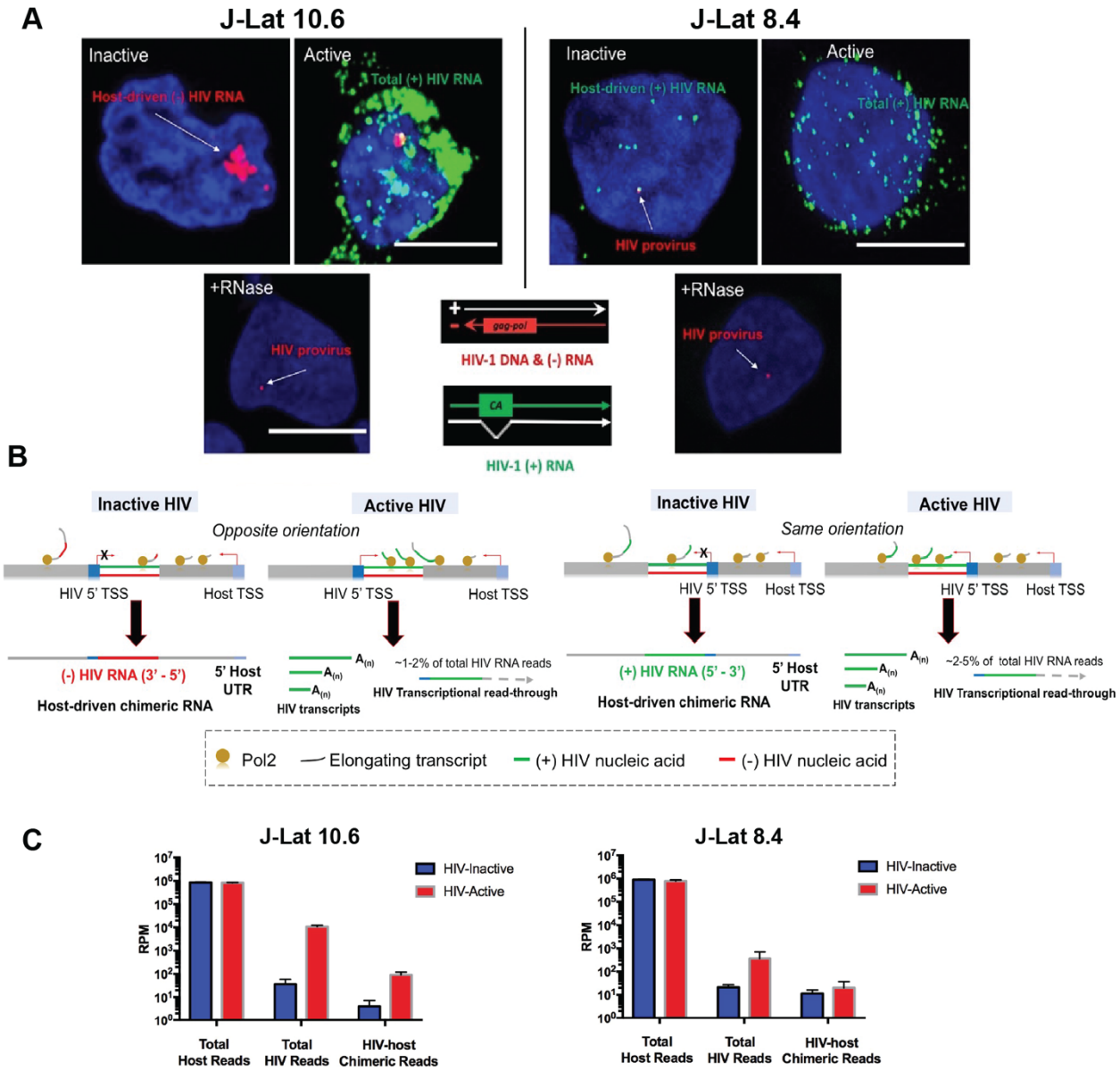


Figure VI.4. Host- & HIV-driven transcriptional read-through at HIV-host genomic boundaries. Visualization of HIV-1 DNA & RNAs of different strandedness *via* single-molecule FISH (smFISH). The oligonucleotide hybridization probe shown in red targets template HIV-1 cDNA (3' ← 5') and antisense (-) HIV-1 transcripts. The probe depicted in green is designed to target positive-sense (+) HIV-1 RNA. Both HIV-1 transcriptional patterns for HIV-inactive & -active cells is shown for J-Lat 10.6 & 8.4 models. RNase-treatment control is included to confirm labeling of viral RNAs, as signal disappears following treatment, staining only the proviral DNA. Scale bars represent 10 µm **(A)**. Schematic of transcriptional state at HIV-host gene boundaries in HIV-inactive & -active states. The diagram color code follows the pseudo-coloring of the smFISH in panel **(A)**. The red nascent transcript in the J-Lat 10.6 inactive cells indicates expression of HIV-1 (-) RNA driven by the flanking host promoter. Similarly, in the inactive 8.4 cells, HIV-1 (+) RNA is expressed with transcription originating at the host promoter and

Pol2 running into the HIV-1 genome. In the active J-Lat cells, (+) RNA is expressed, and both canonical poly(A) transcripts and chimeric HIV-host RNAs are generated **(B)**. Normalized read-counts of total HIV & chimeric HIV-host RNAs using long-read Nanopore sequencing. Normalized read-counts were averaged across four biological replicates/condition (HIV-inactive & -active) for both the J-Lat 10.6 & 8.4 cells **(C)**.

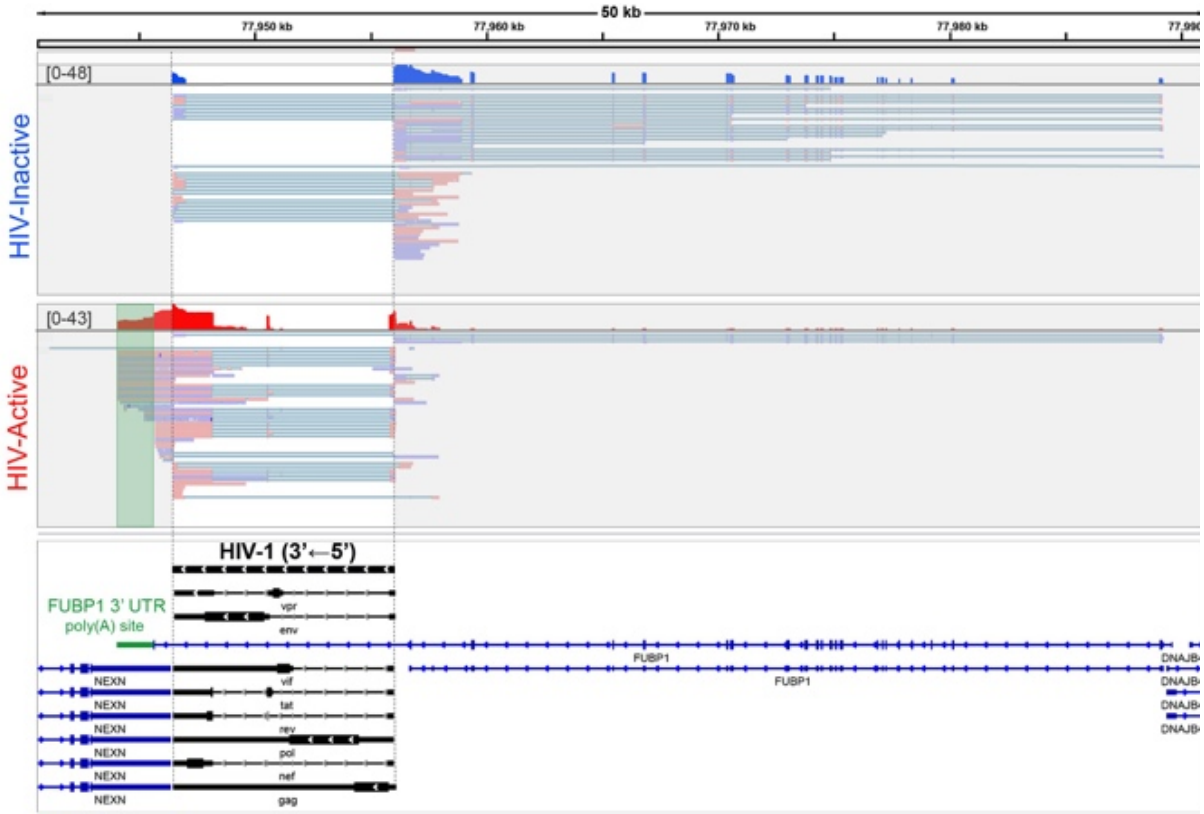


Figure VI.5. Long-read Nanopore RNA-sequencing of J-Lat 8.4 cells enables identification and characterization of long chimeric HIV-host RNA isoforms. Chimeric HIV-host RNA reads were filtered from the total RNA reads and displayed above. Chimeric read pile-up from HIV-inactive (blue) & -active (red) cells are displayed. Under the respective histograms, individual mapped reads are shown. In the HIV-active cell population (red), HIV-driven transcriptional read-through is highlighted by the green box. HIV-driven chimeric reads have a clean transcriptional stopping-pattern at the host *FUBP1* 3' UTR/poly(A) site. HIV-1, cellular, and chimeric transcript structures are shown below the mapped sequencing reads. The HIV-1 and *FUBP1* sequences are in the 3' ← 5' transcriptional orientation.

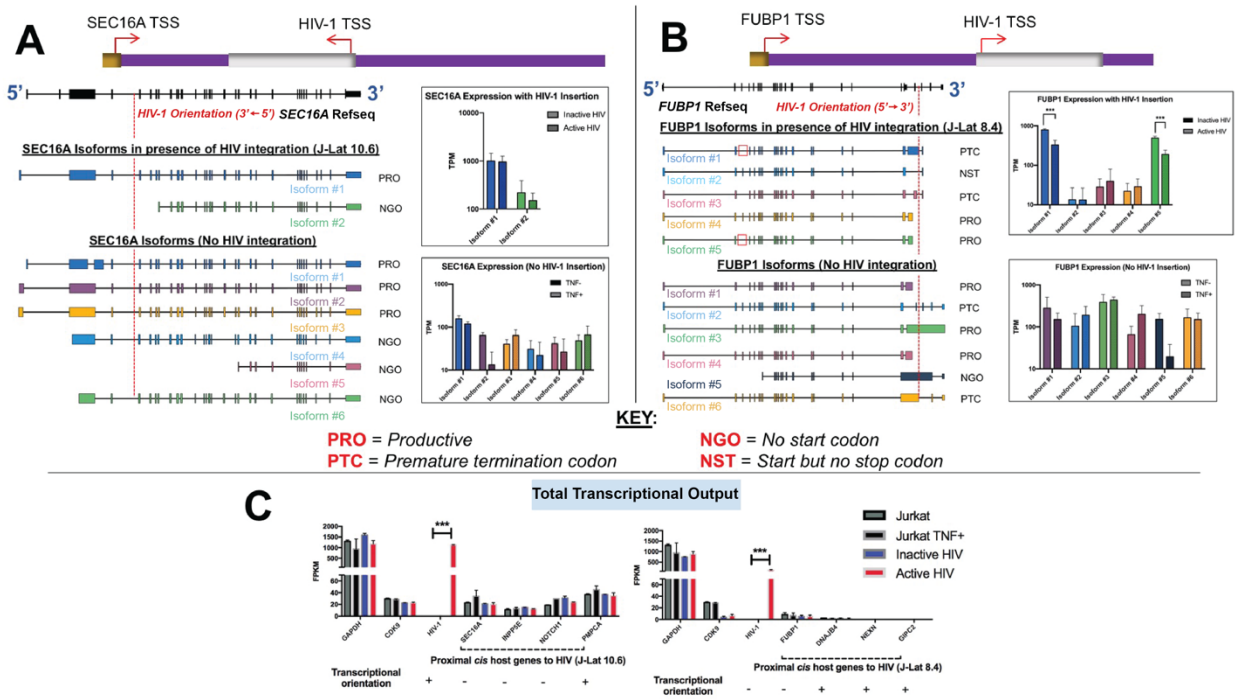


Figure VI.6. RNA splicing analysis of cellular genes in the presence and absence of HIV-1 integration. Isoform expression of the *SEC16A* gene in the presence & absence of HIV-1 integration. To assess how HIV-1 integration may affect expression of *SEC16A*, isoform analysis was performed in the presence/absence of HIV-1 integration. Analysis was done in the J-Lat 10.6 cell (provirus integration in *SEC16A* gene) and the J-Lat 8.4 cell line (HIV-1 integration in *FUBP1*). Similarly, FUBP1 gene expression was assessed in both J-Lat 8.4 & 10.6 cells. The respective Refseq genes (black) depict the exon and intron distribution of both *SEC16A* & *FUBP1*. The red dotted line denotes the sites of integration within the J-Lat models. All *SEC16A* or *FUBP1* transcripts are depicted with a 5' → 3' polarity. We selected isoforms present in both TNF-treated & untreated populations for comparison. Normalized read counts (TPM) for all isoforms were quantified and compared. Isoforms were further functionally characterized as productive (PRO = productive) or potentially translationally defective (NGO = no start codon, PTC = premature termination codon, NST = start but no stop codon). Red boxes in panel **(B)** demarcate exons that were skipped following HIV-1 transcriptional activation in the J-Lat 8.4 model **(A-B)**. Total mRNA read expression for HIV-1 and cellular genes proximal to J-Lat 10.6 & 8.4 integration sites, respectively. mRNA levels of Jurkat T cells (no HIV-1 integration) were also assessed to collectively evaluate the role HIV-1 integration, TNF stimulation, and HIV-1 transcriptional state have on the local cellular RNA landscape. Differential analysis was performed using DESeq2 [401]. *** = p-value ≤ 0.001 **(C).**

DISCUSSION

The HIV-1 genome is integrated into that of its host, and from that point onwards, transcription of viral RNA is carried out by host transcriptional machinery. The nucleus is a heterogeneous environment, and it is increasingly understood that the architecture and organization of the chromatin is an influential factor in the regulation of host gene expression. Thus, chromatin organization is a significant factor when attempting to understand both the transcriptional activity of HIV-1 at different sites in the genome and the potential influence of the HIV-1 provirus on the chromatin and genes in the vicinity of the integration site; however, few studies to date have addressed this relationship. In our work, we have investigated the nuclear environment around two distinct proviral integration sites, characterizing these sites in the presence and absence of the provirus, and addressing the impact of viral transcription on global and local chromatin structure.

Integrative profiling of chromatin organization and RNA transcription at HIV-1 sites of integration allowed us to investigate how HIV-1 can influence the cellular chromatin environment and facilitated identification of signatures associated with transcriptionally active HIV-1 proviruses. Previous studies have demonstrated that the retrovirus, HTLV-1, can alter cellular chromatin structure upon integration *via* induction of differential CTCF-mediated looping, disrupting local transcriptional patterns [337-339]. In the case of HTLV-1, an inherent CTCF binding-site mediates long-range chromatin interactions with the host genome, leading to significant chromatin remodeling. The extent of HIV-mediated chromatin restructuring upon integration is not clear. Extensive studies of HIV-1 gene regulation have identified nuclear factors that can affect proviral transcriptional activity. The HIV-1 provirus can be silenced by repressive epigenetic marks at the Nuc-0 and Nuc-1 nucleosomes flanking the HIV-1 5’ LTR,

causing the proviral promoter to become tightly wound and inaccessible from host transcription factors [212]. Repressive chromatin marks such as H3K9me3, H3K27me3, and CpG methylation are also enriched at latent (transcriptionally inactive) proviruses [181, 213-216]. *Trans*-acting chromatin factors have been implicated in facilitating the recruitment of histone deacetylases (HDACs), polycomb repressive complex-2 (PRC2), and other repressor complexes to the core HIV-1 5' LTR promoter, leading to proviral transcriptional inhibition [217, 218]. In contrast, productive proviruses are associated with open chromatin and active epigenetic markers at the 5' LTR [32, 218, 276, 381]. The relationship between chromatin functional state, HIV-1 integration preference, and HIV-1 transcriptional competency have been established [32, 94, 169, 170, 181], and in this study, we expand our understanding of how integrated HIV-1 can restructure the local nuclear environment.

Through Hi-C analysis at the HIV-1 integration sites in our models, we observe that integrated HIV-1 does not induce large scale genome reorganization. This finding is contrary to what was reported for HTLV-1, demonstrating differences in replication strategies across these two retroviruses. At the HIV-1 integration sites sampled, we did not find significant differences in chromatin interaction profiles and chromatin looping patterns relative to uninfected WT Jurkat T cells. Chromatin looping patterns, nuclear compartmentalization, and inter- & intra-chromosomal interactions were all predominantly preserved in the vicinity of integrated HIV-1 relative to the native chromatin environment. These findings suggested that major chromatin restructuring is not a pre-requisite of HIV-1 integration, and perhaps, differential chromatin interactions with distal cellular regulatory elements are not required for activation of HIV-1 transcriptional activity. Interestingly, we did however, find that transcriptional activation of HIV-1 induced chromatin opening downstream the HIV-1 3' LTR. Since this effect in chromatin

accessibility is not observed in the absence of HIV-1 integration (WT Jurkat T cells +/- TNF-**α**), our results suggested that this subtle change in chromatin organization is an HIV-specific effect and demonstrates that HIV-1 transcriptional activation can induce chromatin opening downstream the 3' LTR. This observation is novel and contrary to our conventional understanding of chromatin accessibility and transcriptional competency, as activated genes typically exhibit increased chromatin opening towards the 5' gene promoter, which then facilitates the recruitment of stable transcription complexes [382-385].

Previous studies may have missed this HIV-specific effect, as proviral chromatin accessibility was assessed following bulk *de novo* infection [386, 387]. Within these studies, the average ATAC signal of all of the HIV-1 proviruses in the infected cellular population is obtained, prohibiting the deconvolution of the accessibility profiles of individual HIV-1 proviruses. In addition, our studies further characterized the identity of chimeric HIV-host RNAs and the transcriptional landscape in the vicinity of integrated HIV-1. Previous studies used RT-qPCR [222] or lower resolution single-cell sequencing approaches to investigate HIV-induced transcriptional perturbation [377]. Utilization of high-throughput long-read RNA-seq methodologies allowed us to demonstrate that HIV-1 integration also appears to have minimal effects on the local transcriptional landscape; however, retroviral integration can promote differential isoform expression at HIV-host intergenic boundaries. While we may not be seeing significant perturbations to local cellular transcriptional output in the presence of integrated HIV-1 in HIV-inactive and -active states **(Figure VI.6C**), it is conceivable that HIV-1 can contribute to defective cellular RNA transcription depending on the site of integration. In the case of the J-Lat 8.4 model, HIV-1 integration enhanced expression of truncated *FUBP1* mRNA (**Figure VI.6B**). Alterations in local cellular transcriptional patterns attributed to HIV-1

integration can subsequently lead to aberrant host protein expression [377]. In addition, when HIV-1 and the respective host gene are in the same transcriptional orientation, as in the J-Lat 8.4 model, the host gene can splice into the HIV-1 genome, generating potentially defective chimeric RNAs (**Figure VI.5**). Thus, our analysis demonstrates another possible avenue for HIV-induced transcriptional dysregulation.

The changes in the chromatin structure at proviral sites of integration following HIV-activation does not appear to be attributed to direct TNF-signaling pathways. In **Figure VI.S.8**, we show that J-Lat cells that are treated with TNF-**α**, but do not activate, still undergo canonical TNF-signaling with an increase in expression of the NFκB subunits. Despite undergoing TNF-signaling, the fact that these cells do not activate suggests that other indirect signaling pathways are also involved in HIV-1 transcriptional activation, and NFκB transcription factor occupancy is likely not the limiting factor for induction of proviral transcription in these cellular models. Expression of chimeric HIV-host RNAs has a stark resemblance to stress-induced transcriptional read-through [388-390] and chromatin remodeling following lytic Herpes Simplex Virus 1 (HSV-1) infection [391, 392].

HIV-1 transcription is highly efficient following assembly of the Tat/TAR transactivation axis [218], a *trans*-acting regulatory mechanism critical for proviral transcriptional elongation. Activation of processive viral transcription enables HIV-1 to dominate the local transcriptional environment. Similarly, HSV-1, a double-stranded DNA virus, hijacks the transcriptional machinery of the host cell during lytic infection. Despite not integrating into the host chromatin, HSV-1 transcriptional activation leads to significant dysregulation of host transcription by inducing genome-wide transcriptional read-through and disruption to transcriptional termination within the 3' UTR of cellular genes. These transcriptional read-through products are not

translated, evident by their absence within polysome fractions [391]. Further studies have demonstrated that transcriptional read-through of host genes following lytic HSV-1 infection also coincides with an increase in chromatin accessibility downstream the 3' UTR of the disrupted host gene. These changes in chromatin accessibility and transcriptional patterns during lytic HSV-1 infection are analogous to what we have demonstrated following HIV-1 transcriptional activation. While cellular transcription may occasionally fail to terminate, possibly due to a "slippery" polymerase, basal levels of transcriptional read-through are very low in cells during periods of cellular homeostasis [388-390]. In **Figure VI.S.10**, we show the read pile-up from RNA-seq at TNF-responsive genes and genes that are highly expressed in our J-Lat models. In these representative sequencing tracks, RNA density or transcriptional read-through is not detectable or at very low incidences. We performed similar analysis looking at read density beyond the 3' UTR of the highest expressed genes from our RNA-seq analysis and did not find significant read-through at these respective genes, suggesting that the read-through present at HIV-1 integration sites and read-through induced by lytic HSV-1 infection is a specific transcriptional signature of these respective viruses.

The Steitz group coined a novel class of transcripts, "Downstream-of-genes" or DoGs, which resulted from transcription failing to terminate at the canonical transcriptional termination site and running extensively beyond the open-reading frame (ORF) of the respective gene [388-390]. DoG production is significantly up-regulated following cellular stress with increased Pol2 occupancy downstream the disrupted genes [393]. Recent studies have demonstrated an association between cellular stress responses and latency reversal (HIV transcriptional activation) [263, 394], with stress-inducing conditions increasing HIV-1 reactivation frequency. Further studies should investigate this potential interaction between HIV-1 transcription and cellular stressors.

Commonalities of transcriptional read-through observed in two different viruses: 1) HIV-1 and 2) HSV-1, as well as under conditions of cellular stress, warrants further investigation into the interplay between cellular stress pathways and chromatin reorganization in mediating viral transcription. Our studies reveal a common chromatin signature associated with HIV-1 transcriptional activation across distinct models, taking into account differences in proviral transcriptional orientation and integration site selection. Thus, this work provides detailed analysis of how HIV-1 integration affects chromatin organization and the local cellular RNA landscape, providing a platform for multiomics NGS-based approaches to study replication across a broad spectrum of viruses.

MATERIALS/METHODS

Cell culture of T cell lines

J-Lat full-length 10.6 cells (NIH AIDS Reagent Program catalog #9849) and J-Lat 8.4 full-length cells (NIH AIDS Reagent Program catalog #9847) are Jurkat-derived human lymphocytic T cells that are latently infected with the packaged retorivral construct HIV-R7/E-/GFP, a full-length HIV-1 minus *env* and *nef*[185]. Cells were cultured in RPMI 1640 medium (Gibco, Waltham, MA, USA), supplemented with 10% heat-inactivated fetal bovine serum (FBS) and 2 mM L-glutamine (Gibco), in a humidified incubator at 37 °C with 5% CO_2. Jurkat T cells were cultured at the same conditions listed above.

Antibodies and Latency-reversing agents

A mouse monoclonal anti-p24 antibody raised against the p24/capsid domain of Gag [395], and a goat anti-mouse secondary antibody conjugated to Alexa 488 (Invitrogen) were used to detect HIV Gag after cell fixation for imaging studies. Compounds used for activation in HIV-inducible cell lines included: (1) 10 ng/µL TNF-**α** (Genscript, Piscataway, NJ, USA) and (2) 81 nM phorbol 12-myristate 13-acetate with 1.34 µM ionomycin (PMA/I) (eBioscience, San Diego, CA, USA). Cells were treated for 24 hours prior to collection for various assays. Treatment conditions were based on parameters previously described [319].

Assay for transposase-accessible chromatin using sequencing (ATAC-seq)

ATAC-seq was performed using the OMNI-ATAC protocol [329]. After counting 50,000 cells/library, nuclei were isolated with Lysis buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$) containing 0.1% NP40, 0.1% Tween-20, and 0.01% digitonin. The purified Jurkat or J-Lat nuclei were then resuspended in a transposase reaction containing 0.05% digitonin and incubated for 30 min at 37 °C. Following incubation, samples were treated with Proteinase K at 55 °C for 2 hr, and genomic DNA was isolated *via* phenol:chloroform:isoamyl alcohol and EtOH precipitation. Library amplification was done with 2x KAPA HiFi mix (Kapa Biosystems) and 1.25 µM indexed Illumina primers using the following PCR conditions: 72 °C for 5 min; 98 °C for 30 s; and 10-11 cycles at 98 °C for 10s, 63 °C for 30 s, and 72 °C for 1 min. Libraries were generated with two biological replicates per condition sampled.

ATAC-seq data analysis

All ATAC-seq libraries were sequenced using an Illumina Hiseq2500 v4 sequencer and 50 bp paired-end format. Paired reads were aligned to the human reference genome hg38 or custom HIV-host genome using Bowtie2 [396]. Further information on generation of our custom J-Lat reference genomes and complementary annotation files (GTF) are discussed below for our Nanopore sequencing pipeline. Reads were aligned using default parameters except -X 2000 -m 1. PCR duplicates were removed using Picard Tools. To adjust for fragment size, we aligned all reads as (+) strands offset by +4 bp and (-) strands offset by -5 bp [328]. The reads corresponding to Tn5 hypersensitive sites and mononuclesomes were separated by filtering for fragments 50-115 bp and 180-247 bp in length, respectively [296]. MACS2 was used for peak calling for Tn5 hypersensitive sites [397]. Quality-control (QC) data was visualized *via* ggplot2 in the R programming language and through ngs.plot [398]. Differential peaks between HIV-inactive and -active populations was performed using MAnorm [399].

Illumina-based RNA-seq library preparation

Total RNA was extracted from one million cells/condition using the RNeasy Plus mini kit (Qiagen; 74134). Preparation included an on-column DNase digestion. Strand-specific cDNA libraries were generated using a TrueSeq Stranded Total RNA sample preparation kit (Illumina). Sequencing was performed on an Illumina Novaseq 6000 to obtain ~150 bp paired-end reads. Each library, performed in replicate, had ≥ 40 million read pairs per sample.

Illumina-based RNA-seq analysis

Reads were mapped to the human reference genome (hg38) or a custom concatenated HIV-host reference genome for the J-Lat 10.6 and 8.4 models. The J-Lat specific sequence that was used for the J-Lat custom reference genomes was generated *de novo* from activated J-Lat 10.6 cells using our Nanopore pipeline discussed below. Reads are mapped using HISAT2 [400]. HTseq was used to quantify RNAs based on annotations [401]. StringTie was used for transcript assembly [402] and for quantification of different isoforms. Differential expression analysis was performed using DESeq2 with read counts normalized to fragments per kilobase of transcript per million mapped reads (FPKM) [401].

Nanopore RNA-sequencing library preparation & sequencing/base-calling

Total RNA was isolated from cell pellets (~50,000 cells) using the RNeasy Mini Kit (QIAGEN, 74134), according to the manufacturer's instruction, with elution in nuclease-free water. Reverse Transcription is carried out with SuperScript IV Reverse Transcriptase (ThermoFisher, 18090010) in 20 µL volume with the following components and final concentrations: 1X Reaction Buffer, 0.5 mM dNTPs, 2U RNAse OUT, 1 µM Oligo-d(T) primer, 5 mM DTT, and 200 U SuperScript IV Reverse Transcriptase. Primer is annealed to template RNA in the presence of dNTPs by heating to 65˚C for 5 mins, followed by snap cooling to 4˚C for at least 2 mins. Rest of the components are added after snap cool step, followed by incubation at 50˚C for 1.5 hours, and heat inactivation at 85˚C for 5 mins. Second-strand synthesis is carried out in a single pot format using modified Gubler and Hoffman procedure from Invitrogen's A48570 kit by direct addition of second strand buffer, dNTPs, E.coli DNA Polymerase I, RNAse H, and E. Coli DNA Ligase to heat-inactivated first strand reaction and incubation at 16˚C for 2 hours.

Second strand products are DNA cleaned with Monarch PCR and DNA Cleanup Kit (NEB, T1030S), and eluted in 0.1X TE. Quality control and yield of double-stranded cDNA is determined with NanoDrop spectrometer.

Double-stranded cDNA samples were barcoded with the Native Barcoding kit (Oxford Nanopore Technologies, EXP-NBD104, EXP-NBD114), and library prepped using the Ligation Sequencing Kit (ONT, SQK-LSK109). Samples were sequenced with MinION Mk1B using a FLO-MIN106D flow cell (R.9.4.1). Reads were basecalled with Guppy Basecaller version 4.2.3+8aca2af with high accuracy mode.

Reference alignment of Nanopore-seq datasets

For human reference alignment, the hg38 UCSC analysis set of December 2013 human genome (GCA_000001405.15) without alt-scaffolds was used along with its associated annotation file in GTF format. For HIV alignments a custom R7 strain reference sequence was generated de-novo from reads from activated J-Lat 10.6 cells, an accompanying custom annotation file in GTF format was also generated containing all canonical splice variants. For alignment of host/viral chimeric reads at integration site, the R7 reference sequence and its associated annotation was inserted into the hg38 reference and annotation file using the Reform Tool web portal (https://reform.bio.nyu.edu/) at the following hg38 positions and R7 orientations according to validated integration sites (PMC5237276): chr1 77,946,384 (R7 antisense orientation) for J-Lat 8.4, and chr9 136,468,579 (R7 sense in sense orientation) for J-Lat 10.6.

Extraction of HIV-host chimeric reads & analysis of HIV-host read-through products

Reads were mapped to R7 HIV reference using *minimap2* with *map-ont* preset and *--secondary=no* option. R7-mapped reads were extracted from sam output using *samtools view* with following options *-h -F4* , and converted to FASTQ using *samtools bam2fq*. R7-mapped reads were aligned to unmodified UCSC hg38 reference using *minimap2* with *splice* preset and *--secondary=no* option. Hg38-mapped reads were extracted and converted into FASTQ as before. The resulting FASTQ file contains HIV/host chimeric reads that map to both R7 HIV and hg38 references. For readthrough analysis, HIV/host chimeric reads were mapped to the custom hg38 reference containing R7 HIV sequence using *minimap2* with *splice* preset and *--secondary=no* option. Mapped reads are then extracted using *samtools view* with *-F4 -h* options, and sorted using *samtools sort*. Sorted bam outputs are then indexed with *samtools sort* and visualized using Integrative Genomics Viewer version 2.8.9.

Quantification and productivity analysis of host cell isoforms at integration site

For this workflow the FLAIR isoform analysis pipeline was used (PMC7080807). Reads were mapped to the UCSC hg38 reference using FLAIR *align* module using option *-p*, followed by splice junction correction using the *correct* module with option *-c* and using hg38 gtf annotation file. Resulting bed files from FLAIR *correct* step are parsed to extract isoforms that overlap with FUBP1 and SEC16A genes, the respective integration sites J-Lat 8.4 and 10.6. For this purpose, a bed file was generated containing the following boundaries for FUBP1 and SEC16A with care to avoid overlap with any neighboring genes: chr1: 77,949,155 - 77,979,130 (FUBP1), and chr9 136,440,096 - 136,482,938 (SEC16A). The *bedtools intersect* command was used with options -

wa and *-a* to extract all isoform that intersect with the query regions, with resulting intersect bed file concatenated according to treatment. The FLAIR *collapse* module is then used with default settings to create a transcript model for FUBP1 and SEC16A isoforms for each treatment. These transcript models are plotted with plot_isoform_usage.py script to obtain isoform structures and cross-referenced with hg38 gtf annotation files to predict productivity of each isoform using the predictProductivity.py script. Following isoform collapse, the FLAIR *quantify* module is used with option *--tpm* to obtain transcript per million (TPM) values for each isoform present in each replicate and treatment.

Analysis of publicly available datasets (ChIP-seq, TSA-seq, & SEduper)

For ChIP-seq analysis of publicly mined data, reads were mapped using Bowtie2 [396] to hg38. PCR duplicates were removed using Picard Tools. MACS2 was used to call peaks using default parameters with IgG ChIP-seq data as input control. ChIP-seq datasets from Jurkat T cells include: H3K27ac (GSM2691418), Pol2 (GSM1850204), and BRD4 (GSM2218755). TSA-seq dataset of SPADs (GSM3111194) in K562 cells were processed as stated in Chen *et al.* [371]. Predicted SE networks were obtained from dbSuper [403].

Flow cytometry and FACS of J-Lat cell lines/flow cytometric analysis

The BC Cytoflex cytometer was used for data collection. 10^5 Jurkat or J-Lat cells were seeded per well in a 96-well plate format. Cells were treated with TNF-**α** or PMA/I and the number of GFP+ cells following treatment were quantified. Gates for GFP+ cells were set by assessing the

fluorescence profiles of uninfected Jurkat T cells. FlowJo software was used for data analysis. Gating strategy is depicted in **Figure III.S.1**.

Fluorescence-activated cell sorting of inactive (GFP-) and active (GFP+) populations of cells was performed using the BD FACS Aria II SORP. Prior to cell sorting, T cells were resuspended in PBS and passed through a 40 µm cell strainer (Corning, 431570). Final cell concentrations were ~ 5 x 10^6 cells/mL prior to sorting.

Single-molecule DNA & RNA FISH

Branched DNA in situ hybridization (bDNA FISH) was used for detection of HIV-1 RNA in J-Lat cells using the RNAscope method, with modifications (1). One million cells per condition were harvested and spun down at 1.5 rpm for 5 minutes. These cells were resuspended in 50 µL of phosphate buffered saline (PBS) and seeded on poly-d-lysine (Gibco, cat: A3890401) coated coverslips for 30 min. Cells were then fixed in 4% paraformaldehyde for 30 min at room temperature (RT) and washed three times in PBS. Samples were then incubated in PBS supplemented with 0.1% Tween-20 (PBS-T) for 10 min at RT and washed twice in PBS. The manufacturer's protease solution (Pretreat 3; Advanced Cell Diagnostics (ACD)) was diluted 1:5 in PBS and applied to samples. Samples were then incubated in a humidified HybEZ oven (ACD) at 40 °C for 15 min. Protease solution was decanted, and samples were washed twice in PBS. A probe that recognizes HIV-1 (+) RNA (HIV-nongagpol-C3; ACD; 317711-C) and HIV-gagpol-C1 (317701) diluted in hybridization buffer was applied to the sample. Incubation with the probe was performed at 40 °C for 2 h in the HybEZ oven. The remaining wash steps and hybridization of preamplifiers, amplifiers, and fluorescent label were performed as performed

previously [317, 318]. HIV-1 cDNA was fluorescently labeled with ATTO 550, and vRNA was labeled with Alexa 647. Nuclei were counter-stained with 4',6'-diamino-2-phenylinndole (DAPI, ACD) for 1 min at RT and washed twice in PBS. Coverslips were mounted on slides using Prolong Gold Antifade (Invitrogen).

Imaging was performed *via* confocal microscopy using a 60x oil-immersion objective. The excitation/emission bandpass wavelengths used to detect DAPI, ATTO 550, and Alexa 647 were set to 405/420-480, 550/560-610, and 647/655-705 nm, respectively.

In-situ Hi-C library preparation & analysis

Hi-C was performed using the *in situ* method with the MboI enzyme as previously described [292]. Hi-C libraries were generated from either wild-type (WT) Jurkat T-cell lines, untreated clonal J-Lat model cell lines, or sorted GFP+ J-Lat cells after stimulation with TNF-**α**. The WT and untreated J-Lat cell lines were fixed in solution using 1% formaldehyde. For sorted samples, GFP+ cells were sorted into PBS and subsequently fixed with 1% formaldehyde. Hi-C libraries were sequenced on Novaseq 6000 S4 flow cells. Reads were aligned to the hg38 genome, filtered, and deduplicated as previously described [372]. After alignment, contact files were generated and processed into hic or cool files using the Juicer or cooler pipelines, respectively. Loop calling was performed using hiccups [292].

ACKNOWLEDGEMENTS

This study was supported in part by the Emory Flow Cytometry Core (EFCC), one of the Emory Integrated Core Facilities (EICF), subsidized by the Emory University School of Medicine. This work was supported in whole or in part by the National Institutes of Health (R01 AI121315, R01 AI146017, R01 AI148382, R01 AI120860, R37 AI076119, and U54 AI150472).

SUPPLEMENTAL DATA

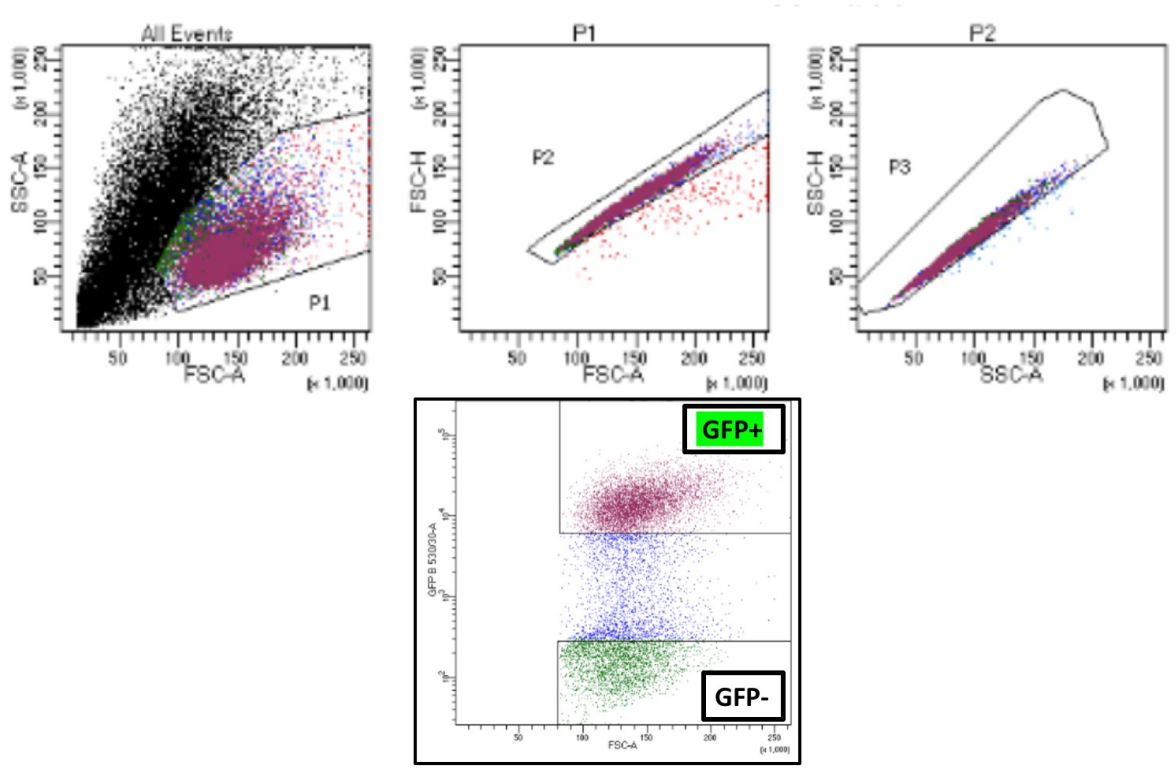


Figure VI.S.1. Gating scheme for sorting reactivated (GFP+) J-Lat cells *via* FACS and for flow cytometric analysis. J-Lat cells are first gated based on cellular granularity, which is followed by gating for singlets. Selection for GFP+ cells is stringent, selecting for the higher-expressing GFP+ cells.

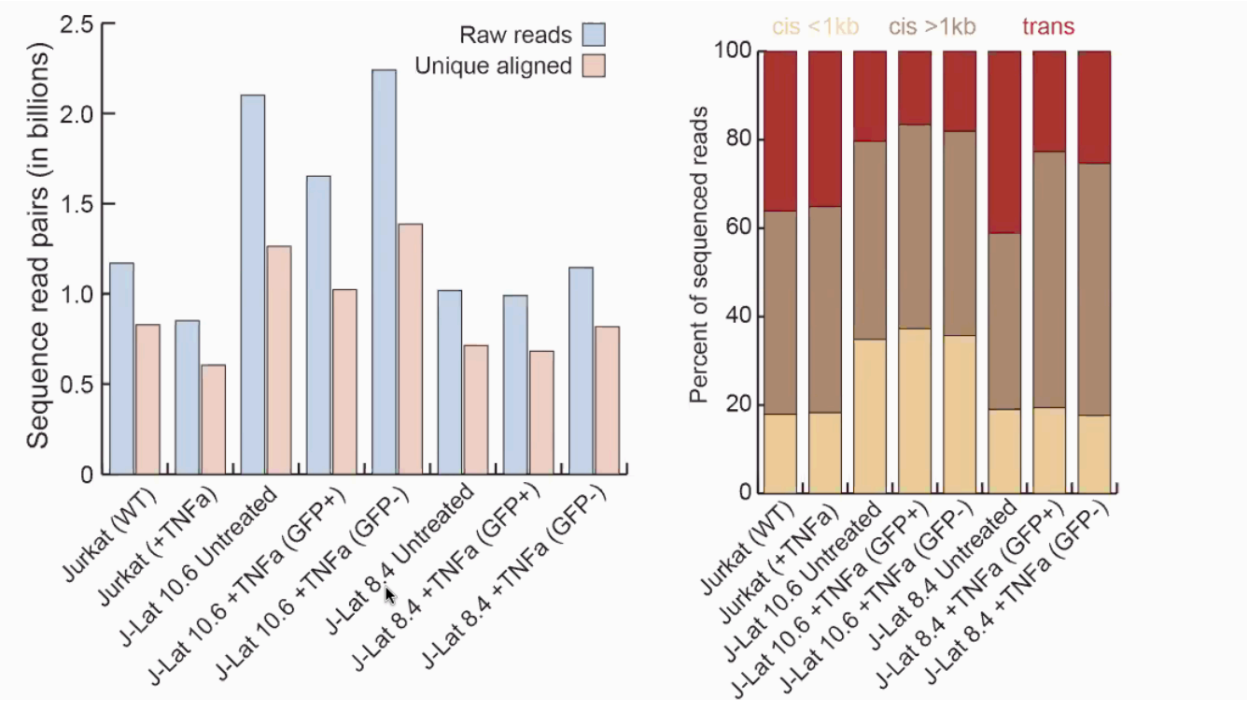
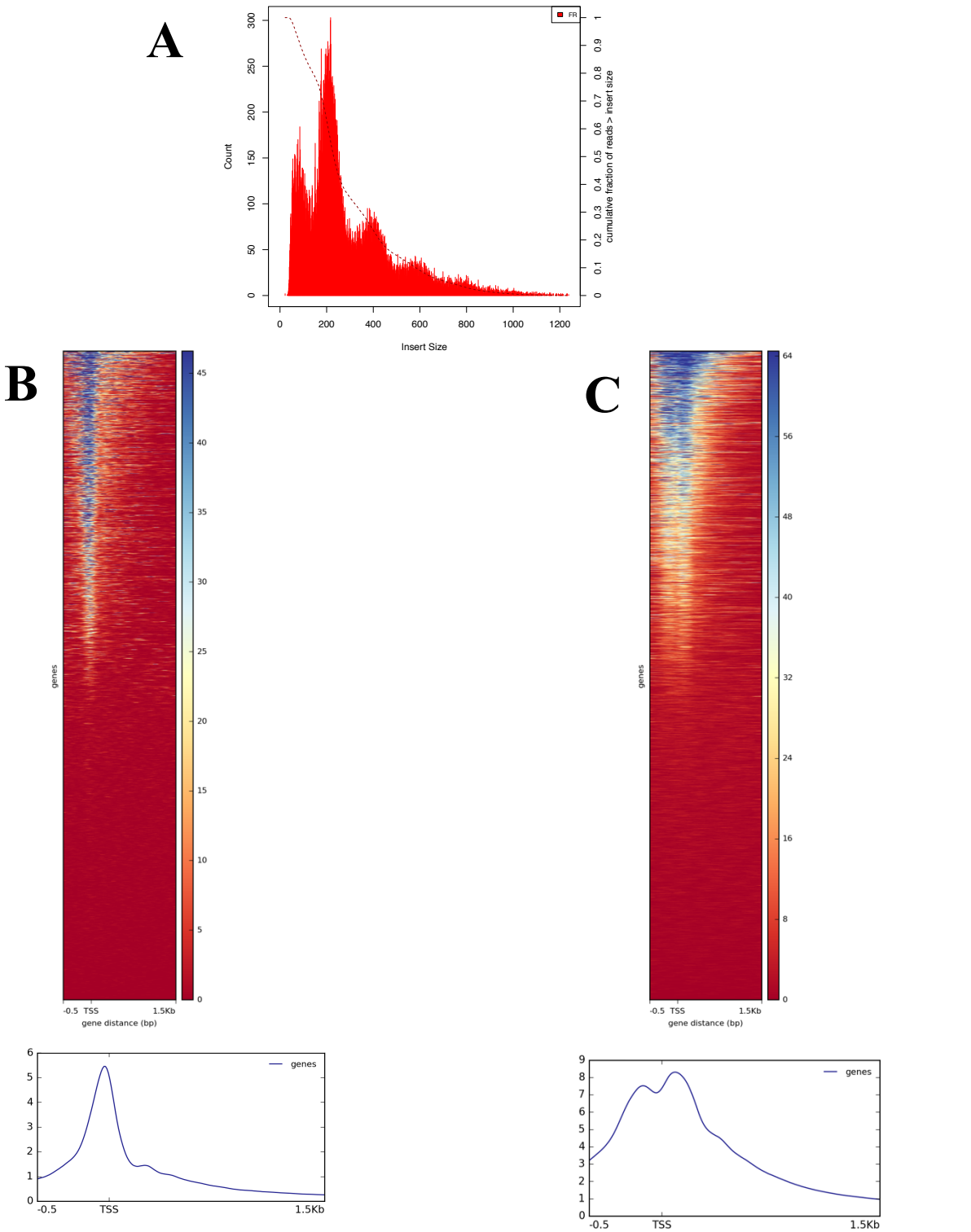


Figure VI.S.2. Hi-C reads per library. Total reads/unique aligned reads per library and percent *cis* & *trans* chromatin-chromatin interactions. These are quality control metrics of Hi-C libraries.



Reads filtered for putative TSS

Reads filtered for putative mononuclesomes

Figure VI.S.3. ATAC-seq QC. Fragment distribution of ATAC-seq reads **(A)**. Heatmap and histogram of reads filtered (1-115 bp) for putative transcriptional start sites (TSS) or promoters **(B)** and reads filtered for mononucleosomes **(C)**. The majority of reads fall within or proximal to annotated TSS. TSS = transcriptional start site.

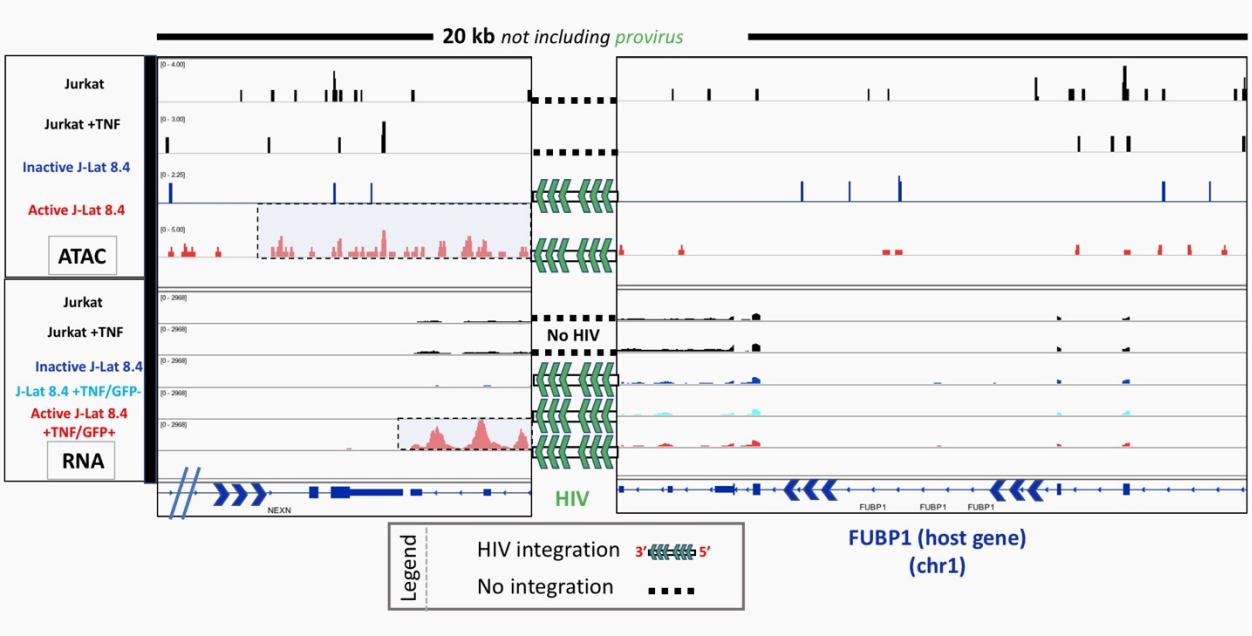
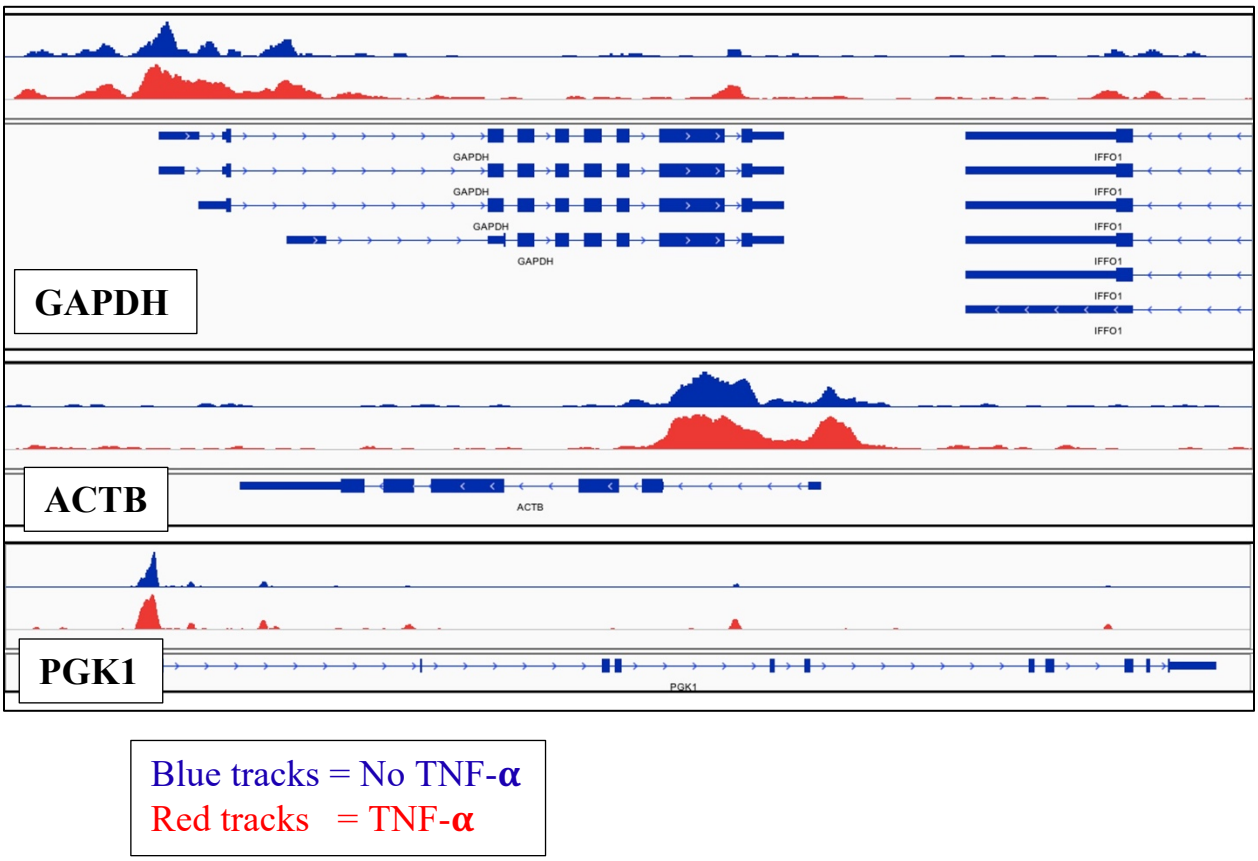


Figure VI.S.4. ATAC- & RNA-seq profiles of inactive & active J-Lat 8.4 cells at site of integration. Jurkat T cell +/- TNF-**α** (black tracks) demonstrate native chromatin and expression profiles of host *FUBP1 gene* and genomic regions directly flanking the provirus. The tracks in blue are the inactive J-Lats, and the tracks in red are the activated J-Lats. The track in cyan are J-Lat cells treated with TNF-**α** but did not activate. The dotted boxes shaded in light blue demarcate regions of the cellular genome where we find differential chromatin accessibility and gene expression. Similar to the J-Lat 10.6, this increase in ATAC & RNA signal is directly downstream the HIV-1 3' LTR. In the J-Lat 8.4 model, the provirus and the host gene (*FUBP*) both have a 3' ← 5' transcriptional orientation. RNA-seq was performed using Illumina-based approaches.



Blue tracks = No TNF-**α**
Red tracks = TNF-**α**

Figure VI.S.5. Chromatin accessibility (ATAC-seq) profiles of constitutively active house-keeping genes +/- TNF-α. There is no significant difference in chromatin accessibility profiles in these genes +/- TNF-**α**, and chromatin is most accessible towards the 5' ends of these genes.

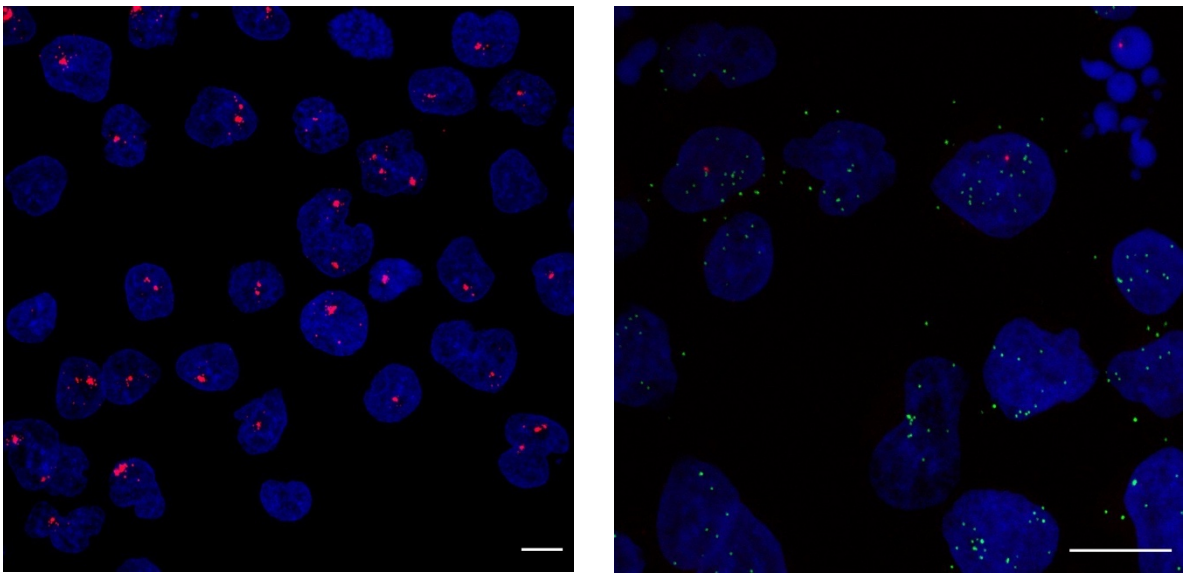


Figure VI.S.6. Maximal intensity projections of inactive J-Lat 10.6 & 8.4 cells. The red staining in the inactive J-Lat 10.6 cells depicts host-driven HIV-1 (-) expression. The green signal in the inactive 8.4 cells demonstrates host-driven HIV-1 (+) expression. The red dots in the 8.4 cells are the provirus. The HIV-1 (-) probe targets antisense HIV-1 transcripts, as well as the template DNA strand (3' ← 5'). Scale bars represent 10 µm. Refer to **Figure III.4** for further discussion of J-Lat single-molecule FISH imaging.

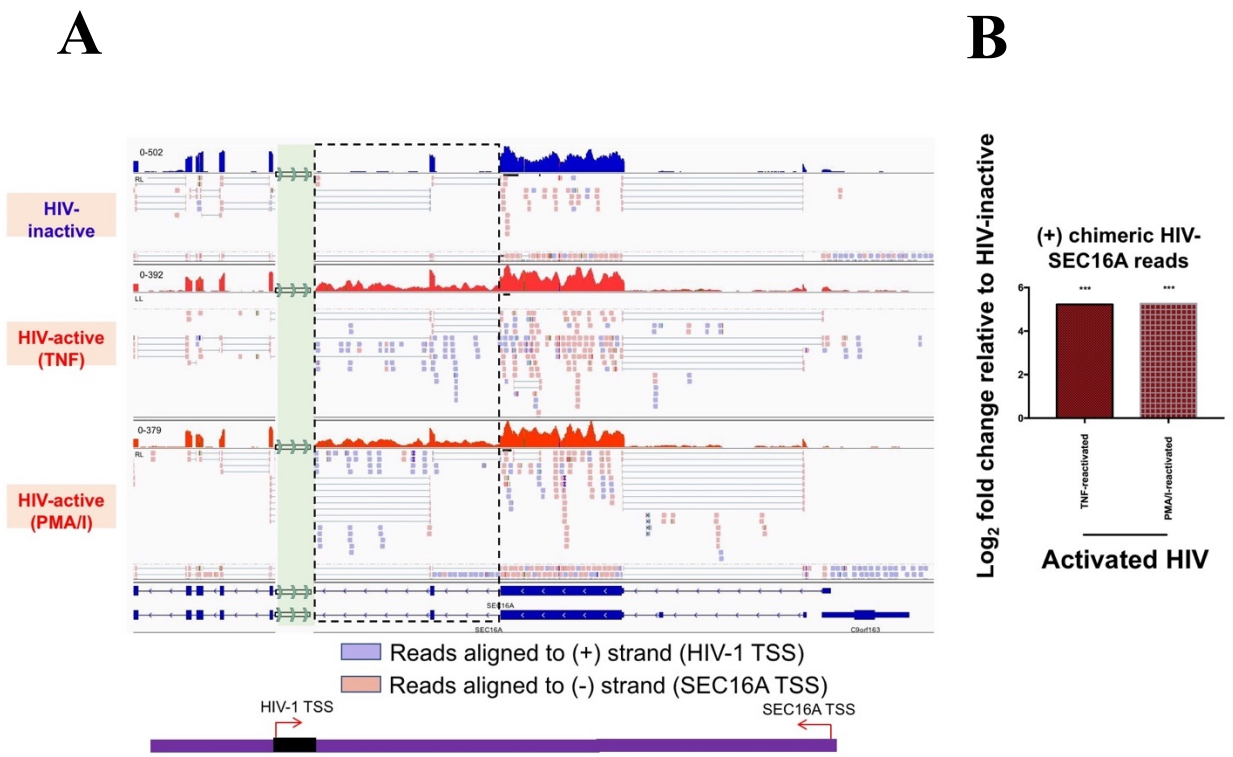


Figure VI.S.7. HIV-host chimeric RNAs induced by HIV-driven transcriptional read-through in the J-Lat 10.6 model. Illumina-based RNA-seq provides high sequencing depth, but the short fragment reads complicate precise transcriptome assemblies of complex RNA species. All reads shown on this genome browser are aligned to the human reference genome (hg38). Tracks to the left are upstream the 5' LTR, and the tracks to the right are downstream HIV-1 3' LTR. The black dotted box demarcates the genomic region directly downstream the 3' LTR where we observe HIV-driven transcriptional read-through into the intron of the *SEC16A* gene. Under the read-pileup, we display aligned reads to hg38. Reads in blue are in the positive-sense (+) orientation and reads in red are in the (-) orientation. HIV-1 has integrated in the 5' → 3' orientation, whereas the host gene (*SEC16A*) has the opposite 3' ← 5' orientation. Therefore, (+) mapped reads (blue) are likely HIV-driven in the area demarcated by the black dotted box. The green highlighted box represents the proviral site of integration in this model. There is an increased number of (+) mapped reads following HIV-activation *via* TNF-**α** or PMA/I **(A)**. Quantification of chimeric HIV-host reads induced by HIV-driven transcriptional read-through. Reads that map to both HIV and the host genome was selected for, and we further filtered for chimeric RNAs that had a (+) orientation at the HIV-host junction. Differential expression of these chimeric RNAs following HIV-activation. *** $p \leq 0.005$ **(B).**

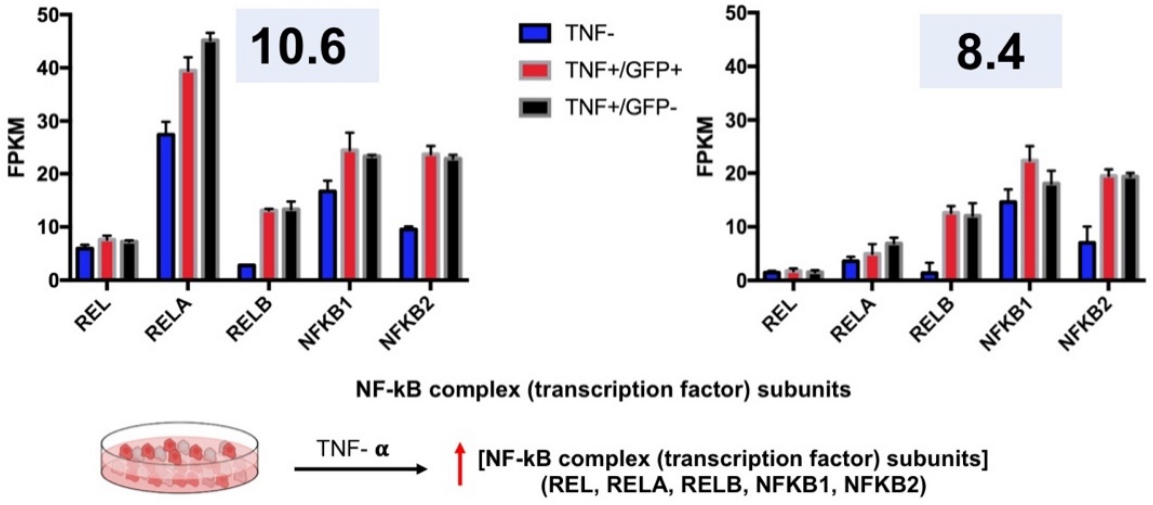


Figure VI.S.8. Expression of NFκB subunits in J-Lat 10.6 & 8.4 models +/- TNF-α & GFP +/-. These results demonstrate that J-Lat cells that do not reactivate are still TNF-responsive, suggesting additional factors are influencing HIV-1 proviral transcriptional activation.

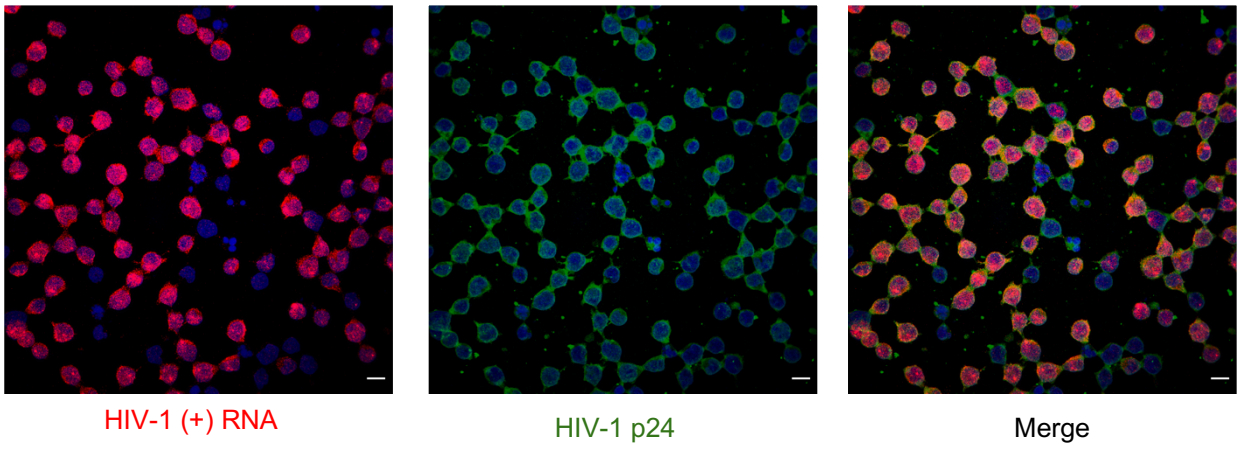


Figure VI.S.9. Activation of J-Lat 10.6 cells treated with TNF-α. Cells were fixed 24 hours post-treatment. Cells were stained for HIV (+) RNA (red) and HIV p24 (green). Images are presented as maximal intensity projections from z-stacks acquired *via* confocal imaging. Scale bars represent 10 µm.

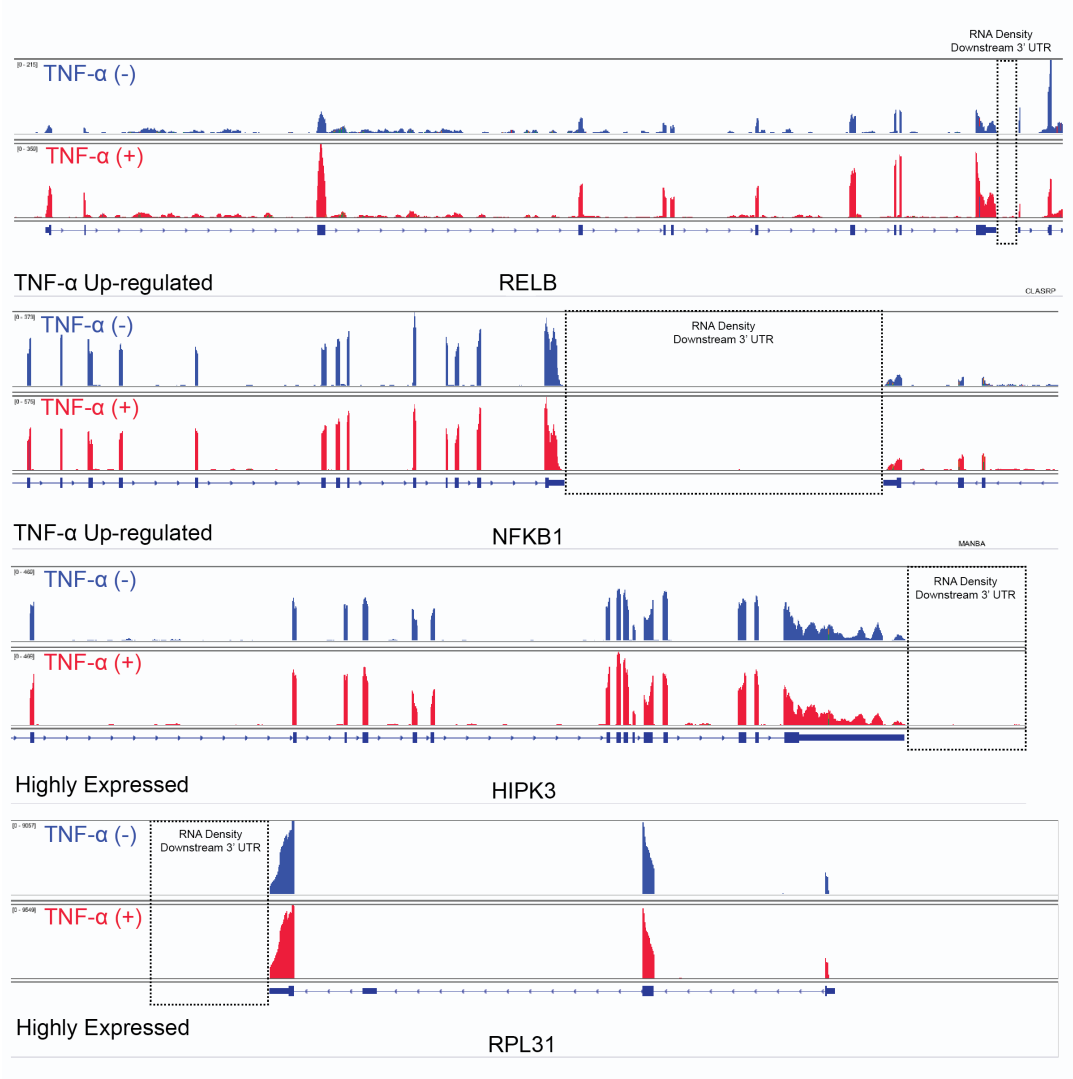


Figure VI.S.10. Lack of RNA density or transcriptional read-through downstream representative TNF-responsive and highly expressed genes in J-Lat cell models. The dotted black box demarcates the region directly downstream the 3' UTR of TNF-responsive genes (RELB and NFKB1) and highly expressed genes (HIPK3 and RPL31) in J-Lat cells +/- TNF-treatment.

CHAPTER VII: DISCUSSION & FUTURE DIRECTIONS

VII-A: Probing Long-Range Chromatin Interactions between HIV-1 & Cellular DNA

In **Chapter VI,** we demonstrated how we can apply the 3C-based method, Hi-C, to map global HIV-host chromatin interactions, as well as host chromatin-chromatin interactions to understand how nuclear ultrastructure around HIV-1 proviral sites of integration are reorganizing in response to integration and proviral transcriptional activation. We were able to interrogate if HIV-1 activation induces changes in chromatin architecture around the integration site by evaluating proviral chromatin functional compartmentalization (Compartment A or B), changes in chromatin looping patterns (CTCF-mediated) proximal to HIV, and HIV-host chromatin interaction networks to ~ 1kb resolution. While these studies successfully allowed us to probe for changes in 3D chromatin organization around HIV sites of integration, this methodology relied on using clonally expanded HIV cell models with defined sites of integration to allow us to achieve the spatial resolution to capture HIV-host chromatin interactions. Chromatin contact frequencies quantified *via* Hi-C are representative of the overall aggregated cellular population, thus limiting us from sampling changes in HIV chromatin structure across several thousands of unique integration sites following bulk *de novo* infection. With the strong assumption that every integration site across a population of thousands of infected cells will be unique, any signal corresponding to an individual integration site would be very low *via* bulk Hi-C. In addition, to determine chromatin structural changes at specific sites of integration, this methodology would have to be multiplexed with integration site identification. These are thus significant bottlenecks for studying direct HIV-host chromatin interactions and native HIV chromatin ultrastructure. To circumvent these complications, we exploited clonal J-Lat cellular models for our studies presented in **Chapter VI**, where we have control over the site of integration and the sensitivity to

quantify the chromatin contact profile with an ensemble average across the cellular population**.** Recent advancements in this methodology have culminated in single-cell approaches that can be multiplexed with other sequencing-based assays [373, 404, 405]. However, the resolution is still limited to observe subtle changes in chromatin organization potentially resulting from HIV-1 integration or proviral activation. To circumvent these issues, we have expanded our studies to more targeted approaches such as 4C-seq and 3C-qPCR.

We have been working on developing a 4C-seq assay to profile changes in direct HIV-host chromatin interaction frequencies, designing our assays based on prior work conducted in the Pintel lab at the University of Missouri-Columbia **(**schematic of assay is presented in **Figure I.F.2**). The Pintel group has demonstrated that they can apply their 4C-seq workflows for efficient capture and mapping of direct virus-host chromatin interactions [308, 309]. We have acquired preliminary data applying our own 4C-seq workflow to capture HIV-host chromatin interactions in the J-Lat 10.6 cell line (single, defined site of HIV-1 integration in *SEC16A* gene), as well as applying this method for kinetic studies tracking differential HBV-host chromatin interactions across a time-course following *de novo* infection.

4C-seq experiments in **Figure VII.A.1** have been performed with two biological replicates per condition (No treatment (HIV-inactive) and 10 ng/mL TNF-α treatment (HIV-active) for the experiments in J-Lat cells. Raw sequencing reads have undergone demultiplexing (removal of HIV "bait" sequence) using custom Python scripts, and reads were then directly aligned to the human reference genome (hg38) at BglII restriction fragments using Bowtie2 [406]. The reads from the biological replicates were then normalized to reads per million and quantile normalized using preprocessCore package on RStudio. For visualization of the 4C-seq data, a running average was calculated using a window size of five contiguous Bglll fragments [407].

Macs2 [408] was used to for peak calling. Distance-normalizations were also applied to the datasets following analysis pipelines previously reported[409]. Similar sample preparation strategies and downstream bioinformatics analyses were imposed on the HBV datasets (**Figure VII.A.2)**. While it is encouraging that we can record changes in interaction profiles between HIV-inactive & -active cellular J-Lat populations (**Figure VII.A.1**) and observe consistent HBV-host chromatin interactions throughout the time-course of HBV infection **(Figure VII.A.2)**, library complexity and sequencing coverage across all of the 4C replicates (including all HIV- and HBV-based samples) needs to be improved. Thus, optimization of the 4C-seq protocol is ongoing.

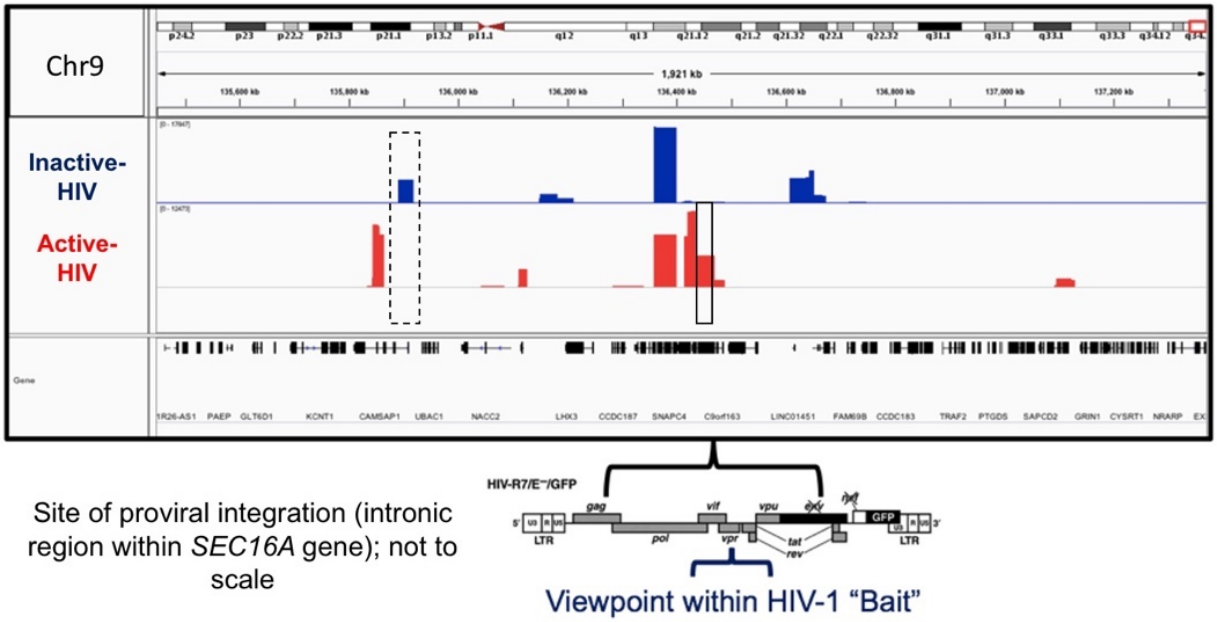
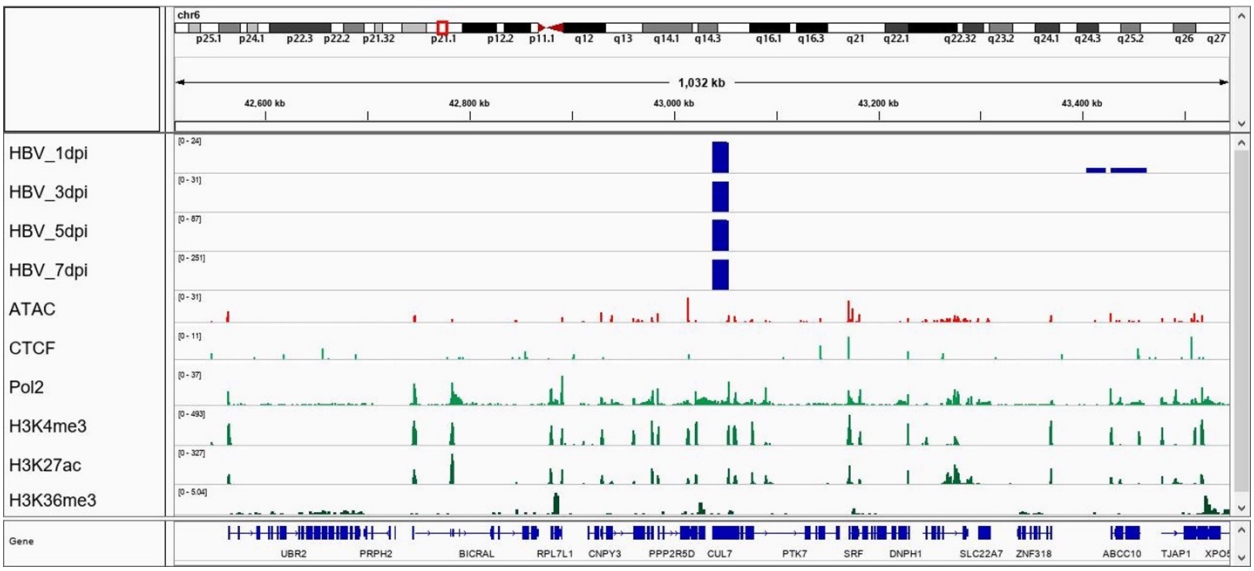


Figure VII.A.1. Preliminary 4C-seq of J-Lat 10.6 samples. The histogram in blue represents the HIV-inactive population, and red represents the activated population. Boxed regions denote genetic *loci* with differential 4C interaction profiles between the HIV-1 "bait" and flanking cellular chromatin. Coverage is very low and libraries lack sufficient complexity. Optimization is underway.

TABLE VII.1 Primers for 4C-seq library preparation of HIV-infected cells.

Target	Sequence
Inverse HindIII	**CGACGAAGAGCTCATCAGAACAGTCA**
Inverse NlaIII	**AAGCCTTAGGCATCTCCTATGGC**
Nested inverse HindIII	**CAGAACAGTCAGACTCATCA**
Nested inverse NlaIII	**GAGATGCCTAAGGCTTTTAT**

chr6-p21.2-3 (HBV Interactions in gene)



chr4-p15.33-1 (HBV Interactions not in gene)

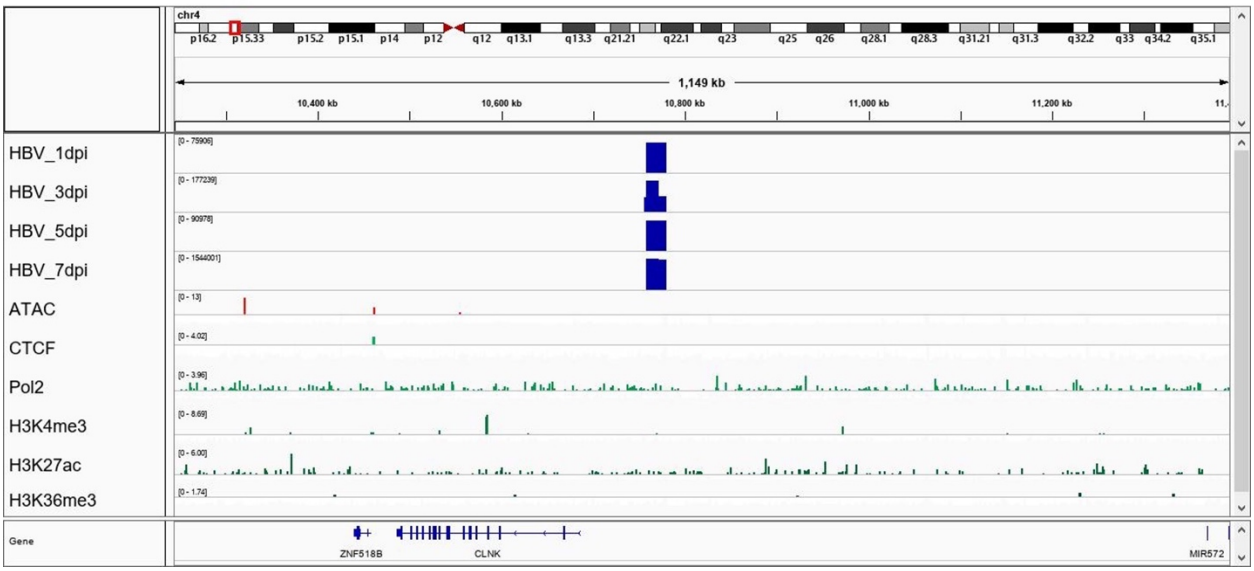


Figure VII.A.2. HBV-host chromatin 4C interaction profiles across the course of 7-day infection. Representative 4C profiles "within" or "outside" genes where interactions between HBV DNA and host chromatin were conserved across 1, 3, 5, & 7 dpi (days post-infection). Infections were performed in HepG2 cells (hepatocytic cell line), and additional ATAC- & ChIP-seq tracks were taken from studies in

HepG2 cells that were deposited in SRA & ENCODE. Similar to the J-Lat datasets in **Figure IV.A.1**, sequencing coverage and read complexity is poor. Optimization is required.

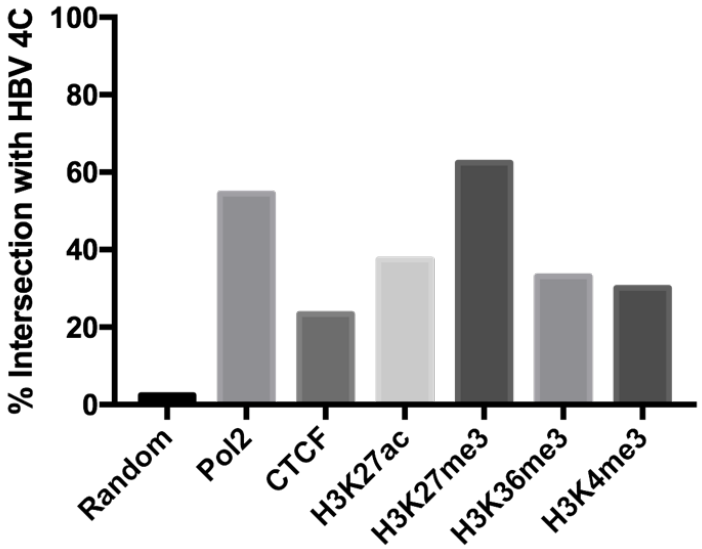


Figure VII.A.3. Jaccard analysis of HBV 4C-seq "hits" & ChIP-seq (Pol2, CTCF, HSK27ac, H3K27me3, H3K36me3, & H3K4me3). Intersection (overlap) between 4C- & ChIP-seq mapped reads. Pan-distribution of intersection across datasets with equal representation of active and repressive epigenetic markers. 4C mapped reads were merged across all four time-points for this analysis.

We began to improve our 4C-seq pipeline, systematically optimizing critical steps in the protocol and using 3C-qPCR as a read-out for our ability to capture specific "bait" and cellular chromatin interactions. Optimization was performed in the J-Lat 10.6 cells. The first parameter of the protocol that was optimized and refined was the first restriction digestion step with a six base-pair cutter. We tested different sodium dodecyl sulfate (SDS) concentrations used for lysing cells following conventional 4C-seq protocols [308, 309] and subsequently performed the first digestion step. The key to this step is obtaining efficient cellular lysis, while also not interfering with enzyme specific activity, by fine-tuning detergent concentrations. SDS is the detergent in this assay used for nuclear permeabilization following cellular lysis, while Triton X-100 is used to sequester SDS following permeabilization. This thus mitigates detergent-mediated enzyme

denaturation, which can otherwise inhibit efficient digestion. Following the first digestion step, I then tested digestion efficiency within the *GAPDH* locus, where I designed primers at both BglII and HindIII restriction sites. At these sites, I collected DNA samples pre- & post-digestion, titrating different concentrations of SDS for nuclei permeabilization and keeping Triton X-100 constant at 2% final concentration. Digestion efficiency was assessed by performing qPCR at the restriction junctions. Efficiency was inversely correlated to signal (C_t value), as less amplification should result if the region was efficiently cleaved. 0.22% SDS concentration while keeping Triton X-100 concentration constant at 2% yielded over 90% digestion efficiency (**Figure VII.A.4)**. Following optimization of digestion of the first step, ligation efficiency was next improved to capture interactions between the 'bait' sequence (Vpr within HIV-1 genome or Ercc3 control) and cellular chromatin. Quantification of chromatin interaction frequency via 3C-qPCR is calculated by normalizing interaction frequency relative to nearest neighbor interactions of contiguous HindIII fragments on the *Ercc3* locus, a site that is constitutively active ubiquitously [309, 410]. From our 3C-qPCR assay, we did not find any significant change in local HIV-host chromatin interactions in the HIV-inactive and -active states **(Figure VII.A.5)**. 4C-seq libraries need to be prepared again for HIV and HBV samples. Current studies in the lab are focusing on the role CTCF plays in HBV replication. Functional cell-based assays are underway, identifying intrinsic CTCF-binding sites within the HBV DNA genome. Applications of an optimized 4C-seq approach can ultimately enable us to interrogate HIV-host chromatin interactions following *de novo* infection, where we can demonstrate if HIV is preferentially interacting with specific genomic regulatory elements across several integration sites. Sorting for HIV-inactive and -active populations, we can further determine if there are differences in HIV-1 chromatin interaction profiles across the two infected cellular populations.

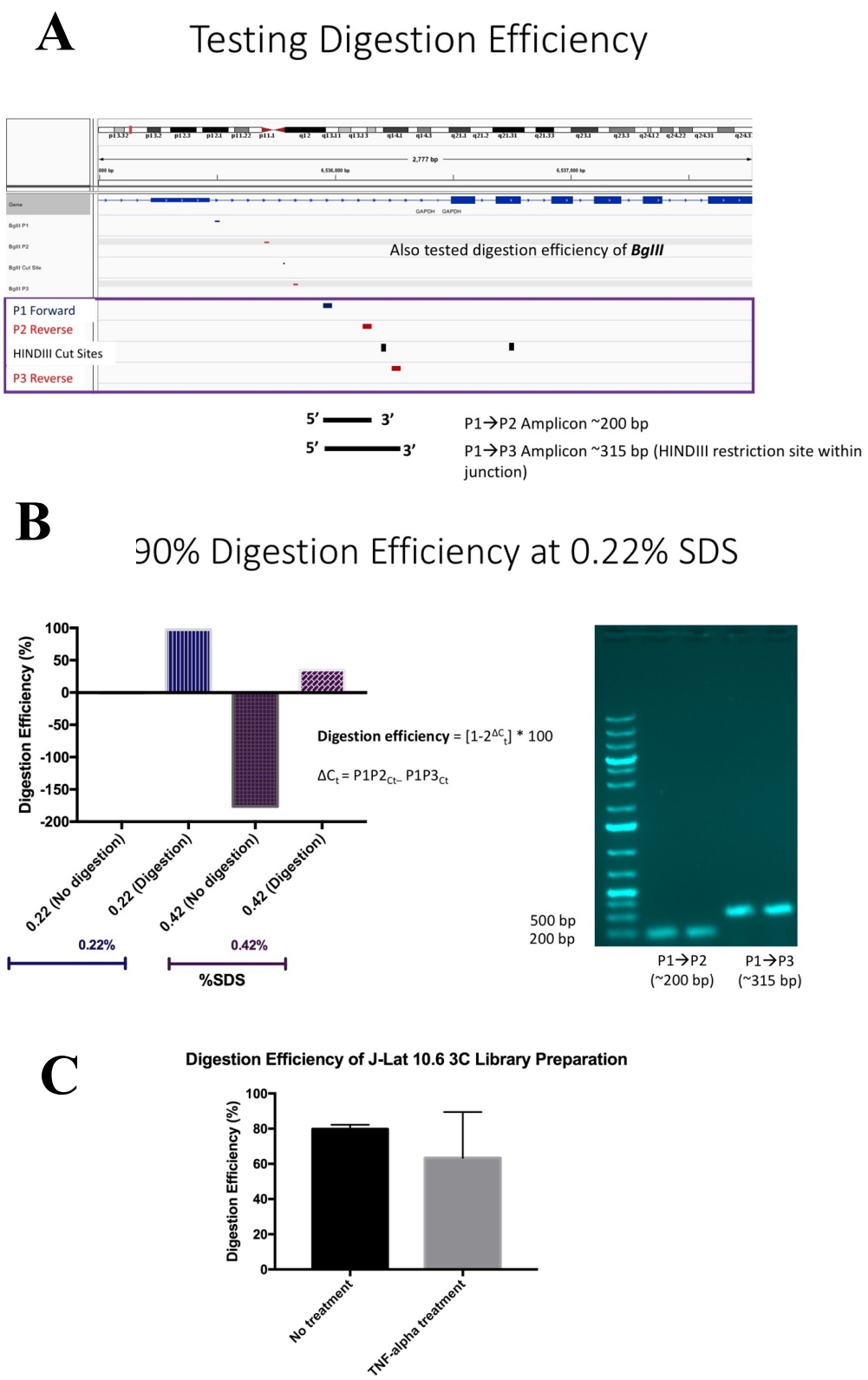


Figure VII.A.4. Optimization of digestion efficiency in 3C/4C assays. Restriction digestion junctions at *GAPDH* locus. We designed primers at "inside" & "outside" HindIII or BglII restriction sites. In **Figures B-C**, we tested digestion efficiency of HindIII **(A)**. Both primer sets were specific to respective

locus of interest. qPCR samples were run on 0.8% agarose gel and single bands detected in lanes corresponding to amplicons "outside" & "inside" HindIII restriction sites, respectively. 0.22% SDS concentration had best digestion efficiency (> 90%) **(B).** Digestions were performed with HINDIII (0.22% SDS & 2% Triton X-100) in J-Lat cells in biological duplicates. Digestion efficiency at 0.22% SDS resulted ~70-80% digestion efficiency in these samples **(C).**

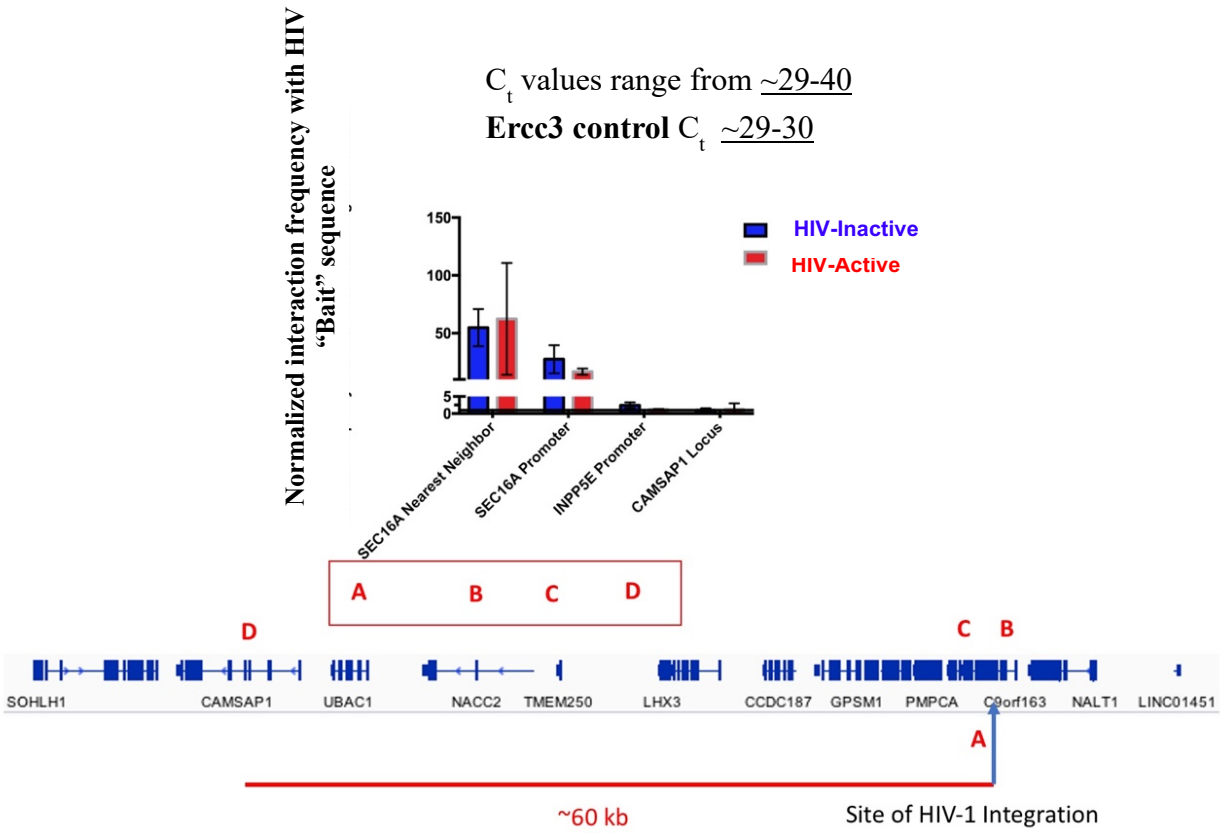


Figure VII.A.5. 3C-qPCR assay in J-Lat 10.6 cells to detect frequency of direct HIV-host chromatin interactions around the site of integration. In HIV-inactive (blue) & -active (red) cellular populations, differential HIV-host chromatin interaction frequencies were quantified *via* qPCR read-out. C_t values were normalized to nearest neighbor interactions at *Ercc3* locus (control). Primer pairs were designed ~50 bp upstream nearest contiguous HindIII site and designed primer ~50 bp downstream contiguous HindIII site in HIV "bait". The HIV "bait" primer set designed at the *vpr* locus is constant across each qPCR reaction. A > B > C > D (Distance from "viewpoint"; A is closest; D is > 60 kb away). Low interaction frequency with flanking cellular genes in both inactive & active populations.

TABLE VII.2. Primers & Taqman probes designed for 3C-qPCR analysis of J-Lat 10.6 cells.

Primer Target	Sequence	Notes
HIV 'bait'	CAAAAGCCTTAGGCATCTCC	Targets HIV 'viewpoint' (vpr)
HIV Taqman probe	ACAGCGACGAAGAGCTCATCa	5' FAM 3' TAMRA
Sec16A	CTCAAGATCCTCCCCCAAAA	Nearest gene to 'viewpoint' in J-Lat 10.6 model
Sec16A	CAGAACACGCAAACTTCCAA	Promoter
INPP5E	CACAGGTGAAGGTGCTTGAA	Promoter
CAMSAP1	CGGAGAGATTTCCGTCACTC	> 60 kb from HIV "viewpoint"
ERCC3 'bait' primer	GTAAGCAGGCTGGAGCTGAG	Control/normalization of libraries
ERCC3 probe	GGGAAAGGGACTGCTGTGTA	5' FAM 3' TAMRA
ERCC3 capture	GGAAGGATCTCTGTTTAATGGAAA	Control

VII-B: Multiomics Single-Cell Profiling of Primary HIV-Infected Cells

Studies mapping individual HIV chromatin conformation states using primary patient infected cells taken *ex vivo* is the penultimate goal for our studies, however, close interrogation of the functional state of HIV and the respective flanking cellular chromatin poses technical hurdles. The first hurdle is that HIV integration sites must be determined if studies are conducted using patient cells taken *ex vivo* or cells infected *de novo*, so multiplexed approaches that enable joint integration site identification and profiling of these respective integration sites is needed [377]. If integration sites can be mapped per infected cell, the second major obstacle of HIV-host chromatin profiling studies is that NGS-based methodologies such as bulk ATAC-seq acquires genome-wide information. This thus imparts a poor signal-to-noise ratio at individual proviral integration sites in bulk infected population, since presumably, each identified integration site will only be represented by a single cell in population of several thousands of cells. The average aggregate signal will be recorded across the cellular population, prohibiting sufficient resolution to interrogate individual proviruses and profile the chromatin landscape at these single integration sites.

Single-cell sequencing can enable closer interrogation of these respective individual proviral integration sites following bulk infection, with pipelines developed that enable multiomics single-cell ATAC- and RNA-seq (scATAC0 & scRNA-seq) profiling [411, 412]. However, to determine proviral integration sites with single-cell resolution, sequencing depth needs to be increased significantly (> 10-fold) relative to conventional scATAC-seq [412], thus driving up sequencing costs. Wang *et al.* has developed a computational pipeline for deriving integration sites from NGS-based data by looking for cellular regions of overlap with a proviral genome of interest during reference alignment [412]. Identifying chimeric provirus-host fragments

("multi-mappers" that align to both host and proviral genomes) following Tn5 tagmentation during ATAC-seq preparation can reveal sites of integration with high precision. A limitation to this methodology at the moment, however, is that single-cell ATAC read pile-up at individual proviral integration sites is relatively low in a single cell [412]. Further enrichment and PCR-based amplification strategies are needed to increase targeted sequencing of HIV and HIV-host genic boundaries to improve the signal collected during single-cell assays to profile HIV and HIV-host chromatin states.

Single-cell sequencing approaches also provide the advantage of probing the heterogeneity of the chromatin and transcriptional landscapes across a diverse cellular population. scATAC- and scRNA-seq of mixed cellular populations such as PBMCs present the opportunity to sample the chromatin accessibility and gene expression networks across a diverse mixture of cells, clustering subpopulations based on single-cell epigenomic and transcriptomic profiles. Profiling different primary T cell subtypes and lymphocytic cells in the context of HIV-1 infection may reveal potential cell-specific factors involved in HIV-1 pathogenesis. Further enrichment of latently infected cells for multiplexed single-cell approaches (integration site analysis & scATAC- or scRNA-seq) can provide specific insight into the functional state surrounding latent proviruses and perhaps cellular factors up-regulated in latently infected cells [377].

VII-C: Motif Analysis at HIV-1 Integration Sites Derived from ART-Suppressed Clinical Isolates

Extensive HIV-1 integration site mapping studies have been conducted, providing the scientific community with several valuable datasets to study characteristics of HIV-1 integration "hotspots" (RIGs). We were especially interested in identifying transcription factor motifs enriched proximally to heavily targeted HIV-1 integration sites. Spearheaded by a doctoral student in the Sarafianos group during his first-year research rotation project, William McFadden developed a computational pipeline for aggregating preferential HIV-1 integration sites with accompanying meta-data (site of integration, directionality, genomic *loci* (intron, exon, intergenic, intragenic)). After collecting integration sites, he extracted the primary DNA sequence (5' – 3') from a fasta file of the human reference genome (hg38 or hg19) at varying window sizes from the respective site of integration (eg. 50 bp upstream/downstream site of integration). Once the primary sequences around the sites of integration were gathered, motif analysis was performed to identify putative transcription factor binding sites "at" and flanking common HIV-1 sites of integration. In addition, a null model was built based on random genic sequences extracted of varying lengths from the hg38 reference genome to control for frequency of random motif matches.

To test this pipeline, sequences were collected from Maldarelli *et al.* [200]. In this longitudinal study, CD4+ T cells were collected from five patients on cART over a span of several years. Integration sites across these cellular samples were aggregated. In our analysis, we selected for integration sites that Maldarelli *et al.* [200] identified as being represented across multiple datasets (RIGs) and appear to be persistent across time-points and from patient-to-patient. After extracting integration sites from this dataset, we compiled the neighboring cellular sequences flanking the respective integration sites at various sizes "downstream" and "upstream"

the site of integration. Motif analysis was performed using the FIMO analysis pipeline [413]. Motif analysis of the integration sites was compared to a null model. The "random" function from the BEDtools bioinformatics suite [414] was used to randomly select genomic coordinates from the human reference genome. Random sequences were generated with varying window sizes to directly compare to the experimentally derived sequences. Random sequences with ambiguous nucleotides (variable nucleotide (N) due to poor sequence coverage at that genomic site in the human genome (e.g., centromeric repeat)) were removed and replaced by another randomly generated sequence. Enrichment analysis was performed to identify putative motifs at persistent HIV-1 sites of integration. This whole pipeline has been automated *via* various R packages (Biostrings v2.58.0, GenomicRanges v1.42.0 [415], & BSgenome v1.58.0) collated into a single R script. Quality control of this pipeline demonstrates that this script can aggregate integration sites, extract sequence information, and perform analysis, as expected. The number of motif matches increases as a function of window size **(Figure VII.C.1).** Conceptually, the probability of motif occurrence should increase as the length of sequence surveyed increases. Preliminary analysis highlights motifs that are up-regulated in experimentally derived integration sites *vs.* the randomly generated sequences. We highlight the transcription factors heat shock factor-1 (HSF1) and Fox01 **(Figure VII.C.2)**, which are nuclear transcription factors that were both recently reported to play a role in influencing the status of viral latency [263, 264]. In addition, we compare integration site frequency "upstream" the site of integration *vs.* "downstream" the 3' LTR, taking into account the orientation of the provirus at the site of integration sampled. We further test the role of HSF1-mediated proviral transcriptional activation in J-Lat models in the next Chapter **(Chapter VII.4).** While we now have a pipeline for efficiently collecting and archiving integration sites from sequencing datasets, we need to refine our statistical rigor to filter for

motifs significantly overrepresented at HIV-1 integration sites. From our output, we are generating contingency tables, taking a tally of motif frequency across all of the integration sites sampled. We will therefore employ an appropriate frequency test (Fischer Exact Test or Wilcoxon Rank Sum Test) for our motif analysis pipeline.

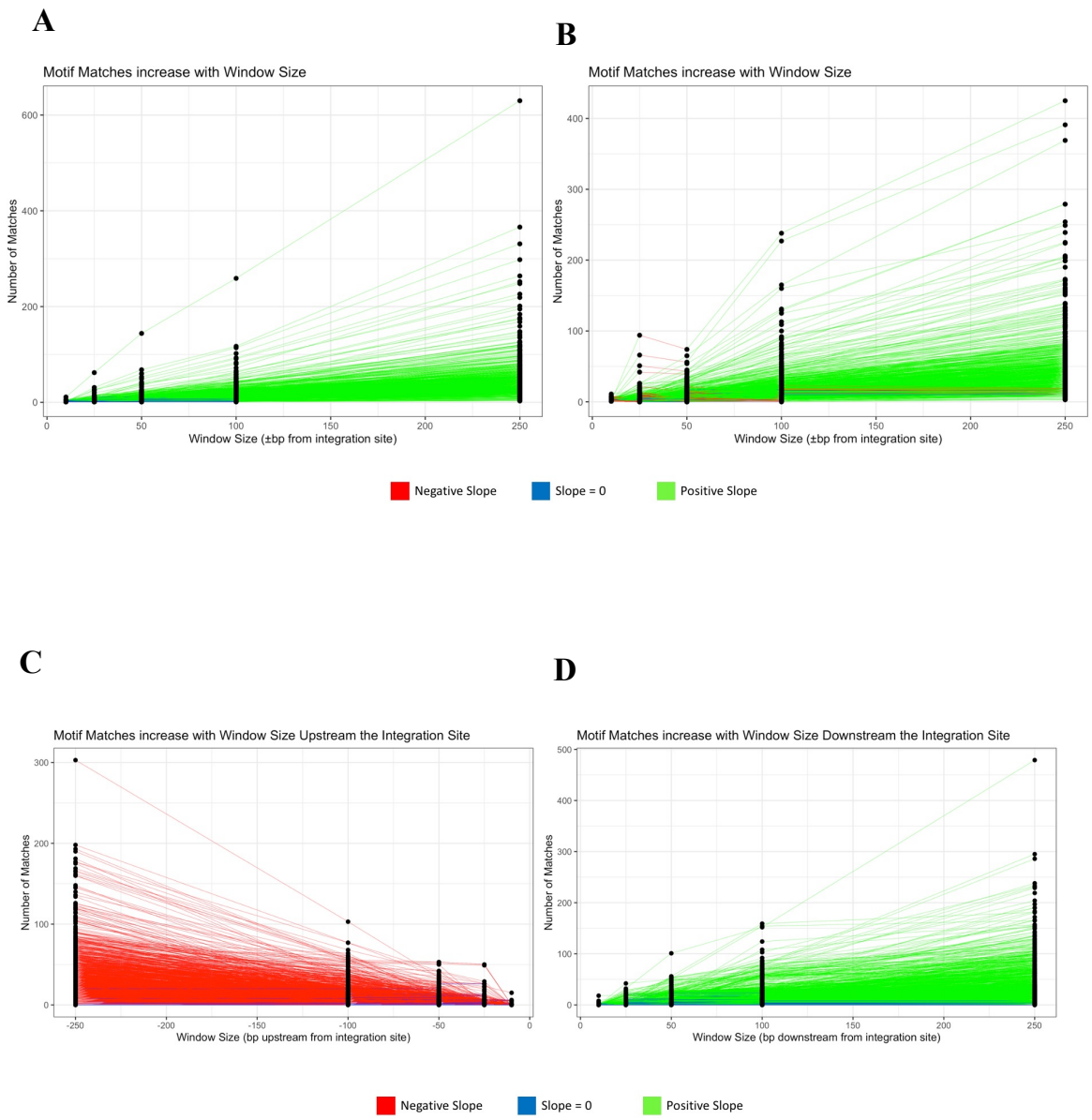
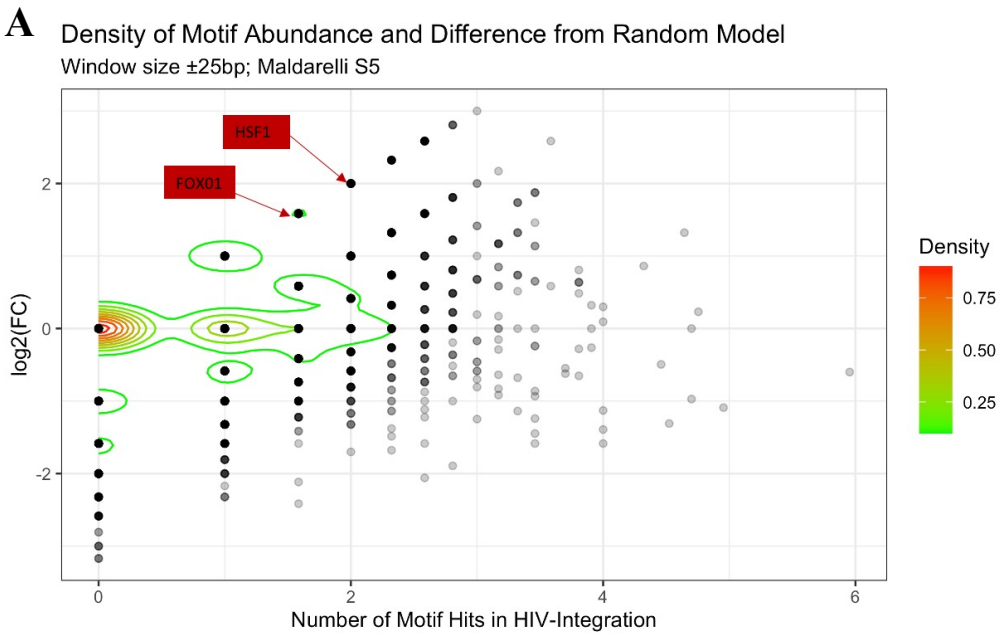


Figure VII.C.1. Quality control of computational pipeline for integration site retrieval & extraction of primary sequences from supplied fasta files (human reference genome; hg19). Motif matches as a function of sequence size for experimentally derived HIV-1 integration sites **(A)** & randomly generated sequences **(B).** Sequences obtained either "upstream" **(C)** or "downstream" **(D)** HIV-1 integration sites. **Credit:** William McFadden conducted analysis and generated plots.



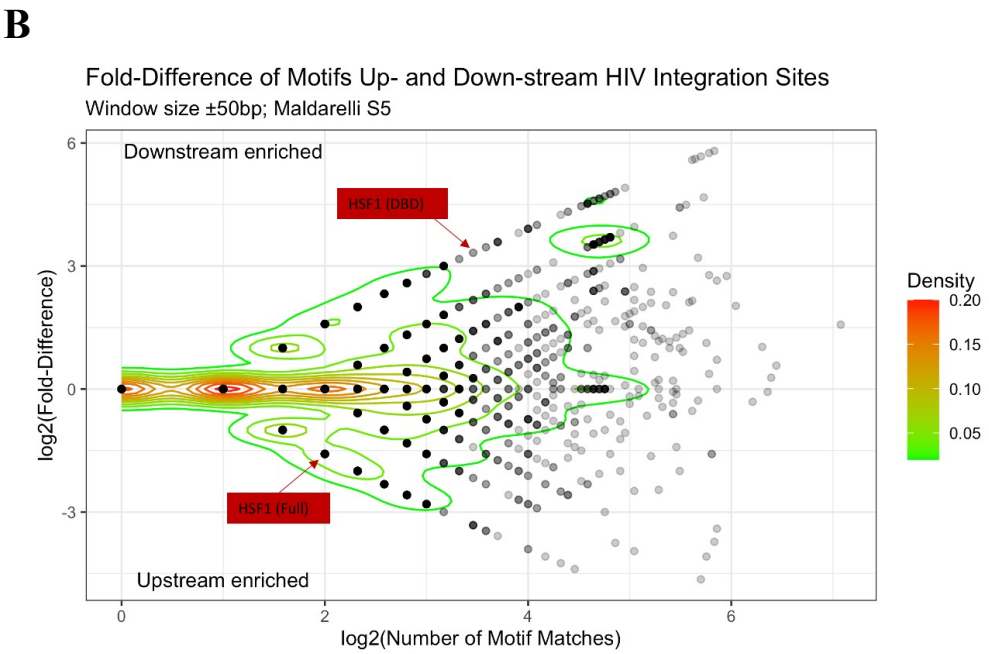


Figure VII.C.2. Density plots of motif occurrences across integration sites sampled. Log_2 fold-change (experimental *vs.* null model) of motif representation as a function of total motif matches across integration sites. Differences in HSF1 & FoxO1 motif occurrence *vs.* random model highlighted **(A).** Differences in motif frequency "downstream" integration site *vs.* "upstream" integration site **(B). Credit:** William McFadden conducted analysis and generated plots.

While our integration site analysis is in its incipient stages, we have demonstrated a proof-of-concept of the functionality of our automated R script and analysis pipeline using a test set of RIGs. Future integration site analysis can apply this computational pipeline to compare motif enrichment at integration sites from latent populations (reactivatable *vs.* non-reactivatable), elite controllers (minority patient population that has intrinsic low viral set point), integration sites obtained following treatment with CA or IN inhibitors, and sites identified following infection with mutations in CA (N74D) or IN. In these mutants, interactions with host factors and either CA or IN are ablated, thus leading to noncanonical integration site targeting [158]. Differences in the properties of integration sites across these multiple populations sampled can provide insight into how variation in integration site selection affects viral replication efficacy. Understanding how aberrant integration impacts proviral transcriptional competency may be important for the development of ALLINIs (allosteric IN inhibitors) that can ultimately suppress viral replication *via* integration mistargeting. An additional application of this script can enable collection of integration sites for epigenomic and transcriptomic profiling (cross-correlation to ChIP-, ATAC-, & RNA-seq datasets) to define the transcriptional and functional states around proviral integration sites.

VII-D: Putative Cellular Pathways Involved in HIV-1 Transcriptional Activation & Chromatin Reorganization

In **Chapter VI**, we identified common patterns of chromatin reorganization and transcriptional reprogramming at activated HIV-1 proviruses using inducible J-Lat cellular models and different NGS- and fluorescence-based approaches. Our observation that chromatin accessibility increases downstream the 3' LTR at the HIV-host intergenic boundary and that host- and HIV-driven transcriptional read-through are present at the proviral locus emulates a similar phenotype demonstrated in cells infected with lytic herpes simplex virus I (HSV-1) and mammalian cells undergoing cellular stress [392]. Hennig *et al.* showed that chromatin accessibility increases at genes where transcriptional read-through is triggered *via* lytic HSV-1 infection or cellular stressors [392]. The Steitz group at Yale University first "coined" these read-through products as "Downstream-of-Genes" or "DoGs" [388, 389]. The biological function(s) of DoGs are not clear; however, they are directly triggered by various cellular stressors, and interestingly, there are differential DoG expression patterns across different stress conditions (heat, osmotic, oxidative) [389]. This suggests that DoGs may be playing specific roles related to the particular cellular stress pathway.

From our differential ATAC-seq of J-Lat cells in HIV-inactive and -active states, we compiled differential peaks "at" and flanking the HIV-1 provirus and performed motif analysis as described in **Figure VII.D.2.** While we were only sampling two independent proviral integration sites (J-Lat 10.6 & 8.4 proviruses), two of our strongest motif "hits" downstream the 3' LTR locus included HSF1 and FoxO1 **(Figure VII.D.1)**. Interestingly, the Siliciano group recently reported that induction of HSF1-signaling pathways promotes latency reversal in

primary CD4+ models [263], and the Ott group reported that inhibition of FoxO1 up-regulates the ER stress response, promoting latency reactivation. These two independent studies found that conditions inflicting cellular stress were increasing reactivation frequency of latently infected cells. We decided to test the role HSF1 and the heat shock pathway may play in HIV-1 transcriptional activation in our J-Lat models.

In preliminary cell-based studies, we found that heat shock pathways appear to improve proviral transcriptional activation within our J-Lat models **(Figure VII.D.3)**, and TNF-**α** stimulation increases HSF1 localization into the nucleus **(Figure VII.D.4)**. TNF-**α** signaling may be playing an indirect role in activation of HSF1-related pathways. HSF1 is the master regulator of the heat shock response and transcription factor that drives expression of HSP70 family of proteins [416]. Interestingly, at our differential ATAC peaks directly downstream the 3' LTR, our motif analysis identified HSF1 as a top "hit". This would thus suggest that HSF1 may have the ability to bind downstream HIV-1 in these models, and the possible role of HSF1 binding downstream activated HIV-1 should be investigated. Conventionally, HSF1 interacts at the 5' gene promoter, recruiting stable transcription initiation complexes and P-TEFb, promoting transcriptional elongation [263, 389, 394]. However, there may be multiple biological functions that HSF1 performs including chromatin remodeling, which has not been investigated.

Perhaps, HSF1 is recruiting chromatin remodeling factors to "open" up chromatin structure downstream HIV-1 genes. To test this hypothesis, biochemical studies need to be conducted. ChIP experiments can be performed pulling-down HSF1 and probing for direct interactions with HIV-1 DNA and cellular chromatin flanking the provirus. We are also interested in biochemically identifying other putative activators or repressors of HIV-1 gene expression. A method "coined" enChIP, utilizes an endonuclease-dead dCas9 fusion protein to

specifically target genetic *loci* of interest and can be subsequently pulled-down by targeting the fusion protein *via* IP. This thus enables pulling down gene regulatory sequences of interest. Prior to IP, cellular samples are fixed *via* cross-linking, maintaining protein-DNA complexes. Proteomic analysis can then be performed followed IP to identify protein factors bound to the DNA locus of interest [417-420]. We have been interested in applying this technology to pull down the HIV-1 5' LTR promoter in differential transcriptional states to identify *cis*-acting nuclear factors associated with proviral suppression or activation. These studies can provide mechanistic insight into cellular chromatin modifying factors that may be influencing the epigenetic and functional state of the provirus.

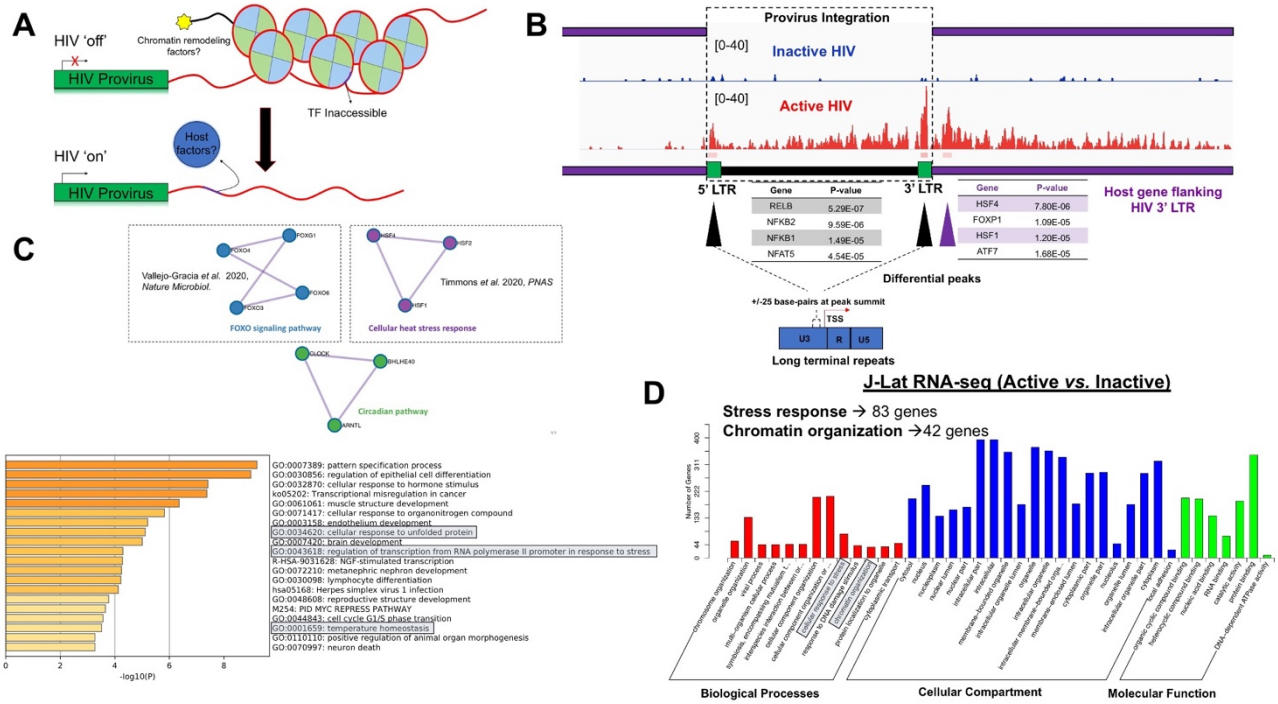


Figure VII.D.1. Predicting cellular host factors and signaling pathways mediating HIV transcriptional activation and local HIV chromatin reorganization. Schematic of cellular chromatin "opening" downstream the active HIV 3' LTR (**A**). Transcription factor (TF) motif analysis at sites of differential chromatin accessibility (5' LTR, 3' LTR, & downstream 3' LTR) at proviruses in inactive *vs.*

active states. Genes in the black table denote shared TFs between the ATAC peaks at the 5' and 3' LTRs. Peaks at these two respective positions were within 20 base pairs of each other, slightly upstream HIV transcriptional start site (TSS). The purple table displays transcription factor motifs enriched at the ATAC peak downstream the active 3' LTR. Motif search window was +/- 25 base pairs from the ATAC peak summit (**B**). Gene ontology (GO) analysis of top motifs enriched downstream the 3' LTR performed using Metascape [421]. Pathways involved in FoxO signaling and the cellular heat stress response were highly represented in the motif analysis, and two recent publications show the role FoxO1 and the heat-shock factor 1 (HSF1) play in HIV transcriptional activation (**C**). GO analysis of RNA-seq of J-Lats (active *vs.* inactive). 83 and 42 genes involved in cellular stress responses and chromatin organization, respectively, differentially up-regulated. These genes are also not differentially up-regulated following TNF treatment in uninfected Jurkats (**D**).

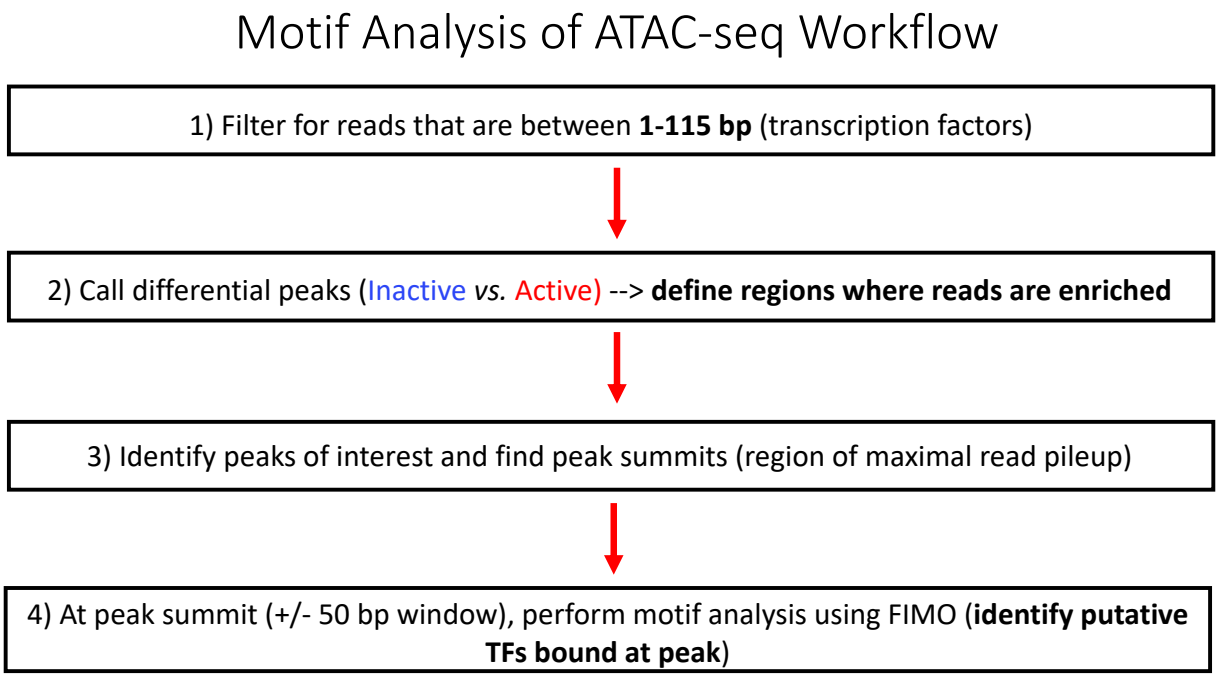


Figure VII.D.2. Workflow for motif analysis of ATAC peaks. Mapped paired-end reads that are 1-115 bp in fragment size are selected. Differential peaks are called to select for regions of increased chromatin accessibility following HIV activation in the J-Lat models. At the summits (maximal number of reads) of the differential peaks, we collect primary sequences of varying window sizes at the selected peak summits. We then perform motif analysis using FIMO [413] to identify putative transcription factor binding sites.

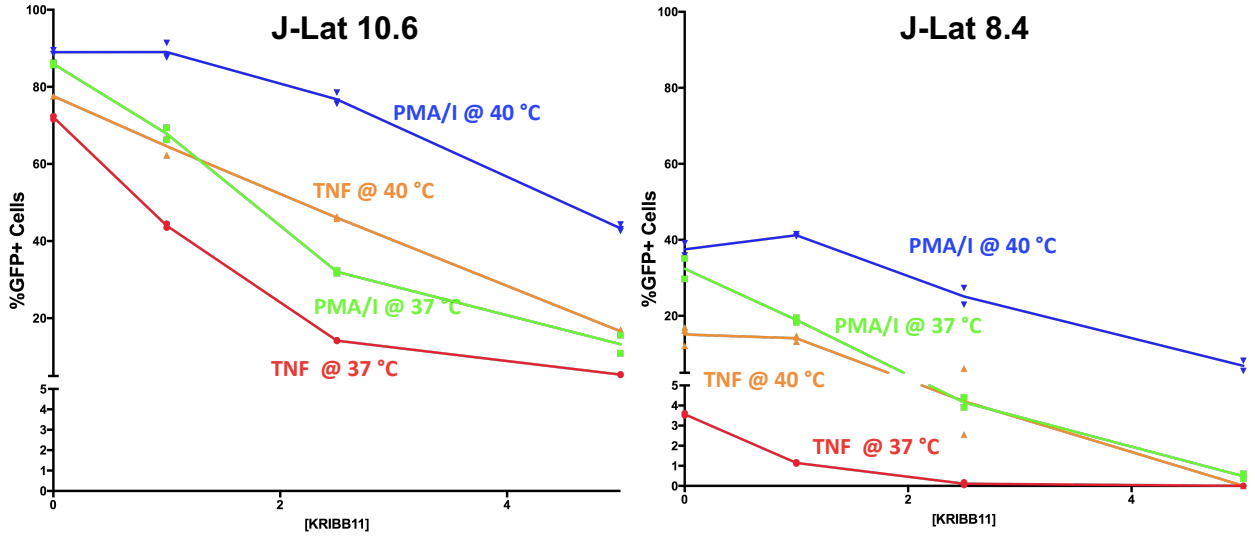
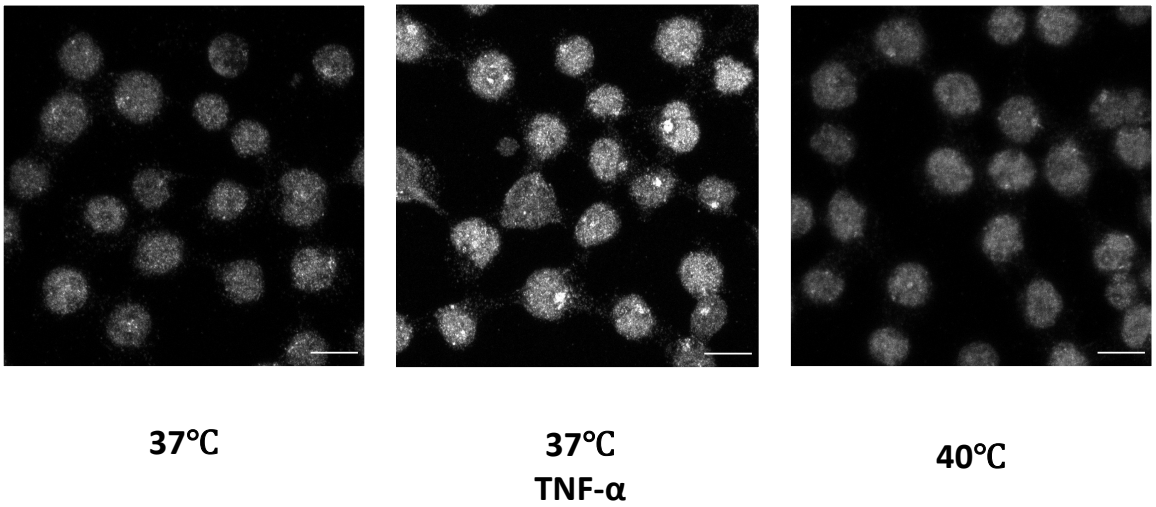


Figure VII.D.3. Effect of HSF1-mediated signaling on latency reversal in J-Lat cells. J-Lat 10.6 (left) and 8.4 cells (right) are treated with two different LRAs (TNF or PMA/I), and we performed a dose-response of HSF1 inhibition *via* the KRIB11 inhibitor (inhibits P-TEFb recruitment by HSF1 at gene promoters). Latency reactivation was blocked in a dose-dependent manner; however, the rate of inhibition was stalled when cells were incubated at 40 °C. A similar trend is observed in both J-Lat models.



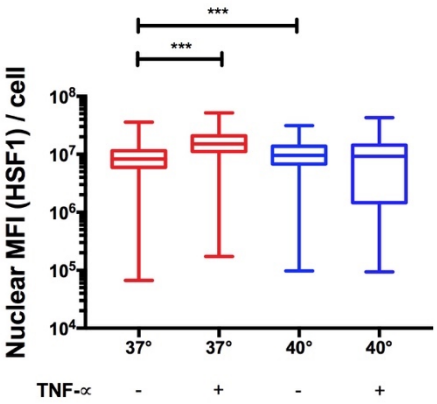


Figure VII.D.4. HSF1 nuclear localization following TNF-α treatment and heat shock. Immunofluorescence of Jurkat T cells incubated at 37°C, 37°C & TNF-α, and 40°C (heat shock condition). HSF1 staining conditions: primary; anti-HSF1 (1:1000) and goat anti-mouse Alexa 647 secondary (1:2000). Scale bars represent 10 µm **(A)**. Nuclear mean fluorescence intensity (MFI) of HSF1 immunostaining (**B**).

VII-E: Concluding remarks

In this presented work, we demonstrate how integrative approaches utilizing deep sequencing and multiplexed *in-situ* imaging can be applied in conjunction to shed important insight into molecular mechanisms underlying viral gene regulation and replication. Much of our work was conducted in HIV-relevant models, demonstrating how lentiviruses can alter local chromatin architecture and host transcriptional patterns. Moving forward, our work presented in **Chapter VI** should be validated in primary cell models following *de novo* infection to verify our findings that HIV-1 proviral activation elicits nucleosome remodeling downstream the HIV-1 3' LTR and further characterize the role chimeric HIV-host RNAs play in viral replication. These studies will depend on obtaining sufficient sequencing depth and coverage using single-cell sequencing approaches to jointly identify individual HIV-1 integration sites across the infected cellular population and probe the RNA landscape (*via* scATAC- or scRNA-seq) around these respective sites of integration. Thus, this would provide a method to survey the heterogeneity of HIV integration sites with single-cell resolution in a more disease-relevant context.

Collectively, this work serves as a foundation for future integrative studies across a broad spectrum of pathogens to study pathogen-host factor interactions, viral gene regulation, and the relationship between nuclear chromatin organization and RNA transcription.

REFERENCES

1. Deeks SG, Overbaugh J, Phillips A and Buchbinder S. HIV infection. *Nature Reviews Disease Primers*. 2015;1:15035.

2. Chun TW, Engel D, Mizell SB, Hallahan CW, Fischette M, Park S, Davey RT, Jr., Dybul M, Kovacs JA, Metcalf JA, Mican JM, Berrey MM, Corey L, Lane HC and Fauci AS. Effect of interleukin-2 on the pool of latently infected, resting CD4+ T cells in HIV-1-infected patients receiving highly active anti-retroviral therapy. *Nature medicine*. 1999;5:651-5.

3. Brenchley JM, Hill BJ, Ambrozak DR, Price DA, Guenaga FJ, Casazza JP, Kuruppu J, Yazdani J, Migueles SA, Connors M, Roederer M, Douek DC and Koup RA. T-cell subsets that harbor human immunodeficiency virus (HIV) in vivo: implications for HIV pathogenesis. *J Virol*. 2004;78:1160-8. PMC321406.

4. Sonza S, Mutimer HP, Oelrichs R, Jardine D, Harvey K, Dunne A, Purcell DF, Birch C and Crowe SM. Monocytes harbour replication-competent, non-latent HIV-1 in patients on highly active antiretroviral therapy. *AIDS (London, England)*. 2001;15:17-22.

5. Wu L and KewalRamani VN. Dendritic-cell interactions with HIV: infection and viral dissemination. *Nature reviews Immunology*. 2006;6:859-68. PMC1796806.

6. Craigie R and Bushman FD. Host Factors in Retroviral Integration and the Selection of Integration Target Sites. *Microbiology spectrum*. 2014;2. PMC4525071.

7. Gallo RC, Salahuddin SZ, Popovic M, Shearer GM, Kaplan M, Haynes BF, Palker TJ, Redfield R, Oleske J, Safai B and et al. Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. *Science*. 1984;224:500-3.

8. Popovic M, Sarngadharan MG, Read E and Gallo RC. Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS. *Science*. 1984;224:497-500.

9. Barré-Sinoussi F, Chermann JC, Rey F, Nugeyre MT, Chamaret S, Gruest J, Dauguet C, Axler-Blin C, Vézinet-Brun F, Rouzioux C, Rozenbaum W and Montagnier L. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science*. 1983;220:868-71.

10. Barré-Sinoussi F, Ross AL and Delfraissy J-F. Past, present and future: 30 years of HIV research. *Nature Reviews Microbiology*. 2013;11:877-883.

11. Balasubramaniam M, Pandhare J and Dash C. Immune Control of HIV. *Journal of life sciences (Westlake Village, Calif)*. 2019;1:4-37. PMC6714987.

12. Reeves JD and Doms RW. Human immunodeficiency virus type 2. *The Journal of general virology*. 2002;83:1253-1265.

13. Faria NR, Rambaut A, Suchard MA, Baele G, Bedford T, Ward MJ, Tatem AJ, Sousa JD, Arinaminpathy N, Pépin J, Posada D, Peeters M, Pybus OG and Lemey P. HIV epidemiology. The early spread and epidemic ignition of HIV-1 in human populations. *Science*. 2014;346:56-61. PMC4254776.

14. Sharp PM and Hahn BH. Origins of HIV and the AIDS pandemic. *Cold Spring Harbor perspectives in medicine*. 2011;1:a006841. PMC3234451.

15. Connor RI, Sheridan KE, Ceradini D, Choe S and Landau NR. Change in coreceptor use correlates with disease progression in HIV-1--infected individuals. *The Journal of experimental medicine*. 1997;185:621-8. PMC2196142.

16. Keele BF, Giorgi EE, Salazar-Gonzalez JF, Decker JM, Pham KT, Salazar MG, Sun C, Grayson T, Wang S, Li H, Wei X, Jiang C, Kirchherr JL, Gao F, Anderson JA, Ping LH, Swanstrom R, Tomaras GD, Blattner WA, Goepfert PA, Kilby JM, Saag MS, Delwart EL, Busch MP, Cohen MS, Montefiori DC, Haynes BF, Gaschen B, Athreya GS, Lee HY, Wood N, Seoighe C, Perelson AS, Bhattacharya T, Korber BT, Hahn BH and Shaw GM. Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proc Natl Acad Sci U S A*. 2008;105:7552-7. PMC2387184.

17. Jones RB and Walker BD. HIV-specific $CD8^+$ T cells and HIV eradication. *J Clin Invest*. 2016;126:455-63. PMC4731167.

18. Maddon PJ, Dalgleish AG, McDougal JS, Clapham PR, Weiss RA and Axel R. The T4 gene encodes the AIDS virus receptor and is expressed in the immune system and the brain. *Cell*. 1986;47:333-48.

19. Landau NR, Warton M and Littman DR. The envelope glycoprotein of the human immunodeficiency virus binds to the immunoglobulin-like domain of CD4. *Nature*. 1988;334:159-62.

20. Hwang SS, Boyle TJ, Lyerly HK and Cullen BR. Identification of the envelope V3 loop as the primary determinant of cell tropism in HIV-1. *Science*. 1991;253:71-4.

21. Deng H, Liu R, Ellmeier W, Choe S, Unutmaz D, Burkhart M, Di Marzio P, Marmon S,

Sutton RE, Hill CM, Davis CB, Peiper SC, Schall TJ, Littman DR and Landau NR. Identification of a major co-receptor for primary isolates of HIV-1. *Nature*. 1996;381:661-6.

22. Chan DC, Fass D, Berger JM and Kim PS. Core structure of gp41 from the HIV envelope glycoprotein. *Cell*. 1997;89:263-73.

23. Weissenhorn W, Dessen A, Harrison SC, Skehel JJ and Wiley DC. Atomic structure of the ectodomain from HIV-1 gp41. *Nature*. 1997;387:426-30.

24. Peng K, Muranyi W, Glass B, Laketa V, Yant SR, Tsai L, Cihlar T, Müller B and Kräusslich HG. Quantitative microscopy of functional HIV post-entry complexes reveals association of replication with the viral capsid. *eLife*. 2014;3:e04114. PMC4293571.

25. Hulme AE, Perez O and Hope TJ. Complementary assays reveal a relationship between HIV-1 uncoating and reverse transcription. *Proc Natl Acad Sci U S A*. 2011;108:9975-80. PMC3116424.

26. Zhong Z, Ning J, Boggs EA, Jang S, Wallace C, Telmer C, Bruchez MP, Ahn J, Engelman AN, Zhang P, Watkins SC and Ambrose Z. Cytoplasmic CPSF6 Regulates HIV-1 Capsid Trafficking and Infection in a Cyclophilin A-Dependent Manner. *mBio*. 2021;12. PMC8092277.

27. Engelman AN. Multifaceted HIV integrase functionalities and therapeutic strategies for their inhibition. *The Journal of biological chemistry*. 2019;294:15137-15157. PMC6791320.

28. Engelman AN and Cherepanov P. Close-up: HIV/SIV intasome structures shed new light on integrase inhibitor binding and viral escape mechanisms. *The FEBS journal*. 2021;288:427-433. PMC7719574.

29. Engelman AN and Singh PK. Cellular and molecular mechanisms of HIV-1 integration targeting. *Cell Mol Life Sci*. 2018;75:2491-2507. PMC6004233.

30. Li C, Burdick RC, Nagashima K, Hu WS and Pathak VK. HIV-1 cores retain their integrity until minutes before uncoating in the nucleus. *Proc Natl Acad Sci U S A*. 2021;118. PMC7958386.

31. Zila V, Margiotta E, Turoňová B, Müller TG, Zimmerli CE, Mattei S, Allegretti M, Börner K, Rada J, Müller B, Lusic M, Kräusslich HG and Beck M. Cone-shaped HIV-1 capsids are transported through intact nuclear pores. *Cell*. 2021;184:1032-1046.e18. PMC7895898.

32. Battivelli E, Dahabieh MS, Abdel-Mohsen M, Svensson JP, Tojal Da Silva I, Cohn LB, Gramatica A, Deeks S, Greene WC, Pillai SK and Verdin E. Distinct chromatin functional states correlate with HIV latency reactivation in infected primary CD4(+) T cells. *eLife*. 2018;7. PMC5973828.

33. Bedwell GJ and Engelman AN. Factors that mold the nuclear landscape of HIV-1 integration. *Nucleic Acids Research*. 2021;49:621-635.

34. Kim SY, Byrn R, Groopman J and Baltimore D. Temporal aspects of DNA and RNA synthesis during human immunodeficiency virus infection: evidence for differential gene expression. *J Virol*. 1989;63:3708-13. PMC250962.

35. Butler SL, Hansen MS and Bushman FD. A quantitative assay for HIV DNA integration in vivo. *Nature medicine*. 2001;7:631-4.

36. Farnet CM and Haseltine WA. Circularization of human immunodeficiency virus type 1 DNA in vitro. *J Virol*. 1991;65:6942-52. PMC250802.

37. Kilzer JM, Stracker T, Beitzel B, Meek K, Weitzman M and Bushman FD. Roles of host cell factors in circularization of retroviral dna. *Virology*. 2003;314:460-7.

38. Espeseth AS, Felock P, Wolfe A, Witmer M, Grobler J, Anthony N, Egbertson M, Melamed JY, Young S, Hamill T, Cole JL and Hazuda DJ. HIV-1 integrase inhibitors that compete with the target DNA substrate define a unique strand transfer conformation for integrase. *Proc Natl Acad Sci U S A*. 2000;97:11244-9. PMC17185.

39. Grobler JA, Stillmock K, Hu B, Witmer M, Felock P, Espeseth AS, Wolfe A, Egbertson M, Bourgeois M, Melamed J, Wai JS, Young S, Vacca J and Hazuda DJ. Diketo acid inhibitor mechanism and HIV-1 integrase: implications for metal binding in the active site of phosphotransferase enzymes. *Proc Natl Acad Sci U S A*. 2002;99:6661-6. PMC124459.

40. Hazuda DJ, Anthony NJ, Gomez RP, Jolly SM, Wai JS, Zhuang L, Fisher TE, Embrey M, Guare JP, Jr., Egbertson MS, Vacca JP, Huff JR, Felock PJ, Witmer MV, Stillmock KA, Danovich R, Grobler J, Miller MD, Espeseth AS, Jin L, Chen IW, Lin JH, Kassahun K, Ellis JD, Wong BK, Xu W, Pearson PG, Schleif WA, Cortese R, Emini E, Summa V, Holloway MK and Young SD. A naphthyridine carboxamide provides evidence for discordant resistance between mechanistically identical inhibitors of HIV-1 integrase. *Proc Natl Acad Sci U S A*. 2004;101:11233-8. PMC509174.

41. Matysiak J, Lesbats P, Mauro E, Lapaillerie D, Dupuy JW, Lopez AP, Benleulmi MS, Calmels C, Andreola ML, Ruff M, Llano M, Delelis O, Lavigne M and Parissi V. Modulation of chromatin structure by the FACT histone chaperone complex regulates HIV-1 integration. *Retrovirology*. 2017;14:39. PMC5534098.

42. Hare S and Cherepanov P. The Interaction Between Lentiviral Integrase and LEDGF:

Structural and Functional Insights. *Viruses*. 2009;1.

43. Fujiwara T and Mizuuchi K. Retroviral DNA integration: structure of an integration intermediate. *Cell*. 1988;54:497-504.

44. Baxter AE, Niessl J, Fromentin R, Richard J, Porichis F, Charlebois R, Massanella M, Brassard N, Alsahafi N, Delgado GG, Routy JP, Walker BD, Finzi A, Chomont N and Kaufmann DE. Single-Cell Characterization of Viral Translation-Competent Reservoirs in HIV-Infected Individuals. *Cell Host Microbe*. 2016;20:368-380. PMC5025389.

45. Cary DC, Fujinaga K and Peterlin BM. Molecular mechanisms of HIV latency. *The Journal of clinical investigation*. 2016;126:448-454.

46. Chavez L, Calvanese V and Verdin E. HIV Latency Is Established Directly and Early in Both Resting and Activated Primary CD4 T Cells. *PLoS Pathog*. 2015;11:e1004955. PMC4466167.

47. Mann DA, Mikaélian I, Zemmel RW, Green SM, Lowe AD, Kimura T, Singh M, Butler PJ, Gait MJ and Karn J. A molecular rheostat. Co-operative rev binding to stem I of the rev-response element modulates human immunodeficiency virus type-1 late gene expression. *Journal of molecular biology*. 1994;241:193-207.

48. Daly TJ, Cook KS, Gray GS, Maione TE and Rusche JR. Specific binding of HIV-1 recombinant Rev protein to the Rev-responsive element in vitro. *Nature*. 1989;342:816-9.

49. Heaphy S, Dingwall C, Ernberg I, Gait MJ, Green SM, Karn J, Lowe AD, Singh M and Skinner MA. HIV-1 regulator of virion expression (Rev) protein binds to an RNA stem-loop structure located within the Rev response element region. *Cell*. 1990;60:685-93.

50. Fischer U, Huber J, Boelens WC, Mattaj IW and Lührmann R. The HIV-1 Rev activation domain is a nuclear export signal that accesses an export pathway used by specific cellular RNAs. *Cell*. 1995;82:475-83.

51. Henderson BR and Percipalle P. Interactions between HIV Rev and nuclear import and export factors: the Rev nuclear localisation signal mediates specific binding to human importin-beta. *Journal of molecular biology*. 1997;274:693-707.

52. Sundquist WI and Kräusslich HG. HIV-1 assembly, budding, and maturation. *Cold Spring Harbor perspectives in medicine*. 2012;2:a006924. PMC3385941.

53. Dong X, Li H, Derdowski A, Ding L, Burnett A, Chen X, Peters TR, Dermody TS, Woodruff E, Wang JJ and Spearman P. AP-3 directs the intracellular trafficking of HIV-1 Gag and

plays a key role in particle assembly. *Cell*. 2005;120:663-74.

54. Martinez NW, Xue X, Berro RG, Kreitzer G and Resh MD. Kinesin KIF4 regulates intracellular trafficking and stability of the human immunodeficiency virus type 1 Gag polyprotein. *J Virol*. 2008;82:9937-50. PMC2566262.

55. Ono A and Freed EO. Plasma membrane rafts play a critical role in HIV-1 assembly and release. *Proc Natl Acad Sci U S A*. 2001;98:13925-30. PMC61143.

56. Jouvenet N, Simon SM and Bieniasz PD. Imaging the interaction of HIV-1 genomes and Gag during assembly of individual viral particles. *Proc Natl Acad Sci U S A*. 2009;106:19114-9. PMC2776408.

57. Chukkapalli V, Oh SJ and Ono A. Opposing mechanisms involving RNA and lipids regulate HIV-1 Gag membrane binding through the highly basic region of the matrix domain. *Proc Natl Acad Sci U S A*. 2010;107:1600-5. PMC2824378.

58. Datta SA, Heinrich F, Raghunandan S, Krueger S, Curtis JE, Rein A and Nanda H. HIV-1 Gag extension: conformational changes require simultaneous interaction with membrane and nucleic acid. *Journal of molecular biology*. 2011;406:205-14. PMC3046808.

59. Rein A, Henderson LE and Levin JG. Nucleic-acid-chaperone activity of retroviral nucleocapsid proteins: significance for viral replication. *Trends in biochemical sciences*. 1998;23:297-301.

60. Boyd PS, Brown JB, Brown JD, Catazaro J, Chaudry I, Ding P, Dong X, Marchant J, O'Hern CT, Singh K, Swanson C, Summers MF and Yasin S. NMR Studies of Retroviral Genome Packaging. *Viruses*. 2020;12. PMC7599994.

61. Keane SC, Heng X, Lu K, Kharytonchyk S, Ramakrishnan V, Carter G, Barton S, Hosic A, Florwick A, Santos J, Bolden NC, McCowin S, Case DA, Johnson BA, Salemi M, Telesnitsky A and Summers MF. RNA structure. Structure of the HIV-1 RNA packaging signal. *Science*. 2015;348:917-21. PMC4492308.

62. Cosson P. Direct interaction between the envelope and matrix proteins of HIV-1. *Embo j*. 1996;15:5783-8. PMC452325.

63. Wyma DJ, Kotov A and Aiken C. Evidence for a stable interaction of gp41 with Pr55(Gag) in immature human immunodeficiency virus type 1 particles. *J Virol*. 2000;74:9381-7. PMC112366.

64. Carlton JG and Martin-Serrano J. The ESCRT machinery: new functions in viral and

cellular biology. *Biochemical Society transactions*. 2009;37:195-9.

65. Hurley JH and Hanson PI. Membrane budding and scission by the ESCRT machinery: it's all in the neck. *Nature reviews Molecular cell biology*. 2010;11:556-66. PMC2922035.

66. Swanstrom R and Wills JW. Synthesis, Assembly, and Processing of Viral Proteins. In: J. M. Coffin, S. H. Hughes and H. E. Varmus, eds. *Retroviruses* Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press
Copyright © 1997, Cold Spring Harbor Laboratory Press.; 1997.

67. Hill M, Tachedjian G and Mak J. The packaging and maturation of the HIV-1 Pol proteins. *Current HIV research*. 2005;3:73-85.

68. Coggins SA, Holler JM, Kimata JT, Kim DH, Schinazi RF and Kim B. Efficient pre-catalytic conformational change of reverse transcriptases from SAMHD1 non-counteracting primate lentiviruses during dNTP incorporation. *Virology*. 2019;537:36-44. PMC6901729.

69. Human Immunodeficiency Virus (HIV). *Transfusion medicine and hemotherapy : offizielles Organ der Deutschen Gesellschaft fur Transfusionsmedizin und Immunhamatologie*. 2016;43:203-22. PMC4924471.

70. Ganser BK, Li S, Klishko VY, Finch JT and Sundquist WI. Assembly and analysis of conical models for the HIV-1 core. *Science*. 1999;283:80-3.

71. Briggs JA, Simon MN, Gross I, Kräusslich HG, Fuller SD, Vogt VM and Johnson MC. The stoichiometry of Gag protein in HIV-1. *Nature structural & molecular biology*. 2004;11:672-5.

72. Ganser-Pornillos BK, Cheng A and Yeager M. Structure of full-length HIV-1 CA: a model for the mature capsid lattice. *Cell*. 2007;131:70-9.

73. Shum K-T, Zhou J and Rossi JJ. Aptamer-Based Therapeutics: New Approaches to Combat Human Viral Diseases. *Pharmaceuticals*. 2013;6.

74. Roebuck KA and Saifuddin M. Regulation of HIV-1 transcription. *Gene expression*. 1999;8:67-84. PMC6157391.

75. Parkin NT, Chamorro M and Varmus HE. Human immunodeficiency virus type 1 gag-pol frameshifting is dependent on downstream mRNA secondary structure: demonstration by expression in vivo. *J Virol*. 1992;66:5147-51. PMC241392.

76. Frankel AD and Young JA. HIV-1: fifteen proteins and an RNA. *Annual review of biochemistry*. 1998;67:1-25.

77. Freed EO. HIV-1 replication. *Somatic cell and molecular genetics*. 2001;26:13-33.

78. Nkeze J, Li L, Benko Z, Li G and Zhao RY. Molecular characterization of HIV-1 genome in fission yeast Schizosaccharomyces pombe. *Cell & Bioscience*. 2015;5:47.

79. Gholizadeh Z, Iqbal MS, Li R and Romerio F. The HIV-1 Antisense Gene ASP: The New Kid on the Block. *Vaccines*. 2021;9. PMC8156140.

80. Tedbury PR and Freed EO. The role of matrix in HIV-1 envelope glycoprotein incorporation. *Trends in microbiology*. 2014;22:372-8. PMC4157688.

81. Tedbury PR and Freed EO. The cytoplasmic tail of retroviral envelope glycoproteins. *Progress in molecular biology and translational science*. 2015;129:253-84. PMC6941201.

82. Gamble TR, Yoo S, Vajdos FF, von Schwedler UK, Worthylake DK, Wang H, McCutcheon JP, Sundquist WI and Hill CP. Structure of the carboxyl-terminal dimerization domain of the HIV-1 capsid protein. *Science*. 1997;278:849-53.

83. Zhuang S and Torbett BE. Interactions of HIV-1 Capsid with Host Factors and Their Implications for Developing Novel Therapeutics. *Viruses*. 2021;13. PMC8001122.

84. Dick RA, Mallery DL, Vogt VM and James LC. IP6 Regulation of HIV Capsid Assembly, Stability, and Uncoating. *Viruses*. 2018;10. PMC6267275.

85. Mallery DL, Márquez CL, McEwan WA, Dickson CF, Jacques DA, Anandapadamanaban M, Bichel K, Towers GJ, Saiardi A, Böcking T and James LC. IP6 is an HIV pocket factor that prevents capsid collapse and promotes DNA synthesis. *eLife*. 2018;7. PMC6039178.

86. Selyutina A, Persaud M, Simons LM, Bulnes-Ramos A, Buffone C, Martinez-Lopez A, Scoca V, Di Nunzio F, Hiatt J, Marson A, Krogan NJ, Hultquist JF and Diaz-Griffero F. Cyclophilin A Prevents HIV-1 Restriction in Lymphocytes by Blocking Human TRIM5α Binding to the Viral Core. *Cell reports*. 2020;30:3766-3777.e6. PMC7363000.

87. Sayah DM, Sokolskaja E, Berthoux L and Luban J. Cyclophilin A retrotransposition into TRIM5 explains owl monkey resistance to HIV-1. *Nature*. 2004;430:569-73.

88. Kim K, Dauphin A, Komurlu S, McCauley SM, Yurkovetskiy L, Carbone C, Diehl WE, Strambio-De-Castillia C, Campbell EM and Luban J. Cyclophilin A protects HIV-1 from restriction by human TRIM5α. *Nature microbiology*. 2019;4:2044-2051. PMC6879858.

89. Ganser-Pornillos BK and Pornillos O. Restriction of HIV-1 and other retroviruses by TRIM5. *Nature reviews Microbiology*. 2019;17:546-556. PMC6858284.

90. Rankovic S, Ramalho R, Aiken C and Rousso I. PF74 Reinforces the HIV-1 Capsid To

Impair Reverse Transcription-Induced Uncoating. *J Virol*. 2018;92. PMC6158434.

91. Sahani RL, Diana-Rivero R, Vernekar SKV, Wang L, Du H, Zhang H, Castaner AE, Casey MC, Kirby KA, Tedbury PR, Xie J, Sarafianos SG and Wang Z. Design, Synthesis and Characterization of HIV-1 CA-Targeting Small Molecules: Conformational Restriction of PF74. *Viruses*. 2021;13. PMC8000227.

92. Wang L, Casey MC, Vernekar SKV, Sahani RL, Kankanala J, Kirby KA, Du H, Hachiya A, Zhang H, Tedbury PR, Xie J, Sarafianos SG and Wang Z. Novel HIV-1 capsid-targeting small molecules of the PF74 binding site. *European journal of medicinal chemistry*. 2020;204:112626. PMC7530112.

93. Scoca V and Di Nunzio F. The HIV-1 Capsid: From Structural Component to Key Factor for Host Nuclear Invasion. *Viruses*. 2021;13. PMC7916756.

94. Francis AC, Marin M, Singh PK, Achuthan V, Prellberg MJ, Palermino-Rowland K, Lan S, Tedbury PR, Sarafianos SG, Engelman AN and Melikyan GB. HIV-1 replication complexes accumulate in nuclear speckles and integrate into speckle-associated genomic domains. *Nat Commun*. 2020;11:3505. PMC7360574.

95. Paillart JC, Skripkin E, Ehresmann B, Ehresmann C and Marquet R. A loop-loop "kissing" complex is the essential part of the dimer linkage of genomic HIV-1 RNA. *Proc Natl Acad Sci U S A*. 1996;93:5572-7. PMC39288.

96. Huang M, Orenstein JM, Martin MA and Freed EO. p6Gag is required for particle production from full-length human immunodeficiency virus type 1 molecular clones expressing protease. *J Virol*. 1995;69:6810-8. PMC189593.

97. Lightfoote MM, Coligan JE, Folks TM, Fauci AS, Martin MA and Venkatesan S. Structural characterization of reverse transcriptase and endonuclease polypeptides of the acquired immunodeficiency syndrome retrovirus. *J Virol*. 1986;60:771-5. PMC288955.

98. Jacobo-Molina A, Ding J, Nanni RG, Clark AD, Jr., Lu X, Tantillo C, Williams RL, Kamer G, Ferris AL, Clark P and et al. Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 A resolution shows bent DNA. *Proc Natl Acad Sci U S A*. 1993;90:6320-4. PMC46920.

99. Oude Essink BB, Das AT and Berkhout B. HIV-1 reverse transcriptase discriminates against non-self tRNA primers. *Journal of molecular biology*. 1996;264:243-54.

100. Tuske S, Sarafianos SG, Clark AD, Jr., Ding J, Naeger LK, White KL, Miller MD, Gibbs

CS, Boyer PL, Clark P, Wang G, Gaffney BL, Jones RA, Jerina DM, Hughes SH and Arnold E. Structures of HIV-1 RT-DNA complexes before and after incorporation of the anti-AIDS drug tenofovir. *Nature structural & molecular biology*. 2004;11:469-74.

101. Sarafianos SG, Clark AD, Jr., Tuske S, Squire CJ, Das K, Sheng D, Ilankumaran P, Ramesha AR, Kroth H, Sayer JM, Jerina DM, Boyer PL, Hughes SH and Arnold E. Trapping HIV-1 reverse transcriptase before and after translocation on DNA. *The Journal of biological chemistry*. 2003;278:16280-8.

102. Julias JG, McWilliams MJ, Sarafianos SG, Arnold E and Hughes SH. Mutations in the RNase H domain of HIV-1 reverse transcriptase affect the initiation of DNA synthesis and the specificity of RNase H cleavage in vivo. *Proc Natl Acad Sci U S A*. 2002;99:9515-20. PMC123172.

103. Hu WS and Temin HM. Retroviral recombination and reverse transcription. *Science*. 1990;250:1227-33.

104. van Wamel JL and Berkhout B. The first strand transfer during HIV-1 reverse transcription can occur either intramolecularly or intermolecularly. *Virology*. 1998;244:245-51.

105. Xu C, Fischer DK, Rankovic S, Li W, Dick RA, Runge B, Zadorozhnyi R, Ahn J, Aiken C, Polenova T, Engelman AN, Ambrose Z, Rousso I and Perilla JR. Permeability of the HIV-1 capsid to metabolites modulates viral DNA synthesis. *PLoS biology*. 2020;18:e3001015. PMC7775124.

106. Perilla JR, Hadden-Perilla JA, Gronenborn AM and Polenova T. Integrative structural biology of HIV-1 capsid protein assemblies: combining experiment and computation. *Curr Opin Virol*. 2021;48:57-64. PMC8187302.

107. Hickman AB, Palmer I, Engelman A, Craigie R and Wingfield P. Biophysical and enzymatic properties of the catalytic domain of HIV-1 integrase. *The Journal of biological chemistry*. 1994;269:29279-87.

108. Passos DO, Li M, Yang R, Rebensburg SV, Ghirlando R, Jeon Y, Shkriabai N, Kvaratskhelia M, Craigie R and Lyumkis D. Cryo-EM structures and atomic model of the HIV-1 strand transfer complex intasome. *Science*. 2017;355:89-92. PMC5508583.

109. Park J and Morrow CD. Mutations in the protease gene of human immunodeficiency virus type 1 affect release and stability of virus particles. *Virology*. 1993;194:843-50.

110. Checkley MA, Luttge BG and Freed EO. HIV-1 envelope glycoprotein biosynthesis,

trafficking, and incorporation. *Journal of molecular biology*. 2011;410:582-608. PMC3139147.

111. Clapham PR and Weiss RA. Spoilt for choice of co-receptors. *Nature*. 1997;388:230-231.

112. Chen B and Chou JJ. Structure of the transmembrane domain of HIV-1 envelope glycoprotein. *The FEBS journal*. 2017;284:1171-1177. PMC5448286.

113. Jones KA. Tat and the HIV-1 promoter. *Current opinion in cell biology*. 1993;5:461-8.

114. Razooky BS, Pai A, Aull K, Rouzine IM and Weinberger LS. A hardwired HIV latency program. *Cell*. 2015;160:990-1001. PMC4395878.

115. Roebuck KA, Rabbi MF and Kagnoff MF. HIV-1 Tat protein can transactivate a heterologous TATAA element independent of viral promoter sequences and the trans-activation response element. *AIDS (London, England)*. 1997;11:139-46.

116. Meyer BE and Malim MH. The HIV-1 Rev trans-activator shuttles between the nucleus and the cytoplasm. *Genes & development*. 1994;8:1538-47.

117. Bogerd HP, Fridell RA, Madore S and Cullen BR. Identification of a novel cellular cofactor for the Rev/Rex class of retroviral regulatory proteins. *Cell*. 1995;82:485-94.

118. Kjems J and Sharp PA. The basic domain of Rev from human immunodeficiency virus type 1 specifically blocks the entry of U4/U6.U5 small nuclear ribonucleoprotein in spliceosome assembly. *J Virol*. 1993;67:4769-76. PMC237863.

119. Dupont L, Bloor S, Williamson JC, Cuesta SM, Shah R, Teixeira-Silva A, Naamati A, Greenwood EJD, Sarafianos SG, Matheson NJ and Lehner PJ. The SMC5/6 complex compacts and silences unintegrated HIV-1 DNA and is antagonized by Vpr. *Cell Host Microbe*. 2021.

120. Agostini I, Navarro JM, Rey F, Bouhamdan M, Spire B, Vigne R and Sire J. The human immunodeficiency virus type 1 Vpr transactivator: cooperation with promoter-bound activator domains and binding to TFIIB. *Journal of molecular biology*. 1996;261:599-606.

121. Wang L, Mukherjee S, Jia F, Narayan O and Zhao LJ. Interaction of virion protein Vpr of human immunodeficiency virus type 1 with cellular transcription factor Sp1 and trans-activation of viral long terminal repeat. *The Journal of biological chemistry*. 1995;270:25564-9.

122. Felzien LK, Woffendin C, Hottiger MO, Subbramanian RA, Cohen EA and Nabel GJ. HIV transcriptional activation by the accessory protein, VPR, is mediated by the p300 co-activator. *Proc Natl Acad Sci U S A*. 1998;95:5281-6. PMC20252.

123. Zhang F and Bieniasz PD. HIV-1 Vpr induces cell cycle arrest and enhances viral gene expression by depleting CCDC137. *eLife*. 2020;9. PMC7295576.

124. Hrecka K, Hao C, Shun MC, Kaur S, Swanson SK, Florens L, Washburn MP and Skowronski J. HIV-1 and HIV-2 exhibit divergent interactions with HLTF and UNG2 DNA repair proteins. *Proc Natl Acad Sci U S A*. 2016;113:E3921-30. PMC4941427.

125. Hrecka K, Gierszewska M, Srivastava S, Kozaczkiewicz L, Swanson SK, Florens L, Washburn MP and Skowronski J. Lentiviral Vpr usurps Cul4-DDB1[VprBP] E3 ubiquitin ligase to modulate cell cycle. *Proc Natl Acad Sci U S A*. 2007;104:11778-83. PMC1906728.

126. Sakai K, Dimas J and Lenardo MJ. The Vif and Vpr accessory proteins independently cause HIV-1-induced T cell cytopathicity and cell cycle arrest. *Proc Natl Acad Sci U S A*. 2006;103:3369-74. PMC1413893.

127. Heinzinger NK, Bukrinsky MI, Haggerty SA, Ragland AM, Kewalramani V, Lee MA, Gendelman HE, Ratner L, Stevenson M and Emerman M. The Vpr protein of human immunodeficiency virus type 1 influences nuclear localization of viral nucleic acids in nondividing host cells. *Proc Natl Acad Sci U S A*. 1994;91:7311-5. PMC44389.

128. Lamb RA and Pinto LH. Do Vpu and Vpr of human immunodeficiency virus type 1 and NB of influenza B virus have ion channel activities in the viral life cycles? *Virology*. 1997;229:1-11.

129. Cohen EA, Subbramanian RA and Göttlinger HG. Role of auxiliary proteins in retroviral morphogenesis. *Current topics in microbiology and immunology*. 1996;214:219-35.

130. Papkalla A, Münch J, Otto C and Kirchhoff F. Nef enhances human immunodeficiency virus type 1 infectivity and replication independently of viral coreceptor tropism. *J Virol*. 2002;76:8455-9. PMC155138.

131. Bukovsky AA, Dorfman T, Weimann A and Göttlinger HG. Nef association with human immunodeficiency virus type 1 virions and cleavage by the viral protease. *J Virol*. 1997;71:1013-8. PMC191151.

132. Rosa A, Chande A, Ziglio S, De Sanctis V, Bertorelli R, Goh SL, McCauley SM, Nowosielska A, Antonarakis SE, Luban J, Santoni FA and Pizzato M. HIV-1 Nef promotes infection by excluding SERINC5 from virion incorporation. *Nature*. 2015;526:212-217.

133. Malim MH. APOBEC proteins and intrinsic resistance to HIV-1 infection. *Philosophical transactions of the Royal Society of London Series B, Biological sciences*. 2009;364:675-87. PMC2660912.

134. Pollpeter D, Parsons M, Sobala AE, Coxhead S, Lang RD, Bruns AM, Papaioannou S,

McDonnell JM, Apolonia L, Chowdhury JA, Horvath CM and Malim MH. Deep sequencing of HIV-1 reverse transcripts reveals the multifaceted antiviral functions of APOBEC3G. *Nature microbiology*. 2018;3:220-233. PMC6014619.

135. Naamati A, Williamson JC, Greenwood EJ, Marelli S, Lehner PJ and Matheson NJ. Functional proteomic atlas of HIV infection in primary human CD4+ T cells. *eLife*. 2019;8. PMC6414203.

136. Salamango DJ, Ikeda T, Moghadasi SA, Wang J, McCann JL, Serebrenik AA, Ebrahimi D, Jarvis MC, Brown WL and Harris RS. HIV-1 Vif Triggers Cell Cycle Arrest by Degrading Cellular PPP2R5 Phospho-regulators. *Cell reports*. 2019;29:1057-1065.e4. PMC6903395.

137. Fischl MA, Richman DD, Grieco MH, Gottlieb MS, Volberding PA, Laskin OL, Leedom JM, Groopman JE, Mildvan D, Schooley RT and et al. The efficacy of azidothymidine (AZT) in the treatment of patients with AIDS and AIDS-related complex. A double-blind, placebo-controlled trial. *The New England journal of medicine*. 1987;317:185-91.

138. Mitsuya H, Weinhold KJ, Furman PA, St Clair MH, Lehrman SN, Gallo RC, Bolognesi D, Barry DW and Broder S. 3'-Azido-3'-deoxythymidine (BW A509U): an antiviral agent that inhibits the infectivity and cytopathic effect of human T-lymphotropic virus type III/lymphadenopathy-associated virus in vitro. *Proc Natl Acad Sci U S A*. 1985;82:7096-100. PMC391317.

139. Furman PA, Fyfe JA, St Clair MH, Weinhold K, Rideout JL, Freeman GA, Lehrman SN, Bolognesi DP, Broder S, Mitsuya H and et al. Phosphorylation of 3'-azido-3'-deoxythymidine and selective interaction of the 5'-triphosphate with human immunodeficiency virus reverse transcriptase. *Proc Natl Acad Sci U S A*. 1986;83:8333-7. PMC386922.

140. Balzarini J, Herdewijn P and De Clercq E. Differential patterns of intracellular metabolism of 2',3'-didehydro-2',3'-dideoxythymidine and 3'-azido-2',3'-dideoxythymidine, two potent anti-human immunodeficiency virus compounds. *The Journal of biological chemistry*. 1989;264:6127-33.

141. Arts EJ and Hazuda DJ. HIV-1 antiretroviral drug therapy. *Cold Spring Harbor perspectives in medicine*. 2012;2:a007161. PMC3312400.

142. Cilento ME, Kirby KA and Sarafianos SG. Avoiding Drug Resistance in HIV Reverse Transcriptase. *Chemical reviews*. 2021;121:3271-3296. PMC8149104.

143. Kawamoto A, Kodama E, Sarafianos SG, Sakagami Y, Kohgo S, Kitano K, Ashida N, Iwai Y, Hayakawa H, Nakata H, Mitsuya H, Arnold E and Matsuoka M. 2′-Deoxy-4′-C-ethynyl-2-halo-

adenosines active against drug-resistant human immunodeficiency virus type 1 variants. *The International Journal of Biochemistry & Cell Biology*. 2008;40:2410-2420.

144. Michailidis E, Marchand B, Kodama EN, Singh K, Matsuoka M, Kirby KA, Ryan EM, Sawani AM, Nagy E, Ashida N, Mitsuya H, Parniak MA and Sarafianos SG. Mechanism of inhibition of HIV-1 reverse transcriptase by 4'-Ethynyl-2-fluoro-2'-deoxyadenosine triphosphate, a translocation-defective reverse transcriptase inhibitor. *The Journal of biological chemistry*. 2009;284:35681-91. PMC2790999.

145. Michailidis E, Huber AD, Ryan EM, Ong YT, Leslie MD, Matzek KB, Singh K, Marchand B, Hagedorn AN, Kirby KA, Rohan LC, Kodama EN, Mitsuya H, Parniak MA and Sarafianos SG. 4'-Ethynyl-2-fluoro-2'-deoxyadenosine (EFdA) inhibits HIV-1 reverse transcriptase with multiple mechanisms. *The Journal of biological chemistry*. 2014;289:24533-48. PMC4148878.

146. Kirby KA, Michailidis E, Fetterly TL, Steinbach MA, Singh K, Marchand B, Leslie MD, Hagedorn AN, Kodama EN, Marquez VE, Hughes SH, Mitsuya H, Parniak MA and Sarafianos SG. Effects of substitutions at the 4' and 2 positions on the bioactivity of 4'-ethynyl-2-fluoro-2'-deoxyadenosine. *Antimicrobial agents and chemotherapy*. 2013;57:6254-64. PMC3837839.

147. Salie ZL, Kirby KA, Michailidis E, Marchand B, Singh K, Rohan LC, Kodama EN, Mitsuya H, Parniak MA and Sarafianos SG. Structural basis of HIV inhibition by translocation-defective RT inhibitor 4′-ethynyl-2-fluoro-2′-deoxyadenosine (EFdA). *Proceedings of the National Academy of Sciences*. 2016;113:9274.

148. Shimura K, Kodama E, Sakagami Y, Matsuzaki Y, Watanabe W, Yamataka K, Watanabe Y, Ohata Y, Doi S, Sato M, Kano M, Ikeda S and Matsuoka M. Broad antiretroviral activity and resistance profile of the novel human immunodeficiency virus integrase inhibitor elvitegravir (JTK-303/GS-9137). *J Virol*. 2008;82:764-74. PMC2224569.

149. Demarest J, Underwood M, St Clair M, Dorey D, Brown D and Zolopa A. Short Communication: Dolutegravir-Based Regimens Are Active in Integrase Strand Transfer Inhibitor-Naive Patients with Nucleoside Reverse Transcriptase Inhibitor Resistance. *AIDS research and human retroviruses*. 2018;34:343-346. PMC5899294.

150. Burlein C, Wang C, Xu M, Bhatt T, Stahlhut M, Ou Y, Adam GC, Heath J, Klein DJ, Sanders J, Narayan K, Abeywickrema P, Heo MR, Carroll SS, Grobler JA, Sharma S, Diamond TL, Converso A and Krosky DJ. Discovery of a Distinct Chemical and Mechanistic Class of Allosteric HIV-1 Integrase Inhibitors with Antiretroviral Activity. *ACS chemical biology*.

2017;12:2858-2865.

151. Christ F, Shaw S, Demeulemeester J, Desimmie BA, Marchand A, Butler S, Smets W, Chaltin P, Westby M, Debyser Z and Pickford C. Small-molecule inhibitors of the LEDGF/p75 binding site of integrase block HIV replication and modulate integrase multimerization. *Antimicrobial agents and chemotherapy*. 2012;56:4365-74. PMC3421592.

152. Fenwick C, Amad M, Bailey MD, Bethell R, Bös M, Bonneau P, Cordingley M, Coulombe R, Duan J, Edwards P, Fader LD, Faucher AM, Garneau M, Jakalian A, Kawai S, Lamorte L, LaPlante S, Luo L, Mason S, Poupart MA, Rioux N, Schroeder P, Simoneau B, Tremblay S, Tsantrizos Y, Witvrouw M and Yoakim C. Preclinical profile of BI 224436, a novel HIV-1 non-catalytic-site integrase inhibitor. *Antimicrobial agents and chemotherapy*. 2014;58:3233-44. PMC4068430.

153. Choi E, Mallareddy JR, Lu D and Kolluru S. Recent advances in the discovery of small-molecule inhibitors of HIV-1 integrase. *Future science OA*. 2018;4:Fso338. PMC6222271.

154. Shafer RW, Dupnik K, Winters MA and Eshleman SH. A Guide to HIV-1 Reverse Transcriptase and Protease Sequencing for Drug Resistance Studies. *HIV sequence compendium*. 2001;2001:1-51. PMC3274565.

155. Bester SM, Wei G, Zhao H, Adu-Ampratwum D, Iqbal N, Courouble VV, Francis AC, Annamalai AS, Singh PK, Shkriabai N, Van Blerkom P, Morrison J, Poeschla EM, Engelman AN, Melikyan GB, Griffin PR, Fuchs JR, Asturias FJ and Kvaratskhelia M. Structural and mechanistic bases for a potent HIV-1 capsid inhibitor. *Science*. 2020;370:360-364. PMC7831379.

156. Haynes BF, Gilbert PB, McElrath MJ, Zolla-Pazner S, Tomaras GD, Alam SM, Evans DT, Montefiori DC, Karnasuta C, Sutthent R, Liao HX, DeVico AL, Lewis GK, Williams C, Pinter A, Fong Y, Janes H, DeCamp A, Huang Y, Rao M, Billings E, Karasavvas N, Robb ML, Ngauy V, de Souza MS, Paris R, Ferrari G, Bailer RT, Soderberg KA, Andrews C, Berman PW, Frahm N, De Rosa SC, Alpert MD, Yates NL, Shen X, Koup RA, Pitisuttithum P, Kaewkungwal J, Nitayaphan S, Rerks-Ngarm S, Michael NL and Kim JH. Immune-correlates analysis of an HIV-1 vaccine efficacy trial. *The New England journal of medicine*. 2012;366:1275-86. PMC3371689.

157. Passos DO, Li M, Jóźwik IK, Zhao XZ, Santos-Martins D, Yang R, Smith SJ, Jeon Y, Forli S, Hughes SH, Burke TR, Jr., Craigie R and Lyumkis D. Structural basis for strand-transfer inhibitor binding to HIV intasomes. *Science*. 2020;367:810-814. PMC7357238.

158. Achuthan V, Perreira JM, Sowd GA, Puray-Chavez M, McDougall WM, Paulucci-

Holthauzen A, Wu X, Fadel HJ, Poeschla EM, Multani AS, Hughes SH, Sarafianos SG, Brass AL and Engelman AN. Capsid-CPSF6 Interaction Licenses Nuclear HIV-1 Trafficking to Sites of Viral DNA Integration. *Cell Host Microbe*. 2018;24:392-404.e8. PMC6368089.

159. Achuthan V, Perreira JM, Ahn JJ, Brass AL and Engelman AN. Capsid-CPSF6 interaction: Master regulator of nuclear HIV-1 positioning and integration. *J Life Sci (Westlake Village)*. 2019;1:39-45. PMC6707730.

160. Ciuffi A, Llano M, Poeschla E, Hoffmann C, Leipzig J, Shinn P, Ecker JR and Bushman F. A role for LEDGF/p75 in targeting HIV DNA integration. *Nature medicine*. 2005;11:1287-9.

161. Poeschla EM. Integrase, LEDGF/p75 and HIV replication. *Cell Mol Life Sci*. 2008;65:1403-24. PMC3902792.

162. Cherepanov P, Maertens G, Proost P, Devreese B, Van Beeumen J, Engelborghs Y, De Clercq E and Debyser Z. HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells. *The Journal of biological chemistry*. 2003;278:372-81.

163. Emiliani S, Mousnier A, Busschots K, Maroun M, Van Maele B, Tempé D, Vandekerckhove L, Moisant F, Ben-Slama L, Witvrouw M, Christ F, Rain JC, Dargemont C, Debyser Z and Benarous R. Integrase mutants defective for interaction with LEDGF/p75 are impaired in chromosome tethering and HIV-1 replication. *The Journal of biological chemistry*. 2005;280:25517-23.

164. Huang H, Santoso N, Power D, Simpson S, Dieringer M, Miao H, Gurova K, Giam CZ, Elledge SJ and Zhu J. FACT Proteins, SUPT16H and SSRP1, Are Transcriptional Suppressors of HIV-1 and HTLV-1 That Facilitate Viral Latency. *The Journal of biological chemistry*. 2015;290:27297-27310. PMC4646377.

165. Liu Y, Zhou K, Zhang N, Wei H, Tan YZ, Zhang Z, Carragher B, Potter CS, D'Arcy S and Luger K. FACT caught in the act of manipulating the nucleosome. *Nature*. 2020;577:426-431.

166. Gurova K, Chang HW, Valieva ME, Sandlesh P and Studitsky VM. Structure and function of the histone chaperone FACT - Resolving FACTual issues. *Biochimica et biophysica acta Gene regulatory mechanisms*. 2018. PMC6349528.

167. Schröder AR, Shinn P, Chen H, Berry C, Ecker JR and Bushman F. HIV-1 integration in the human genome favors active genes and local hotspots. *Cell*. 2002;110:521-9.

168. Wang GP, Ciuffi A, Leipzig J, Berry CC and Bushman FD. HIV integration site selection: analysis by massively parallel pyrosequencing reveals association with epigenetic modifications.

Genome Res. 2007;17:1186-94. PMC1933515.

169. Marini B, Kertesz-Farkas A, Ali H, Lucic B, Lisek K, Manganaro L, Pongor S, Luzzati R, Recchia A, Mavilio F, Giacca M and Lusic M. Nuclear architecture dictates HIV-1 integration site selection. *Nature*. 2015;521:227-31.

170. Lucic B, Chen HC, Kuzman M, Zorita E, Wegner J, Minneker V, Wang W, Fronza R, Laufs S, Schmidt M, Stadhouders R, Roukos V, Vlahovicek K, Filion GJ and Lusic M. Spatially clustered loci with multiple enhancers are frequent targets of HIV-1 integration. *Nat Commun*. 2019;10:4059. PMC6731298.

171. Dieudonne M, Maiuri P, Biancotto C, Knezevich A, Kula A, Lusic M and Marcello A. Transcriptional competence of the integrated HIV-1 provirus at the nuclear periphery. *EMBO J*. 2009;28:2231-43. PMC2726691.

172. Chin CR, Perreira JM, Savidis G, Portmann JM, Aker AM, Feeley EM, Smith MC and Brass AL. Direct Visualization of HIV-1 Replication Intermediates Shows that Capsid and CPSF6 Modulate HIV-1 Intra-nuclear Invasion and Integration. *Cell reports*. 2015;13:1717-31. PMC5026322.

173. Burdick RC, Li C, Munshi M, Rawson JMO, Nagashima K, Hu WS and Pathak VK. HIV-1 uncoats in the nucleus near sites of integration. *Proc Natl Acad Sci U S A*. 2020;117:5486-5493. PMC7071919.

174. Stultz RD, Cenker JJ and McDonald D. Imaging HIV-1 Genomic DNA from Entry through Productive Infection. *J Virol*. 2017;91. PMC5391475.

175. Tsopoulidis N, Kaw S, Laketa V, Kutscheidt S, Baarlink C, Stolp B, Grosse R and Fackler OT. T cell receptor-triggered nuclear actin network formation drives CD4(+) T cell effector functions. *Science immunology*. 2019;4.

176. Brady T, Agosto LM, Malani N, Berry CC, O'Doherty U and Bushman F. HIV integration site distributions in resting and activated CD4+ T cells infected in culture. *AIDS (London, England)*. 2009;23:1461-71. PMC2862484.

177. Sowd GA, Serrao E, Wang H, Wang W, Fadel HJ, Poeschla EM and Engelman AN. A critical role for alternative polyadenylation factor CPSF6 in targeting HIV-1 integration to transcriptionally active chromatin. *Proc Natl Acad Sci U S A*. 2016;113:E1054-63. PMC4776470.

178. Singh PK, Plumb MR, Ferris AL, Iben JR, Wu X, Fadel HJ, Luke BT, Esnault C, Poeschla EM, Hughes SH, Kvaratskhelia M and Levin HL. LEDGF/p75 interacts with mRNA splicing

factors and targets HIV-1 integration to highly spliced genes. *Genes & development*. 2015;29:2287-97. PMC4647561.

179. Ocwieja KE, Brady TL, Ronen K, Huegel A, Roth SL, Schaller T, James LC, Towers GJ, Young JA, Chanda SK, König R, Malani N, Berry CC and Bushman FD. HIV integration targeting: a pathway involving Transportin-3 and the nuclear pore protein RanBP2. *PLoS Pathog*. 2011;7:e1001313. PMC3053352.

180. Li W, Singh PK, Sowd GA, Bedwell GJ, Jang S, Achuthan V, Oleru AV, Wong D, Fadel HJ, Lee K, KewalRamani VN, Poeschla EM, Herschhorn A and Engelman AN. CPSF6-Dependent Targeting of Speckle-Associated Domains Distinguishes Primate from Nonprimate Lentiviral Integration. *mBio*. 2020;11. PMC7527728.

181. Vansant G, Chen HC, Zorita E, Trejbalová K, Miklík D, Filion G and Debyser Z. The chromatin landscape at the HIV-1 provirus integration site determines viral expression. *Nucleic Acids Res*. 2020;48:7801-7817. PMC7641320.

182. de Mendoza C. Hot News: HIV Epidemics - Current Burden and Future Prospects. *AIDS reviews*. 2017;19:239.

183. Deleage C, Wietgrefe SW, Del Prete G, Morcock DR, Hao XP, Piatak M, Jr., Bess J, Anderson JL, Perkey KE, Reilly C, McCune JM, Haase AT, Lifson JD, Schacker TW and Estes JD. Defining HIV and SIV Reservoirs in Lymphoid Tissues. *Pathogens & immunity*. 2016;1:68-106. PMC4943335.

184. Khoury G, Darcis G, Lee MY, Bouchat S, Van Driessche B, Purcell DFJ and Van Lint C. The Molecular Biology of HIV Latency. *Advances in experimental medicine and biology*. 2018;1075:187-212.

185. Jordan A, Bisgrove D and Verdin E. HIV reproducibly establishes a latent infection after acute infection of T cells in vitro. *Embo j*. 2003;22:1868-77. PMC154479.

186. Perelson AS, Essunger P, Cao Y, Vesanen M, Hurley A, Saksela K, Markowitz M and Ho DD. Decay characteristics of HIV-1-infected compartments during combination therapy. *Nature*. 1997;387:188-91.

187. Gulick RM, Mellors JW, Havlir D, Eron JJ, Gonzalez C, McMahon D, Richman DD, Valentine FT, Jonas L, Meibohm A, Emini EA and Chodakewitz JA. Treatment with indinavir, zidovudine, and lamivudine in adults with human immunodeficiency virus infection and prior antiretroviral therapy. *The New England journal of medicine*. 1997;337:734-9.

188. Siliciano JD, Kajdas J, Finzi D, Quinn TC, Chadwick K, Margolick JB, Kovacs C, Gange SJ and Siliciano RF. Long-term follow-up studies confirm the stability of the latent reservoir for HIV-1 in resting CD4+ T cells. *Nature medicine*. 2003;9:727-8.

189. Rouzioux C and Richman D. How to best measure HIV reservoirs? *Curr Opin HIV AIDS*. 2013;8:170-5. PMC3763804.

190. Finzi D, Hermankova M, Pierson T, Carruth LM, Buck C, Chaisson RE, Quinn TC, Chadwick K, Margolick J, Brookmeyer R, Gallant J, Markowitz M, Ho DD, Richman DD and Siliciano RF. Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy. *Science*. 1997;278:1295-300.

191. Strain MC, Günthard HF, Havlir DV, Ignacio CC, Smith DM, Leigh-Brown AJ, Macaranas TR, Lam RY, Daly OA, Fischer M, Opravil M, Levine H, Bacheler L, Spina CA, Richman DD and Wong JK. Heterogeneous clearance rates of long-lived lymphocytes infected with HIV: intrinsic stability predicts lifelong persistence. *Proc Natl Acad Sci U S A*. 2003;100:4819-24. PMC153639.

192. Ho YC, Shan L, Hosmane NN, Wang J, Laskey SB, Rosenbloom DI, Lai J, Blankson JN, Siliciano JD and Siliciano RF. Replication-competent noninduced proviruses in the latent reservoir increase barrier to HIV-1 cure. *Cell*. 2013;155:540-51. PMC3896327.

193. Bruner KM, Murray AJ, Pollack RA, Soliman MG, Laskey SB, Capoferri AA, Lai J, Strain MC, Lada SM, Hoh R, Ho YC, Richman DD, Deeks SG, Siliciano JD and Siliciano RF. Defective proviruses rapidly accumulate during acute HIV-1 infection. *Nature medicine*. 2016;22:1043-9. PMC5014606.

194. Imamichi H, Dewar RL, Adelsberger JW, Rehm CA, O'Doherty U, Paxinos EE, Fauci AS and Lane HC. Defective HIV-1 proviruses produce novel protein-coding RNA species in HIV-infected patients on combination antiretroviral therapy. *Proc Natl Acad Sci U S A*. 2016;113:8783-8. PMC4978246.

195. Siliciano JD and Siliciano RF. Assays to Measure Latency, Reservoirs, and Reactivation. *Current topics in microbiology and immunology*. 2018;417:23-41.

196. Cohn LB, Silva IT, Oliveira TY, Rosales RA, Parrish EH, Learn GH, Hahn BH, Czartoski JL, McElrath MJ, Lehmann C, Klein F, Caskey M, Walker BD, Siliciano JD, Siliciano RF, Jankovic M and Nussenzweig MC. HIV-1 integration landscape during latent and active infection. *Cell*. 2015;160:420-32. PMC4371550.

197. Pollack RA, Jones RB, Pertea M, Bruner KM, Martin AR, Thomas AS, Capoferri AA, Beg SA, Huang SH, Karandish S, Hao H, Halper-Stromberg E, Yong PC, Kovacs C, Benko E, Siliciano RF and Ho YC. Defective HIV-1 Proviruses Are Expressed and Can Be Recognized by Cytotoxic T Lymphocytes, which Shape the Proviral Landscape. *Cell Host Microbe*. 2017;21:494-506.e4. PMC5433942.

198. Chun TW, Carruth L, Finzi D, Shen X, DiGiuseppe JA, Taylor H, Hermankova M, Chadwick K, Margolick J, Quinn TC, Kuo YH, Brookmeyer R, Zeiger MA, Barditch-Crovo P and Siliciano RF. Quantification of latent tissue reservoirs and total body viral load in HIV-1 infection. *Nature*. 1997;387:183-8.

199. Wagner TA, McLaughlin S, Garg K, Cheung CY, Larsen BB, Styrchak S, Huang HC, Edlefsen PT, Mullins JI and Frenkel LM. HIV latency. Proliferation of cells with HIV integrated into cancer genes contributes to persistent infection. *Science*. 2014;345:570-3. PMC4230336.

200. Maldarelli F, Wu X, Su L, Simonetti FR, Shao W, Hill S, Spindler J, Ferris AL, Mellors JW, Kearney MF, Coffin JM and Hughes SH. HIV latency. Specific HIV integration sites are linked to clonal expansion and persistence of infected cells. *Science*. 2014;345:179-83. PMC4262401.

201. Hosmane NN, Kwon KJ, Bruner KM, Capoferri AA, Beg S, Rosenbloom DI, Keele BF, Ho YC, Siliciano JD and Siliciano RF. Proliferation of latently infected CD4(+) T cells carrying replication-competent HIV-1: Potential role in latent reservoir dynamics. *The Journal of experimental medicine*. 2017;214:959-972. PMC5379987.

202. Bui JK, Sobolewski MD, Keele BF, Spindler J, Musick A, Wiegand A, Luke BT, Shao W, Hughes SH, Coffin JM, Kearney MF and Mellors JW. Proviruses with identical sequences comprise a large fraction of the replication-competent HIV reservoir. *PLoS Pathog*. 2017;13:e1006283. PMC5378418 WS, BTL, and BFK are employees of Leidos Biomedical Research, Inc. No other authors report competing financial interests.

203. Lorenzi JC, Cohen YZ, Cohn LB, Kreider EF, Barton JP, Learn GH, Oliveira T, Lavine CL, Horwitz JA, Settler A, Jankovic M, Seaman MS, Chakraborty AK, Hahn BH, Caskey M and Nussenzweig MC. Paired quantitative and qualitative assessment of the replication-competent HIV-1 reservoir and comparison with integrated proviral DNA. *Proc Natl Acad Sci U S A*. 2016;113:E7908-e7916. PMC5150408.

204. Pace MJ, Graf EH, Agosto LM, Mexas AM, Male F, Brady T, Bushman FD and O'Doherty

U. Directly infected resting CD4+T cells can produce HIV Gag without spreading infection in a model of HIV latency. *PLoS Pathog*. 2012;8:e1002818. PMC3406090 (patents, consultation, products in development etc). This does not alter our adherence to all PLoS Pathogens policies on sharing data and materials.

205. Shan L, Deng K, Gao H, Xing S, Capoferri AA, Durand CM, Rabi SA, Laird GM, Kim M, Hosmane NN, Yang HC, Zhang H, Margolick JB, Li L, Cai W, Ke R, Flavell RA, Siliciano JD and Siliciano RF. Transcriptional Reprogramming during Effector-to-Memory Transition Renders CD4(+) T Cells Permissive for Latent HIV-1 Infection. *Immunity*. 2017;47:766-775.e3. PMC5948104.

206. Nabel G and Baltimore D. An inducible transcription factor activates expression of human immunodeficiency virus in T cells. *Nature*. 1987;326:711-3.

207. Duh EJ, Maury WJ, Folks TM, Fauci AS and Rabson AB. Tumor necrosis factor alpha activates human immunodeficiency virus type 1 through induction of nuclear factor binding to the NF-kappa B sites in the long terminal repeat. *Proc Natl Acad Sci U S A*. 1989;86:5974-8. PMC297754.

208. Kaufman JD, Valandra G, Roderiquez G, Bushar G, Giri C and Norcross MA. Phorbol ester enhances human immunodeficiency virus-promoted gene expression and acts on a repeated 10-base-pair functional enhancer element. *Molecular and cellular biology*. 1987;7:3759-66. PMC368032.

209. Chatila T, Silverman L, Miller R and Geha R. Mechanisms of T cell activation by the calcium ionophore ionomycin. *Journal of immunology (Baltimore, Md : 1950)*. 1989;143:1283-9.

210. Zhang X, Zhang R, Zhao H, Cai H, Gush KA, Kerr RG, Pettit GR and Kraft AS. Preclinical pharmacology of the natural product anticancer agent bryostatin 1, an activator of protein kinase C. *Cancer research*. 1996;56:802-8.

211. Korin YD, Brooks DG, Brown S, Korotzer A and Zack JA. Effects of prostratin on T-cell activation and human immunodeficiency virus latency. *J Virol*. 2002;76:8118-23. PMC155166.

212. Verdin E, Paras P, Jr. and Van Lint C. Chromatin disruption in the promoter of human immunodeficiency virus type 1 during transcriptional activation. *Embo j*. 1993;12:3249-59. PMC413592.

213. du Chene I, Basyuk E, Lin YL, Triboulet R, Knezevich A, Chable-Bessia C, Mettling C, Baillat V, Reynes J, Corbeau P, Bertrand E, Marcello A, Emiliani S, Kiernan R and Benkirane M.

Suv39H1 and HP1gamma are responsible for chromatin-mediated HIV-1 transcriptional silencing and post-integration latency. *Embo j*. 2007;26:424-35. PMC1783455.

214. Kauder SE, Bosque A, Lindqvist A, Planelles V and Verdin E. Epigenetic regulation of HIV-1 latency by cytosine methylation. *PLoS Pathog*. 2009;5:e1000495. PMC2695767.

215. Blazkova J, Trejbalova K, Gondois-Rey F, Halfon P, Philibert P, Guiguen A, Verdin E, Olive D, Van Lint C, Hejnar J and Hirsch I. CpG methylation controls reactivation of HIV from latency. *PLoS Pathog*. 2009;5:e1000554. PMC2722084.

216. Friedman J, Cho WK, Chu CK, Keedy KS, Archin NM, Margolis DM and Karn J. Epigenetic silencing of HIV-1 by the histone H3 lysine 27 methyltransferase enhancer of Zeste 2. *J Virol*. 2011;85:9078-89. PMC3165831.

217. He G and Margolis DM. Counterregulation of chromatin deacetylation and histone deacetylase occupancy at the integrated promoter of human immunodeficiency virus type 1 (HIV-1) by the HIV-1 repressor YY1 and HIV-1 activator Tat. *Molecular and cellular biology*. 2002;22:2965-73. PMC133763.

218. Mbonye U and Karn J. Transcriptional control of HIV latency: cellular signaling pathways, epigenetics, happenstance and the hope for a cure. *Virology*. 2014;454-455:328-39. PMC4010583.

219. Conrad RJ, Fozouni P, Thomas S, Sy H, Zhang Q, Zhou MM and Ott M. The Short Isoform of BRD4 Promotes HIV-1 Latency by Engaging Repressive SWI/SNF Chromatin-Remodeling Complexes. *Molecular cell*. 2017;67:1001-1012.e6. PMC5610089.

220. Han Y, Lin YB, An W, Xu J, Yang HC, O'Connell K, Dordai D, Boeke JD, Siliciano JD and Siliciano RF. Orientation-dependent regulation of integrated HIV-1 expression by host gene transcriptional readthrough. *Cell Host Microbe*. 2008;4:134-46. PMC2604135.

221. Shan L, Yang HC, Rabi SA, Bravo HC, Shroff NS, Irizarry RA, Zhang H, Margolick JB, Siliciano JD and Siliciano RF. Influence of host gene transcription level and orientation on HIV-1 latency in a primary-cell model. *J Virol*. 2011;85:5384-93. PMC3094997.

222. Lenasi T, Contreras X and Peterlin BM. Transcriptional interference antagonizes proviral gene expression to promote HIV latency. *Cell host & microbe*. 2008;4:123-133.

223. Besnard E, Hakre S, Kampmann M, Lim HW, Hosmane NN, Martin A, Bassik MC, Verschueren E, Battivelli E, Chan J, Svensson JP, Gramatica A, Conrad RJ, Ott M, Greene WC, Krogan NJ, Siliciano RF, Weissman JS and Verdin E. The mTOR Complex Controls HIV Latency. *Cell Host Microbe*. 2016;20:785-797. PMC5354304.

224. Jiang C, Lian X, Gao C, Sun X, Einkauf KB, Chevalier JM, Chen SMY, Hua S, Rhee B, Chang K, Blackmer JE, Osborn M, Peluso MJ, Hoh R, Somsouk M, Milush J, Bertagnolli LN, Sweet SE, Varriale JA, Burbelo PD, Chun TW, Laird GM, Serrao E, Engelman AN, Carrington M, Siliciano RF, Siliciano JM, Deeks SG, Walker BD, Lichterfeld M and Yu XG. Distinct viral reservoirs in individuals with spontaneous control of HIV-1. *Nature*. 2020;585:261-267. PMC7837306.

225. Golumbeanu M, Cristinelli S, Rato S, Munoz M, Cavassini M, Beerenwinkel N and Ciuffi A. Single-Cell RNA-Seq Reveals Transcriptional Heterogeneity in Latent and Reactivated HIV-Infected Cells. *Cell reports*. 2018;23:942-950.

226. Morton EL, Forst CV, Zheng Y, DePaula-Silva AB, Ramirez NP, Planelles V and D'Orso I. Transcriptional Circuit Fragility Influences HIV Proviral Fate. *Cell reports*. 2019;27:154-171.e9. PMC6461408.

227. Weinberger LS, Burnett JC, Toettcher JE, Arkin AP and Schaffer DV. Stochastic gene expression in a lentiviral positive-feedback loop: HIV-1 Tat fluctuations drive phenotypic diversity. *Cell*. 2005;122:169-82.

228. Hughes SH and Coffin JM. What Integration Sites Tell Us about HIV Persistence. *Cell Host Microbe*. 2016;19:588-98. PMC4900157.

229. Rani A and Murphy JJ. STAT5 in Cancer and Immunity. *Journal of interferon & cytokine research : the official journal of the International Society for Interferon and Cytokine Research*. 2016;36:226-37.

230. Yang L, Chen S, Zhao Q, Sun Y and Nie H. The Critical Role of Bach2 in Shaping the Balance between CD4(+) T Cell Subsets in Immune-Mediated Diseases. *Mediators of inflammation*. 2019;2019:2609737. PMC7012215.

231. Sasaki S, Ito E, Toki T, Maekawa T, Kanezaki R, Umenai T, Muto A, Nagai H, Kinoshita T, Yamamoto M, Inazawa J, Taketo MM, Nakahata T, Igarashi K and Yokoyama M. Cloning and expression of human B cell-specific transcription factor BACH2 mapped to chromosome 6q15. *Oncogene*. 2000;19:3739-49.

232. Berkhout B and Jeang KT. trans activation of human immunodeficiency virus type 1 is sequence specific for both the single-stranded bulge and loop of the trans-acting-responsive hairpin: a quantitative analysis. *J Virol*. 1989;63:5501-4. PMC251225.

233. Cullen BR, Lomedico PT and Ju G. Transcriptional interference in avian retroviruses--

implications for the promoter insertion model of leukaemogenesis. *Nature*. 1984;307:241-5.

234. Berkhout B, Silverman RH and Jeang KT. Tat trans-activates the human immunodeficiency virus through a nascent RNA target. *Cell*. 1989;59:273-82.

235. Core LJ and Lis JT. Transcription regulation through promoter-proximal pausing of RNA polymerase II. *Science*. 2008;319:1791-2. PMC2833332.

236. Wagschal A, Rousset E, Basavarajaiah P, Contreras X, Harwig A, Laurent-Chabalier S, Nakamura M, Chen X, Zhang K, Meziane O, Boyer F, Parrinello H, Berkhout B, Terzian C, Benkirane M and Kiernan R. Microprocessor, Setx, Xrn2, and Rrp6 co-operate to induce premature termination of transcription by RNAPII. *Cell*. 2012;150:1147-57. PMC3595997.

237. Wen Y and Shatkin AJ. Transcription elongation factor hSPT5 stimulates mRNA capping. *Genes & development*. 1999;13:1774-9. PMC316881.

238. Missra A and Gilmour DS. Interactions between DSIF (DRB sensitivity inducing factor), NELF (negative elongation factor), and the Drosophila RNA polymerase II transcription elongation complex. *Proc Natl Acad Sci U S A*. 2010;107:11301-6. PMC2895096.

239. Jadlowsky JK, Wong JY, Graham AC, Dobrowolski C, Devor RL, Adams MD, Fujinaga K and Karn J. Negative elongation factor is required for the maintenance of proviral latency but does not induce promoter-proximal pausing of RNA polymerase II on the HIV long terminal repeat. *Molecular and cellular biology*. 2014;34:1911-28. PMC4019061.

240. Natarajan M, Schiralli Lester GM, Lee C, Missra A, Wasserman GA, Steffen M, Gilmour DS and Henderson AJ. Negative elongation factor (NELF) coordinates RNA polymerase II pausing, premature termination, and chromatin remodeling to regulate HIV transcription. *The Journal of biological chemistry*. 2013;288:25995-26003. PMC3764804.

241. Herrmann CH and Rice AP. Lentivirus Tat proteins specifically associate with a cellular protein kinase, TAK, that hyperphosphorylates the carboxyl-terminal domain of the large subunit of RNA polymerase II: candidate for a Tat cofactor. *J Virol*. 1995;69:1612-20. PMC188757.

242. Wei P, Garber ME, Fang SM, Fischer WH and Jones KA. A novel CDK9-associated C-type cyclin interacts directly with HIV-1 Tat and mediates its high-affinity, loop-specific binding to TAR RNA. *Cell*. 1998;92:451-62.

243. Tahirov TH, Babayeva ND, Varzavand K, Cooper JJ, Sedore SC and Price DH. Crystal structure of HIV-1 Tat complexed with human P-TEFb. *Nature*. 2010;465:747-51. PMC2885016.

244. Czudnochowski N, Bösken CA and Geyer M. Serine-7 but not serine-5 phosphorylation

primes RNA polymerase II CTD for P-TEFb recognition. *Nat Commun*. 2012;3:842.

245. Bourgeois CF, Kim YK, Churcher MJ, West MJ and Karn J. Spt5 cooperates with human immunodeficiency virus type 1 Tat by preventing premature RNA release at terminator sequences. *Molecular and cellular biology*. 2002;22:1079-93. PMC134635.

246. Yamada T, Yamaguchi Y, Inukai N, Okamoto S, Mura T and Handa H. P-TEFb-mediated phosphorylation of hSpt5 C-terminal repeats is critical for processive transcription elongation. *Molecular cell*. 2006;21:227-37.

247. Nguyen Quang N, Goudey S, Ségéral E, Mohammad A, Lemoine S, Blugeon C, Versapuech M, Paillart JC, Berlioz-Torrent C, Emiliani S and Gallois-Montbrun S. Dynamic nanopore long-read sequencing analysis of HIV-1 splicing events during the early steps of infection. *Retrovirology*. 2020;17:25. PMC7433067.

248. Kutluay SB, Emery A, Penumutchu SR, Townsend D, Tenneti K, Madison MK, Stukenbroeker AM, Powell C, Jannain D, Tolbert BS, Swanstrom RI and Bieniasz PD. Genome-Wide Analysis of Heterogeneous Nuclear Ribonucleoprotein (hnRNP) Binding to HIV-1 RNA Reveals a Key Role for hnRNP H1 in Alternative Viral mRNA Splicing. *J Virol*. 2019;93. PMC6803249.

249. Emery A, Zhou S, Pollom E and Swanstrom R. Characterizing HIV-1 Splicing by Using Next-Generation Sequencing. *J Virol*. 2017;91. PMC5331825.

250. Sertznig H, Hillebrand F, Erkelenz S, Schaal H and Widera M. Behind the scenes of HIV-1 replication: Alternative splicing as the dependency factor on the quiet. *Virology*. 2018;516:176-188.

251. Archin NM, Liberty AL, Kashuba AD, Choudhary SK, Kuruc JD, Crooks AM, Parker DC, Anderson EM, Kearney MF, Strain MC, Richman DD, Hudgens MG, Bosch RJ, Coffin JM, Eron JJ, Hazuda DJ and Margolis DM. Administration of vorinostat disrupts HIV-1 latency in patients on antiretroviral therapy. *Nature*. 2012;487:482-5. PMC3704185.

252. Yeh YJ and Ho YC. Shock-and-kill versus block-and-lock: Targeting the fluctuating and heterogeneous HIV-1 gene expression. *Proc Natl Acad Sci U S A*. 2021;118. PMC8072369.

253. Zhu J, Gaiha GD, John SP, Pertel T, Chin CR, Gao G, Qu H, Walker BD, Elledge SJ and Brass AL. Reactivation of latent HIV-1 by inhibition of BRD4. *Cell reports*. 2012;2:807-16. PMC3523124.

254. Archin NM, Espeseth A, Parker D, Cheema M, Hazuda D and Margolis DM. Expression

of latent HIV induced by the potent HDAC inhibitor suberoylanilide hydroxamic acid. *AIDS research and human retroviruses*. 2009;25:207-12. PMC2853863.

255. Park SY, Kim KC, Hong KJ, Kim SS and Choi BS. Histone deactylase inhibitor SAHA induces a synergistic HIV-1 reactivation by 12-O-tetradecanoylphorbol-13-acetate in latently infected cells. *Intervirology*. 2013;56:242-8.

256. Rasmussen TA, Schmeltz Søgaard O, Brinkmann C, Wightman F, Lewin SR, Melchjorsen J, Dinarello C, Østergaard L and Tolstrup M. Comparison of HDAC inhibitors in clinical development: effect on HIV production in latently infected cells and T-cell activation. *Human vaccines & immunotherapeutics*. 2013;9:993-1001. PMC3899169.

257. Wei DG, Chiang V, Fyne E, Balakrishnan M, Barnes T, Graupe M, Hesselgesser J, Irrinki A, Murry JP, Stepan G, Stray KM, Tsai A, Yu H, Spindler J, Kearney M, Spina CA, McMahon D, Lalezari J, Sloan D, Mellors J, Geleziunas R and Cihlar T. Histone deacetylase inhibitor romidepsin induces HIV expression in CD4 T cells from patients on suppressive antiretroviral therapy at concentrations achieved by clinical dosing. *PLoS Pathog*. 2014;10:e1004071. PMC3983056 Gilead Sciences, Inc. JM is a scientific advisor of Gilead Sciences, Inc. JL is a director of Quest Clinical Research that provided clinical specimens. EF, JS, MK, CAS, and DM have no competing interests. Part of the study was funded by Gilead Sciences, Inc. Employees of the company were involved in study design, data collection and analysis, and writing the manuscript. However, none of the therapeutics characterized in this study are products of Gilead Sciences, and studies were designed, executed, and objectively interpreted purely with a purpose to advance basic research in the field of HIV latency and cure. This does not alter our adherence to all PLOS Pathogens policies on sharing data and materials.

258. Novis CL, Archin NM, Buzon MJ, Verdin E, Round JL, Lichterfeld M, Margolis DM, Planelles V and Bosque A. Reactivation of latent HIV-1 in central memory $CD4^+$ T cells through TLR-1/2 stimulation. *Retrovirology*. 2013;10:119. PMC3826617.

259. Pache L, Dutra MS, Spivak AM, Marlett JM, Murry JP, Hwang Y, Maestre AM, Manganaro L, Vamos M, Teriete P, Martins LJ, König R, Simon V, Bosque A, Fernandez-Sesma A, Cosford ND, Bushman FD, Young JA, Planelles V and Chanda SK. BIRC2/cIAP1 Is a Negative Regulator of HIV-1 Transcription and Can Be Targeted by Smac Mimetics to Promote Reversal of Viral Latency. *Cell Host Microbe*. 2015;18:345-53. PMC4617541.

260. Nixon CC, Mavigner M, Sampey GC, Brooks AD, Spagnuolo RA, Irlbeck DM, Mattingly

C, Ho PT, Schoof N, Cammon CG, Tharp GK, Kanke M, Wang Z, Cleary RA, Upadhyay AA, De C, Wills SR, Falcinelli SD, Galardi C, Walum H, Schramm NJ, Deutsch J, Lifson JD, Fennessey CM, Keele BF, Jean S, Maguire S, Liao B, Browne EP, Ferris RG, Brehm JH, Favre D, Vanderford TH, Bosinger SE, Jones CD, Routy JP, Archin NM, Margolis DM, Wahl A, Dunham RM, Silvestri G, Chahroudi A and Garcia JV. Systemic HIV and SIV latency reversal via non-canonical NF-κB signalling in vivo. *Nature*. 2020;578:160-165. PMC7111210.

261. DeChristopher BA, Loy BA, Marsden MD, Schrier AJ, Zack JA and Wender PA. Designed, synthetically accessible bryostatin analogues potently induce activation of latent HIV reservoirs in vitro. *Nature chemistry*. 2012;4:705-10. PMC3428736.

262. Yang HC, Xing S, Shan L, O'Connell K, Dinoso J, Shen A, Zhou Y, Shrum CK, Han Y, Liu JO, Zhang H, Margolick JB and Siliciano RF. Small-molecule screening using a human primary cell model of HIV latency identifies compounds that reverse latency without cellular activation. *J Clin Invest*. 2009;119:3473-86. PMC2769176.

263. Timmons A, Fray E, Kumar M, Wu F, Dai W, Bullen CK, Kim P, Hetzel C, Yang C, Beg S, Lai J, Pomerantz JL, Yukl SA, Siliciano JD and Siliciano RF. HSF1 inhibition attenuates HIV-1 latency reversal mediated by several candidate LRAs In Vitro and Ex Vivo. *Proc Natl Acad Sci U S A*. 2020;117:15763-15771.

264. Vallejo-Gracia A, Chen IP, Perrone R, Besnard E, Boehm D, Battivelli E, Tezil T, Krey K, Raymond KA, Hull PA, Walter M, Habrylo I, Cruz A, Deeks S, Pillai S, Verdin E and Ott M. FOXO1 promotes HIV latency by suppressing ER stress in T cells. *Nature microbiology*. 2020;5:1144-1157. PMC7483895.

265. Yeh Y-HJ, Jenike KM, Calvi RM, Chiarella J, Hoh R, Deeks SG and Ho Y-C. Filgotinib suppresses HIV-1–driven gene transcription by inhibiting HIV-1 splicing and T cell activation. *The Journal of Clinical Investigation*. 2020;130:4969-4984.

266. Gavegnano C, Detorio M, Montero C, Bosque A, Planelles V and Schinazi RF. Ruxolitinib and tofacitinib are potent and selective inhibitors of HIV-1 replication and virus reactivation in vitro. *Antimicrobial agents and chemotherapy*. 2014;58:1977-86. PMC4023721.

267. Vansant G, Bruggemans A, Janssens J and Debyser Z. Block-And-Lock Strategies to Cure HIV Infection. *Viruses*. 2020;12. PMC7019976.

268. Vranckx LS, Demeulemeester J, Saleh S, Boll A, Vansant G, Schrijvers R, Weydert C, Battivelli E, Verdin E, Cereseto A, Christ F, Gijsbers R and Debyser Z. LEDGIN-mediated

Inhibition of Integrase-LEDGF/p75 Interaction Reduces Reactivation of Residual Latent HIV. *EBioMedicine*. 2016;8:248-264. PMC4919729.

269. Janssens J, Bruggemans A, Christ F and Debyser Z. Towards a Functional Cure of HIV-1: Insight Into the Chromatin Landscape of the Provirus. *Frontiers in microbiology*. 2021;12:636642. PMC8044952.

270. Spina CA, Anderson J, Archin NM, Bosque A, Chan J, Famiglietti M, Greene WC, Kashuba A, Lewin SR, Margolis DM, Mau M, Ruelas D, Saleh S, Shirakawa K, Siliciano RF, Singhania A, Soto PC, Terry VH, Verdin E, Woelk C, Wooden S, Xing S and Planelles V. An in-depth comparison of latent HIV-1 reactivation in multiple cell model systems and resting CD4+ T cells from aviremic patients. *PLoS Pathog*. 2013;9:e1003834. PMC3873446.

271. Swiggard WJ, Baytop C, Yu JJ, Dai J, Li C, Schretzenmair R, Theodosopoulos T and O'Doherty U. Human immunodeficiency virus type 1 can establish latent infection in resting CD4+ T cells in the absence of activating stimuli. *J Virol*. 2005;79:14179-88. PMC1280214.

272. Bosque A and Planelles V. Induction of HIV-1 latency and reactivation in primary memory CD4+ T cells. *Blood*. 2009;113:58-65. PMC2614643.

273. Lassen KG, Hebbeler AM, Bhattacharyya D, Lobritz MA and Greene WC. A flexible model of HIV-1 latency permitting evaluation of many primary CD4 T-cell reservoirs. *PloS one*. 2012;7:e30176. PMC3265466.

274. Cameron PU, Saleh S, Sallmann G, Solomon A, Wightman F, Evans VA, Boucher G, Haddad EK, Sekaly RP, Harman AN, Anderson JL, Jones KL, Mak J, Cunningham AL, Jaworowski A and Lewin SR. Establishment of HIV-1 latency in resting CD4+ T cells depends on chemokine-induced changes in the actin cytoskeleton. *Proc Natl Acad Sci U S A*. 2010;107:16934-9. PMC2947912.

275. Symons J, Cameron PU and Lewin SR. HIV integration sites and implications for maintenance of the reservoir. *Curr Opin HIV AIDS*. 2018;13:152-159. PMC5808998.

276. Battivelli E and Verdin E. HIV(GKO): A Tool to Assess HIV-1 Latency Reversal Agents in Human Primary CD4(+) T Cells. *Bio-protocol*. 2018;8. PMC7909466.

277. Tyagi M, Pearson RJ and Karn J. Establishment of HIV latency in primary CD4+ cells is due to epigenetic transcriptional silencing and P-TEFb restriction. *J Virol*. 2010;84:6425-37. PMC2903277.

278. Bosque A and Planelles V. Studies of HIV-1 latency in an ex vivo model that uses primary

central memory T cells. *Methods (San Diego, Calif)*. 2011;53:54-61. PMC3031099.

279. Falcinelli SD, Ceriani C, Margolis DM and Archin NM. New Frontiers in Measuring and Characterizing the HIV Reservoir. *Frontiers in microbiology*. 2019;10:2878. PMC6930150.

280. Stone M, Rosenbloom D, Bacchetti P, Deng X, Dimapasoc M, Keating S, Bakkour S, Richman D, Mellors J, Deeks S, Lai J, Beg S, Siliciano J, Pagliuzza A, Chomont N, Lackman-Smith C, Ptak RG and Busch MP. Assessing suitability of next-generation viral outgrowth assays as proxies for classic QVOA to measure HIV-1 latent reservoir size. *The Journal of infectious diseases*. 2020.

281. Fujinaga K and Cary DC. Experimental Systems for Measuring HIV Latency and Reactivation. *Viruses*. 2020;12. PMC7696534.

282. Massanella M and Richman DD. Measuring the latent reservoir in vivo. *J Clin Invest*. 2016;126:464-72. PMC4731179.

283. Towbin BD, Gonzalez-Sandoval A and Gasser SM. Mechanisms of heterochromatin subnuclear localization. *Trends in biochemical sciences*. 2013;38:356-63.

284. Peric-Hupkes D, Meuleman W, Pagie L, Bruggeman SW, Solovei I, Brugman W, Gräf S, Flicek P, Kerkhoven RM, van Lohuizen M, Reinders M, Wessels L and van Steensel B. Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. *Molecular cell*. 2010;38:603-13. PMC5975946.

285. Andrey G and Mundlos S. The three-dimensional genome: regulating gene expression during pluripotency and development. *Development (Cambridge, England)*. 2017;144:3646-3658.

286. Marsman J and Horsfield JA. Long distance relationships: enhancer-promoter communication and dynamic gene transcription. *Biochimica et biophysica acta*. 2012;1819:1217-27.

287. Kleinjan DA and van Heyningen V. Long-range control of gene expression: emerging mechanisms and disruption in disease. *American journal of human genetics*. 2005;76:8-32. PMC1196435.

288. Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, Piolot T, van Berkum NL, Meisig J, Sedat J, Gribnau J, Barillot E, Blüthgen N, Dekker J and Heard E. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*. 2012;485:381-385.

289. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS and Ren B. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*.

2012;485:376-80. PMC3356448.

290. Rowley MJ and Corces VG. Organizational principles of 3D genome architecture. *Nat Rev Genet*. 2018;19:789-800. PMC6312108.

291. Hansen AS, Cattoglio C, Darzacq X and Tjian R. Recent evidence that TADs and chromatin loops are dynamic structures. *Nucleus*. 2018;9:20-32. PMC5990973.

292. Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES and Aiden EL. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*. 2014;159:1665-80. PMC5635824.

293. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, Sandstrom R, Bernstein B, Bender MA, Groudine M, Gnirke A, Stamatoyannopoulos J, Mirny LA, Lander ES and Dekker J. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*. 2009;326:289-93. PMC2858594.

294. Filippova GN, Fagerlie S, Klenova EM, Myers C, Dehner Y, Goodwin G, Neiman PE, Collins SJ and Lobanenkov VV. An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. *Molecular and cellular biology*. 1996;16:2802-13. PMC231272.

295. Phillips JE and Corces VG. CTCF: master weaver of the genome. *Cell*. 2009;137:1194-211. PMC3040116.

296. Jung YH, Kremsky I, Gold HB, Rowley MJ, Punyawai K, Buonanotte A, Lyu X, Bixler BJ, Chan AWS and Corces VG. Maintenance of CTCF- and Transcription Factor-Mediated Interactions from the Gametes to the Early Mouse Embryo. *Molecular cell*. 2019;75:154-171.e5.

297. Ong CT and Corces VG. CTCF: an architectural protein bridging genome topology and function. *Nat Rev Genet*. 2014;15:234-46. PMC4610363.

298. Xiang JF and Corces VG. Regulation of 3D chromatin organization by CTCF. *Current opinion in genetics & development*. 2021;67:33-40. PMC8084898.

299. Rae MM and Franke WW. The interphase distribution of satellite DNA-containing heterochromatin in mouse nuclei. *Chromosoma*. 1972;39:443-56.

300. Schermelleh L, Carlton PM, Haase S, Shao L, Winoto L, Kner P, Burke B, Cardoso MC, Agard DA, Gustafsson MG, Leonhardt H and Sedat JW. Subdiffraction multicolor imaging of the nuclear periphery with 3D structured illumination microscopy. *Science*. 2008;320:1332-6.

PMC2916659.

301. Huisinga KL, Brower-Toland B and Elgin SC. The contradictory definitions of heterochromatin: transcription and silencing. *Chromosoma*. 2006;115:110-22.

302. Guelen L, Pagie L, Brasset E, Meuleman W, Faza MB, Talhout W, Eussen BH, de Klein A, Wessels L, de Laat W and van Steensel B. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature*. 2008;453:948-51.

303. Lund E, Oldenburg AR, Delbarre E, Freberg CT, Duband-Goulet I, Eskeland R, Buendia B and Collas P. Lamin A/C-promoter interactions specify chromatin state-dependent transcription outcomes. *Genome Res*. 2013;23:1580-9. PMC3787256.

304. Pickersgill H, Kalverda B, de Wit E, Talhout W, Fornerod M and van Steensel B. Characterization of the Drosophila melanogaster genome at the nuclear lamina. *Nature genetics*. 2006;38:1005-14.

305. Wang H, Xu X, Nguyen CM, Liu Y, Gao Y, Lin X, Daley T, Kipniss NH, La Russa M and Qi LS. CRISPR-Mediated Programmable 3D Genome Positioning and Nuclear Organization. *Cell*. 2018;175:1405-1417.e14. PMC6239909.

306. Dekker J, Rippe K, Dekker M and Kleckner N. Capturing chromosome conformation. *Science*. 2002;295:1306-11.

307. Han J, Zhang Z and Wang K. 3C and 3C-based techniques: the powerful tools for spatial genome organization deciphering. *Molecular cytogenetics*. 2018;11:21. PMC5845197.

308. Majumder K, Boftsi M and Pintel DJ. Viral Chromosome Conformation Capture (V3C) Assays for Identifying Trans-interaction Sites between Lytic Viruses and the Cellular Genome. *Bio-protocol*. 2019;9. PMC6482961.

309. Majumder K, Wang J, Boftsi M, Fuller MS, Rede JE, Joshi T and Pintel DJ. Parvovirus Minute Virus of Mice Interacts with Sites of Cellular DNA Damage to Establish and Amplify its Lytic Infection. *eLife*. 2018;7.

310. Simonis M, Kooren J and de Laat W. An evaluation of 3C-based methods to capture DNA interactions. *Nature methods*. 2007;4:895-901.

311. Splinter E, de Wit E, Nora EP, Klous P, van de Werken HJ, Zhu Y, Kaaij LJ, van Ijcken W, Gribnau J, Heard E and de Laat W. The inactive X chromosome adopts a unique three-dimensional conformation that is dependent on Xist RNA. *Genes & development*. 2011;25:1371-83. PMC3134081.

312. van de Werken HJ, Landan G, Holwerda SJ, Hoichman M, Klous P, Chachik R, Splinter E, Valdes-Quezada C, Oz Y, Bouwman BA, Verstegen MJ, de Wit E, Tanay A and de Laat W. Robust 4C-seq data analysis to screen for regulatory DNA interactions. *Nature methods*. 2012;9:969-72.

313. Kolovos P, Brouwer RWW, Kockx CEM, Lesnussa M, Kepper N, Zuin J, Imam AMA, van de Werken HJG, Wendt KS, Knoch TA, van Ijcken WFJ and Grosveld F. Investigation of the spatial structure and interactions of the genome at sub-kilobase-pair resolution using T2C. *Nature Protocols*. 2018;13:459-477.

314. Kolovos P, van de Werken HJ, Kepper N, Zuin J, Brouwer RW, Kockx CE, Wendt KS, van IWF, Grosveld F and Knoch TA. Targeted Chromatin Capture (T2C): a novel high resolution high throughput method to detect genomic interactions and regulatory elements. *Epigenetics & chromatin*. 2014;7:10. PMC4100494.

315. Bickmore WA. The spatial organization of the human genome. *Annual review of genomics and human genetics*. 2013;14:67-84.

316. Wang F, Flanagan J, Su N, Wang LC, Bui S, Nielson A, Wu X, Vo HT, Ma XJ and Luo Y. RNAscope: a novel in situ RNA analysis platform for formalin-fixed, paraffin-embedded tissues. *J Mol Diagn*. 2012;14:22-9. PMC3338343.

317. Shah R, Lan S, Puray-Chavez MN, Liu D, Tedbury PR and Sarafianos SG. Single-cell Multiplexed Fluorescence Imaging to Visualize Viral Nucleic Acids and Proteins and Monitor HIV, HTLV, HBV, HCV, Zika Virus, and Influenza Infection. *Journal of visualized experiments : JoVE*. 2020. PMC7774762.

318. Puray-Chavez M, Tedbury PR, Huber AD, Ukah OB, Yapo V, Liu D, Ji J, Wolf JJ, Engelman AN and Sarafianos SG. Multiplex single-cell visualization of nucleic acids and protein during HIV infection. *Nat Commun*. 2017;8:1882. PMC5709414.

319. Ukah OB, Puray-Chavez M, Tedbury PR, Herschhorn A, Sodroski JG and Sarafianos SG. Visualization of HIV-1 RNA Transcription from Integrated HIV-1 DNA in Reactivated Latently Infected Cells. *Viruses*. 2018;10. PMC6212899.

320. Chen J, Liu Y, Wu B, Nikolaitchik OA, Mohan PR, Chen J, Pathak VK and Hu WS. Visualizing the translation and packaging of HIV-1 full-length RNA. *Proc Natl Acad Sci U S A*. 2020;117:6145-6155. PMC7084099.

321. Shim SH. Super-resolution microscopy of genome organization. *Genes & genomics*.

2021;43:281-287.

322. Sun NH, Chen DY, Ye LP, Sheng G, Gong JJ, Chen BH, Lu YM and Han F. CRISPR-Sunspot: Imaging of endogenous low-abundance RNA at the single-molecule level in live cells. *Theranostics*. 2020;10:10993-11012. PMC7532675.

323. Knight SC, Tjian R and Doudna JA. Genomes in Focus: Development and Applications of CRISPR-Cas9 Imaging Technologies. *Angewandte Chemie (International ed in English)*. 2018;57:4329-4337. PMC6014596.

324. Ma S and Zhang Y. Profiling chromatin regulatory landscape: insights into the development of ChIP-seq and ATAC-seq. *Molecular Biomedicine*. 2020;1:9.

325. Skene PJ, Henikoff JG and Henikoff S. Targeted in situ genome-wide profiling with high efficiency for low cell numbers. *Nat Protoc*. 2018;13:1006-1019.

326. Boyle AP, Song L, Lee BK, London D, Keefe D, Birney E, Iyer VR, Crawford GE and Furey TS. High-resolution genome-wide in vivo footprinting of diverse transcription factors in human cells. *Genome Res*. 2011;21:456-64. PMC3044859.

327. Schones DE, Cui K, Cuddapah S, Roh TY, Barski A, Wang Z, Wei G and Zhao K. Dynamic regulation of nucleosome positioning in the human genome. *Cell*. 2008;132:887-98.

328. Buenrostro JD, Giresi PG, Zaba LC, Chang HY and Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature methods*. 2013;10:1213-8. PMC3959825.

329. Corces MR, Trevino AE, Hamilton EG, Greenside PG, Sinnott-Armstrong NA, Vesuna S, Satpathy AT, Rubin AJ, Montine KS, Wu B, Kathiria A, Cho SW, Mumbach MR, Carter AC, Kasowski M, Orloff LA, Risca VI, Kundaje A, Khavari PA, Montine TJ, Greenleaf WJ and Chang HY. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nature methods*. 2017;14:959-962. PMC5623106.

330. Kukurba KR and Montgomery SB. RNA Sequencing and Analysis. *Cold Spring Harb Protoc*. 2015;2015:951-69. PMC4863231.

331. Sessegolo C, Cruaud C, Da Silva C, Cologne A, Dubarry M, Derrien T, Lacroix V and Aury J-M. Transcriptome profiling of mouse samples using nanopore sequencing of cDNA and RNA molecules. *Scientific Reports*. 2019;9:14908.

332. Philpott M, Watson J, Thakurta A, Brown T, Brown T, Oppermann U and Cribbs AP. Nanopore sequencing of single-cell transcriptomes with scCOLOR-seq. *Nature biotechnology*.

2021.

333. Gallardo CM, Wang S, Montiel-Garcia DJ, Little SJ, Smith DM, Routh AL and Torbett BE. MrHAMER yields highly accurate single molecule viral sequences enabling analysis of intra-host evolution. *Nucleic Acids Res*. 2021.

334. Kono N and Arakawa K. Nanopore sequencing: Review of potential applications in functional genomics. *Development, growth & differentiation*. 2019;61:316-326.

335. McNaughton AL, Roberts HE, Bonsall D, de Cesare M, Mokaya J, Lumley SF, Golubchik T, Piazza P, Martin JB, de Lara C, Brown A, Ansari MA, Bowden R, Barnes E and Matthews PC. Illumina and Nanopore methods for whole genome sequencing of hepatitis B virus (HBV). *Sci Rep*. 2019;9:7081. PMC6506499.

336. Kang H, Wiedmer A, Yuan Y, Robertson E and Lieberman PM. Coordination of KSHV latent and lytic gene control by CTCF-cohesin mediated chromosome conformation. *PLoS Pathog*. 2011;7:e1002140. PMC3158054.

337. Melamed A, Yaguchi H, Miura M, Witkover A, Fitzgerald TW, Birney E and Bangham CR. The human leukemia virus HTLV-1 alters the structure and transcription of host chromatin in cis. *eLife*. 2018;7. PMC6019074.

338. Satou Y, Miyazato P, Ishihara K, Yaguchi H, Melamed A, Miura M, Fukuda A, Nosaka K, Watanabe T, Rowan AG, Nakao M and Bangham CRM. The retrovirus HTLV-1 inserts an ectopic CTCF-binding site into the human genome. *Proceedings of the National Academy of Sciences*. 2016;113:3054-3059.

339. Cook L, Melamed A, Yaguchi H and Bangham CR. The impact of HTLV-1 on the cellular genome. *Curr Opin Virol*. 2017;26:125-131.

340. Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES and Aiden EL. Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. *Cell systems*. 2016;3:99-101. PMC5596920.

341. Zheng Q and Lavis LD. Development of photostable fluorophores for molecular imaging. *Current opinion in chemical biology*. 2017;39:32-38.

342. Liu D, Tedbury PR, Lan S, Huber AD, Puray-Chavez MN, Ji J, Michailidis E, Saeed M, Ndongwe TP, Bassit LC, Schinazi RF, Ralston R, Rice CM and Sarafianos SG. Visualization of Positive and Negative Sense Viral RNA for Probing the Mechanism of Direct-Acting Antivirals against Hepatitis C Virus. *Viruses*. 2019;11. PMC6893808.

343. Xia C, Babcock HP, Moffitt JR and Zhuang X. Multiplexed detection of RNA using MERFISH and branched DNA amplification. *Sci Rep*. 2019;9:7721. PMC6531529.

344. Player AN, Shen LP, Kenny D, Antao VP and Kolberg JA. Single-copy gene detection using branched DNA (bDNA) in situ hybridization. *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society*. 2001;49:603-12.

345. Battich N, Stoeger T and Pelkmans L. Image-based transcriptomics in thousands of single human cells at single-molecule resolution. *Nature methods*. 2013;10:1127-33.

346. Bushnell S, Budde J, Catino T, Cole J, Derti A, Kelso R, Collins ML, Molino G, Sheridan P, Monahan J and Urdea M. ProbeDesigner: for the design of probesets for branched DNA (bDNA) signal amplification assays. *Bioinformatics (Oxford, England)*. 1999;15:348-55.

347. Brown K. Visualizing nuclear proteins together with transcribed and inactive genes in structurally preserved cells. *Methods (San Diego, Calif)*. 2002;26:10-8.

348. Dharan A, Bachmann N, Talley S, Zwikelmaier V and Campbell EM. Nuclear pore blockade reveals that HIV-1 completes reverse transcription and uncoating in the nucleus. *Nature microbiology*. 2020;5:1088-1095.

349. Hamid FB, Kim J and Shin CG. Distribution and fate of HIV-1 unintegrated DNA species: a comprehensive update. *AIDS research and therapy*. 2017;14:9. PMC5314604.

350. Geis FK and Goff SP. Unintegrated HIV-1 DNAs are loaded with core and linker histones and transcriptionally silenced. *Proc Natl Acad Sci U S A*. 2019;116:23735-23742. PMC6876238.

351. Gutierrez-Escribano P, Hormeño S, Madariaga-Marcos J, Solé-Soler R, O'Reilly FJ, Morris K, Aicart-Ramos C, Aramayo R, Montoya A, Kramer H, Rappsilber J, Torres-Rosell J, Moreno-Herrero F and Aragon L. Purified Smc5/6 Complex Exhibits DNA Substrate Recognition and Compaction. *Molecular cell*. 2020;80:1039-1054.e6. PMC7758880.

352. Rossi F, Helbling-Leclerc A, Kawasumi R, Jegadesan NK, Xu X, Devulder P, Abe T, Takata M, Xu D, Rosselli F and Branzei D. SMC5/6 acts jointly with Fanconi anemia factors to support DNA repair and genome stability. *EMBO reports*. 2020;21:e48222. PMC7001510.

353. Balak JRA, de Graaf N, Zaldumbide A, Rabelink TJ, Hoeben RC, de Koning EJP and Carlotti F. Highly efficient ex vivo lentiviral transduction of primary human pancreatic exocrine cells. *Sci Rep*. 2019;9:15870. PMC6825235.

354. Ge X, Rubinstein WS, Jung YC and Wu Q. Genome-wide analysis of antisense transcription with Affymetrix exon array. *BMC genomics*. 2008;9:27. PMC2257944.

355. van Duin M, van Den Tol J, Hoeijmakers JH, Bootsma D, Rupp IP, Reynolds P, Prakash L and Prakash S. Conserved pattern of antisense overlapping transcription in the homologous human ERCC-1 and yeast RAD10 DNA repair gene regions. *Molecular and cellular biology*. 1989;9:1794-8. PMC362600.

356. Yu W, Gius D, Onyango P, Muldoon-Jacobs K, Karp J, Feinberg AP and Cui H. Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA. *Nature*. 2008;451:202-6. PMC2743558.

357. Xu Z, Wei W, Gagneur J, Clauder-Münster S, Smolik M, Huber W and Steinmetz LM. Antisense expression increases gene expression variability and locus interdependency. *Molecular systems biology*. 2011;7:468. PMC3063692.

358. Faghihi MA, Modarresi F, Khalil AM, Wood DE, Sahagan BG, Morgan TE, Finch CE, St Laurent G, 3rd, Kenny PJ and Wahlestedt C. Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of beta-secretase. *Nature medicine*. 2008;14:723-30. PMC2826895.

359. Li R, Sklutuis R, Groebner JL and Romerio F. HIV-1 Natural Antisense Transcription and Its Role in Viral Persistence. *Viruses*. 2021;13. PMC8145503.

360. Mancarella A, Procopio FA, Achsel T, De Crignis E, Foley BT, Corradin G, Bagni C, Pantaleo G and Graziosi C. Detection of Human Immunodeficiency Virus Type 1 (HIV-1) Antisense Protein (ASP) RNA Transcripts in Patients by Strand-Specific RT-PCR. *Journal of visualized experiments : JoVE*. 2019.

361. Mancarella A, Procopio FA, Achsel T, De Crignis E, Foley BT, Corradin G, Bagni C, Pantaleo G and Graziosi C. Detection of antisense protein (ASP) RNA transcripts in individuals infected with human immunodeficiency virus type 1 (HIV-1). *The Journal of general virology*. 2019;100:863-876.

362. Ma G, Yasunaga JI, Shimura K, Takemoto K, Watanabe M, Amano M, Nakata H, Liu B, Zuo X and Matsuoka M. Human retroviral antisense mRNAs are retained in the nuclei of infected cells for viral persistence. *Proc Natl Acad Sci U S A*. 2021;118. PMC8092383.

363. Hu B, Guo H, Zhou P and Shi Z-L. Characteristics of SARS-CoV-2 and COVID-19. *Nature Reviews Microbiology*. 2021;19:141-154.

364. Snijder EJ, Limpens R, de Wilde AH, de Jong AWM, Zevenhoven-Dobbe JC, Maier HJ, Faas F, Koster AJ and Bárcena M. A unifying structural and functional model of the coronavirus

replication organelle: Tracking down RNA synthesis. *PLoS biology*. 2020;18:e3000715. PMC7302735.

365. Sola I, Almazán F, Zúñiga S and Enjuanes L. Continuous and Discontinuous RNA Synthesis in Coronaviruses. *Annual Review of Virology*. 2015;2:265-288.

366. Alexandersen S, Chamings A and Bhatta TR. SARS-CoV-2 genomic and subgenomic RNAs in diagnostic samples are not an indicator of active replication. *Nature Communications*. 2020;11:6059.

367. Case JB, Bailey AL, Kim AS, Chen RE and Diamond MS. Growth, detection, quantification, and inactivation of SARS-CoV-2. *Virology*. 2020;548:39-48. PMC7293183.

368. Chen HC, Martinez JP, Zorita E, Meyerhans A and Filion GJ. Position effects influence HIV latency reversal. *Nature structural & molecular biology*. 2017;24:47-54.

369. Dahabieh MS, Battivelli E and Verdin E. Understanding HIV latency: the road to an HIV cure. *Annual review of medicine*. 2015;66:407-21. PMC4381961.

370. Bushman FD, Fujiwara T and Craigie R. Retroviral DNA integration directed by HIV integration protein in vitro. *Science (New York, NY)*. 1990;249:1555-8.

371. Chen Y, Zhang Y, Wang Y, Zhang L, Brinkman EK, Adam SA, Goldman R, van Steensel B, Ma J and Belmont AS. Mapping 3D genome organization relative to nuclear compartments using TSA-Seq as a cytological ruler. *The Journal of cell biology*. 2018;217:4025-4048. PMC6219710.

372. Dixon JR, Xu J, Dileep V, Zhan Y, Song F, Le VT, Yardımcı GG, Chakraborty A, Bann DV, Wang Y, Clark R, Zhang L, Yang H, Liu T, Iyyanki S, An L, Pool C, Sasaki T, Rivera-Mulia JC, Ozadam H, Lajoie BR, Kaul R, Buckley M, Lee K, Diegel M, Pezic D, Ernst C, Hadjur S, Odom DT, Stamatoyannopoulos JA, Broach JR, Hardison RC, Ay F, Noble WS, Dekker J, Gilbert DM and Yue F. Integrative detection and analysis of structural variation in cancer genomes. *Nature genetics*. 2018;50:1388-1398. PMC6301019.

373. Lee DS, Luo C, Zhou J, Chandran S, Rivkin A, Bartlett A, Nery JR, Fitzpatrick C, O'Connor C, Dixon JR and Ecker JR. Simultaneous profiling of 3D genome structure and DNA methylation in single human cells. *Nature methods*. 2019;16:999-1006. PMC6765423.

374. Tchasovnikarova IA, Timms RT, Matheson NJ, Wals K, Antrobus R, Göttgens B, Dougan G, Dawson MA and Lehner PJ. GENE SILENCING. Epigenetic silencing by the HUSH complex mediates position-effect variegation in human cells. *Science*. 2015;348:1481-1485. PMC4487827.

375. Moquin SA, Thomas S, Whalen S, Warburton A, Fernandez SG, McBride AA, Pollard KS and Miranda JJL. The Epstein-Barr virus episome maneuvers between nuclear chromatin compartments during reactivation. *J Virol*. 2017. PMC5774889.

376. Tan X, Ravasio A, Ong HT, Wu J and Hew CL. White spot syndrome viral protein VP9 alters the cellular higher-order chromatin structure. *FASEB bioAdvances*. 2020;2:264-279. PMC7133739.

377. Liu R, Yeh YJ, Varabyou A, Collora JA, Sherrill-Mix S, Talbot CC, Jr., Mehta S, Albrecht K, Hao H, Zhang H, Pollack RA, Beg SA, Calvi RM, Hu J, Durand CM, Ambinder RF, Hoh R, Deeks SG, Chiarella J, Spudich S, Douek DC, Bushman FD, Pertea M and Ho YC. Single-cell transcriptional landscapes reveal HIV-1-driven aberrant host gene transcription as a potential therapeutic target. *Science translational medicine*. 2020;12. PMC7453882.

378. Fernandez G and Zeichner SL. Cell line-dependent variability in HIV activation employing DNMT inhibitors. *Virology journal*. 2010;7:266. PMC2964676.

379. Depledge DP, Srinivas KP, Sadaoka T, Bready D, Mori Y, Placantonakis DG, Mohr I and Wilson AC. Direct RNA sequencing on nanopore arrays redefines the transcriptional complexity of a viral pathogen. *Nat Commun*. 2019;10:754. PMC6376126.

380. Kao SY, Calman AF, Luciw PA and Peterlin BM. Anti-termination of transcription within the long terminal repeat of HIV-1 by tat gene product. *Nature*. 1987;330:489-93.

381. Miller-Jensen K, Dey SS, Pham N, Foley JE, Arkin AP and Schaffer DV. Chromatin accessibility at the HIV LTR promoter sets a threshold for NF-κB mediated viral gene expression. *Integrative biology : quantitative biosciences from nano to macro*. 2012;4:661-71. PMC3362694.

382. Miyamoto K, Nguyen KT, Allen GE, Jullien J, Kumar D, Otani T, Bradshaw CR, Livesey FJ, Kellis M and Gurdon JB. Chromatin Accessibility Impacts Transcriptional Reprogramming in Oocytes. *Cell reports*. 2018;24:304-311.

383. Teif VB, Vainshtein Y, Caudron-Herger M, Mallm J-P, Marth C, Höfer T and Rippe K. Genome-wide nucleosome positioning during embryonic stem cell development. *Nature structural & molecular biology*. 2012;19:1185-1192.

384. Lu F, Liu Y, Inoue A, Suzuki T, Zhao K and Zhang Y. Establishing Chromatin Regulatory Landscape during Mouse Preimplantation Development. *Cell*. 2016;165:1375-1388.

385. Wu J, Huang B, Chen H, Yin Q, Liu Y, Xiang Y, Zhang B, Liu B, Wang Q, Xia W, Li W, Li Y, Ma J, Peng X, Zheng H, Ming J, Zhang W, Zhang J, Tian G, Xu F, Chang Z, Na J, Yang X

and Xie W. The landscape of accessible chromatin in mammalian preimplantation embryos. *Nature*. 2016;534:652-657.

386. Einkauf KB, Lee GQ, Gao C, Sharaf R, Sun X, Hua S, Chen SM, Jiang C, Lian X, Chowdhury FZ, Rosenberg ES, Chun TW, Li JZ, Yu XG and Lichterfeld M. Intact HIV-1 proviruses accumulate at distinct chromosomal positions during prolonged antiretroviral therapy. *J Clin Invest*. 2019;129:988-998. PMC6391088.

387. Jefferys SR, Burgos SD, Peterson JJ, Selitsky SR, Turner AW, James LI, Tsai YH, Coffey AR, Margolis DM, Parker J and Browne EP. Epigenomic characterization of latent HIV infection identifies latency regulating transcription factors. *PLoS Pathog*. 2021;17:e1009346. PMC7946360.

388. Rosa-Mercado NA, Zimmer JT, Apostolidi M, Rinehart J, Simon MD and Steitz JA. Hyperosmotic stress alters the RNA polymerase II interactome and induces readthrough transcription despite widespread transcriptional repression. *Molecular cell*. 2021;81:502-513.e4. PMC7867636.

389. Vilborg A, Sabath N, Wiesel Y, Nathans J, Levy-Adam F, Yario TA, Steitz JA and Shalgi R. Comparative analysis reveals genomic features of stress-induced transcriptional readthrough. *Proc Natl Acad Sci U S A*. 2017;114:E8362-e8371. PMC5635911.

390. Vilborg A and Steitz JA. Readthrough transcription: How are DoGs made and what do they do? *RNA biology*. 2017;14:632-636. PMC5449079.

391. Rutkowski AJ, Erhard F, L'Hernault A, Bonfert T, Schilhabel M, Crump C, Rosenstiel P, Efstathiou S, Zimmer R, Friedel CC and Dolken L. Widespread disruption of host transcription termination in HSV-1 infection. *Nat Commun*. 2015;6:7126. PMC4441252.

392. Hennig T, Michalski M, Rutkowski AJ, Djakovic L, Whisnant AW, Friedl MS, Jha BA, Baptista MAP, L'Hernault A, Erhard F, Dölken L and Friedel CC. HSV-1-induced disruption of transcription termination resembles a cellular stress response but selectively increases chromatin accessibility downstream of genes. *PLoS Pathog*. 2018;14:e1006954. PMC5886697.

393. Cardiello JF, Goodrich JA and Kugel JF. Heat Shock Causes a Reversible Increase in RNA Polymerase II Occupancy Downstream of mRNA Genes, Consistent with a Global Loss in Transcriptional Termination. *Molecular and cellular biology*. 2018;38. PMC6113597.

394. Pan XY, Zhao W, Zeng XY, Lin J, Li MM, Shen XT and Liu SW. Heat Shock Factor 1 Mediates Latent HIV Reactivation. *Sci Rep*. 2016;6:26294. PMC4870680.

395. Simon JH, Fouchier RA, Southerling TE, Guerra CB, Grant CK and Malim MH. The Vif and Gag proteins of human immunodeficiency virus type 1 colocalize in infected human T cells. *J Virol*. 1997;71:5259-67. PMC191762.

396. Langmead B and Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nature methods*. 2012;9:357-359.

397. Liu T. Use Model-Based Analysis of ChIP-Seq (MACS) to Analyze Short Reads Generated by Sequencing Protein–DNA Interactions in Embryonic Stem Cells. In: B. L. Kidder, ed. *Stem Cell Transcriptional Networks: Methods and Protocols* New York, NY: Springer New York; 2014: 81-95.

398. Shen L, Shao N, Liu X and Nestler E. ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. *BMC genomics*. 2014;15:284. PMC4028082.

399. Shao Z, Zhang Y, Yuan GC, Orkin SH and Waxman DJ. MAnorm: a robust model for quantitative comparison of ChIP-Seq data sets. *Genome biology*. 2012;13:R16. PMC3439967.

400. Kim D, Paggi JM, Park C, Bennett C and Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature biotechnology*. 2019;37:907-915.

401. Anders S, Pyl PT and Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics (Oxford, England)*. 2015;31:166-9. PMC4287950.

402. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT and Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature biotechnology*. 2015;33:290-5. PMC4643835.

403. Khan A and Zhang X. dbSUPER: a database of super-enhancers in mouse and human genome. *Nucleic Acids Res*. 2016;44:D164-71. PMC4702767.

404. Wolff J, Backofen R and Grüning B. Robust and efficient single-cell Hi-C clustering with approximate k-nearest neighbor graphs. *Bioinformatics (Oxford, England)*. 2021.

405. Fiorillo L, Musella F, Conte M, Kempfer R, Chiariello AM, Bianco S, Kukalev A, Irastorza-Azcarate I, Esposito A, Abraham A, Prisco A, Pombo A and Nicodemi M. Comparison of the Hi-C, GAM and SPRITE methods using polymer models of chromatin. *Nature methods*. 2021;18:482-490.

406. Langdon WB. Performance of genetic programming optimised Bowtie2 on genome comparison and analytic testing (GCAT) benchmarks. *BioData Min*. 2015;8:1. PMC4304608.

407. Medvedovic J, Ebert A, Tagoh H, Tamir IM, Schwickert TA, Novatchkova M, Sun Q, Huis In 't Veld PJ, Guo C, Yoon HS, Denizot Y, Holwerda SJ, de Laat W, Cogne M, Shi Y, Alt FW and Busslinger M. Flexible long-range loops in the VH gene region of the Igh locus facilitate the generation of a diverse antibody repertoire. *Immunity*. 2013;39:229-44. PMC4810778.

408. Thomas R, Thomas S, Holloway AK and Pollard KS. Features that define the best ChIP-seq peak calling algorithms. *Brief Bioinform*. 2017;18:441-450. PMC5429005.

409. van de Werken HJ, de Vree PJ, Splinter E, Holwerda SJ, Klous P, de Wit E and de Laat W. 4C technology: protocols and data analysis. *Methods Enzymol*. 2012;513:89-112.

410. Hagege H, Klous P, Braem C, Splinter E, Dekker J, Cathala G, de Laat W and Forne T. Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nat Protoc*. 2007;2:1722-33.

411. Chen S, Lake BB and Zhang K. High-throughput sequencing of the transcriptome and chromatin accessibility in the same cell. *Nature biotechnology*. 2019;37:1452-1457. PMC6893138.

412. Wang W, Fasolino M, Cattau B, Goldman N, Kong W, Frederick MA, McCright SJ, Kiani K, Fraietta JA and Vahedi G. Joint profiling of chromatin accessibility and CAR-T integration site analysis at population and single-cell levels. *Proc Natl Acad Sci U S A*. 2020;117:5442-5452. PMC7071901.

413. Grant CE, Bailey TL and Noble WS. FIMO: scanning for occurrences of a given motif. *Bioinformatics (Oxford, England)*. 2011;27:1017-8. PMC3065696.

414. Quinlan AR and Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics (Oxford, England)*. 2010;26:841-2. PMC2832824.

415. Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, Morgan MT and Carey VJ. Software for computing and annotating genomic ranges. *PLoS computational biology*. 2013;9:e1003118. PMC3738458.

416. Kainth AS, Chowdhary S, Pincus D and Gross DS. Primordial super-enhancers: heat shock-induced chromatin organization in yeast. *Trends in cell biology*. 2021.

417. Fujita T and Fujii H. Purification of specific DNA species using the CRISPR system. *Biology methods & protocols*. 2019;4:bpz008. PMC7200925.

418. Fujita T, Kitaura F, Yuno M, Suzuki Y, Sugano S and Fujii H. Locus-specific ChIP combined with NGS analysis reveals genomic regulatory regions that physically interact with the

Pax5 promoter in a chicken B cell line. *DNA research : an international journal for rapid publication of reports on genes and genomes*. 2017;24:537-548. PMC5737561.

419. Fujita T, Yuno M, Suzuki Y, Sugano S and Fujii H. Identification of physical interactions between genomic regions by enChIP-Seq. *Genes to cells : devoted to molecular & cellular mechanisms*. 2017;22:506-520.

420. Hamidian A, Vaapil M, von Stedingk K, Fujita T, Persson CU, Eriksson P, Veerla S, De Preter K, Speleman F, Fujii H, Påhlman S and Mohlin S. Promoter-associated proteins of EPAS1 identified by enChIP-MS - A putative role of HDX as a negative regulator. *Biochemical and biophysical research communications*. 2018;499:291-298.

421. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C and Chanda SK. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun*. 2019;10:1523. PMC6447622.