Inaugural dissertation

for

obtaining the doctoral degree

of the

Combined Faculty of Mathematics, Engineering and Natural Sciences of the

Ruprecht - Karls - University

Heidelberg

Presented by

M.Sc. Mona Rheinberger

born in Mainz

Oral examination: 26.01.2024

Characterization of the HIV-1 integration site landscape and chromatin determinants of integration in microglia cellular models and $CD4^+$ T cells

Referees: Prof. Dr. Oliver T. Fackler

Dr. Marina Lusic

Acknowledgments

Here I would like to thank all the people that supported me throughout the time of my PhD and thereby contributed to the successful completion of this journey.
First, I would like to thank Prof. Dr. Oliver T. Fackler that he agreed to be my first referee. Thanks a lot for all the helpful comments during TAC meetings and seminars.

I am also deeply grateful to Dr. Marina Lusic for affording me the opportunity to pursue my PhD thesis in her lab, for her support, and advice. Thank you for believing in me and entrusting me with such an incredible project.

Thanks a lot also to Dr. Eileen Furlong for being part of my TAC and defense committee, as well as Prof. Dr. Elmar Schiebel to agree to be my fourth referee during my defense.

Special acknowledgment goes to Dr. Carl Herrmann and Ana Luisa Costa for their indispensable bioinformatic analysis, without which this thesis would not have been possible. Thanks for the close collaboration on the project and all the fruitful discussion to achieve the best from our data.

Furthermore, I would like to express my deep appreciation to Dr. Bojana Lusic for her exceptional supervision in the lab. Beyond being a fantastic colleague, you became a true friend, always ready to assist with any questions, problems, or concerns. Your support throughout the project, including answering my queries and helping troubleshoot issues, was invaluable. I thoroughly enjoyed my time in the lab, and I will miss our shared moments, from coffee breaks to Radler sessions in Botanik, and our lengthy writing and discussion sessions in the office. Thanks to all of that I learned so much, both scientifically as well as personally. Hvala od sveg srca!

Of course, I also want to thank all members of the Lusic and Fackler labs for supporting me and making the lab time enjoyable and entertaining. Without you all would have been half the fun. Special thanks to Katharina, going through this together with you made many things much more fun.

Last but not least many thanks to my family and friends. Science is not everything :-)
I would have not survived this journey without you, M&M

Summary

The establishment of HIV-1 latent reservoirs, residing primarily in T cells of memory phenotype, represent the main barrier to an HIV-1 cure. However, also cells of the central nervous system, including astrocytes and microglia, may contribute to the reservoir, and are particularly important sanctuaries of latent HIV-1 due to low penetration of antiretroviral drugs, lack of resident T cells, and permanent viral integration. Moreover, as many as half of the people living with HIV-1 display mild to moderate neurocognitive disorders, even under antiretroviral therapy. The aim of this PhD thesis was to determine HIV-1 integration sites in the microglia genome, and to profile global genomic and cell type specific chromatin signatures that may play an important role in the establishment of productive versus latent infections in the brain microglia model. In the first part of the thesis, I established a linker-mediated PCR based method to create a first map of HIV-1 integration patterns in microglia. Subsequently, I generated multi-omics data sets to profile HIV-1 insertions on different scales together with transcriptome by RNA-Seq, epigenome by ChIP-Seq and chromatin accessibility by ATAC-Seq in uninfected microglia cellular model. Comparison of the ISs with insertions of the main cellular targets of HIV-1 $CD4^+$ T cells and macrophages revealed similar genomic feature distribution and shared gene repertoires. Intronic targeting of highly transcribed genes, demarcated by H3K36me3 and genic enhancers along with cellular cofactors requirements were found to be universal integration properties within different HIV-1 reservoirs.

The second aim of my thesis focused on the investigation of potential differences in transcriptional networks between cells harboring actively replicating virus versus cells with latent HIV-1 by performing ATAC-Seq. Transcription factor footprinting as well as biochemical fractionations revealed the architectural protein CCCTC-binding factor (CTCF) as a dynamic mark of infection in microglia cellular model. Given CTCF's significant role in determining 3D chromatin structures, the thesis explored the possibility of a connection between HIV-1 integration sites and topologically associated domains (TADs) by mapping HIV-1 insertion sites onto previously published Hi-C maps of microglia-containing tissues and T cells. The analysis uncovered TAD boundaries delimited with H3K36me3 chromatin mark as a new 3D genome determinant of HIV-1 integration in microglia and $CD4^+$ T cells. Moreover, CTCF was found to interact with viral integrase in a LEDGF/p75 dependent manner, which is an essential integration host factor and integrase interactor. To explore the potential involvement of CTCF in

HIV-1 integration, I transiently depleted CTCF in HIV-1 target cells and observed a reduction in HIV-1 integration efficiency and redistributed insertion sites, indicating that the 3D arrangement of chromatin may play a significant role in HIV-1 infection.

In conclusion, this thesis provides valuable insights into the role of chromatin shaping proteins and 3D genome architecture in viral integration and latency, which holds increasing importance in HIV-1 research. Furthermore, it demonstrated that integration patterns and chromatin characteristics in microglia, a central nervous system HIV-1 target cell type, resemble those found in blood reservoirs. This underscores the potential of using common therapeutic strategies for different HIV-1 reservoirs. Nonetheless, future research using primary tissues could unveil new, cell-type-specific factors as possible targets for effective and safe brain-targeting therapeutics.

Zusammenfassung

Die Etablierung von latenten HIV-1 Reservoiren, die hauptsächlich in T-Zellen des Gedächtnisphänotyps angesiedelt sind, stellt das Haupthindernis für eine HIV-1-Heilung dar. Aber auch Zellen des Zentralnervensystems, einschließlich Astrozyten und Mikroglia, können zum Reservoir beitragen und sind aufgrund der geringen Penetration antiretroviraler Medikamente, des Mangels an ansässigen T-Zellen und der permanenten viralen Integration besonders wichtige Zufluchtsorte für latentes HIV-1. Darüber hinaus weist die Hälfte der Menschen, die mit HIV-1 leben, leichte bis mittelschwere neurokognitive Störungen auf, selbst unter antiretroviraler Therapie. Ziel dieser Doktorarbeit war es, HIV-1-Integrationsstellen im Mikroglia-Genom zu bestimmen und ein Profil globaler genomischer und zelltypspezifischer Chromatinsignaturen zu erstellen, die eine wichtige Rolle bei der Etablierung produktiver bzw. latenter Infektionen im Mikroglia-Modell des Gehirns spielen könnten. Im ersten Teil der Arbeit habe ich eine auf Linker-basierende PCR-Methode etabliert, um ein erstes Profil der HIV-1-Integrationsmuster in Mikroglia zu erstellen. Anschließend erstellte ich Multi-omics-Datensätze, um HIV-1-Insertionen auf verschiedenen Ebenen zusammen mit dem Transkriptom (RNA-Seq), dem Epigenom (ChIP-Seq) und der Chromatin-Zugänglichkeit (ATAC-Seq) in einem nicht infizierten Mikroglia-Zellmodell zu charakterisieren.

Der Vergleich der Integrationsstellen mit den Insertionen der Hauptzielzellen von HIV-1, $CD4^+$ T-Zellen und Makrophagen, ergab eine ähnliche Verteilung der genomischen Merkmale und gemeinsame Genrepertoires. Als universelle Integrationseigenschaften in verschiedenen HIV-1-Reservoirs erwiesen sich die Insertion in intronische Bereiche stark exprimierter Gene, markiert durch H3K36me3 und Enhancer, sowie die Anforderungen an zelluläre Kofaktoren.

Das zweite Ziel meiner Dissertation war die Untersuchung möglicher Unterschiede in den Transkriptionsnetzwerken zwischen Zellen, die ein aktiv replizierendes Virus beherbergen, und Zellen mit latentem HIV-1 durch ATAC-Seq. Transkriptionsfaktor-Fußabdrücke und biochemische Zellfraktionierungen ergaben, dass das Strukturprotein CCCTC-Bindungsfaktor (CTCF) eine dynamische Markierung der Infektion im Mikroglia-Zellmodell darstellt. Angesichts der bedeutenden Rolle von CTCF bei der Bestimmung von 3D-Chromatinstrukturen wurde in dieser Arbeit die Möglichkeit einer Verbindung zwischen HIV-1-Integrationsstellen und topologisch assoziierten Domänen (TADs) untersucht, indem HIV-1-Insertionsstellen auf zuvor

veröffentlichten Hi-C Intervallen von Mikroglia-haltigen Geweben und T-Zellen abgebildet wurden. Die Analyse ergab, dass die TAD-Grenzen, die durch die Chromatinmarkierung H3K36me3 abgegrenzt sind, eine neue 3D-Genom-Determinante der HIV-1-Integration in Mikroglia und $CD4^+$ T-Zellen darstellen. Außerdem wurde festgestellt, dass CTCF mit der viralen Integrase in einer LEDGF/p75-abhängigen Weise interagiert, die ein wesentlicher Integrations-Wirtsfaktor und Integrase-Interaktor ist. Um die mögliche Beteiligung von CTCF an der HIV-1-Integration zu untersuchen, habe ich CTCF in HIV-1-Zielzellen vorübergehend runter reguliert und eine Verringerung der HIV-1-Integrationseffizienz sowie eine Umverteilung der Insertionsstellen beobachtet, was darauf hindeutet, dass die 3D-Anordnung des Chromatins eine wichtige Rolle bei der HIV-1-Infektion spielen könnte. Zusammenfassend lässt sich sagen, dass diese Arbeit wertvolle Einblicke in die Rolle der Chromatin formenden Proteine und der dreidimensionalen Genomarchitektur bei der viralen Integration und Latenz liefert, was in der HIV-1-Forschung zunehmend an Bedeutung gewinnt. Darüber hinaus konnte gezeigt werden, dass Integrationsmuster und Chromatinmerkmale in Mikroglia, einem HIV-1-Zelltyp des Zentralnervensystems, denen in Blutreservoirs ähneln. Dies unterstreicht das Potenzial gemeinsamer therapeutischer Strategien für verschiedene HIV-1-Reservoirs. Dennoch könnten künftige Forschungen unter Verwendung von Primärgeweben neue, zelltypspezifische Faktoren als mögliche Ziele für wirksame und sichere, auf das Gehirn ausgerichtete Therapeutika ans Licht bringen.

Table of content

1 Introduction

1.1 Human immunodeficiency virus

1.1.1 HIV and AIDS

Human immunodeficiency virus (HIV), first discovered in 1983 (Barré-Sinoussi et al., 1983), is a lentivirus belonging to the Retroviridae family. Based on the viral genome and their replication strategy, viruses can be categorized using the Baltimore scheme (Baltimore, 1971). HIV is a retrovirus that encodes for a RNA-dependent DNA polymerase (reverse transcriptase (RT)) and classifies into class VI of the Baltimore scheme: single stranded RNA-RT virus with a + sense strand genome and a DNA intermediate (Baltimore, 1971; BALTIMORE, 1970; Temin and Mizutani, 1970).

HIV evolved from cross species transmission from simian immunodeficiency virus (SIV) from chimpanzees or sooty mangabeys to humans and genomic comparisons suggest that HIV subtype 1 (HIV-1) stems from chimpanzees and the less prevalent HIV subtype 2 (HIV-2) from sooty mangabeys (Clavel et al., 1987; Gao et al., 1992, 1999; Peeters et al., 1989; Sharp and Hahn, 2011). HIV-1 is classified into 4 lineages: M, N, O, and P, where M is causing 95% of HIV-1 infections (Hemelaar et al., 2011).

HIV-1 primarily targets immune cells of the hematopoietic/lymphoid lineage, including $CD4^+$ helper T cells, macrophages, gut-associated lymphoid tissue (GALT) and cells of the central nervous system (CNS) (Le Douce et al., 2010; Kumar et al., 2014; Wallet et al., 2019). If untreated, the virus causes the Acquired Immune deficiency Syndrome (AIDS), and infected individuals require lifelong anti-retroviral therapy (ART) to control viral load. ART interruption can lead to viral rebound due to the persistence of the virus in latent reservoirs throughout the body, making it essential for patients to adhere to their medication regimen all lifelong to prevent viral transmission and lower mortality, as so far there is no treatment that targets the stably integrated virus.

1.1.2 Transmission and Pathogenesis

HIV-1 is mainly sexually transmitted via contact with infected semen or mucosal surfaces. Less frequently infections can be passed on via blood transmissions, drug injection or from an infected mother to the infant. Viral particles pass through epithelial barriers of the mucosa after which they reach myeloid cells. The first HIV-1 target cells for cell-free viral entry are macrophages and dendritic cells, leading to moderate

amplification of the infection (Kolodkin-Gal et al., 2013). When reaching the secondary lymphoid tissues, harboring a high T cell count, virus is spread via cell-to-cell contact leading to systemic infection (Brenchley et al., 2004; Felts et al., 2010; Guadalupe et al., 2003; Hübner et al., 2009; Imle et al., 2019).

HIV-1 infections can be divided into 3 phases: the acute infection, asymptomatic infection, and AIDS development. Two to four weeks after infection, RNA levels in the patient's plasma peaks, while $CD4^+$ T cell counts decrease due to massive viral replication, leading to the development of flu-like symptoms in the patients. During the asymptomatic phase without treatment plasma RNA levels drop to a steady state for several years, causing a constant loss of T cells due to direct killing and apoptosis of infected cells. This impairment of the patient's immune system is followed by the development of opportunistic disease as for example secondary pathogen infections or tumors, leading to AIDS (Deeks et al., 2015). Worldwide there are about 39 million active HIV-1 infection cases in 2022, with around 1.3 million new cases every year (UNAIDS).

1.1.3 HIV-1 genome structure

HIV-1 viral genome is 10 kb in length and consist of 9 open reading frames encoding for 15 different viral proteins (Frankel and Young, 1998). The genome is flanked by 5' long terminal repeat (LTR) and 3' LTR, both of which consist of identical nucleotide sequences. These sequences are divided into three identical regions (U3, R, U5) (Figure 1).

Common to all retroviruses, HIV-1 genome encodes for the polyproteins group-specific antigen (gag), polymerase (pol), and envelope (env) which are cleaved by viral protease (Coffin et al., 1997) (Figure 1).

Gag produces proteins for the inner structure of the virion, e.g. capsid (CA), matrix (MA), nucleocapsid (NC), p6, spacer peptide 1 (sp1), and spacer peptide 2 (sp2). Gag pol polyprotein is cleaved into protease, reverse transcriptase (RT) to transcribe single stranded RNA (ssRNA) into viral double stranded DNA (dsDNA) and integrase (IN). Env gives rise to the glycoprotein gp120 and gp41 to build the viral envelope (Checroune et al., 1995; Jacks et al., 1988; Kondo and Göttlinger, 1996; Turner and Summers, 1999). In addition, HIV-1 viral genome encodes for nonstructural proteins viral protein R (Vpr), viral protein U (Vpu), viral infectivity factor (Vif) and negative factor (Nef) which are mainly involved in the regulation of innate immune response to HIV-1.

Trans-activator of transcription (tat) and regulator of virion (Rev) have important roles in regulating and enhancing viral gene expression (Figure 1). Rev is involved in exporting unspliced pre-mRNAs from the nucleus to the cytosol (Daly et al., 1989; Turner and Summers, 1999). Altogether, the viral proteins and 2 copies of viral RNA assemble and undergo maturation into virions with a closed cone shaped capsid and ~ 100-200 nm in diameter.

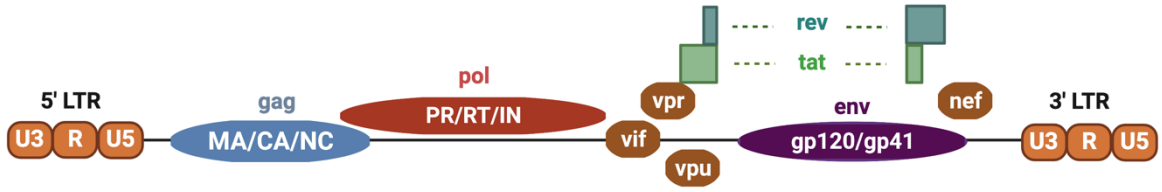


Figure 1: HIV-1 genomic structure. Schematic representation of the HIV-1 genome structure. Genome is flanked by 5' and 3' LTR which are composed of the identical U3, R and U5 regions. Gag polyprotein (red) encodes for matrix (MA), capsid (CA) and nucleocapsid (NC), while pol is cleaved after transcription into protease (PR), reverse transcriptase (RT) and integrase (IN). Viral envelope (env) is transcribed into gp120 and gp41. In addition, viral genome encodes for the accessory proteins vif, vpr, vpu, nef, and regulatory proteins tat and rev. Figure was created with BioRender.com.

1.1.4 HIV-1 Replication cycle

The viral replication cycle of HIV-1 is initiated by viral entry into the host cell. The viral surface protein gp120 interacts with the CD4 receptor on the host cell plasma membrane (PM) (Dalgleish et al., 1984), together with the co-receptors C-C chemokine receptor type 5 (CCR5) or C-X-C motif chemokine receptor 4 (CXCR4) leading to viral entry (Alkhatib et al., 1996; Choe et al., 1996; Oberlin et al., 1996). After the membrane fusion, the viral capsid core is released into the cytoplasm, where it is transported to the nuclear pore complex (NPC) by the cytoskeleton network (McClure et al., 1988; Melikyan, 2008; Stein et al., 1987). As HIV-1 is capable of infecting non-dividing cells, the pre-integration complex (PIC) can enter the nucleus through the nuclear pore. Several host factors, that interact with viral CA protein, are involved in the nuclear import of HIV-1. CA interacts with nucleoporins 358 (Nup358, also known as RAN Binding Protein 2 (RanBP2)) through FG repeats as well as with nucleoporin 153 (Nup153), a component of the inner nuclear basket (Matreyek and Engelman, 2013; Di Nunzio et al., 2013; Schaller et al., 2011). Beside nucleoporins, various host proteins including Transportin 3, Cyclophilin A, and cleavage and polyadenylation specificity factor 6 (CPSF6) were found to be involved in nuclear import (Shen et al., 2021).

In the recent years it was shown that the intact capsid core passes through the NPC (Zila et al., 2021) and that the capsid core uncoating occurs inside the nucleus (Burdick et al.; Muller et al., 2022; Schifferdecker et al., 2022). During trafficking of the capsid core from the cytosol to the NPC, reverse transcription of the viral ssRNA into double dsDNA starts within the capsid core, but reverse transcription is only completed after nuclear entry (Dharan et al., 2020; Francis et al., 2020; Müller et al., 2021; Selyutina et al., 2020). The reverse transcription of the viral genome into dsDNA is catalyzed by the viral enzyme reverse transcriptase (RT) (Sousa et al., 1993). The pre-integration complex (PIC) consisting of vDNA, capsid protein and cellular proteins, is released into the nucleoplasm and viral dsDNA is integrated into the host chromatin (Lusic and Siliciano, 2017), catalyzed by the activity of retroviral IN (Vink and Plasterk, 1993). Upon integration of viral DNA (vDNA) into the host genome, during the later phase of infection, provirus is expressed, and assembly of viral particles takes place. Therefore, synthesized viral envelope proteins are trafficked via the secretory pathway passing the rough endoplasmic reticulum and the Golgi apparatus to the PM. Additionally, the Gag polyprotein, which was synthesized in the cytosol, recruits the viral genomic ssRNA, multimerizes and reaches the PM. These HIV-1 proteins recruit several host factors to promote virus assembly and release at the PM. The viral particle undergoes maturation through the action of the viral protease, which cleaves the Gag and Gag-Pol proteins, resulting in a restructuring of the viral protein's morphology (Freed, 2015) (Figure 2).

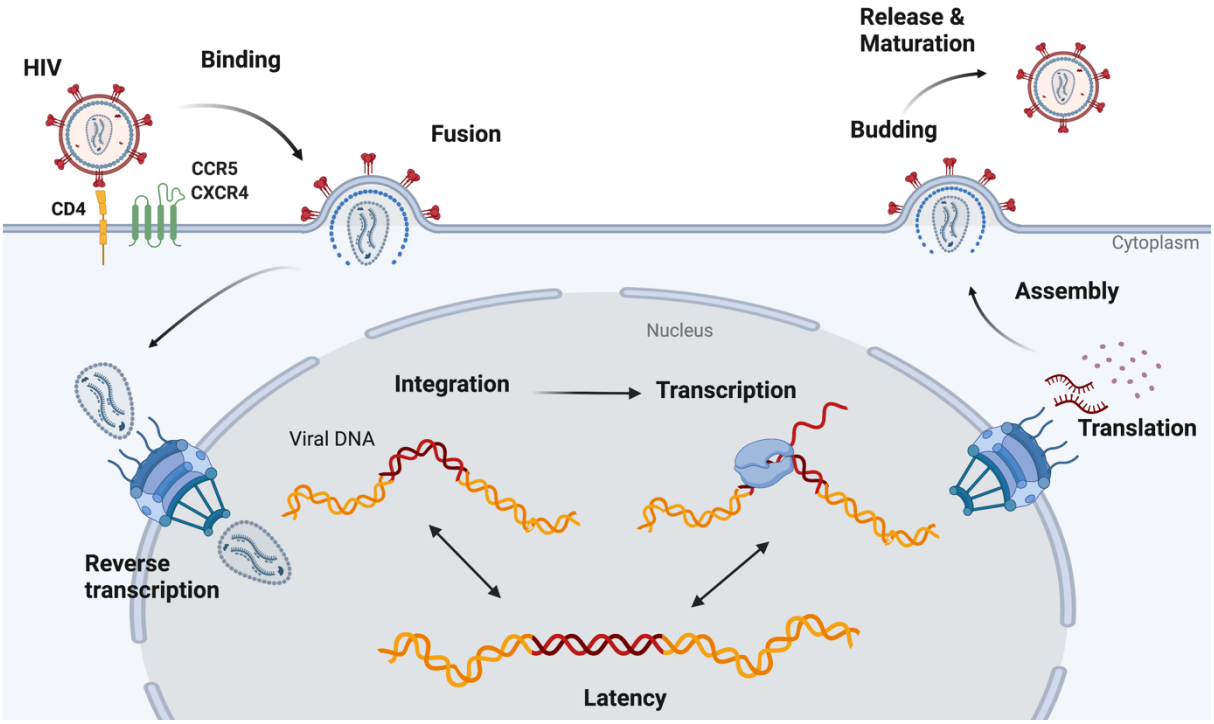


Figure 2: HIV-1 replication cycle. HIV-1 enters the host cell via binding to the CD4 and CCR5 or CXCR4 co-receptors and membrane fusion. The viral capsid transfers through the nuclear pore complex into the nucleus, where reverse transcription from the viral RNA into viral cDNA is completed. In the nucleus, pre-integration complex is released, and provirus integrates with the help of viral integrase into the host genome. On the transcriptional level, virus either goes into latency, or provirus is transcribed, new viral proteins are translated, assembled, and released for a new round of infection. Figure was prepared with BioRender.com.

1.1.5 Integration

After CA core enters the nucleus via the nuclear pore complex, HIV-1 integrates its genome into the host genome. Newest research revealed that RT is completed in the nucleus followed by the release of vDNA into the nucleoplasm where integration is happening (Burdick et al., 2020; Müller et al., 2021; Schifferdecker et al., 2022; Zila et al., 2021). The main protein mediating HIV-1 integration is the viral integrase, which is indispensable and sufficient to facilitate HIV-1 integration reaction, comprised of 3′-processing and DNA strand transfer, in vitro (Brown et al., 1987). HIV-1 integration process takes place within stable nucleoprotein complexes known as intasomes that are formed by a dodecameric assembly of viral IN bound to both ends of vDNA (Ballandras-Colas et al., 2017; Hare et al., 2010; Maertens et al., 2010; Passos et al., 2017, 2020). Further binding of host factors to the intasome builds the large nucleoprotein pre-integration complex (PIC) (Bowerman et al., 1989; Llano et al., 2004; Miller et al., 1997; Shun et al., 2007).

HIV-1 integrase is encoded by the pol gene and contains three independent domains: (1) The amino-terminal zinc-binding (NTD) domain that contributes to protein multimerization and to the catalytic activity, (2) the central catalytic core domain (CCD) that contains a DDE (aspartic acid – aspartic acid – glutamic acid) motif which is indispensable for the catalytic activity and (3) the carboxy-terminal domain (CTD) that has an unspecific DNA binding motif which is involved in 3' processing and strand transfer (Chen et al., 2000; Khan et al., 1990; Kulkosky et al., 1992; Leavitt et al., 1993). The first step of the integration process, called 3' end processing, is the endonucleolytic cleavage of the 3'-OH ends of vDNA (Figure 3). During the second step, the strand transfer reaction occurs in which IN uses the vDNA CAOH hydroxyl end to cut the target DNA major groove in staggered fashion followed by the 5' phosphate joining with vDNA by transesterification. Single stranded gaps are repaired through host DNA polymerase and host ligase leading to a 5 bp duplication in the host genome (Brown et al., 1987; Bushman et al., 1990; Engelman and Cherepanov, 2017; Engelman et al., 1991; Hare et al., 2010, 2012; Maertens et al., 2010; Vink and Plasterk, 1993).

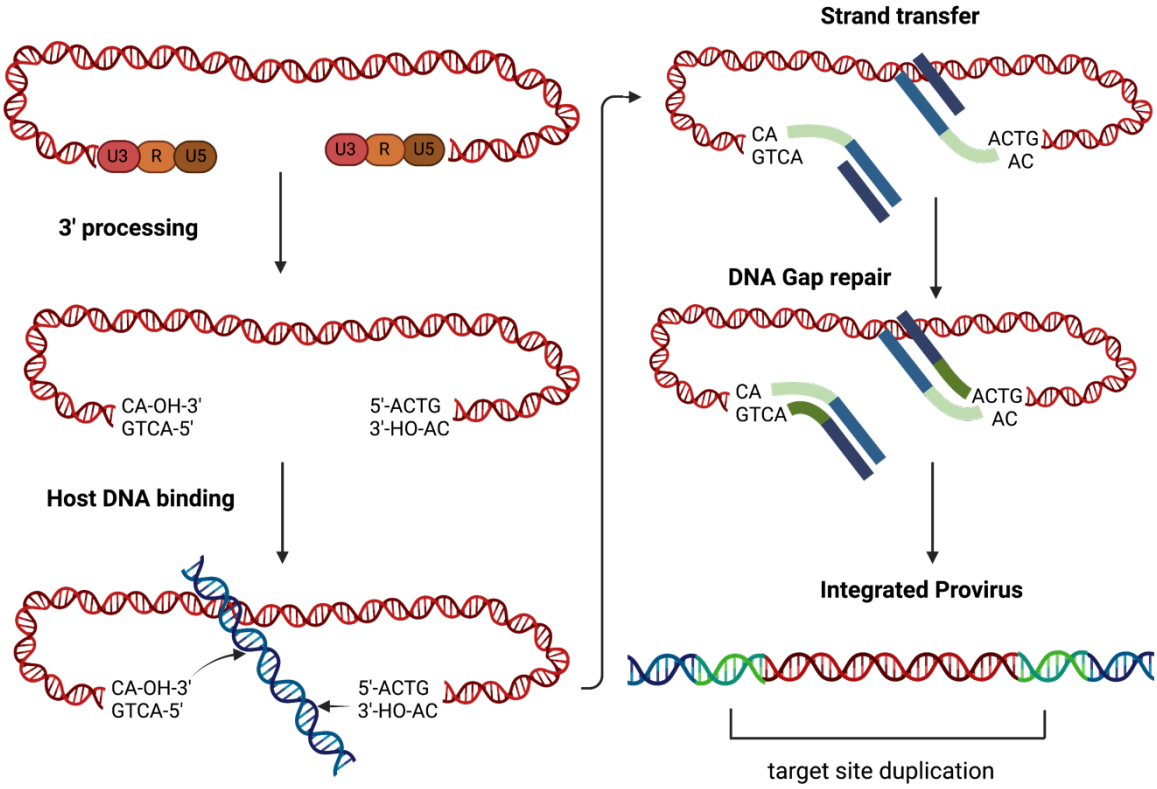


Figure 3: HIV-1 integration. HIV-1 IN is catalyzing integration of the viral cDNA into the host genome. IN binds to the LTRs of the viral cDNA through its C- and N-terminus to from the PIC. The first step of integration is the 3' processing where the 3' OH ends of the LTR are removed by endonucleolytic

cleavage. The second step is the strand transfer reaction, in which the CAOH hydroxyl group attacks the human major DNA groove. The 5' phosphate is joined with the viral DNA via transesterification and single stranded gaps are repaired through the host DNA repair machinery. Figure was created with BioRender.com.

1.1.5.1 Viral cDNA forms

The prerequisite for productive viral infection is integration of the viral cDNA. But before viral integration, cDNA can persist as either linear form of unintegrated viral DNA or 2 forms of circular unintegrated cDNA identified so far: 1- and 2- LTR circles. The main way how circular forms are build, is through autointegration of the viral cDNA into itself by ligation of the processed 3' ends with internal viral DNA (Hamid et al., 2017; Lee and Craigie, 1994). 1-LTR circles are formed by homologous recombination of the 5' and 3' LTR or interrupted RT intermediates whereas 2-LTR circles form through non-homologous end joining (NHEJ) of linear cDNA (Farnet and Haseltine, 1991; Miller et al., 1995). Several cellular host factors are involved in the formation of LTR circles. RAD50 Double Strand Break Repair Protein, MRE11 Homolog, Double Strand Break Repair Nuclease and Nijmegen breakage syndrome (NBS) 1 (NBS1) were shown to play a role in the formation of 1-LTR circles and silencing of the NHEJ factors Ku70/80, ligase IV and X-Ray Repair Cross Complementing 4 (XRCC4) impaired 2-LTR circle establishment (Sloan and Wainberg, 2011) (Figure 4).

Persistence of unintegrated HIV-1 cDNA forms is debated. Due to the absence of an origin of replication, they are not maintained during cellular division. 2-LTR levels are decreasing in the total $CD4^+$ T cell population in patients (Sharkey et al., 2005). In non-dividing cells, like macrophages, circular cDNA forms were found to persist up to 30 days, suggesting that they indeed can be maintained in slow or non-dividing cell types (Kelly et al., 2008). Circular forms of cDNA were thought to be dead-end products of viral replication cycle, but they were found to support viral transcription. The viral accessory protein vpr can promote viral transcription from unintegrated DNA without the influence of tat (Poon et al., 2007). On the contrary, unintegrated DNA forms were found to be chromatinized by core histones and linker histones early upon infection, leading to pre-integration silencing of the virus through H3K9me3 histone mark deposition, prevention of POL-II binding or compaction by SMC5/6 complex (Dupont et al., 2021; Geis and Goff, 2019; Geis et al., 2022a; Machida et al., 2020), highlighting the possible role of unintegrated DNA forms in latency establishment of HIV-1.

Furthermore, a recent study showed that 2-LTR circles can even serve as a substrate for viral integration both *in vitro* and in eukaryotic cells, suggesting that unintegrated circular forms of viral cDNA are not a dead-end product of viral replication cycle, but might rather constitute a retained supply for proviral integration (Richetta et al., 2019).

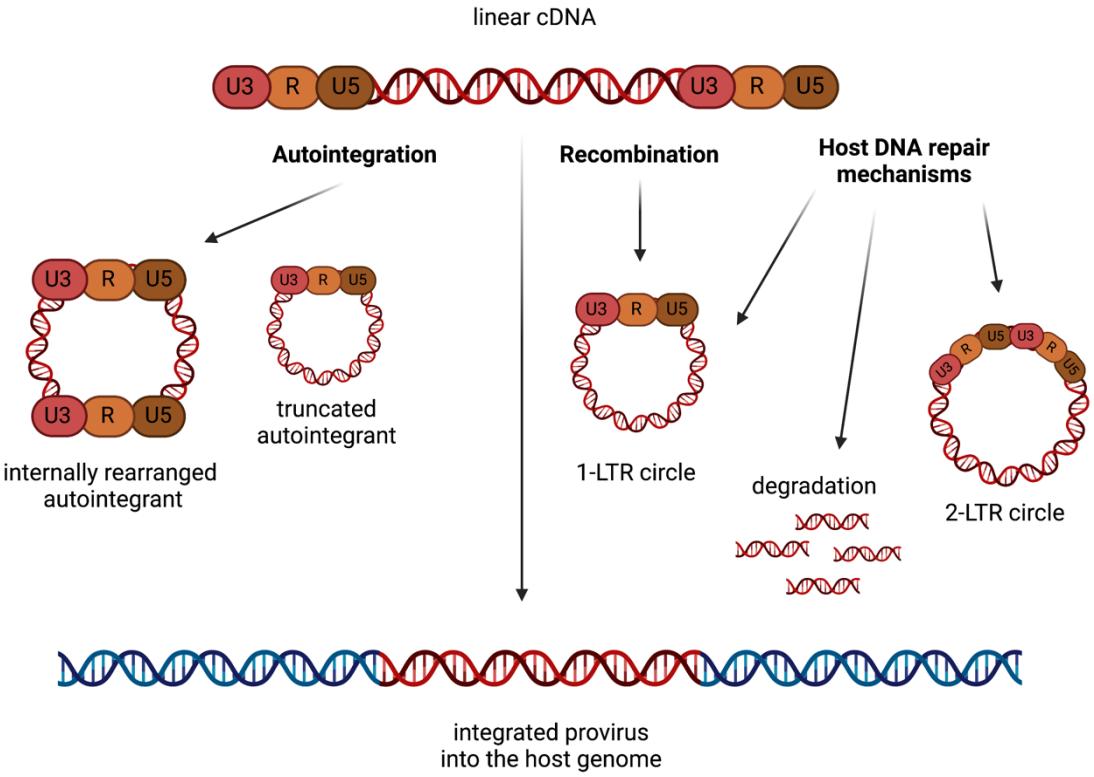


Figure 4: HIV-1 cDNA forms. HIV-1 proviral cDNA can exist in several variations. Immediately after reverse transcription, HIV-1 cDNA is present as a linear strand of DNA. Through autointegration into itself, HIV-1 can form truncated and internally rearranged circular DNA rings. When the linear viral cDNA recombines through its LTRs, 1-LTR circles are build. In addition, host factors can aid the formation of 2-LTR circles via non homologous end joining mechanisms and the host DNA repair machinery. DNA repair factors and host restriction factors can also directly degrade viral cDNA. Figure was adapted from (Sloan and Wainberg, 2011) and created with BioRender.com.

1.1.5.2 Host factors for integration

HIV-1 integration process is supported by several host factors but extensive research in the recent years has shown that HIV-1 integration targeting is largely mediated by the interaction of two virus proteins, IN and CA with cognate cellular host factors, chromatin reader lens epithelium-derived growth factor p75 (LEDGF/p75) and cleavage and polyadenylation specificity factor 6 (CPSF6), respectively (Figure 5).

Host cell protein LEDGF/p75, encoded by the PC4 and SFRS1 Interacting Protein 1 (PSIP1) gene, contains an integrase binding domain (IBD) that binds to two specific domains of viral IN: the CCD and NTD (Cherepanov et al., 2004; Hare and Cherepanov, 2009; Maertens et al., 2003). Additionally, LEDGF/p75 is a member of the hepatoma-derived growth factor (HDGF)-related protein (HRP) family, which contains an N-terminal proline- tryptophan-tryptophan-proline (PWWP) domain that binds to methylated histones H3K36me2 and 3 (Izumoto et al., 1997). As a chromatin reader protein, LEDGF/p75 binds chromatin (Cherepanov et al., 2003; Maertens et al., 2003; Nishizawa et al., 2001), recognizes methylation of lysine 36 on the Histone H3 (H3K36me3) via the PWWP domain (Eidahl et al., 2013; van Nuland et al., 2013; Pradeepa et al., 2012) and therefore tethers the PIC to the bodies of transcribed genes (Demeulemeester et al., 2015; Llano et al., 2006; De Rijck et al., 2006; Schrijvers et al., 2012; Shun et al., 2007). In addition to its tethering function, LEDGF/p75 enhances the affinity between IN and host DNA through its IBD and protects IN from proteolytic degradation (Busschots et al., 2005; Cherepanov et al., 2003; Llano et al., 2004). LEDGF/p75 knock down (KD) and knock out (KO) experiments showed that the preferred gene body integration pattern of HIV-1 shifted towards the 5' end of the genes in the absence of LEDGF/p75 (Singh et al., 2015; Sowd et al., 2016).
Although more than 200 cellular host factors were identified as IN interaction partners (Engelman and Singh, 2018), it is notable that the impact on HIV-1 integration patterns has hitherto been attributed exclusively to LEDGF/p75. Other factors that could also play a role are IN interactor 1 (INI1)/SMARCB1 (Kalpana et al., 1994), which is a component of BAF and PBAF chromatin remodeling complexes binding to the CTD of IN, and was found to mimic TAR RNA stem loop influencing later steps of the viral replication cycle (Dixit et al., 2021). Histone acetyltransferase enzyme EP300 (Cereseto et al., 2005) is another IN interactor, but it remains for further studies to investigate their possible influence on HIV-1 integration patterns.

The second important virus-host interaction contributing to integration site selection is CA with cellular partner CPSF6 which directs HIV-1 to actively transcribed euchromatic regions of the nucleus close to nuclear speckles called speckle-associated domains (SPADS) (Bedwell et al., 2021; Francis et al., 2020; Sowd et al., 2016). Furthermore, experiments with viral constructs harboring N74D and A77V mutations, which abolish CA interaction, phenocopied the CPSF6 KD IS distribution and led to integration into

gene sparse regions (Koh et al., 2013; Lee et al., 2010; Saito et al., 2016; Schaller et al., 2011; Sowd et al., 2016). Moreover, microscopy imaging of fluorescent viruses in CPSF6 depleted cells or using N74D and A77V CA mutant viruses led to an accumulation of PIC in the nuclear periphery nearby lamin associated domains (LADs) (Achuthan et al., 2018; Francis et al., 2020).

The viral CA protein also directly interacts with several nuclear and cytoplasmatic factors including RANBP2 (Lee et al., 2010; Schaller et al., 2011), NUP153 (Brass et al., 2008; Bushman et al., 2009; Matreyek and Engelman, 2011) and cyclophilin A (CypA) (Gamble et al., 1996; De Iaco and Luban, 2014; Luban et al., 1993).

While these additional factors potentially contribute to the viral integration sites selection, integrase-LEDGF/p75 and capsid-CPSF6 interactions are the most established, as they promote insertions into gene bodies and localization of the viral genome in euchromatic regions of the nucleus.

1.1.5.3 Integration site selection of HIV-1

Stable integration of the provirus into the host genome is the main determinant of retroviruses, where different genera were found integrated in different regions of the host genome. While gamma-retroviruses like MLV were found at DNA-hypersensitivity and promoter sites (LaFave et al., 2014; Lewinski et al., 2006; Panet and Cedar, 1977; De Ravin et al., 2014; Rohdewohld et al., 1987; Vijaya et al., 1986; Wu et al., 2003) and delta-retroviruses like HTLV-1 preferably around TSS, promoters and transcription regulatory elements (Gillet et al., 2011), alpha- and beta- retroviruses showed nearly random IS distribution (Faschinger et al., 2008; Mitchell et al., 2004; Narezkina et al., 2004).

Large sequencing studies of virus- genome junctions in the recent years substantially contributed to the understanding of the integration site selection of the lentivirus HIV-1. From the first integration sites (IS) obtained in HIV-1 infected SupT1 cells (Schroder et al., 2002) up to date, large sets of genomic data were generated in relevant HIV-1 target cell lines, primary T cells and in patient material (Barr et al., 2006; Bedwell et al., 2021; Cohn et al., 2015; Einkauf et al., 2022; Lucic et al., 2019; Maldarelli et al., 2014; Sherrill-Mix et al., 2013; Wagner et al., 2014; Wang et al., 2007). These studies showed that HIV-1 integration is a non-random process incorporating elements of chromatin structure and function, ranging from DNA cis-elements to the three-dimensional (3D) genome organization. Indeed, HIV-1 mostly integrates into actively

transcribed regions of the genome, in the vicinity of activating epigenetic marks as H3K36me3, H3K4me1 or H4K16ac, CpG islands and intronic regions and away from repressive marks H3K27me3 and H3K9me3 (Figure 5) (Marini et al., 2015; Roth et al., 2011; Wang et al., 2007). Although it seems that HIV-1 does not target any particular genomic sequence, it is known to integrate near Alu repeats (Cohn et al., 2015; Schroder et al., 2002). The integration motif (INT-motif) TG-(N)5-7-CA has been found in association with HIV-1 integration sites near Alu repeats, but not a consensus sequence (Brady et al., 2009a; Holman and Coffin, 2005; Serrao et al., 2014). Cohn et al. found a 30 bp INT-motif within 100 bp of 7 % of all integrations, which was located at the 3' end of Alu elements (Cohn et al., 2015). Furthermore, enrichment of HIV-1 integration in Alu repeats was found, both in genic and intergenic regions, suggesting that HIV-1 integration has a bias towards Alu repeats independent of its preference for integration into genes.

HIV-1 integration takes place in a complex setting of the human genome, packed into the nucleus, which is surrounded by the inner and outer nuclear envelope. The inner envelope is associated with lamins and other proteins, forming the nuclear lamina (Pombo and Dillon, 2015). Chromatin forms contacts with the surface of the nuclear lamina, containing mainly repressed chromatin in the lamin-associated domains (LADs) and open chromatin marks underneath the NPCs (Ibarra et al., 2016). The HIV-1 provirus was found excluded from LADs and associated with several proteins of the NPC, where KD of NUP153 led to diminished integration levels and a shift of integrations to the interior of the T cell nucleus (Marini et al., 2015), suggesting that the nuclear periphery is an important determinant for HIV-1 integration.

Although no sequence specificity was found for HIV-1, ISs are enriched in certain genomic regions over others. Recent efforts to characterize HIV-1 integration repertoires showed that HIV-1 recurrently integrates into certain genes (Bedwell et al., 2021; Lucic et al., 2019). In one study, these recurrent integration genes (RIGs) were defined by a gene being present in at least two different integration studies both *in vitro* and in patients (Lucic et al., 2019). Another way of characterizing recurrently targeted genes was done by comparing across 10 different IS studies (both patient and in vitro studies) which genes were experimentally targeted more or less frequently than expected by chance (Bedwell et al., 2021). This analysis showed that RIGs and RAGs (recurrently avoided genes) are rather similar between in vitro studies, ART treated

and untreated patient IS data sets, suggesting that HIV-1 frequently targets a subset of genes independent of HIV-1 target cell background (Figure 5).

Robust genomic analysis of HIV-1 hot-spots of integration in primary $CD4^+$ T cells together with Hi-C in Jurkat cells revealed that genes with HIV-1 insertions build 3D clusters in close proximity to enhancer and super-enhancer (SE) regions (genomic regions enriched in H3K27ac, H3K4me1 and transcription factor binding sites (TF)). This enrichment in the active sub-compartment of the nucleus was still preserved when corrected for gene expression, suggesting that nuclear architecture could have a critical role in integration site selection (Figure 5) (Lucic et al., 2019). As super-enhancers are known to regulate cell identity genes (Hnisz et al 2013), the association of ISs with super-enhancers might be cell-type specific and might lead to the integration into cell-type specific genes. Furthermore, it was shown that HIV-1 targets speckle associated domains (SPADs, 40% overlap of IS with SPADs), which contain a high proportion of SE (70 %) (Bedwell et al., 2021; Chen et al., 2018b; Francis et al., 2020). The only available SPAD data set so far has been generated in K562 cells (Chen et al., 2018b) and further studies in T cells are needed to deconvolute the influence of SPADs and enhancers on HIV-1 integration.

In HIV-1 IS research it is not only important to study the actual IS but also information about the corresponding proviral sequence is crucial. In particular, when studying patient data and latent reservoirs, the intactness of the viral sequence can provide insights if the respective insertion could potentially give rise to a functional viral particle to fuel new infections. A recent study analyzed viral transcription, integration site and viral genome sequence from the same cell in ART treated patients. They could show that intact provirus was found more frequently in non-genic regions than defective ones. Furthermore, by integrating their sequencing data with reference chromatin immunoprecipitation followed by sequencing (ChIP-Seq), Assay for Transposase-Accessible Chromatin using sequencing (ATAC-Seq) and 3D genome contact Hi-C data of primary T cells they could show that non-genic regions harboring IS showed an increased distance to frequently interacting regions (FIREs), as determined by Hi-C (Einkauf et al., 2022). In addition, regions with active integration sites, meaning with a baseline viral expression, had higher contact frequencies in 3D to active regions as characterized by open chromatin marks, ATAC-Seq and RNA-Seq signal.

Over the recent years, research studies indicate a structure-function relationship between integration sites, the encompassing 3D genome, and the expression of the

HIV-1 provirus. This underscores the potential significance of 3D chromatin and cis genomic elements in influencing the selection of HIV-1 integration sites.

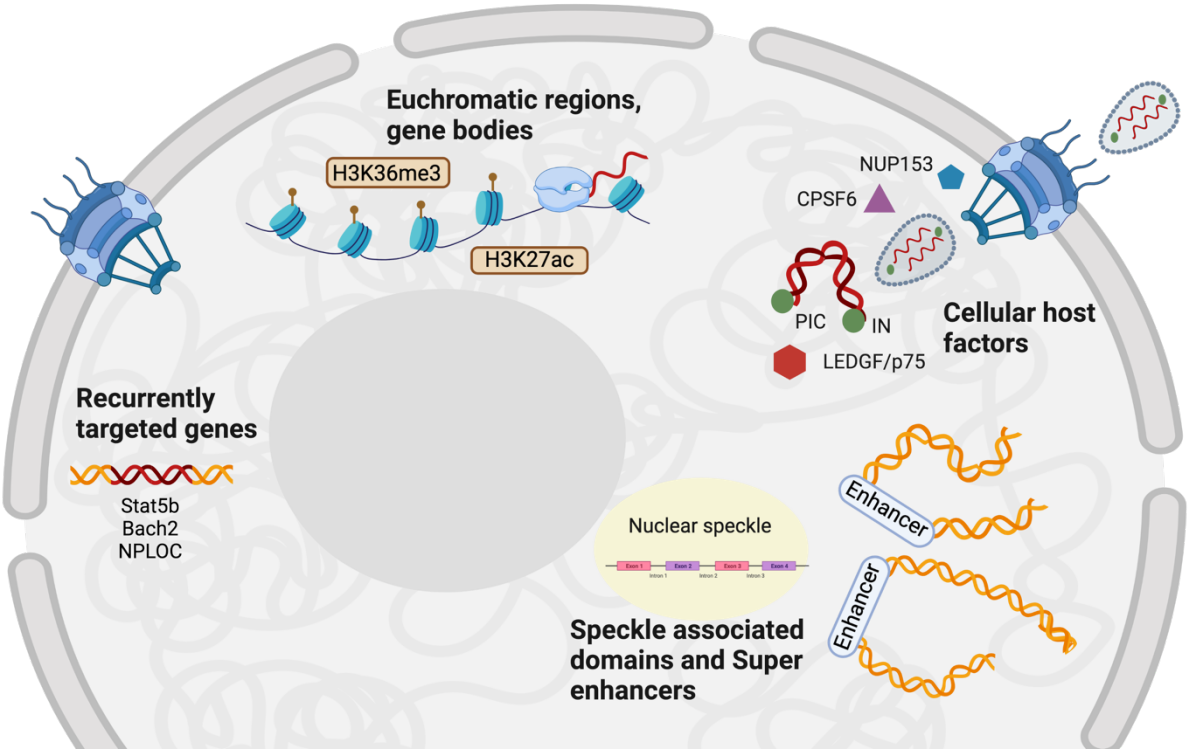


Figure 5: HIV-1 integration determinants. HIV-1 integration and IS landscape is determined by several viral and host factors. It is known that HIV-1 integration is dependent on LEDGF/p75 and CPSF6. On the chromatin level, provirus integrates into euchromatic regions and actively transcribing genes marked by H3K36me3 and H3K27ac. Integration sites are frequently found in certain genes e.g. STAT5b, BACH2 or NPLOC. In recent years it evolved that HIV-1 proviruses frequently reside in proximity to several subnuclear compartments such as for example PML nuclear bodies and nuclear speckle associated domains. In addition, recurrently targeted HIV-1 genes were found to cluster together in the 3D nuclear periphery of the A1 subcompartment of the T cell nucleus close to super enhancers. Figure was prepared with BioRender.com.

1.1.6 HIV-1 transcription

Once the viral DNA is integrated into the host genome, the proviral genome is either expressed by the cellular machinery, or transcriptionally silenced leading to the establishment of post-integration latency (described in the following chapter). The HIV-1 LTR is composed of the transcription start side (TSS), the TATA box and the SP1 enhancer region, containing binding sites for the constitutive TF SP1, and the inducible transcription factors activator protein-1 (AP-1), nuclear factor of activated T-cells (NFAT) and nuclear factor κb (NF-κb) (Dutilleul et al., 2020; Ne et al., 2018). The viral genome is also chromatinized upon integration with nuc-0, nuc-1 and nuc-2

nucleosomes precisely positioned at the viral LTR regions around the TSS and having important roles in TF recruitment and transcriptional regulation of HIV-1 (Lusic et al., 2003; Verdin et al., 1993).

The first genes expressed of the viral genome are tat, rev and nef transcripts. These mRNAs are fully spliced, exported to the cytoplasm, translated, and imported to the nucleus where they regulate expression of the full genome. The remaining transcripts of the HIV-1 genome are either unspliced or partially spliced (Ocwieja et al., 2012; Purcell and Martin, 1993). The unspliced transcripts are recognized by the viral protein rev via the rev responsive element (RRE), regulating transcript export (Malim et al., 1989).

One important factor enhancing HIV-1 transcription is the transactivator protein tat. It regulates transcription via binding to the transactivation response element (TAR) which is formed at the 5' end of viral transcripts. The cellular RNA polymerase II (POL-II) pauses after TAR RNA loop in the absence of tat. For the full activation of the viral LTR, several factors need to be recruited to the regulatory region upstream of the LTR including AP-1, NFAT, Upstream stimulatory factor 1 (USF), Ets, Lymphoid enhancer-binding factor 1 (LEF-1) and NF-κb. Furthermore, tat recruits Positive Transcription Elongation Factor-b (P-TEFb) to the LTR, consisting of cyclin dependent kinase 9 (CDK9) and cyclin T1 (CycT1), via disruption of the inactive P-TEFb:7SK/snRNP/Hexim1 complex. It was found that P-TEFb is recruited to the LTR as part of the larger super elongation complex (SEC), consisting of scaffolding subunits AFF1 and AFF4 and transcription elongation factors ELL1 and 2 (He et al., 2010; Sobhian et al., 2010). The recruitment together with the SEC is necessary for the full tat-dependent transcriptional activation of the LTR (Lu et al., 2014). P-TEFb/CDK9 phosphorylates the CTD of polymerase II at Ser2, which leads to transcription elongation. Tat competes with Hexim1 for CycT1 binding. If the level of tat is high, HIV-1 transcription continues even if there are high levels of 7SK snRNP (Herrmann and Rice, 1995; Kinoshita et al., 1997; Nabel and Baltimore, 1987; Reines et al., 1996; Tong-Starksen et al., 1987; Wei et al., 1998). The availability of the LTR activating components is crucial for viral transcription to occur.

1.1.7 HIV-1 viral latency

On the transcriptional level, HIV-1 infection can have two different outcomes; either the virus gets transcribed and new viral particles are assembled or the virus becomes

silenced and goes into post-integration latency, which is a state where the virus is transcriptionally silent but replication competent and can be reactivated pharmacologically or naturally (Halvas et al., 2020; Lusic and Siliciano, 2017).

As HIV-1 becomes an intrinsic component of the host genome, there is so far no treatment that targets the stably integrated provirus and therefore latency is still a bottleneck on the way to an HIV-1 cure.

In case of HIV-1, post-integration latency is attributed to the transcriptionally silent and highly stable proviral reservoir formed mostly in resting memory $CD4^+$ T cells and macrophages. Apart from post-integration latency, latency can be established before the PIC integrates into the host genome and is achieved by incomplete RT due to a lack of dNTPs or chromatinization and epigenetic silencing of unintegrated HIV-1 DNA forms (Geis and Goff, 2019; Geis et al., 2022b; Pierson et al., 2002; Swiggard et al., 2005; Wang et al., 2016a). A recent study suggested that SMC5/6 complex is contributing to epigenetic silencing of unintegrated HIV-1 DNA by SUMOylation (Irwan et al., 2022).

Post-integration latency is a multifactorial process, and there are several mechanisms that block viral transcription after integration into the host genome, thus leading to viral latency. In transcriptional interference, HIV-1 integration in cis to the host genome can lead to promoter occlusion, and convergent antisense transcription of the viral genome can also inhibit successful transcription (Lenasi et al., 2008). Latency can be established also through decreased transcription factor (TF) binding, either because they are defective or excluded from nucleus during the resting $CD4^+$ T cell state (Marcello et al., 2003; Sabo et al., 2008; Williams et al., 2007). Other mechanisms contribute to the latency establishment such as proviral repression through chromatin silencing by H3K27me3 or H3K9me3 (Chéné et al., 2007; Friedman et al., 2011; Imai et al., 2010) or CpG promoter methylation (Blazkova et al., 2009; Kauder et al., 2009). Transcriptionally repressed provirus was found distal from activating chromatin marks (Chen et al., 2017; Vansant et al., 2020).

Heterochromatin is maintained through histone modifications as for example methylation or acetylation (Zhang et al., 2015). Though, the heterochromatic environment can be decondensed through histone deacetylase inhibitors (HDAC), histone methyltransferase inhibitors (HMT's) or inhibitors of the Bromo- and Extra-Terminal domain (BET) family (chromatin reader proteins of acetylated lysines) and therefore the viral replication is reactivated (Colin and Van Lint, 2009; Lucic and Lusic,

2016). Intensive research in the last decade showed that latency reversing agents (LRAs) do not achieve a full reactivation of the reservoirs, that they are less effective in vivo than ex vivo and that they can have strong side effects due to a global T cell boost (Archin et al., 2012; Bullen et al., 2014; Kulkosky et al., 2002; Shan et al., 2012). Therefore, the 'shock and kill' approach for HIV-1 cure, where transcriptionally activated provirus is subsequently targeted with ART, is not fully effective to eradicate viral infection. Another approach to achieve a functional cure is the ‘block and lock’ strategy, where the idea is to suppress viral transcription to silence HIV-1, preventing viral rebound. Two examples of small molecules to achieve this are Didehydro-cortistatin A (dCA) which is a tat inhibitor or LEDGINs, which are small inhibitors of the LEDGF/p75- IN interaction. dCA was shown to inhibit tat and to promote a tight interaction of nucleosome and viral DNA by increasing H3 deacetylation (Li et al., 2019; Mousseau et al., 2015). By inhibiting the interaction of LEDGF/p75 and IN, viral integration was substantially reduced and at the same time the remaining IS were shifted to genomic sites with less susceptibility of reactivation, ‘locking’ them in viral sanctuaries (Christ et al., 2010; Debyser et al., 2018; Vranckx et al., 2016). Nonetheless, also with this concept, scientists faced the problem of not reaching and affecting the full latent reservoir, due to the high heterogeneity (Clark et al., 2023; Dufour et al., 2023; Wu et al., 2023).

Apart from the 'shock and kill' and ‘block and lock’ approaches, genome-editing based methods to eliminate the latent reservoir have been used in the past decade. As an integral part of the human genome, latent proviral DNA can be directly targeted or the host factors necessary for HIV-1 replication can be disrupted. Editing via zinc finger nuclease- targeted disruption of the HIV-1 co-receptor CCR5 was used to introduce double strand breaks into the CCR5 receptor gene (Perez et al., 2008). In an alternative approach, CRISPR/Cas9 system is used to design gRNAs against the viral LTRs to eliminate the viral genome from the host cell (Cong et al., 2013; Ebina et al., 2013; Li et al., 2015, 2022; Mali et al., 2013). However, despite the promising research results it is still unclear if the high specificity and efficiency in vitro can be transferred into the clinics. In addition to the high risk of off target effects, the accessibility of the latent reservoir in vivo is still unclear. For the improvement of currently available HIV-1 treatments (functional cure), and with the goal to achieve a sterilizing cure, it is crucial to get a better understanding of how latency is established and maintained. Despite the effort that has been made in this direction, it is still not clear whether HIV-1 actively

selects the genes into which it integrates, what are their potential regulatory features and how the integration sites contribute to the latency state.

1.1.8 HIV-1 reservoirs and clonal expansion

Latent cells can build latent reservoirs throughout the body. A reservoir can be defined as a cell type or anatomical site that allows longer-term persistence of replication-competent virus in the setting of optimal cART (Sengupta and Siliciano, 2018). The most known and well-studied HIV-1 reservoir are long-lived memory $CD4^+$ T cells of resting phenotype that are main barrier to cure. The proposed, and well accepted mechanism how they become a reservoir is that activated infected T cells revert to a resting memory state and through heterochromatin formation, transcription factor sequestration and lack of dNTPs for reverse transcription the virus becomes silenced (Lee and Lichterfeld, 2016; Wong et al., 2019). In addition, resting primary $CD4^+$ T cells can also directly establish latent infection (Agosto et al., 2018; Cameron et al., 2010; Chavez et al., 2015; Moso et al., 2019; Pace et al., 2012). Apart from $CD4^+$ T cells, macrophages, gut associated lymphoid tissue (GALT) and cells of the central nervous system (CNS) represent plausible latent reservoirs due to their epigenetic and metabolic characteristics (Kruize and Kootstra, 2019; Marban et al., 2016; Veazey et al., 1998; Yukl et al., 2010).

Viral reservoirs are established early during acute infection (Chun et al., 1998) and the virus can persist in the body for a long time without any noticeable symptoms, when the patient is under antiretroviral therapy. However, when the therapy is interrupted, the virus load rebounds and the patient can develop AIDS (Broder, 2010). Long term persistence of the resistant latent reservoir cells is achieved by clonal expansion of infected cells. This process can be initiated by antigen-driven proliferation (Gantner et al., 2020; Mendoza et al., 2020; Simonetti et al., 2021), homeostatic proliferation of infected latent T cells (Chomont et al., 2009) or the integration site and enhanced cell proliferation driven by insertional mutagenesis into proliferation genes (Cesana et al., 2017; Cohn et al., 2015; Liu et al., 2020; Maldarelli et al., 2014; Wagner et al., 2014) (Figure 6).

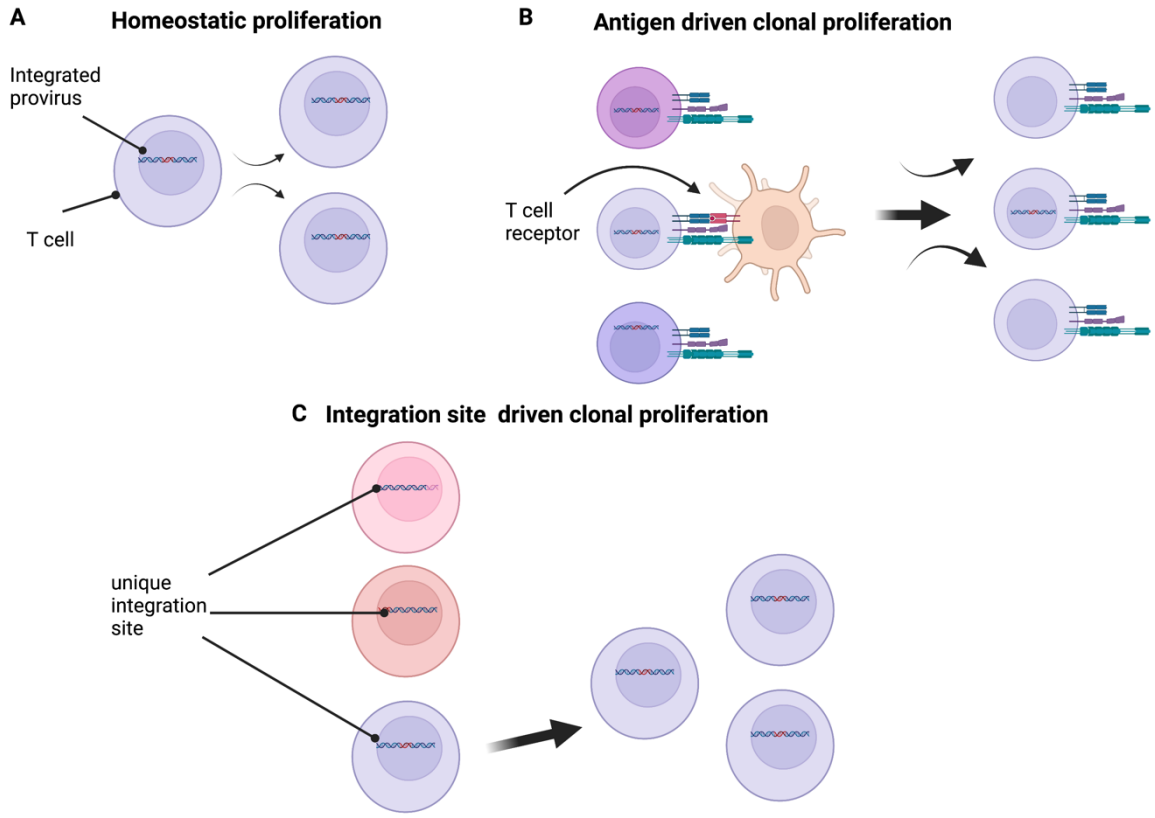


Figure 6: Mechanisms of clonal expansion. (A) $CD4^+$ T cells harboring integrated provirus are proliferating and maintained by cell division. (B) Antigen binding to the T cell receptor stimulates proliferation of infected T cells generating large T cell clones. (C) Integration into cell survival genes can also lead to integration site driven T cell proliferation. Figure was adapted from (Lau et al., 2021) and created with BioRender.com.

The identification of a significant proportion of chronically infected cells within patients receiving ART as constituents of expansive cellular clones has substantially altered our perspective on the reservoir's composition and its origin. In support of integration site driven clonal expansion, it was shown that IS selection was closely linked to viral persistence. Longitudinal patient studies showed several independent integration sites in the same gene as well as identical insertions derived from different cells (most prominent genes detected in multiple patient studies were BACH2, STAT5B, MKL2) upstream of the translation initiation site of these genes or in introns between two exons (Cesana et al., 2017; Coffin et al., 2021; Cohn et al., 2015; Maldarelli et al., 2014; Wagner et al., 2014). The in vivo selection process might be influenced by the fact that these genes are involved in cell growth or cell cycle regulation and that their dysregulation (altered gene expression levels or expression of altered protein forms) is associated with cancer (Scholz et al., 2017). These findings suggest a clonal

expansion phenotype of a cell bearing the particular proviral insertion that leads to viral persistence (Cohn et al., 2015; Maldarelli et al., 2014; Wagner et al., 2014).

It was found in patient studies over the last years that not all latent reservoirs are transcriptionally silent and that the active IS clones make up 30% of all proviruses in people living with HIV (PLWH). Sequencing of the IS together with the viral sequence and the viral expression levels revealed that RNA expression is associated with active epigenetic marks both in cis and in 3D (Einkauf et al., 2022).

Under ART, the majority of integrated virus is defective due to deletions and hypermutations and the defective viruses accumulate over time in patients under ART (Bruner et al., 2016; Cohn et al., 2015; Ho et al., 2013; Simonetti et al., 2016). While most of the clonally expanded cells contain non-intact proviruses, even the small fraction of intact transcriptionally active proviruses can fuel viral rebound upon treatment interruption (Aamer et al., 2020; Bui et al., 2017; Cole et al., 2022; Hosmane et al., 2017; Lorenzi et al., 2016; Simonetti et al., 2016). Among HIV-1 infected patients a unique population exists which can spontaneously control HIV-1 infection without treatment. These patients are called elite controllers (EC) because they have undetectable levels of viremia (Gonzalo-Gil et al., 2017). The mechanisms how they control infection are poorly understood, and further studies of these individuals could give important insights into how to achieve a functional cure of HIV-1. Many EC were found to have host gene polymorphisms in the CCR5-Δ32 gene which renders cells not permissive for viruses using CCR5 co-receptor for viral entry (Antoni et al., 2013). A recent study focusing on the IS in the latent reservoir in EC found that EC have more IS in the silent ‘deep sleep’ regions of the genome like non-coding centromeric satellite DNA (Chomont, 2020; Jiang et al., 2020). Intact viral insertions were found located closer to the repressive chromatin mark H3K9me3 and further away from the enhancer mark H3K4me1 (Jiang et al., 2020), suggesting that these dense heterochromatin regions can contribute to the st ate of deep latency (Lichterfeld et al., 2022).

Intact integrations into non-silent euchromatic regions of the genome leading to production of viral transcripts results in sensing by the cell intrinsic immunorecognition and clearing by the immune system (Li et al., 2016). As a consequence, a deep latency reservoir is selected over time (Jiang et al., 2020). Additionally, intact viral insertions decline faster over time in patients than defective viral clones, possibly due to a more sensitive recognition by the immune system (Cho et al., 2022; Falcinelli et al., 2021; Garcia-Broncano et al., 2019; Peluso et al., 2020; Pinzone et al., 2019). All together

these studies highlight a prominent role of the insertion site and the resulting viral transcription level on viral persistence and clonal expansion (Cole et al., 2022; Einkauf et al., 2019, 2022; Huang et al., 2021). The establishment of a viral deep latent reservoir in EC to control viral infection supports a 'block and lock' like strategy to achieve a functional HIV-1 cure.

Due to the large heterogeneity and rarity of the latent reservoirs, in the recent year patient studies focused on advanced and costly single cell methods to investigate reservoir composition. Focused interrogation of cells by nucleic acid detection and sequencing (FIND-Seq) is a microfluidic based technology performing RNA-Seq only on cells harboring HIV-1 DNA. Host cell transcriptomic profiles of HIV DNA$^+$ cells were highly diverse but showed shared gene expression patterns of resistance to cell death and resistance to anti-proliferative signaling, likely promoting persistence of infected clones (Clark et al., 2023). ATAC-Seq in combination select antigen profiling by sequencing (ASAP-Seq) revealed a large variation of surface and epigenetic markers between and across infected memory CD4$^+$ T cells (Wu et al., 2023). Another recent study on the phenotype of latent reservoir cells showed a large phenotypic diversity of latent CD4$^+$ T cells harboring the same intact HIV-1 genome (Dufour et al., 2023), further suggesting that it is difficult to find a druggable latency marker for HIV-1.

1.2 Microglia as latent HIV-1 reservoir

1.2.1 Microglia

The main and well-studied HIV-1 latent reservoir are memory CD4$^+$ T cells of a resting phenotype. Nonetheless, latent reservoirs can also be established within other tissue compartments expressing CD4 receptor. Notably, HIV-1 invades the central nervous system (CNS) early in infection, leading to HIV-associated neurological disorders (HAND) in up to 50% of individuals receiving ART (Davis et al., 1992; Heaton et al., 2011; Rappaport and Volsky, 2015). Studies using laser microdissection coupled with polymerase chain reaction (PCR) have detected integrated HIV-1 provirus in perivascular macrophages, astrocytes, and microglia derived from HIV-1 patients (Churchill et al., 2006; Thompson et al., 2011). However, a recent study raised doubts about the presence of HIV-1 DNA in astrocytes of ART-treated patients (Ko et al., 2019). Conversely, evidence from the past three decades strongly supports the in vivo and in vitro replication of HIV-1 in microglia (Albright et al., 2000; Cosenza et al., 2002;

Edén et al., 2010, 2016; Ferretti et al., 2015; Garcia-Mesa et al., 2017; McCarthy et al., 1998; Neuen-Jacob et al., 1993; Stoler et al., 1986; Tang et al., 2023; Wiley et al., 1986), indicating that microglia, brain resident mononuclear phagocytes crucial for the innate immune response in the brain, are the main CNS cell type supporting HIV-1 replication.

Microglia cells originate from the yolk sac and migrate to the brain during embryogenesis, constituting approximately 10% of the cellular population in the brain (Ginhoux et al., 2010). They play a role in tissue structure, dynamics, and provide innate immune protection against microorganisms and physical trauma in the brain (Ginhoux et al., 2010; Sousa et al., 2018). When activated by pathogens or inflammatory signals in their environment (Davalos et al., 2005; Nimmerjahn et al., 2005), they release various cytokines, chemokines, and neurotoxic proteins, both pro-inflammatory (e.g., tumor necrosis factor α (TNFα), interleukin 1β (IL1β), interleukin 10 (IL10)) and anti-inflammatory (e.g., interferon α/β (IFNα/β)) (Lee et al., 2002). P2Y purinoceptor 12 (P2Y12) and Transmembrane Protein 119 (TREM119) serve as molecular markers on the surface of microglia cells, and they are involved in tissue remodeling, clearance of neurons, pathogens, and cellular debris (Böttcher et al., 2019; Priller and Prinz, 2019; Spittau, 2017).

Microglia cells get infected either via cell-free viral infection or through transmigration of infected T lymphocytes or monocytes (Murooka et al., 2012; Veenstra et al., 2017). Several studies suggest that microglia present a potential latent reservoir for HIV-1 given their resistance to apoptosis, their year-long lifespan and self-renewal properties (Bilimoria and Stevens, 2015; Hashimoto et al., 2013; Kumar et al., 2014; Tay et al., 2017). Anatomically, microglia cells are embedded in the complex brain tissue and are thus protected via the blood brain barrier, making infected cells difficult to target by ART, contributing to viral persistence. Indeed, recent studies support the existence of replication competent viral reservoirs in the CNS also during ART (Avalos et al., 2017; Joseph et al., 2019; Tang et al., 2023).

1.2.2 HIV associated neurological disorders and latency in microglia cells

Untreated HIV-1 infection can lead to brain damage, HIV associated encephalitis (HIVE) or dementia. Moreover, even individuals on successful ART with undetectable plasma viral load develop pathological neuroinflammation, which leads to HAND. Whether symptoms arise directly due to viral activity, are mediated indirectly through

systemic inflammatory processes, or emerge as a result of a complex interplay between viral presence and systemic inflammation is still under debate. Non-viral mechanisms supporting HAND development include high correlation of systemic and CNS inflammation with HAND in patients infected with HIV-1 (Livelli et al., 2019; Swanta et al., 2020; Williams et al., 2021). The release of neurotoxins and reactive oxygen species (ROS) causing neuroinflammation was both observed in patient brains and ex vivo microglia infections (Alvarez-Carbonell et al., 2017, 2019; Garcia-Mesa et al., 2017; Ginsberg et al., 2018; Mahajan et al., 2021). Furthermore, it was shown that the viral proteins tat, nef and envelop glycoprotein gp120 can act directly as neurotoxins leading to microglia activation (Kraft-Terry et al., 2009). Gene expression analysis of post mortem brain samples from the National NeuroHIV Tissue consortium revealed an increased expression of proinflammatory cytokine and interferon signaling pathway genes in patients with high viral load (Sanna et al., 2021) (Figure 7).

One mechanism proposed for virus induced establishment of HAND is the periodic emergence of HIV-1 from latency in microglia cells (Alvarez-Carbonell et al., 2017, 2019). Cytokines elevated in inflammation were also found to be able to reactivate HIV-1 and microglia (Alvarez-Carbonell et al., 2017, 2019). Several factors have been proposed to contribute to latency reactivation in microglia cells including neuronal damage or inflammatory signals (Dahl et al., 2014; Eggers et al., 2017; Yadav and Collman, 2009). Toll-like receptor 3 (TLR3) and the glucocorticoid receptor were shown to be crucial for HIV-1 activation in infected microglia, leading to translocation of NF-κB p65/p50 form the cytosol to the nucleus binding to the viral LTR (Alvarez-Carbonell et al., 2017, 2019; Garcia-Mesa et al., 2017). In a co-culture of induced pluripotent stem cell (iPSC) derived microglia and neurons, neuronal damage led to HIV-1 reactivation from latency (Alvarez-Carbonell et al., 2020). On the contrary, binding of nuclear hormone receptor family protein Nurr1 and epigenetic silencing by histone methyltransferases results in silencing of HIV-1 provirus (Nguyen et al., 2021; Ye et al., 2022). Interestingly, a recent study highlighted the role of microglia as sites of membrane-associated amyloid precursor protein (APP) that, when cleaved by HIV-1 Gag, deposits beta-amyloid (Aβ), which is a classic hallmark of Alzheimer's disease (AD) and dementia (Chai et al., 2017).

Moreover, HIV-1 RNA and DNA persistence in cerebrospinal fluid (CSF) samples, including microglia-like cells, has been linked to neurocognitive decline under virological suppression (Farhadian et al., 2018; Spudich et al., 2019). It could well be

that low levels of HIV-1 RNA detected in CSF and plasma from patients on ART support chronic inflammation, and together with ART toxicity, microglial activation and neuroinflammation lead to HAND in PLWH (Anderson et al., 2017; Ginsberg et al., 2018; Robertson et al., 2020; Spudich et al., 2019; Vera et al., 2016). Another important contributor to HAND development in PLWH is substance abuse, which is a common comorbidity of PLWH, also leading to poor ART adherence. Methamphetamine, cocaine or opioid abuse was found to influence BBB integrity enhancing neuroinflammation, neuronal injury and cognitive disorders (Chilunda et al., 2019). These clinical implications underscore the significance of further research on the primary HIV-1 latent reservoir in the CNS.

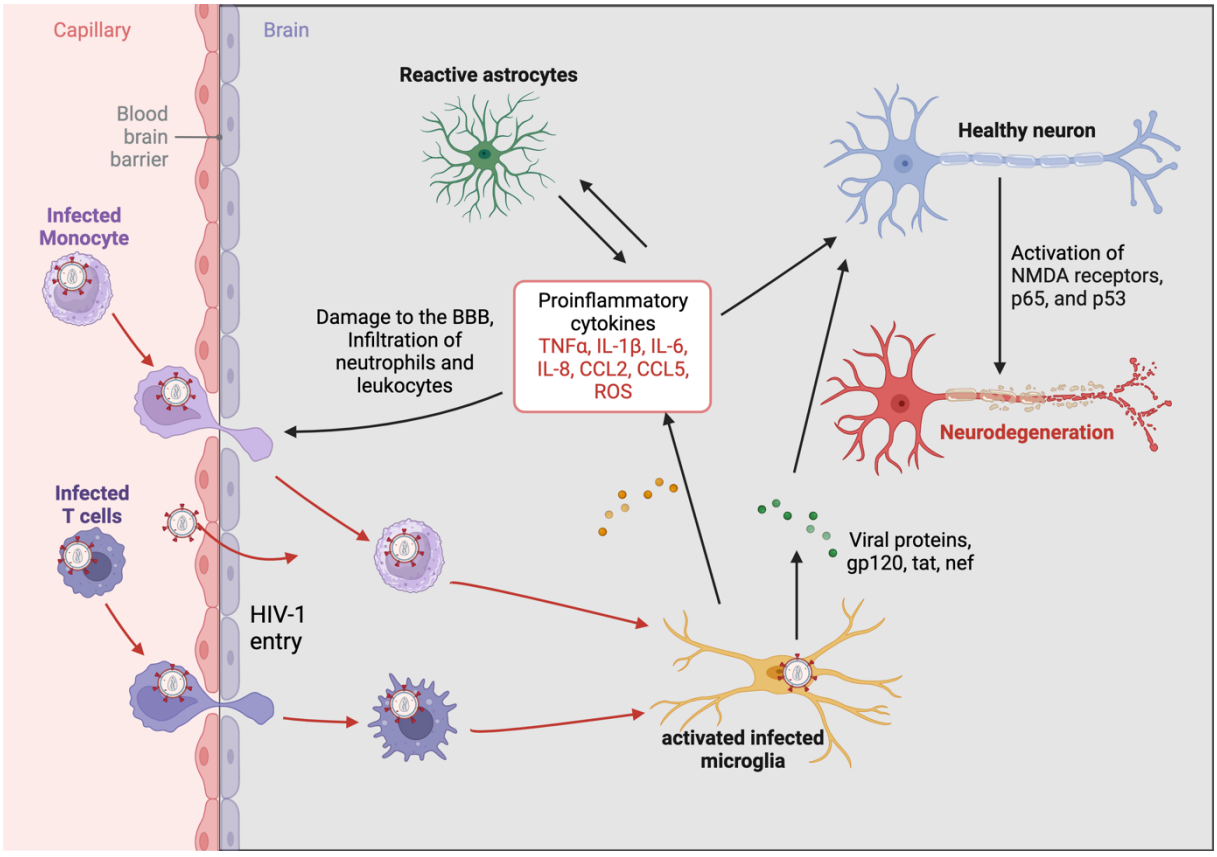


Figure 7: Implications causing HIV associated disorders in the brain. HIV-1 enters the brain via infected monocytes or lymphocytes or directly via leaky blood brain barrier and can then infect microglia. Infected microglia with replicating HIV-1 become activated and release proinflammatory cytokines like TNFα, IL-1β, IL-6, IL-8, CCL2, CCL5 and ROS. These cytokines lead to the activation of astrocytes which in response further release proinflammatory cytokines. Proinflammatory cytokines cause neuronal damage by activation of NMDA receptors, increasing neurodegeneration in the brain. In addition, viral proteins like gp120, tat and nef can directly cause neuronal damage, augmenting neuroinflammation in the brain. Figure was adapted from (Sreeram et al., 2022) and created with BioRender.com.

1.3 Nuclear architecture and 3D genome organization

1.3.1 DNA compaction and histone modifications

Human DNA is approximately 2 meter in length and has to be tightly packaged in the mammalian nucleus. Therefore, 147 bp of DNA is wrapped around a histone core made of two H3-H4 and two H2A– H2B histone dimers to form the nucleosome octamer. Nucleosomes are connected by linker DNA (~ 10-80 bp) and organized in a string like structure. The nucleosome fiber is further condensed into higher-order structures forming the chromosome (Alberts, Bruce; Heald, Rebecca; Johnson, 2022). Nucleosome position can influence transcription due to different DNA accessibility for transcription factors and POL-II. Chromatin remodelers can therefore move nucleosomes along the DNA and change their positioning. Histones do not only influence DNA compaction but also gene expression by histone displacement and histone tail amino acid residue modifications. Histones can be posttranslationally modified by a plethora of different modifications and mechanisms, most widely studied being acetylation, methylation, phosphorylation, crotonylation, ubiquitinylation or SUMOylation (Alberts, Bruce; Heald, Rebecca; Johnson, 2022; Fischle et al., 2005; Galisson et al., 2011; Hendriks et al., 2014; Joseph et al., 2022; Tan et al., 2011; Tsukiyama, 2002).

1.3.1.1 Histone acetylation

Histone acetylation is catalyzed by histone acetyl transferases (HATs) which add a negatively charged acetyl group to lysins on the histone tails, mostly H3 and H4. This diminishes the electrostatic affinity between DNA and histones leading to a greater permissiveness for transcription (Brownell et al., 1996; Kouzarides, 2007; Shahbazian and Grunstein, 2007). HATs can be classified into Gcn5-related N-acetyltransferases (GNAT), MYST and p300/CBP protein families (Berndsen and Denu, 2008; Marmorstein and Roth, 2001; Sterner and Berger, 2000). Acetyl groups are removed from histones by histone deacetylases (HDACs) that are classified into zinc-dependent class I, II and IV and NAD-dependent class III protein family (Haberland et al., 2009). Histone lysine code can be read by Bromodomains (BRD) e.g. Bromodomain and extraterminal domain proteins (BET) BRD2, BRD3 and BRD4. They activate gene transcription via recruitment of p-TEFb to RNA polymerase II (Chiang, 2009; Yap and Zhou, 2010). BRD4 also has important roles in regulation of HIV-1 transcription. It was

shown to induce phosphorylation of CDK9, leading to the inhibition of its kinase function thereby inhibiting HIV-1 transcription (Zhou et al., 2009). HDACs were shown to have an important role in the HIV-1 replication cycle, as their inhibition with HDAC inhibitors led to HIV-1 reactivation from latency. HDAC1 is binding to the HIV-1 LTR and by recruiting transcriptional repressors silencing HIV-1 transcription (Archin et al., 2009; Romerio et al., 1997). Several HDAC inhibitors have been tested in clinical trials as latency reversing agents to treat HIV-1 in a 'shock and kill' approach but up to date without significantly reducing HIV-1 reservoir size (Archin et al., 2009, 2014, 2017; Barton et al., 2014; Newhard et al., 2021). A recent preprint study by Peterson et al provides an interesting alternative scenario, according to which HDACs might be important for the establishment of latency rather than latency maintenance, and proposes the use of HDAC inhibitors at the initiation of ART to reduce the formation of latent reservoirs (Peterson et al., 2022).

In line with the function of histone acetylation in modifying the histone backbone to facilitate active transcription, one of the initial events upon activation of the LTR is the recruitment of histone acetyltransferases CBP, GCN5 or P/CAF by tat. HIV-1 LTR was found to be acetylated before actual HIV-1 transcription starts further promoting viral transcription by tat-mediated transactivation of the LTR (Benkirane et al., 1998; Lusic et al., 2003; Marzio et al., 1998).

1.3.1.2 Histone methylation

Histones can be methylated at lysine, arginine, and histidine residues. Addition of a methyl group is not changing the charge but rather position and number of the added methyl group is determining the outcome for gene transcription (Byvoet et al., 1972; Fischle et al., 2008; MURRAY, 1964). Lysine residues can be mono-, di- or trimethylated whereas arginine and histidine residues can be monomethylated (Borun et al., 1972; Gershey et al., 1969; Hempel et al., 1968; Paik and Kim, 1969). The most studied and well characterized histones which are methylated are H3K4, H3K9, H3K27, H3K36, H3K79 and H4K20. Histone methylation is catalyzed by histone methyltransferases (HMTs), transferring a methyl group donated form S-adenosylmethionine. HMTs are categorized into 3 families: SET-domain containing proteins (SETD), DOT1-like proteins and protein-arginine N-methyltransferases (PRMT) (Bannister and Kouzarides, 2011; Feng et al., 2002; Rea et al., 2000). H3K4 methylations are performed by SETD1 and mixed lineage leukemia (MLL) HMTs.

H3K9me3 is deposited by SUV39H, H3K9me2 by euchromatic histone lysine methyltransferase 2 (EHMT2) also known as G9a and SUV39H and H3K9me1 by G9a and euchromatic histone lysine methyltransferase 1 (EHMT1). H3K27 is methylated by enhancer of zeste homolog 2 (EZH) proteins as part of the Polycomb Repressive Complex 2 (PRC2), and H3K36me3 is marked by SETD2.

Demethylases can be classified into amine oxidases and jumonjiC (JmjC) domain containing dioxygenases (Shi et al., 2004; Tsukada et al., 2006), with a big protein family being represented by lysine demethylases (KDM).

Histone methylation can be recognized by proteins with methyl binding domains e.g.: PHD fingers, WD40 repeats, CW domains, ankyrin repeats and PWWP domains (Collins et al., 2008; Hoppmann et al., 2011; Margueron et al., 2009; Shi et al., 2006; Taverna et al., 2007; Wang et al., 2009). An important factor in HIV-1 infection that contains a PWWP domain is LEDGF/p75. The protein binds viral integrase and is leading the PIC to H3K36me3 rich gene bodies (Van Maele et al., 2006).

Histone methylation has different effects on gene transcription, depending on degree and position of the methyl group and the combinations of different marks. H3K36me3 chromatin mark has important roles in transcription elongation, splicing regulation by the recruitment of splicing factors as well as protection of genes from genotoxic stress (Bleuyard et al., 2017; Iwamori et al., 2016; Luco et al., 2010; Pradeepa et al., 2012; Yu et al., 2022). Moreover, through the binding of LEDGF/p75 to H3K36me3 during transcription, double strand breaks recruit C-terminal binding protein Interacting Protein (CtIP) to LEDGF/p75, facilitating homologous recombination and genome stability (Daugaard et al., 2012). H3K4me3 marks active gene promoters, while H3K4me1 is associated with enhancers (Bernstein et al., 2002; Heintzman et al., 2007; Santos-Rosa et al., 2002). H3K27me3 and H3K9me3 mark polycomb and constitutively repressed chromatin, whereas H9K9me2 marks facultative heterochromatin. Combination of H3K4me1 and H3K27me3 are associated with poised enhancers, H3K4me1 and H3K27ac mark active enhancers, while H3K4me1 but absent H3K27ac mark inactive enhancers (Creyghton et al., 2010; Heintzman et al., 2009; Rada-Iglesias et al., 2011).

1.3.1.3 Enhancers

Enhancers are cis-regulatory elements ensuring temporal and cell type specific regulation of gene expression. While the human genome encodes for approximately

20,000 protein coding genes, it contains more than 13 millions of enhancer regulatory elements (Abascal et al., 2020; Gao and Qian, 2020; Kundaje et al., 2015; Omenn et al., 2018). Pioneering studies from several research groups have elucidated a set of epigenetic modifications that delineate enhancer elements. These modifications include an array of histone marks that serve as indicators for cell and tissue-specific enhancers as well as their functional dynamics (Heintzman et al., 2007, 2009; Rada-Iglesias et al., 2011; Visel et al., 2009; Zentner et al., 2011). H3K4me1 and H3K27me3 are marking poised enhancers. H3K4me1 alone marks primed enhancers, while H3K4me1 together with H3K27ac is found on active enhancers (Spicuglia and Vanhille, 2012). Enhancers are activated in a step wise manner by recruitment of cell-type specific transcription factors, chromatin remodelers, coactivators and/or POL-II and communicate with their cognate gene promoter through physical interaction mediated by mediator complex, cohesin, CCCTC-Binding Factor (CTCF) and ying yang 1 (YY1) (Kagey et al., 2010; Spitz and Furlong, 2012; Weintraub et al., 2017). The regulated gene promoters can be located either close or several 100 kb away from the enhancer (Zuin et al., 2022). Multiple enhancers can regulate one gene, as well as multiple genes can be regulated by one enhancer (Furlong and Levine, 2018). Furthermore, enhancers can build highly active clusters, called super enhancers or stretched enhancers, which mostly regulate cell type specific genes (Hnisz et al., 2013). However, in the field a debate about the usefulness of the term SE started to emerge in the recent years, questioning the functional distinctness of SE from other enhancers (Blobel et al., 2021). The complex regulatory enhancer networks allow for the precise coordination of gene expression in development and homeostasis. Lineage- and developmental-stage specific enhancers can be primed in progenitor cells, defining the gene expression potential of a cell in response to external stimuli through stimulus responsive TF's (Glass and Natoli, 2016; Ma et al., 2020; Madsen et al., 2020). Enhancer dysfunction through genetic variants, point mutations or structural variations was associated with several common diseases and cancer (Bradner et al., 2017; Chen et al., 2014; Gasperini et al., 2020; Maurano et al., 2012; Nasser et al., 2021; Waszak et al., 2015).

In HIV-1 infection, integration sites were found clustered close to SE in T cells, with an effect on HIV-1 transcription depending on the distance to host enhancers (Chen et al., 2017; Lucic et al., 2019). For microglia it was recently shown that only a small fraction of active enhancers was shared between different brain cell types, suggesting that the

enhancers repertoire is a landmark for cell type specificity in CNS, and therefore might also be important in HIV-1 integration and persistence (Nott et al., 2019). Moreover, Alzheimer disease risk variants were found to be most prominent in microglia-specific enhancers, making it even more important to study the IS profiles of the so far understudied latent reservoir in the CNS, microglia cells.

1.3.1.4 Chromatin accessibility

Chromatin accessibility describes the extent to which chromatin is available for physical contact with nuclear macromolecules, and it is dependent on nucleosome occupancy as wells other chromatin binding factors (Klemm et al., 2019). Nucleosomes can be post translationally modified which impacts chromatin accessibility by forming a sterical barrier for TF binding or determining affinity for chromatin remodelers (Allis and Jenuwein, 2016; Dann et al., 2017). While facultative and constitutive chromatin are marked by dense nucleosomal packaging, regulatory loci like enhancers, insulators or transcribed gene bodies are nucleosome depleted (Lee et al., 2004; Thurman et al., 2012). The internucleosomal space is mostly marked by transcription factors, architectural factors and POL-II binding and accessibility is highly dynamic (Poirier et al., 2008).

In recent years, several sequencing based methods were developed to measure chromatin accessibility. DNase I hypersensitivity site sequencing (DNase-Seq) measures the abundance of DNase sensitive sites by cutting open chromatin with DNase enzyme followed by library generation and sequencing of accessibly DNA (Boyle et al., 2008; Hesselberth et al., 2009). ATAC-Seq uses a hyperactive Tn5 transposase which is cutting accessible chromatin while at the same time inserting adapters for library generation (Buenrostro et al., 2015). The method has a high resolution, high efficiency (it is suitable for cell numbers as low as 500 cells) and the protocol is fast and can be performed in a high-throughput manner (Corces et al., 2017). Micrococcal nuclease sequencing (MNase-Seq) cleaves the internucleosomal DNA with the endonuclease activity of MNase, and degrades the unoccupied DNA with the exonuclease activity leading to the sequencing of isolated fragments that span single nucleosomes (Mieczkowski et al., 2016; Mueller et al., 2017). Another popular method to determine accessibility is nucleosome occupancy and methylome sequencing (NOMe-Seq) that uses a GpC methyltransferase (MTase) to methylate accessibly DNA followed by bisulfite conversion of unmethylated DNA and sequencing.

Compared to the beforementioned methods, NOMe-Seq is not enrichment based, and therefore more quantitative and has a single molecule character, which requires deep sequencing (Kelly et al., 2012). More recently, single cell approaches for ATAC-Seq, DNase-Seq and NOMe-Seq have been developed to be able to capture short time scale fluctuations and the high heterogeneity among cells (Allis and Jenuwein, 2016; Clark et al., 2018; Jin et al., 2015).

When studying chromatin accessibility and its relation to enhancers, one has to bear in mind that open chromatin alone is not determining activity. Inactive enhancers and promoters of silenced genes are often accessibly but poised by certain chromatin marks, which emphasizes that chromatin accessibility is necessary for activity but not sufficient (Corces et al., 2017; Dogan et al., 2015; Heinz et al., 2015; Lara-Astiaso et al., 2014).

1.3.2 Territories, compartments and topologically associated domains

Over the past decade, significant technological advancements to investigate the three-dimensional structure of the genome have led to new insights into the fundamental principles that govern genome organization. The current model predicts the involvement of multi scale hierarchical 3D folding into higher order chromatin structures at different levels in order to assure gene expression at the right timing and order (Chang et al., 2020; Dixon et al., 2012; Hansen et al., 2018; Lieberman-aiden et al., 2009; McArthur and Capra, 2021; Nora et al., 2012; Rao et al., 2014; Zuin et al., 2014). Key technological advancement has been made by development of chromosome conformation capture methods, including Hi-C, which use chemical crosslinking followed by proximity ligation to detect 3D DNA interactions, that helped to map genome wide contact frequencies (Dixon et al., 2012; Lieberman-aiden et al., 2009; Nora et al., 2012). Over the years, also microscopy-based methods (Boettiger et al., 2016; Finn et al., 2019) and orthogonal methodologies like Genome Architecture Mapping (GAM) complemented our understanding of 3D genome organization (Beagrie et al., 2017).

Chromosomes occupy distinct regions within the nucleus, called chromosome territories. The genome is spatially divided into open and closed chromatin, with A compartments including transcriptionally active euchromatin and B compartments including compact heterochromatin based on long range interactions (Lieberman-aiden et al., 2009) (Figure 8). A and B compartments are further divided into sub

compartments with A1 and A2 and B1-4, being located closer to the nuclear periphery or to nucleoli. A1 sub compartment is rich in highly expressed genes, H3K27ac chromatin mark, H3K79me2 and H3K26me3, whereas A2 contains longer genes and lower GC content. B1 is enriched in facultative heterochromatin, B2 in pericentromeric heterochromatin marks. B3 sub compartment is located at the nuclear lamina and B4 sub compartment is only present at chromosome 19 (Lieberman-aiden et al., 2009; Rao et al., 2014).

Hi-C revealed that on a smaller scale, chromosomes fold into topologically associated domains (TADs) that have a high frequency of internal DNA interactions compared to interaction across boundaries, which facilitate contacts of regulatory features e.g. enhancers or TF with their target promoters and genes (Dixon et al., 2012; Hnisz et al., 2013; Lieberman-aiden et al., 2009; Nora et al., 2012; Rao et al., 2014) (Figure 8). The human genome contains about 4000 to 9000 domains with TAD sizes between 40 kb up to 3 Mb (Dixon et al., 2012; Nora et al., 2012; Rao et al., 2014). The development of more fine-scale chromatin contact methods like micro-C, which includes the digestion with micrococcal nuclease, allows for a high resolution of TAD and loop maps of around 200 bp (compared to 1 kb with Hi-C), supporting investigations of precise enhancer-promoter contacts (Hsieh et al., 2020).

TADs and loops allow for the precise temporal and spatial regulation of gene clusters, were the removal of a gene can lead to transcriptional misregulation of co-associated genes (Fanucchi et al., 2013). Genes in the same TAD were found to be locally co regulated (Hsieh et al., 2020; Krietenstein et al., 2020). Locally, DNA is folded into chromatin loops, that regulate interaction of regulatory elements like SE (Dekker and Misteli, 2015; Halfon, 2020; Hnisz et al., 2013; Vermunt et al., 2019). At the margins between chromatin domains, TAD borders restrict the interaction between regulatory elements and target genes between domains (Beagrie et al., 2017).

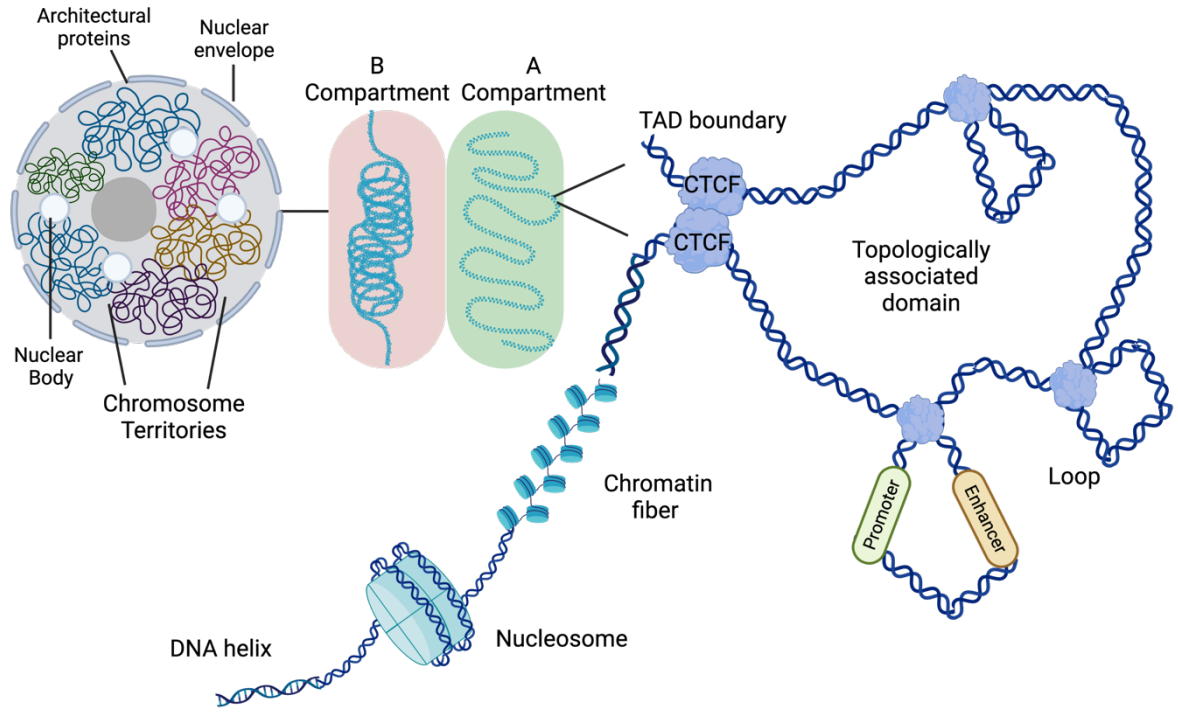


Figure 8: 3D genome Topology. Schematic depiction of 3D genome topology in the human genome. Human chromosomes are organized into chromosome territories and on a smaller scale into A and B compartments (with active and repressive chromatin marks, respectively). Within these compartments DNA is folded into topologically associated domains (TADs), with their margins being marked by CTCF binding. TADs, and more locally genome loops, facilitate enhancer promoter contacts, regulating gene transcription. Figure adapted from (Deng et al., 2022; Misteli, 2020) and created with BioRender.com.

One of the main factors involved in 3D genome organization is CCCTC-binding factor (CTCF) together with the cohesin complex, which are both enriched at TAD margins (Dixon et al., 2012; Nora et al., 2012, 2020; Wutz et al., 2017). The proposed mechanism involves the formation of chromatin loops via cohesin complex ATP-dependent molecular motor activity also known as the loop extrusion model (Fudenberg et al., 2016). Cohesin is a multi-protein complex consisting of SMC1 (structural maintenance of chromosome protein 1) and SMC3 (structural maintenance of chromosome protein 3) that form an open-ended heterodimer, RAD21 that bridges the open end, and SA1 (stromal antigen 1) and SA2 (stromal antigen 2) proteins. It is loaded onto the DNA which is extruded until it encounters two convergent CTCF binding sites bound by CTCF (Figure 9). CTCF is enriched at TAD boundaries, and deletion of CTCF binding sites or altered cohesin binding destroys TAD structure and loop domains (Nuebler et al., 2018; Rao et al., 2017; Wutz et al., 2017) but has no influence on compartments (Haarhuis et al., 2017; Nora et al., 2017; Rao et al., 2017;

Schwarzer et al., 2017). Disrupted TAD structure can lead to misregulation of genes (Lupiáñez et al., 2015; Narendra et al., 2015).

In addition to CTCF and cohesin binding at TAD boundaries, also active chromatin marks, repeat elements, nascent transcripts, housekeeping genes and transfer RNA were found at boundary margins (Dixon et al., 2012; Rao et al., 2014). It is likely that also other factors contribute to higher order chromatin structure, e.g. non coding RNAs, POL-II binding, mediator complex and TF machinery, through the formation of phase-separated condensates building 'hubs' for transcriptional regulation and thereby shaping 3D chromatin structure (Boija et al., 2018; Chong et al., 2018; Quinodoz et al., 2018).

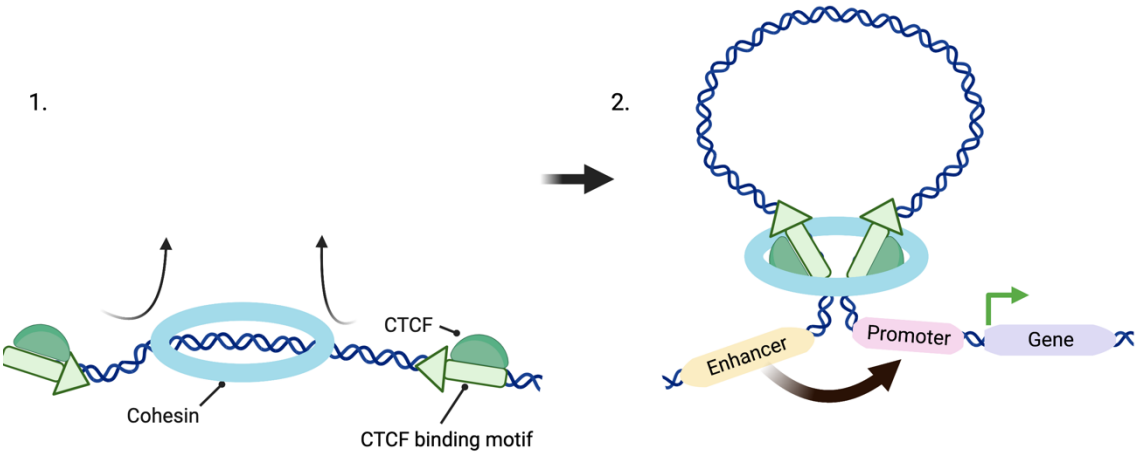


Figure 9: Loop extrusion model. Schematic depiction of the proposed model of loop extrusion. 1.) The cohesin complex is loaded onto the DNA and symmetrically extrudes the DNA strand until it encounters two convergent CTCF binding sites bound by CTCF. 2.) By loop formation and stopping loop extrusion upon CTCF encounter, cohesin and CTCF bring together enhancers with their cognate promoter to activate gene transcription. Figure was created with BioRender.com.

1.3.2.1 The architectural protein CTCF

CTCF plays crucial roles in many cellular processes. It was first identified as transcriptional repressor of the chicken c-*myc* promotor before over time many further studies attributed additional regulatory capabilities to the protein (Klenova et al., 1993). The protein is 82 kDa large and consists of an N-terminal domain, a C-terminal domain and the central DNA and RNA binding domain comprised of 11 zinc fingers (ZF) (Filippova et al., 1996; Vostrov et al., 2002). ZF4-7 are ensuring DNA binding and ZF1, ZF8-11 allow for RNA binding with an additional RNA binding domain in the C-terminus (Hansen et al., 2019; Nakahashi et al., 2013; Nora et al., 2020; Saldaña-Meyer et al.,

2014, 2019). Mutations in ZF1 and ZF10 affect CTCF's capacity to form loops and chromatin insulation, therefore affecting gene expression patterns (Saldaña-Meyer et al., 2019). All of the protein domains can be post translationally modified by SUMOylation, phosphorylation or PARylation (Farrar et al., 2010; Klenova et al., 2001; MacPherson et al., 2009). CTCF SUMOylation was shown to enhance CTCF's repressive function, while phosphorylation enhanced its activator function (Klenova et al., 2001; MacPherson et al., 2009). PARylation of CTCF was found to be involved in DNA damage response and nucleolar transcription (Han et al., 2017; Torrano et al., 2006). In the human genome, CTCF can bind to 55,000 to 65,000 unmethylated DNA binding sites (Chen et al., 2012b).

The distinct functions of CTCF include context-dependent gene promoter activation/repression, enhancer blocking and barrier insulation, hormone-responsive silencing, X-chromosome inactivation, genomic imprinting, blocking long-range chromatin interactions and separating silent from active chromatin (Chao et al., 2002; Hou et al., 2008; Phillips and Corces, 2009; Schmidt et al., 2012).

The fundamental principle of gene expression regulation in eukaryotes is the interaction of enhancers with promoters. These regulatory elements can act over long distances, with a median of 4 long range interactions per promoter (Javierre et al., 2016). Multiple regulatory elements contribute to gene expression highlighting the need to regulate and prevent uncontrolled enhancer activity. CTCF was the first protein identified as an enhancer-blocking insulator to block the activity of cis-acting enhancer elements at the chicken β-globin locus (Bell et al., 1999).

The second important locus having a fundamental role in the discovery of CTCF functions is the H19/Igf2 locus where CTCF plays a role in gene imprinting, where a gene on one parental allele is silenced, while the other one is expressed (Bajrami and Spiroski, 2016). Between the 2 gene loci, the imprinting control region (ICR) is located, which is differentially methylated depending if it is on the maternal or paternal allele. CTCF only binds to the unmethylated ICR on the maternal allele and thereby preventing the binding of the H19 enhancer to the Igf2 promoter leading to gene inactivation. Conversely, on the methylated paternal allele CTCF binding is prevented and the H19 enhancer can therefore activate transcription of Igf2 gene (Hark et al., 2000). These early discovered examples highlight a direct role for CTCF in gene expression regulation.

1.3.2.2 The cohesin complex

The mechanisms that mediate genome topology are currently subject of extensive research and include loop extrusion, phase separation and formation of condensates (Boija et al., 2018; Chong et al., 2018; Dixon et al., 2012; Nora et al., 2012; Quinodoz et al., 2018; Wutz et al., 2017). In the case of TADs, most of these domains are formed by loop extrusion, mediated by CTCF and cohesin complex interaction (Banigan et al., 2020; Fudenberg et al., 2016; Gabriele et al., 2022; Ganji et al., 2018; Wutz et al., 2017). The complex consists of the subunits SMC1 and SMC3, forming a ring like structure with RAD21 (also known as SCC1), and the subunits STAG1 or STAG2 (also known as SA1/2) (Yatskevich *et al*., 2019) (Figure 10). It was originally recognized as a complex being involved in chromosome cohesion between sister chromosomes during mitosis (Galli et al., 2013; Sagai et al., 2005).

In 3D chromatin organization, loading of the cohesin complex onto the U-formed tip of the DNA is mediated by the heterodimer Nipped-B-like protein (NIPBL) and MAU2/SCC4 which stays associated to the complex during loop extrusion. SMC1/3 are ATPases that work like molecular motors to mediate moving of the cohesin complex along the DNA. ATP hydrolysis is stimulated by DNA, leading to a conformational switch of SMC1/3, extruding the DNA in a symmetric bidirectional manner. Cohesin extrudes DNA into loops until it encounters 2 convergent CTCF proteins binding to the DNA (Davidson et al., 2019; Kim et al., 2019; Li et al., 2020b; Pugacheva et al., 2020). The release of the complex is wings apart-like protein homolog (WAPL) dependent. CTCF interaction with RAD21 is stabilizing the cohesin complex by protecting it from WAPL release. CTCF is interacting with cohesin via its N-terminal domain residues 222-231 and 23-47. The former one is binding to SA2 and RAD21 subunits, whereas the latter one is binding to the PDS5A subunit (Li et al., 2020b; Pugacheva et al., 2020). Cohesin release from the chromatin is regulated by WAPL binding to PDS5A, meaning that WAPL and CTCF are competing for cohesin binding allowing CTCF to protect cohesin from WAPL release (Nora et al., 2020).

One member of the cohesin complex is RAD21. The protein consists of 631 amino acids and has 3 binding domains interacting with SMC3, STAG and SMC1 (Gligoris et al., 2014). Apart from its role in the cohesin complex and 3D organization, the protein is involved in double strand break repair by promoting the use of sister chromatid as a template for recombination based DNA damage repair (Kong et al., 2014).

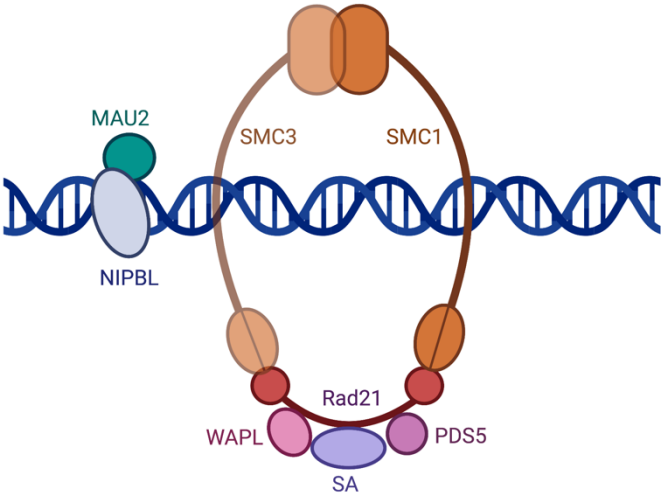


Figure 10: The cohesin complex. Schematic depiction of the cohesin complex. Cohesin is loaded onto DNA by NIPBL and MAU2. SMC3 and SMC1 form a ring like structure around the DNA, which is closed by Rad21 and SA. Cohesin complex stays on the DNA upon release by WAPL and PDS5. Figure was created by BioRender.com.

1.3.2.3 Influence of CTCF and cohesin on TAD structure

The influence of CTCF depletion and removal in 3D chromatin structure and organization is debated in the field, as it is very much context and cell-type dependent. Depletion of CTCF in mouse embryonic stem cells (mESC) and human cell lines showed the involvement of the protein in looping and TAD formation, but compartment structure was unaffected (Haarhuis et al., 2017; Hyle et al., 2019; Nora et al., 2017; Nuebler et al., 2018; Rao et al., 2017; Schwarzer et al., 2017; Wutz et al., 2017). Restoration of CTCF re-established the correct architecture on altered chromosomes, indicating an important function of CTCF in chromatin folding (Nora et al., 2017). However, a subset comprising less than 20% of TAD boundaries were found to be resistant to CTCF loss, underscoring the prominent role of CTCF as a principle determinant of TAD insulation, while acknowledging the existence of additional mechanisms at play (Nora et al., 2017). TAD structure disruption can lead to gene expression misregulation (Lupiáñez et al., 2015; Narendra et al., 2015), but to a minor extent and not necessarily connected to the depleted CTCF sites, highlighting the pleiotropic effects of CTCF (Kubo et al., 2021). CTCF absence leading to disrupted 3D genome organization in B cells still allowed for trans differentiation into macrophages but regulation of activation of transcriptional response following LPS stimulation was disturbed (Stik et al., 2020).

Understanding the importance of chromatin structure in relation to genome function is possibly one of the most important focuses of ongoing research in the field. Deletion of a TAD boundary can lead to fusion of two neighboring TADs where the outcome on gene expression is locus dependent (Despang et al., 2019; Narendra et al., 2016). The fusion might either have no effect on enhancer promoter contacts or an enhancer finds a new target promoter leading to enhancer hijacking, being a rare case. Deletion of a selective CTCF site at the α-globin enhancer, extended a TAD to the adjacent subTAD, leading to interaction of the α-globin enhancer with the neighboring chromatin upregulating genes expression (Hanssen et al., 2017). Loss of CTCF at a domain boundary led to an aberrant activation of platelet-derived growth factor receptor A (PDGFRA), a prominent glioma oncogene (Flavahan et al., 2016). In the HoxA/C cluster during development from mouse embryonic stem cells into motoneurons, long range enhancer contacts and chromatin marks were found to be restricted within a TAD influencing correct gene expression, impacted by dynamic CTCF boundary elements (Narendra et al., 2015, 2016).

Cohesin depletion also leads to chromatin loop domain disruption, with only a small subset of genes being transcriptionally dysregulated. After protein recovery, loops were reformed. Furthermore, upon cohesin depletion newly super-enhancer enriched loops were formed (Rao et al., 2017).

When disrupting other cofactors involved in loop extrusion process like NIPBL, a factor required for cohesin loading onto chromatin, chromatin loop structure and TADs disappeared, while compartment structure was preserved and even refined. Gene expression was moderately misregulated, indicating a cohesin independent mechanism of compartmentalization (Schwarzer et al., 2017). On the other hand, when depleting the cohesin release factor WAPL, the median loop size increased by around 200 kb leading to differential expression of about 1000 genes (Haarhuis et al., 2017). Polymer modeling approaches with the aim to explain the opposing results of TAD destruction and finer compartmentalization revealed that loop extrusion suppresses inherent compartmentalization on a small scale, as compartments segregate by phase separation due to epigenetic marks (Nuebler et al., 2018). Another study systemically investigated the effect of NIPBL and WAPL depletion in HCT116 cells. They depleted both factors independent from each other and together. The investigations showed that knock out of either of the two led to dysregulation of about 2000 genes, where around 30% of them were shared. When both factors are depleted, majority of gene

expression was restored, indicating that a balance between the factors regulating cohesin dynamics are important for correct gene expression (Luppino et al., 2022).
One of the main questions in the field of 3D genome organization that merits further investigations is the relation of genome structure to function. As discussed in this paragraph, perturbation studies over the years showed opposing results about the influence of genome topology on gene transcription. Currently data indicate that TADs and loops build a structural framework rather than a strong determinant of gene function (Misteli, 2020). In addition, this influence works in both directions. Loop structure can have influence on function, maybe in some cases only in a modest degree, but also transcription can influence looping structure. In addition, one has to keep in mind when interpreting functional studies on genome structure that genome-wide contact frequency approaches, like Hi-C, are population based. Gene regulation and chromatin structures are dynamic processes which can vary from cell to cell, which makes it important to further study 3D genome on a single cell level (Bintu et al., 2018; Finn et al., 2019). It seems that the possible functional roles of genome topology are highly specific to the developmental timing, organism and genomic context (Misteli, 2020).

1.3.3 3D chromatin organization in viral infections

Whenever viral pathogens infect a host cell and enter the nucleus, they encounter the hierarchically organized 3D genome, and are exposed to genome folding and to separating principles of nuclear body formation. Several studies over the past years focused on how viruses make use of the host chromatin to navigate the nuclear environment and modulate the topology of host chromatin to establish viral infection (Lucic et al., 2021). Most of these studies found DNA viruses to exploit host architectural proteins to organize their genomes and facilitate essential processes such as infection, lytic replication, and latency. CTCF and other architectural proteins are used to either change host genome topology and transcription or to organize its own genome to obtain optimal transcriptional control.
The DNA virus Hepatitis B virus (HBV) contains two CTCF binding sites and disruption of CTCF binding leads to an accumulation of activating epigenetic histone modifications, resulting in increased transcription of the virus (D'Arienzo et al., 2021). The covalently closed circular DNA (cccDNA) was also shown to be tethered more frequently to active chromatin regions, and interacting more with genes dysregulated

during infection, highlighting that HBV might interfere with cellular gene expression (Moreau et al., 2018). Another study showed a preferred targeting of the cccDNA to the enhancer element 19p13.11, influencing viral transcription (Shen et al., 2020). All studies are highlighting that HBV is exploiting the 3D host genome for its own transcription, but also influencing 3D chromatin environment of the host.

For herpes simplex virus (HSV), CTCF insulators are located at the boundaries of the latency and lytic genes, preventing inappropriate activation of lytic genes (Amelio et al., 2006; Chen et al., 2007; Ertel et al., 2012). Additionally, HSV genome was found to form multiple loops structure which are maintained by CTCF and are essential for viral latency and reactivation (Washington et al., 2018a, 2018b). Similar regulatory mechanisms involving CTCF have been observed in other herpesviruses, including human cytomegalovirus, Epstein-Barr virus (EBV), and Kaposi sarcoma-associated herpesvirus (Arvey et al., 2012; Holdorf et al., 2011; Tempera et al., 2010; Varghese et al., 2022). In the case of human papillomaviruses (HPV) 18, CTCF mediates a chromatin loop between the CTCF binding site in the early gene region and the upstream viral enhancer leading to the dampening of viral transcription (Paris et al., 2015; Pentland et al., 2018). HPV integrations were found to alter TAD border structure in the IS vicinity and change genome topology, leading to the formation of new virus-host DNA interactions including host enhancers (Cao et al., 2020; Tian et al., 2023).

To summarize, DNA viruses employ host 3D chromatin structure and CTCF as an insulator protein to govern epigenetic modifications and regulate transcription.

1.3.3.1 CTCF and genome topology in retroviruses

The first oncogenic human retrovirus discovered was the human T-lymphotropic virus type 1 (HTLV-1) (Poiesz et al., 1980). The virus can cause adult T-cell leukaemia/lymphoma (ATL) and HTLV-1- associated myelopathy (Proietti et al., 2005). In addition to viral enzymes and structural proteins, the HTLV- 1 genome encodes regulatory and auxiliary genes in the pX region (Boxus and Willems, 2009). CTCF plays a role on the expression regulation of 2 important viral mRNAs important for viral pathogenesis: the mRNA encoding for sHBZ protein, involved in oncogenesis, and the mRNA encoding for immunogenic tax protein, necessary for viral transcription. The spliced sHBZ mRNA is constitutively expressed from the antisense strand in HTLV-1 infections, while tax mRNA expression is suppressed by sHBZ protein to evade the host immune response (Clerc et al., 2008; Nyborg et al., 2010; Philip et al., 2014). Plus strand transcription is epigenetically silenced by methylation of the 5’ LTR, whereas

the 3' LTR is only moderately methylated (Koiwa et al., 2002), with CTCF binding between those differentially methylated regions in the pX region (Satou et al., 2006). CTCF bound to the pX region showed enhancer blocking activity and affected proviral transcription. Mutations in the CTCF binding site within the pX region caused significantly reduced sHBZ transcription, suggesting that CTCF is a boundary for repressive epigenetic modifications allowing for correct mRNA expression (Satou et al., 2016). Furthermore, CTCF bound to the HTLV-1 genome was found to mediate the formation of virus-host gene loops, blocking enhancer-promoter contacts, altering transcription and splicing more than 300 kb away from the IS (Melamed et al., 2018).

A similar role of CTCF could also be found in regulating bovine leukemia virus (BLV) latency (Bellefroid et al., 2022). Latent infections of this oncogenic deltaretrovirus are characterized by the absence of viremia and plus strand transcription likely allowing the escape from the host immune system while causing tumor development by minus strand transcription (Gillet et al., 2013). Latency is induced by suppression of the POL-II promoter on the 5' LTR (Pluta et al., 2020). BLV provirus contains three putative CTCF binding sites, with one being located in the 5' LTR and one in 3' LTR. Mutations in the CTCF binding sites showed that CTCF activates 3' LTR antisense promoter while suppressing 5' LTR sense promoter, promoting viral latency. The proposed mechanism behind proviral silencing was the prevention of active chromatin mark spreading from the 3' LTR to the 5' LTR from the CTCF binding site situated within the second exon of Tax/Rex (Bellefroid et al., 2022).
Altogether, HTLV-1 und BLV use CTCF binding to the provirus for epigenetic regulation allowing escape from host immune surveillance.

1.3.3.2 CTCF and genome topology in HIV-1 infection

In contrast to HTLV-1, consensus sequence for CTCF could not be detected within the proviral gene sequence and CTCF was not found to bind the HIV-1 provirus (Satou et al., 2016). A recent study performed ATAC-Seq in sorted active and latent HIV-1 infected T cells and revealed that CTCF is required for establishing HIV-1 latency. Increased occupancy of CTCF in latent cells and transcriptional repression mediated by CTCF suggested that dynamic changes in CTCF binding and possibly chromatin organization could influence HIV-1 viral expression and persistence (Jefferys et al., 2021). These findings indicate that in the context of HIV-1 infection, CTCF plays a role

in the viral life cycle via a different mechanism compared to the previously mentioned viruses, but further research is necessary.

Few studies in the recent years started addressing the role of 3D genome in HIV-1 infection. Lucic and colleagues showed that HIV-1 insertion hotspots cluster together in the 3D A1 sub compartment of the T cell nucleus in proximity to super enhancers (Lucic et al., 2019), by using Hi-C of uninfected Jurkat cells and HIV-1 integration sites. Genome organization assessment in patient derived $CD4^+$ T cells revealed a distinct chromatin accessibility in the vicinity of intact HIV-1 proviruses possibly contributing to a selective advantage during long-term ART (Einkauf et al., 2019). More recently, Einkauf et al. performed IS sequencing, proviral sequencing and RNA-Seq from the same cell in patient derived $CD4^+$ T cells and analyzed the data together with ChIP-Seq, ATAC-Seq and Hi-C data sets of uninfected T cells. They could demonstrate that non genic IS had increased distance to frequently interacting regions (FIREs) and TADs and reduced numbers 3D inter- and intrachromosomal interactions. Regions with transcriptionally active proviruses showed reduced distance to FIREs, higher frequency of 3D contacts and high levels of RNA-Seq, ATAC-Seq, and activating ChIP-Seq signals in the 3D contact regions, suggesting that proviral gene expression is facilitated by integration into interactive and more active chromosomal regions (Einkauf et al., 2022).

These studies highlight a possibly link between 3D host chromatin organization and HIV-1 transcriptional control.

1.4 Aims of the study

The stably integrated provirus that persists under ART in latent reservoirs throughout the body is still the main bottleneck for an HIV cure. To this end, the predominant non-tissue and well-studied latent reservoir of HIV-1 are memory $CD4^+$ T cells, where the majority of our insights into HIV-1 integration sites, including the locations, genomic profiles and transcriptional output, have been obtained. Microglia cells are the main HIV-1 target cells in the brain, therefore difficult to target with ART and a possible reason for neurological disorders of HIV-1 infected patients. Thus, examining integration patterns and mechanisms of establishment and/or maintenance of HIV-1 latency in this brain cell type is pivotal for understanding the highly complex and dynamic nature of HIV-1 latent reservoirs in the CNS for the development and improvement of new HIV-1 therapies and optimization of therapy delivery systems.

AIM I:

- Are HIV-1 integration sites in microglia reservoir cell type specific?

As patterns of HIV-1 integration and genome regulation linked to viral insertions in microglia still remain undetermined, the first aim of my PhD thesis is the determination of the integration sites landscape and their chromatin signatures in microglia cells as well as the comparative analysis of these profiles with integration site patters previously obtained in $CD4^+$ T cells and macrophages.
To sequence HIV-1 ISs in microglia cells, the first goal is the establishment of an IS sequencing method in the lab, based on linker mediated polymerase chain reaction (LM PCR), adapted from (Serrao et al., 2016). The obtained ISs in microglia cells will be compared with IS data sets of the main HIV-1 target cell types $CD4^+$ T cells (GEO: GSE134382, (Lucic et al., 2019)) and macrophages (Kok et al., 2016) together with RNA-Seq and ChIP-Seq for several histone marks in uninfected microglia with the aim to investigate cell type specificity of HIV-1 integration.

AIM 2

- How do host chromatin and transcriptional activity differ between microglia cells harboring transcriptionally active and latent proviruses?
- How do these differences contribute to latency establishment and maintenance?

In the second part of my PhD thesis the aim is to investigate whether and how the host chromatin and regulatory networks might differ between active and latent infection in microglia early upon infection (3 d.p.i.). To do so I will use a dual-labeled orange-green HIV-1 reporter (HIV_{GKO}) virus, to distinguish cells with actively replicating virus, from latent cells (Battivelli and Verdin, 2018; Battivelli et al., 2018). To characterize chromatin state differences, chromatin accessibility of genomic DNA regions from sorted early productive and latent microglia population (and uninfected controls) will be assessed by ATAC-Seq (Buenrostro et al., 2015). In this way, I will be able to obtain a dynamic time course of the chromatin changes in HIV-1 infection to answer the question if distinct transcriptional programs and/or chromatin states define early latency establishment.

2 Results

All bioinformatic analysis of this study was performed by Ana Luisa Costa, PhD student, and Prof Dr Carl Herrmann from the Health Data Science unit, University clinics Heidelberg in close collaboration with our lab.
Parts of the results presented in this thesis have been published in Cell Reports (Rheinberger et al., 2023) under the Creative Commons CC-BY-NC-ND license and are reproduced here in accordance with the rights of open-access publishing from Elsevier.

2.1 HIV-1 integration site landscape in microglia cellular model

2.1.1 Establishment of HIV-1 infection in microglia cellular model

For this study, I was using an immortalized microglia cell line C20 (Garcia-Mesa et al., 2017) from cryopreserved primary human microglia infected with viral particles containing SV large antigen and human telomerase reverse transcriptase (hTERT). This cell line expresses microglia cell surface markers and shows the same morphology as primary cells (Garcia-Mesa et al., 2017). It is well known that transcriptional signatures change upon transfer of primary microglia cells into cell culture conditions due to sensitivity to environmental brain signals (Gosselin et al., 2017). Yet, this cell line represents a valid tool to start investigating HIV-1 infection in microglia, where the lack of relevant human models and the limited availability of primary brain samples has restrained large scale genomic studies (Alvarez-Carbonell et al., 2019, 2020; Garcia-Mesa et al., 2017; Ingram et al., 2020). C20 cells are less susceptible to HIV-1 wild type infections because they lose the CD4 receptor for viral entry over time after several passages (Garcia-Mesa et al., 2017). Therefore, I infected the microglia cells with viral particles pseudotyped with vesicular stomatitis virus G protein (VSV-G), which allows the viral particles to fuse with any cell membrane (Ci et al., 2018). After three days post infection (3 d.p.i.), cells were harvested to perform FACS analysis to control for p24 capsid protein, showing an efficient infection rate of around 20% -30% p24 positive microglia cells (Figure 11A). Viral gene expression was also confirmed by measuring gag mRNA levels by quantitative PCR (qPCR) (Figure 11B). Next, the integration level was assessed through Alu PCR, which is a nested

PCR with primers annealing in the viral LTR region and genomic Alu repeats (Liszewski et al., 2009; Tan et al., 2006). In the first PCR, the HIV LTR specific primer contains a phage lambda-specific heel sequence. The second PCR is a qPCR with primers annealing to the lambda-specific sequence, internal LTR primer and TaqMan probe, to ensure that only proviral sequences from the first round are amplified.

The microglia cells showed high integration levels over treated samples with integrase (Raltegravir) and reverse transcriptase (Efavirenz) inhibitors (Figure 11C). For the visualization of the integrated provirus, I performed HIV-1 DNA FISH with directly labelled HIV-1 DNA probes. In addition, nuclear pore complex proteins (mAB414) were stained to visualize the nuclear rim (Figure 11C, HIV-1 (green), mAB414 (red)). In summary, I was able to establish productive HIV-1 infection in microglia cell model.

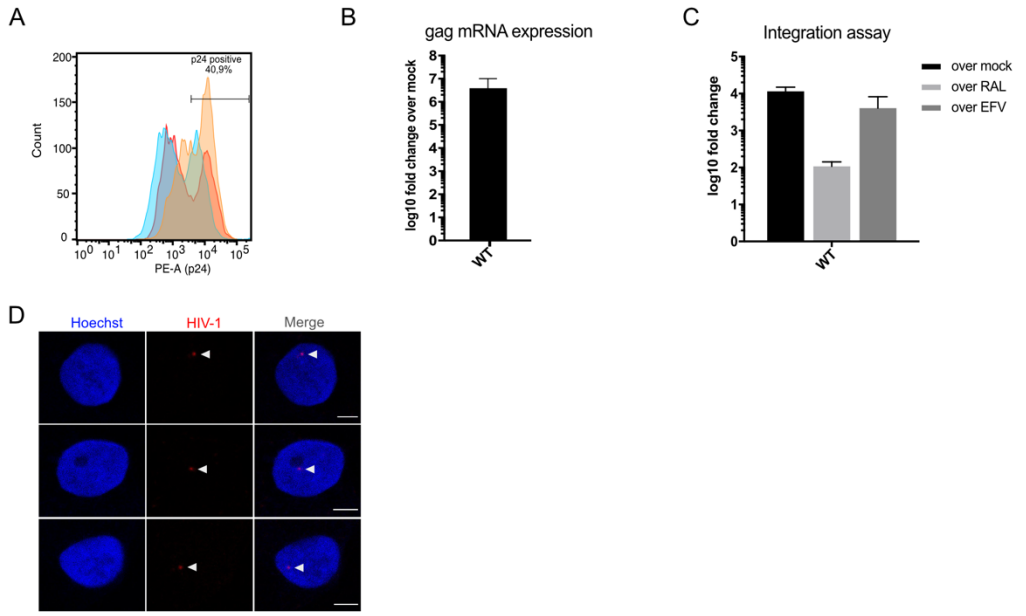


Figure 11: Establishment of HIV-1 infection in microglia cellular model. (A) FACS plot of 3 biological replicates (indicated in blue, yellow, red). Infected C20 cells (250 ng p24/ 1 mill cells) were fixed with 4% PFA and stained for intracellular p24 capsid protein with PE antibody for 30 min and then read on a FACS celesta instrument. Plot was generated with FlowJo Software. (B) C20 cells were infected with 250 ng p24/ 1 mill cells. Bar plot of gag mRNA expression over mock measured by qPCR. Plot shows mean of 3 biological replicates ± SEM and was generated with GraphPad prism. (C) Bar plot of HIV-1 integration assay determined by Alu PCR. Integration levels are expressed over mock, over Raltegravir (RAL) treated samples (10 μM) or samples treated with Efavirenz (EFV) (20 μM). Data represent mean of 3 technical replicates ± SEM and was generated with GraphPad prism. (D) HIV-1 3D DNA immuno-FISH of HIV-1 infected C20 cells (250 ng p24/ 1 mill cells). Cells were fixed, permeabilized and hybridized for 48h with directly labeled HIV-1 FISH probe. HIV-1 signal in red, DNA stained with Höchst in blue, scale bar 5 μm. Figure was adapted from (Rheinberger et al., 2023).

2.1.2 Establishing an Integration site sequencing method

To sequence HIV-1 integration sites in microglia cellular model, I established linker mediated (LM) PCR method (adopted from (Serrao et al., 2016)). At 3 d.p.i. (peak of the productive infection), infected genomic DNA was isolated and sheared by sonication to obtain fragments smaller than 500 bp. A partially double stranded asymmetric linker with a 3' -T overhang was ligated to the DNA to amplify fragments from HIV 3'-LTR and the single stranded linker overhang in a first PCR step. The second nested PCR was next performed with an inner LTR primer pair to introduce Illumina sequence adaptors and barcodes (Figure 12A). To get a more precise fragment size distribution of the library compared to an agarose gel (Figure 12B), the libraries were loaded onto a Bioanalyzer DNA high sensitivity Chip. Average size distribution of the libraries ranged from 364 bp to 427 bp (Table 1, Figure 12C). To ensure that after the second nested PCR sequencing adapters were correctly added to the viral-genome junctions, I used the NEBNext® Library Quant Kit for Illumina® from NEB to perform a qPCR utilizing primers complementary to the P5 and P7 adapters. By including a library standard, molarities can be calculated. For the first sequencing run, molarities ranged from 22 nM to 33 nM.

To determine integration sites (ISs), obtained reads were mapped onto the human and HIV-1 chimeric genome and the chimeric reads were further processed and filtered to obtain the IS position. Reads shorter than 15 bp were excluded from the alignment and only uniquely mapped reads were further processed[1]. From the first sequencing run 2368 IS out of 3 biological replicates were obtained. One issue was that a large fraction of fragments had very short genomic read left after adapters, linker and LTR region were trimmed off. This resulted in a substantial portion of the reads that could not be accurately mapped to the human genome. Looking at the Bioanalyzer profiles of the sent libraries, one can observe a sharp peak at around 200 bp, indicating an overamplification of short DNA fragments. Furthermore, a lot of reads had linker duplications and concatemers. Based on these initial settings, the LM-PCR protocol was optimized to obtain larger fragments, which are also less prone to overamplification.

The first parameter changed was the purification method after linker ligation. Instead of purifying the sample with the PCR clean up kit, I purified it with 0.9x AMpure beads,

[1]Analysis was performed by Ana Luisa Costa, Health Data Science unit, University Clinics Heidelberg.

which allows for the efficient removal of unligated linker sequences (Figure 12D, first panel). To restrict overamplification, sonication time was shortened, primer concentration was lowered from 1.5 µM to 0.8 µM and cycle number of the second nested PCR was reduced from 30 cycles to 13 cycles. Moreover, the input material of the second PCR was lowered and the DNA of the first LM PCR was split into several smaller reactions, to ensure proper amplification and no overload of DNA input. All these adjustments led to more homogenous library size distribution (Figure 12D, middle and right panel). 3 more microglia infections were performed and subjected to the adjusted LM-PCR protocol. The 3 new IS libraries showed average library sizes from 442 bp to 479 bp with library molarities form 164 nM to 334 nM (Figure 12E,F, Table 1). Together with the 3 previously sequenced IS libraries 4,590 IS out of 6 biological replicates were obtained (Table 1, GEO accession number GSE205915)[1].

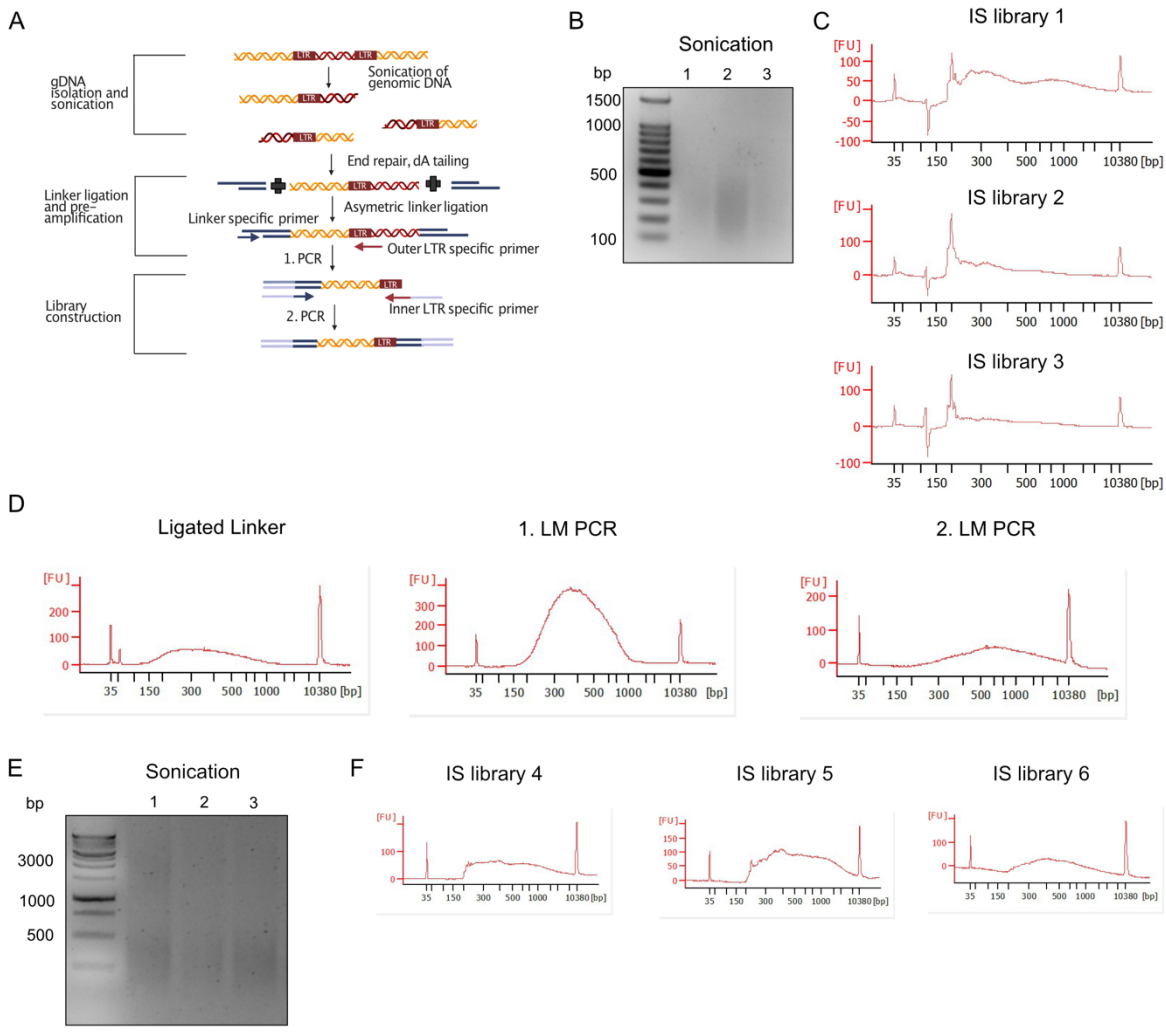


Figure 12: Integration sites sequencing optimization. (A) Schematic overview of linker-mediated PCR method. Genomic DNA is sonicated, end repaired, and an asymmetric linker is annealed. In a first PCR step, viral genome junctions are amplified with a primer pair annealing to the linker and the LTR region. In a second PCR with an inner primer pair, Illumina indices and adapters are introduced, and PCR products can be sent for Illumina paired end sequencing. Scheme was generated with BioRender. (B) Sonicated infected genomic DNA was loaded onto a 2% agarose gel to check sonication efficiency. 100 bp DNA ladder served as DNA size standard. Gel was run at constant 90 V for 45 min and image was taken with UV light. 1: Infection replicate 1; 2: infection replicate 2; 3: infection replicate 3. (C) Bioanalyzer profiles of 3 LM-PCR ISs libraries. X-axis shows size in bp and Y-axis displays fluorescent units. 35 bp marks the lower marker, 10380 bp marks the upper marker. (D) Bioanalyzer profiles of ligated linker sample, 1. LM PCR and 2. LM PCR after protocol adjustments. X-axis shows size in bp and Y-axis displays fluorescent units. 35 bp marks the lower marker, 10380 bp marks the upper marker. (E) Sonicated infected genomic DNA was loaded onto a 2% agarose gel to check sonication efficiency. 10o0 bp DNA ladder served as DNA size standard. Gel was run at constant 90 V for 45 min and image was taken with UV light. 1: Infection replicate 4; 2: infection replicate 5; 3: infection replicate 6. (F) Bioanalyzer profiles of 3 LM-PCR ISs libraries. X-axis shows size in bp and Y-axis displays fluorescent units. 35 bp marks the lower marker, 10380 bp marks the upper marker.

Sample	Concentration (ng/µl)	Average fragment size (bp)	Molarity based on qPCR (nM)
IS library 1	9.3	427	33.5
IS library 2	6.8	364	28.7
IS library 3	5.6	389	22.1
IS library 4	18	442	334
IS library 5	7.5	469	164
IS library 6	18	479	255

Table 1: IS libraries of C20 WT infections.

2.1.3 HIV-1 integrates into actively transcribing genes in microglia cellular model

ISs profiles were annotated and further characterized with respect to their chromosomal distribution and genic position compared to published IS profiles of $CD4^+$ T cells and macrophages (Kok et al., 2016; Lucic et al., 2019) (Figure 13A and B)[1]. Most of the ISs were found on the gene dense chromosomes 17 and 19 in all 3 cellular targets of HIV-1 (Figure 13A). 60.2 % of integrations were found in introns in microglia (MDMs: 56.3%; $CD4^+$ T cells: 57.2%) and 14.5% in distal intergenic regions (MDMs: 11.7%; $CD4^+$ T cells: 17.3%) (Figure 13B). When comparing genic integration sites between the 3 cellular HIV-1 targets, C20 genic ISs displayed higher similarity to $CD4^+$ T cells (20% overlap, Jaccard index 0.209) than to MDM (5% overlap, Jaccard index 0.096) (Figure 13C)[1].

In order to determine transcriptional output of integration genes, RNA-Seq in uninfected C20 cells was performed. Genes were stratified according to their expression levels into highly expressed genes (top 10%), low expressed genes (bottom 10%), mid expressed genes (remaining 80%) and non-expressed genes (Figure 13D). Integrating the expression levels and the ISs data showed that 25% of the genes where HIV-1 integrates are highly expressed genes in microglia, and about 60% are mid expressed genes, which goes along with the finding that HIV-1 integrates into actively transcribing genes, previously shown for $CD4^+$ T cells and macrophages (Figure 13D)[1].

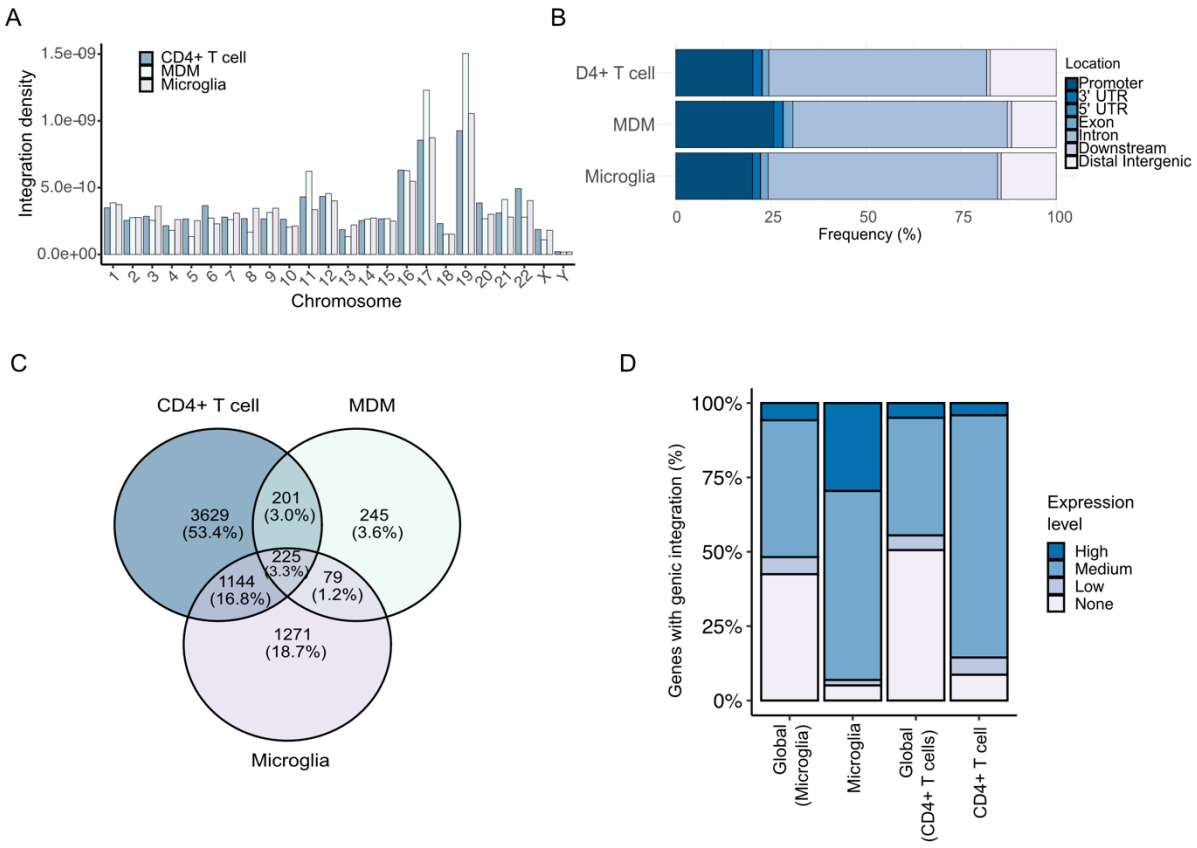


Figure 13: Integration site sequencing in microglia cellular model. (A) Integration site density per chromosome for CD4⁺ T cell IS (blue bar) (Lucic et al., 2019), MDM IS (light green bar) (Kok et al., 2016) and C20 microglia IS from this study (purple bar)[1]. (B) Bar plots showing frequency of integration into Promoter regions, 3' UTR, 5' UTR, Exons, Introns, downstream and distal intergenic regions for CD4⁺ T cell IS (Lucic et al., 2019), MDM IS (Kok et al., 2016), and C20 microglia IS (this study)[1]. (C) Overlap between genic integrations between CD4⁺ T cells (blue, n=5199), MDM (light green, n=750) and C20 microglia (purple, n=2719)[1]. (D) Genes were stratified into high (dark blue), medium (light blue), low (grey) and non-expressed (light grey) genes based on RNA-Seq (global) of uninfected C20 microglia cells. Integration genes (n=2719) were stratified in the same way to determine percentage of genes with genic integration per expression class[1]. Figure was adapted from (Rheinberger et al., 2023).

Gene annotation using genome assembly hg38 showed that the top 30 genes where genic integrations most frequently occurred in microglia (genes harboring ≥ 5 HIV-1 insertions) were also recurrently targeted in CD4⁺ T cells (56.8%) and were highly expressed genes as assessed by RNA-Seq (Figure 14A and B)[1]. Gene ontology enrichment analysis of integration genes revealed that numerous genes were part of the histone modification and chromatin organization pathway in both microglia and CD4⁺ T cells, suggesting similar integration patterns for the two cellular reservoirs.

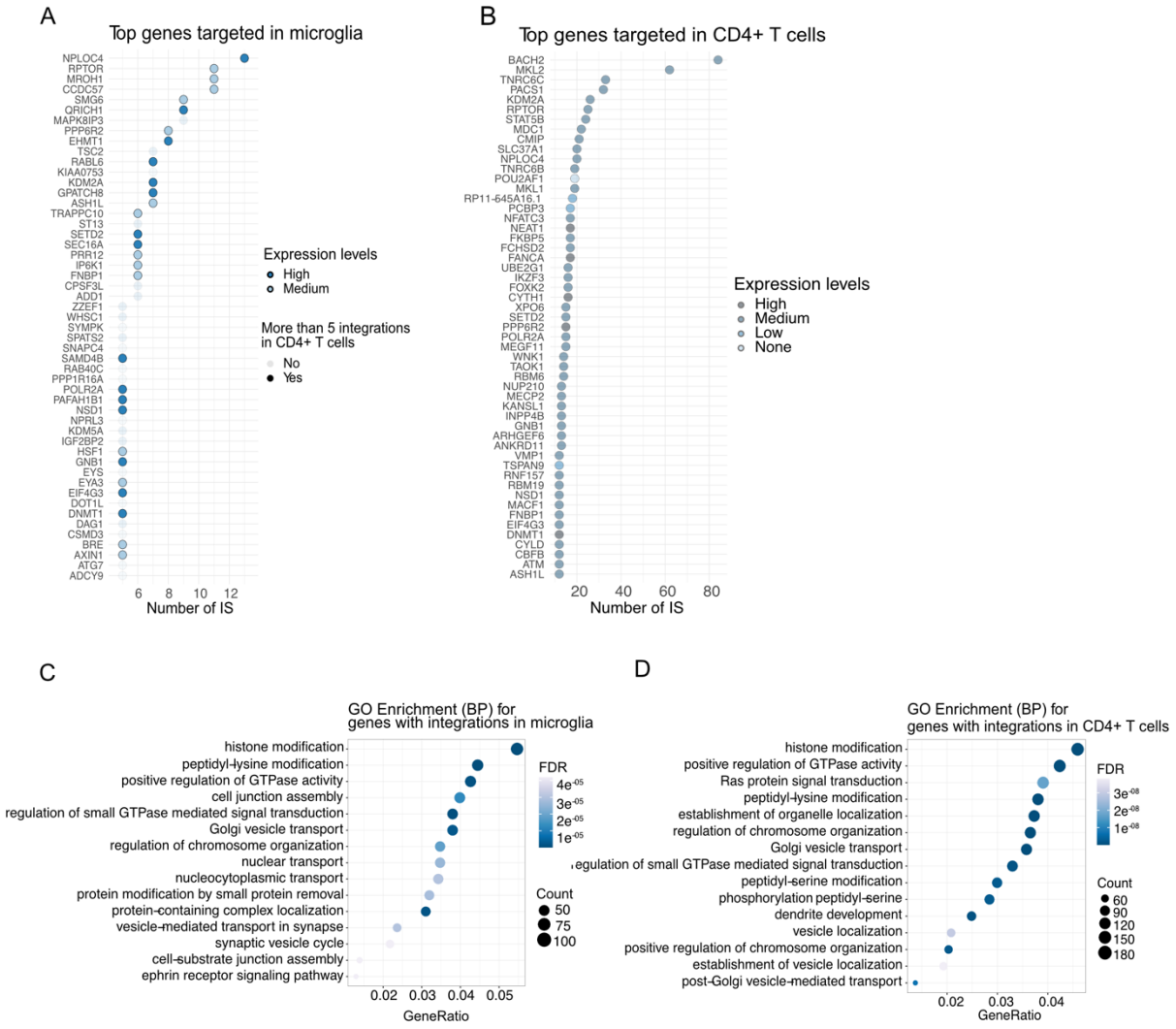


Figure 14: Genes highly targeted in microglia cellular model are also frequently targeted in CD4⁺ T cells. (A) Top 30 IS genes in microglia targeted with ≥ 5 IS. Dark blue: high expression level of the gene, light blue: medium expression of the gene. Circled dot: gene harbors also > 5 integrations in CD4⁺ T cells[1]. (B) Top 30 IS genes in CD4⁺ T targeted with ≥ 12 IS. Dark blue: high expression level of the gene, light blue: medium expression of the gene, grey: low expression, light grey: no expression[1]. Gene ontology analysis of biological processes of IS target genes in microglia (C) or CD4⁺ T cells (D). Color indicates significance (FDR), size of the dot shows counts of the genes part in the respective pathway[1]. Figure was adapted from (Rheinberger et al., 2023).

2.1.4 Genic integration requirements are corroborated in primary microglia cell model

Gene expression profiles of microglia cells change due to environmental signals e.g., when culturing them for a longer time ex vivo. This might also influence HIV-1 integration profiles. Therefore, in addition to the immortalized microglia cell model, integration sites from induced pluripotent stem cell (iPSC) derived microglia cells were

sequenced to obtain ISs from a model system that is closer to primary microglia cells. The cells were infected with the VSVG-HIV-1 GFP virus at MOI of 0.5 in the laboratory of Jonathan Karn and the obtained DNA was subjected to the LM-PCR protocol in our laboratory, enabling the sequencing of virus-host genome junctions. From one infection, I subsampled 2 LM libraries, with an average fragment size of 411 bp and 426 bp and molarities of 49 nM and 91 nM (Figure 15A,B). 2639 integration sites out of 2 technical replicates were obtained. Chromosome distribution showed similar enrichment of IS to C20 with most of IS on chromosome 17 and 19 (Figure 15C)[1]. Using a pseudobulk scRNA-Seq data of uninfected iPSC-derived microglia (also performed in the lab of Jonathan Karn) expression levels of all genes were segregated into four classes based on their expression levels and showed that large proportion of IS mapped to high and mid expressed genes (95%) (Figure 15D)[1]. Overall feature distribution of IS displayed similar integration frequency into introns in iPSC-derived microglia (61%) and C20 cells (60%), as well as distal intergenic regions (iPSC-derived microglia 14.2%, C20 14.5%) and promoter regions (iPSC-derived microglia 20.1%, C20 20.1%), thereby indicating a comparable HIV-1 integration landscape in brain microglia cellular models (Figure 15E)[1].

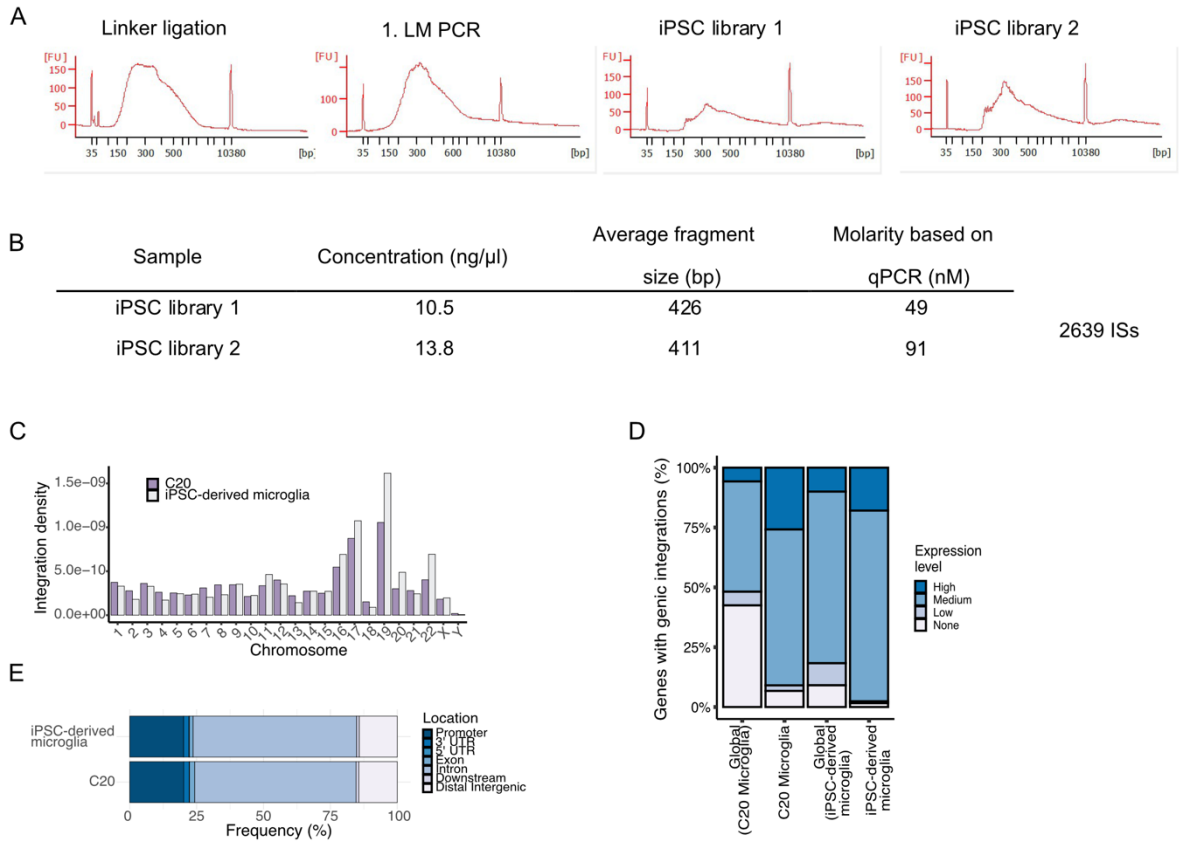


Figure 15: HIV-1 integrates into introns of active genes in iPSC-derived microglia. (A) Bioanalyzer profiles of linker ligation, 1. LM PCR, and 2 technical replicates of second LM PCR. X-axis shows fragment length in bp, Y-axis displays peaks of certain fragments size in fluorescent units. (B) Table showing the QC values of the iPSC integration sites libraries including concentration, average size and library molarity. (C) Integration density per chromosome for C20 microglia (purple bars) and iPSC-derived microglia (light grey bars)[1]. (D) Genes were stratified into high (dark blue), medium (light blue), low (grey) and non-expressed (light grey) genes based on RNA-Seq of C20 cells or pseudobulk scRNA-Seq for iPSC-derived microglia cells (global). Integration genes were stratified in the same way to determine % of genes with genic integration per expression class[1]. (E) Bar plots showing frequency of integration into Promoter regions (Microglia: 20.1%, iPSC-derived microglia: 20.1%), 3' UTR (Microglia: 2%, iPSC-derived microglia: 2%), 5' UTR (Microglia: 0.1%, iPSC-derived microglia: 0.2%), Exons (Microglia: 1.9%, iPSC-derived microglia: 1.4%), Introns (Microglia: 60.2%, iPSC-derived microglia: 61%), downstream (Microglia: 0.9%, iPSC-derived microglia: 1%, from transcription end site (TES) to 3 kb downstream) and distal intergenic regions (Microglia: 14.5%, iPSC- derived microglia: 14.2%) for C20 and iPSC-derived microglia[1]. Figure was adapted from (Rheinberger et al., 2023).

2.1.5 HIV-1 target genes are associated with active chromatin marks in microglia cellular model

To map the chromatin profiles surrounding the integration site, chromatin immunoprecipitation in uninfected C20 microglia cells was performed followed by deep sequencing of the chromatin mark H3K36me3, found at active gene bodies, H3K27ac and H3K4me1, which identify active enhancers, H3K27me3 and H3K9me3, which are associated with heterochromatin and H3K9me2 which marks facultative repressed chromatin. Chromatin IPs of H3K27ac and one replicate of H3K9me2 were generated by Martin Kampmann during his master thesis.

For chromatin IP, C20 cells were crosslinked with 1% formaldehyde for 7 min, stopped with glycine and after nuclei isolation subjected to sonication. For efficient chromatin IP and subsequent sequencing DNA fragment size of the input should be around 500 bp (Figure 16A). After chromatin IP overnight, several washing steps and decrosslinking, ChIP efficiency was tested by qPCR. ChIP qPCR results of H3K36me3 showed a good enrichment over input of GAPDH intron 7 positive control region and absence of signal from the negative primer gene desert region. For H3K27me3 signal was enriched over input at the promoter region of myelin transcription factor 1 (MYT1) and depleted from GAPDH and gene desert (Figure 16B). After library preparation, average fragment size was around 500 bp for all 4 libraries (Figure 16C).

H3K4me1 ChIP qPCR displayed enrichment over input at GAPDH intron 1 and no signal at the satellite 2 (SAT2) repeat region, while H3K9me3 ChIP showed the opposite trend. Negative gene desert region had no signal for both (Figure 16D). Size distribution showed an average of 500 bp for the H3K4me1 libraries and around 700 bp for the H3K9me3 libraries. H3K9me2 ChIP displayed enrichment over input of SAT2 and negative primer set and depletion at the GAPDH intron 1 region, as well as a size distribution of around 600 bp (Figure 16F,G). For all ChIPs, rabbit and mouse IgG controls gave minor to no signal. All libraries were sent out for sequencing together with the respective inputs.

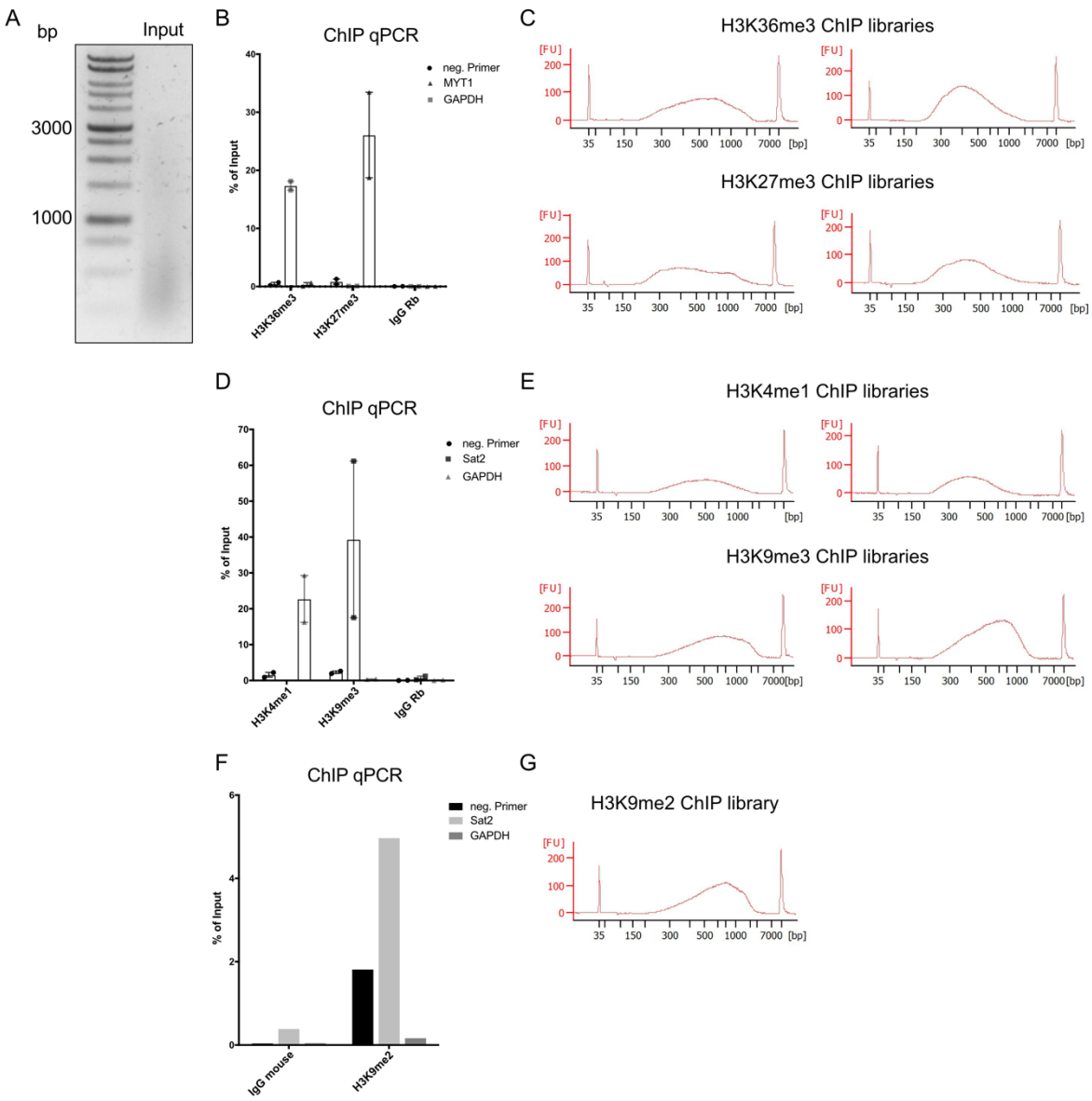


Figure 16: Quality controls of histone chromatin immunoprecipitation libraries. (A) Fraction of de-crosslinked input chromatin loaded onto 2% agarose gel to probe fragment size after sonication. Gel was run at constant 90 V for 45 min. 1000 bp DNA ladder served as fragments size control. (B) Bar plots showing percentage of input results after H3K36me3 and H3K27me3 ChIP in C20 cells of 2 biological replicates. Rabbit IgG (IgG Rb) served as negative control. GAPDH primer set was used as positive control for H3K36me3 and negative control for H3K27me3, MYT1 primer set served as positive control for H3K27me3 and negative control for H3K36me3. Negative primer set was used as negative control for both ChIPs. Data represent mean ± SEM and plot was generated with GraphPad prism. (C) Bioanalyzer profiles of 2 biological replicates of H3K36me3 and H3K27me3 ChIP-Seq libraries. X-axis shows fragment length in bp, Y-axis displays fluorescent units. (D) Bar plots showing percentage of input results after H3K4me1 and H3K9me3 ChIP in C20 cells of 2 biological replicates. Rabbit IgG (IgG Rb) served as negative control. GAPDH primer set was used as positive control for H3K4me1 and negative control for H3K9me3, SAT2 primer set served as positive control for H3K9me3 and negative control for H3K4me1. Negative primer set was used as negative control for both ChIPs. Data represent mean ± SEM and plot was generated with GraphPad prism. (E) Bioanalyzer profiles of 2 biological replicates of

H3K4me1 and H3K9me3 ChIP-Seq libraries. X-axis shows fragment length in bp, Y-axis displays fluorescent units. (F) Bar plots showing percentage of input results after H3K9me2 ChIP in C20 cells of 1 biological replicate. Mouse IgG (IgG mouse) served as negative control. GAPDH primer set was used as negative control and SAT2 and negative primer set served as positive control for H3K9me2. Plot was generated with GraphPad prism. (G) Bioanalyzer profiles of 1 biological replicate of H3K9me2 ChIP-Seq library. X-axis shows fragment length in bp, Y-axis displays fluorescent units.

In addition to chromatin marks, I also profiled accessible chromatin regions in C20 cells using ATAC-Seq in uninfected cells. ATAC-Seq is an assay to profile genomic regions that exhibit accessible chromatin, i.e., regions that are not occupied by nucleosomes using a protocol developed by Buenrostro et al. (Buenrostro et al., 2015).
For ATAC-Seq, nuclei were isolated and tagmented with Tn5 transposase, which cuts and inserts adapters at accessible genomic sites. DNA is then isolated and amplified by PCR, to introduce sequencing indices (Figure 17). Quality assessment of ATAC-Seq library is described in Figure 24.

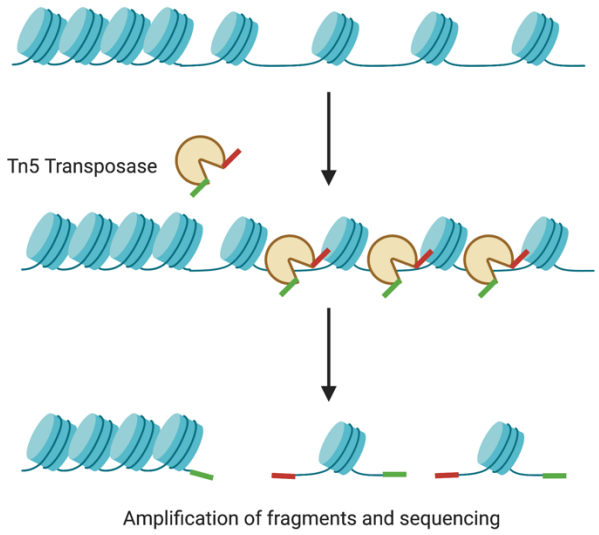


Figure 17: ATAC sequencing. Schematic representation of ATAC-Seq protocol. Nuclei are isolated and digested with Tn5 transposase, which is cutting the DNA that is not occupied by nucleosomes and at the same time inserting illumine sequencing adapters. Subsequently, DNA is isolated and amplified by PCR, to introduce indexes. Scheme was generated with BioRender.

First, integration genes were plotted from the TSS to the TES, stratified according to their expression levels into high, mid, low and non-expressed, and mapped to all chromatin marks tested. H3K36me3 was enriched at gene bodies of high and mid expressed integration genes, whereas H4K4me1 and H3K27ac were enriched at TSS

scalable with transcription levels (Figure 18A)[1]. The repressive chromatin marks H3K9me2, H3K9me3 and H3K27me3 were underrepresented at integration genes at all expression classes (Figure 18B)[1]. ATAC-Seq signal was enriched the most at TSS of integration genes in high and mid expression class, and to a lesser extend in low and non-expressed (Figure 18C)[1].

Next, genes were stratified into genes with and without ISs in high, mid, low and non-expressed class and epigenetic marks and ATAC-Seq signal were profiled. H3K36me3 and H3K27me3 were similarly enriched or depleted, respectively, at genes with and without IS in high, mid and low expression class, whereas in the non-expressed genes, genes harboring IS were enriched in H3K36me3 and depleted in H3K27me3, suggesting that in addition to expression levels of the integration genes, surrounding chromatin plays a role in IS targeting (Figure 18D)[1]. Also, ATAC-Seq signal and enhancer marks (H3K4me1 and H3K27ac) were enriched at TSS of IS target genes in the low and non-expressed genes (Figure 18E,F)[1], further supporting the role of chromatin in addition to gene expression on HIV-1 integration genes.

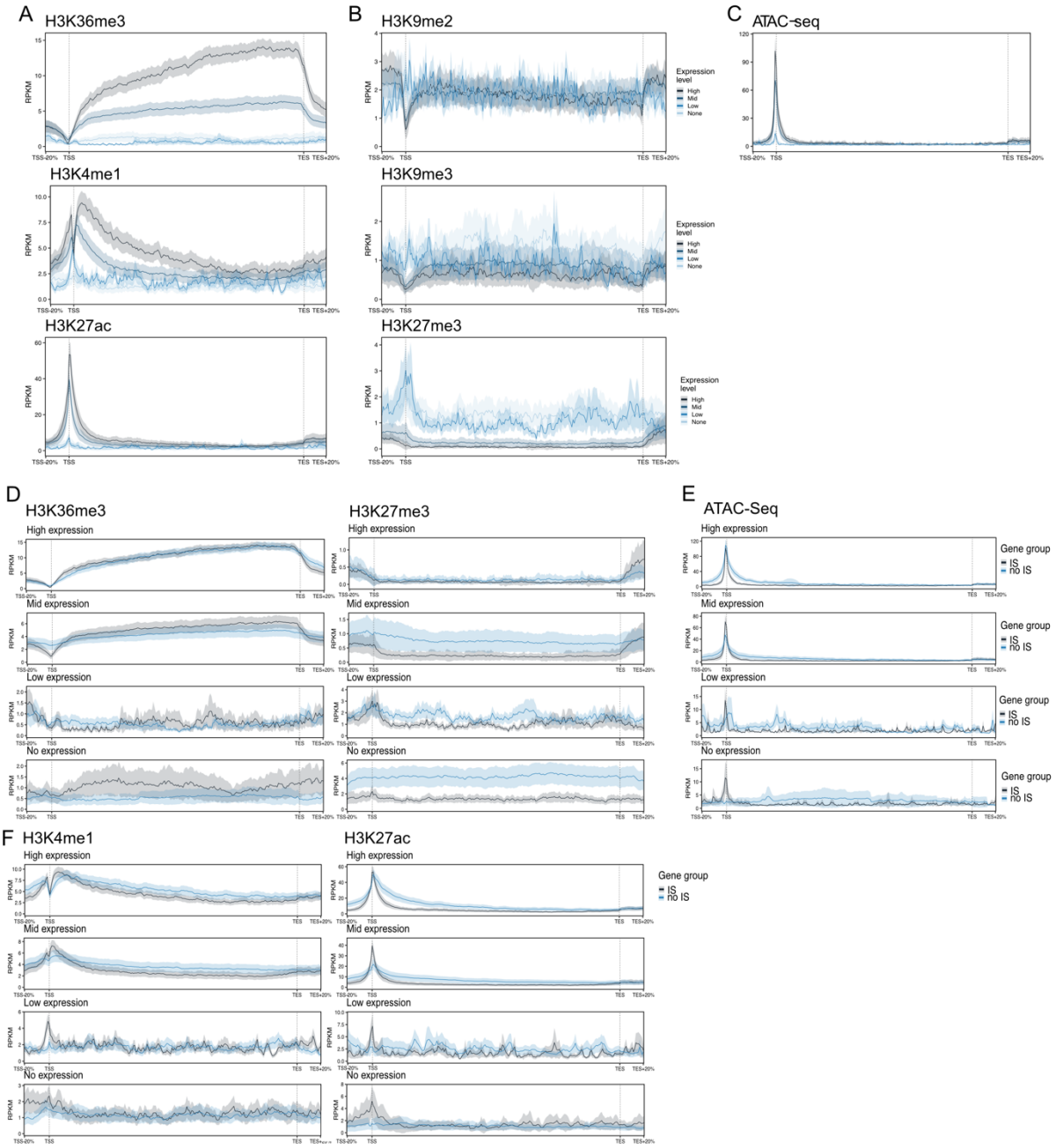


Figure 18: HIV-1 target genes are associated with active chromatin marks. Metagene plots of ChIP-Seq signal (in Reads per kilobase per million mapped reads (RPKM)) of genes containing IS in microglia divided according to their expression levels into 4 groups: high (black), mid (dark blue), low (light blue) and non-expressed (gray). Plots show the full gene body and 20% of its length upstream of the transcription start site (TSS) and downstream of the transcription end site (TES). Confidence intervals (95%) are shown as shaded regions. (A) H3K36me3, H3K4me1, H3K27ac (B) H3K9me2, H3K9me3 and H3K27me3 C) ATAC-Seq signals[1].

Metagene plots of ChIP-Seq signal (in RPKM) of genes containing IS in microglia by transcription level (N(high)=710; N(mid)=1,788; N(low)=64; and N(none)=185). Plots show the full gene body plus 20% of its length upstream of the TSS and downstream of the TES. Averaged signal along genes with IS in microglia is shown in dark gray. Averaged signal along randomly sampled gene set of equal size is shown in light blue. Confidence intervals (95%) are represented as the light-colored regions. (D) for

chromatin marks H3K36me3 and H3K27me3. (E) for ATAC-Seq. (F) for chromatin mark H3K4me1 andH3K27ac[1]. Figure was adapted from (Rheinberger et al., 2023).

2.1.6 Chromatin landscape surrounding HIV-1 insertions in microglia cellular model is characterized by high H3K36me3 signal

In addition to genic characterization of HIV-1 IS in microglia, the chromatin encompassing the precise insertion sites within a range of plus/minus 10 kb compared to random matching controls (RMC) was subsequently inspected. The RMC IS data set was generated to ensure that the sites had equivalent distances to a transcription start site (TSS) as the actual sites, as well as a comparable ratio of genic to intergenic regions. H3K36me3 chromatin mark was enriched at the IS when compared to the RMC, while enhancer marks H3K4me1 and H3K27ac were depleted at IS but enriched in the vicinity regions. All repressive chromatin marks tested (H3K9me2, H3K9me3 and H3K27me3) were underrepresented at the IS (Figure 19A)[1].

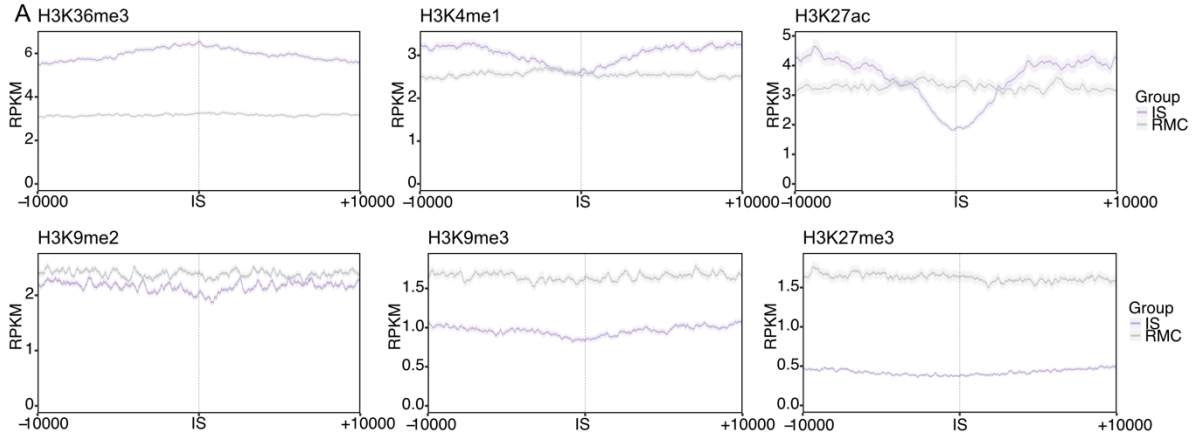


Figure 19: Chromatin mark distribution surrounding HIV-1 integration sites. (A) ChIP-Seq signal of histone marks in C20 cells in RPKM for the chromatin marks H3K36me3, H3K4me1, H3K27ac, H3K9me2, H3K9me3, and H3K27me3. X-axis shows the IS vicinity (±10 kb). Microglia IS in purple, random matched controls (RMCs) in gray. Confidence intervals (95%) are shown as light-colored regions. RMC were generated with the same distance to the nearest TSS and the same intergenic/genic ratio[1]. Figure was adapted from (Rheinberger et al., 2023).

When comparing genic and intergenic IS with their corresponding random matched control (RMC) set, a higher enrichment of H3K36me3 was observed at the genic sites compared to the intergenic sites (Figure 20A left panel). In the comparison between intergenic IS and their intergenic RMC, a higher level of gene body chromatin mark was detected at IS. The same was true for the enhancer marks, intergenic IS had less

enrichment compared to genic ones but more enriched compared to intergenic RMC (Figure 20A middle and right panel). For the repressive chromatin marks H3K9me3 and H3K27me3 they were also underrepresented for the intergenic IS compared to RMC (Figure 10B middle and right panel). H3K9me2 was unchanged (Figure 20B left panel)[1]. In summary, HIV-1 IS in C20 cells associate with the gene body chromatin mark H3K36me3 and have enriched enhancer marks within ± 10 kb in their vicinity.

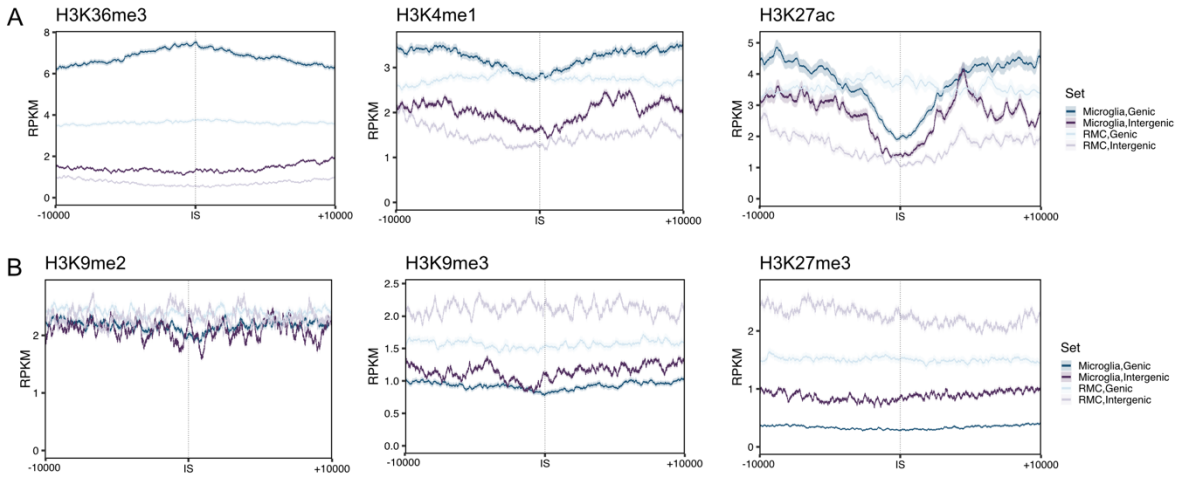


Figure 20: Chromatin mark distribution surrounding HIV-1 integration sites at genic and intergenic sites. (A) ChIP-Seq signal of histone marks in RPKM for the chromatin marks H3K36me3, H3K4me1 and H3K27ac. Microglia IS were divided into genic (blue) and intergenic sites (purple). X-axis shows the IS vicinity (±10 kb). RMC sets are shown in lighter blue for genic sites and lighter purple for intergenic IS. Confidence intervals (95%) are shown as light-colored regions. RMC were generated with the same distance to the nearest TSS. (B) ChIP-Seq signal of histone marks in RPKM for the chromatin marks H3K9me2, H3K9me3 and H3K27me3. Microglia IS were divided into genic (blue) and intergenic sites (purple). X-axis shows the IS vicinity (±10 kb). RMC sets are shown in lighter blue for genic sites and lighter purple for intergenic IS. Confidence intervals (95%) are shown as light-colored regions. RMC were generated with the same distance to the nearest TSS[1].

2.1.7 HIV-1 IS partition with active transcription and genic and super enhancers in microglia cellular model and CD4⁺ T cells

To obtain a comprehensive overview of the HIV-1 integration landscape in microglia and to achieve a detailed characterization at high resolution (200 bp bins, corresponding to nucleosomal scale), histone profiles and chromatin accessibility datasets generated in this work were integrated using the 10-state ChromHMM chromatin model (Ernst and Kellis, 2010, 2017). According to the emission signal of histone chromatin IP, RNA-Seq or ATAC-Seq per genomic bin, the genome was

classified into 10 different states. Global overlap of the genome with each state in percentage is stated in the all states column (Figure 21A)[1]. The predominant patterns associated with HIV-1 IS were linked to four specific chromatin states, namely active transcription (35.1%), genic enhancers (5.7%), weak enhancers (5.6%), and a state enriched with H3K27ac and H3K9me2 (4.3%) (One-sided binomial test, p-value ≤ 0.05). Conversely, quiescent states (44.8%), heterochromatin (0.9%), and both high and low Polycomb states (0.3% and 0.5%) were significantly less targeted than expected when compared to the overall distribution of chromatin states (One-sided binomial test, displayed as *All states*, p-value ≤ 0.05) (Figure 21A). In summary, these findings suggest that microglia ISs tend to be located within genomic regions characterized by genic enhancers or regions with high levels of transcriptional activity. It has been established that a key characteristic of HIV-1 integration is the occurrence of hotspots for genic insertion, which are regions of local IS enrichments. HIV-1 hotspots have been identified in both in-vitro studies and in patients (Bedwell et al., 2021; Brady et al., 2009b; Cohn et al., 2015; Francis et al., 2020; Ikeda et al., 2007; Kok et al., 2016; Lucic et al., 2019; Maldarelli et al., 2014; Schroder et al., 2002; Sherrill-Mix et al., 2013; Wagner et al., 2014; Wang et al., 2007). Previous research has described the association of HIV-1 integration with both genic and distal enhancers, including SEs, in T cells (Bedwell et al., 2021; Chen et al., 2017; Francis et al., 2020; Lucic et al., 2019). In order to investigate the previously observed connection between IS and genic enhancers in microglia at the nucleosomal level (as shown in Figure 21A) and to identify additional genomic signatures that are associated with IS on a larger genomic scale, a recently developed non-negative matrix factorization (NMF)-based method was applied (Quintero et al., 2021)[1]. In this approach, epigenetic signatures were assigned to all 50 kb partitions of the genome. The bins were distinguished by their RNA-Seq, ChIP-Seq, and ATAC-Seq signals, encompassing a of total of 8 features. To simplify the initial data matrix, NMF was employed to decompose it into two matrices: the exposure matrix (H) (depicted in Figure 21B) and the signature matrix (W) (shown in Figure 21C)[1]. These matrices share a common dimension, the signatures (k=4). The exposure matrix (H) represents how much each chromatin feature contributes to the signatures, enabling the identification of the biological characteristics present in each signature. On the other hand, the signature matrix (W) assigns each genome bin to a signature, based on the

chromatin features observed in each signature. As a result, the distribution of relevant biological traits (such as HIV-1 IS and SEs) is linked to the identified signatures.

The majority of microglia integrations were found in two signatures: signature 1 (H3K36me3, high transcription/RNA-Seq with 15.3% of bins containing IS) and signature 4 (H3K27ac, H3K4me1, accessible chromatin/ATAC-Seq with 8.3% of bins containing IS). In contrast, there were fewer integrations in signatures 2 (H3K9me2 and H3K9me3 with 1.3% of bins containing IS) and 3 (H3K27me3 with 1.1% of bins containing IS) (Figure 21C). Bins that couldn't be assigned to any specific signature formed a distinct fifth signature.

Comparing the chromatin signatures obtained in the NMF analysis with the locations of SE revealed that signature 4 targeted by HIV-1 was also enriched in SE (12.7% in signature 4, compared to 0.8%, 0.1% and 0.2% in the signature 1, 2 and 3, respectively) (Figure 21C).

The main HIV-1 target cell type are $CD4^+$ T cells with well characterized integration site profiles. To compare if chromatin requirements are preserved in the brain target cell type of HIV-1 infection, NMF analysis was applied to $CD4^+$ T cell chromatin data[1]. NMF-based decomposition identified 4 signatures: signature 1 (H3K27me3), signature 2 (H3K36me3), signature 3 (H3K4me1 and me3, H3K27ac), and signature 4 (H3K9me3). HIV-1 IS in $CD4^+$ T cells partitioned with signatures 2 (H3K36me3, 28.32% of IS overlap) and signature 3 (H3K27ac, H3K4me1, H3K4me3 rich, 32.66%) (Figure 21D), with signature 3 also having a high overlap with SE (27.13%). This analysis confirms that the chromatin landscape of HIV-1 integration profiles is largely conserved between T cells and microglia cellular model (Bedwell et al., 2021; Francis et al., 2020; Lucic et al., 2019).

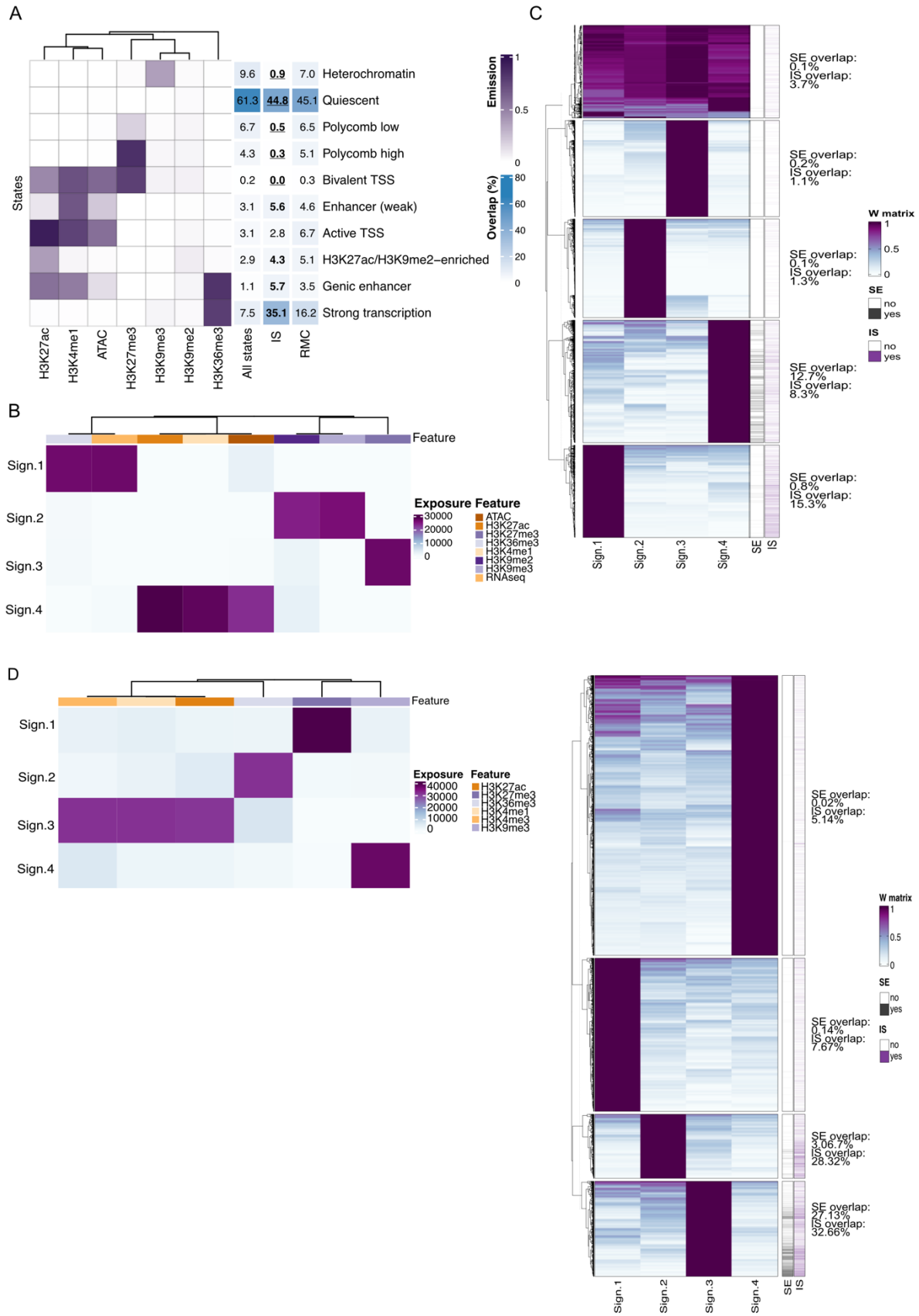


Figure 21: HIV-1 IS associates with genic and super enhancers in microglia cellular model and CD4⁺ T cells. (A) 10-state ChromHMM model generated on microglia genomic data sets. The model includes emissions (in purple) and percentage of overlap (in blue) for each state. Each row represents a different state based on chromatin modifications and chromatin accessibility, while columns indicate

the chromatin features used to create the model (ChIP-Seq and ATAC-Seq). Darker colors in the right panel indicate higher emission probabilities for each feature. The columns also show the global state ratios for the entire genome (All states) and the percentage of overlap between each chromatin state, the IS and the RMC set. Darker colors represent a larger overlap between the regions displayed and the corresponding state. States targeted by integrations significantly more than expected are represented in bold, while states targeted significantly less than expected are represented in bold and underlined text (exact binomial one-sided test, $p < 0.05$)[1]. (B) Summarized ATAC-Seq, ChIP-Seq, and RNA-Seq signal (8 features) over the 50-kb windows (n = 57,238) in C20 microglia cells. Resulting H matrix was visualized as a heatmap. The colors in the heatmap display the exposure values of each feature (columns) in the signatures (rows)[1]. (C) The W matrix of the NMF performed on the summarized ATAC-Seq, ChIP-Seq, and RNA-seq signal over the 50 kb windows (n=57,238) in C20 microglia cells was visualized as a heatmap. In this heatmap, the columns indicate the signatures, and the rows represent the features (50-kb windows). The color in this heatmap displays the contribution of the feature to each signature. The columns on the right side of the heatmap indicate ISs (purple) and SEs (gray) that were mapped to the signature bins, and their fractions are shown as percentages. Unassigned bins were clustered in the upper section[1]. (D) The H matrix resulting from NMF applied to the summarized ChIP-Seq signal of $CD4^+$ T cells over 50 kb windows (n=57,238) is represented as a heatmap. The colors indicate the exposure values of the six features in the signatures. Heatmap representing the W matrix resulting from NMF applied to the summarized ChIP-Seq signal of $CD4^+$ T cells over 50 kb windows (n=57,238). The rows represent the features (50 kb windows) and the columns represent the signatures. The color indicates the contribution of each feature to each signature. The right-hand columns display the fractions of IS (purple) and SE (gray) mapped to the signature bins, with percentages indicated[1]. Figure was adapted from (Rheinberger et al., 2023).

2.1.8 LEDGF/p75 and CPSF6 host factors are involved in HIV-1 integration process in microglia cell model

LEDGF/p75 is the main viral integrase host interaction partner, leading the preintegration complex towards H3K36me3 gene bodies and therefore being essential for gene body targeting of HIV-1 in $CD4^+$ T cells (Debyser et al., 2015; Van Maele et al., 2006). Depletion of LEDGF/p75 in T cells or HEK293T cells diminishes integration and redirect the remaining IS to less gene dense regions (Bedwell et al., 2021; Sowd et al., 2016; Vranckx et al., 2016). To test whether HIV-1 integration process in microglia cells depends on LEDGF/p75, I employed siRNA knockdown to transiently deplete LEDGF/p75 in C20 cells, followed by infection after 48h with VSVG HIV-1 (Figure 22A,B). Integration levels were assayed by Alu PCR and showed a significant decrease in HIV-1 integration to 0.1 FC over NT ($p\text{-value} \leq 0.0001$) in the absence of LEDGF/p75 (Figure 22C).

A second important host factor in HIV-1 integration in T cells is CPSF6, a CA interaction partner (Lee et al., 2012; Sowd et al., 2016). CPSF6 is known to direct the PIC into speckles associated domains and depletion of CPSF6 halts PIC at the nuclear periphery in T cells and macrophages (Bedwell et al., 2021; Francis et al., 2020; Li et al., 2020a). Transient depletion of CSPF6 by siRNA transfection in C20 cells and subsequent HIV-1 infection also led to a significant decrease in HIV-1 integration levels to 0.2 FC over NT (p-value ≤ 0.001) (Figure 22C).

In summary, these findings support the notion that LEDGF/p75 and CPSF6 are universal HIV-1 integration host factors.

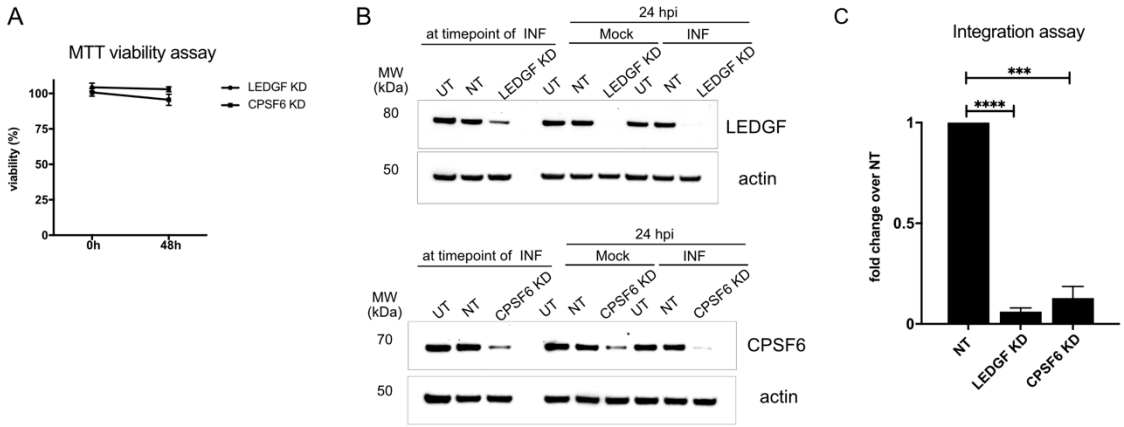


Figure 22: LEDGF/p75 and CPSF6 are important for HIV-1 integration in microglia cellular model. (A) MTT assay in C20 cells was performed at 0h and 48h post transfection with siRNA pool targeting LEDGF/p75 (50 nM) or CPSF6 (50 nM) and compared to non-targeting (NT) siRNA pool (50 nM). Data represent mean of 3 biological replicates ± SE and plot was generated with GraphPad prism. (B) Representative immunoblot image for LEDGF KD (upper panel) and CPSF6 KD (lower panel) at timepoint of infection and 24 h post infection. Proteins were detected by immunoblotting using the following antibodies: LEDGF/p75, CPSF6 and beta actin. Actin served as loading control. UT: untreated, NT: non-targeting. (C) Bar plot representing 3 biological replicates of integration assay by Alu PCR. C20 cells were infected for 24 h with 250 ng p24 / 1 mill of cells after LEDGF KD or CPSF6 KD for 48 h by siRNA transfection. Data are presented as mean ± SEM. Unpaired student's t-test was performed using GraphPad prism; **** = p≤0.0001; *** = p≤0.001.

2.2 HIV-1 latency in microglia cellular model

2.2.1 Establishing HIV_{GKO} virus infection in microglia cellular model

To date there is a lack of treatment that specifically targets the stably integrated virus, leading to the persistence of latent reservoirs throughout the body, including the CNS and microglia cells. The development of HAND in HIV-1 infected people has a significant impact on the quality of life and overall health of individuals living with HIV-1 (Clifford and Ances, 2013). Therefore, it is important to study HIV-1 latency and to discern the differences between actively replicating and latent microglia cells.

To start addressing this question I employed a dual color virus that harbors a csGFP under the viral LTR promoter and a kusabira orange fluorophore under the ef1alpha promotor (Battivelli and Verdin, 2018). The latter is constitutively expressed when the virus is integrated, whereas the GFP is only expressed when HIV-1 is replicating. C20 cells were infected with VSVG pseudotyped HIV_{GKO} virus, and 3 d.p.i. integrated virus was visualized by HIV-1 DNA FISH (Figure 23A) and integration levels were assayed by Alu PCR (Figure 23B). In addition, mRNA was extracted to test for gag mRNA expression (Figure 23C). Altogether, HIV_{GKO} established a productive infection in C20 microglia cellular model.

Due to the reporter expression, latent and actively replicating cells can be distinguished and sorted using a FACS sorter (Figure 23D). For the uninfected and actively replicating cells (10% of the population) approximately 50,000 cells were sorted, for the latent population (0.7 – 1 %) from 8,000 to 45,000 cells were sorted and checked for purity (Figure 23E,F). RNA extraction was performed of the uninfected, the actively replicating and the latent cells and by qPCR expression levels of viral gag mRNA was tested (Figure 23G). As expected, viral gag mRNA was most expressed in actively replicating cells, and only on a basal level expressed in the latent cells, when compared to the uninfected cells (Figure 23G, unpaired student's t-test; ** = $p \leq 0.01$).

In response to viral infection, cells upregulate inflammatory response genes (Catalfamo et al., 2012; Deeks et al., 2013; Kamat et al., 2012; Vandergeeten et al., 2012). To investigate inflammatory response within microglia latency model system, mRNA expression of key pro-inflammatory cytokines, CXCL8, IL6 and IL1b, was tested. Inflammation marks were upregulated in active and latent cells compared to uninfected (between 2 to 4 fold increase) (Figure 23H). For IL6 and IL1b there was a

trend of more expression in the latent cells compared to the active ones, without reaching statistical significance.

Altogether, integration, gene expression and FACS assays support that HIV_{GKO} infection is a good model system to study HIV-1 latency in microglia cellular model.

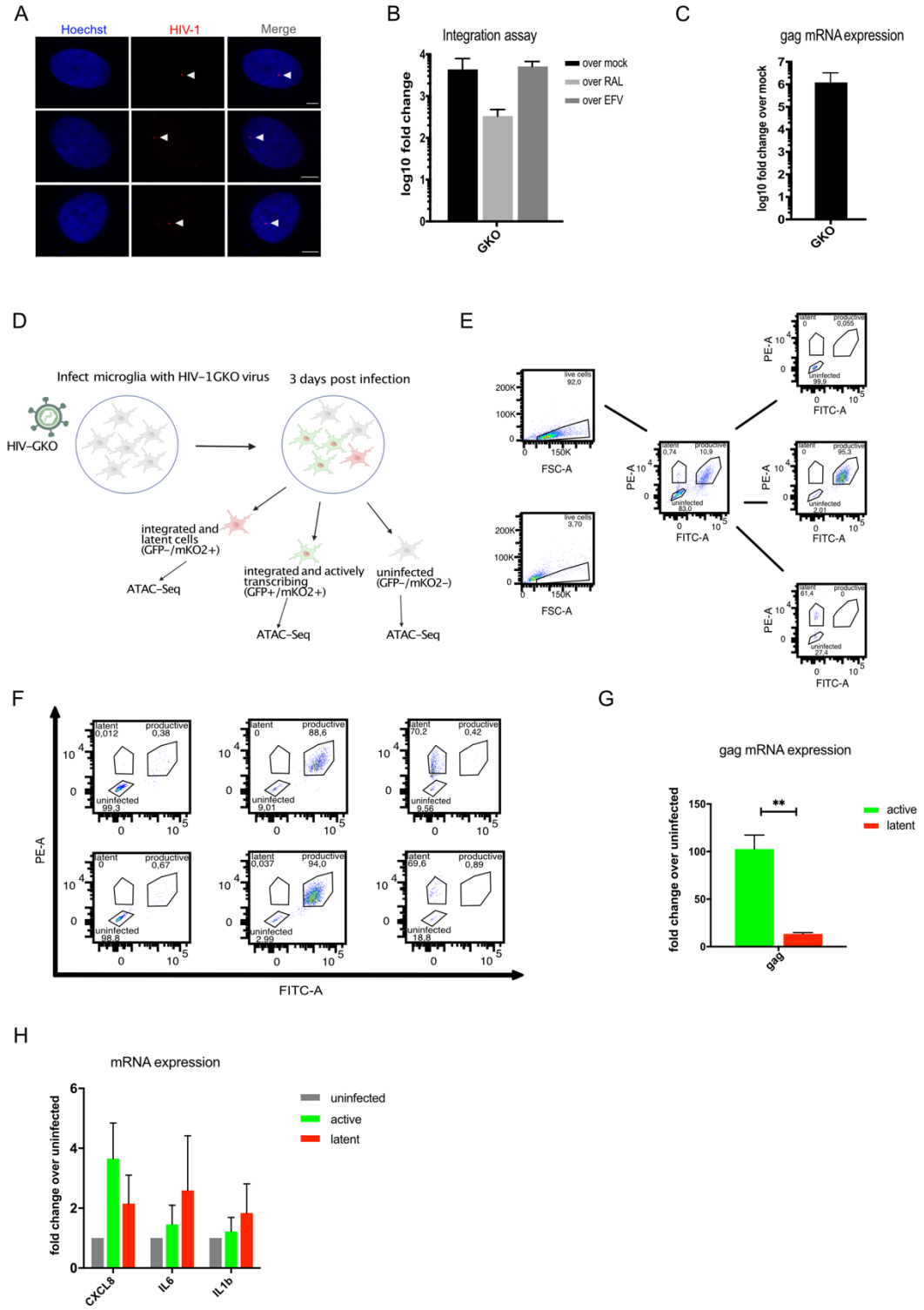


Figure 23: Sorting actively and latently infected C20 cells using HIV_{GKO} virus. (A) HIV-1 3D DNA Immuno-FISH of HIV-1 infected C20 microglia cells (250 ng p24/ 1 mill cells). Cells were fixed, permeabilized and hybridized for 48h with directly labeled HIV-1 FISH probe. Nucleus stained with

Höchst, HIV-1 DNA in red; scale bar 5 μm. (B) Bar plot of HIV-1 integration assay determined by Alu PCR of C20 microglia cells infected with HIV_{GKO} (250 ng p24/ 1 mill of cells). Integration levels are expressed over mock, over Raltegravir (RAL) treated samples (10 μM) or samples treated with Efavirenz (EFV) (20 μM). Data represent mean of 3 technical replicates ± SEM. Graph was generated with GraphPad prism. (C) C20 cells infected with HIV_{GKO} (250 ng p24/ 1 mill cells). RNA of infected cells was harvested 3 dpi and viral gag mRNA levels were assay by qPCR over mock. Bar plot shows mean of 3 biological replicates ± SEM. Graph was generated with GraphPad prism. (D) Schematic representation of sorting set up. C20 cells were infected with HIV_{GKO} virus, and 3 dpi cells were sorted on a BD FACS Aria II instrument into uninfected cells (grey), actively transcribing cells (green and red) and latent cells (red) to perform ATAC-Seq. Scheme was generated with BioRender. (E) Gating strategy for sorting of C20 cells infected with HIV_{GKO} virus. Cells were gated on live cells using SSC-FSC channel. Based on FITC-A and PE-A signal cells were sorted into uninfected ($FITC\text{-}A^-/PE\text{-}A^-$), double positive ($FITC\text{-}A^+/PE\text{-}A^+$) and single positive ($FITC^-A\text{-}/PE\text{-}A^+$). Plot was generated using FlowJo software. (F) Resorting of sorted HIV_{GKO} C20 cells. 5,000 cells were sorted to check sorting purity. Plot was generated using FlowJo software. (G) mRNA was extracted from sorted cell populations and assayed by qPCR for viral gag mRNA expression over uninfected. Bar plot shows mean of 3 biological replicates ± SEM. Statistical analyses was performed with GraphPad prism. Unpaired student's t-test was performed; ** = p≤0.01. (H) mRNA was extracted from sorted HIV_{GKO} infected C20 cell populations and assayed by qPCR for mRNA expression of CXCL8, IL6 and IL1b in uninfected (grey), active (green) and latent (red) C20 cells. Bar plot shows mean of 3 biological replicates ± SEM. Graph was generated with GraphPad prism. Figure was adapted from (Rheinberger et al., 2023).

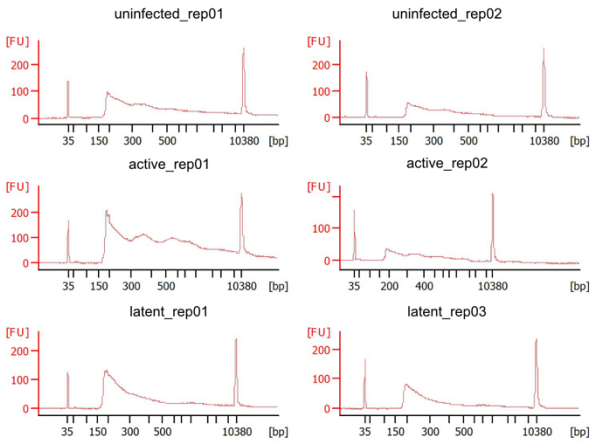
2.2.2 The global chromatin accessibility exhibits comparable patterns in latent and actively replicating cells of microglia cellular model

After successful establishment of HIV_{GKO} infection in C20 cells, the aim was to investigate changes on the chromatin between uninfected, active and latent cells that might influence the outcome of viral infection and latency establishment. To do so, I used this viral construct for sorting uninfected, latent and actively replicating cell populations and subsequently performed ATAC-Seq with the isolated nuclei.

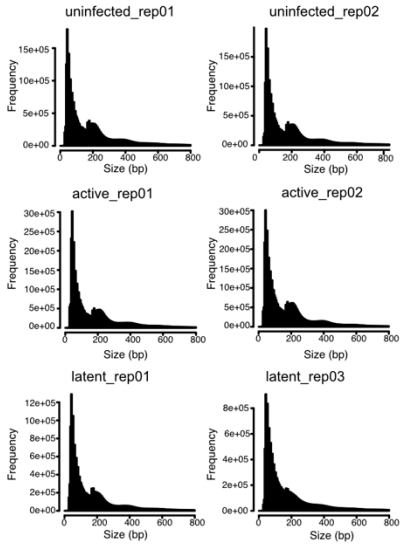
C20 cells were infected with the viral construct and 3 d.p.i. sorted into uninfected, actively replicating, and latent cells. For the uninfected condition, about 80,000 cells were sorted, for the $FITC^-/PE^+$ population 8,000 to 45,000 cells were sorted. Nuclei were isolated and processed for ATAC-Seq assay, which involves the utilization of Tn5 transposase to profile genomic regions that exhibit accessible chromatin, i.e., regions that are not occupied by nucleosomes using a protocol developed by Buenrostro et al. (Buenrostro et al., 2015) (Figure 17). Quality assessment of the ATAC-Seq libraries by Bioanalyzer displayed the characteristic mono-, di- and trinucleosomal size distribution

for the uninfected and active population libraries and to a lesser extend in the latent cells (Figure 24A). After sending the libraries for sequencing, first quality check (QC) of the mapped reads by plotting their read length distribution resembled the previously observed size distribution for the transposase reactions in all 3 cell populations (Figure 24B). The ATAC-Seq analysis yielded a total of 107,949,036 uniquely mapped reads for uninfected samples (replicates merged), 175,833,394 reads for active cells (replicates merged) and 67,889,466 reads for latent cells (replicates merged)[2]. Out of that, 92,812 peaks in uninfected samples, 95,401 peaks in active cells, and 54,744 peaks in latent cells could be called, using a peak calling cutoff of MACS2 Q < 0.001. The fraction of reads in the called peaks (FRIP score) was > 0.2, apart from one latent replicate, which could be due to the low number of nuclei isolated and the risk of over tagmentation (Figure 24C)[1].

A



B



C

Condition	Replicate	Uniquely mapped reads	FRIP score	Number of peaks (score > 30)
Uninfected	Rep01	56,411,459	0.42	
Uninfected	Rep02	56,428,705	0.45	
Uninfected	**merged**	107,949,036		92,812
Active	Rep01	81,547,710	0.48	
Active	Rep02	94,285,684	0.44	
Active	**merged**	175,833,394		95,401
Latent	Rep01	37,225,217	0.25	
Latent	Rep03	30,664,249	0.08	
Latent	**merged**	67,889,466		54,744

Figure 24: Quality controls of ATAC sequencing. (A) Bioanalyzer profiles of ATAC-Seq libraries of 2 sorted uninfected C20 samples (upper panel), 2 sorted actively replicating C20 samples (middle panels) and 2 sorted latent C20 samples (lower panel). X-axis shows fragments length in bp, Y-axis displays fluorescent units. (B) Frequency of read length distribution of uninfected, actively replicating, and latent

[2] Analysis was performed by Dr Carl Herrmann, Health Data Science unit, University Clinics Heidelberg.

ATAC-Seq libraries. ATAC-Seq was performed with 80,000 nuclei of uninfected C20 cells, about 80,000 nuclei of $FITC^{+}/PE^{+}$ actively replicating cell population and 8,000 to 45,000 nuclei of $FITC^{-}/PE^{+}$ latent cell population[2]. (C) Quality controls of ATAC-Seq analysis. Table displays uniquely mapped reads per replicate, FRIP score, and number of peaks called[2]. Figure was adapted from (Rheinberger et al., 2023).

Gene ontology (GO) analysis of the global open genomic regions of each infection population showed in the uninfected cells, 1,285 biological processes (BP) being significantly enriched (FDR < 0.5), in the active cells 1,381 BP terms and in the latent cells 1,290 BP terms (Supplementary Table 2, (Rheinberger et al., 2023)). Looking at the top 30 enriched GO terms showed an enrichment of glia and neurological terms in biological processes in all 3 cell populations (Figure 25A,B and C), supporting the microglia identity of the C20 cells[2].

To investigate the chromatin accessibility differences among active, latent, and uninfected cell populations, differential peak analysis was performed by Ana Luisa Costa and Carl Herrmann. Surprisingly, no significant global changes in chromatin accessibility between the 3 cell states were observed, and only a limited number of peaks, namely 6, 161 and 379 peaks exhibited differential accessibility between uninfected and latent, active and latent, and uninfected and active populations, respectively (Figure 25D, red dots represent significantly changed regions; false discovery rate [FDR] < 0.1)[2]. The 6 genes upregulated in latent cells compared to uninfected included microRNAs, the transmembrane glycoprotein embigin, lysine methyltransferase 2C and pterin-4 alpha-carbinolamine dehydratase 2. For the differential expressed genes in active versus latent cells, they were involved in processes like cell-cell adhesion, cell-junction organization or actin cytoskeleton, while in the active versus uninfected cells genes of glycoprotein metabolic processes, adherence junction organization or negative regulation of cell adhesion were changed. Altogether, no global changes of gene accessibility could be observed between uninfected, actively replicating and latent C20 cells.

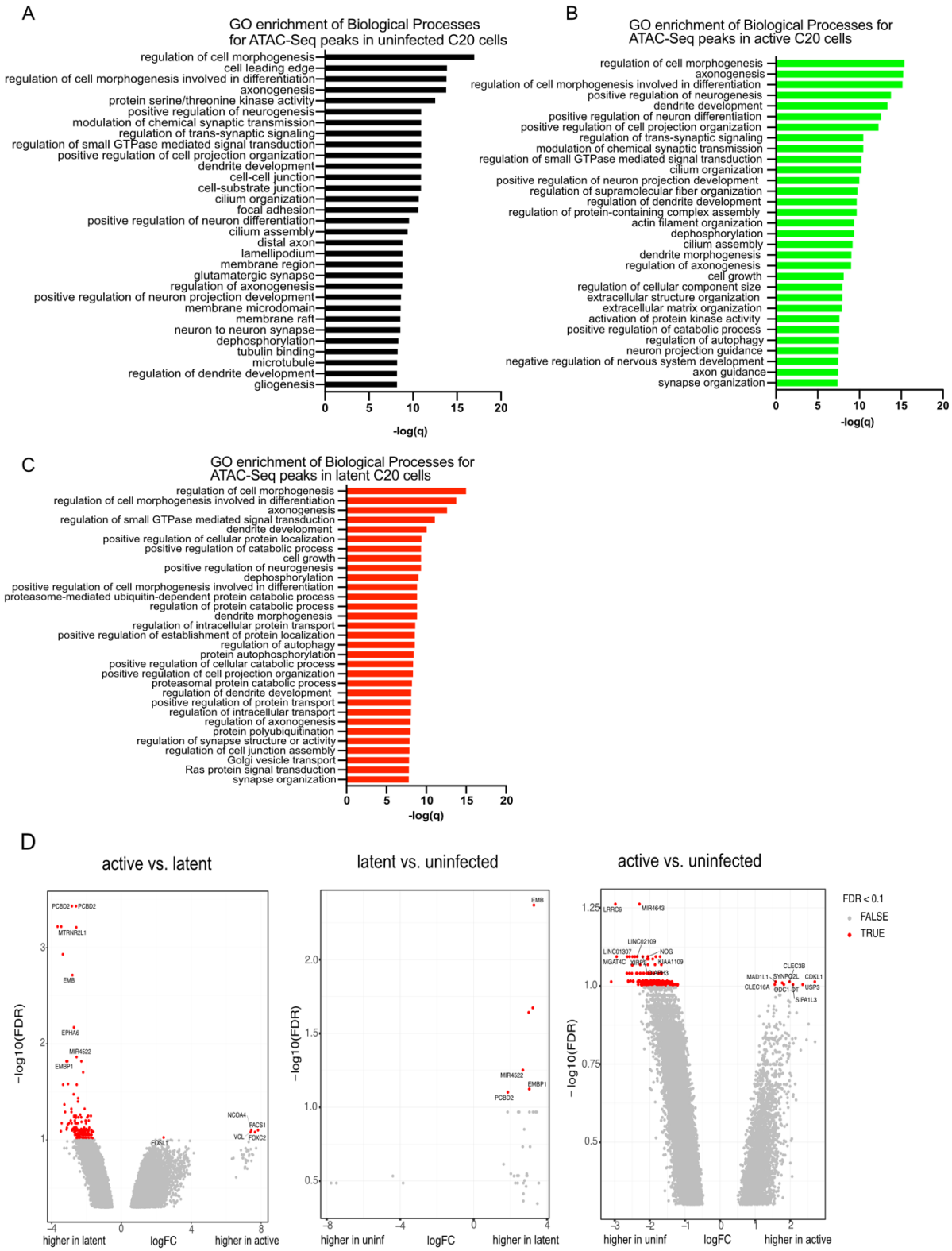


Figure 25: Chromatin accessibility is similar between active and latent HIV-1 infected microglia cellular model. (A) Gene ontology analysis for biological processes of uninfected ATAC-Seq accessible peaks (n=92,812). On the X-axis is displayed -log10 adjusted p-value. Plot was generated with GraphPad prism. (B) Gene ontology analysis for biological processes of active accessible peaks (n=95,401). On the X-axis is displayed -log10 adjusted p-value. Plot was generated with GraphPad prism. (C) Gene ontology analysis for biological processes of latent accessible peaks (n=54,744). On

the X-axis is displayed -log10 adjusted p-value. Plot was generated with GraphPad prism. (D) Differential peak analysis of 2 independent ATAC-Seq experiments. Volcano plots represent log fold change (logFC) versus false discovery rate (FDR) of differentially accessible regions of the cellular genome identified in the three cell states. Significantly changed peaks are displayed in red (FDR <0.1)[2]. Figure was adapted from (Rheinberger et al., 2023).

2.2.3 CTCF is released from the genome in active viral replication

Chromatin accessibility is commonly recognized as a characteristic feature of active regulatory elements that serve as sites where transcription factors are recruited (Tsompana and Buck, 2014). In order to identify candidate regulatory genomic regions in productively and latently HIV-1 infected microglia cells, transcription factor footprinting on the three cell populations was performed using TOBIAS tool (Bentsen et al., 2020)[1]. TOBIAS uses ATAC-Seq data, transcription factor motifs and sequence annotations as input for the analysis. The tool is calculating TF binding scores based on the accessibility of the region and the depth of the TF footprint and based on a threshold bound and unbound sites can be distinguished (Bentsen et al., 2020).

When comparing all 3 cell populations, several transcription factor binding sites (TFBS) were detected to be differentially occupied which have well established roles in HIV-1 transcription and general cell activation, e.g., nuclear factor kB (NF-kB), signal transducer and activator of transcription 1 (STAT1), and activator protein 1 (AP-1) component FOS (Figure 26A,B,C, colored dots indicate overrepresented footprints in the 3 states, FDR < 0.05, predicted differential binding score > |0.05|) (Gosselin et al., 2017; Van Lint et al., 2013; Stroud et al., 2020).

NF-kB was found to be more occupied in the actively replicating and latent cells when compared to the uninfected cells (Figure 26A,C), while STAT1 displayed increased footprint scores in active versus latent cells (Figure 26B), supporting the overall active cell state of the C20 cells and HIV-1 infection. FOS showed enriched occupancy in uninfected cells both compared to latent and to active cells (Figure 26A,C) and enrichment in the latent cells compared to the active ones (Figure 26B).

Interestingly TFBS of the main architectural protein CTCF, which is involved in 3D genome organization and transcriptional regulation, were predicted to be more occupied in the latent and uninfected cells when compared to the active cell population. Looking at the differential occupied CTCF footprints confirmed an overall reduction of predicted CTCF footprints in productive infection (800 footprints in active cells, 2621 in latent and 1556 in uninfected) (Figure 26D)[1].

Next, the focus was on CTCF footprints at least occupied in one of the 3 infection states and profiling them in uninfected, latent and actively replicating cells to further investigate a possible role of CTCF in HIV-1 infection. A large fraction of the TFBS were predicted to be stably bound in all 3 infection populations (blue bars, Figure 26E), but there was also a fraction of TFBS which was dynamically binding. A large portion of the CTCF TFBS stably bound in uninfected became unbound in HIV-1 actively replicating cells, whereas the dynamic sites between active and latent cells were predicted to have an inverse CTCF binding pattern (Figure 26E)[1].

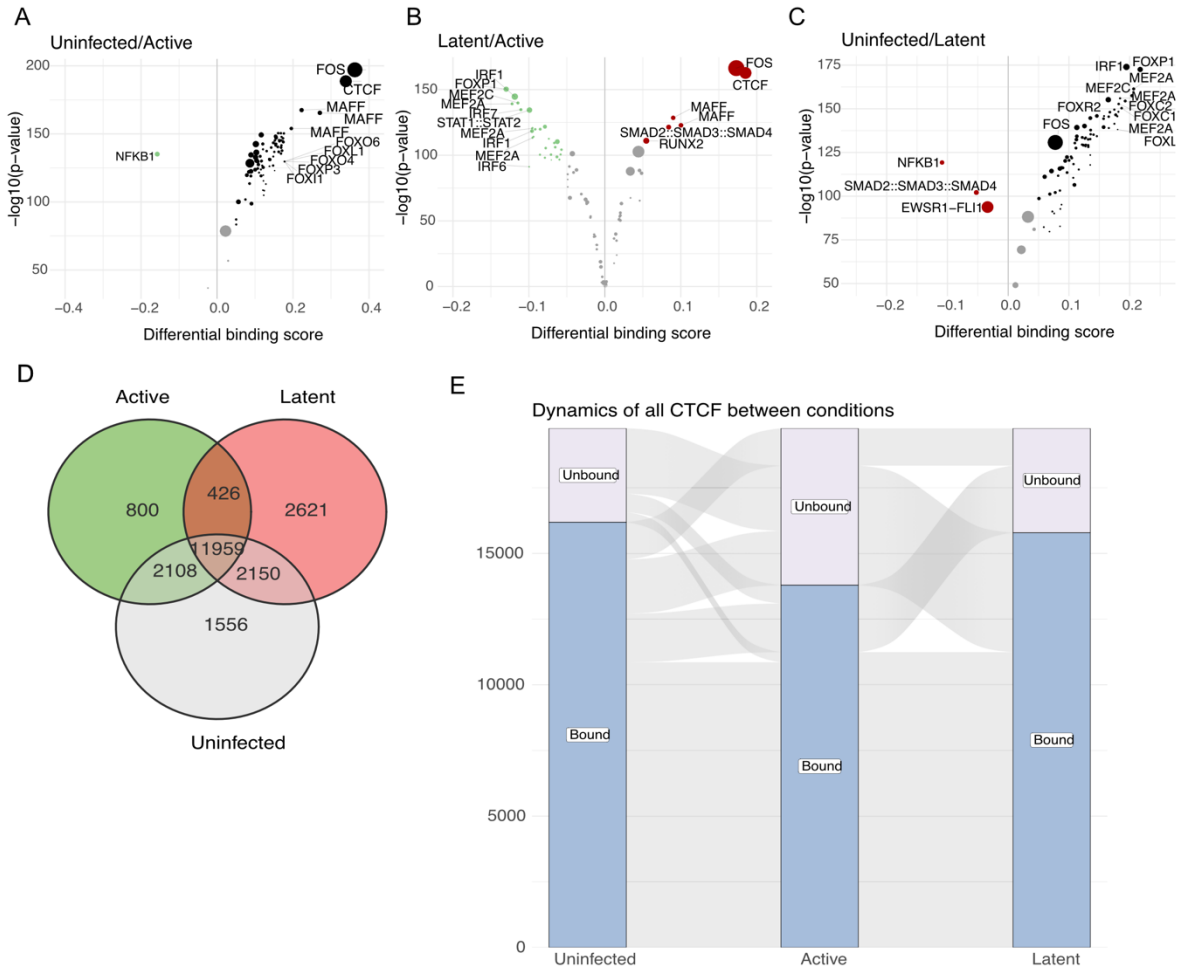


Figure 26: CTCF is released from the genome in active viral replication. Differentially bound TFBS between cell populations presented as volcano plot against the −log10(p- value) of all investigated TF motifs. Higher accessibility is indicated by positive x values, while lower accessibility is indicated by negative values in the first mentioned population. Each dot on the graph represents a single motif. Dot size displays the number of binding sites. (A) Uninfected/active infection (right/left represents a higher signal in uninfected/active, respectively). (B) Latent/active infection. (C) Uninfected/latent infection[1]. (D) Venn diagram displaying the numbers of predicted CTCF footprints found to be shared between the three conditions or uniquely bound. Green: active cell population (800 footprints); red: latent cell

population (2621 footprints); grey: uninfected cell population (1556 footprints)[1]. (E) River chart displaying the binding dynamics of all predicted CTCF TFBSs bound in at least one infection condition. Purple color indicates TFBSs are predicted as bound, lilac color indicates TFBSs are predicted as unbound for the conditions represented on the x axis. Changes between the conditions can be traced by the links between condition bars[1]. Figure was adapted from (Rheinberger et al., 2023).

In order to validate the predicted global depletion of CTCF occupancy on chromatin in productively infected cells, I sought to analyze the protein content from sorted cells. However, low frequency of latent cells (~1%) at 3 d.p.i., precluded obtaining sufficient cell number of the sorted latent cells to perform established protein characterization assay by subcellular protein fractionation. As an alternative approach, I employed biochemical fractionation on bulk infected cell population to isolate tightly chromatin-bound factors by gradually increasing the salt concentration. Therefore, microglia cells were infected with HIV_{GKO} virus at a high infection rate (from 40% to 60% of infection) to obtain most of the infected cells actively producing viral proteins (Figure 27A). 3 d.p.i. biochemical fractionations of the bulk infected cells were performed, to isolate cytoplasm from the nuclear fraction and further separating the nuclear fraction into low salt, high salt, and insoluble fractions (Figure 27B). Alpha tubulin and histone H2B showed a clear separation of cytoplasm and the nuclear fractions, serving as cytoplasmatic and nuclear markers, respectively. Comparison of the nuclear high salt chromatin bound fractions between mock and infected samples showed a release of CTCF from the genome upon HIV-1 infection (Figure 27C,D). Of note, inhibition of viral integration by Raltegravir (10 μM) showed chromatin bound CTCF as in mock samples, supporting the notion obtained by ATAC-Seq that CTCF is released from the genome upon active viral transcription.

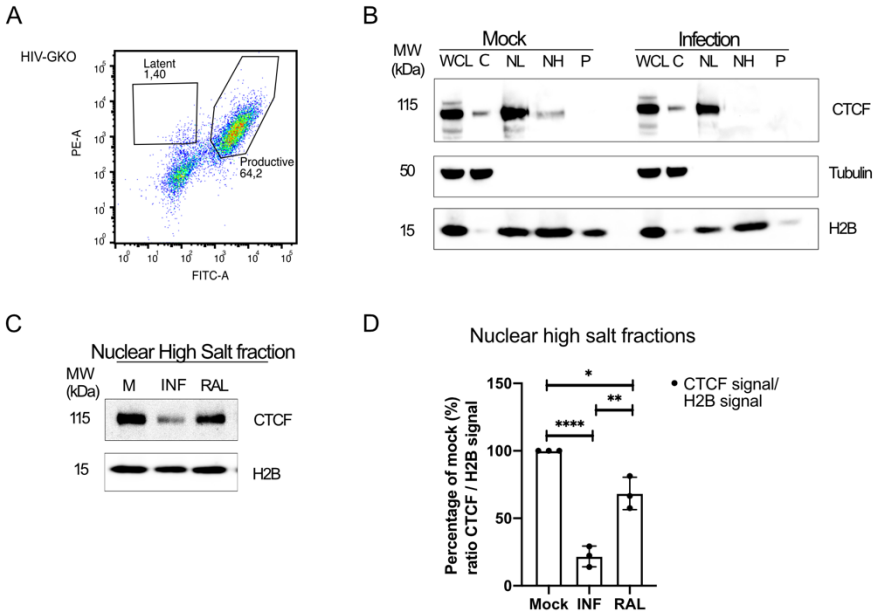


Figure 27: CTCF is released from the nuclear bound fraction in productive HIV-1 infection. (A) Representative FACS plot of C20 cells infected with HIV_{GKO}. Cells were gated for $FITC^+/PE^+$ (productive infection) and $FITC^-/PE^+$ (latent infection). Plot was generated with FlowJo software. (B) Representative immunoblot image of biochemical fractionation of mock and infected C20 cells (250 ng p24 / 1 mill cells). Proteins were detected by immunoblotting using the following antibodies: CTCF, alpha tubulin and histone H2B. WCL: whole cell lysate; C: cytoplasm; NL: nuclear low salt; NH: nuclear high salt; P: insoluble pellet. Alpha tubulin served as cytoplasmic marker, histone H2B served as nuclear marker. (C) Representative immunoblot image of nuclear high salt fractions of mock, infected and infected in the presence of Raltegravir (RAL, 10 μM) C20 cells. Proteins were detected by immunoblotting using the following antibodies: CTCF and histone H2B. (D) Bar plot displaying quantifications of nuclear high salt fractions of 3 biological replicates. CTCF signal was normalized to H2B and expressed as percentage of mock. Data are represented as mean ± SEM. Statistical analyses was performed using GraphPad prism. Unpaired student's t-test was performed; **** = p≤0.0001; ** = p≤0.01; * = p≤0.05.

2.3 3D genome organization plays an important role in HIV-1 integration

2.3.1 HIV-1 integrates in close proximity to the boundaries of TADs in microglia and CD4$^+$ T cells

CTCF is an important factor for 3D genome organization (Wang et al., 2019) and has essential roles in the formation of TADs. TADs are genomic regions that have a high frequency of internal DNA interactions which facilitate contacts of regulatory features e.g. enhancers or TF with their target promoters and genes (Dixon et al., 2012; Hnisz et al., 2013; Lieberman-aiden et al., 2009; Nora et al., 2012; Rao et al., 2014). The architectural protein CTCF is enriched at TAD boundaries, and depletion of the protein can influence TAD structure and loop domains leading to misregulation of genes (Lupiáñez et al., 2015; Narendra et al., 2015; Nuebler et al., 2018; Rao et al., 2017; Wutz et al., 2017).

The differential binding of CTCF in infection identified by ATAC-Seq analysis pointed to a possible involvement of CTCF in HIV-1 infection, including the influence of the 3D chromatin structure. To investigate that possibility, we obtained published Hi-C data of NeuN$^-$ cells of primary brain tissue samples, available through the PsychEncode portal (https://psychencode.synapse.org/). The Hi-C was performed on sorted nuclei negative for the NeuN marker of brains from healthy patients (Hu et al., 2021). NeuN is a neuronal marker absent in glia cells within the brain, meaning that all NeuN negative cells of the brain were sorted (microglia, oligodendrocytes, Astrocytes) to perform Hi-C. At the time of the analysis, this data set represented the sole available Hi-C data of primary microglia-containing cells.

The TAD coordinates were used to plot the relative distance of TAD border, the region insulating two TADs from each other, to TAD midpoint regions[2]. For the analysis, boundary regions were defined as the midpoints between two consecutive TAD intervals. First, microglia ATAC-Seq derived CTCF footprints were mapped to the NeuN-derived TAD coordinates, using relative distances for each TAD (boundary to midpoint). The distribution density plots of CTCF footprints with respect to NeuN$^-$ TAD coordinates showed a characteristic enrichment of CTCF at the borders, as expected (Dixon et al., 2012) (Figure 28A, upper panel). Mapping microglia HIV-1 IS (dark blue line) showed a prominent enrichment towards TAD borders when comparing it to 10 IS subsampling experiments (light blue lines) and to the random control sites (black

dotted line), which is a random IS data set generated based on the same distance to the TSS and genic to intergenic ratio as the microglia IS data set obtained in this study (Figure 28A, middle panel).

Additionally, ChIP-Seq profiles in C20 microglia demonstrated strong enrichment of the gene body and HIV-1 IS-associated histone mark, H3K36me3, in proximity to TAD borders. There were no noticeable variations in the densities of genic and active enhancers (H3K4me1 and H3K27ac) or the heterochromatin mark H3K27me3. However, the signal for H3K9me3 was reduced at the boundaries (Figure 28A, lower panel)[2].

In order to investigate whether the observed association in microglia between TAD borders and enriched HIV-1 ISs is maintained in the major blood target cell type, TAD coordinates from $CD4^+$ T cells were used (Javierre et al., 2016; Nanni et al., 2020). Mapping of $CD4^+$ T cell-derived CTCF ChIP-Seq peaks (Figure 28B, upper panel; GEO: GSE13105515, (Qi et al., 2021)) confirmed TAD border regions. Notably, ISs from ex vivo infections of $CD4^+$ T cells and patient samples (curated previously, (Lucic et al., 2019)) showed enriched density profiles towards TAD borders (Figure 28B, middle panel, dark blue line), followed by H3K36me3, enhancer marks and promoter proximal H3K4me3 (Figure 28B, lower panel)[2]. These results suggest that IS enrichment at transcriptionally active chromatin at TAD borders is a common feature of HIV-1 integration.

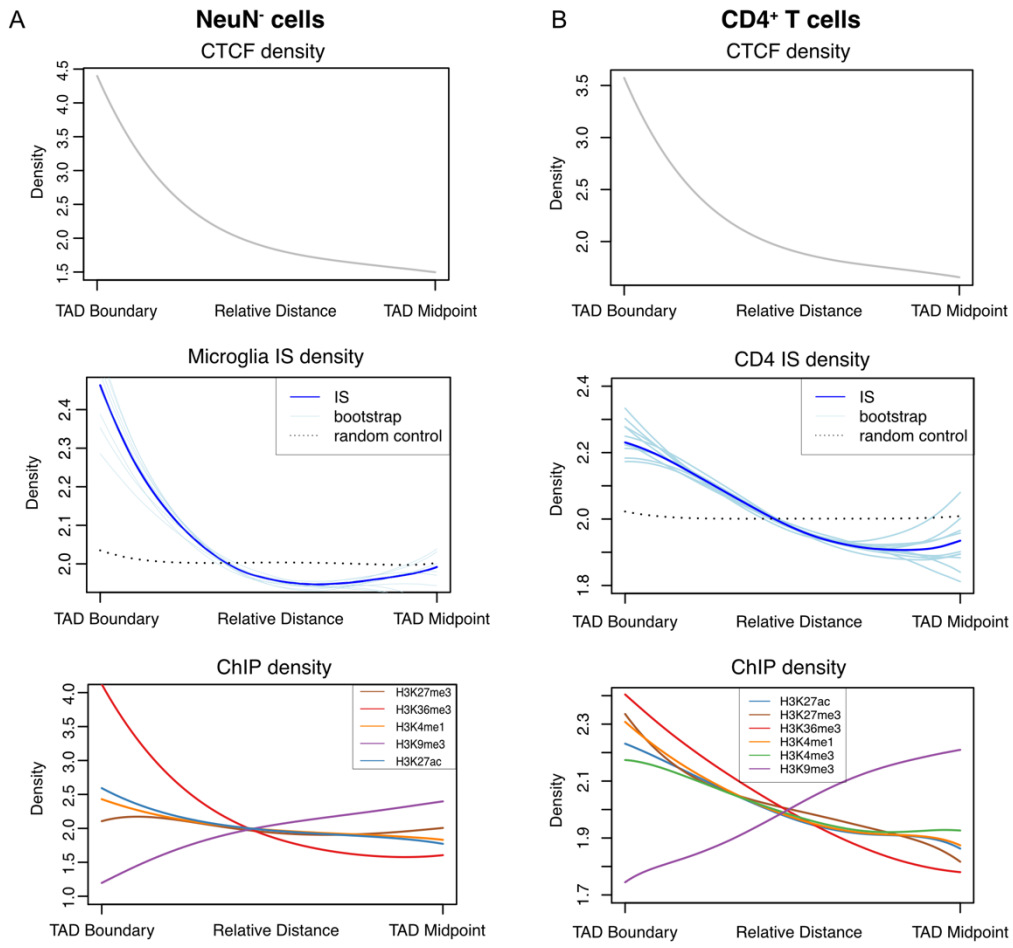


Figure 28: HIV-1 integrates close to TAD boundaries in microglia cellular model and CD4⁺ T cells. (A) CTCF TFBS footprints (gray line, upper panel), ISs in microglia (dark blue line, middle panel), and microglia ChIP-Seq peaks for H3K27me3 (brown line), H3K36me3 (red line), H3K4me1 (orange line), H3K9me3 (violet line), and H3K27ac (blue line) (lower panel) plotted as density location over the boundary-to-midpoint regions of NeuN⁻ (oligodendrocytes, astrocytes, microglia) TAD boundaries (n = 2,077) (Hu et al., 2021). Light blue lines (middle panel) represent 10 bootstrapped subsamples obtained based on the real IS data set and black dotted line represents a random IS control data set[2]. (B) CD4⁺ T cell CTCF binding sites (gray line, upper panel), ISs in CD4⁺ T cells (dark blue line, middle panel), and CD4⁺ T cell ChIP-Seq peaks for H3K27ac (blue line), H3K27me3 (brown line), H3K36me3 (red line), H3K4me1 (orange line), H3K4me3 (green line), and H3K9me3 (violet line) (lower panel) plotted as density location over the boundary-to-midpoint regions of consensus TAD boundaries (n = 2,171) (Javierre et al., 2016; Nanni et al., 2020). Light blue lines (middle panel) represent 10 bootstrapped subsamples obtained based on the real IS data set and black dotted line represents a random IS control data set[2]. Figure was adapted from (Rheinberger et al., 2023).

2.3.2 HIV-1 integration close to TAD boundaries is independent of gene expression level

Apart from enriched CTCF binding and active chromatin marks, TAD boundary regions are also characterized by high transcriptional output and active transcription start sites (Dixon et al., 2012). To assess the transcription levels of integration genes at TAD boundaries, genic integration genes were stratified according to their expression levels into normalized expression categories using C20 and CD4$^+$ T cell RNA-Seq data sets (this study and (Lucic et al., 2019)) (Figure 29A and B, microglia k=9/ CD4$^+$ T cells k=10, 9/10 highest expression class based on logTPM)[1]. The fraction of IS per expression category was then plotted against 10 bins along the TAD boundary-to-midpoint regions of either NeuN$^-$ or consensus CD4$^+$ T cell TAD boundaries, showing that HIV-1 insertions are found enriched at TAD boundaries independent of gene expression, when comparing boundary to midpoint in C20 and T cells (paired t test comparing frequencies of integration at boundary versus midpoint, p = 0.035) (Figure 29C and D)[2].

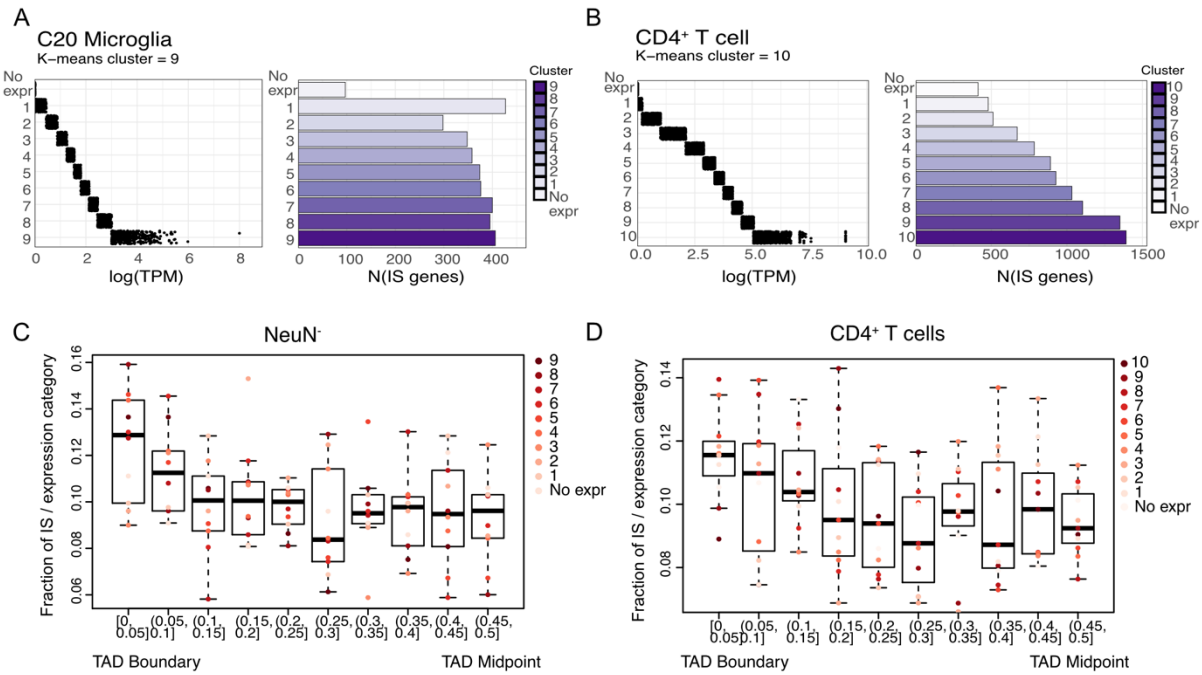


Figure 29: HIV-1 integration close to TAD boundaries is independent of gene expression level. (A) HIV-1 integration genes in C20 cells were scored and clustered based on logTPM using RNA-Seq from this study (k=9). Maximum 5,000 genes per cluster to ensure comparable gene sets and non-expressed genes were assigned to a separate cluster[1]. (B) HIV-1 integration genes in CD4$^+$ T cells were scored and clustered based on logTPM using RNA-Seq (Lucic et al., 2019) (k=10). Maximum 5,000 genes per cluster to ensure comparable gene sets and non-expressed genes were assigned to a

separate cluster[1]. (C) TAD boundary-to-midpoint regions were separated in 10 bins, and frequency of IS per host gene expression class stratified by expression strength was plotted along the TAD regions for NeuN$^-$ (oligodendrocytes, astrocytes, microglia). Paired t test comparing frequencies of integration at boundary versus midpoint ($p < 0.05$)[2]. (D) TAD boundary-to-midpoint regions were separated in 10 bins, and frequency of IS per host gene expression class stratified by expression strength was plotted along the TAD regions for CD4$^+$ T cells. Paired t test comparing frequencies of integration at boundary versus midpoint ($p < 0.05$)[2]. Figure was adapted from (Rheinberger et al., 2023)

2.3.3 HIV-1 ISs TAD border proximity is H3K36me3 dependent

In addition to highly transcribed genes, HIV-1 targets H3K36me3 marked gene bodies directed by the LEDG/p75 – IN interaction. Considering H3K36me3 enrichment at TAD borders, C20 ISs were stratified based on whether they are genic or intergenic, or if they overlapped H3K36me3 ChIP-Seq peaks or not. Interestingly, intergenic ISs (728 of 4,590 IS) and ISs not overlapping H3K36me3 peaks (2,140 of 4,590 ISs) were depleted at NeuN$^-$ derived TAD boundaries and rather enriched towards the TAD midpoint, pointing to a H3K36me3 dependency of IS enrichment towards TAD boundaries (Figure 30A)[2]. TAD Insertion dependency on H3K36me3 could be confirmed in the T cell background by using previously curated IS in Jurkat cells either in LEDGF/p75 knockout conditions (LKO) or expressing an IN binding mutant of LEDGF/p75 (IBD) (Li et al., 2020a). In both conditions IS were depleted at CD4$^+$ T cell TAD boundaries compared to WT Jurkat IS (Figure 30B)[2]. TAD-proximal HIV-1 insertion biases seemed to be CPSF6 independent as ISs sequenced from CPSF6 knockdown (CKD) Jurkat cells or infected with mutant viruses with impaired CA-CPSF6 interaction did not influence TAD-proximal proviral positioning (Figure 30B) (Li et al., 2020a). These data are strengthening the H3K36me3 dependent HIV-1 insertions into TAD boundaries.

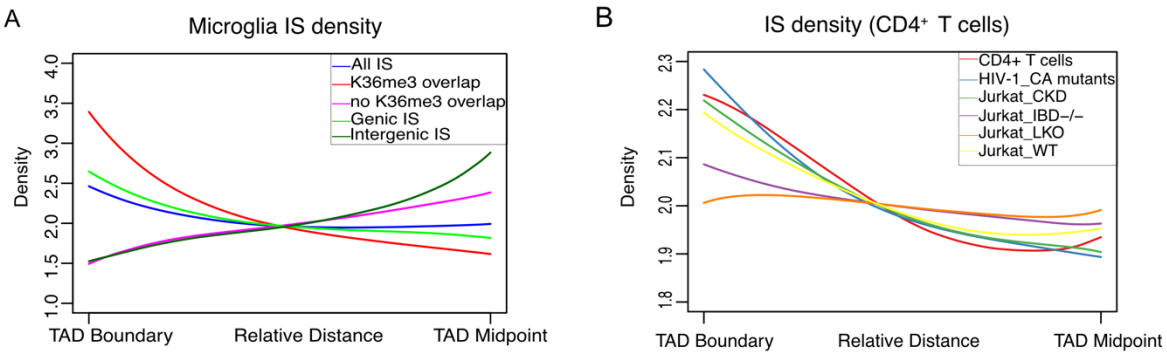


Figure 30: IS TAD border proximity is H3K36me3 dependent. (A) HIV-1 ISs in microglia overlapping (red line, n = 2,486) and non-overlapping (pink line, n = 2,104) H3K36me3 ChIP-Seq peaks, or genic (light green line, n = 3,862) and intergenic (dark green line, n = 728) ISs were plotted as density location over the boundary-to-midpoint regions of NeuN$^-$ (oligodendrocytes, astrocytes, microglia) TAD boundaries (n = 2,077)[2]. (B) CD4$^+$ T cell ISs (red line, n = 13,544), Jurkat wild type (WT) HIV-1 ISs (yellow line, N = 823,169), HIV-1 N74D and A77V CA mutants ISs in Jurkat cells (blue line, n = 726,694), ISs in Jurkat CKD (green line, n = 91,081), ISs in Jurkat IBD/ (violet line, n = 81,346), and ISs in Jurkat LKO (orange line, n = 65,717) were plotted as density location over the boundary-to-midpoint regions of consensus CD4$^+$ T cell TAD boundaries (n=2,171). IS were taken from (Li et al., 2020a)[2]. Figure was adapted from (Rheinberger et al., 2023).

2.3.4 CTCF co-immunoprecipitates with LEDGF/p75 and HIV-1 integrase

HIV-1 PIC is targeted to H3K36me3 enriched regions via LEDGF/p75 binding to viral integrase (Van Maele et al., 2006). Due to the strong dependency of IS TAD border proximity and H3K36me3, I probed for the possibility of an interaction between CTCF and LEDGF/p75 by co-immunoprecipitations. HEK293T cells were transfected with CTCF-eGFP or eGFP alone and CTCF was precipitated with GFP-Trap beads. Immunoblots showed that LEDGF/p75 co-immunoprecipitated with CTCF as well as RAD21, which is known to interact with CTCF (Li et al., 2020b; Nanni et al., 2020) (Figure 31A). Reciprocal interaction was shown by using HEK293T cells expressing LEDGF/p75-HA and immunoprecipitating with HA-antibody (Figure 31D). CTCF binding to LEDG/p75 could also be shown by immunoprecipitation of endogenous CTCF both in C20 and Jurkat T cell line (Figure 31C and D).

As LEDGF/p75 is one of the main interacting partners of viral integrase, the next question arising was whether CTCF also interacts with viral integrase. Co-immunoprecipitation of transfected Flag-IN in HEK293T cells showed interaction of CTCF with viral integrase (Figure 31E). Co-transfection of Flag-IN and siRNA targeting LEDGF/p75 revealed that IN-CTCF interaction is lost upon LEDGF/p75 KD, indicating

that integrase CTCF interaction is LEDGF/p75 dependent (Figure 31F). As expected, silencing of LEDGF/p75 led to a loss of binding to the chromatin mark H3K36me3 (Figure 31F). Interestingly, interaction of CA interactor CPSF6 with CTCF could not be detected (Figure 31A-C), in agreement with the previous observation that IS TAD border enrichment is lost in LKO but not in CKO background (Figure 30B).

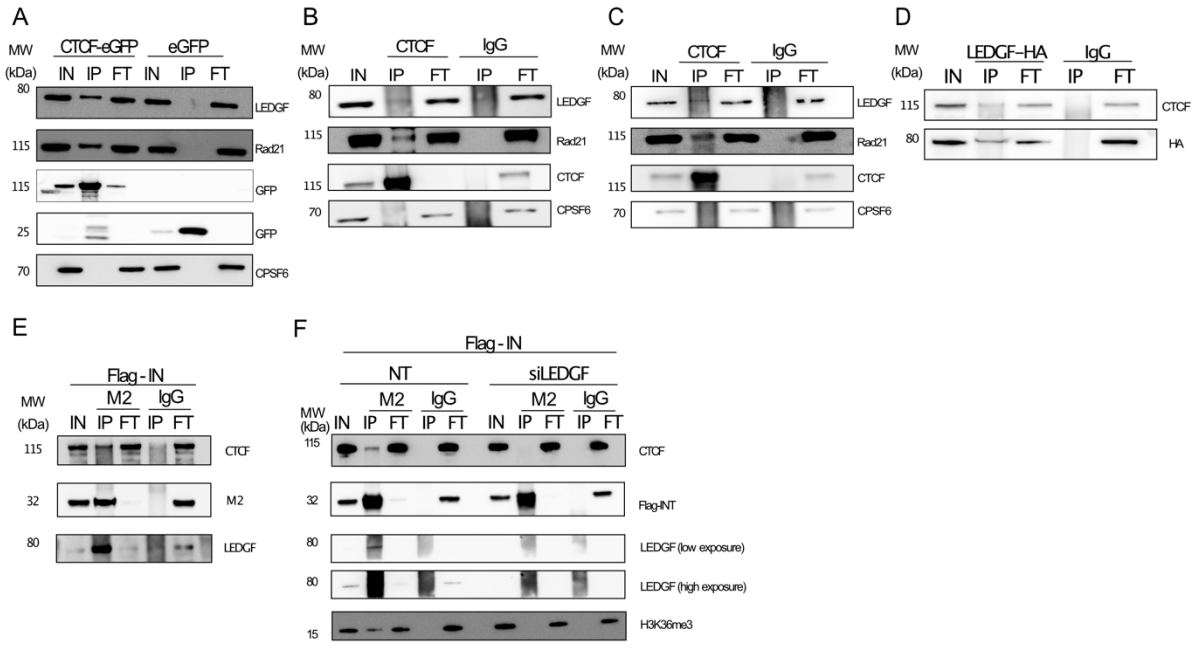


Figure 31: CTCF co-immunoprecipitates with LEDGF/p75 and HIV-1 integrase. (A) HEK293T cells were transfected with either CTCF-eGFP or eGFP. Immunoprecipitations of the nuclear extracts were performed using GFP-TRAP beads. Associated proteins were detected by immunoblotting using the following antibodies: LEDGF/p75, RAD21, CTCF, CPSF6, Anti-GFP. (B) Immunoblot image of co-immunoprecipitation experiment of endogenous CTCF in C20 microglia cell model using CTCF antibody. IgG served as a negative control for the IP. Associated proteins were detected by LEDGF/p75, Rad21 and CPSF6 antibody. (C) Immunoblot image of co-immunoprecipitation experiment of endogenous CTCF in Jurkat T cells using CTCF antibody. IgG served as a negative control for the IP. Associated proteins were detected by LEDGF/p75, Rad21 and CPSF6 antibody. Input (IN), IP (immunoprecipitation), FT (flow through). (D) HEK293T cells were transfected with LEDGF/p75-HA and immunoprecipitation of nuclear extracts was performed using anti-HA antibody (C29F499) or IgG (negative control). Proteins were detected by anti-HA or CTCF antibody. Hemagglutinin (HA). (E) HEK293T cells were transfected with Flag-IN (integrase) and immunoprecipitations of nuclear extracts were performed with M2-antibody (anti-Flag) or IgG antibody (negative control). Proteins were detected using LEDGF/p75, CTCF and M2 antibodies. (F) HEK293T cells were transfected with Flag-IN and either siRNA pool targeting LEDGF/p75 or NT control. After 48h, immunoprecipitations of nuclear extracts were performed using M2-antibody or IgG. Associated proteins were detected by CTCF, M2, LEDGF/p75 or H3K36me3 antibodies. Figure was adapted from (Rheinberger et al., 2023).

2.3.5 CTCF depletion is influencing H3K36me3 chromatin mark at TAD boundaries

CTCF binding together with active epigenetic chromatin marks is enriched at TAD boundaries (Dixon et al., 2012). To test a possible role for CTCF in influencing the chromatin around TAD borders, a Bayesian model network to infer the potential influence of CTCF on the distribution of the repressive marks H3K27me3 and H3K9me3 and the active chromatin marks H3K4me1, H3K27ac, H3K4me3 and H3K36me3 was trained. Therefore, 20 kb genomic windows around the TAD midpoints were used of consensus $CD4^+$ T cell TAD boundaries and published ChIP-Seq data sets of histone marks and CTCF in $CD4^+$ T cells (Lucic et al., 2019; Qi et al., 2021)[2]. 1,800 TAD boundary regions were randomly sampled and bootstrapped to infer direction and strength of the influence indicated by arrow size. It was observed that polycomb repressive mark H3K27me3 was directed towards CTCF, while H3K9me3 heterochromatin histone PTM had an unresolved directionality. The model predicted CTCF directionality towards most transcriptionally active chromatin marks, including H3K36me3, H3K27ac, and H3K4me3, while it had a less defined directionality for H3K4me1 (Figure 32A). In conclusion, the Bayesian network analysis indicated a driving role of CTCF in shaping the chromatin at TAD margins, specifically for active chromatin marks.

To test this hypothesis experimentally, I focused on the main HIV-1 integration determining chromatin mark H3K36me3. I transiently depleted CTCF in C20 cellular model by siRNA transfection (Figure 32B) and fixed the cells and extracted chromatin for CTCF and H3K36me3 ChIP after 48 h (Figure 32C). As positive control regions for ChIP-qPCR, GAPDH intron 7 was chosen for H3K36me3, and hg19 imprinting locus (Kurukuti et al., 2006) served as positive control for CTCF ChIP. Gene desert negative primer set was used as negative control for both. ChIP-qPCR in NT conditions showed a clear enrichment over input signal of H3K36me3 ChIP and CTCF ChIP at their positive control regions compared to isotype control (Figure 32D, all panels, dot). Significant reduction in CTCF signal was observed in CTCF KD conditions compared to NT control at the hg19 imprinting locus (Figure 32D, middle panel). H3K36me3 ChIP showed a reduced signal in CTCF KD conditions at GAPDH intron 7 region, without reaching statistical significance (Figure 32D, left panel). Two replicates of CTCF and H3K36me3 ChIP in NT and CTCF KD conditions were chosen and libraries were prepared, with an average size distribution of around 500 bp (Figure 32E,F).

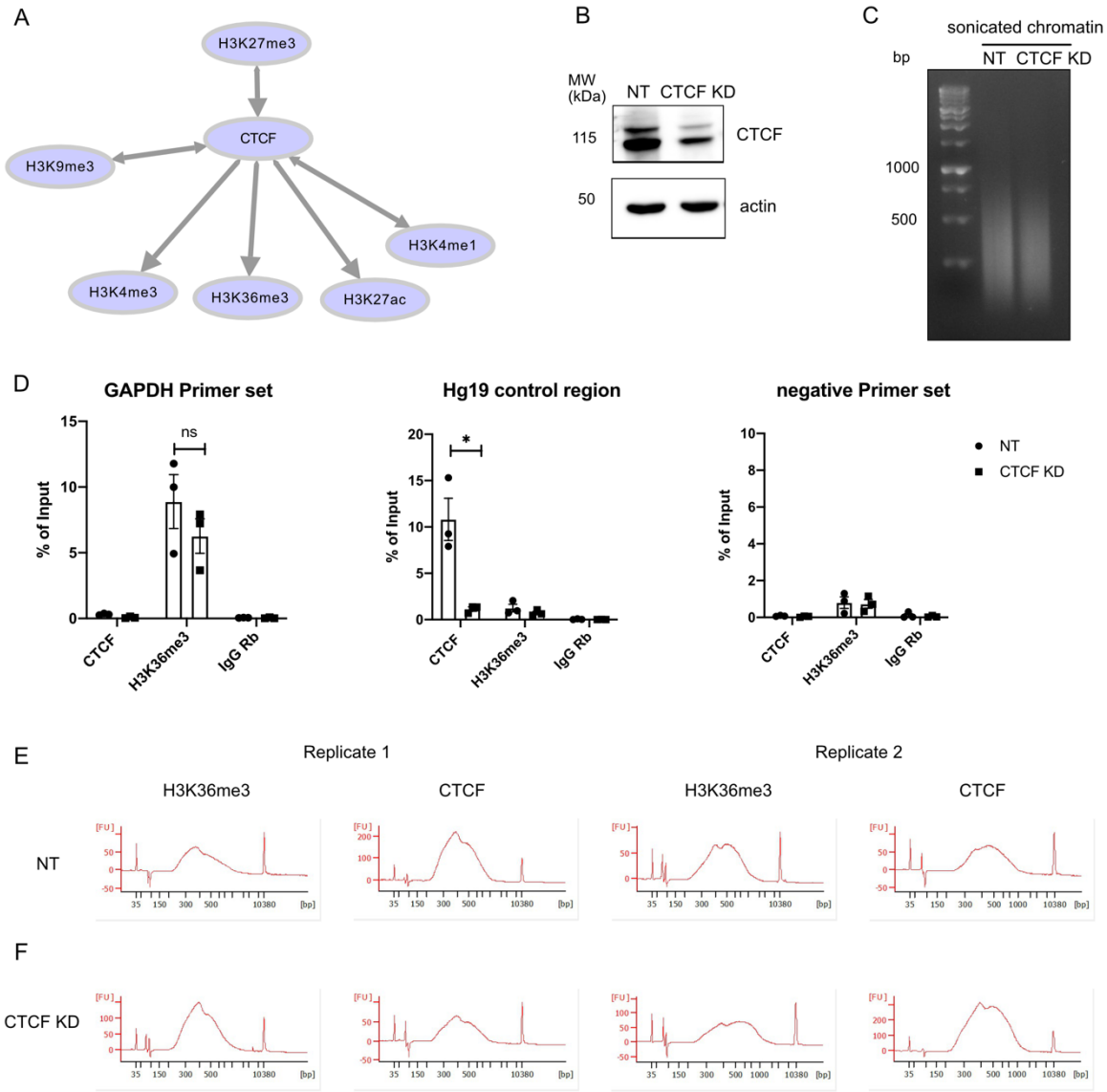


Figure 32: CTCF Bayesian modeling and ChIP-Seq quality control analysis in NT and CTCF KD conditions. (A) Bayesian network based on TAD centered regions in CD4⁺ T cells (20 kb windows, n = 2,171). Directed inference between CTCF and six histone modifications at TAD boundaries was generated by 2,000 bootstraps randomly sampling 1,800 TAD boundary regions each time. Arrow size indicates the direction of the influence. (B) Representative immunoblot image of CTCF KD in C20 cells. Cells were transfected with either NT siRNA pool or siRNA pool targeting CTCF (CTCF KD) for 48h before being crosslinked with 1% formaldehyde for ChIP-Seq. Proteins were detected by immunoblotting using the following antibodies: CTCF, beta actin. (C) Representative 2% agarose gel image of sonicated de-crosslinked chromatin in NT and CTCF KD conditions. 1000 bp DNA ladder served as fragment size standard. Gel was run at constant 90V for 45 min. (D) Bar plots showing percentage of input results after CTCF and H3K36me3 ChIP in C20 cells (NT and CTCF KD) of 3 biological replicates. Rabbit IgG (IgG Rb) served as negative control. GAPDH primer set was used as positive control for H3K36me3, Hg19 imprinting region served as positive control for CTCF and negative primer set was used as negative control. Statistical analysis was performed with GraphPad prism. Data represent mean ± SEM and were analyzed by unpaired t test (*= $p \leq 0.05$, ns= $p > 0.05$). (E) Bioanalyzer

profiles of 2 biological replicates of CTCF and H3K36me ChIP in NT condition. X-axis shows fragment length in bp. Y-axis displays amount of each fragment size in fluorescent units. (F) Bioanalyzer profiles of 2 biological replicates of CTCF and H3K36me ChIP in CTCF KD condition. X-axis shows fragment length in bp. Y-axis displays amount of each fragment size in fluorescent units. Figure was adapted from (Rheinberger et al., 2023).

Differential peak analysis of the CTCF ChIP-Seq between NT and KD conditions showed depletion of 90% of CTCF (FDR ≤ 0.05), both genome-wide (Figure 33A) and at TAD boundary regions (Figure 33B)[1]. Correlation analysis of H3K36me3 ChIP-Seq peaks between NT and CTCF KD revealed moderate differences between the replicates (Spearman correlations coefficients above 0.7) and between the conditions (Spearman's correlation coefficients below 0.5) (Figure 33C)[1]. Analysis of H3K36me3 distribution specifically at TAD boundaries showed significant reduction of this signature at genomic sites where CTCF was depleted and strongly bound in the NT condition (One-sided Wilcoxon test, $p = 4.029e^{13}$, comparing H3K36me3 differences between very low and very high initial CTCF levels in NT condition), suggesting that CTCF is indeed influencing H3K36me3 deposition and/or stability at TAD boundaries and therefore shaping chromatin environment at TAD margins (Figure 33D)[1].

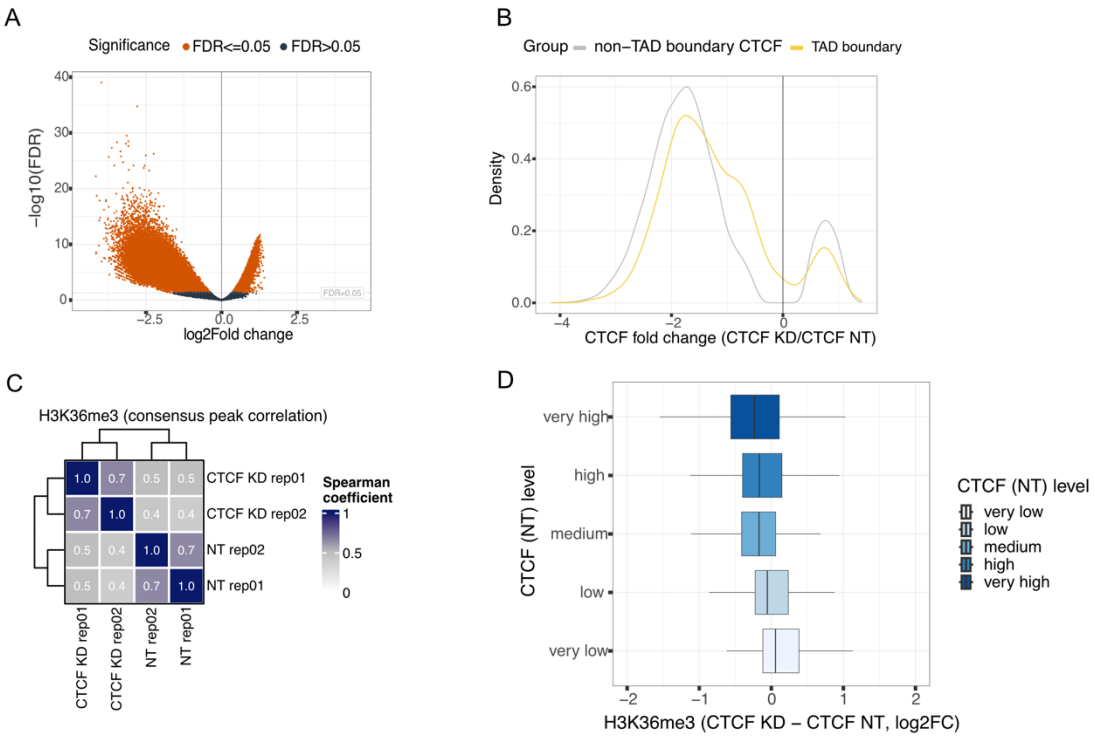


Figure 33: ChIP-Seq analysis of H3K36me3 binding to chromatin in CTCF KD. (A) CTCF ChIP-Seq differential peak analysis between CTCF KD and NT conditions displayed as volcano plot (-log10(FDR) vs log2Fold change). Orange dots indicate significantly differential peaks between NT and CTCF KD (FDR≤0.05, dashed line)[1]. (B) Averaged mean fold change of CTCF binding between CTCF KD and NT condition (CTCF-KD/-NT peaks FDR≤0.05) for each TAD boundary (yellow) and fold change of non-TAD boundary associated genome-wide CTCF (gray) displayed as density plot[1]. (C) Spearman's correlation coefficients for pairwise comparisons of ChIP-Seq peaks in NT H3K36me3 ChIP-Seq and CTCF KD H3K36me3 ChIP-Seq shown as heat map for each replicate[1]. (D) Boxplot comparing H3K36me3 level differences between CTCF KD and NT (log2 fold change (log2FC)) at CTCF peaks with significant loss of CTCF binding upon KD (FDR % 0.05, fold change < 1) sorted by the initial CTCF binding level in NT condition (from very low to very high, in blue shades). One-sided Wilcoxon test, p = $4.029e^{13}$, comparing H3K36me3 differences between very low and very high initial CTCF levels in NT condition[1]. Figure was adapted from (Rheinberger et al., 2023).

2.4 Identification of CTCF and RAD21 as putative host factors influencing HIV-1 integration

2.4.1 CTCF affects HIV-1 integration levels in microglia cellular model and CD4$^+$ T cells

Upon integration, HIV-1 is subjected to the hierarchically organized host chromatin and the underlying principles of genome folding and organization, influencing both HIV-1 integration but also transcription (Einkauf et al., 2019, 2022; Lucic et al., 2019). Given my previous findings that CTCF plays a role in HIV-1 infection (see Figures 26 and 27), interacts with LEDGF/p75 and IN (Figure 31), and IS are enriched at CTCF bound TAD border regions, a subsequent question emerged regarding a potential influence of CTCF on HIV-1 integration. To test this hypothesis, I transiently depleted CTCF in C20 cells by siRNA transfection for 48 h, infected the cells with HIV-1 and subsequently measure integration levels by Alu PCR (Figure 34A). Integration levels were decreased upon CTCF KD to 0.4 over NT (**= $p \leq 0.01$), while total viral levels remained unchanged and viability was not impaired (Figure 34B,C and D). To test if the influence of CTCF on HIV-1 integration was preserved in the main HIV-1 blood cellular target, CD4$^+$ T cells, I transiently depleted CTCF by siRNA in primary CD4$^+$ T cells (Figure 34E and F). Integration levels in T cells exhibited a significant reduction to 0.5 (*= $p \leq 0.05$) in CTCF KD when normalized to total viral levels, suggesting that CTCF influence on HIV-1 integration levels is a common feature of HIV-1 integration (Figure 34G).

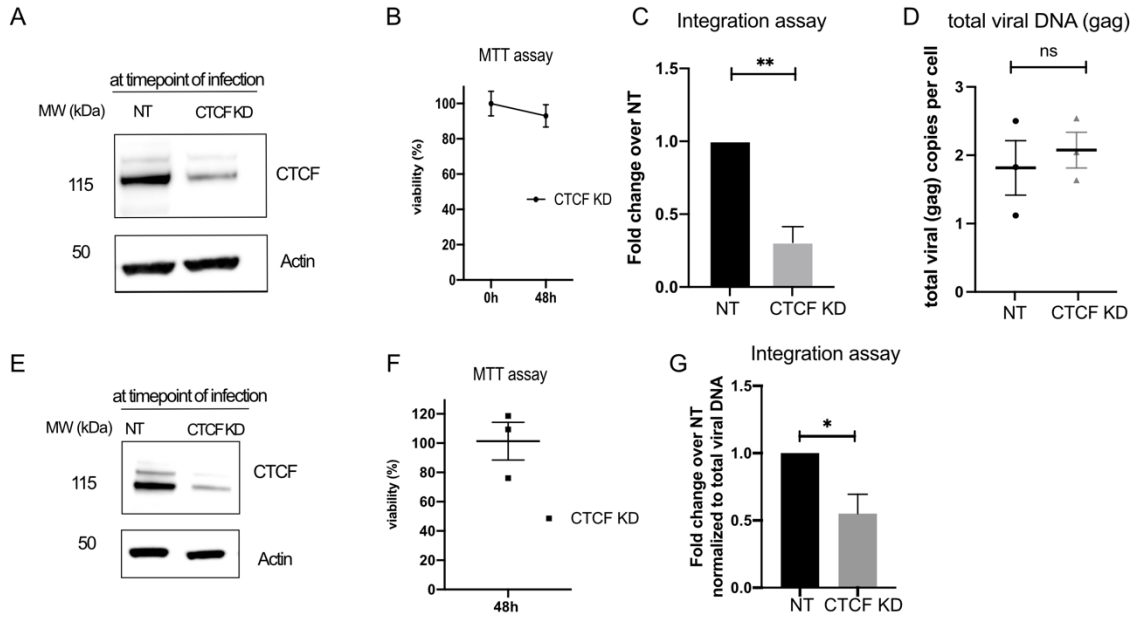


Figure 34: CTCF influences integration efficiency in microglia like cells and CD4⁺ T cells. (A) Representative immunoblot image of CTCF KD in C20 cells. Cells were transfected with either non-targeting (NT) siRNA pool or siRNA pool targeting CTCF (CTCF KD) (50 nM) for 48h before being infected with HIV-1. Proteins were detected by immunoblotting using the following antibodies: CTCF, beta actin. (B) MTT assay of CTCF KD in C20 cells. Cells were transfected with either NT siRNA pool or siRNA pool targeting CTCF for 48h. Viability was calculated as percentage of NT. Data represent mean ± SEM of 3 biological replicates. Graph was generated with GraphPad prism. (C) Boxplot displaying HIV-1 integration levels measured by Alu PCR in microglia cells transfected with the CTCF siRNA pool (CTCF KD) relative to NT siRNA control transfection for 48h and subsequent HIV-1 infection of 3 biological replicates. Data represent mean ± SEM and were analyzed by unpaired t test (**= $p \leq 0.01$) using GraphPad prism. (D) Scatterplot representing total viral DNA copies (gag) per cell in microglia cells transfected with the CTCF siRNA pool (CTCF KD) or NT siRNA control transfection for 48h and subsequent HIV-1 infection of 3 biological replicates. Data are displayed as mean ± SEM and analyzed by unpaired t test (ns, not significant, $p > 0.05$) using GraphPad prism. (E) Representative immunoblot image of CTCF KD in CD4⁺ T cells. Cells were transfected with either NT siRNA pool or siRNA pool targeting CTCF (CTCF KD) (600 nM) for 48h before being infected with HIV-1. Proteins were detected by immunoblotting using the following antibodies: CTCF, beta actin. (F) MTT assay of CTCF KD in CD4⁺ T cells. Cells were transfected with either NT siRNA pool or siRNA pool targeting CTCF for 48h. Viability was calculated as percentage of NT. Data represent mean ± SEM of 3 biological replicates and plot was generated using GraphPad prism. (G) Boxplot displaying HIV-1 integration levels measured by Alu PCR in CD4⁺ T cells transfected with CTCF siRNA pool (CTCF KD) relative to NT siRNA control transfection for 48h and subsequent HIV-1 infection of 3 biological replicates. Integration levels were normalized to total viral DNA (gag) levels measured by qPCR. Data represent mean ± SEM and were analyzed by unpaired t test (*= $p \leq 0.05$) using GraphPad prism. Figure was adapted from (Rheinberger et al., 2023).

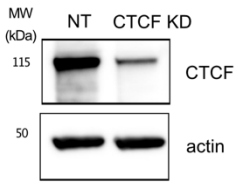
2.4.2 CTCF depletion redistributes HIV-1 insertion sites in microglia cellular model

To further investigate into the potential impact of CTCF on HIV-1 IS distribution, I transiently depleted CTCF in C20 cells by siRNA transfection, as previously described. CTCF KD or control cells were subsequently infected with HIV-1 and the genomic DNA was processed using LM-PCR protocol to sequence genome-wide viral-genome junctions (Figure 35A). Fragments size distribution analysis with Bioanalyzer showed uniform libraries with average sizes ranging from 441 bp to 519 bp (Figure 35B,C). Quality control for proper Illumina adapter addition to the 6 prepared libraries, performed by amplifying DNA fragments with the P5 and P7 adapter primers, estimated full-length library molarities between 2.87 nM to 209 nM.

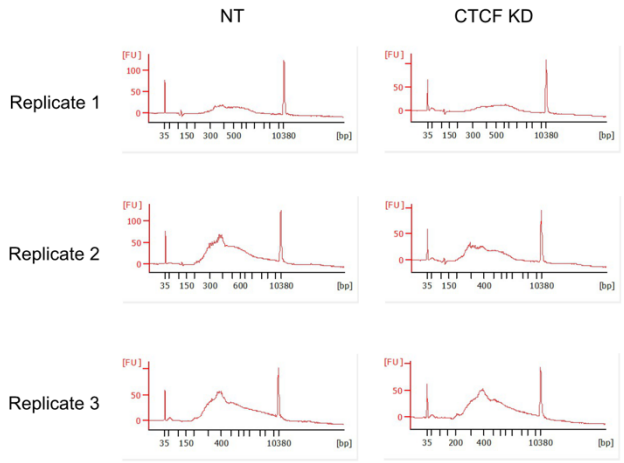
2,814 IS in NT and 2,326 IS in CTCF KD conditions were obtained from 3 biological replicates. Profiling H3K36me3 chromatin mark within ± 10 kb region surrounding the respective IS revealed a decreased signal of H3K36me3 at HIV-1 IS in the CTCF KD compared to NT control (Figure 35D)[1], suggesting that CTCF is contributing to HIV-1 targeting to H3K36me3 enriched chromatin regions.

When mapping the integration sites onto the TAD boundary to TAD midpoint regions of human brain $NeuN^-$ TAD borders, CTCF KD IS were depleted at the boundaries (Figure 35E, brown line) compared to NT IS (yellow line, Figure 35E) and previously sequenced ISs (green, Figure 35E)[1], hinting that CTCF integration site distribution is changing upon CTCF KD.

A



B



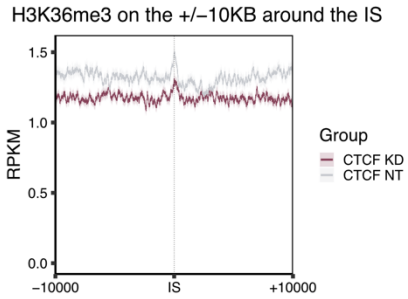
C

Sample	Concentration (ng/µl)	Average fragment size (bp)	Molarity based on qPCR (nM)
NT 1	6.6	489	14.25
CTCF 2	5.5	519	2.87
NT 2	14.5	457	154.4
CTCF 2	8.9	441	49.4
NT 3	22.4	463	209.3
CTCF 3	22	462	204.5

NT
2,814 ISs

CTCF KD
2,326 ISs

D



E

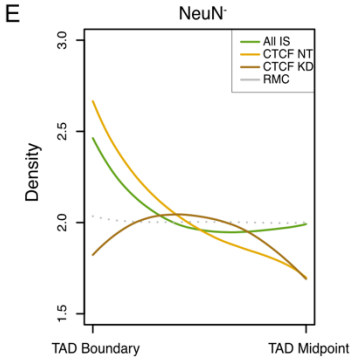


Figure 35: CTCF KD redistributes HIV-1 integration sites. (A) Representative immunoblot image of CTCF KD in C20 cells. Cells were transfected with either NT siRNA pool or siRNA pool targeting CTCF (CTCF KD) for 48h before being infected with HIV-1 and harvested for DNA extraction and ISs sequencing. Proteins were detected by immunoblotting using the following antibodies: CTCF, beta actin. (B) Bioanalyzer profiles of 3 biological replicates of LM-PCR IS libraries in NT and CTCF KD conditions. X-axis shows fragment length in bp. Y-axis displays amount of each fragment size in fluorescent units. (C) Table displaying quality controls of IS libraries in NT and CTCF KD conditions, including concentrations, average fragment size and library molarity. (D) H3K36me3 ChIP-Seq profile plot in RPKM (on CTCF KD and CTCF NT) at vicinity (±10 kb) of ISs obtained in CTCF KD (red, n=2,326) and CTCF NT (gray, n=2,814). Confidence intervals (95%) are indicated as light-colored regions[1]. (E) Microglia IS (green), NT IS (yellow) or CTCF KD IS (brown) plotted as density locations over the boundary-to-midpoint regions of NeuN$^-$ (Oligodendrocytes, Astrocytes, Microglia) TAD boundaries (n=2,077)[1]. Figure was adapted from (Rheinberger et al., 2023).

2.4.3 RAD21 influences HIV-1 integration levels

My findings so far support the role of structural protein CTCF in modulating the levels of HIV-1 integration and in redistributing HIV-1 insertions. Although we observed a depletion of CTCF KD integration sites at the TAD boundaries compared to NT control sites, whether CTCF's role in 3D genome organization and TAD boundary formation contributes to this phenotype still needs to be investigated.

The mechanism how CTCF is contributing to TAD border and loop formation is by loop extrusion. The cohesin complex is loaded onto the DNA by NIPL. DNA is extruded until the complex encounters two convergently bound CTCF proteins. The complex consists of SMC3, SMC1, SA1/SA1 and Rad21 proteins (Fudenberg et al., 2016; Wutz et al., 2017). In order to start investigating whether RAD21, as part of the cohesin complex being involved in TAD border establishment, would also influence HIV-1 integration levels, I transiently depleted RAD21 in C20 cells by siRNA transfection (Figure 36A, B). After 48 h cells were infected with HIV-1 and 24h post infection, integration levels were assessed by Alu PCR. HIV-1 integration was diminished by RAD21 KD, while total viral levels stayed unchanged (Figure 36C and D). Quantification of 2LTR circle number as an indicator of insufficient integration showed a decreased trend when comparing RAD21 with NT condition, without reaching statistical significance.

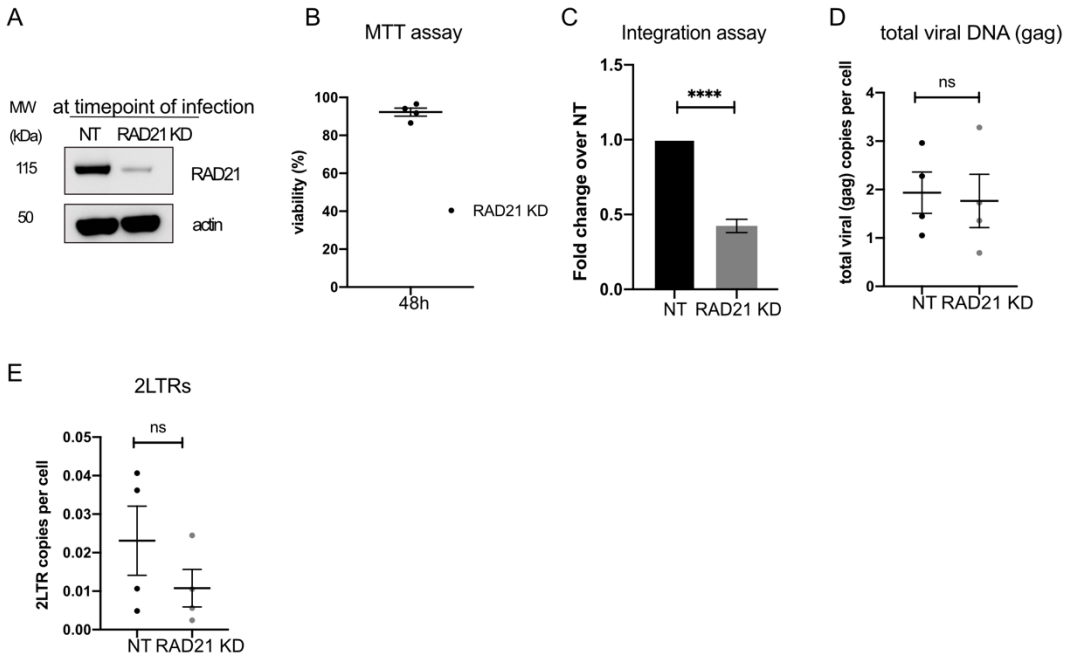


Figure 36: RAD21 influences HIV-1 integration efficiency. (A) Representative immunoblot image of Rad21 KD in C20 cells. Cells were transfected with either NT siRNA pool or siRNA pool targeting Rad21 (Rad21 KD) (10 nM) for 48h before being infected with HIV-1. Proteins were detected by immunoblotting using the following antibodies: Rad21, beta actin. (B) MTT assay of Rad21 KD in C20 cells. Cells were

transfected with either NT siRNA pool or siRNA pool targeting Rad21 for 48h. Viability was calculated as percentage of NT. Data represent mean ± SEM of 4 biological replicates using GraphPad prism. **(C)** Boxplot displaying HIV-1 integration levels measured by Alu PCR in HIV-1 infected microglia cells transfected with the Rad21 siRNA pool (Rad21 KD) relative to NT siRNA control transfection for 48h of 4 biological replicates. Data represent mean ± SEM and were analyzed by unpaired t test (****= $p \leq 0.0001$) using GraphPad prism. (D) Scatterplot representing total viral DNA copies (gag) per cell in HIV-1 infected microglia cells transfected with the Rad21 siRNA pool (Rad21 KD) or NT siRNA control transfection of 4 biological replicates. Data are displayed as mean ± SEM and analyzed by unpaired t test (ns, not significant, $p > 0.05$) using GraphPad prism. (E) Scatterplot representing 2LTR copies per cell in HIV-1 infected microglia cells transfected with the Rad21 siRNA pool (Rad21 KD) or NT siRNA control transfection of 4 biological replicates. Data are displayed as mean ± SEM and analyzed by unpaired t test (ns, not significant, $p > 0.05$) using GraphPad prism.

3 Discussion

HIV-1 mainly targets cells of the hematopoietic lineage, but it can also infect cells of the CNS, mainly microglia cells. The persistence of latent HIV-1 reservoirs throughout the human body is still the main impediment for an HIV-1 cure. Upon treatment interruption in patients previously receiving ART for viral suppression, viral replication can resume from latently infected cells, leading to high viral loads within weeks. The main latent reservoir, and the most characterized one, are memory $CD4^+$ T cells of a resting phenotype. Over the past three decades, the primary focus of research has been directed towards either reactivating or 'locking' the latent reservoir in HIV-1 infected patients. Additionally, there have been approaches utilizing CRISPR/Cas9 technology to either remove the provirus or target viral receptors. Up to now the main difficulty is that current therapies do not achieve the full reactivation of the whole latent reservoir pool in the absence of general T cell activation, as well as effective killing of the reactivated cells, due to the reservoir heterogeneity. In the recent years, it became increasingly evident that also other cell types and tissues might contribute to fueling the latent reservoir, including brain microglia cells (Tang et al., 2023). The development of HIV-associated neurological disorders in PLWH emphasized even more the importance of studying this so far unexplored latent reservoir of HIV-1. Therefore, this thesis focused on obtaining the first HIV-1 insertion data in microglia cellular models, investigation of the chromatin integration sites landscape and the possible contribution of genomic and transcriptional determinants to IS selection and early latency events.

3.1 Characterization of microglia integration sites landscape

3.1.1 Establishment of an integration site sequencing method

The first step in characterizing HIV-1 insertions in the microglia cellular model was the establishment of a method to sequence ISs. I used the linker mediated PCR method, which is a well-known sensitive and efficient protocol for sequencing retroviral integration sites both in in vitro infections and patient material (Gillet et al., 2017; Maldarelli et al., 2014; Plaza-Jennings et al., 2022; Schroder et al., 2002; Sunshine et al., 2016; Wagner et al., 2014; Wu et al., 2003). The method employs a double stranded asymmetric linker that is annealed to the sonicated DNA fragments to prevent

linker-linker ligations, and which serves as an annealing point for the subsequent PCR steps. After the first ISs sequencing of HIV-1 infected C20, bioinformatic analysis revealed that the genomic fragments after trimming the LTR and linker sequences were too short for precise mapping to the human genome, and many duplicated linker sequences were formed. Therefore, the LM-PCR protocol was optimized at several steps. Sonication time was shortened to increase fragment length for improved mapping in the analysis. A more stringent clean up by AMpure beads was introduced after linker ligation to remove unligated linker sequences, which can, if still present, serve as primers and lead to unwanted PCR artifacts. Furthermore, primer concentration was reduced, cycle number of the second nested PCR and the sample input was lowered as well as the annealing temperature increased to prevent PCR recombination and mispriming. PCR recombination is an event that happens during incomplete elongation step and a partial DNA fragment acts as a primer, annealing to a DNA fragment elsewhere in the genome (Wells et al., 2020). In PCR reactions with mixed templates, as it is the case with different ISs in the pool of host genome fragments, homologous DNA strands amplified from different chromosomes can cross-hybridize and anneal via the LTR and linker regions to form large PCR bubbles (Kalle et al., 2014). By lowering of the input DNA material per PCR reaction and the PCR cycle number both artifacts could be reduced in the IS libraries.

A way of ensuring correct IS assignment is the paired end sequencing. Both genomic fragments have to align on the same chromosome to come from a true IS. Another way of avoiding false positive ISs is the implementation of 3' LTR and 5' LTR primers, to get genomic fragments from both ends of the viral insertion site. The difficulty here is, that the 5' LTR of the here used pNL4.3 virus strain is highly mutated, which makes it hard for primer annealing (Wells et al., 2020). In the first IS sequencing of my thesis, 5' LTR primers were included but due to the very low IS yield, for the following runs only 3' LTR primers were used, as in most of the viral-genome junction mapping methods (Sunshine et al., 2016; Wagner et al., 2014).

3.1.2 Integration site landscape in microglia follows the determinants observed in blood reservoirs

Comparison of C20 ISs with the main HIV-1 cellular reservoirs, $CD4^+$ T cells and macrophages, showed that HIV-1 integration has certain genomic and chromatin determinants that are shared among the different cellular and tissue reservoirs. The

analysis support gene bodies of actively transcribed genes as predominant target of HIV-1 integration, as well as the dependency of integration on LEDGF/p75 and CPSF6 host factors, mediating the target biases towards active genomic regions (Bedwell et al., 2021). Nonetheless, limited numbers of samples that were sequenced in C20 cells, as well as the limited number of available ISs in MDMs (only 1000 sites), made the analysis possibly less robust (Kok et al., 2016). Precise definition of the extent and molecular basis of differences in the targeted genes between HIV-1 target reservoirs needs further investigations, especially because previous work showed that also not all HIV-1 target genes in $CD4^+$ T cells have been discovered yet (Lucic et al., 2019).

Within gene bodies, HIV-1 insertions are found mostly in intronic regions (Schroder et al., 2002; Singh et al., 2015; Sowd et al., 2016). Additionally, in T cells, HIV-1 was found to reside close to SPADs, which are regions close to nuclear speckles, where splicing factors are stored (Bedwell et al., 2021; Chen et al., 2018b; Francis et al., 2020; Li et al., 2020a). Whether and how splicing processes are impacting HIV-1 integration biases is still unclear and needs further investigations. Moreover, SPAD proximity has so far only been shown for T cells and macrophages and still needs to be explored if it is also the case in microglia cells.

The relationship between cis and trans regulatory elements, transcription, and regions associated with speckles is evident, yet additional research is needed to understand the individual contributions of each these features to the IS.

Genome-wide integration of genomic analysis showed that association of HIV-1 IS with gene body H3K36me3 mark and enhancers is a common feature of HIV-1 integration sites both in microglia cellular model and T cells. Interestingly, in the brain, genomic variants of disease risk genes were found to be enriched in microglia-enhancers, most prominently for Alzheimer disease (Nott et al., 2019). The observation that in C20 cells HIV-1 insertions are close to enhancers, opens the possibility that also HIV-1 insertions close to enhancers could disrupt important microglia genes leading to the development of Alzheimer or other neurological diseases. It remains for future studies to gather more integration sites and enhancer landscapes in primary brain tissues for a more comprehensive analysis, including patient material.

To help identifying common retroviral integration determinants, the principle of defining recurrently targeted genes informed a lot in recent years about HIV-1 insertion biases. One way of defining RIGs was initially based on the cross comparison of different IS studies and the occurrence of a gene in 2 or more of these studies (Lucic et al., 2019;

Marini et al., 2015). A recent large scale IS study from Bedwell and colleagues defined RIGs and recurrent avoided genes (RAG) by setting a statistical threshold for insertion across 10 in vitro (also including non-target cells of HIV-1) and patient studies. Genes that were experimentally targeted significantly more than expected by chance were considered a RIG, genes that were targeted less than expected by chance were considered as a RAG (Bedwell et al., 2021). Both ways of RIG definitions come to the same conclusion that HIV-1 targets certain genes, irrespective of the target cell type, highlighting that other characteristics of these genes as chromatin marks, nuclear position or regulatory elements are defining integration.

Gene ontology analysis of RAGs derived from non-neuronal cells revealed neuronal gene ontology terms to be enriched. It is noteworthy that comparison of the here identified microglia target genes with the RAGs gene list revealed a tendency for microglia target genes to align more closely with RAGs (one sixth of RAGs were targeted in microglia). While this increased targeting of RAGs in microglia may be attributed to the specific integration pattern of these cells, it's important to note that, as of now, this analysis is based on the only available dataset for microglia. Further investigations with integration sites in primary microglia are needed to confirm these findings.

While microglia cellular models like C20 and iPSC-derived microglia used in this study permit studying several concepts of HIV-1 integration process including large-scale genomic studies and IS sequencing, they do not fully recapitulate microglia properties in vivo (Alvarez-Carbonell et al., 2017; Garcia-Mesa et al., 2017; Gosselin et al., 2017). It is known that microglia cells highly depend on the environmental signals, as transferring primary microglia form brain surgery into in vitro culture is changing transcriptional profiles (Gosselin et al., 2017). Future experiments involving microglia inflammation and activation studies would certainly benefit from valuable primary microglia samples.

In the course of my PhD thesis, a study was published where HIV-1 ISs from physiological microglia were obtained from HIV-infected patients with and without encephalitis (HIVE) (Plaza-Jennings et al., 2022). The authors sequenced around 1200 ISs from sorted $NeuN^-$ cells and observed similar overall IS distribution between T cells and $NeuN^-$ cells. Chromosome 19 showed the highest density of HIV-1 ISs, integration genes were highly expressed and integration into A subcompartement was favored, as observed in T cells and C20. Moreover, comparison of active enhancer

marks (H3K4me3, H3K27ac), transcriptionally repressive modifications (H3K9me3, H3K27me3) and chromatin accessibility (ATAC-Seq) clearly showed that HIV-1 integrates close to active and away from repressive chromatin to the same extent in T cells and microglia containing primary cells (~10^4 bp and 10^6-10^8 bp from the nearest active and repressive marks, respectively) (Plaza-Jennings et al., 2022).

Remarkably, the authors of the Plaza-Jennings study generated 81 IS libraries using LM-PCR method from 34 unique donors (7 from HIVE and 27 from HIV^+ patients) obtaining only around 1200 ISs, predominantly from HIVE samples. This is emphasizing how technically challenging IS sequencing in the brain is, as HIV-1 integration is a rare event, and the sample quality of frozen post-mortem tissues is limited. In addition to in vitro infections of C20 and iPSC-derived microglia that highlighted that the main epigenetic signatures from microglia containing tissues and T cells are common, primary sample ISs are valuable to investigate reactivation potential, intactness of the virus and clonal expansion of ISs, as this cannot be studied in vitro. Plaza-Jennings and co-authors did observe less recurrent integration in $NeuN^-$ cells compared to T cells, which is not surprising considering the low IS number and the fact that it is the only primary microglia data set so far. Of note, they observed similar clonal expansion percentages as in T cells (Plaza-Jennings et al., 2022). Since viral load rebounds after treatment interruption, it is of great importance to get further insights into the capacity of the provirus for reactivation and expansion in microglia. It would be valuable to obtain information about the proviral sequence of ISs as intactness of the virus is impacting the capacity of viral reactivation and viral rebound. Intact proviruses in blood cells have been found to be enriched in ZNF genes in the heterochromatic B2 and B4 compartment of the genome under long-term ART (Huang et al., 2021; Lian et al., 2023). It remains for future studies, to implement single cell methods in the microglia infection research to further investigate its contribution to HAND development and viral rebound after ART interruption. Applying scATAC-Seq together with RNA-Seq could identify transcriptionally active proviruses together with the IS (Ma et al., 2020; Wang et al., 2020).

New integration requirements and differences could be identified in future studies using HIV-1 patient-derived postmortem tissue, considering the role of immune signaling and environmental signaling cues in shaping epigenetic and transcriptional landscape of brain tissue in health and disease.

3.2 Early latency events in HIV-1 infection

The establishment of latent reservoirs, cell subsets or anatomical sites harboring long-lasting replication competent virus compared to the actively replicating viral pool, is still the main bottleneck for an HIV-1 cure. Latency is established early on in infection, also in the CNS (Chavez et al., 2015; Churchill et al., 2015), making it important to assess the establishment of early latency events, where preexisting chromatin states that HIV-1 encounters could influence the outcome of proviral transcription.

3.2.1 Transcription factor networks in active and latent C20 cells

Unexpectedly, as observed by ATAC-Seq, the global chromatin accessibility did not show changes between the uninfected, latent, and productive infection states. Chromatin accessibility is highly dynamic and is known to change during cell differentiation and mitosis (Martin et al., 2023; Yu et al., 2023). It cannot be excluded that as ATAC-Seq was performed in bulk, and single cells are heterogenous, possible overall differences in accessibility were evened out over the populations. In future studies, scATAC-Seq approaches could help to clarify if the lack of global changes in accessibility were due to the chosen timepoint or technical limitations (Wang et al., 2020).

In contrast, several transcription factors whose binding sites were distinctively accessible with respect to the viral presence and/or gene expression could be identified in microglia. In particular, we found several TFs that have well-established roles in regulating HIV-1 transcription, neuronal tissue development and function and host cell nuclear organization, including NF-κB, Members of the Myocyte enhancer factor 2 (MEF2) family and CTCF (Gosselin et al., 2017; Van Lint et al., 2013). In line with a main role of NF-κB family in HIV-1 LTR transcription and reactivation from latency, an increased footprint in productive infection compared to uninfected and latent cells was observed. NF-κB is an inducible transcription factor which is sequestered in the cytosol during latency, and translocation into the nucleus reactivates viral transcription by binding to the LTR (Van Lint et al., 2013). A TF motif found to be occupied more in uninfected and latent cells compared to active is the TF family FOS. The FOS TF family consists of c-FOS and FOSb, which dimerize with c-Jun to form AP-1 TF complex (Milde-Langosch, 2005). AP-1 has well-known roles in

activating HIV-1 transcription, as it has a binding site in the HIV-1 promoter (Duverger et al., 2013). The formation of dimers cannot be distinguished in the footprinting tool TOBIAS, which was used for the analysis, as it is based on TF motifs that are fed to the algorithm. It does not provide information about cooperative binding of TFs, dimerization or can distinguish between different TF family members (Bentsen et al., 2020). Therefore, the predicted increased FOS occupancy in the latent and uninfected cells represents an interesting target for further investigation.

Another TF class which was predicted to be more occupied in the latent cells compared to uninfected and active cells were SMADs. SMADs are phosphorylated following TGFβ signaling and translocate to the nucleus where they regulate anti-inflammatory response genes (Hata and Chen, 2016). TGFβ/SMAD signaling has recognized roles in neuroprotection in Alzheimer's disease and could therefore also well be involved in the development of HAND during HIV-1 infection in the brain (Von Bernhardi et al., 2015; Tesseur et al., 2006; Ueberham et al., 2012).

Neuronal and microglia homeostasis regulators, the MEF2 TF family, were found to be less bound in microglia latently infected cells. Interestingly, MEF2 TFs were linked to heritability and disease progression of Alzheimer's disease (AD) (Tansey et al., 2018; Young et al., 2021), as well as microglia-mediated immune responses and loss of cognitive ability in aged mice (Deczkowska et al., 2017; Yang et al., 2015). MEF2A and D were also found to be key regulators of the aging transcriptional signature in the brain, where their suppression was involved in neuronal cell death, thus opening a possibility that MEF2 TFs might contribute to HIV-1 latency in microglia and/or HAND establishment (González-Velasco et al., 2020). This interesting candidate merits further investigations in the future, especially validations by ChIP-Seq for the different paralogs A, B, C and D, which cannot be distinguished by TF footprinting.

In conclusion, assessment of chromatin accessibility did not reveal gross changes, but rather pointed to plausible differential TF networks between active, latent and uninfected HIV-1 infected C20 cells, which would need further investigations by ChIP-Seq binding assay.

3.2.2 CTCF is released from the genome in productive infection

CTCF was revealed as a dynamic factor in HIV-1 infection. CTCF is a transcription factor that is involved in transcriptional activation, repression and 3D chromatin organization via DNA looping and insulation (Cuddapah et al., 2009; Hansen et al.,

2017; Hou et al., 2008; Luan et al., 2022; Nuebler et al., 2018; Oomen et al., 2019; Zuin et al., 2014).

Because HIV-1 proviruses do not harbor CTCF binding sites, it is tempting to speculate that CTCF contributes to viral gene expression via its pivotal role in TAD formation and looping (Dixon et al., 2012; Nora et al., 2012). If the unbound CTCF in active viral transcription has an influence on 3D chromatin organization or if it is TAD structure independent still needs to be investigated. Ongoing viral replication could perturb genome organization and/or CTCF and other architectural proteins, as previously found in HIV-1 infection in the brain in patients that developed HIVE. The Plaza-Jennings study comparing Hi-C in microglia of HIV$^-$ patients and HIVE patients found a switch of interferon genes from heterochromatic regions (B) into the euchromatic A compartment leading to higher gene expression in HIVE, and a significant switch from A to B compartment of homeostatic genes, being lower expressed. Furthermore, the HIVE patients showed genome wide weaker TAD boundaries. The authors concluded that HIV infection in the brain induces open chromatin and compartment switching which is accompanied by increase in inflammatory gene expression (Plaza-Jennings et al., 2022). Another example of compartment switch comes from influenza infection, where the viral protein NS1 causes a transcriptional readthrough of POL-II thereby resulting in compartment switching from B to A (Heinz et al., 2018).

The question that remains unresolved is if CTCF release from the genome is occurring in the regions of the genome proximal to the provirus, and/or whether it is genome-wide in response to the viral replication. Either way, CTCF, alone or as a part of larger complex could contribute to transcriptional networks reprogramming and 3D genome remodeling, either locally or genome wide. Of note, one indication that CTCF is released due to viral replication and not due to sensing the viral particle is the finding that Raltegravir treatment, which blocks viral integration, restores CTCF on the genome comparable to mock levels.

The data gathered thus far solely present evidence for quantitative changes in CTCF on chromatin in HIV-1 replicating cells (fractionation) or predicted TF binding (ATAC-Seq). It remains for future studies to precisely map CTCF genome binding by ChIP-Seq or Cleavage Under Targets and Tagmentation (Cut&Tag) (suitable for low cell numbers) during productive and latent infection in different HIV-1 target cells to obtain information about CTCF differentially occupied positions in the genome.

Apart from having global genome-wide effects, CTCF's role in HIV-1 latency could be through different mechanisms. It has been shown that CTCF is highly dynamic and in constant exchange with cofactors and the genome (Agarwal et al., 2017; Hansen et al., 2017, 2018; Oomen et al., 2019). A small subset of CTCF TFBS can rapidly respond to environmental signals in development and differentiation (Arzate-Mejía et al., 2018; Hansen et al., 2018; Wang et al., 2019). In pluripotent stem cells differentiating into neurons a pool of stress sensitive CTCF was found in nuclear speckles, and was downregulated by proteasomal degradation upon oxidative stress (Lehman et al., 2021). This interesting study points to a dynamic regulation of CTCF upon stress by rapid changes in protein stability. An earlier study proposed that oxidative stress and hypoxia induces de-SUMOylation of a pool of CTCF, rendering it available for gene expression regulation (Kitchen and Schoenherr, 2010; Wang et al., 2012). Similarly, HIV-1 infection induces oxidative stress, which could lead to some influence on CTCF protein stability on DNA to induce specific gene expression profiles (Shytaj et al., 2020). Along the lines of posttranslational protein modifications, Investigations in human cell lines revealed that phosphorylation of CTCF zinc finger domains upon starvation stress by sensitive large tumor suppressor (LATS) kinases impairs CTCF DNA binding. This led to locus-specific release of a subset of CTCF TFBS, disrupting local chromatin domains and consequently decreasing gene expression of genes within that domain (Luo et al., 2020). It could well be that also during active HIV-1 infection, the host cell stress response is regulated by a subset of CTCF sites that might be induced by posttranslational modifications.

Apart from influencing the host genome, CTCF could also have an active role in HIV-1 latency. A recent study demonstrated by ATAC-Seq that CTCF also plays a role in HIV-1 latency in T cells, established by reversal to latency of actively HIV-1 replicating T cells after 12 weeks (Jefferys et al., 2021). Knockout of CTCF in actively replicating T cells reduced the level of latency establishment in cells depleted for CTCF over time, highlighting a necessity of CTCF for latency establishment in T cells. In my PhD thesis the focus was on understanding the contribution of early latency events, rather than reversal to latency as in activated $CD4^+$ T cells after longer culturing time period (Jefferys et al., 2021), thus adding an additional layer to the role of architectural protein CTCF early in the HIV-1 replication cycle. It could be that CTCF rebinding to the host genome during reversal to latency is important to re-establish silencing of the virus.

A role for CTCF in the viral life cycle was proposed for several DNA viruses by binding of CTCF to their genomes. For KSHV, CTCF binds to several regions on the viral genome, repressing lytic replication (Chen et al., 2012a). In HPV 18 infection, CTCF was proposed to serve as a host cell restriction factor, as mutation of the CTCF binding site led to an increase in E6 and E7 viral transcripts (Paris et al., 2015). CTCF binding between OriP and Cp promoter of EBV genome negatively affects transcription of the latency protein Ebna2, which defines latency types of EBV virus infection (Chau et al., 2006). Altogether, different to HIV-1, for these viruses it is known that via direct binding of CTCF, latency is maintained.

Genome binding profiling of CTCF throughout HIV-1 infection will confirm experimentally at which genomic positions CTCF is dynamically binding during HIV-1 infection, and further clarifying if CTCF has a direct influence on latency or response to the virus is affecting CTCF.

3.3 3D genome topology in HIV-1 infection

During viral infections, viruses are subjected to the hierarchically organized host genome, which they can exploit for their viral life cycle. Conversely, the host 3D chromatin organization can be influenced by the viral infection. For HIV-1, the impact of 3D host chromatin on infection is largely underexplored but has been in focus in the recent years. In my PhD thesis TAD boundaries were determined as a new feature of HIV-1 integration, and possible outcomes for viral replication and latency establishment are discussed below.

3.3.1 HIV-1 integrates close to TAD boundaries

HIV-1 ISs, both in T cells and C20 microglia cellular model, were found enriched close to TAD boundaries, which are regions insulating two neighboring TADs. TAD boundaries are characterized by a high transcriptional output, active chromatin marks, CTCF and cohesin binding, housekeeping genes and replication origins (Dixon et al., 2012; Emerson et al., 2022; Hnisz et al., 2013; McArthur and Capra, 2021; Nanni et al., 2020; Nora et al., 2012; Rao et al., 2014). The analysis indicated that enrichment of ISs at border regions is H3K36me3 dependent, but independent of the high transcriptional level, indicating that the chromatin and structural determinants at the boundaries are important for integration. A recent study cross comparing 37 Hi-C data

sets could show that TAD boundaries are highly evolutionary conserved among different cell types and that these stable TAD boundaries are further enriched in CTCF binding and immunologic heritability traits (McArthur and Capra, 2021). Moreover, clustered CTCF TFBS at TAD boundaries were found to promote long-term residence of cohesin and loop extrusion blocking leading to strong TAD boundary insulation (Chang et al., 2023). These factors could contribute to HIV-1 insertion biases at TAD borders by providing a stable component of 3D genome among different HIV-1 targets for correct integration. Cis-regulatory elements like promoters and enhancers are unevenly distributed throughout the genome and can be bound by master transcription factors regulating highly transcribed cell identity genes and forming clusters at TAD boundaries (Madani Tonekaboni et al., 2021). Because HIV-1 insertions were found to cluster close to enhancers it is tempting to speculate that these attributes could contribute to the integration at TAD boundaries. A recent study in peripheral blood samples of HIV-1 patients also revealed decreased distance of genic ISs to frequently interacting regions (FIREs), supporting the here presented insertion biases towards TAD boundaries (Einkauf et al., 2022).

The proviral positioning close to TAD boundaries could influence the outcome of HIV-1 infection through facilitating 3D long-range contacts. Viral transcription could be enhanced or silenced, leading to productive viral infection or latency. The gene expression of HIV-1 is dependent and regulated by TF binding to the 5’ LTR, the HIV-1 tat protein and the chromatin state of the provirus. Furthermore, the activation of the promoter occurs as a result of the creation of a gene loop that connects the 5' LTR with the 3' LTR poly(A) signal (Van Lint et al., 2013; Perkins et al., 2008). It could therefore be envisaged that positioning of the provirus close to the insulating regions between interacting genomic regions in the complex 3D genome of the host cell could promote additional looping mechanisms between the viral LTR promoter and the host genome. For example, an adoption of a distal enhancer by the HIV-1 LTR could promote active viral transcription or reactivation from latency. This mechanism, termed ‘enhancer hijacking’ was previously shown to be involved in oncogenic transformation of cancer cells (Drier et al., 2016; Gröschel et al., 2014; Haller et al., 2019). Interestingly, Einkauf and colleagues found a higher 3D contact frequency of active HIV-1 provirus in patients, as well as more active epigenetic signals, a higher accessibility and increased transcription in the 3D contact regions, suggesting that 3D genomic cis- and trans regulatory elements could influence HIV-1 transcription (Einkauf et al., 2022). Another

recent study investigating chromatin accessibility, mRNA expression, protein levels, genome wide 3D genomic contacts as well as HIV-1 LTR 3D contacts in HIV-1 infected Jurkat clones observed increased chromatin accessibility at host and proviral genome in HIV-1 actively transcribing Jurkat clones. At the genomic level, HIV-1 integration and transcription did not induce alterations in chromatin conformation, however, 3D contacts of the HIV-1 genome were detected up to 300 kb away from the provirus on the same chromosome as the IS. Moreover, in the accessible chromatin in the 3D contact regions of the LTR the authors found enriched motifs of several zinc finger TFs, including CTCF, concluding that virus-host interactions might be mediated by cis regulatory elements and TFs in a host environment dependent manner to regulate HIV-1 accessibility (Collora and Ho, 2023).

In contrary to switching on transcription, HIV-1 provirus could also be silenced by being embedded adjacent to TAD boundaries, for example by heterochromatin spreading from neighboring TADs upon interruption of CTCF insulator binding or through the high transcriptional output at boundary regions, having an insulation effect (Dixon et al., 2012, 2016; McArthur and Capra, 2021). Latently infected HIV-1 cells are heterogenous and the IS plus the surrounding host genome influence the unique reactivation potential of each of them (Chen et al., 2017; Einkauf et al., 2022). Even within latent isogenic clones that contain the same IS, reaction potential differs (Collora and Ho, 2023). Indeed, in HIV-1 latency, cell fate decision is stochastically driven by noise. Viral transcription is regulated by stochastic fluctuations of tat protein, and if initial transcript burst reaches a critical threshold, viral fate changes to exponential viral growth (Pai and Weinberger, 2017; Rouzine et al., 2015; Weinberger et al., 2005). Proviral positioning close to TAD boundaries together with the cis regulatory elements like enhancers or CTCF binding could lead to an influence of tat transcriptional circuits by increasing transcriptional activation events. Thereby, the IS surrounding genome at any given time within a cell is potentially influencing proviral fate. It remains for future studies to investigate the role of TAD boundary targeting of HIV-1 for viral infection outcome and reactivation potential.

For several DNA viruses it was shown that they use the 3D host genome to influence their transcription. HBV forms a cccDNA that was found to interact with host DNA regions being rich in active enhancers and promoter signals (H3K4me3, H3K9ac, H3K4me1, H3K27ac) (Yang et al., 2020), while inactive viral DNA was found close to chromosome 19 at heterochromatic regions 12, 27, 44, 52 and 58 Mb (Tang et al.,

2021). More specifically, cccDNA was found at 19p13.11 locus in hepatoma cells, where it uses an enhancer element to switch on viral transcription (Shen et al., 2020). Hi-C in cells infected with EBV showed 15,000 contacts between EBV and the human genome, mostly contacts enriched in H3K27ac and H3K4me1 (Wang et al., 2020). When the virus reactivates from latency, the interaction sites were found to change from heterochromatin to euchromatin (Moquin et al., 2018).

Conversely to using 3D chromatin for the viral life cycle, viral infections can also influence host 3D genome organization and host gene transcription. The DNA virus HPV can disrupt the host genome by dividing one TAD into two, and thereby locating the PEGβ enhancer towards the CCDC10 gene, increasing its expression (Cao et al., 2020). Interaction between HPV integrant and the host chromosomes leads to long range interactions of more than 500 bp, inducing host gene dysregulation through host/host interaction disruption in TADs (Groves et al., 2021). The retrovirus HTLV-1 disrupts the host chromatin structure by loop formation between virus and host genome through insertion of an ectopic CTCF site introducing long-range interactions (Melamed et al., 2018). A recent study on a non-integrating virus SARS-CoV-2 showed a direct influence of viral infection on host genome. They discovered A compartment weakening upon infection as well as A and B compartment mixing. Furthermore, intra TAD contacts were reduced, H3K27ac signal decreased and cohesin was depleted from intra TAD regions. These genomic perturbations correlated with transcriptional suppression of interferon response genes, while increasing H3K4me3 signal in promoters of proinflammatory genes, with possible implications for long covid development (Wang et al., 2023).

In their recent study, Plaza-Jennings and colleagues investigated genome-wide 3D contacts of brain microglia nuclei of uninfected and HIV^+ patients with encephalitis brains. They revealed that in microglia cells of chronically HIV-1 infected patients with inflammation in the brain, parts of the genome shifted from B to A compartment. As the provirus does not harbor CTCF binding sites and is thereby not inserting ectopic CTCF sites into the host genome with the possibility of changing 3D contacts, the authors went on and found in particular type 1 interferon response genes to be switched from B to A compartment (Plaza-Jennings et al., 2022). These 3D changes driven by immune activation could render these regions susceptible for new viral integrations (Plaza-Jennings et al., 2022). Genome wide chromatin contacts measured by Hi-C in

4 HIV-1 infected Jurkat cell clones did not show global changes and only mild increase in H3K27ac contacts between infected and uninfected clones, but not at ISs (Collora and Ho, 2023). The authors conclude that this could on one side be due to the high degree of heterogeneity between single cells and the hypotetraploidy of the Jurkat cell line diluting out possible effects of the IS. On the other side, it could be that HIV-1 provirus is subjected to the existing host architecture without changing 3D chromatin upon integration. It is known that within the transcription unit, actively transcribing provirus is changing chromatin accessibility of host chromatin at IS, that could lead to aberrant host gene expression (Collora and Ho, 2023; Liu et al., 2020).

Collectively, these studies suggest that viral infections can indeed impact 3D host chromatin organization, influencing host gene expression, in particular of viral response genes. Whether this is the case for HIV-1 infection, remains to be further investigated. One of the primary unresolved questions in the field of 3D genome organization is the connection between genome structure and its functionality, also during viral infections. At present, available data suggest that rather than serving as a dominant determinant of gene function, TADs and loops establish a foundational structural framework, where structure could have influence on transcription and transcription can also affect looping structure (Cattoni et al., 2017; Finn and Misteli, 2019; Haarhuis et al., 2017; Lupiáñez et al., 2015; Mirny et al., 2019; Nagano et al., 2013; Narendra et al., 2015; Nora et al., 2017; Nuebler et al., 2018; Rao et al., 2017; Wutz et al., 2017). Cell populations are highly heterogenous, especially in HIV-1 infection and potential functional roles of genome topology are tightly linked to the developmental timing, the specific organism, and genomic context. With the advancement of more sophisticated single-cell techniques (like single cell Hi-C (Ramani et al., 2017)), 3D genome organization studies in HIV-1 infection could give more insights into the structure to function relationship for viral infections (Bintu et al., 2018; Boettiger et al., 2016; Cattoni et al., 2017; Finn et al., 2019; Mateo et al., 2019; Nagano et al., 2013; Stevens et al., 2017; Wang et al., 2016b).

3.3.2 CTCF interacts with LEDGF/p75 and viral integrase

By immunoprecipitating CTCF in HEK293T cells or endogenously in C20 and Jurkat cells I could show that CTCF partners with the host factor LEDGF/p75, as well as with viral IN in a LEDGF/p75 dependent manner. Interaction of CTCF protein with known HIV-1 integration co-factors has not been reported so far and interaction can be direct

or indirect, mediated through a large protein complex, where additional partners would still have to be identified. A large proteome-wide quantitative screen in HeLa cells using sucrose gradient centrifugation identified CTCF interaction with LEDGF/p75 (Caudron-Herger et al., 2019), concluding that most chromatin CTCF interactions are RNA dependent.

LEDGF/p75 is a known HIV-1 viral IN interactor and is recognizing via its PWWP domain the host chromatin mark H3K36me3, H3K36me2 and an additional site at the interface between nucleosomal DNA and the histone H2A-H2B acidic patch (Acke et al., 2022; Eidahl et al., 2013; Pradeepa et al., 2012; Vansant et al., 2020). Although the protein interaction was detected in the absence of HIV-1 infection, CTCF binding to LEDGF/p75 could contribute to HIV-1 integration process. This is supported by the finding of impairment of integration efficiency observed upon CTCF loss in C20 and T cells. Moreover, the observed HIV-1 IS enrichment at TAD boundaries, which, together with the protein binding of CTCF to LEDGF/p75 and the H3K36me3 and CTCF enrichment at TAD margins further support a role for CTCF in HIV-1 targeting. It is still unclear when during viral replication cycle and where in the nucleus the interaction of CTCF, LEDGF/p75 and viral IN is happening. Whether this interplay is taking place in the context of PIC remains to be further investigated. Microscopy based techniques like proximity ligation assay could aid in further clarification. Published ISs generated in LEDGF/p75 KO conditions in Jurkat cells shifted away from TAD boundaries, indicating that the H3K36me3 signal as well as the lost CTCF binding to viral IN and LEDGF/p75 could be driving factors in HIV-1 targeting towards TAD boundaries. It remains elusive if TAD boundaries or H3K36me3 signal are changing upon LEDGF/p75 KO, which so far was not investigated.

CPSF6 interaction with CTCF was not detected in any of the cellular backgrounds tested, in line with the notion that it contributes to integration site selection through the association with viral CA rather than with viral IN (Sowd et al., 2016). CPSF6 binds to viral CA and directs the PIC towards highly transcribed regions close to nuclear speckles. Analysis of IS distribution in Jurkat and HEK293T cells with LEDGF/p75 KO showed HIV-1 insertions shifted away from gene bodies and closer to the TSS, as well as less recurrent IS targeting, while sites sequenced in CPSF6 depletion or from infection with viral mutants impaired for the CA/CPSF6 interaction revealed insertions outside of gene dense regions in the nuclear periphery close to LADs (Achuthan et al., 2018; Bedwell et al., 2021; Burdick et al., 2020; Chin et al., 2015; Francis et al., 2020;

Li et al., 2020a; Singh et al., 2022). In the analysis performed by Carl Herrmann and Ana Luisa Costa, TAD border proximal targeting in T cells seemed not to be affected by CPSF6 impairments, in contrast to ISs in LKO conditions that were shifted towards TAD midpoints. Considering the finding that CPSF6 KO ISs were found in the nuclear periphery close to LADs, it is tempting to speculate that PIC is hold back at the nuclear pore regions because CPSF6 is missing to navigate towards SPADs, but might be targeted towards LAD border regions, which are also rich in CTCF binding (Chen et al., 2018a; Guelen et al., 2008; Kaczmarczyk et al., 2022). LADs and inter-LAD regions are separated by CTCF binding to prevent heterochromatin spreading, and previous studies showed that LAD regions can be determined by Hi-C methods (Barski et al., 2007; Fraser et al., 2015; Kind et al., 2015). Although majority of HIV-1 insertions were found outside of LADs (Marini et al., 2015), a possible scenario in CPSF6 depleted conditions could be that stalled PICs are released at the nuclear periphery and the CTCF-LEDGF/p75-IN protein complex directs HIV-1 into LAD border regions, which are also enriched in active chromatin marks (Barski et al., 2007; Guelen et al., 2008). To test this hypothesis, genome-wide contact frequencies complemented by Lamin Dam-ID in CPSF6 KO, in CTCF KD or double KD conditions together with ISs mapping would be important to clarify if insertions in the absence of CPSF6 end up in LAD border regions.

3.3.3 CTCF KD influences H3K36me3 deposition

Identifying the genome-wide patterns of histone post-translational modifications and CTCF and their consequential functional implications for HIV-1 integration and persistence would significantly contribute to our understanding of how 3D genome organization influences HIV-1 integration biases. Therefore, a Bayesian network approach was applied that not only infers co-occurrence but also the causal relationships of chromatin modifications and a driving role for CTCF in influencing the chromatin around it was observed. CTCF KD in C20 cells followed by H3K36me3 ChIP-Seq indicated indeed an influence on H3K36me3 deposition at CTCF sites that are highly bound in the NT conditions. One hypothesis of how this could be achieved is through heterochromatin spreading into active chromatin regions upon removal of CTCF's barrier function. It is known that CTCF insulates active from repressive chromatin regions, which are dynamically regulated by chromatin modifying enzymes (Chung et al., 1993; Cuddapah et al., 2009; Dowen et al., 2014; Groner et al., 2012;

Lensch et al., 2022; Li et al., 2002; Probst et al., 2009). To experimentally address this hypothesis, further investigation by performing Hi-C in CTCF depleted conditions to obtain information about changes in TAD border formation and looping would be needed. CTCF depletion in ChIP my experiments was performed transiently by siRNA transfection, resulting in depletion of 90% of CTCF from TFBS upon KD. Previous functional studies on CTCF showed differences between transient depletion, protein-based depletion, or KO. Hi-C in partially depleted HEK293T cells showed only mild insulation loss (Zuin et al., 2014). Inducible degron system in mESCs revealed differences between complete and near complete loss of CTCF (Nora et al., 2017). Furthermore, transient depletion showed resistance of approximately 20% of CTCF binding sites, being constitutively bound (Khoury et al., 2020). Persistent sites were shown to depend on position and cohesin binding and are mostly at enhancers and involved in genome architecture (Luan et al., 2021). This emphasized the importance of complete depletion of CTCF for functional studies, as the remaining CTCF can still take over functions.

Furthermore, H3K27me3 ChIP-Seq would help to clarify if the positions previously occupied by H3K36me3 are now modified with H3K27me3. It is also known that H3K36me3 prevents the binding of PRC2 complex and is therefore prohibiting the deposition of polycomb repressive chromatin mark H3K27me3 (Schmitges et al., 2011; Yuan et al., 2011).

It cannot be excluded that the influence of CTCF on H3K36me3 deposition is through other mechanisms, independent of the structural role of CTCF. For example, CTCF could bind or indirectly influence the function of SETD2, the only methyltransferase depositing H3K36me3 (Edmunds et al., 2008), and thereby affect H3K36me3 deposition. SETD2 binds to CTD of POL-II and catalyzes co-transcriptional H3K36me3 modification (Kizer et al., 2005; Xiao et al., 2003). In support of this scenario is the finding that promoters of genes downregulated upon CTCF depletion in mouse embryonic stem cells (mESCs) were shown to contain CTCF TFBS (Nora et al., 2017). RNA-Seq of CTCF KD in C20 could aid to clarify if the genes were H3K36me3 signal was reduced are also transcriptionally downregulated. It is likely that CTCF depletion disrupts additional processes in the cell other than 3D genome organization (Luan et al., 2022). Moreover, CTCF and H3K36me3 were found to be involved in alternative splicing regulation (Luco et al., 2010; Shukla et al., 2011). At CD45 DNA locus it has been shown that CTCF binds to unmethylated exon DNA, pausing POL-II and thereby

leading to exon recognition and incorporation into pre-mRNA (Shukla et al., 2011). Heterogenous ribonucleoprotein L (hnRNPL) is a splicing factor binding to skipped exons during alternative splicing and is binding to SETD2 as well, mediating a crosstalk between splicing and transcription machinery (Bhattacharya et al., 2021). It could well be that CTCF influences the crosstalk between co-transcriptional epigenetic modifications and the splicing machinery which needs further investigations.

3.4 Host factors influencing integration

HIV-1 infection is dependent on many host factors to complete its replication cycle. Extensive research has been done so far to identify host factors required for efficient HIV-1 replication, including functional high throughput screenings by siRNA and CRISPR/Cas9 (Brass et al., 2008; Hiatt et al., 2022; König et al., 2008; Park et al., 2017), as well as assessment of physical interactions of viral protein interactors (Jäger et al., 2011). Among the common reported factors important for HIV-1 are well studied host cell proteins including LEDGF/p75, CPSF6, TNPO3, NUP153 or CypA. All of them come into place at different steps in the viral replication cycle mostly at a level of nuclear import and HIV-1 integration. During my thesis, I identified two additional host factors playing a role in HIV-1 integration: the architectural proteins CTCF and RAD21. How they mechanistically influence HIV-1 integration levels still needs to be further investigated.

3.4.1 CTCF

CTCF is a protein with many essential roles in the complex environment of the human cell. It is a transcription factor that can through its looping function result in activating or restricting gene expression (Fudenberg et al., 2016; Hou et al., 2008; Nora et al., 2017). Furthermore, it is an architectural protein involved in 3D genome organization by influencing TAD border formation (Wutz et al., 2017).

CTCF depletion during HIV-1 infection by siRNA transfection reduced HIV-1 integration levels significantly, both in C20 cells and T cells, suggesting that CTCF is an important host factor engaged in HIV-1 integration. One of the possible scenarios by which CTCF could influence HIV-1 integration is through its binding to LEDGF/p75, however the mechanism through which CTCF contributes to insertional biases is still unknown. Comparison of ISs in the absence of LEDGF/p75 to the ones in CTCF depletion would

provide additional insights if IS distribution is shifted in the same way. Moreover, protein domain deletion experiments could clarify through which domains of LEDGF/p75 and CTCF the interaction is mediated. In the absence of CTCF, HIV-1 could not be efficiently targeted to TAD boundaries and be integrated there. This is supported by the observation that ISs in the absence of CTCF are redirected away from NeuN$^-$ TAD boundaries and to regions with lower H3K36me3 signal. To take into account that upon CTCF KD, 3D structure and H3K36me3 deposition could change (discussed in the previous chapter) (Nora et al., 2017; Zuin et al., 2014), Hi-C in CTCF depleted conditions would be necessary to investigate probable 3D genome organization changes and the possibility that IS are shifted due to the disturbed 3D structure.

Certainly, it cannot be excluded that the effect of CTCF depletion on integration is independent of CTCF's functional role in 3D genome architecture, as CTCF has pleiotropic effects and can affect several processes in the cell. One possibility could be that CTCF is influencing indirectly integration efficiency via its function as a transcription factor and/or recruiter of other TFs on chromatin including tissue-specific activators, repressors, or POL-II, impacting the expression of important HIV-1 integration co-factors. Moreover, the RNA binding capacity of CTCF was found to be essential for CTCF's functions, suggesting the possibility that RNA targeting could also be important for CTCF's role in HIV-1 integration efficiency (Caudron-Herger et al., 2019; Saldaña-Meyer et al., 2014, 2019).

3.4.2 RAD21

RAD21 is part of the cohesin complex being involved in loop extrusion and 3D chromatin organization (Davidson et al., 2019; Fudenberg et al., 2016; Hansen et al., 2018; Kim et al., 2019; Krietenstein et al., 2020; Nuebler et al., 2018; Wutz et al., 2017). Due to the finding that CTCF KD reduced integration efficiency and the IS enrichment close to TAD boundaries, the question arose if also other factors involved in the 3D genome organization could have a role in the integration process. A genome wide siRNA screen identified RAD21 as a putative factor involved in RT kinetics or viral DNA stability (König et al., 2008). siRNA KD of RAD21 and subsequent HIV-1 infection in my thesis showed a reduction in integration. These results support the notion that 3D genome organization and TAD border structure might be important for the virus to successfully integrate. Nonetheless, also for RAD21 differences between KD and full protein depletion on 3D genome organization need to be considered. While CTCF

reduction leads to insulation loss at TAD boundaries, RAD21 KD reduced loop domains (Nora et al., 2017; Rao et al., 2017; Schwarzer et al., 2017; Wutz et al., 2017; Zuin et al., 2014). Hi-C in RAD21 depleted conditions in C20 as well as IS sequencing would aid to investigate if changes in 3D genome organization influences HIV-1 integration and if ISs would be shifted to similar positions as in CTCF KD. To further investigate if 3D genome is the determining factor, depletion of other factors involved in the loop extrusion process and cohesin complex as SMC1, SMC3, PDS5, SCC3, WAPL or NIPBL/MAU could be probed in integration assay and IS sequencing (Gligoris et al., 2014; Grubert et al., 2020; Li et al., 2020b). HIV-1 infection experiments using a delta N-terminus mutant of CTCF (being responsible for binding RAD21) as well as double KD of RAD21 and CTCF could help to investigate if reduction in integration efficiency is due to the cohesin – CTCF interaction and their role in loop anchoring (Li et al., 2020b; Pugacheva et al., 2020). Hsieh and colleagues recently suggested a 'time buffering model' for enhancer-promoter loops mediated by RAD21. A loop can sustain short term depletion (several hours) of structural proteins, possibly through epigenetic signals and TF binding, while the establishment of new loops is impaired (Hsieh et al., 2022). In my experiments, HIV-1 infection was performed at 48 h of RAD21 KD, which, following this model, would mean that loop structure might indeed be impaired, influencing integration efficiency. The beforementioned experiments will aid to further investigate the role of cohesin – CTCF loop anchoring in HIV-1 integration.

Eukaryotes have three different SMC complexes all involved in different biological processes in the cell. The cohesin complex contains SMC1 and SMC3 and plays a role in sister chromatid cohesion, gene expression regulation and DNA repair (Nishiyama, 2019). Chromatid cohesion is unlikely to contribute to HIV-1 integration as HIV-1 infects interphase nuclei. SMC2 and 4 are part of the condensin complex regulating interphase chromatin compaction into mitotic chromosomes, while SMC5/6 play a role in DNA replication and repair, but it is less characterized as the other SMC members (Palecek, 2019; Paul et al., 2019). Two recent studies investigated a role for SMC5/6 in latency establishment by silencing expression from unintegrated lentiviral DNA through compaction and SUMOylation (Dupont et al., 2021; Irwan et al., 2022). Having this in mind it is tempting to speculate that also other SMC and associated proteins could be involved in other steps of the viral replication cycle (like integration).

It cannot be excluded that RAD21 implications in HIV-1 integration are independent of its structural role together with CTCF. RAD21 is not only part of the cohesin complex as a structural protein but is also involved in replication stress and double strand break repair through non-homologous end joining (NHEJ) pathway. Apart from reducing integration efficiency, I could also observe a trend in reduction of 2LTR circles, suggesting that RAD21 not only has influence on HIV-1 integration but also on the formation and/or stability of 2LTR circles. 2LTR circle formation has been shown to be dependent on host factors involved in NHEJ like Ku80, XRCC4, ligase 4, Mre11 or NBS1 (Jeanson et al., 2002; Kilzer et al., 2003; Li et al., 2001), implying that also RAD21 could be a host factor involved in 2LTR circle formation.

4 Conclusion and Perspectives

HIV-1 infects target cells of the hematopoietic lineage, mainly $CD4^+$ T cells and macrophages, but also other tissue types e.g., the CNS, early during HIV-1 infection. The establishment of latent reservoirs through the stable integration into the host genome is the main barrier for an HIV-1 cure, which makes it important to study the nature of these latent cells for the development of new therapeutic strategies. Cellular determinants of HIV-1 integration and latency have been mostly studied in T cells, but to obtain potential cell type specific differences in the IS profiles and latency programs between tissue types it is of great importance to explore other target types.
In my PhD study, I started to investigate IS profiles together with chromatin and transcriptional profiles using microglia cellular models, concluding that IS determinants are shared between T cells and microglia cells.
Nonetheless, the conclusions obtained in the cellular model would benefit from further studies in patient-derived primary microglia from post-mortem tissue samples. Using additional approaches as for example matched IS and proviral sequencing (MIP-Seq) or long read sequencing to obtain additional information about the intactness of the HIV-1 provirus could help to understand the fraction of replication competent virus in the brain (Einkauf et al., 2019). Central questions that remain to be addressed are how the latent clones in the brain contribute to the viral rebound and inflammation and whether and how this is influenced by the integration site.

HAND development in PLWH is a severe complication of HIV-1 infection in the brain but the detailed mechanism of how it is developing is still not fully clear. Although C20 cells are a good starting point to investigate HIV-1 infection and latency in microglia cells, more advanced models are required, especially to study immune signaling and inflammation pathways. The advancements in differentiation protocols of pluripotent stem cells and organoid developments in the recent years helped to establish systems that fully recapitulate brain microenvironments to study HIV-1 infection in in vitro conditions (Dos Reis et al., 2020, 2023).

Chromatin accessibility in active and latently infected cells identified differentially occupied TF between the conditions highlighting different transcriptional networks. One of the factors prominently involved was CTCF, with a dynamic binding behavior

throughout infection. The findings are based on sequencing and biochemical methods. Further insights into the mechanism of CTCF being involved in HIV-1 infection could be gained through high resolution imaging of CTCF in infection to see the spatiotemporal resolution of the interaction with PIC. The central questions that remain open are if CTCF is influencing viral replication and/or if it is involved in the response to the virus. I started establishing Cut&Tag method in the lab, with the aim to profile CTCF binding in active, latent and uninfected sorted cells, under conditions of limited cell numbers. With that approach one could investigate qualitative changes in CTCF binding throughout infection to analyze genomic regions affected by infection. In addition, scCut&Tag and obtaining IS from active and latent single cells by scATAC-Seq could inform about the role of CTCF in HIV-1 infection and transcription (Bartosovic et al., 2021; Janssens et al., 2022; Wang et al., 2020). Nonetheless, CTCF as a potential influencer of HIV-1 transcription, is not targetable therapeutically due to its pleiotropic effects in 3D genome organization, looping, gene transcription and splicing (Fudenberg et al., 2016; Hou et al., 2008; Nora et al., 2017; Saldaña-Meyer et al., 2019; Shukla et al., 2011; Wutz et al., 2017).

The thesis also suggested an influence of CTCF in H3K36me3 chromatin mark deposition around TAD boundaries. This observation needs further investigations by additional profiling of other chromatin marks as H3K27ac or H3K4me3 to investigate if CTCF has a general role in shaping the chromatin landscape around TAD boundaries. The mode of action of this influence remains unclear and further imaging and IP experiments including for example epigenetic remodeling enzymes such as SETD2 could give insights in the mechanism of CTCF on epigenetic marks.

3D genome is important in HIV-1 infection and gained more appreciation in the recent years in HIV-1 research. Many viruses exploit or influence 3D genome and nuclear architecture on multiple scales. HIV-1 viral positioning close to TAD boundaries could influence the outcome of viral infection. Also recent studies identified increased 3D contacts of the provirus supporting the notion that 3D environment is an important factor in HIV-1 infection (Collora and Ho, 2023; Einkauf et al., 2022). To specifically address if TAD structure rearrangements are influencing integration and latency, chromosome conformation capture-based approaches are necessary on infected cells and would be the most informative on a single cell level. Hi-C or 4C in HIV-1 infected single clones could give insights if the virus does change 3D host genome, on the

global or local level, respectively. By generating cell clones via targeting viral insertions through CRISPR/Cas9 at defined positions with different proximities from TAD boundaries, the influence on viral transcription could be inferred. Under physiological conditions HIV-1 integration is a single event per cell, thus single cell data are necessary to further identify the possible influence of host genome architecture on viral positioning and transcription.

Up to now there is still no HIV-1 cure due to the establishment of life-long HIV-1 latent reservoirs in different cell and tissue types throughout the human body. Microglia cells are the main HIV-1 target cells in the brain, difficult to target with ART and a possible contributor to the development of neurological disorders in HIV-1 infected patients. Therefore, it is important to get further insights into the IS landscape and the latency establishment mechanisms of this cell type for the development of new HIV-1 therapies. Through the integration of the IS profiles, together with transcription, epigenetics, and chromatin accessibility I defined the first IS landscape in microglia cellular models and revealed that the main hallmarks of HIV-1 integration were corroborated in the brain background. Future studies in patient samples and iPSC-derived cellular systems are needed to deepen our understanding of the contribution of latent provirus in the brain for viral rebound, clonal expansion, and brain inflammation. The latent reservoir in patients is characterized by complex transcriptional and phenotypic signatures, with different susceptibility to immune responses (Chomont et al., 2009; Dufour et al., 2023; Gantner et al., 2020). In the blood, less than 10% of the latent clones harbor intact proviral sequences, which are vulnerable to host immune activity (Gaebler et al., 2019; Hiener et al., 2017; Ho et al., 2013; Lee et al., 2017). Whether this is the case for latent proviruses in the brain, to which extent they clonally expand and how immune cascades act to control viral replication in the brain still requires additional high throughput single cell IS analysis studies. My findings in the second part of my thesis also point to the importance of further exploring the 3D genome organization and its functional implications for HIV-1 infection in both blood and brain reservoirs, which could contribute to the future therapy development.

5 Material and Methods

5.1 Materials

Antibodies

Antibody	Company	Catalog number
Anti-Histone H3 (tri methyl K36) antibody - ChIP Grade	Abcam	ab9050
Anti-Histone H3 (acetyl K27) antibody - ChIP Grade	Abcam	ab4729
Tri-Methyl-Histone H3 (Lys27) (C36B11) Rabbit mAb	CellSignaling	#9733
Anti-Histone H3 (mono methyl K4) antibody - ChIP Grade	Abcam	ab8895
Anti-Histone H3 (tri methyl K9) antibody - ChIP Grade	Abcam	ab8898
Anti-Histone H3 (di methyl K9) antibody [mAbcam 1220] - ChIP Grade	Abcam	ab1220
CTCF (D31H2) XP® Rabbit mAb	CellSignaling	#3418
Mouse IgG2a, Kappa Monoclonal [MOPC-173] - Isotype Control - ChIP Grade	Abcam	ab18413
Rabbit Control IgG-ChIP grade	Abcam	ab46540
anti-beta actin	Sigma	a2228
LEDGF/p75 Antibody	Bethyl labs	A300-848A
Anti-CPSF6 antibody	Abcam	ab99347
Anti-HIV-1 p24-PE (KC57)	Beckman Coulter	6604667
Anti- RAD21	Upstate, Sigma	05-908
Anti-GFP antibody	Abcam	ab6556
Anti-HA-tag antibody (C29F4)	CellSignaling	#3724
Monoclonal anti-Flag M2 antibody produced in mouse	Sigma	F3165-1MG

Plasmids

Name	Source	Identifier
pNL4.3	NIH Aids Reagent Program	ARP-114
HIVGKO	(Battivelli and Verdin, 2018)	Addgene, #112234
pMD2.G	Addgene	12259
pWPI-eGFP	Addgene, kindly provided by Prof. Dr. Oliver T. Fackler	12254
pcDNA3.1(+)-HA-LEDGF/p75	synthesized at GenScript Biotech	
pKS070 - pCAGGS-3XFLAG-(human)CTCF-eGFP	Addgene	Addgene, #156448
pCMV-2b Flag-Integrase	(Ali et al., 2019)	

Cell lines

Name	Reference
C20 microglia	(Garcia-Mesa et al., 2017), Kindly provided by Dr Alvarez-Carbonell
Jurkat E6.1	ATCC, TIB-152
HEK293T cells	ATCC, CRL-3216
TZM-bl cells	(Wei et al., 2002)
(iPSC)-derived human microglial cells (iMG)	TempoBioscience, Cat#SKU 1001.1

Viral strains

Virus	Source	Identifier
HIV-1 (NL4.3)	NIH AIDS reagent program	ARP-114
HIVGKO	(Battivelli and Verdin, 2018)	Addgene, #112234
HIV-EGFP	(Garcia-Mesa et al., 2017)	

Kits

Kit	Company
Arcturus® PicoPure® RNA isolation kit	Applied Biosystems
CellTiter 96® Non-Radioactive Cell Proliferation Assay	Promega
ChIP DNA Clean & Concentrator	Zymo Research
DNeasy® Blood and tissue Kit	QIAGEN
End-It™ DNA End-Repair Kit	Epicentre
High-Capacity cDNA Reverse Transcription Kit	Applied Biosystems
Illumina Tagment DNA Enzyme and Buffer Small Kit	Illumina
NEBNext® dA-Tailing Module	New England Biolabs GmbH
NEBNext® Ultra™ II DNA Library Prep Kit for Illumina®	New England Biolabs GmbH
NEBNext® Multiplex Oligos for Illumina® (Index Primers Set 1)	New England Biolabs GmbH
NEBNext® Multiplex Oligos for Illumina® (96 Dual Index Primer Pairs set 1)	New England Biolabs GmbH
NucleoSpin Gel and PCR Clean-up	Machery-Nagel
NucleoSpin RNA Mini kit	Machery-Nagel
Taq DNA polymerase and taq PCR Core Kit	QIAGEN
Phusion® High-Fidelity PCR Kit	New England Biolabs GmbH

Equipment

Name	Company
CFX96 Touch™ Real-Time PCR Detection System	BioRad
Heraeus PICO 1700	Thermo scientific
Biometra TS1 ThermoShaker	Analytic Jena
Biometra TSC ThermoShaker	Analytic Jena
NanoPhotometer® NP80	Implen
Bioruptor® Plus sonication device	Diagenode

FlexCycler^2	Analytic Jena
Thermal Cycler	Applied Biosystems by lifetechnologies
Qubit 4	Thermo Fisher Scientific
UVP UVsolo touch	Analytic Jena
Vortexer	Neolabs

Software

Software name	Source
BioRender	BioRender
CFX Manager	BioRad
Excel	Microsoft
FlowJo V10	FLOWJO software
GraphPad Prism 9	GraphPad software
Image J	NIH

5.2 Molecular and Cell Biology methods

5.2.1 Cell culture and cell lines

HEK293T cells for virus production and TZM-bl cells were cultivated in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% FBS and 1% penicillin–streptomycin at 37°C and 5% CO_2.

Microglia C20 cell line (kindly provided by Dr Alvarez-Carbonell) was cultured in BrainPhys Neuronal medium supplemented with 1% penicillin–streptomycin, 10% FBS, 1% N2-Supplement, 500 mg Normocin and 1% L-Glutamine at 37°C and 5% CO_2.

Jurkat cells were cultured in Roswell Park Memorial Institute (RPMI) medium supplemented with 1% penicillin–streptomycin, 10% FBS at 37°C and 5% CO_2.

5.2.2 Induced pluripotent stem cells (iPSC)-derived human microglial cells

Sheetal Sreeram, from our collaborators at Jonathan Karn lab at Case Western Reserve University, obtained and infected iPSC-derived microglia. Briefly, iPSC-

derived human microglial (iMG) cells (Tempo Bioscience, Cat#SKU 1001.1) were plated in Tempo iMG maintenance media on plates pre-coated with Matrigel matrix (Corning, Cat#356254) according to the manufacturer's instructions. 3 days post culture, approximately 1 million iMG cells per well were infected with vesicular stomatitis virus glycoprotein (VSV-G) pseudotyped eGFP HIV-1 reporter virus at ~300 infectious units (IFU)/ 1 million cells by directly adding viral supernatants to the culture media. 24 hours post-infection (hpi), the virus-containing media was exchanged with fresh media. Percentage of GFP positive iMG cells was quantified by flow cytometry using LSR Fortessa instrument 5 days post infection (dpi). Therefore, cells were harvested, washed, and resuspended in Phosphate-Buffered Saline (PBS) buffer before fluorescence measurement. DNA from infected and uninfected iMG were harvested at 5 dpi (for infected cells) and DNA was extracted using Qiagen blood and tissue kit according to the manufacturer's instructions.

5.2.3 $CD4^+$ T cell isolation from whole blood

$CD4^+$ T cells were isolated from buffy coats obtained from healthy anonymous blood donors at the Heidelberg University Hospital Blood bank, using Ficoll density gradient purification. First, buffy coat was mixed with RosetteSep Human $CD4^+$ T cell enrichment cocktail beads (StemCell Technologies) and incubated for 20 min at RT. Next, blood was mixed in a 1:1 ratio with PBS/2% Fetal Bovine Serum (FBS), overlayed onto Ficoll (Cytiva) and centrifuged at 2400 rpm for 30 min, 37°C with breaks off. T cells were collected, washed, and incubated for 5 min at RT with ACK buffer (150 mM NH_4Cl, 10 mM $KHCO_3$, 0.1 mM Na_2EDTA). After 2 more washes with PBS /2% FBS cells were resuspend in complete T cell medium (RPMI-1640+10% FBS + IL-2 (10 ng/ml)) and activated with Dynabeads Human T-Activator CD3/CD28 (Gibco) for 72h in humidified incubators at 37°C and 5% CO_2.

5.2.4 DNA Plasmids preparation and construct assessment by restriction enzyme digestion

VSVG envelope DNA plasmids and CTCF-GFP and HA-LEDGF were transformed into DH5α competent cells (NEB). Proviral constructs were transformed into MAX Efficiency™ Stbl2™ competent cells (ThermoFisher). 1 µl plasmid was added to 30 µl competent cells and incubated for 30 min on ice. Heat shock was performed for 45 sec at 42°C and the transformation reaction was stored on ice for 2 min. 500 µl Lysogeny

broth (LB) medium was added, and cells were incubated for 1 hour at 37 °C with 120 rpm agitation. Cells were then plated on LB Agar plates containing ampicillin (50 µg/ml) and incubated over night at 37°C.

To obtain large-scale amounts of plasmids, colonies were picked and incubated at 37°C and 120 rpm in 7 ml LB medium supplemented with ampicillin (50 µg/ml) for 5-6 hours. All non-viral plasmids were cultivated in 250 ml LB-medium plus ampicillin (50 µg/ml) overnight at 37°C after inoculation with 2 ml of the small over day bacterial culture.

For proviral constructs, colony was picked and incubated at 37 °C and 120 rpm in 7 ml terrific broth (TB) medium supplemented with ampicillin (50 µg/ml) for 16 hours before transferring it into a 250 ml of TB medium larger overnight culture. Before harvesting, 150 µg/ml chloramphenicol was added to the cultures for 5-6 hours.

Plasmids were isolated with NucleoBond Xtra Maxi Plus kit (MacheryNagel) according to the manufacturer instructions. The plasmid pellet was air dried for one hour and eluted in water, vortexed and frozen at -20°C. The next day the concentration was determined with NanoPhotometer® NP80 (Implen) at 260 nm.

1 µg of the plasmids were digested with the corresponding restriction enzymes (pNL4.3 WT: NheI + EcoRI, fragment size 13319 bp+1506 bp; HIV_{GKO}: KpnI, fragment size 7000 bp, 6753 bp, 300 bp; pMD.2G: KpnI, fragment size 1594 bp + 4228 bp) and the recommended reaction buffer for 2 h at 37 °C.

5.2.5 DNA fragments separation by agarose gel electrophoresis

The digested vectors were separated by gel electrophoresis with a 1% agarose gel containing MIDORI Green Advance DNA stain (NIPPON Genetics, 1:10’000) at constant 90 V for 45 min and appropriate fragment size was verified under UV light.

5.2.6 Virus production

The day before transfection, $5x10^6$ HEK293T cells were seeded in 15 cm cell culture dishes. Viral stocks were produced by transfecting either 25 µg of pNL4.3 viral plasmid DNA (AIDS Reagent Program, ARP-114) or 45 µg HIV_{GKO} (Addgene Plasmid#112234) viral DNA together with 10 µg pMD.2G packaging plasmid (Addgene Plasmid, #12259) with polyethylenimine (PEI) transfection method. Briefly, plasmid DNA and PEI were diluted in Opti-Mem medium (Gibco) in a 1:3 ratio and added dropwise to cells after 20 min of incubation. Supernatants were collected after 48 h and filtered with a 0.45 µm

syringe filter (Roth) followed by DNase treatment (15 U/ml (Sigma) and 10 mM $MgCl_2$) for 2 h at 37°C. Viral particles were purified by 20% sucrose in PBS gradient purification at 28'000 rpm, 1.5 h at 4 °C and resuspended in PBS, before storing them at -80°C.

5.2.7 SG-Pert RT activity assay

To determine virus titers, SG-PERT assay was performed as previously described in (Pizzato et al., 2009) to quantify active RT units in the viral stocks.

5 µl of concentrated virus were lysed with 5 µl 2x lysis buffer (0.25% Triton X-100, 50 mM KCl, 100 mM TrisHCl pH 7.4, 40% glycerol) for 10 min. Ribolock™ RNase inhibitor (Fermentas, 0.4 U/µL) was added immediately prior to use. 90 µl of 1X dilution buffer (50 mM $(NH_4)_2SO_4$, 200 mM KCl and 200 mM Tris–HCl, pH 8.3) were added to the lysed sample and 10 µl of the dilution were immediately mixed with 10 µl of 2X PCR reaction mix (10 mM $(NH_4)_2SO_4$, 40 mM KCl and 40 mM Tris–Cl pH 0.3, 10 mM $MgCl_2$, 0.2 mg/ml BSA, 1/10,000 SYBR Green I, 400 µM dNTPs, 1 µM forward primer (5′-TAGTTGTTGGGCTTCGCTTT-3′), 1 µM reverse primer (reverse: 5′-TTGTCGGCTTTACCTGCTTT-3'), 1.2 µg/ml BMV RNA). HotStartTaq polymerase (0.2 U/reaction) was added immediately prior to PCR reaction.

Cycler conditions were the following using the CFX96 Touch platform (Bio Rad):

37°C	30 min	RT reaction
95°C	5 min	
95°C	5 s	
55°C	5 s	
72°C	20 s	45x
83°C	11 s	

Data analysis was performed with the CFX Maestro Software (Bio Rad). Real time Ct values of 10-fold serial dilutions of a recombinant HIV-1 RT standard were compared to the analyzed sample (activity of undiluted standard is 5,088 * 10^9 pUnits RT/µL).

5.2.8 HIV-1 infectivity assay

To assess HIV-1 infectivity, TZM-bl cell line was infected with HIV-1 virus. TZM-bl cells are engineered to harbor a β-galactosidase gene and a luciferase reporter under the

HIV-1 LTR and they express CD4, CCR5 and CXCR4 receptors for viral entry on their plasma membrane.

The day before infection, 5'000 cells per 96-well were plated. The next day, virus was diluted in 10-fold serial dilutions in DMEM medium, and medium was exchanged with virus containing medium. 48 h later, medium was removed, and cells were fixed with 4% PFA for 10 min. After washing once with PBS, β-galactosidase activity was assessed by addition of 100 µl β-galactosidase/X-gal substrate to the cells and subsequent incubation 2 h – 4 h. To calculate multiplicity of infection (MOI), blue cells were counted and infected cells / µl were calculated.

5.2.9 HIV-1 infection

Microglia cells were infected with 250 ng p24 /1 million cells of vesicular stomatitis virus glycoprotein (VSV-G) pseudotyped pNL4.3 WT or HIV_{GKO} virus through spinoculation for 90 min at 2'300 rpm and 37°C or by only mixing cells and virus. Infected cells were cultivated for 3 days at 37°C and 5% CO_2.

5.2.10 Flow cytometry

Productive infection was measured by flow cytometry analysis on BD FACSCelesta™. Therefore, cells were harvested by trypsinization and fixed with 4% PFA for 90 min to inactivate the virus. After washing with 2% FBS in PBS, HIV-1 p24 was stained with Anti-HIV-1 p24-PE (KC57) (Beckman Coulter) antibody in 0.1% Triton X-100/PBS for 30 min on ice and washed again 3 times before measuring on the BD FACSCelesta™.

5.2.11 FACS sorting

Approximately 24 million C20 microglia cells were infected with 250 ng p24 HIV_{GKO} virus/ 1 million cells by spinoculation (2'300 rpm, 1.5 h, 37°C). 3 dpi, cells were washed with PBS, trypsinized and filtered in 2% FBS/PBS through a 45 µm filter before sorting on BD FACSAria™ Fusion Cell Sorter into 3 populations: $FITC^-/PE^-$, $FITC^+/PE^+$ and $FITC^-/PE^+$. For each cell population, about 80'000 cells were sorted. For the latent $FITC^-/PE^+$ population 8'000 to 45'000 cells were sorted. 5'000 cells were resorted to check for sorting purity and used for ATAC-Seq as in (Buenrostro et al., 2015).

5.2.12 3D Immuno DNA FISH

Probe labeling

HIV-1 FISH probes were generated by labeling HIV-1 DNA (pHXB2) plasmid. Biotin-dUTP nucleotide mix containing 0.25 mM dATP, 0.25 mM dCTP, 0.25 mM dGTP, 0.17 mM dTTP and 0.08 mM biotin-16-dUTP in H_2O was prepared. 3 µg of pHXB2 were diluted in a final volume of 12 µl with H_2O and combined with 4 µl of nucleotide mix and 4 µl Nick translation mix (Roche). Labeling was performed at 15°C for 5 h. Probes were precipitated in 100% ethanol with sodium acetate (3 M), 10 µl herring sperm (Sigma) and 10 µl human COT DNA (Sigma) overnight and resuspended in 2x SSC/10% dextran sulfate/50% formamide, denatured at 95°C for 5 min and stored at -20°C until use.

Hybridization and probe development

Approximately $1.5x10^5$ microglia cells were plated on coverslips on the day of infection in a 24-well plate. 3 dpi cells were washed with PBS and fixed in 4% PFA in PBS for 10 min. Coverslips were extensively washed with PBS and cells were permeabilized in 0.5% triton X-100/PBS for 10 min, followed by another 3 washes with PBS-T. Subsequently, the cells were incubated in a solution consisting of 0.5% Triton X-100 and 0.5% saponin in PBS for 10 min. After three more PBS-T washes, the coverslips were treated with 0.1 M HCl for 10 min, followed by three additional PBS-T washes. To further permeabilize the cells, they were exposed to 0.5% Triton X-100 and 0.5% saponin in PBS for 10 min, followed by another set of three PBS-T washes. Next, RNA digestion was carried out using RNAse A (100 µg/ml) for 30 min at 37°C. The coverslips were then equilibrated for 5 minutes in 2x SSC and incubated overnight at 4°C in a hybridization solution composed of deionized formamide, 20x SSC, sterile water (pH 7.0). Probe hybridization was performed with 2 – 4 µl of HIV-1 probe in 6 µl reaction in 2x SSC/10% dextran sulfate/50% formamide, denatured at 95°C for 5 min and then placed on ice for 1 min. After spotting the probe on a glass slide and sealing the coverslip with rubber cement, the slide was placed at 80°C for 5 min on a heating plate for denaturation. Subsequently, the hybridization reaction was incubated at 37°C for 48 to 72 h in a closed 15 cm dish in an incubator. Probe detection was carried out by washing the coverslips with 2x SSC and 0.5x SSC at 37°C and 65°C, respectively; 1h blocking in TSA blocking buffer (TNB) and detection with streptavidin-HRP in TNB (1:1500) for 40 min at 37°C. Coverslips were then washed with TNT wash buffer four

times at RT, before incubation with Fluorescein Plus amplification reagent (1:500 in TSA Plus amplification diluent, part of TSA Plus Fluorescein kit) for 5 min at RT. Following this, 5 additional washes with TNT buffer were conducted, and the cell nuclei were counterstained with Hoechst 33342 (diluted 1:10,000 in PBS) followed by two PBS washes. Finally, coverslips were mounted with mowiol. For confocal microscopy and manual image analysis, 3D stacks were acquired with a Leica TCS SP8 confocal microscope using a x63 oil immersion objective.

5.2.13 Small interfering RNA (siRNA) gene expression silencing in microglia cells

The day before transfection, 2x 10^5 microglia cells were seeded in 6-well plates or 2 mill cells into 10 cm dishes (for IS sequencing and ChIP-Seq) to ensure a confluency of 50% on the day of transfection.

Transfection was performed with jetPRIME® (Polyplus transfection®) according to the manufacturer's instructions. In brief, 110 pmol (for 50 nm siRNA end concentration) of either ON-TARGETplus Non-targeting Control siRNA (Horizon Discovery), ON-TARGETplus Human PSIP1 siRNA SMARTPool, ON-TARGETplus Human CPSF6 SMARTPool (Horizon Discovery), ON-TARGETplus Human CTCF SMARTPool (Horizon Discovery) or ON-TARGETplus Human RAD21 SMARTPool were mixed with 200 µl jetPRIME® buffer. Then 4 µl jetPRIME® reagent were added to the siRNA mix, incubated for 10 min at RT and added dropwise to the cells.

For IS sequencing and ChIP-Seq in CTCF KD conditions, Lipofectamine™ RNAiMAX transfection reagent was used. In brief, 22 pmol (for 10 nm siRNA end concentration) of either ON-TARGETplus Non-targeting Control siRNA (Horizon Discovery) or ON-TARGETplus Human CTCF SMARTPool (Horizon Discovery) were diluted in 1 ml Opti-MEM medium (Gibco). 30 µl Lipofectamine® RNAiMAX Reagent were diluted in 1 ml Opti-MEM medium, mixed with the diluted siRNA and incubated for 10 min at RT. Mixture was added dropwise to the cells.

48 h post transfection, cells were trypsinized, counted and replated for infection. Cell pellets for western blot analysis to confirm KD were harvested. Microglia cells were infected with 250 ng p24/ 1 million of cells of VSV-G pseudotyped HIV-1 WT or GKO virus and harvested after 24h of infection for western blot analysis and DNA extraction.

5.2.14 siRNA CTCF expression silencing in CD4+ T cells

siRNA transfection in CD4+ T cells was performed using the Amaxa™ 4D-Nucleofector™ Protocol for stimulated Human T Cells (Lonza) according to manufacturer's instructions. Briefly, 10 million cells were one time washed in PBS and then resuspended in 100 µL P3 Primary Cell 4D-Nucleofector™ Solution plus 600 nM siRNA pool targeting CTCF or non-targeting control (Horizon) and were transfected with program EH100 using Single Nucleocuvette™ in a 4D-Nucleofector® X Unit (Lonza). Transfected cells were incubated 10 min with RPMI-1640 medium without FBS or antibiotics at 37°C before resuspension in complete RPMI-1640 medium.
48 h after transfection, CD4+ T cells were infected with HIV-1_{GKO} by spininfection at 2'150 rpm, 2 h at 32°C to establish single round infections.

5.2.15 MTT assay

Cell viability upon siRNA transfection was measured through the CellTiter 96® Non-Radioactive Cell Proliferation Assay (3-[4,5-dimethylthiazol-2-yl]-2,5 diphenyl tetrazolium bromide (MTT)) (Promega). 1x 10^4 microglia cells were plated per well in a 96 well plate and transfected with either ON-TARGETplus Non-targeting Control siRNA (Horizon Discovery), ON-TARGETplus Human PSIP1 siRNA SMARTPool, ON-TARGETplus Human CPSF6 SMARTPool (Horizon Discovery), ON-TARGETplus Human CTCF SMARTPool (Horizon Discovery) or ON-TARGETplus Human RAD21 SMARTPool with jetPRIME® (Polyplus transfection®) according to the manufacturer's instructions for each time point of interest. On the day of transfection and the time point of infection, new medium plus MTT solution (15 µl) was added to the cells and after 2 h the reaction was stopped by the addition of 100 µl of the solubilization/stop solution. Absorbance values at 570 nm were acquired the next day with an Infinite 200 PRO (Tecan) multimode plate reader. Reactions were conducted in triplicate, and the average signal of the triplicates was normalized over the matched NT controls and expressed as percentage.

5.3 Nucleic acid based methods

5.3.1 RNA isolation of sorted RNA

RNA was isolated for mRNA expression analysis with Arcturus™PicoPure™ RNA Isolation Kit according to manufacturer's instructions. Concentration and purity were

determined using a NanoPhotometer® NP80 (Implen). The concentration was measured by evaluating the optical density at 260 nm and the OD260nm/OD280nm ratio was verified to be within the range of 1.8 and 2.0 to ensure sample purity.

5.3.2 Reverse transcription

RNA was isolated from uninfected and HIV-1 infected cells (WT and HIV_{GKO}) 3 dpi with NucleoSpin RNA Mini kit (MacheryNagel) according to manufacturer's instructions. 250 to 500 ng of RNA were reverse transcribed using High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems). In brief, 10x RT buffer, 25x dNTP Mix (100 nM), 10x RT Random primers, RNase inhibitor and MultiScribe Reverse Transcriptase were mixed with 250 ng to 500 ng of RNA. RNA was reversed transcribed using SimpliAmp Thermal Cycler under the following conditions:

25°C	10 min
37°C	120 min
85°C	5 min
4°C	∞

5.3.3 Quantitative real time PCR

cDNA of HIV-1 infected microglia was diluted 1:10 and subjected to quantitative polymerase chain reaction (qPCR) analysis in triplicates using iQ™ SuperMix (BioRad) and commercially available TaqMan Gene Expression Assays (Thermo Fisher Scientific). Activation was measured by IL-6 (HS00174131_m1), CXCL8 (HS00174103_m1), IL1b (HS00174128_m1) and TNFα (HS1555410_m1) expression. Viral mRNA expression was measured with HIV-1 gag primer probe mix (500nM of each primer, 200 nM probe) (Appendix Primer List) and normalized to Eukaryotic 18S rRNA Endogenous Control (VIC™/TAMRA™ probe, primer limited) (Applied Biosystems) using the following program:

98°C	3 min	
98°C	10 sec	
60°C	40 sec	45x
72°C	10 min	

Relative expression was calculated using the $\Delta\Delta C_t$ method (Livak and Schmittgen, 2001). Statistical analysis was performed using GraphPad Prism.

5.3.4 DNA isolation

DNA was extracted using Blood and tissue kit (Qiagen) according to the manufacturer's instructions and concentration and purity determination was performed using a NanoPhotometer® NP80 (Implen). The concentration was measured by evaluating the optical density at 260 nm and the OD260nm/OD280nm ratio was verified to be within the range of 1.8 and 2.0 to ensure sample purity.

5.3.5 Integration assay by Alu-PCR

The amount of integrated HIV-1 vDNA was quantified through *Alu*-PCR (Liszewski et al., 2009; Tan et al., 2006). Briefly, in a nested PCR, an integrated virus is amplified with an *Alu* specific primer harboring a λ tag and a HIV-1 LTR specific primer in the first reaction (Appendix Primer List). In the second qPCR, proviral sequences are amplified using lambda-specific primer (λT) and an internal LTR primer. First round PCR was performed with Alu1 primer (0.1 µM), HIV LTR specific primer LM667 (0.3 µM), 10 mM dNTPs, 10x buffer, AmpliTaq DNA polymerase (Thermo Fisher Scientific) and 100 ng DNA. PCR was performed using SimpliAmp Thermal Cycler (Thermo Fisher Scientific) with the following reaction parameters:

94°C	15 sec	
94°C	15 sec	20x
55°C	30 sec	
70°C	2 min	
72°C	10 min	

The first PCR reaction was diluted (1:100), and the second round was performed by qPCR using iQ™ SuperMix (BioRad), lambda-specific primer λT, internal LTR primer LR and TaqMan probe (900 nM each primer, 200 nM probe, Appendix Primer List). The same qPCR was performed with 10 ng genomic DNA and a housekeeping gene, i.e., lamin B2, B13 region (Appendix Primer List) which was used to normalize relative integration levels by the $\Delta\Delta C_t$ method (Livak and Schmittgen, 2001). Reaction was

performed on a CFX96 C1000 Touch Thermal Cycler (BioRad) in triplicates with the following conditions:

98°C	3 min	
98°C	10 sec	45x
60°C	40 sec	
98°C	10 min	

5.3.6 Total viral DNA quantification by qPCR

Relative total viral DNA levels between CTCF KD and NT conditions in $CD4^+$ T cells were quantified by qPCR using iQ™ SuperMix (BioRad), primers targeting viral gag DNA sequence and TaqMan probe (900 nM each primer, 200 nM probe, Appendix Primer List) and 200 ng input infected genomic DNA. The same qPCR was performed with 10 ng genomic DNA and a housekeeping gene, i.e., lamin B2, B13 region (Appendix Primer List) used to normalize total viral levels by the $\Delta\Delta C_t$ method (Livak and Schmittgen, 2001). Reaction was performed on a CFX96 C1000 Touch Thermal Cycler (BioRad) in triplicates with the following conditions:

98°C	3 min	
98°C	10 sec	45x
60°C	40 sec	
98°C	10 min	

5.3.7 Digital droplet PCR

HIV-1 total viral DNA products and 2LTR circles of infected cells were detected by Digital droplet PCR (ddPCR), as described previously (Bejarano et al., 2018; Morón-López et al., 2017). Total genomic DNA was isolated with Qiagen blood and tissue kit according to manufacturer's instructions, and between 50-100 ng DNA were used as input. As an alternative, cell lysate of 50'000 cells (lysis buffer: 10 mM Tris HCl pH=9, 0.1% Triton X-100, 400 µg/mL proteinase K) was used to detect total viral DNA and 2LTR products.

Viral DNA products were detected with a primer-probe set annealing to the HIV-1 gag sequence and one for 2LTR circles. Sample input was normalized by detection of host ribonuclease P protein subunit p30 (RPP30) (Sequences of primers and probes are

listed in Appendix Primer List). For each reaction 900 nM of each primer, 200 nM probe and 1x ddPCR Supermix for probes (no dUTP) (BioRad, Hercules, CA, USA) were prepared and droplets were generated using QX200™ Droplet Generator (BioRad). Targets were amplified by PCR amplification before droplets sorting and analysis in a QX200 droplet reader (BioRad). Target counts were analyzed by software QuantaSoft v1.6 (BioRad) using absolute quantification settings and HIV-1 copy numbers and 2LTRs were normalized using RPP30 housekeeping gene.

5.4 Sequencing methods

5.4.1 Chromatin immunoprecipitation (ChIP)

$10x10^6$ million microglia cells were cultured in a 15 cm cell culture dish. Medium was removed and cells washed once in PBS (plus 1 mM sodium butyrate for H3K27ac IP to block deacetylases). Cells were fixed with 1% formaldehyde/PBS (plus 1 mM sodium butyrate for K27ac) for 7 min at RT followed by quenching with 0.125 M glycine/PBS for 7 min at RT. After removing all liquid, cells were scraped 2 times in 5 ml ice cold PBS pelleted by centrifugation at 1700 rpm, 7 min. After 2 washings with 10 ml of cold PBS, pellet was resuspended in swelling buffer (10 mM HEPES/KOH pH 7.9, 85 mM KCl, 1mM EDTA, 0.5% IGEPAL CA-630, 1x protease inhibitor cocktail (Roche)) and incubated for 10 min rotating at 4°C. Following this, the pellet was dounced ten times before being centrifuged at 3500xg for 10 minutes at 4°C. An additional wash with swelling buffer, minus IGEPAL CA-630, was performed before resuspending the nuclei in a cold sonication buffer (TE with pH=8, 0.1% SDS, and a protease inhibitor tablet). Sonication was was carried out using a Covaris S220 Focused Ultrasonicator for 18 min (Duty cycle 20%, Intensity 5, Cycles/burst 200) or a Covaris M220 Focused-ultrasonicator for 8 min (Duty cycle 20%, peak power 75, Cycles/burst 200). DNA size was followed by 2% agarose gel. DNA fragments should be between 200-500 bp. Triton-X was added to the lysate to a final concentration of 1% and incubated for 10 min on ice. Lysate was cleared by centrifugation at 18'000xg 4°C for 5 minutes. Magna ChIP Protein A and G magnetic beads from Millipore were washed twice with TE containing 0.1% SDS and 1% TritonX, and then added to the lysate for preclearing for 1 hour at 4°C with rotation. Subsequently, 2-8 µg of chromatin were incubated overnight at 4°C with the corresponding amount of antibody. A 1% aliquot of the chromatin was saved as input. The used antibodies were the following:

H3K36me3 (ab9050, Abcam), H3K27ac (ab4729, Abcam), H3K27me3 (C36B11, Cell Signaling), H3K4me1 (ab8895, Abcam), H3K9me3 (ab8898, Abcam), H3K9me2 (ab1220, Abcam), CTCF (D31H2) XP® Rabbit mAb (#3418, Cell Signaling), Rabbit control IgG (ab46540, Abcam), Mouse control IgG2a (ab18413, Abcam).

On the following day, Protein A and G magnetic beads were washed twice with sonication buffer containing 1% TritonX-100 and then incubated with the lysates for 2 hours at 4°C. Subsequently, beads were washed twice 10 min with cold buffer I (150 mM NaCl, 1% Triton X-100, 0.1% SDS, 2 mM EDTA, protease inhibitor cocktail (Roche)), once 10 min with cold buffer II (10 mM Tris/HCl pH 7.5, 250 mM LiCl, 1% IGEPAL CA-630, 0.7% Deoxycholate, 1 mM EDTA, protease inhibitor cocktail (Roche)), twice 10 min with cold TET buffer (10 mM Tris/HCl pH7.5, 1 mM EDTA, 0.1% Tween-20, protease inhibitor cocktail (Roche)) and eluted with TE buffer, 1% SDS, 100 mM NaCl. 0.5 mg/ml Proteinase K were added, and samples were incubated for 2 h at 55°C and overnight at 65°C. The next day, 0.33 mg/ml RNAse A (Thermo Fisher Scientific) was added and incubated for 1 h at 37°C. Supernatant was removed from the magnetic beads and DNA was purified with AMPure beads XP clean up according to manufacturer's instructions. Concentrations were determined by Qubit Fluorometer and enrichment was determined by qPCR on a CFX96 C1000 Touch Thermal Cycler (BioRad). ChIP libraries were prepared using NEBNext® Ultra™ II DNA Library Prep Kit for Illumina® (NEB) and NEBNext® Multiplex Oligos for Illumina® (Index Primers Set 1) or NEBNext® Multiplex Oligos for Illumina® (96 Unique Dual Index Primer Pairs, Set1) (NEB) according to manufacturer's instructions. Libraries were sequenced at c.ATG sequencing core facility at Tübingen University on a NextSeq instrument 2x75 bp or on a NovaSeq6000 instrument 2x100 bp.

5.4.2 ChIP-qPCR

ChIP DNA and input was diluted 1:10 and subjected to qPCR analysis in triplicates using iTaq™ Universal SYBR® Green Supermix (BioRad) or SsoFast™ EvaGreen® Supermix (BioRad) and commercially available Human Chromatin IP Control qPCR Primer Sets (Active Motif): Human Negative Control Primer Set 2 (Active Motif), Human Positive Control Primer Set GAPDH-1 (Active Motif), Human Positive Control Primer Set MYT1 (Active Motif), Human Positive Control Primer Set ACTB-2 (Active Motif), Simple Chip® Human Sat2 Repeat Element Primers (Cell Signaling).

H3K36me3 and CTCF IP were controlled by Human Positive Control Primer Set GAPDH-1 and Human Negative Control Primer Set 2 (Active Motif). CTCF control primers were binding to human H19 internal control region and are listed in Appendix Primer List (500 nM of each primer). Reaction was performed on a CFX96 C1000 Touch Thermal Cycler (BioRad) in triplicates with the following conditions:

98°C	3 min	
98°C	10 sec	45x
60°C	40 sec	
72°C	10 min	

Input samples were adjusted to 100% and IP enrichment was expressed as percentage of input using $100*2^{(\text{Adjusted input - Ct (IP)})}$. Statistical analysis was performed using GraphPad Prism.

5.4.3 ATAC-Seq

ATAC-sequencing was performed as in (Buenrostro et al., 2015). Cells were pelleted at 500xg, 4°C for 5 min, washed once with cold PBS and resuspended in 50 µl cold lysis buffer (10 mM Tris HCl, pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.1% (v/v) Igepal CA-630). The reaction was centrifuged immediately for 10 min at 500xg, 4°C. Supernatant was removed and transposition reaction was set up with 25 µl TD (2x reaction buffer, Nextera Kit, Illumina), 2.5 µl TDE1 (Tn5 transposase, Nextera Kit, Illumina) and 22.5 µl nuclease free H_2O. The reaction was incubated for 30 min at 37°C shaking on a thermal block (350 rpm). DNA was purified with Zymo ChIP clean and concentrator kit (Zymo research) according to manufacturer's instructions. Eluted DNA fragments were amplified by PCR with 1.25 µM Primer 1, 1.25 µM Primer 2 (Appendix Primer List), 25 µl NEBNext® Ultra™ II Q5® Master Mix (NEB) and the whole purified DNA sample with the following conditions:

72°C	5 min	
98°C	30 sec	
98°C	10 sec	5x
63°C	30 sec	
72°C	1 min	

To determine the remaining cycle number for appropriate amplification, 5 µl of the pre-amplified reactions were amplified by qPCR with SsoFast™ EvaGreen® Supermix (BioRad) and 0.25 µl (25 µM) Primer 1, 0.25 µl 25 µM Primer 2 under the following conditions:

98°C	30 sec	
98°C	10 sec	
63°C	30 sec	20x
72°C	1 min	

Linear RN was plotted versus cycle number to determine the cycle number corresponding to one-third of the maximum fluorescent intensity. Subsequently, the remaining PCR reaction was conducted using the determined cycle number, employing the same cycling conditions as previously. The resulting DNA was then purified using the Zymo ChIP Clean and Concentrator Kit (Zymo Research), following the manufacturer's instructions. The libraries were sequenced at the c.ATG Sequencing Core Facility at Tübingen University using a NovaSeq instrument with 2x50 bp sequencing settings.

5.4.4 Integration site sequencing

Infected C20 microglia cells were harvested at 3 dpi and DNA was isolated with Qiagen blood and tissue kit according to manufacturer’s instructions. DNA was sonicated with Covaris S220 Focused Ultrasonicator 450 s (Duty factor 10%, 200 cycles per burst) or Covaris M220 Focused Ultrasonicator for 2 min (Duty cycle 20%, peak power 75, Cycles/burst 200) and size was monitored by agarose (2%) gel analysis. Desired fragment size was below 500 bp.

The LM PCR protocol was adapted from (Serrao et al., 2016). Sonicated gDNA ends were repaired using End-It™ DNA End-Repair Kit (Epicentre) according to manufacturer’s instructions. Sample was purified with PCR purification Kit (MacheryNagel) or 1.8x AMPure beads XP according to manufacturer’s instructions. A-tailing was performed using NEBNext® dA-Tailing Module (NEB). The reaction was incubated in a thermal cycler for 30 minutes at 37°C and purified with PCR purification Kit (MN) or 1.8x AMPure beads XP according to manufacturer’s instructions. An asymmetric double stranded linker was annealed overnight at 12°C (800 U T4 ligase,

10x ligase buffer, 1.5 µM linker). After purification with PCR purification kit or 0.9x AMPure XP clean up beads according to manufacturer's instructions, first nested PCR with linker and LTR specific primers was performed (Appendix Primer List). 25 µl PCR reaction was set up with 0.8 µM LTR specific primer, 0.8 µM linker specific primer, 5x buffer, 2.5 mM dNTPs, Phusion polymerase (NEB) and 100 ng ligation reaction. The thermal cycler program was the following:

98°C	30 sec	
98°C	10 sec	30x
61°C	30 sec	
72°C	30 sec	
72°C	10 min	

PCR reactions were purified with PCR purification kit (MN) according to manufacturer's instructions (or 0.9x AMPure XP clean up beads*)*. The second nested PCR with the linker specific primer and an inner LTR specific primer containing Illumina sequencing indices (Appendix Primer List) was performed under the following conditions:

98°C	30 sec	
98°C	10 sec	15x to 13x
64°C	30 sec	
72°C	30 sec	
72°C	10 min	

PCR reactions were purified with PCR purification kit (MN) according to manufacturer's instructions (or 0.9x AMPure XP clean up beads). After quality assessment, IS libraries were sent for sequencing to c.ATG sequencing core facility at Tübingen University and sequenced on a MiSeq instrument 2x150 bp.

5.4.5 Library quality assessment

Quality of the sequencing libraries was analyzed by Bioanalyzer (Agilent) and NEBNext® Library Quant Kit for Illumina® (NEB). To control DNA fragment size distribution in high resolution, Agilent High Sensitivity DNA Kit (Agilent) was used according to manufacturer's instructions. 1 µl of library sample was loaded onto a

prepared Bioanalyzer Chip and read on a Bioanalyzer 2100 instrument (Agilent). For integration site and ChIP-Seq libraries, fragment size had to be around 500 bp, ATAC-Seq libraries had to display mono-, di- and trinucleosomal size distribution.
Complete adapter ligation in the IS library preparation was assayed using NEBNext® Library Quant Kit for Illumina® (NEB) according to manufacturer's instructions. Library fragments were amplified by qPCR using primers complementary to P5 and P7 Illumina sequencing adapters and library molarity was calculated using the kit's standard.

5.4.6 RNA Sequencing

RNA was isolated from 2.5x10^6 untreated microglia cells (3 biological replicates) with NucleoSpin RNA Mini kit (MacheryNagel) according to manufacturer's instructions and RNA concentration and quality was assessed using a NanoPhotometer® NP80 (Implen). Samples were sent for RNA sequencing at the Next Generation Sequencing Core Facility of Heidelberg University. Library preparation and rRNA depletion was performed at the facility. Sequencing was performed on a Next Seq instrument, Mid output 2x150bp.

5.5 Biochemistry methods

5.5.1 SDS-PAGE and Immunoblot analysis

Cell pellets were lysed in 1x Radioimmunoprecipitation assay (RIPA) buffer (Abcam, ab156034) supplemented with 1x protease inhibitor cocktail (Roche) for 10 min at 4 °C and homogenized by sonication in a water bath for 10 min. Protein concentration was assayed using the Micro BCA Protein Assay Kit (Thermo Fisher Scientific) and equal total protein amounts were loaded and run on a precast NuPAGE Bis-Tris 4–12% (Thermo Fisher Scientific) SDS–PAGE at 120 V. Protein transfer was performed with the Trans-Blot® Turbo™ Transfer System (BioRad) using Trans-Blot Turbo Mini 0.2 µm Nitrocellulose Transfer Packs (BioRad). Proteins were transferred on a nitrocellulose membrane for 10 min, 25V, 2.5A. After blocking the membrane for 1 h with 5% milk in 0.1% PBS-Tween (PBS-T) on a rocker at RT, they were incubated with primary antibody overnight at 4°C on a rocker (anti-beta actin (1:10000), a2228, Sigma Aldrich; LEDGF/p75 Antibody (1:2000), A300-848A, Bethyl labs; Anti-CPSF6 antibody (1:2000), ab99347, Abcam; CTCF (D31H2) XP® Rabbit mAb (1:1000) #3418, Cell

Signaling). The primary antibodies were diluted in 5 ml of 5 % milk in 0.1% PBS-T. The following day, the membranes were washed 3 times for 5 min with PBS-T and incubated with horseradish peroxidase linked secondary antibody (mouse or rabbit) in a 1:5000 dilution in 5% milk in 0.1% PBS-T on a rocker at RT. After washing 3 times for 5 min with PBS-T, proteins were visualized with SuperSignal™ West Pico PLUS Chemiluminescent Substrate (Thermo Fisher Scientific).

5.5.2 Co Immunoprecipitation with GFP-Trap® Agarose beads

$10x10^6$ million HEK293T cells were transfected with 5 µg CTCF-GFP-Flag plasmid (Addgene) or 5 µg pWPI-eGFP control plasmid (Addgene, kindly provided by Prof. Dr. Oliver T. Fackler) with jetPRIME® (Polyplus transfection®). Corresponding plasmid DNA was mixed with 200 µl jetPRIME® buffer. Then 2x of plasmid amount jetPRIME® reagent was added to the DNA mix, incubated for 10 min at RT and added dropwise to the cells.

24 h post transfection, cells were trypsinized, washed once with PBS and lysed in 200 µl RIPA buffer supplemented with 1x protease inhibitor cocktail (Roche) and 15 U/ml DNAse for 1 h on ice. Lysate was sonicated in water bath for 25 min (30' on 30' off in ice cold water bath) and cleared at 17’000x g for 10 minutes at 4°C. Volume was adjusted with dilution buffer (10 mM TRIS/Cl pH 7.5, 150 mM NaCl, 0.5 mM EDTA, 1 mM PMSF, 1x protease inhibitor cocktail) to 500 µl. To prepare the GFP-Trap® beads, 25 µl of the beads were equilibrated in dilution buffer by centrifugation at 2500x g for 5 minutes at 4°C, followed by two washes with dilution buffer. The cell lysate was then added to the equilibrated GFP-Trap® beads, and the mixture was incubated on a rotating wheel for 1 hour at 4°C. After incubation, the beads were washed three times with a low salt wash buffer (10 mM TRIS/Cl pH 7.5, 150 mM NaCl, 0.5 mM EDTA, 0.05% NP-40, and 1x protease inhibitor cocktail), followed by one wash with a high salt wash buffer (250 mM NaCl). The immunoprecipitated proteins were eluted from the beads using 80 µl of 2x loading buffer (50 mM Tris HCl pH 6.8, 15% Sucrose, 2 mM EDTA, 3% SDS, 0.01% Bromphenol blue). Elution was achieved by boiling for 30 minutes at 65°C, followed by a 5-minute boil at 95°C. Finally, the eluted proteins were loaded onto an SDS page for further analysis.

5.5.3 Endogenous CTCF, HA-LEDGF/p75 and Flag-IN immunoprecipitation

20-40x10^6 million C20 microglia like cells, or Jurkat cells were lysed in 600 µl hypotonic lysis buffer (HEPES-KOH pH 7.9 10 mM, $MgCl_2$ 1.5 mM, KCl 10 mM, DTT 0.5 mM, PMSF 0.2 mM, 1x protease inhibitor cocktail) for 30 min on ice. For LEDGF/p75 immunoprecipitation, 10x10^6 million HEK293T cells were transfected with 5 µg pcDNA3.1(+)-HA-LEDGF/p75 plasmid (synthesized at GenScript Biotech) using jetPRIME® (Polyplus transfection®). For Flag-IN immunoprecipitation, 10x10^6 million HEK293T cells were transfected with 5 µg IN-Flag plasmid (REF) with jetPRIME® (Polyplus transfection®). For Flag-IN immunoprecipitation in LEDGF/p75 KD conditions, 10x10^6 million HEK293T cells were transfected with 5 µg IN-Flag plasmid (REF) and either 10 nM ON-TARGETplus Non-targeting Control siRNA (Horizon Discovery) or 10 nM ON-TARGETplus Human PSIP1 siRNA SMARTPool Horizon Discovery). 48 h post transfection, cells were lysed for immunoprecipitation.

Lysates were dounced 15x and cytoplasm was extracted by centrifuging 5 min at 1.100xg at 4°C. Nuclei were lysed in 200 µl of nuclei lysis buffer (HEPES-KOH pH 7.9 20 mM, $MgCl_2$ 1.5 mM, NaCl 200 mM, DTT 0.5 mM, PMSF 0.2 mM, EDTA 0.2 mM, NP-40 0.5%, Glycerol 25%, 1x protease inhibitor cocktail) for 1 h on ice. Lysate was sonicated using Covaris M220 sonicator, 200 burst, 10%, 75 peak power for 5 min. DNA was digested with 150 units/ml of DNase and 2.5 mM $MgCl_2$ for 30 min on ice before lysate was cleared at 17'000xg, 10 min, at 4°C. Three parts dilution buffer (HEPES-KOH pH 7.9 20 mM, NP-40 0.5%, NaCl 150 mM, 1x protease inhibitor cocktail) were added to the lysate and pre-cleared with Dynabeads™ Protein A+G (10 µl each per reaction, Invitrogen™) for 1 h at 4°C. Antibodies (CTCF (D31H2) XP® Rabbit mAb (Cell Signaling, #3418), Rabbit control IgG (ab46540, Abcam), anti HA-tag antibody (C29F4) (Cell Signaling, 3724), Monoclonal anti-Flag M2 antibody produced in mouse (Sigma, F3165-1MG)), were pre-coupled to Dynabeads™ Protein A+G (25 µl each per reaction) in TBS-T buffer (TRIS-HCl pH 7.5 50 mM, Tween 0.1%, NaCl 150 mM, 1x protease inhibitor cocktail) for 1 h at 4°C rotating on a wheel.

Coupled antibodies were washed one time carefully with TBS-T buffer before combining them with the pre-cleared lysate. They were incubated over night at 4°C rotating on a wheel.

The next day, beads were washed 3 times with low salt low salt wash buffer (HEPES-KOH pH 7.9 20 mM, $MgCl_2$ 1.5 mM, NP-40 0.1%, NaCl 150 mM, 1x protease inhibitor cocktail) followed by 2 times high salt wash buffer (250 mM NaCl).

Immunoprecipitated proteins were eluted with 50 µl of 2x loading buffer (50 mM Tris HCl pH 6.8, 15% Sucrose, 2 mM EDTA, 3% SDS, 0.01% Bromphenol blue) by boiling 30 min at 65°C, 5 min at 95 °C and loaded onto an SDS page for analysis.

5.5.4 Biochemical fractionation

Protocol was modified from (Aprile-Garcia et al., 2019; Shytaj et al., 2020). 2 million C20 cells were infected with 1 µg p24 of HIV_{GKO} virus per 1 million cells in 10 cm dishes. 3 dpi, cells were washed with PBS, trypsinized and harvested in low binding Eppendorf tubes. Pellet was washed twice with ice-cold PBS and resuspended in 750 µl buffer A (0.1% NP40, 1 mM DTT, protease inhibitor tablet) and incubated on ice for 10 min. After swelling, an aliquot was taken for whole cell lysate before centrifugation 30 sec at 4°C 15'000 rpm. Cytoplasmatic fraction was saved, and nuclei were washed in the same buffer without the detergent. Nuclear pellet was resuspended in 250 µl of 0.5 M buffer B (20 mM HEPES, pH 7.9, 0.5 M KCl, 1.5 mM $MgCl_2$, 0.1 mM EDTA, 0.5% NP-40, 10% glycerol, protease inhibitor tablet) and incubated on a wheel for 30 min at 4°C. The suspension was centrifuged at 20,000g for 30 min and an aliquot of the soluble chromatin fraction was saved. Chromatin pellet was resuspended in 250 µl of 2 M buffer B (20 mM HEPES, pH 7.9, 2 M KCl, 1.5 mM MgCl2,0.1 mM EDTA, 0.5% NP-40, 10% glycerol, protease inhibitor tablet), passed through a 200-µl cut tip ten times and incubated on wheel for 30 min at 4°C. Chromatin fractions were then sonicated for 10-15 cycles (30 s on, 30 s off, high power) in a water bath. High salt and insoluble fractions were centrifuged at 20'000xg for 30 min at 4°C. Supernatant was saved as high salt fraction and pellet was resuspended in 125 µl 1x Pierce™ Lane Marker Reducing Sample buffer (ThermoFisher).

All fractions were loaded onto SDS page and probed for CTCF (D31H2) XP® Rabbit mAb (1:1000) #3418, Cell Signaling, pAb to Histone H2B (1:10'000), ab1790, Abcam and alpha-Tubulin, T6074, Sigma.

5.6 Statistical analysis

The statistical analysis of the datasets was performed with GraphPad Prism software. For datasets with a normal distribution, the Student's t-test was employed to determine statistical significance. Significance levels were denoted as follows: "ns" for not significant, "*" for p-value < 0.05, "**" for p-value ≤ 0.01, "***" for p-value ≤ 0.001 and "****" for p-value ≤ 0.0001.

5.7 Bioinformatic analysis

All Bioinformatic analysis of this study was performed by Ana Luisa Costa and Carl Herrmann form the Health Data science unit, University clinics Heidelberg and are described in Rheinberger et al., 2023.

Data sets used in this study

Data set	Source	Identifier
RNA-Seq (C20)	(Rheinberger et al., 2023)	GSE205915
scRNA-Seq (iPSC-derived microglia)	(Rheinberger et al., 2023)	GSE205915
RNA-Seq (CD4+ T cells)	(Lucic et al., 2019)	GSE122735
ChIP-Seq (C20) for: - H3K36me3 - H3K27ac - H3K4me1 - H3K9me2 - H3K9me3 - H3K27me3	(Rheinberger et al., 2023)	GSE205915
ChIP-Seq (C20) for CTCF upon CTCF knock-down and CTCF NT	(Rheinberger et al., 2023)	GSE205915
ChIP-Seq (C20) for H3K36me3 upon CTCF knock-down and CTCF NT	(Rheinberger et al., 2023)	GSE205915
ChIP-Seq ($CD4^+$ T cells) for: - H3K36me3 - H3K27ac - H3K4me1 - H3K9me2 - H3K9me3 - H3K27me3	The ENCODE Project Consortium	Reference epigenome ENCSR479XLD

ChIP-Seq ($CD4^+$ T cells) for: - H3K36me3 - H3K27ac	(Lucic et al., 2019)	GSE122826
ChIP-Seq ($CD4^+$ T cells) for CTCF	(Javierre et al., 2016)	GSE131055
ATAC-Seq (C20) for: - Actively infected cells - Latently infected cells - Uninfected cells	(Rheinberger et al., 2023)	GSE205915
Integration site sequencing (C20)	(Rheinberger et al., 2023)	GSE205915
Integration site sequencing iPSC-derived microglia	(Rheinberger et al., 2023)	GSE205915
Integration site sequencing CTCF KD and NT (C20)	(Rheinberger et al., 2023)	GSE205915
Integration sites ($CD4^+$ T cells)	(Lucic et al., 2019)	GSE134382
Integration sites (MDM)	(Kok et al., 2016)	
Integration sites (Jurkat)	(Li et al., 2020a)	PRJNA647337
TADs ($NeuN^-$)	(Hu et al., 2021)	PsychENCODE Knowledge Portal, accession number syn4921369
TADs ($CD4^+$ T cells)	(Qi et al., 2021; Schmitt et al., 2016)	https://osf.io/u8tzp

6 Appendices

6.1 List of Abbreviations

Abbreviation	Description
7SK snRNP	7SK small nuclear ribonucleoprotein
3D	Three-dimensional
AIDS	Acquired Immune Deficiency Syndrome
AP-1	Activator protein 1
ART	antiretroviral therapy
ATAC-Seq	Assay for Transposase-Accessible Chromatin using sequencing
BACH2	BTB Domain And CNC Homolog 2
BBB	Blood brain barrier
BET	bromodomain and extra-terminal domain
bp	base pair
CA	capsid protein
CCD	catalytic core domain
CD4	cluster of differentiation 4
CDK9	cyclin dependent kinase 9
cDNA	complementary DNA
ChIP-Seq	chromatin immunoprecipitation followed by sequencing
CNS	Central nervous system
CPSF6	cleavage and polyadenylation specificity factor subunit 6
CSF	cerebrospinal fluid
CTCF	CCCTC-binding factor
CTD	carboxyl-terminal domain
CYPA	cyclosporine A
DMEM	Dulbecco’s Modified Eagle Medium
dNTP	deoxyribonucleotide triphosphate
EBV	Epstein-Barr virus
EC	Elite controller
EDTA	ethylenediaminetetraacetate
EHMT2	euchromatic histone lysine methyltransferase 2
env	envelope

EZH	enhancer of zeste homolog
FBS	fetal bovine serum
FIRE	Frequently interacting regions
gag	Group specific antigen
GALT	Gut associated lymphoid tissue
h	hours
HAND	HIV-associated neurological disorders
HBV	Hepatitis B virus
HDAC	histone deacetylase
HDGF	hepatoma-derived growth factor
HEXIM-1	hexamethylene bis-acetamide-inducible protein 1
HIV-1	Human Immunodeficiency Virus 1
HMG	High mobility genes
HMT	histone methyltransferase
HRP	HDGF-related protein
HSV	herpes simplex virus
HTLV-1	human T-lymphotropic virus 1
ICR	imprinting control region
IN	integrase
IL1β	interleukin 1β
IP	immunoprecipitation
iPSC	induced pluripotent stem cell
IS	integration site
kb	kilo base
kDa	kilodalton
KD	Knock down
KO	Knock out
µl	microliter
µM	micromolar
LADs	lamin associated domains
LAM	linear amplification mediated
LEDGF/p75	lens epithelium-derived growth factor
LM	linker mediated
LTR	long terminal repeats

LRA	latency reversing agents
MA	matrix
mESC	Mouse embryonic stem cells
ml	milliliter
min	minute
MKL2	Myocardin-like protein2
MLV	murine leukemia virus
NELF	negative elongation factor
NF-κB	nuclear factor 'kappa-light-chain-enhancer' of activated B-cells
NFATc	Nuclear factor of activated T-cells, cytoplasmic
ng	nanogram
NHEJ	Non-homologous end joining
NIPBL	Nipped-B-like protein
nm	nanometer
nt	nucleotide
NTD	amino-terminal domain
NPC	nuclear pore complex
NUP	nucleoporin
P2Y12	P2Y purinoceptor 12
PBS	phosphate buffered saline
PIC	Pre-integration complex
PLWH	People living with HIV
pol	polymerase
PRC	Polycomb Repressive Complex
PSIP1	PC4 and SFRS1 interacting protein 1
PTEFb	positive transcription elongation factor b
PWWP	proline-tryptophan-tryptophan-proline
RANBP2	RAN binding protein 2
Rev	Regulator of virion
RIGs	recurrent integration genes
RNA	ribonucleic acid
POL-II	RNA polymerase II
RT	reverse transcriptase

RT	room temperature
rpm	rounds per minute
sec	second
SE	Super-enhancer
SETD	SET-domain containing proteins
SMC	Structural Maintenance of Chromosomes
SPAD	Speckle-associated domain
SSC	saline sodium citrate
STAT5B	Signal transducer and activator of transcription 5B
TAD	Topologically associated domain
TAR	trans-activation response element
tat	Trans-activator of transcription
TF	transcription factor
TLR-3	Toll-like receptor 3
TNFα	tumor necrosis factor α
TNPO3	transportin 3
TRIS	tris(hydroxymethyl)aminomethane
TSS	transcription start site
vDNA	viral DNA
V	voltage
Vif	Viral infectivity factor
Vpr	Viral protein R
Vpu	Viral protein U
WAPL	wings apart-like protein homolog
WT	wild type
YY1	Ying Yang 1
ZF	Zinc finger

6.2 Primer list

Primer	Sequence
Alu1	TCCCAGCTACTGGGGAGGCTGAGG
LM667	ATGCCACGTAAGCGAAACTCTGGCTAACTAGGGAACCCA CTG
Alu2 λT	ATGCCACGTAAGCGAAACT

LR	TCCACACTGACTAAAAGGGTCTGA
Tag man	6-FAM- TGTGACTCTGGTAACTAG
B13 forward	CCCCAGGGAGTAGGTTGTGA
B13 reverse	TGTTATTTGAGAAAAGCCCAA
Tag man	6-FAM- CAGCAGGAAAGGAC
HIV-1 gag forward	ACATCAAGCAGCCATGCAAAA
HIV-1 gag reverse	TGGATGCAATCTATCCCATTCTG
Taq man	6-FAM-AAGAGACCATCAATGAGGAA
RPP30 forward	GATTTGGACCTGCGAGCG
RPP30 reverse	GCGGCTGTCTCCACAAGT
RPP30 FAM-BHQ probe	CTGACCTGAAGGCTCT
ddPCR HIV-1 Gag reverse	TGCTTGATGTCCCCCCACT
ddPCR HIV-1 Gag forward	CATGTTTTCAGCATTATCAGAAGGA
HIV-1 gag Probe FAM-BHQ	CCACCCCACAAGATTTAAACACCATGCTAA
CTCF ChIP primer forward	CCCATCTTGCTGACCTCAC
CTCF ChIP primer reverse	AGACCTGGGACGTTTCTGTG
ATAC i5_N501	AATGATACGGCGACCACCGAGATCTACACTAGATCGCTC GTCGGCAGCGTCAGATGTG
ATAC i5_N502	AATGATACGGCGACCACCGAGATCTACACCTCTCTATTCG TCGGCAGCGTCAGATGTG
ATAC i5_N503	AATGATACGGCGACCACCGAGATCTACACTATCCTCTTCG TCGGCAGCGTCAGATGTG
ATAC i5_N504	AATGATACGGCGACCACCGAGATCTACACAGAGTAGATC GTCGGCAGCGTCAGATGTG
ATAC i5_N505	AATGATACGGCGACCACCGAGATCTACACGTAAGGAGTC GTCGGCAGCGTCAGATGTG
ATAC i5_N506	AATGATACGGCGACCACCGAGATCTACACACTGCATATCG TCGGCAGCGTCAGATGTG

ATAC i5_N507	AATGATACGGCGACCACCGAGATCTACACAAGGAGTATC GTCGGCAGCGTCAGATGTG
ATAC i5_N508	AATGATACGGCGACCACCGAGATCTACACCTAAGCCTTC GTCGGCAGCGTCAGATGTG
ATAC i5_N510	AATGATACGGCGACCACCGAGATCTACACCGTCTAATTCG TCGGCAGCGTCAGATGTG
ATAC i5_N511	AATGATACGGCGACCACCGAGATCTACACTCTCTCCGTC GTCGGCAGCGTCAGATGTG
ATAC i5_N513	AATGATACGGCGACCACCGAGATCTACACTCGACTAGTC GTCGGCAGCGTCAGATGTG
ATAC i5_N515	AATGATACGGCGACCACCGAGATCTACACTTCTAGCTTCG TCGGCAGCGTCAGATGTG
ATAC i7_N701	CAAGCAGAAGACGGCATACGAGATTCGCCTTAGTCTCGT GGGCTCGGAGATGT
ATAC i7_N702	CAAGCAGAAGACGGCATACGAGATCTAGTACGGTCTCGT GGGCTCGGAGATGT
ATAC i7_N703	CAAGCAGAAGACGGCATACGAGATTTCTGCCTGTCTCGT GGGCTCGGAGATGT
ATAC i7_N704	CAAGCAGAAGACGGCATACGAGATGCTCAGGAGTCTCGT GGGCTCGGAGATGT
ATAC i7_N705	CAAGCAGAAGACGGCATACGAGATAGGAGTCCGTCTCGT GGGCTCGGAGATGT
ATAC i7_N706	CAAGCAGAAGACGGCATACGAGATCATGCCTAGTCTCGT GGGCTCGGAGATGT
ATAC i7_N707	CAAGCAGAAGACGGCATACGAGATGTAGAGAGGTCTCGT GGGCTCGGAGATGT
ATAC i7_N708	CAAGCAGAAGACGGCATACGAGATCCTCTCTGGTCTCGT GGGCTCGGAGATGT
ATAC i7_N709	CAAGCAGAAGACGGCATACGAGATAGCGTAGCGTCTCGT GGGCTCGGAGATGT
ATAC i7_N710	CAAGCAGAAGACGGCATACGAGATCAGCCTCGGTCTCGT GGGCTCGGAGATGT
ATAC i7_N711	CAAGCAGAAGACGGCATACGAGATTGCCTCTTGTCTCGT GGGCTCGGAGATGT

ATAC i7_N712	CAAGCAGAAGACGGCATACGAGATTTCCTCTACGTCTCG TGGGCTCGGAGATGT
short linker	GTCCCTTAAGCGGAG[SpcC3] NH2
long linker	GTAATACGACTCACTATAGGGCCTCCGCTTAAGGGACT
1st LM PCR	
linker specific primer	CAAGCAGAAGACGGCATACGAGAT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT GTAATACGACTCACTATAGGGC
First round LTR 3	TGTGACTCTGGTAACTAGAGATCCCTC
Sec round LTR3 i1	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTAC ACGACGCTCTTCCGATCT CGTGATGAGATCCCTCAGACCCTTTTAGTCAG
Sec round LTR3 i2	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTAC ACGACGCTCTTCCGATCTACATCGGAGATCCCTCAGACC CTTTTAGTCAG
Sec round LTR3 i3	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTAC ACGACGCTCTTCCGATCTGCCTAAGAGATCCCTCAGACC CTTTTAGTCAG
Sec round LTR3 i4	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTAC ACGACGCTCTTCCGATCTTGGTCAGAGATCCCTCAGACC CTTTTAGTCAG
Sec round LTR3 i5	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTAC ACGACGCTCTTCCGATCTCACTGTGAGATCCCTCAGACC CTTTTAGTCAG
Sec round LTR3 i6	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTAC ACGACGCTCTTCCGATCTATTGGCGAGATCCCTCAGACC CTTTTAGTCAG
Sec round LTR3 i7	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTAC ACGACGCTCTTCCGATCTGATCTGGAGATCCCTCAGACC CTTTTAGTCAG
Sec round LTR3 i8	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTAC ACGACGCTCTTCCGATCTTCAAGTGAGATCCCTCAGACC CTTTTAGTCAG

Sec round LTR3 i9	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTAC ACGACGCTCTTCCGATCTCTGATCGAGATCCCTCAGACC CTTTTAGTCAG
Sec round LTR3 i10	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTAC ACGACGCTCTTCCGATCTAAGCTAGAGATCCCTCAGACC CTTTTAGTCAG
Sec round LTR3 i11	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTAC ACGACGCTCTTCCGATCTGTAGCCGAGATCCCTCAGACC CTTTTAGTCAG
Sec round LTR3 i12	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTAC ACGACGCTCTTCCGATCTTACAAGGAGATCCCTCAGACC CTTTTAGTCAG

2nd LM PCR

Linker i701	CAAGCAGAAGACGGCATACGAGAT CGAGTAAT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT GTAATACGACTCACTATAGGGC
Linker i702	CAAGCAGAAGACGGCATACGAGAT TCTCCGGA GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT GTAATACGACTCACTATAGGGC
Linker i703	CAAGCAGAAGACGGCATACGAGAT AATGAGCG GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT GTAATACGACTCACTATAGGGC
Linker i704	CAAGCAGAAGACGGCATACGAGAT GGAATCTC GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT GTAATACGACTCACTATAGGGC
Linker i705	CAAGCAGAAGACGGCATACGAGAT TTCTGAAT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT GTAATACGACTCACTATAGGGC
Linker i706	CAAGCAGAAGACGGCATACGAGAT ACGAATTC GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT GTAATACGACTCACTATAGGGC
sec 3LTR 501	AATGATACGGCGACCACCGAGATCTACAC TATAGCCT ACACTCTTTCCCTACACGACGCTCTTCCGATCT GAGATCCCTCAGACCCTTTTAGTCAG

sec 3LTR 502	AATGATACGGCGACCACCGAGATCTACAC ATAGAGGC ACACTCTTTCCCTACACGACGCTCTTCCGATCT GAGATCCCTCAGACCCTTTTAGTCAG
sec 3LTR 503	AATGATACGGCGACCACCGAGATCTACAC CCTATCCT ACACTCTTTCCCTACACGACGCTCTTCCGATCTGAGATCC CTCAGACCCTTTTAGTCAG
sec 3LTR 504	AATGATACGGCGACCACCGAGATCTACAC GGCTCTGA ACACTCTTTCCCTACACGACGCTCTTCCGATCT GAGATCCCTCAGACCCTTTTAGTCAG
sec 3LTR 505	AATGATACGGCGACCACCGAGATCTACAC AGGCGAAG ACACTCTTTCCCTACACGACGCTCTTCCGATCT GAGATCCCTCAGACCCTTTTAGTCAG
sec 3LTR 506	AATGATACGGCGACCACCGAGATCTACAC TAATCTTA ACACTCTTTCCCTACACGACGCTCTTCCGATCT GAGATCCCTCAGACCCTTTTAGTCAG
First linker specific primer	GTAATACGACTCACTATAGGGC

6.3 Publications

Rheinberger M*, Costa AL*, Kampmann M, Glavas D, Shytaj IL, Sreeram S, Penzo C, Tibroni N, Garcia-Mesa Y, Leskov K, Fackler OT, Vlahovicek K, Karn J, Lucic B, Herrmann C, Lusic M. *Genomic profiling of HIV-1 integration in microglia cells links viral integration to the topologically associated domains.* Cell Rep. 2023 Feb 13;42(2):112110. doi: 10.1016/j.celrep.2023.112110.

* denotes shared first authors

7 References

Aamer, H.A., McClure, J., Ko, D., Maenza, J., Collier, A.C., Coombs, R.W., Mullins, J.I., and Frenkel, L.M. (2020). Cells producing residual viremia during antiretroviral treatment appear to contribute to rebound viremia following interruption of treatment. PLoS Pathog. *16*, e1008791.

Abascal, F., Acosta, R., Addleman, N.J., Adrian, J., Afzal, V., Ai, R., Aken, B., Akiyama, J.A., Jammal, O. Al, Amrhein, H., et al. (2020). Expanded encyclopaedias of DNA elements in the human and mouse genomes. Nature *583*, 699–710.

Achuthan, V., Perreira, J.M., Sowd, G.A., Puray-Chavez, M., McDougall, W.M., Paulucci-Holthauzen, A., Wu, X., Fadel, H.J., Poeschla, E.M., Multani, A.S., et al. (2018). Capsid-CPSF6 Interaction Licenses Nuclear HIV-1 Trafficking to Sites of Viral DNA Integration. Cell Host Microbe *24*, 392-404.e8.

Acke, A., Van Belle, S., Louis, B., Vitale, R., Rocha, S., Voet, T., Debyser, Z., and Hofkens, J. (2022). Expansion microscopy allows high resolution single cell analysis of epigenetic readers. Nucleic Acids Res. *50*, e100–e100.

Agarwal, H., Reisser, M., Wortmann, C., and Gebhardt, J.C.M. (2017). Direct Observation of Cell-Cycle-Dependent Interactions between CTCF and Chromatin. Biophys. J. *112*, 2051–2055.

Agosto, L.M., Herring, M.B., Mothes, W., and Henderson, A.J. (2018). HIV-1-Infected CD4+ T Cells Facilitate Latent Infection of Resting CD4+ T Cells through Cell-Cell Contact. Cell Rep. *24*, 2088–2100.

Alberts, Bruce; Heald, Rebecca; Johnson, A. (2022). Molecular Biology of the cell.

Albright, A. V, Shieh, J.T., O'Connor, M.J., and González-Scarano, F. (2000). Characterization of cultured microglia that can be infected by HIV-1. J. Neurovirol. *6 Suppl 1*, S53-60.

Ali, H., Mano, M., Braga, L., Naseem, A., Marini, B., Vu, D.M., Collesi, C., Meroni, G., Lusic, M., and Giacca, M. (2019). Cellular TRIM33 restrains HIV-1 infection by targeting viral integrase for proteasomal degradation. Nat. Commun. *10*, 1–15.

Alkhatib, G., Combadiere, C., Broder, C.C., Feng, Y., Kennedy, P.E., Murphy, P.M., and Berger, E.A. (1996). CC CKR5: a RANTES, MIP-1alpha, MIP-1beta receptor as a fusion cofactor for macrophage-tropic HIV-1. Science *272*, 1955–1958.

Allis, C.D., and Jenuwein, T. (2016). The molecular hallmarks of epigenetic control. Nat. Rev. Genet. *17*, 487–500.

Alvarez-Carbonell, D., Garcia-Mesa, Y., Milne, S., Das, B., Dobrowolski, C., Rojas, R., and Karn, J. (2017). Toll-like receptor 3 activation selectively reverses HIV latency in microglial cells. Retrovirology *14*, 9.

Alvarez-Carbonell, D., Ye, F., Ramanath, N., Dobrowolski, C., and Karn, J. (2019). The Glucocorticoid Receptor Is a Critical Regulator of HIV Latency in Human Microglial Cells. J Neuroimmune Pharmacol *14*, 94–109.

Alvarez-Carbonell, D., Ye, F., Ramanath, N., Garcia-Mesa, Y., Knapp, P.E., Hauser, K.F., and Karn, J. (2020). Cross-talk between microglia and neurons regulates HIV latency. PLOS Pathog. *15*, e1008249.

Amelio, A.L., McAnany, P.K., and Bloom, D.C. (2006). A chromatin insulator-like element in the herpes simplex virus type 1 latency-associated transcript region binds CCCTC-binding factor and displays enhancer-blocking and silencing activities. J. Virol. *80*, 2358–2368.

Anderson, A.M., Muñoz-Moreno, J.A., McClernon, D.R., Ellis, R.J., Cookson, D., Clifford, D.B., Collier, A.C., Gelman, B.B., Marra, C.M., McArthur, J.C., et al. (2017). Prevalence and Correlates of Persistent HIV-1 RNA in Cerebrospinal Fluid During Antiretroviral Therapy. J. Infect. Dis. *215*, 105–113.

Antoni, G., Guergnon, J., Meaudre, C., Samri, A., Boufassa, F., Goujard, C., Lambotte, O., Autran, B., Rouzioux, C., Costagliola, D., et al. (2013). MHC-driven HIV-1 control on the long run is not systematically determined at early times post-HIV-1 infection. AIDS *27*, 1707–1716.

Aprile-Garcia, F., Tomar, P., Hummel, B., Khavaran, A., and Sawarkar, R. (2019). Nascent-protein ubiquitination is required for heat shock-induced gene downregulation in human cells. Nat. Struct. Mol. Biol. *26*, 137–146.

Archin, N.M., Espeseth, A., Parker, D., Cheema, M., Hazuda, D., and Margolis, D.M. (2009). Expression of latent HIV induced by the potent HDAC inhibitor suberoylanilide hydroxamic acid. AIDS Res. Hum. Retroviruses *25*, 207–212.

Archin, N.M., Liberty, A.L., Kashuba, A.D., Choudhary, S.K., Kuruc, J.D., Crooks, A.M., Parker, D.C., Anderson, E.M., Kearney, M.F., Strain, M.C., et al. (2012). Administration of vorinostat disrupts HIV-1 latency in patients on antiretroviral therapy. Nature *487*, 482.

Archin, N.M., Bateson, R., Tripathy, M.K., Crooks, A.M., Yang, K.-H., Dahl, N.P., Kearney, M.F., Anderson, E.M., Coffin, J.M., Strain, M.C., et al. (2014). HIV-1 expression within resting CD4+ T cells after multiple doses of vorinostat. J. Infect. Dis.

210, 728–735.

Archin, N.M., Kirchherr, J.L., Sung, J.A.M., Clutton, G., Sholtis, K., Xu, Y., Allard, B., Stuelke, E., Kashuba, A.D., Kuruc, J.D., et al. (2017). Interval dosing with the HDAC inhibitor vorinostat effectively reverses HIV latency. J. Clin. Invest. *127*, 3126–3135.

Arvey, A., Tempera, I., Tsai, K., Chen, H.-S., Tikhmyanova, N., Klichinsky, M., Leslie, C., and Lieberman, P.M. (2012). An atlas of the Epstein-Barr virus transcriptome and epigenome reveals host-virus regulatory interactions. Cell Host Microbe *12*, 233–245.

Arzate-Mejía, R.G., Recillas-Targa, F., and Corces, V.G. (2018). Developing in 3D: the role of CTCF in cell differentiation. Development *145*.

Avalos, C.R., Abreu, C.M., Queen, S.E., Li, M., Price, S., Shirk, E.N., Engle, E.L., Forsyth, E., Bullock, B.T., Mac Gabhann, F., et al. (2017). Brain Macrophages in Simian Immunodeficiency Virus-Infected, Antiretroviral-Suppressed Macaques: a Functional Latent Reservoir. MBio *8*.

Bajrami, E., and Spiroski, M. (2016). Genomic Imprinting. Open Access Maced. J. Med. Sci. *4*, 181–184.

Ballandras-Colas, A., Maskell, D.P., Serrao, E., Locke, J., Swuec, P., Jonsson, S.R., Kotecha, A., Cook, N.J., Pye, V.E., Taylor, I.A., et al. (2017). A supramolecular assembly mediates lentiviral DNA integration. Science (80-.). *355*, 93–95.

Baltimore, D. (1971). Expression of animal virus genomes. Bacteriol. Rev. *35*, 235–241.

BALTIMORE, D. (1970). Viral RNA-dependent DNA Polymerase: RNA-dependent DNA Polymerase in Virions of RNA Tumour Viruses. Nature *226*, 1209–1211.

Banigan, E.J., van den Berg, A.A., Brandão, H.B., Marko, J.F., and Mirny, L.A. (2020). Chromosome organization by one-sided and two-sided loop extrusion. Elife *9*, e53558.

Bannister, A.J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. Cell Res. *21*, 381–395.

Barr, S.D., Ciuffi, A., Leipzig, J., Shinn, P., Ecker, J.R., and Bushman, F.D. (2006). HIV Integration Site Selection: Targeting in Macrophages and the Effects of Different Routes of Viral Entry. Mol. Ther. *14*, 218–225.

Barré-Sinoussi, F., Chermann, J.C., Rey, F., Nugeyre, M.T., Chamaret, S., Gruest, J., Dauguet, C., Axler-Blin, C., Vézinet-Brun, F., Rouzioux, C., et al. (1983). Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science *220*, 868–871.

Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Schones, D.E., Wang, Z., Wei, G.,

Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823–837.

Barton, K.M., Archin, N.M., Keedy, K.S., Espeseth, A.S., Zhang, Y., Gale, J., Wagner, F.F., Holson, E.B., and Margolis, D.M. (2014). Selective HDAC inhibition for the disruption of latent HIV-1 infection. PLoS One *9*, e102684.

Bartosovic, M., Kabbe, M., and Castelo-Branco, G. (2021). Single-cell CUT&Tag profiles histone modifications and transcription factors in complex tissues. Nat. Biotechnol. *39*, 825–835.

Battivelli, E., and Verdin, E. (2018). HIVGKO: A Tool to Assess HIV-1 Latency Reversal Agents in Human Primary CD4+ T Cells. Bio-Protocol *8*, e3050.

Battivelli, E., Dahabieh, M.S., Abdel-Mohsen, M., Svensson, J.P., Tojal Da Silva, I., Cohn, L.B., Gramatica, A., Deeks, S., Greene, W.C., Pillai, S.K., et al. (2018). Distinct chromatin functional states correlate with HIV latency reactivation in infected primary CD4+ T cells. Elife *7*, e34655.

Beagrie, R.A., Scialdone, A., Schueler, M., Kraemer, D.C.A., Chotalia, M., Xie, S.Q., Barbieri, M., de Santiago, I., Lavitas, L.-M., Branco, M.R., et al. (2017). Complex multi-enhancer contacts captured by genome architecture mapping. Nature *543*, 519–524.

Bedwell, G.J., Jang, S., Li, W., Singh, P.K., and Engelman, A.N. (2021). rigrag : high-resolution mapping of genic targeting preferences during HIV-1 integration in vitro and in vivo. *75*, 1–17.

Bejarano, D.A., Puertas, M.C., Börner, K., Martinez-Picado, J., Müller, B., and Kräusslich, H.-G. (2018). Detailed Characterization of Early HIV-1 Replication Dynamics in Primary Human Macrophages. Viruses *10*.

Bell, A.C., West, A.G., and Felsenfeld, G. (1999). The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. Cell *98*, 387–396.

Bellefroid, M., Rodari, A., Galais, M., Krijger, P.H.L., Tjalsma, S.J.D., Nestola, L., Plant, E., Vos, E.S.M., Cristinelli, S., Van Driessche, B., et al. (2022). Role of the cellular factor CTCF in the regulation of bovine leukemia virus latency and three-dimensional chromatin organization. Nucleic Acids Res. *50*, 3190–3202.

Benkirane, M., Chun, R.F., Xiao, H., Ogryzko, V. V, Howard, B.H., Nakatani, Y., and Jeang, K.T. (1998). Activation of integrated provirus requires histone acetyltransferase. p300 and P/CAF are coactivators for HIV-1 Tat. J. Biol. Chem. *273*, 24898–24905.

Bentsen, M., Goymann, P., Schultheis, H., Klee, K., Petrova, A., Wiegandt, R., Fust,

A., Preussner, J., Kuenne, C., Braun, T., et al. (2020). ATAC-seq footprinting unravels kinetics of transcription factor binding during zygotic genome activation. Nat. Commun. *11*.

Berndsen, C.E., and Denu, J.M. (2008). Catalysis and substrate selection by histone/protein lysine acetyltransferases. Curr. Opin. Struct. Biol. *18*, 682–689.

Von Bernhardi, R., Cornejo, F., Parada, G., and Eugenin, J. (2015). Role of TGFβ signaling in the pathogenesis of Alzheimer's disease . Front. Cell. Neurosci. *9*.

Bernstein, B.E., Humphrey, E.L., Erlich, R.L., Schneider, R., Bouman, P., Liu, J.S., Kouzarides, T., and Schreiber, S.L. (2002). Methylation of histone H3 Lys 4 in coding regions of active genes. Proc. Natl. Acad. Sci. U. S. A. *99*, 8695–8700.

Bhattacharya, S., Wang, S., Reddy, D., Shen, S., Zhang, Y., Zhang, N., Li, H., Washburn, M.P., Florens, L., Shi, Y., et al. (2021). Structural basis of the interaction between SETD2 methyltransferase and hnRNP L paralogs for governing co-transcriptional splicing. Nat. Commun. *12*, 6452.

Bilimoria, P.M., and Stevens, B. (2015). Microglia function during brain development: New insights from animal models. Brain Res *1617*, 7–17.

Bintu, B., Mateo, L.J., Su, J.-H., Sinnott-Armstrong, N.A., Parker, M., Kinrot, S., Yamaya, K., Boettiger, A.N., and Zhuang, X. (2018). Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. Science *362*.

Blazkova, J., Trejbalova, K., Gondois-Rey, F., Halfon, P., Philibert, P., Guiguen, A., Verdin, E., Olive, D., Van Lint, C., Hejnar, J., et al. (2009). CpG methylation controls reactivation of HIV from latency. PLoS Pathog *5*, e1000554.

Bleuyard, J.-Y., Fournier, M., Nakato, R., Couturier, A.M., Katou, Y., Ralf, C., Hester, S.S., Dominguez, D., Rhodes, D., Humphrey, T.C., et al. (2017). MRG15-mediated tethering of PALB2 to unperturbed chromatin protects active genes from genotoxic stress. Proc. Natl. Acad. Sci. U. S. A. *114*, 7671–7676.

Blobel, G.A., Higgs, D.R., Mitchell, J.A., Notani, D., and Young, R.A. (2021). Testing the super-enhancer concept. Nat. Rev. Genet. *22*, 749–755.

Boettiger, A.N., Bintu, B., Moffitt, J.R., Wang, S., Beliveau, B.J., Fudenberg, G., Imakaev, M., Mirny, L.A., Wu, C., and Zhuang, X. (2016). Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. Nature *529*, 418–422.

Boija, A., Klein, I.A., Sabari, B.R., Dall'Agnese, A., Coffey, E.L., Zamudio, A. V, Li, C.H., Shrinivas, K., Manteiga, J.C., Hannett, N.M., et al. (2018). Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains.

Cell *175*, 1842-1855.e16.

Borun, T.W., Pearson, D., and Paik, W.K. (1972). Studies of histone methylation during the HeLa S-3 cell cycle. J. Biol. Chem. *247*, 4288–4298.

Böttcher, C., Schlickeiser, S., Sneeboer, M.A.M., Kunkel, D., Knop, A., Paza, E., Fidzinski, P., Kraus, L., Snijders, G.J.L., Kahn, R.S., et al. (2019). Human microglia regional heterogeneity and phenotypes determined by multiplexed single-cell mass cytometry. Nat. Neurosci. *22*, 78–90.

Bowerman, B., Brown, P.O., Bishop, J.M., and Varmus, H.E. (1989). A nucleoprotein complex mediates the integration of retroviral DNA. Genes Dev. *3*, 469–478.

Boxus, M., and Willems, L. (2009). Mechanisms of HTLV-1 persistence and transformation. Br. J. Cancer *101*, 1497–1501.

Boyle, A.P., Davis, S., Shulha, H.P., Meltzer, P., Margulies, E.H., Weng, Z., Furey, T.S., and Crawford, G.E. (2008). High-resolution mapping and characterization of open chromatin across the genome. Cell *132*, 311–322.

Bradner, J.E., Hnisz, D., and Young, R.A. (2017). Transcriptional Addiction in Cancer. Cell *168*, 629–643.

Brady, T., Lee, Y.N., Ronen, K., Malani, N., Berry, C.C., Bieniasz, P.D., and Bushman, F.D. (2009a). Integration target site selection by a resurrected human endogenous retrovirus. Genes Dev *23*, 633–642.

Brady, T., Agosto, L.M., Malani, N., Berry, C.C., O'Doherty, U., and Bushman, F. (2009b). HIV integration site distributions in resting and activated CD4+ T cells infected in culture. Aids *23*, 1461–1471.

Brass, A.L., Dykxhoorn, D.M., Benita, Y., Yan, N., Engelman, A., Xavier, R.J., Lieberman, J., and Elledge, S.J. (2008). Identification of host proteins required for HIV infection through a functional genomic screen. Science (80-.). *319*, 921–926.

Brenchley, J.M., Schacker, T.W., Ruff, L.E., Price, D.A., Taylor, J.H., Beilman, G.J., Nguyen, P.L., Khoruts, A., Larson, M., Haase, A.T., et al. (2004). CD4+ T cell depletion during all stages of HIV disease occurs predominantly in the gastrointestinal tract. J. Exp. Med. *200*, 749–759.

Broder, S. (2010). The development of antiretroviral therapy and its impact on the HIV-1/AIDS pandemic. Antiviral Res. *85*, 1–18.

Brown, P.O., Bowerman, B., Varmus, H.E., and Bishop, J.M. (1987). Correct integration of retroviral DNA in vitro. Cell *49*, 347–356.

Brownell, J.E., Zhou, J., Ranalli, T., Kobayashi, R., Edmondson, D.G., Roth, S.Y., and

Allis, C.D. (1996). Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. Cell *84*, 843–851.

Bruner, K.M., Murray, A.J., Pollack, R.A., Soliman, M.G., Laskey, S.B., Capoferri, A.A., Lai, J., Strain, M.C., Lada, S.M., Hoh, R., et al. (2016). Defective proviruses rapidly accumulate during acute HIV-1 infection. Nat. Med. *22*, 1043–1049.

Buenrostro, J.D., Wu, B., Chang, H.Y., and Greenleaf, W.J. (2015). ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. Curr. Protoc. Mol. Biol. *109*, 21.29.1-21.29.9.

Bui, J.K., Sobolewski, M.D., Keele, B.F., Spindler, J., Musick, A., Wiegand, A., Luke, B.T., Shao, W., Hughes, S.H., Coffin, J.M., et al. (2017). Proviruses with identical sequences comprise a large fraction of the replication-competent HIV reservoir. PLOS Pathog. *13*, e1006283.

Bullen, C.K., Laird, G.M., Durand, C.M., Siliciano, J.D., and Siliciano, R.F. (2014). New ex vivo approaches distinguish effective and ineffective single agents for reversing HIV-1 latency in vivo. Nat. Med. *20*, 425.

Burdick, R.C., Deleage, C., Duchon, A., Estes, J.D., and Hu, W. Intranuclear Positions of HIV-1 Proviruses Are Dynamic and Do Not Correlate with Transcriptional Activity.

Burdick, R.C., Li, C., Munshi, M., Rawson, J.M.O., Nagashima, K., Hu, W.-S., and Pathak, V.K. (2020). HIV-1 uncoats in the nucleus near sites of integration. Proc. Natl. Acad. Sci. U. S. A. *117*, 5486–5493.

Bushman, F.D., Fujiwara, T., and Craigie, R. (1990). Retroviral DNA integration directed by HIV integration protein in vitro. Science *249*, 1555–1558.

Bushman, F.D., Malani, N., Fernandes, J., D'Orso, I., Cagney, G., Diamond, T.L., Zhou, H., Hazuda, D.J., Espeseth, A.S., Konig, R., et al. (2009). Host cell factors in HIV replication: meta-analysis of genome-wide studies. PLoS Pathog *5*, e1000437.

Busschots, K., Vercammen, J., Emiliani, S., Benarous, R., Engelborghs, Y., Christ, F., and Debyser, Z. (2005). The interaction of LEDGF/p75 with integrase is lentivirus-specific and promotes DNA binding. J Biol Chem *280*, 17841–17847.

Byvoet, P., Shepherd, G.R., Hardin, J.M., and Noland, B.J. (1972). The distribution and turnover of labeled methyl groups in histone fractions of cultured mammalian cells. Arch. Biochem. Biophys. *148*, 558–567.

Cameron, P.U., Saleh, S., Sallmann, G., Solomon, A., Wightman, F., Evans, V.A., Boucher, G., Haddad, E.K., Sekaly, R.-P., Harman, A.N., et al. (2010). 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. *107*, 16934–16939.

Cao, C., Hong, P., Huang, X., Lin, D., Cao, G., Wang, L., Feng, B., Wu, P., Shen, H., Xu, Q., et al. (2020). HPV-CCDC106 integration alters local chromosome architecture and hijacks an enhancer by three-dimensional genome structure remodeling in cervical cancer. J. Genet. Genomics *47*, 437–450.

Catalfamo, M., Le Saout, C., and Lane, H.C. (2012). The role of cytokines in the pathogenesis and treatment of HIV infection. Cytokine Growth Factor Rev. *23*, 207–214.

Cattoni, D.I., Cardozo Gizzi, A.M., Georgieva, M., Di Stefano, M., Valeri, A., Chamousset, D., Houbron, C., Déjardin, S., Fiche, J.-B., González, I., et al. (2017). Single-cell absolute contact probability detection reveals chromosomes are organized by multiple low-frequency yet specific interactions. Nat. Commun. *8*, 1753.

Caudron-Herger, M., Rusin, S.F., Adamo, M.E., Seiler, J., Schmid, V.K., Barreau, E., Kettenbach, A.N., and Diederichs, S. (2019). R-DeeP: Proteome-wide and Quantitative Identification of RNA-Dependent Proteins by Density Gradient Ultracentrifugation. Mol. Cell *75*, 184-199.e10.

Cereseto, A., Manganaro, L., Gutierrez, M.I., Terreni, M., Fittipaldi, A., Lusic, M., Marcello, A., and Giacca, M. (2005). Acetylation of HIV-1 integrase by p300 regulates viral integration. EMBO J. *24*, 3070–3081.

Cesana, D., Santoni de Sio, F.R., Rudilosso, L., Gallina, P., Calabria, A., Beretta, S., Merelli, I., Bruzzesi, E., Passerini, L., Nozza, S., et al. (2017). HIV-1-mediated insertional activation of STAT5B and BACH2 trigger viral reservoir in T regulatory cells. Nat. Commun. *8*, 498.

Chai, Q., Jovasevic, V., Malikov, V., Sabo, Y., Morham, S., Walsh, D., and Naghavi, M.H. (2017). HIV-1 counteracts an innate restriction by amyloid precursor protein resulting in neurodegeneration. Nat. Commun. *8*, 1522.

Chang, L.H., Ghosh, S., and Noordermeer, D. (2020). TADs and Their Borders: Free Movement or Building a Wall? J. Mol. Biol. *432*, 643–652.

Chang, L.H., Ghosh, S., Papale, A., Luppino, J.M., Miranda, M., Piras, V., Degrouard, J., Edouard, J., Poncelet, M., Lecouvreur, N., et al. (2023). Multi-feature clustering of CTCF binding creates robustness for loop extrusion blocking and Topologically Associating Domain boundaries. Nat. Commun. *14*.

Chao, W., Huynh, K.D., Spencer, R.J., Davidow, L.S., and Lee, J.T. (2002). CTCF, a candidate trans-acting factor for X-inactivation choice. Science *295*, 345–347.

Chau, C.M., Zhang, X.-Y., McMahon, S.B., and Lieberman, P.M. (2006). Regulation of Epstein-Barr virus latency type by the chromatin boundary factor CTCF. J. Virol. *80*, 5723–5732.

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

Checroune, F., Yao, X.J., Göttlinger, H.G., Bergeron, D., and Cohen, E.A. (1995). Incorporation of Vpr into human immunodeficiency virus type 1: role of conserved regions within the P6 domain of Pr55gag. J. Acquir. Immune Defic. Syndr. Hum. Retrovirology Off. Publ. Int. Retrovirology Assoc. *10*, 1–7.

Chen, C.-Y., Chang, I.-S., Hsiung, C.A., and Wasserman, W.W. (2014). On the identification of potential regulatory variants within genome wide association candidate SNP sets. BMC Med. Genomics *7*, 34.

Chen, H.-S., Wikramasinghe, P., Showe, L., and Lieberman, P.M. (2012a). Cohesins repress Kaposi's sarcoma-associated herpesvirus immediate early gene transcription during latency. J. Virol. *86*, 9454–9464.

Chen, H., Tian, Y., Shu, W., Bo, X., and Wang, S. (2012b). Comprehensive identification and annotation of cell type-specific and ubiquitous CTCF-binding sites in the human genome. PLoS One *7*, e41374.

Chen, H.C., Martinez, J.P., Zorita, E., Meyerhans, A., and Filion, G.J. (2017). Position effects influence HIV latency reversal. Nat Struct Mol Biol *24*, 47–54.

Chen, J.C., Krucinski, J., Miercke, L.J., Finer-Moore, J.S., Tang, A.H., Leavitt, A.D., and Stroud, R.M. (2000). Crystal structure of the HIV-1 integrase catalytic core and C-terminal domains: a model for viral DNA binding. Proc. Natl. Acad. Sci. U. S. A. *97*, 8233–8238.

Chen, Q., Lin, L., Smith, S., Huang, J., Berger, S.L., and Zhou, J. (2007). CTCF-dependent chromatin boundary element between the latency-associated transcript and ICP0 promoters in the herpes simplex virus type 1 genome. J. Virol. *81*, 5192–5201.

Chen, S., Luperchio, T.R., Wong, X., Doan, E.B., Byrd, A.T., Roy Choudhury, K., Reddy, K.L., and Krangel, M.S. (2018a). A Lamina-Associated Domain Border Governs Nuclear Lamina Interactions, Transcription, and Recombination of the Tcrb Locus. Cell Rep. *25*, 1729-1740.e6.

Chen, Y., Zhang, Y., Wang, Y., Zhang, L., Brinkman, E.K., Adam, S.A., Goldman, R.,

Van Steensel, B., Ma, J., and Belmont, A.S. (2018b). Mapping 3D genome organization relative to nuclear compartments using TSA-Seq as a cytological ruler. J. Cell Biol. *217*, 4025–4048.

Chéné, I. du, Basyuk, E., Lin, Y.-L., Triboulet, R., Knezevich, A., Chable-Bessia, C., Mettling, C., Baillat, V., Reynes, J., Corbeau, P., et al. (2007). Suv39H1 and HP1γ are responsible for chromatin-mediated HIV-1 transcriptional silencing and post-integration latency. Embo J *26*, 424–435.

Cherepanov, P., Maertens, G., Proost, P., Devreese, B., Van Beeumen, J., Engelborghs, Y., De Clercq, E., and Debyser, Z. (2003). HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells. J Biol Chem *278*, 372–381.

Cherepanov, P., Devroe, E., Silver, P.A., and Engelman, A. (2004). Identification of an evolutionarily conserved domain in human lens epithelium-derived growth factor/transcriptional co-activator p75 (LEDGF/p75) that binds HIV-1 integrase. J Biol Chem *279*, 48883–48892.

Chiang, C.-M. (2009). Brd4 engagement from chromatin targeting to transcriptional regulation: selective contact with acetylated histone H3 and H4. F1000 Biol. Rep. *1*, 98.

Chilunda, V., Calderon, T.M., Martinez-Aguado, P., and Berman, J.W. (2019). The impact of substance abuse on HIV-mediated neuropathogenesis in the current ART era. Brain Res. *1724*, 146426.

Chin, C.R., Perreira, J.M., Savidis, G., Portmann, J.M., Aker, A.M., Feeley, E.M., Smith, M.C., and Brass, A.L. (2015). Direct Visualization of HIV-1 Replication Intermediates shows that Viral Capsid and CPSF6 Modulate HIV-1 Intra-nuclear Invasion and Integration. Cell Rep. *13*, 1717–1731.

Cho, A., Gaebler, C., Olveira, T., Ramos, V., Saad, M., Lorenzi, J.C.C., Gazumyan, A., Moir, S., Caskey, M., Chun, T.-W., et al. (2022). Longitudinal clonal dynamics of HIV-1 latent reservoirs measured by combination quadruplex polymerase chain reaction and sequencing. Proc. Natl. Acad. Sci. U. S. A. *119*.

Choe, H., Farzan, M., Sun, Y., Sullivan, N., Rollins, B., Ponath, P.D., Wu, L., Mackay, C.R., LaRosa, G., Newman, W., et al. (1996). The β-Chemokine Receptors CCR3 and CCR5 Facilitate Infection by Primary HIV-1 Isolates. Cell *85*, 1135–1148.

Chomont, N. (2020). HIV enters deep sleep in people who naturally control the virus. Nature *585*, 190–191.

Chomont, N., El-Far, M., Ancuta, P., Trautmann, L., Procopio, F.A., Yassine-Diab, B., Boucher, G., Boulassel, M.-R., Ghattas, G., Brenchley, J.M., et al. (2009). HIV reservoir size and persistence are driven by T cell survival and homeostatic proliferation. Nat. Med. *15*, 893–900.

Chong, S., Dugast-Darzacq, C., Liu, Z., Dong, P., Dailey, G.M., Cattoglio, C., Heckert, A., Banala, S., Lavis, L., Darzacq, X., et al. (2018). Imaging dynamic and selective low-complexity domain interactions that control gene transcription. Science *361*.

Christ, F., Voet, A., Marchand, A., Nicolet, S., Desimmie, B.A., Marchand, D., Bardiot, D., Van der Veken, N.J., Van Remoortel, B., Strelkov, S. V, et al. (2010). Rational design of small-molecule inhibitors of the LEDGF/p75-integrase interaction and HIV replication. Nat. Chem. Biol. *6*, 442–448.

Chun, T.W., Engel, D., Berrey, M.M., Shea, T., Corey, L., and Fauci, A.S. (1998). Early establishment of a pool of latently infected, resting CD4(+) T cells during primary HIV-1 infection. Proc. Natl. Acad. Sci. U. S. A. *95*, 8869–8873.

Chung, J.H., Whiteley, M., and Felsenfeld, G. (1993). A 5' element of the chicken beta-globin domain serves as an insulator in human erythroid cells and protects against position effect in Drosophila. Cell *74*, 505–514.

Churchill, M.J., Gorry, P.R., Cowley, D., Lal, L., Sonza, S., Purcell, D.F.J., Thompson, K.A., Gabuzda, D., McArthur, J.C., Pardo, C.A., et al. (2006). Use of laser capture microdissection to detect integrated HIV-1 DNA in macrophages and astrocytes from autopsy brain tissues. J. Neurovirol. *12*, 146–152.

Churchill, M.J., Deeks, S.G., Margolis, D.M., Siliciano, R.F., and Swanstrom, R. (2015). HIV reservoirs: What, where and how to target them. Nat. Rev. Microbiol. *14*, 55–60.

Ci, Y., Yang, Y., Xu, C., and Shi, L. (2018). Vesicular stomatitis virus G protein transmembrane region is crucial for the hemi-fusion to full fusion transition. Sci. Rep. *8*, 10669.

Clark, I.C., Mudvari, P., Thaploo, S., Smith, S., Abu-Laban, M., Hamouda, M., Theberge, M., Shah, S., Ko, S.H., Pérez, L., et al. (2023). HIV silencing and cell survival signatures in infected T cell reservoirs. Nature *614*, 318–325.

Clark, S.J., Argelaguet, R., Kapourani, C.-A., Stubbs, T.M., Lee, H.J., Alda-Catalinas, C., Krueger, F., Sanguinetti, G., Kelsey, G., Marioni, J.C., et al. (2018). scNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells. Nat. Commun. *9*, 781.

Clavel, F., Mansinho, K., Chamaret, S., Guetard, D., Favier, V., Nina, J., Santos-

Ferreira, M.O., Champalimaud, J.L., and Montagnier, L. (1987). Human immunodeficiency virus type 2 infection associated with AIDS in West Africa. N. Engl. J. Med. *316*, 1180–1185.

Clerc, I., Polakowski, N., André-Arpin, C., Cook, P., Barbeau, B., Mesnard, J.-M., and Lemasson, I. (2008). An Interaction between the Human T Cell Leukemia Virus Type 1 Basic Leucine Zipper Factor (HBZ) and the KIX Domain of p300/CBP Contributes to the Down-regulation of Tax-dependent Viral Transcription by HBZ*. J. Biol. Chem. *283*, 23903–23913.

Clifford, D.B., and Ances, B.M. (2013). HIV-associated neurocognitive disorder. Lancet. Infect. Dis. *13*, 976–986.

Coffin, J.M., Hughes, S.H., and Varmus, H.E. (1997). Retroviruses (Cold Spring Harbor (NY)).

Coffin, J.M., Bale, M.J., Wells, D., Guo, S., Luke, B., Zerbato, J.M., Sobolewski, M.D., Sia, T., Shao, W., Wu, X., et al. (2021). Integration in oncogenes plays only a minor role in determining the in vivo distribution of HIV integration sites before or during suppressive antiretroviral therapy. PLoS Pathog. *17*, 1–28.

Cohn, L.B., Silva, I.T., Oliveira, T.Y., Rosales, R.A., Parrish, E.H., Learn, G.H., Hahn, B.H., Czartoski, J.L., McElrath, M.J., Lehmann, C., et al. (2015). HIV-1 integration landscape during latent and active infection. Cell *160*, 420–432.

Cole, B., Lambrechts, L., Boyer, Z., Noppe, Y., De Scheerder, M.-A., Eden, J.-S., Vrancken, B., Schlub, T.E., McLaughlin, S., Frenkel, L.M., et al. (2022). Extensive characterization of HIV-1 reservoirs reveals links to plasma viremia before and during analytical treatment interruption. Cell Rep. *39*, 110739.

Colin, L., and Van Lint, C. (2009). Molecular control of HIV-1 postintegration latency: implications for the development of new therapeutic strategies. Retrovirology *6*, 111.

Collins, R.E., Northrop, J.P., Horton, J.R., Lee, D.Y., Zhang, X., Stallcup, M.R., and Cheng, X. (2008). The ankyrin repeats of G9a and GLP histone methyltransferases are mono- and dimethyllysine binding modules. Nat. Struct. Mol. Biol. *15*, 245–250.

Collora, J.A., and Ho, Y.-C. (2023). Integration site-dependent HIV-1 promoter activity shapes host chromatin conformation. Genome Res. gr.277698.123.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science (80-.). *339*, 819–823.

Corces, M.R., Trevino, A.E., Hamilton, E.G., Greenside, P.G., Sinnott-Armstrong, N.A.,

Vesuna, S., Satpathy, A.T., Rubin, A.J., Montine, K.S., Wu, B., et al. (2017). An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. Nat. Methods *14*, 959–962.

Cosenza, M.A., Zhao, M.-L., Si, Q., and Lee, S.C. (2002). Human brain parenchymal microglia express CD14 and CD45 and are productively infected by HIV-1 in HIV-1 encephalitis. Brain Pathol. *12*, 442–455.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc. Natl. Acad. Sci. U. S. A. *107*, 21931–21936.

Cuddapah, S., Jothi, R., Schones, D.E., Roh, T.-Y., Cui, K., and Zhao, K. (2009). Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. Genome Res. *19*, 24–32.

D'Arienzo, V., Ferguson, J., Giraud, G., Chapus, F., Harris, J.M., Wing, P.A.C., Claydon, A., Begum, S., Zhuang, X., Balfe, P., et al. (2021). The CCCTC-binding factor CTCF represses hepatitis B virus enhancer I and regulates viral transcription. Cell. Microbiol. *23*, e13274.

Dahl, V., Peterson, J., Fuchs, D., Gisslen, M., Palmer, S., and Price, R.W. (2014). Low levels of HIV-1 RNA detected in the cerebrospinal fluid after up to 10 years of suppressive therapy are associated with local immune activation. Aids *28*, 2251–2258.

Dalgleish, A.G., Beverley, P.C.L., Clapham, P.R., Crawford, D.H., Greaves, M.F., and Weiss, R.A. (1984). The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. Nature *312*, 763–767.

Daly, T.J., Cook, K.S., Gray, G.S., Maione, T.E., and Rusche, J.R. (1989). Specific binding of HIV-1 recombinant Rev protein to the Rev-responsive element in vitro. Nature *342*, 816–819.

Dann, G.P., Liszczak, G.P., Bagert, J.D., Müller, M.M., Nguyen, U.T.T., Wojcik, F., Brown, Z.Z., Bos, J., Panchenko, T., Pihl, R., et al. (2017). ISWI chromatin remodellers sense nucleosome modifications to determine substrate preference. Nature *548*, 607–611.

Daugaard, M., Baude, A., Fugger, K., Povlsen, L.K., Beck, H., Sorensen, C.S., Petersen, N.H., Sorensen, P.H., Lukas, C., Bartek, J., et al. (2012). LEDGF (p75) promotes DNA-end resection and homologous recombination. Nat Struct Mol Biol *19*, 803–810.

Davalos, D., Grutzendler, J., Yang, G., Kim, J. V, Zuo, Y., Jung, S., Littman, D.R., Dustin, M.L., and Gan, W.B. (2005). ATP mediates rapid microglial response to local brain injury in vivo. Nat Neurosci *8*, 752–758.

Davidson, I.F., Bauer, B., Goetz, D., Tang, W., Wutz, G., and Peters, J.-M. (2019). DNA loop extrusion by human cohesin. Science *366*, 1338–1345.

Davis, L.E., Hjelle, B.L., Miller, V.E., Palmer, D.L., Llewellyn, A.L., Merlin, T.L., Young, S.A., Mills, R.G., Wachsman, W., and Wiley, C.A. (1992). Early viral brain invasion in iatrogenic human immunodeficiency virus infection. Neurology *42*, 1736 LP – 1736.

Debyser, Z., Christ, F., De Rijck, J., and Gijsbers, R. (2015). Host factors for retroviral integration site selection. Trends Biochem Sci *40*, 108–116.

Debyser, Z., Vansant, G., Bruggemans, A., Janssens, J., and Christ, F. (2018). Insight in HIV Integration Site Selection Provides a Block-and-Lock Strategy for a Functional Cure of HIV Infection. Viruses *11*.

Deczkowska, A., Matcovitch-Natan, O., Tsitsou-Kampeli, A., Ben-Hamo, S., Dvir-Szternfeld, R., Spinrad, A., Singer, O., David, E., Winter, D.R., Smith, L.K., et al. (2017). Mef2C restrains microglial inflammatory response and is lost in brain ageing in an IFN-I-dependent manner. Nat. Commun. *8*.

Deeks, S.G., Tracy, R., and Douek, D.C. (2013). Systemic Effects of Inflammation on Health during Chronic HIV Infection. Immunity *39*, 633–645.

Deeks, S.G., Overbaugh, J., Phillips, A., and Buchbinder, S. (2015). HIV infection. Nat. Rev. Dis. Prim. *1*, 15035.

Dekker, J., and Misteli, T. (2015). Long-Range Chromatin Interactions. Cold Spring Harb. Perspect. Biol. *7*, a019356.

Demeulemeester, J., De Rijck, J., Gijsbers, R., and Debyser, Z. (2015). Retroviral integration: Site matters: Mechanisms and consequences of retroviral integration site selection. Bioessays *37*, 1202–1214.

Deng, S., Feng, Y., and Pauklin, S. (2022). 3D chromatin architecture and transcription regulation in cancer. J. Hematol. Oncol. *15*, 49.

Despang, A., Schöpflin, R., Franke, M., Ali, S., Jerković, I., Paliou, C., Chan, W.-L., Timmermann, B., Wittler, L., Vingron, M., et al. (2019). Functional dissection of the Sox9-Kcnj2 locus identifies nonessential and instructive roles of TAD architecture. Nat. Genet. *51*, 1263–1271.

Dharan, A., Bachmann, N., Talley, S., Zwikelmaier, V., and Campbell, E.M. (2020). Nuclear pore blockade reveals that HIV-1 completes reverse transcription and

uncoating in the nucleus. Nat. Microbiol. *5*, 1088–1095.

Dixit, U., Bhutoria, S., Wu, X., Qiu, L., Spira, M., Mathew, S., Harris, R., Adams, L.J., Cahill, S., Pathak, R., et al. (2021). INI1/SMARCB1 Rpt1 domain mimics TAR RNA in binding to integrase to facilitate HIV-1 replication. Nat. Commun. *12*, 2743.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376–380.

Dixon, J.R., Gorkin, D.U., and Ren, B. (2016). Chromatin Domains: The Unit of Chromosome Organization. Mol. Cell *62*, 668–680.

Dogan, N., Wu, W., Morrissey, C.S., Chen, K.-B., Stonestrom, A., Long, M., Keller, C.A., Cheng, Y., Jain, D., Visel, A., et al. (2015). Occupancy by key transcription factors is a more accurate predictor of enhancer activity than histone modifications or chromatin accessibility. Epigenetics Chromatin *8*, 16.

Le Douce, V., Herbein, G., Rohr, O., and Schwartz, C. (2010). Molecular mechanisms of HIV-1 persistence in the monocyte-macrophage lineage. Retrovirology *7*, 32.

Dowen, J.M., Fan, Z.P., Hnisz, D., Ren, G., Abraham, B.J., Zhang, L.N., Weintraub, A.S., Schuijers, J., Lee, T.I., Zhao, K., et al. (2014). Control of Cell Identity Genes Occurs in Insulated Neighborhoods in Mammalian Chromosomes. Cell *159*, 374–387.

Drier, Y., Cotton, M.J., Williamson, K.E., Gillespie, S.M., Ryan, R.J.H., Kluk, M.J., Carey, C.D., Rodig, S.J., Sholl, L.M., Afrogheh, A.H., et al. (2016). An oncogenic MYB feedback loop drives alternate cell fates in adenoid cystic carcinoma. Nat. Genet. *48*, 265–272.

Dufour, C., Richard, C., Pardons, M., Massanella, M., Ackaoui, A., Murrell, B., Routy, B., Thomas, R., Routy, J.P., Fromentin, R., et al. (2023). Phenotypic characterization of single CD4+ T cells harboring genetically intact and inducible HIV genomes. Nat. Commun. *14*, 1–15.

Dupont, L., Bloor, S., Williamson, J.C., Cuesta, S.M., Shah, R., Teixeira-Silva, A., Naamati, A., Greenwood, E.J.D., Sarafianos, S.G., Matheson, N.J., et al. (2021). The SMC5/6 complex compacts and silences unintegrated HIV-1 DNA and is antagonized by Vpr. Cell Host Microbe *29*, 792-805.e6.

Dutilleul, A., Rodari, A., and Van Lint, C. (2020). Depicting HIV-1 Transcriptional Mechanisms: A Summary of What We Know. Viruses *12*.

Duverger, A., Wolschendorf, F., Zhang, M., Wagner, F., Hatcher, B., Jones, J., Cron, R.Q., van der Sluis, R.M., Jeeninga, R.E., Berkhout, B., et al. (2013). An AP-1 binding

site in the enhancer/core element of the HIV-1 promoter controls the ability of HIV-1 to establish latent infection. J. Virol. *87*, 2264–2277.

Ebina, H., Misawa, N., Kanemura, Y., and Koyanagi, Y. (2013). Harnessing the CRISPR/Cas9 system to disrupt latent HIV-1 provirus. Sci. Rep. *3*, 2510.

Edén, A., Fuchs, D., Hagberg, L., Nilsson, S., Spudich, S., Svennerholm, B., Price, R.W., and Gisslén, M. (2010). HIV-1 viral escape in cerebrospinal fluid of subjects on suppressive antiretroviral treatment. J. Infect. Dis. *202*, 1819–1825.

Edén, A., Nilsson, S., Hagberg, L., Fuchs, D., Zetterberg, H., Svennerholm, B., and Gisslén, M. (2016). Asymptomatic Cerebrospinal Fluid HIV-1 Viral Blips and Viral Escape During Antiretroviral Therapy: A Longitudinal Study. J. Infect. Dis. *214*, 1822–1825.

Edmunds, J.W., Mahadevan, L.C., and Clayton, A.L. (2008). Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation. EMBO J. *27*, 406–420.

Eggers, C., Arendt, G., Hahn, K., Husstedt, I.W., Maschke, M., Neuen-Jacob, E., Obermann, M., Rosenkranz, T., Schielke, E., and Straube, E. (2017). HIV-1-associated neurocognitive disorder: epidemiology, pathogenesis, diagnosis, and treatment. J Neurol *264*, 1715–1727.

Eidahl, J.O., Crowe, B.L., North, J.A., McKee, C.J., Shkriabai, N., Feng, L., Plumb, M., Graham, R.L., Gorelick, R.J., Hess, S., et al. (2013). Structural basis for high-affinity binding of LEDGF PWWP to mononucleosomes. Nucleic Acids Res *41*, 3924–3936.

Einkauf, K.B., Lee, G.Q., Gao, C., Sharaf, R., Sun, X., Hua, S., Chen, S.M.Y., Jiang, C., Lian, X., Chowdhury, F.Z., et al. (2019). Intact HIV-1 proviruses accumulate at distinct chromosomal positions during prolonged antiretroviral therapy. J. Clin. Invest. *129*, 988–998.

Einkauf, K.B., Osborn, M.R., Gao, C., Sun, W., Sun, X., Lian, X., Parsons, E.M., Gladkov, G.T., Seiger, K.W., Blackmer, J.E., et al. (2022). Parallel analysis of transcription, integration, and sequence of single HIV-1 proviruses. Cell 1–17.

Emerson, D.J., Zhao, P.A., Cook, A.L., Barnett, R.J., Klein, K.N., Saulebekova, D., Ge, C., Zhou, L., Simandi, Z., Minsk, M.K., et al. (2022). Cohesin-mediated loop anchors confine the locations of human replication origins. Nature *606*, 812–819.

Engelman, A.N., and Cherepanov, P. (2017). Retroviral intasomes arising. Curr. Opin. Struct. Biol. *47*, 23–29.

Engelman, A.N., and Singh, P.K. (2018). Cellular and molecular mechanisms of HIV-

1 integration targeting. Cell. Mol. Life Sci. *75*, 2491–2507.

Engelman, A., Mizuuchi, K., and Craigie, R. (1991). HIV-1 DNA integration: mechanism of viral DNA cleavage and DNA strand transfer. Cell *67*, 1211–1221.

Ernst, J., and Kellis, M. (2010). Discovery and characterization of chromatin states for systematic annotation of the human genome. Nat. Biotechnol. *28*, 817–825.

Ernst, J., and Kellis, M. (2017). Chromatin-state discovery and genome annotation with ChromHMM. Nat. Protoc. *12*, 2478–2492.

Ertel, M.K., Cammarata, A.L., Hron, R.J., and Neumann, D.M. (2012). CTCF occupation of the herpes simplex virus 1 genome is disrupted at early times postreactivation in a transcription-dependent manner. J. Virol. *86*, 12741–12759.

Falcinelli, S.D., Kilpatrick, K.W., Read, J., Murtagh, R., Allard, B., Ghofrani, S., Kirchherr, J., James, K.S., Stuelke, E., Baker, C., et al. (2021). Longitudinal Dynamics of Intact HIV Proviral DNA and Outgrowth Virus Frequencies in a Cohort of Individuals Receiving Antiretroviral Therapy. J. Infect. Dis. *224*, 92–100.

Fanucchi, S., Shibayama, Y., Burd, S., Weinberg, M.S., and Mhlanga, M.M. (2013). Chromosomal contact permits transcription between coregulated genes. Cell *155*, 606–620.

Farhadian, S.F., Mehta, S.S., Zografou, C., Robertson, K., Price, R.W., Pappalardo, J., Chiarella, J., Hafler, D.A., and Spudich, S.S. (2018). Single-cell RNA sequencing reveals microglia-like cells in cerebrospinal fluid during virologically suppressed HIV. JCI Insight *3*, 1–7.

Farnet, C.M., and Haseltine, W.A. (1991). Circularization of human immunodeficiency virus type 1 DNA in vitro. J Virol *65*, 6942–6952.

Farrar, D., Rai, S., Chernukhin, I., Jagodic, M., Ito, Y., Yammine, S., Ohlsson, R., Murrell, A., and Klenova, E. (2010). Mutational analysis of the poly(ADP-ribosyl)ation sites of the transcription factor CTCF provides an insight into the mechanism of its regulation by poly(ADP-ribosyl)ation. Mol. Cell. Biol. *30*, 1199–1216.

Faschinger, A., Rouault, F., Sollner, J., Lukas, A., Salmons, B., Günzburg, W.H., and Indik, S. (2008). Mouse mammary tumor virus integration site selection in human and mouse genomes. J. Virol. *82*, 1360–1367.

Felts, R.L., Narayan, K., Estes, J.D., Shi, D., Trubey, C.M., Fu, J., Hartnell, L.M., Ruthel, G.T., Schneider, D.K., Nagashima, K., et al. (2010). 3D visualization of HIV transfer at the virological synapse between dendritic cells and T cells. Proc. Natl. Acad. Sci. U. S. A. *107*, 13336–13341.

Feng, Q., Wang, H., Ng, H.H., Erdjument-Bromage, H., Tempst, P., Struhl, K., and Zhang, Y. (2002). Methylation of H3-lysine 79 is mediated by a new family of HMTases without a SET domain. Curr. Biol. *12*, 1052–1058.

Ferretti, F., Gisslen, M., Cinque, P., and Price, R.W. (2015). Cerebrospinal Fluid HIV Escape from Antiretroviral Therapy. Curr. HIV/AIDS Rep. *12*, 280–288.

Filippova, G.N., Fagerlie, S., Klenova, E.M., Myers, C., Dehner, Y., Goodwin, G., Neiman, P.E., Collins, S.J., and Lobanenkov, V. V (1996). An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. Mol. Cell. Biol. *16*, 2802–2813.

Finn, E.H., and Misteli, T. (2019). Molecular basis and biological function of variability in spatial genome organization. Science *365*.

Finn, E.H., Pegoraro, G., Brandão, H.B., Valton, A.-L., Oomen, M.E., Dekker, J., Mirny, L., and Misteli, T. (2019). Extensive Heterogeneity and Intrinsic Variation in Spatial Genome Organization. Cell *176*, 1502-1515.e10.

Fischle, W., Tseng, B.S., Dormann, H.L., Ueberheide, B.M., Garcia, B.A., Shabanowitz, J., Hunt, D.F., Funabiki, H., and Allis, C.D. (2005). Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation. Nature *438*, 1116–1122.

Fischle, W., Franz, H., Jacobs, S.A., Allis, C.D., and Khorasanizadeh, S. (2008). Specificity of the chromodomain Y chromosome family of chromodomains for lysine-methylated ARK(S/T) motifs. J. Biol. Chem. *283*, 19626–19635.

Flavahan, W.A., Drier, Y., Liau, B.B., Gillespie, S.M., Venteicher, A.S., Stemmer-Rachamimov, A.O., Suvà, M.L., and Bernstein, B.E. (2016). Insulator dysfunction and oncogene activation in IDH mutant gliomas. Nature *529*, 110–114.

Francis, A.C., Marin, M., Singh, P.K., Achuthan, V., Prellberg, M.J., Palermino-Rowland, K., Lan, S., Tedbury, P.R., Sarafianos, S.G., Engelman, A.N., et al. (2020). HIV-1 replication complexes accumulate in nuclear speckles and integrate into speckle-associated genomic domains. Nat. Commun. *11*, 3505.

Frankel, A.D., and Young, J.A. (1998). HIV-1: fifteen proteins and an RNA. Annu. Rev. Biochem. *67*, 1–25.

Fraser, J., Ferrai, C., Chiariello, A.M., Schueler, M., Rito, T., Laudanno, G., Barbieri, M., Moore, B.L., Kraemer, D.C.A., Aitken, S., et al. (2015). Hierarchical folding and reorganization of chromosomes are linked to transcriptional changes in cellular

differentiation. Mol. Syst. Biol. *11*, 852.

Freed, E.O. (2015). HIV-1 assembly, release and maturation. Nat. Rev. Microbiol. *13*, 484–496.

Friedman, J., Cho, W.-K., Chu, C.K., Keedy, K.S., Archin, N.M., Margolis, D.M., and Karn, J. (2011). Epigenetic Silencing of HIV-1 by the Histone H3 Lysine 27 Methyltransferase Enhancer of Zeste 2. J. Virol. *85*, 9078–9089.

Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N., and Mirny, L.A. (2016). Formation of Chromosomal Domains by Loop Extrusion. Cell Rep. *15*, 2038–2049.

Furlong, E.E.M., and Levine, M. (2018). Developmental enhancers and chromosome topology. Science *361*, 1341–1345.

Gabriele, M., Brandão, H.B., Grosse-Holz, S., Jha, A., Dailey, G.M., Cattoglio, C., Hsieh, T.-H.S., Mirny, L., Zechner, C., and Hansen, A.S. (2022). Dynamics of CTCF- and cohesin-mediated chromatin looping revealed by live-cell imaging. Science *376*, 496–501.

Gaebler, C., Lorenzi, J.C.C., Oliveira, T.Y., Nogueira, L., Ramos, V., Lu, C.-L., Pai, J.A., Mendoza, P., Jankovic, M., Caskey, M., et al. (2019). Combination of quadruplex qPCR and next-generation sequencing for qualitative and quantitative analysis of the HIV-1 latent reservoir. J. Exp. Med. *216*, 2253–2264.

Galisson, F., Mahrouche, L., Courcelles, M., Bonneil, E., Meloche, S., Chelbi-Alix, M.K., and Thibault, P. (2011). A novel proteomics approach to identify SUMOylated proteins and their modification sites in human cells. Mol. Cell. Proteomics *10*, M110.004796.

Galli, G.G., Carrara, M., Francavilla, C., de Lichtenberg, K.H., Olsen, J.V., Calogero, R.A., and Lund, A.H. (2013). Genomic and proteomic analyses of Prdm5 reveal interactions with insulator binding proteins in embryonic stem cells. Mol. Cell. Biol. *33*, 4504–4516.

Gamble, T.R., Vajdos, F.F., Yoo, S., Worthylake, D.K., Houseweart, M., Sundquist, W.I., and Hill, C.P. (1996). Crystal structure of human cyclophilin A bound to the amino-terminal domain of HIV-1 capsid. Cell *87*, 1285–1294.

Ganji, M., Shaltiel, I.A., Bisht, S., Kim, E., Kalichava, A., Haering, C.H., and Dekker, C. (2018). Real-time imaging of DNA loop extrusion by condensin. Science *360*, 102–105.

Gantner, P., Pagliuzza, A., Pardons, M., Ramgopal, M., Routy, J.-P., Fromentin, R.,

and Chomont, N. (2020). Single-cell TCR sequencing reveals phenotypically diverse clonally expanded cells harboring inducible HIV proviruses during ART. Nat. Commun. *11*, 4089.

Gao, T., and Qian, J. (2020). EnhancerAtlas 2.0: an updated resource with enhancer annotation in 586 tissue/cell types across nine species. Nucleic Acids Res. *48*, D58–D64.

Gao, F., Yue, L., White, A.T., Pappas, P.G., Barchue, J., Hanson, A.P., Greene, B.M., Sharp, P.M., Shaw, G.M., and Hahn, B.H. (1992). Human infection by genetically diverse SIVSM-related HIV-2 in west Africa. Nature *358*, 495–499.

Gao, F., Bailes, E., Robertson, D.L., Chen, Y., Rodenburg, C.M., Michael, S.F., Cummins, L.B., Arthur, L.O., Peeters, M., Shaw, G.M., et al. (1999). Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes. Nature *397*, 436–441.

Garcia-Broncano, P., Maddali, S., Einkauf, K.B., Jiang, C., Gao, C., Chevalier, J., Chowdhury, F.Z., Maswabi, K., Ajibola, G., Moyo, S., et al. (2019). Early antiretroviral therapy in neonates with HIV-1 infection restricts viral reservoir size and induces a distinct innate immune profile. Sci. Transl. Med. *11*.

Garcia-Mesa, Y., Jay, T.R., Checkley, M.A., Luttge, B., Dobrowolski, C., Valadkhan, S., Landreth, G.E., Karn, J., and Alvarez-Carbonell, D. (2017). Immortalization of primary microglia: a new platform to study HIV regulation in the central nervous system. J Neurovirol *23*, 47–66.

Gasperini, M., Tome, J.M., and Shendure, J. (2020). Towards a comprehensive catalogue of validated and target-linked human enhancers. Nat. Rev. Genet. *21*, 292–310.

Geis, F.K., and Goff, S.P. (2019). Unintegrated HIV-1 DNAs are loaded with core and linker histones and transcriptionally silenced. Proc. Natl. Acad. Sci. U. S. A. *116*, 23735–23742.

Geis, F.K., Sabo, Y., Chen, X., Li, Y., Lu, C., and Goff, S.P. (2022a). CHAF1A/B mediate silencing of unintegrated HIV-1 DNAs early in infection. Proc. Natl. Acad. Sci. U. S. A. *119*.

Geis, F.K., Kelenis, D.P., and Goff, S.P. (2022b). Two lymphoid cell lines potently silence unintegrated HIV-1 DNAs. Retrovirology *19*, 16.

Gershey, E.L., Haslett, G.W., Vidali, G., and Allfrey, V.G. (1969). Chemical studies of histone methylation. Evidence for the occurrence of 3-methylhistidine in avian erythrocyte histone fractions. J. Biol. Chem. *244*, 4871–4877.

Gillet, N.A., Malani, N., Melamed, A., Gormley, N., Carter, R., Bentley, D., Berry, C., Bushman, F.D., Taylor, G.P., and Bangham, C.R. (2011). The host genomic environment of the provirus determines the abundance of HTLV-1-infected T-cell clones. Blood *117*, 3113–3122.

Gillet, N.A., Gutiérrez, G., Rodriguez, S.M., de Brogniez, A., Renotte, N., Alvarez, I., Trono, K., and Willems, L. (2013). Massive depletion of bovine leukemia virus proviral clones located in genomic transcriptionally active sites during primary infection. PLoS Pathog. *9*, e1003687.

Gillet, N.A., Melamed, A., and Bangham, C.R. (2017). High-Throughput Mapping and Clonal Quantification of Retroviral Integration Sites. Methods Mol Biol *1582*, 127–141.

Ginhoux, F., Greter, M., Leboeuf, M., Nandi, S., See, P., Gokhan, S., Mehler, M.F., Conway, S.J., Ng, L.G., Stanley, E.R., et al. (2010). Fate mapping analysis reveals that adult microglia derive from primitive macrophages. Science (80-.). *330*, 841–845.

Ginsberg, S.D., Alldred, M.J., Gunnam, S.M., Schiroli, C., Lee, S.H., Morgello, S., and Fischer, T. (2018). Expression profiling suggests microglial impairment in human immunodeficiency virus neuropathogenesis. Ann. Neurol. *83*, 406–417.

Glass, C.K., and Natoli, G. (2016). Molecular control of activation and priming in macrophages. Nat. Immunol. *17*, 26–33.

Gligoris, T.G., Scheinost, J.C., Bürmann, F., Petela, N., Chan, K.-L., Uluocak, P., Beckouët, F., Gruber, S., Nasmyth, K., and Löwe, J. (2014). Closing the cohesin ring: structure and function of its Smc3-kleisin interface. Science *346*, 963–967.

González-Velasco, O., Papy-García, D., Le Douaron, G., Sánchez-Santos, J.M., and De Las Rivas, J. (2020). Transcriptomic landscape, gene signatures and regulatory profile of aging in the human brain. Biochim. Biophys. Acta - Gene Regul. Mech. *1863*, 194491.

Gonzalo-Gil, E., Ikediobi, U., and Sutton, R.E. (2017). Mechanisms of Virologic Control and Clinical Characteristics of HIV+ Elite/Viremic Controllers. Yale J. Biol. Med. *90*, 245–259.

Gosselin, D., Skola, D., Coufal, N.G., Holtman, I.R., Schlachetzki, J.C.M., Sajti, E., Jaeger, B.N., O'Connor, C., Fitzpatrick, C., Pasillas, M.P., et al. (2017). An environment-dependent transcriptional network specifies human microglia identity. Science (80-.). *356*.

Groner, A.C., Tschopp, P., Challet, L., Dietrich, J.-E., Verp, S., Offner, S., Barde, I., Rodriguez, I., Hiiragi, T., and Trono, D. (2012). The Krüppel-associated box repressor

domain can induce reversible heterochromatization of a mouse locus in vivo. J. Biol. Chem. *287*, 25361–25369.

Gröschel, S., Sanders, M.A., Hoogenboezem, R., de Wit, E., Bouwman, B.A.M., Erpelinck, C., van der Velden, V.H.J., Havermans, M., Avellino, R., van Lom, K., et al. (2014). A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in leukemia. Cell *157*, 369–381.

Grubert, F., Srivas, R., Spacek, D. V., Kasowski, M., Ruiz-Velasco, M., Sinnott-Armstrong, N., Greenside, P., Narasimha, A., Liu, Q., Geller, B., et al. (2020). Landscape of cohesin-mediated chromatin loops in the human genome. Nature *583*, 737–743.

Guadalupe, M., Reay, E., Sankaran, S., Prindiville, T., Flamm, J., McNeil, A., and Dandekar, S. (2003). Severe CD4+ T-cell depletion in gut lymphoid tissue during primary human immunodeficiency virus type 1 infection and substantial delay in restoration following highly active antiretroviral therapy. J. Virol. *77*, 11708–11717.

Guelen, L., Pagie, L., Brasset, E., Meuleman, W., Faza, M.B., Talhout, W., Eussen, B.H., de Klein, A., Wessels, L., de Laat, W., et al. (2008). Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. Nature *453*, 948–951.

Haarhuis, J.H.I., van der Weide, R.H., Blomen, V.A., Yáñez-Cuna, J.O., Amendola, M., van Ruiten, M.S., Krijger, P.H.L., Teunissen, H., Medema, R.H., van Steensel, B., et al. (2017). The Cohesin Release Factor WAPL Restricts Chromatin Loop Extension. Cell *169*, 693-707.e14.

Haberland, M., Montgomery, R.L., and Olson, E.N. (2009). The many roles of histone deacetylases in development and physiology: implications for disease and therapy. Nat. Rev. Genet. *10*, 32–42.

Halfon, M.S. (2020). Silencers, Enhancers, and the Multifunctional Regulatory Genome. Trends Genet. *36*, 149–151.

Haller, F., Bieg, M., Will, R., Körner, C., Weichenhan, D., Bott, A., Ishaque, N., Lutsik, P., Moskalev, E.A., Mueller, S.K., et al. (2019). Enhancer hijacking activates oncogenic transcription factor NR4A3 in acinic cell carcinomas of the salivary glands. Nat. Commun. *10*, 368.

Halvas, E.K., Joseph, K.W., Brandt, L.D., Guo, S., Sobolewski, M.D., Jacobs, J.L., Tumiotto, C., Bui, J.K., Cyktor, J.C., Keele, B.F., et al. (2020). HIV-1 viremia not suppressible by antiretroviral therapy can originate from large T cell clones producing

infectious virus. J Clin Invest *130*, 5847–5857.

Hamid, F. Bin, Kim, J., and Shin, C.-G. (2017). Distribution and fate of HIV-1 unintegrated DNA species: a comprehensive update. AIDS Res. Ther. *14*, 9.

Han, D., Chen, Q., Shi, J., Zhang, F., and Yu, X. (2017). CTCF participates in DNA damage response via poly(ADP-ribosyl)ation. Sci. Rep. *7*, 43530.

Hansen, A.S., Pustova, I., Cattoglio, C., Tjian, R., and Darzacq, X. (2017). CTCF and cohesin regulate chromatin loop stability with distinct dynamics. Elife *6*, e25776.

Hansen, A.S., Cattoglio, C., Darzacq, X., and Tjian, R. (2018). Recent evidence that TADs and chromatin loops are dynamic structures. Nucleus *9*, 20–32.

Hansen, A.S., Hsieh, T.-H.S., Cattoglio, C., Pustova, I., Saldaña-Meyer, R., Reinberg, D., Darzacq, X., and Tjian, R. (2019). Distinct Classes of Chromatin Loops Revealed by Deletion of an RNA-Binding Region in CTCF. Mol. Cell *76*, 395-411.e13.

Hanssen, L.L.P., Kassouf, M.T., Oudelaar, A.M., Biggs, D., Preece, C., Downes, D.J., Gosden, M., Sharpe, J.A., Sloane-Stanley, J.A., Hughes, J.R., et al. (2017). Tissue-specific CTCF-cohesin-mediated chromatin architecture delimits enhancer interactions and function in vivo. Nat. Cell Biol. *19*, 952–961.

Hare, S., and Cherepanov, P. (2009). The Interaction Between Lentiviral Integrase and LEDGF: Structural and Functional Insights. Viruses *1*, 780–801.

Hare, S., Gupta, S.S., Valkov, E., Engelman, A., and Cherepanov, P. (2010). Retroviral intasome assembly and inhibition of DNA strand transfer. Nature *464*, 232–236.

Hare, S., Maertens, G.N., and Cherepanov, P. (2012). 3'-processing and strand transfer catalysed by retroviral integrase in crystallo. EMBO J. *31*, 3020–3028.

Hark, A.T., Schoenherr, C.J., Katz, D.J., Ingram, R.S., Levorse, J.M., and Tilghman, S.M. (2000). CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. Nature *405*, 486–489.

Hashimoto, D., Chow, A., Noizat, C., Teo, P., Beasley, M.B., Leboeuf, M., Becker, C.D., See, P., Price, J., Lucas, D., et al. (2013). Tissue-resident macrophages self-maintain locally throughout adult life with minimal contribution from circulating monocytes. Immunity *38*, 792–804.

Hata, A., and Chen, Y.-G. (2016). TGF-β Signaling from Receptors to Smads. Cold Spring Harb. Perspect. Biol. *8*.

He, N., Liu, M., Hsu, J., Xue, Y., Chou, S., Burlingame, A., Krogan, N.J., Alber, T., and Zhou, Q. (2010). HIV-1 Tat and host AFF4 recruit two transcription elongation factors into a bifunctional complex for coordinated activation of HIV-1 transcription. Mol. Cell

38, 428–438.

Heaton, R.K., Franklin, D.R., Ellis, R.J., McCutchan, J.A., Letendre, S.L., LeBlanc, S., Corkran, S.H., Duarte, N.A., Clifford, D.B., Woods, S.P., et al. (2011). HIV-associated neurocognitive disorders before and during the era of combination antiretroviral therapy: Differences in rates, nature, and predictors. J. Neurovirol. *17*, 3–16.

Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat. Genet. *39*, 311–318.

Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W., et al. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature *459*, 108–112.

Heinz, S., Romanoski, C.E., Benner, C., and Glass, C.K. (2015). The selection and function of cell type-specific enhancers. Nat. Rev. Mol. Cell Biol. *16*, 144–154.

Heinz, S., Texari, L., Hayes, M.G.B., Urbanowski, M., Chang, M.W., Givarkes, N., Rialdi, A., White, K.M., Albrecht, R.A., Pache, L., et al. (2018). Transcription Elongation Can Affect Genome 3D Structure. Cell *174*, 1522-1536.e22.

Hemelaar, J., Gouws, E., Ghys, P.D., and Osmanov, S. (2011). Global trends in molecular epidemiology of HIV-1 during 2000-2007. AIDS *25*, 679–689.

Hempel, K., Lange, H.W., and Birkofer, L. (1968). [Epsilon-N-trimethyllysine, a new amino acid in histones]. Naturwissenschaften *55*, 37.

Hendriks, I.A., D'Souza, R.C.J., Yang, B., Verlaan-de Vries, M., Mann, M., and Vertegaal, A.C.O. (2014). Uncovering global SUMOylation signaling networks in a site-specific manner. Nat. Struct. Mol. Biol. *21*, 927–936.

Herrmann, C.H., and Rice, A.P. (1995). 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. *69*, 1612–1620.

Hesselberth, J.R., Chen, X., Zhang, Z., Sabo, P.J., Sandstrom, R., Reynolds, A.P., Thurman, R.E., Neph, S., Kuehn, M.S., Noble, W.S., et al. (2009). Global mapping of protein-DNA interactions in vivo by digital genomic footprinting. Nat. Methods *6*, 283–289.

Hiatt, J., Hultquist, J.F., McGregor, M.J., Bouhaddou, M., Leenay, R.T., Simons, L.M., Young, J.M., Haas, P., Roth, T.L., Tobin, V., et al. (2022). A functional map of HIV-

host interactions in primary human T cells. Nat. Commun. *13*, 1752.

Hiener, B., Horsburgh, B.A., Eden, J.-S., Barton, K., Schlub, T.E., Lee, E., von Stockenstrom, S., Odevall, L., Milush, J.M., Liegler, T., et al. (2017). Identification of Genetically Intact HIV-1 Proviruses in Specific CD4(+) T Cells from Effectively Treated Participants. Cell Rep. *21*, 813–822.

Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-André, V., Sigova, A.A., Hoke, H.A., and Young, R.A. (2013). Super-enhancers in the control of cell identity and disease. Cell *155*, 934–947.

Ho, Y.C., Shan, L., Hosmane, N.N., Wang, J., Laskey, S.B., Rosenbloom, D.I., Lai, J., Blankson, J.N., Siliciano, J.D., and Siliciano, R.F. (2013). Replication-competent noninduced proviruses in the latent reservoir increase barrier to HIV-1 cure. Cell *155*, 540–551.

Holdorf, M.M., Cooper, S.B., Yamamoto, K.R., and Miranda, J.J.L. (2011). Occupancy of chromatin organizers in the Epstein-Barr virus genome. Virology *415*, 1–5.

Holman, A.G., and Coffin, J.M. (2005). Symmetrical base preferences surrounding HIV-1, avian sarcoma/leukosis virus, and murine leukemia virus integration sites. Proc Natl Acad Sci U S A *102*, 6103–6107.

Hoppmann, V., Thorstensen, T., Kristiansen, P.E., Veiseth, S.V., Rahman, M.A., Finne, K., Aalen, R.B., and Aasland, R. (2011). The CW domain, a new histone recognition module in chromatin proteins. EMBO J. *30*, 1939–1952.

Hosmane, N.N., Kwon, K.J., Bruner, K.M., Capoferri, A.A., Beg, S., Rosenbloom, D.I.S., Keele, B.F., Ho, Y.-C., Siliciano, J.D., and Siliciano, R.F. (2017). Proliferation of latently infected CD4(+) T cells carrying replication-competent HIV-1: Potential role in latent reservoir dynamics. J. Exp. Med. *214*, 959–972.

Hou, C., Zhao, H., Tanimoto, K., and Dean, A. (2008). CTCF-dependent enhancer-blocking by alternative chromatin loop formation. Proc. Natl. Acad. Sci. U. S. A. *105*, 20398–20403.

Hsieh, T.-H.S., Cattoglio, C., Slobodyanyuk, E., Hansen, A.S., Darzacq, X., and Tjian, R. (2022). Enhancer–promoter interactions and transcription are largely maintained upon acute loss of CTCF, cohesin, WAPL or YY1. Nat. Genet. *54*, 1919–1932.

Hsieh, T.H.S., Cattoglio, C., Slobodyanyuk, E., Hansen, A.S., Rando, O.J., Tjian, R., and Darzacq, X. (2020). Resolving the 3D Landscape of Transcription-Linked Mammalian Chromatin Folding. Mol. Cell *78*, 539-553.e8.

Hu, B., Won, H., Mah, W., Park, R.B., Kassim, B., Spiess, K., Kozlenkov, A., Crowley,

C.A., Pochareddy, S., Ashley-Koch, A.E., et al. (2021). Neuronal and glial 3D chromatin architecture informs the cellular etiology of brain disorders. Nat. Commun. *12*, 3968.

Huang, A.S., Ramos, V., Oliveira, T.Y., Gaebler, C., Jankovic, M., Nussenzweig, M.C., and Cohn, L.B. (2021). Integration features of intact latent HIV-1 in CD4+ T cell clones contribute to viral persistence. J. Exp. Med. *218*.

Hübner, W., McNerney, G.P., Chen, P., Dale, B.M., Gordon, R.E., Chuang, F.Y.S., Li, X.-D., Asmuth, D.M., Huser, T., and Chen, B.K. (2009). Quantitative 3D video microscopy of HIV transfer across T cell virological synapses. Science *323*, 1743–1747.

Hyle, J., Zhang, Y., Wright, S., Xu, B., Shao, Y., Easton, J., Tian, L., Feng, R., Xu, P., and Li, C. (2019). Acute depletion of CTCF directly affects MYC regulation through loss of enhancer-promoter looping. Nucleic Acids Res. *47*, 6699–6713.

De Iaco, A., and Luban, J. (2014). Cyclophilin A promotes HIV-1 reverse transcription but its effect on transduction correlates best with its effect on nuclear entry of viral cDNA. Retrovirology *11*, 11.

Ibarra, A., Benner, C., Tyagi, S., Cool, J., and Hetzer, M.W. (2016). Nucleoporin-mediated regulation of cell identity genes. Genes Dev *30*, 2253–2258.

Ikeda, T., Shibata, J., Yoshimura, K., Koito, A., and Matsushita, S. (2007). Recurrent HIV-1 integration at the BACH2 locus in resting CD4+ T cell populations during effective highly active antiretroviral therapy. J Infect Dis *195*, 716–725.

Imai, K., Togami, H., and Okamoto, T. (2010). Involvement of histone H3 lysine 9 (H3K9) methyltransferase G9a in the maintenance of HIV-1 latency and its reactivation by BIX01294. J Biol Chem *285*, 16538–16545.

Imle, A., Kumberger, P., Schnellbächer, N.D., Fehr, J., Carrillo-Bustamante, P., Ales, J., Schmidt, P., Ritter, C., Godinez, W.J., Müller, B., et al. (2019). Experimental and computational analyses reveal that environmental restrictions shape HIV-1 spread in 3D cultures. Nat. Commun. *10*, 2144.

Ingram, Z., Taylor, M., Okland, G., Martin, R., and Hulme, A.E. (2020). Characterization of HIV-1 uncoating in human microglial cell lines. Virol. J. *17*, 31.

Irwan, I.D., Bogerd, H.P., and Cullen, B.R. (2022). Epigenetic silencing by the SMC5/6 complex mediates HIV-1 latency. Nat. Microbiol.

Iwamori, N., Tominaga, K., Sato, T., Riehle, K., Iwamori, T., Ohkawa, Y., Coarfa, C., Ono, E., and Matzuk, M.M. (2016). MRG15 is required for pre-mRNA splicing and

spermatogenesis. Proc. Natl. Acad. Sci. U. S. A. *113*, E5408-15.

Izumoto, Y., Kuroda, T., Harada, H., Kishimoto, T., and Nakamura, H. (1997). Hepatoma-derived growth factor belongs to a gene family in mice showing significant homology in the amino terminus. Biochem Biophys Res Commun *238*, 26–32.

Jacks, T., Power, M.D., Masiarz, F.R., Luciw, P.A., Barr, P.J., and Varmus, H.E. (1988). Characterization of ribosomal frameshifting in HIV-1 gag-pol expression. Nature *331*, 280–283.

Jäger, S., Cimermancic, P., Gulbahce, N., Johnson, J.R., McGovern, K.E., Clarke, S.C., Shales, M., Mercenne, G., Pache, L., Li, K., et al. (2011). Global landscape of HIV-human protein complexes. Nature *481*, 365–370.

Janssens, D.H., Otto, D.J., Meers, M.P., Setty, M., Ahmad, K., and Henikoff, S. (2022). CUT&Tag2for1: a modified method for simultaneous profiling of the accessible and silenced regulome in single cells. Genome Biol. *23*, 81.

Javierre, B.M., Burren, O.S., Wilder, S.P., Kreuzhuber, R., Hill, S.M., Sewitz, S., Cairns, J., Wingett, S.W., Várnai, C., Thiecke, M.J., et al. (2016). Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. Cell *167*, 1369-1384.e19.

Jeanson, L., Subra, F., Vaganay, S., Hervy, M., Marangoni, E., Bourhis, J., and Mouscadet, J.-F. (2002). Effect of Ku80 Depletion on the Preintegrative Steps of HIV-1 Replication in Human Cells. Virology *300*, 100–108.

Jefferys, S.R., Burgos, S.D., Peterson, J.J., Selitsky, S.R., Turner, A.M.W., James, L.I., Tsai, Y.H., Coffey, A.R., Margolis, D.M., Parker, J., et al. (2021). Epigenomic characterization of latent HIV infection identifies latency regulating transcription factors.

Jiang, C., Lian, X., Gao, C., Sun, X., Einkauf, K.B., Chevalier, J.M., Chen, S.M.Y., Hua, S., Rhee, B., Chang, K., et al. (2020). Distinct viral reservoirs in individuals with spontaneous control of HIV-1. Nature *585*, 261–267.

Jin, W., Tang, Q., Wan, M., Cui, K., Zhang, Y., Ren, G., Ni, B., Sklar, J., Przytycka, T.M., Childs, R., et al. (2015). Genome-wide detection of DNase I hypersensitive sites in single cells and FFPE tissue samples. Nature *528*, 142–146.

Joseph, K.W., Halvas, E.K., Brandt, L.D., Patro, S.C., Rausch, J.W., Chopra, A., Mallal, S., Kearney, M.F., Coffin, J.M., and Mellors, J.W. (2022). Deep Sequencing Analysis of Individual HIV-1 Proviruses Reveals Frequent Asymmetric Long Terminal Repeats. J. Virol.

Joseph, S.B., Trunfio, M., Kincer, L.P., Calcagno, A., and Price, R.W. (2019). What can characterization of cerebrospinal fluid escape populations teach us about viral reservoirs in the central nervous system? Aids *33 Suppl 2*, S171-s179.

Kaczmarczyk, L.S., Levi, N., Segal, T., Salmon-Divon, M., and Gerlitz, G. (2022). CTCF supports preferentially short lamina-associated domains. Chromosom. Res. an Int. J. Mol. Supramol. Evol. Asp. Chromosom. Biol. *30*, 123–136.

Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S., et al. (2010). Mediator and cohesin connect gene expression and chromatin architecture. Nature *467*, 430–435.

Kalle, E., Kubista, M., and Rensing, C. (2014). Multi-template polymerase chain reaction. Biomol. Detect. Quantif. *2*, 11–29.

Kalpana, G. V, Marmon, S., Wang, W., Crabtree, G.R., and Goff, S.P. (1994). Binding and Stimulation of HIV-1 Integrase by a Human Homolog of Yeast Transcription Factor SNF5. Science (80-.). *266*, 2002–2006.

Kamat, A., Misra, V., Cassol, E., Ancuta, P., Yan, Z., Li, C., Morgello, S., and Gabuzda, D. (2012). A plasma biomarker signature of immune activation in HIV patients on antiretroviral therapy. PLoS One *7*, e30881.

Kauder, S.E., Bosque, A., Lindqvist, A., Planelles, V., and Verdin, E. (2009). Epigenetic regulation of HIV-1 latency by cytosine methylation. PLoS Pathog *5*, e1000495.

Kelly, J., Beddall, M.H., Yu, D., Iyer, S.R., Marsh, J.W., and Wu, Y. (2008). Human macrophages support persistent transcription from unintegrated HIV-1 DNA. Virology *372*, 300–312.

Kelly, T.K., Liu, Y., Lay, F.D., Liang, G., Berman, B.P., and Jones, P.A. (2012). Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. Genome Res. *22*, 2497–2506.

Khan, E., Mack, P.G.J., Katz, A.R., Kulkosky, J., and Skalka, A.M. (1990). Retroviral integrase domains: DNA binding and the recognition of LTR sequences. Nucleic Acids Res. *19*, 851–860.

Khoury, A., Achinger-Kawecka, J., Bert, S.A., Smith, G.C., French, H.J., Luu, P.L., Peters, T.J., Du, Q., Parry, A.J., Valdes-Mora, F., et al. (2020). Constitutively bound CTCF sites maintain 3D chromatin architecture and long-range epigenetically regulated domains. Nat. Commun. *11*.

Kilzer, J.M., Stracker, T., Beitzel, B., Meek, K., Weitzman, M., and Bushman, F.D. (2003). Roles of host cell factors in circularization of retroviral dna. Virology *314*, 460–

467.

Kim, Y., Shi, Z., Zhang, H., Finkelstein, I.J., and Yu, H. (2019). Human cohesin compacts DNA by loop extrusion. Science (80-.). *366*, 1345–1349.

Kind, J., Pagie, L., de Vries, S.S., Nahidiazar, L., Dey, S.S., Bienko, M., Zhan, Y., Lajoie, B., de Graaf, C.A., Amendola, M., et al. (2015). Genome-wide maps of nuclear lamina interactions in single human cells. Cell *163*, 134–147.

Kinoshita, S., Su, L., Amano, M., Timmerman, L.A., Kaneshima, H., and Nolan, G.P. (1997). The T cell activation factor NF-ATc positively regulates HIV-1 replication and gene expression in T cells. Immunity *6*, 235–244.

Kitchen, N.S., and Schoenherr, C.J. (2010). Sumoylation modulates a domain in CTCF that activates transcription and decondenses chromatin. J. Cell. Biochem. *111*, 665–675.

Kizer, K.O., Phatnani, H.P., Shibata, Y., Hall, H., Greenleaf, A.L., and Strahl, B.D. (2005). A novel domain in Set2 mediates RNA polymerase II interaction and couples histone H3 K36 methylation with transcript elongation. Mol. Cell. Biol. *25*, 3305–3316.

Klemm, S.L., Shipony, Z., and Greenleaf, W.J. (2019). Chromatin accessibility and the regulatory epigenome. Nat. Rev. Genet. *20*, 207–220.

Klenova, E.M., Nicolas, R.H., Paterson, H.F., Carne, A.F., Heath, C.M., Goodwin, G.H., Neiman, P.E., and Lobanenkov, V. V (1993). CTCF, a conserved nuclear factor required for optimal transcriptional activity of the chicken c-myc gene, is an 11-Zn-finger protein differentially expressed in multiple forms. Mol. Cell. Biol. *13*, 7612–7624.

Klenova, E.M., Chernukhin, I. V, El-Kady, A., Lee, R.E., Pugacheva, E.M., Loukinov, D.I., Goodwin, G.H., Delgado, D., Filippova, G.N., León, J., et al. (2001). Functional phosphorylation sites in the C-terminal region of the multivalent multifunctional transcriptional factor CTCF. Mol. Cell. Biol. *21*, 2221–2234.

Ko, A., Kang, G., Hattler, J.B., Galadima, H.I., Zhang, J., Li, Q., and Kim, W.K. (2019). Macrophages but not Astrocytes Harbor HIV DNA in the Brains of HIV-1-Infected Aviremic Individuals on Suppressive Antiretroviral Therapy. J. Neuroimmune Pharmacol. *14*, 110–119.

Koh, Y., Wu, X., Ferris, A.L., Matreyek, K.A., Smith, S.J., Lee, K., KewalRamani, V.N., Hughes, S.H., and Engelman, A. (2013). Differential effects of human immunodeficiency virus type 1 capsid and cellular factors nucleoporin 153 and LEDGF/p75 on the efficiency and specificity of viral DNA integration. J Virol *87*, 648–658.

Koiwa, T., Hamano-Usami, A., Ishida, T., Okayama, A., Yamaguchi, K., Kamihira, S., and Watanabe, T. (2002). 5'-long terminal repeat-selective CpG methylation of latent human T-cell leukemia virus type 1 provirus in vitro and in vivo. J. Virol. *76*, 9389–9397.

Kok, Y.L., Vongrad, V., Shilaih, M., Di Giallonardo, F., Kuster, H., Kouyos, R., Günthard, H.F., and Metzner, K.J. (2016). Monocyte-derived macrophages exhibit distinct and more restricted HIV-1 integration site repertoire than CD4(+) T cells. Sci. Rep. *6*, 24157.

Kolodkin-Gal, D., Hulot, S.L., Korioth-Schmitz, B., Gombos, R.B., Zheng, Y., Owuor, J., Lifton, M.A., Ayeni, C., Najarian, R.M., Yeh, W.W., et al. (2013). Efficiency of cell-free and cell-associated virus in mucosal transmission of human immunodeficiency virus type 1 and simian immunodeficiency virus. J. Virol. *87*, 13589–13597.

Kondo, E., and Göttlinger, H.G. (1996). A conserved LXXLF sequence is the major determinant in p6gag required for the incorporation of human immunodeficiency virus type 1 Vpr. J. Virol. *70*, 159–164.

Kong, X., Ball, A.R.J., Pham, H.X., Zeng, W., Chen, H.-Y., Schmiesing, J.A., Kim, J.-S., Berns, M., and Yokomori, K. (2014). Distinct functions of human cohesin-SA1 and cohesin-SA2 in double-strand break repair. Mol. Cell. Biol. *34*, 685–698.

König, R., Zhou, Y., Elleder, D., Diamond, T.L., Bonamy, G.M.C., Irelan, J.T., Chiang, C. yuan, Tu, B.P., De Jesus, P.D., Lilley, C.E., et al. (2008). Global Analysis of Host-Pathogen Interactions that Regulate Early-Stage HIV-1 Replication. Cell *135*, 49–60.

Kouzarides, T. (2007). SnapShot: Histone-modifying enzymes. Cell *128*, 802.

Kraft-Terry, S.D., Buch, S.J., Fox, H.S., and Gendelman, H.E. (2009). A coat of many colors: neuroimmune crosstalk in human immunodeficiency virus infection. Neuron *64*, 133–145.

Krietenstein, N., Abraham, S., Venev, S. V, Abdennur, N., Gibcus, J., Hsieh, T.-H.S., Parsi, K.M., Yang, L., Maehr, R., Mirny, L.A., et al. (2020). Ultrastructural Details of Mammalian Chromosome Architecture. Mol. Cell *78*, 554-565.e7.

Kruize, Z., and Kootstra, N.A. (2019). The Role of Macrophages in HIV-1 Persistence and Pathogenesis. Front. Microbiol. *10*, 2828.

Kubo, N., Ishii, H., Xiong, X., Bianco, S., Meitinger, F., Hu, R., Hocker, J.D., Conte, M., Gorkin, D., Yu, M., et al. (2021). Promoter-proximal CTCF binding promotes distal enhancer-dependent gene activation. Nat. Struct. Mol. Biol. *28*, 152–161.

Kulkosky, J., Jones, K.S., Katz, R.A., Mack, J.P., and Skalka, A.M. (1992). Residues

critical for retroviral integrative recombination in a region that is highly conserved among retroviral/retrotransposon integrases and bacterial insertion sequence transposases. Mol Cell Biol *12*, 2331–2338.

Kulkosky, J., Nunnari, G., Otero, M., Calarota, S., Dornadula, G., Zhang, H., Malin, A., Sullivan, J., Xu, Y., DeSimone, J., et al. (2002). Intensification and Stimulation Therapy for Human Immunodeficiency Virus Type 1 Reservoirs in Infected Persons Receiving Virally Suppressive Highly Active Antiretroviral Therapy. J Infect Dis *186*, 1403–1411.

Kumar, A., Abbas, W., and Herbein, G. (2014). HIV-1 latency in monocytes/macrophages. Viruses *6*, 1837–1860.

Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M.J., et al. (2015). Integrative analysis of 111 reference human epigenomes. Nature *518*, 317–330.

Kurukuti, S., Tiwari, V.K., Tavoosidana, G., Pugacheva, E., Murrell, A., Zhao, Z., Lobanenkov, V., Reik, W., and Ohlsson, R. (2006). CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2. Proc. Natl. Acad. Sci. U. S. A. *103*, 10684–10689.

LaFave, M.C., Varshney, G.K., Gildea, D.E., Wolfsberg, T.G., Baxevanis, A.D., and Burgess, S.M. (2014). MLV integration site selection is driven by strong enhancers and active promoters. Nucleic Acids Res. *42*, 4257–4269.

Lara-Astiaso, D., Weiner, A., Lorenzo-Vivas, E., Zaretsky, I., Jaitin, D.A., David, E., Keren-Shaul, H., Mildner, A., Winter, D., Jung, S., et al. (2014). Immunogenetics. Chromatin state dynamics during blood formation. Science *345*, 943–949.

Lau, C.Y., Adan, M.A., and Maldarelli, F. (2021). Why the hiv reservoir never runs dry: Clonal expansion and the characteristics of hiv-infected cells challenge strategies to cure and control hiv infection. Viruses *13*.

Leavitt, A.D., Shiue, L., and Varmus, H.E. (1993). Site-directed mutagenesis of HIV-1 integrase demonstrates differential effects on integrase functions in vitro. J Biol Chem *268*, 2113–2119.

Lee, G.Q., and Lichterfeld, M. (2016). Diversity of HIV-1 reservoirs in CD4+ T-cell subpopulations. Curr. Opin. HIV AIDS *11*, 383–387.

Lee, M.S., and Craigie, R. (1994). Protection of retroviral DNA from autointegration: Involvement of a cellular factor. Proc. Natl. Acad. Sci. U. S. A. *91*, 9823–9827.

Lee, C.-K., Shibata, Y., Rao, B., Strahl, B.D., and Lieb, J.D. (2004). Evidence for nucleosome depletion at active regulatory regions genome-wide. Nat. Genet. *36*, 900–

905.

Lee, G.Q., Orlova-Fink, N., Einkauf, K., Chowdhury, F.Z., Sun, X., Harrington, S., Kuo, H.-H., Hua, S., Chen, H.-R., Ouyang, Z., et al. (2017). Clonal expansion of genome-intact HIV-1 in functionally polarized Th1 CD4+ T cells. J. Clin. Invest. *127*, 2689–2696.

Lee, K., Ambrose, Z., Martin, T.D., Oztop, I., Mulky, A., Julias, J.G., Vandegraaff, N., Baumann, J.G., Wang, R., Yuen, W., et al. (2010). Flexible use of nuclear import pathways by HIV-1. Cell Host Microbe *7*, 221–233.

Lee, K., Mulky, A., Yuen, W., Martin, T.D., Meyerson, N.R., Choi, L., Yu, H., Sawyer, S.L., and Kewalramani, V.N. (2012). HIV-1 capsid-targeting domain of cleavage and polyadenylation specificity factor 6. J Virol *86*, 3851–3860.

Lee, Y.B., Nagai, A., and Kim, S.U. (2002). Cytokines, chemokines, and cytokine receptors in human microglia. J Neurosci Res *69*, 94–103.

Lehman, B.J., Lopez-Diaz, F.J., Santisakultarm, T.P., Fang, L., Shokhirev, M.N., Diffenderfer, K.E., Manor, U., and Emerson, B.M. (2021). Dynamic regulation of CTCF stability and subnuclear localization in response to stress.

Lenasi, T., Contreras, X., and Peterlin, B.M. (2008). Transcriptional Interference Antagonizes Proviral Gene Expression to Promote HIV Latency. Cell Host Microbe *4*, 123–133.

Lensch, S., Herschl, M.H., Ludwig, C.H., Sinha, J., Hinks, M.M., Mukund, A., Fujimori, T., and Bintu, L. (2022). Dynamic spreading of chromatin-mediated gene silencing and reactivation between neighboring genes in single cells. Elife *11*.

Lewinski, M.K., Yamashita, M., Emerman, M., Ciuffi, A., Marshall, H., Crawford, G., Collins, F., Shinn, P., Leipzig, J., Hannenhalli, S., et al. (2006). Retroviral DNA integration: viral and cellular determinants of target-site selection. PLoS Pathog. *2*, e60.

Li, C., Guan, X., Du, T., Jin, W., Wu, B., Liu, Y., Wang, P., Hu, B., Griffin, G.E., Shattock, R.J., et al. (2015). Inhibition of HIV-1 infection of primary CD4+ T-cells by gene editing of CCR5 using adenovirus-delivered CRISPR/Cas9. J. Gen. Virol. *96*, 2381–2393.

Li, C., Mousseau, G., and Valente, S.T. (2019). Tat inhibition by didehydro-Cortistatin A promotes heterochromatin formation at the HIV-1 long terminal repeat. Epigenetics Chromatin *12*, 23.

Li, J.Z., Etemad, B., Ahmed, H., Aga, E., Bosch, R.J., Mellors, J.W., Kuritzkes, D.R.,

Lederman, M.M., Para, M., and Gandhi, R.T. (2016). The size of the expressed HIV reservoir predicts timing of viral rebound after treatment interruption. AIDS *30*, 343–353.

Li, L., Olvera, J.M., Yoder, K.E., Mitchell, R.S., Butler, S.L., Lieber, M., Martin, S.L., and Bushman, F.D. (2001). Role of the non-homologous DNA end joining pathway in the early steps of retroviral infection. Embo J *20*, 3272–3281.

Li, Q., Zhang, M., Han, H., Rohde, A., and Stamatoyannopoulos, G. (2002). Evidence that DNase I hypersensitive site 5 of the human beta-globin locus control region functions as a chromosomal insulator in transgenic mice. Nucleic Acids Res. *30*, 2484–2491.

Li, S., Holguin, L., and Burnett, J.C. (2022). CRISPR-Cas9-mediated gene disruption of HIV-1 co-receptors confers broad resistance to infection in human T cells and humanized mice. Mol. Ther. Methods Clin. Dev. *24*, 321–331.

Li, W., Singh, P.K., Sowd, G.A., Bedwell, G.J., Jang, S., Achuthan, V., Oleru, A. V., Wong, D., Fadel, H.J., Lee, K., et al. (2020a). CPSF6-dependent targeting of speckle-associated domains distinguishes primate from nonprimate lentiviral integration. MBio *11*, 1–20.

Li, Y., Haarhuis, J.H.I., Sedeño Cacciatore, Á., Oldenkamp, R., van Ruiten, M.S., Willems, L., Teunissen, H., Muir, K.W., de Wit, E., Rowland, B.D., et al. (2020b). The structural basis for cohesin–CTCF-anchored loops. Nature *578*, 472–476.

Lian, X., Seiger, K.W., Parsons, E.M., Gao, C., Sun, W., Gladkov, G.T., Roseto, I.C., Einkauf, K.B., Osborn, M.R., Chevalier, J.M., et al. (2023). Progressive transformation of the HIV-1 reservoir cell profile over two decades of antiviral therapy. Cell Host Microbe *31*, 83-96.e5.

Lichterfeld, M., Gao, C., and Yu, X.G. (2022). An ordeal that does not heal: understanding barriers to a cure for HIV-1 infection. Trends Immunol. *43*, 608–616.

Lieberman-aiden, E., Berkum, N.L. Van, Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). of the Human Genome. *33292*, 289–294.

Van Lint, C., Bouchat, S., and Marcello, A. (2013). HIV-1 transcription and latency: an update. Retrovirology *10*, 67.

Liszewski, M.K., Yu, J.J., and O'Doherty, U. (2009). Detecting HIV-1 integration by repetitive-sampling Alu-gag PCR. Methods *47*, 254–260.

Liu, R., Yeh, Y.-H.J., Varabyou, A., Collora, J.A., Sherrill-Mix, S., Talbot, C.C.J., Mehta,

S., Albrecht, K., Hao, H., Zhang, H., et al. (2020). Single-cell transcriptional landscapes reveal HIV-1-driven aberrant host gene transcription as a potential therapeutic target. Sci. Transl. Med. *12*.

Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods *25*, 402–408.

Livelli, A., Vaida, F., Ellis, R.J., Ma, Q., Ferrara, M., Clifford, D.B., Collier, A.C., Gelman, B.B., Marra, C.M., McArthur, J.C., et al. (2019). Correlates of HIV RNA concentrations in cerebrospinal fluid during antiretroviral therapy: a longitudinal cohort study. Lancet. HIV *6*, e456–e462.

Llano, M., Vanegas, M., Fregoso, O., Saenz, D., Chung, S., Peretz, M., and Poeschla, E.M. (2004). LEDGF/p75 determines cellular trafficking of diverse lentiviral but not murine oncoretroviral integrase proteins and is a component of functional lentiviral preintegration complexes. J Virol *78*, 9524–9537.

Llano, M., Vanegas, M., Hutchins, N., Thompson, D., Delgado, S., and Poeschla, E.M. (2006). Identification and characterization of the chromatin-binding domains of the HIV-1 integrase interactor LEDGF/p75. J Mol Biol *360*, 760–773.

Lorenzi, J.C.C., Cohen, Y.Z., Cohn, L.B., Kreider, E.F., Barton, J.P., Learn, G.H., Oliveira, T., Lavine, C.L., Horwitz, J.A., Settler, A., et al. (2016). 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. *113*, E7908–E7916.

Lu, H., Li, Z., Xue, Y., Schulze-Gahmen, U., Johnson, J.R., Krogan, N.J., Alber, T., and Zhou, Q. (2014). AFF1 is a ubiquitous P-TEFb partner to enable Tat extraction of P-TEFb from 7SK snRNP and formation of SECs for HIV transactivation. Proc. Natl. Acad. Sci. U. S. A. *111*, E15—24.

Luan, J., Xiang, G., Gómez-García, P.A., Tome, J.M., Zhang, Z., Vermunt, M.W., Zhang, H., Huang, A., Keller, C.A., Giardine, B.M., et al. (2021). Distinct properties and functions of CTCF revealed by a rapidly inducible degron system. Cell Rep. *34*.

Luan, J., Vermunt, M.W., Syrett, C.M., Coté, A., Tome, J.M., Zhang, H., Huang, A., Luppino, J.M., Keller, C.A., Giardine, B.M., et al. (2022). CTCF blocks antisense transcription initiation at divergent promoters. Nat. Struct. Mol. Biol. *29*, 1136–1144.

Luban, J., Bossolt, K.L., Franke, E.K., Kalpana, G. V, and Goff, S.P. (1993). Human immunodeficiency virus type 1 Gag protein binds to cyclophilins A and B. Cell *73*,

1067–1078.

Lucic, B., and Lusic, M. (2016). Connecting HIV-1 integration and transcription: a step toward new treatments. FEBS Lett *590*, 1927–1939.

Lucic, B., Chen, H.-C., Kuzman, M., Zorita, E., Wegner, J., Minneker, V., Wang, W., Fronza, R., Laufs, S., Schmidt, M., et al. (2019). Spatially clustered loci with multiple enhancers are frequent targets of HIV-1 integration. Nat. Commun. *10*, 4059.

Lucic, B., de Castro, I.J., and Lusic, M. (2021). Viruses in the Nucleus. Cold Spring Harb. Perspect. Biol. *13*.

Luco, R.F., Pan, Q., Tominaga, K., Blencowe, B.J., Pereira-Smith, O.M., and Misteli, T. (2010). Regulation of alternative splicing by histone modifications. Science *327*, 996–1000.

Luo, H., Yu, Q., Liu, Y., Tang, M., Liang, M., Zhang, D., Xiao, T.S., Wu, L., Tan, M., Ruan, Y., et al. (2020). LATS kinase–mediated CTCF phosphorylation and selective loss of genomic binding. Sci. Adv. *6*.

Lupiáñez, D.G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J.M., Laxova, R., et al. (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell *161*, 1012–1025.

Luppino, J.M., Field, A., Nguyen, S.C., Park, D.S., Shah, P.P., Abdill, R.J., Lan, Y., Yunker, R., Jain, R., Adelman, K., et al. (2022). Co-depletion of NIPBL and WAPL balance cohesin activity to correct gene misexpression. PLOS Genet. *18*, e1010528.

Lusic, M., and Siliciano, R.F. (2017). Nuclear landscape of HIV-1 infection and integration. Nat Rev Microbiol *15*, 69–82.

Lusic, M., Marcello, A., Cereseto, A., and Giacca, M. (2003). Regulation of HIV-1 gene expression by histone acetylation and factor recruitment at the LTR promoter. EMBO J. *22*, 6550–6561.

Ma, S., Zhang, B., LaFave, L.M., Earl, A.S., Chiang, Z., Hu, Y., Ding, J., Brack, A., Kartha, V.K., Tay, T., et al. (2020). Chromatin Potential Identified by Shared Single-Cell Profiling of RNA and Chromatin. Cell *183*, 1103-1116.e20.

Machida, S., Depierre, D., Chen, H.-C., Thenin-Houssier, S., Petitjean, G., Doyen, C.M., Takaku, M., Cuvier, O., and Benkirane, M. (2020). Exploring histone loading on HIV DNA reveals a dynamic nucleosome positioning between unintegrated and integrated viral genome. Proc. Natl. Acad. Sci. U. S. A. *117*, 6822–6830.

MacPherson, M.J., Beatty, L.G., Zhou, W., Du, M., and Sadowski, P.D. (2009). The

CTCF insulator protein is posttranslationally modified by SUMO. Mol. Cell. Biol. *29*, 714–725.

Madani Tonekaboni, S.A., Haibe-Kains, B., and Lupien, M. (2021). Large organized chromatin lysine domains help distinguish primitive from differentiated cell populations. Nat. Commun. *12*, 1–9.

Madsen, J.G.S., Madsen, M.S., Rauch, A., Traynor, S., Van Hauwaert, E.L., Haakonsson, A.K., Javierre, B.M., Hyldahl, M., Fraser, P., and Mandrup, S. (2020). Highly interconnected enhancer communities control lineage-determining genes in human mesenchymal stem cells. Nat. Genet. *52*, 1227–1238.

Van Maele, B., Busschots, K., Vandekerckhove, L., Christ, F., and Debyser, Z. (2006). Cellular co-factors of HIV-1 integration. Trends Biochem Sci *31*, 98–105.

Maertens, G., Cherepanov, P., Pluymers, W., Busschots, K., De Clercq, E., Debyser, Z., and Engelborghs, Y. (2003). LEDGF/p75 is essential for nuclear and chromosomal targeting of HIV-1 integrase in human cells. J Biol Chem *278*, 33528–33539.

Maertens, G.N., Hare, S., and Cherepanov, P. (2010). The mechanism of retroviral integration from X-ray structures of its key intermediates. Nature *468*, 326–329.

Mahajan, S.D., Ordain, N.S., Kutscher, H., Karki, S., and Reynolds, J.L. (2021). HIV Neuroinflammation: The Role of Exosomes in Cell Signaling, Prognostic and Diagnostic Biomarkers and Drug Delivery . Front. Cell Dev. Biol. *9*.

Maldarelli, F., Wu, X., Su, L., Simonetti, F.R., Shao, W., Hill, S., Spindler, J., Ferris, A.L., Mellors, J.W., Kearney, M.F., et al. (2014). HIV latency. Specific HIV integration sites are linked to clonal expansion and persistence of infected cells. Science (80-.). *345*, 179–183.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-guided human genome engineering via Cas9. Science (80-.). *339*, 823–826.

Malim, M.H., Böhnlein, S., Hauber, J., and Cullen, B.R. (1989). Functional dissection of the HIV-1 Rev trans-activator—Derivation of a trans-dominant repressor of Rev function. Cell *58*, 205–214.

Marban, C., Forouzanfar, F., Ait-Ammar, A., Fahmi, F., El Mekdad, H., Daouad, F., Rohr, O., and Schwartz, C. (2016). Targeting the Brain Reservoirs: Toward an HIV Cure. Front. Immunol. *7*.

Marcello, A., Ferrari, A., Pellegrini, V., Pegoraro, G., Lusic, M., Beltram, F., and Giacca, M. (2003). Recruitment of human cyclin T1 to nuclear bodies through direct interaction

with the PML protein. Embo J *22*, 2156–2166.

Margueron, R., Justin, N., Ohno, K., Sharpe, M.L., Son, J., Drury, W.J. 3rd, Voigt, P., Martin, S.R., Taylor, W.R., De Marco, V., et al. (2009). Role of the polycomb protein EED in the propagation of repressive histone marks. Nature *461*, 762–767.

Marini, B., Kertesz-Farkas, A., Ali, H., Lucic, B., Lisek, K., Manganaro, L., Pongor, S., Luzzati, R., Recchia, A., Mavilio, F., et al. (2015). Nuclear architecture dictates HIV-1 integration site selection. Nature *521*, 227–231.

Marmorstein, R., and Roth, S.Y. (2001). Histone acetyltransferases: function, structure, and catalysis. Curr. Opin. Genet. Dev. *11*, 155–161.

Martin, E.W., Rodriguez y Baena, A., Reggiardo, R.E., Worthington, A.K., Mattingly, C.S., Poscablo, D.M., Krietsch, J., McManus, M.T., Carpenter, S., Kim, D.H., et al. (2023). Dynamics of Chromatin Accessibility During Hematopoietic Stem Cell Differentiation Into Progressively Lineage-Committed Progeny. Stem Cells *41*, 520–539.

Marzio, G., Tyagi, M., Gutierrez, M.I., and Giacca, M. (1998). HIV-1 tat transactivator recruits p300 and CREB-binding protein histone acetyltransferases to the viral promoter. Proc. Natl. Acad. Sci. U. S. A. *95*, 13519–13524.

Mateo, L.J., Murphy, S.E., Hafner, A., Cinquini, I.S., Walker, C.A., and Boettiger, A.N. (2019). Visualizing DNA folding and RNA in embryos at single-cell resolution. Nature *568*, 49–54.

Matreyek, K.A., and Engelman, A. (2011). The requirement for nucleoporin NUP153 during human immunodeficiency virus type 1 infection is determined by the viral capsid. J. Virol. *85*, 7818–7827.

Matreyek, K.A., and Engelman, A. (2013). Viral and cellular requirements for the nuclear entry of retroviral preintegration nucleoprotein complexes. Viruses *5*, 2483–2511.

Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J., et al. (2012). Systematic localization of common disease-associated variation in regulatory DNA. Science *337*, 1190–1195.

McArthur, E., and Capra, J.A. (2021). Topologically associating domain boundaries that are stable across diverse cell types are evolutionarily constrained and enriched for heritability. Am. J. Hum. Genet. *108*, 269–283.

McCarthy, M., He, J., and Wood, C. (1998). HIV-1 strain-associated variability in infection of primary neuroglia. J. Neurovirol. *4*, 80–89.

McClure, M.O., Marsh, M., and Weiss, R.A. (1988). Human immunodeficiency virus infection of CD4-bearing cells occurs by a pH-independent mechanism. EMBO J. *7*, 513–518.

Melamed, A., Yaguchi, H., Miura, M., Witkover, A., Fitzgerald, T.W., Birney, E., and Bangham, C.R. (2018). The human leukemia virus HTLV-1 alters the structure and transcription of host chromatin in cis. Elife *7*.

Melikyan, G.B. (2008). Common principles and intermediates of viral protein-mediated fusion: the HIV-1 paradigm. Retrovirology *5*, 111.

Mendoza, P., Jackson, J.R., Oliveira, T.Y., Gaebler, C., Ramos, V., Caskey, M., Jankovic, M., Nussenzweig, M.C., and Cohn, L.B. (2020). Antigen-responsive CD4+ T cell clones contribute to the HIV-1 latent reservoir. J. Exp. Med. *217*.

Mieczkowski, J., Cook, A., Bowman, S.K., Mueller, B., Alver, B.H., Kundu, S., Deaton, A.M., Urban, J.A., Larschan, E., Park, P.J., et al. (2016). MNase titration reveals differences between nucleosome occupancy and chromatin accessibility. Nat. Commun. *7*, 11485.

Milde-Langosch, K. (2005). The Fos family of transcription factors and their role in tumourigenesis. Eur. J. Cancer *41*, 2449–2461.

Miller, M.D., Wang, B., and Bushman, F.D. (1995). Human immunodeficiency virus type 1 preintegration complexes containing discontinuous plus strands are competent to integrate in vitro. J. Virol. *69*, 3938–3944.

Miller, M.D., Farnet, C.M., and Bushman, F.D. (1997). Human immunodeficiency virus type 1 preintegration complexes: studies of organization and composition. J. Virol. *71*, 5382–5390.

Mirny, L.A., Imakaev, M., and Abdennur, N. (2019). Two major mechanisms of chromosome organization. Curr. Opin. Cell Biol. *58*, 142–152.

Misteli, T. (2020). The Self-Organizing Genome: Principles of Genome Architecture and Function. Cell *183*, 28–45.

Mitchell, R.S., Beitzel, B.F., Schroder, A.R., Shinn, P., Chen, H., Berry, C.C., Ecker, J.R., and Bushman, F.D. (2004). Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. PLoS Biol *2*, E234.

Moreau, P., Cournac, A., Palumbo, G.A., Marbouty, M., Mortaza, S., Thierry, A., Cairo, S., Lavigne, M., Koszul, R., and Neuveut, C. (2018). Tridimensional infiltration of DNA viruses into the host genome shows preferential contact with active chromatin. Nat. Commun. *9*.

Morón-López, S., Puertas, M.C., Gálvez, C., Navarro, J., Carrasco, A., Esteve, M., Manyé, J., Crespo, M., Salgado, M., and Martinez-Picado, J. (2017). Sensitive quantification of the HIV-1 reservoir in gut-associated lymphoid tissue. PLoS One *12*, e0175899.

Moso, M.A., Anderson, J.L., Adikari, S., Gray, L.R., Khoury, G., Chang, J.J., Jacobson, J.C., Ellett, A.M., Cheng, W.J., Saleh, S., et al. (2019). HIV latency can be established in proliferating and nonproliferating resting CD4+ T cells in vitro: implications for latency reversal. Aids *33*, 199–209.

Mousseau, G., Kessing, C.F., Fromentin, R., Trautmann, L., Chomont, N., and Valente, S.T. (2015). The Tat Inhibitor Didehydro-Cortistatin A Prevents HIV-1 Reactivation from Latency. MBio *6*, e00465.

Mueller, B., Mieczkowski, J., Kundu, S., Wang, P., Sadreyev, R., Tolstorukov, M.Y., and Kingston, R.E. (2017). Widespread changes in nucleosome accessibility without changes in nucleosome occupancy during a rapid transcriptional induction. Genes Dev. *31*, 451–462.

Muller, T.G., Zila, V., Muller, B., and Krausslich, H.G. (2022). Nuclear Capsid Uncoating and Reverse Transcription of HIV-1. Annu. Rev. Virol. *9*, 261–284.

Müller, T.G., Zila, V., Peters, K., Schifferdecker, S., Stanic, M., Lucic, B., Laketa, V., Lusic, M., Müller, B., and Kräusslich, H.-G. (2021). HIV-1 uncoating by release of viral cDNA from capsid-like structures in the nucleus of infected cells. Elife *10*, e64776.

Murooka, T.T., Deruaz, M., Marangoni, F., Vrbanac, V.D., Seung, E., von Andrian, U.H., Tager, A.M., Luster, A.D., and Mempel, T.R. (2012). HIV-infected T cells are migratory vehicles for viral dissemination. Nature *490*, 283–287.

MURRAY, K. (1964). THE OCCURRENCE OF EPSILON-N-METHYL LYSINE IN HISTONES. Biochemistry *3*, 10–15.

Nabel, G., and Baltimore, D. (1987). An inducible transcription factor activates expression of human immunodeficiency virus in T cells. Nature *326*, 711–713.

Nagano, T., Lubling, Y., Stevens, T.J., Schoenfelder, S., Yaffe, E., Dean, W., Laue, E.D., Tanay, A., and Fraser, P. (2013). Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. Nature *502*, 59–64.

Nakahashi, H., Kieffer Kwon, K.-R., Resch, W., Vian, L., Dose, M., Stavreva, D., Hakim, O., Pruett, N., Nelson, S., Yamane, A., et al. (2013). A genome-wide map of CTCF multivalency redefines the CTCF code. Cell Rep. *3*, 1678–1689.

Nanni, L., Ceri, S., and Logie, C. (2020). Spatial patterns of CTCF sites define the

anatomy of TADs and their boundaries. Genome Biol. *21*, 1–25.

Narendra, V., Rocha, P.P., An, D., Raviram, R., Skok, J.A., Mazzoni, E.O., and Reinberg, D. (2015). CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. Science (80-.). *347*, 1017–1021.

Narendra, V., Bulajić, M., Dekker, J., Mazzoni, E.O., and Reinberg, D. (2016). CTCF-mediated topological boundaries during development foster appropriate gene regulation. Genes Dev. *30*, 2657–2662.

Narezkina, A., Taganov, K.D., Litwin, S., Stoyanova, R., Hayashi, J., Seeger, C., Skalka, A.M., and Katz, R.A. (2004). Genome-wide analyses of avian sarcoma virus integration sites. J. Virol. *78*, 11656–11663.

Nasser, J., Bergman, D.T., Fulco, C.P., Guckelberger, P., Doughty, B.R., Patwardhan, T.A., Jones, T.R., Nguyen, T.H., Ulirsch, J.C., Lekschas, F., et al. (2021). Genome-wide enhancer maps link risk variants to disease genes. Nature *593*, 238–243.

Ne, E., Palstra, R.-J., and Mahmoudi, T. (2018). Transcription: Insights From the HIV-1 Promoter. Int. Rev. Cell Mol. Biol. *335*, 191–243.

Neuen-Jacob, E., Arendt, G., Wendtland, B., Jacob, B., Schneeweis, M., and Wechsler, W. (1993). Frequency and topographical distribution of CD68-positive macrophages and HIV-1 core proteins in HIV-associated brain lesions. Clin. Neuropathol. *12*, 315–324.

Newhard, W., Patel, M., Cassaday, J., Ballard, J., Squadroni, B., Wu, G., Liu, J., Yu, W., Kozlowski, J., Zuck, P., et al. (2021). In Vitro Pharmacokinetic/Pharmacodynamic Modeling of HIV Latency Reversal by Novel HDAC Inhibitors Using an Automated Platform. SLAS Discov. *26*, 642–654.

Nguyen, K., Dobrowolski, C., Shukla, M., Cho, W.-K., Luttge, B., and Karn, J. (2021). Inhibition of the H3K27 demethylase UTX enhances the epigenetic silencing of HIV proviruses and induces HIV-1 DNA hypermethylation but fails to permanently block HIV reactivation. PLoS Pathog. *17*, e1010014.

Nimmerjahn, A., Kirchhoff, F., and Helmchen, F. (2005). Resting microglial cells are highly dynamic surveillants of brain parenchyma in vivo. Science (80-.). *308*, 1314–1318.

Nishiyama, T. (2019). Cohesion and cohesin-dependent chromatin organization. Curr. Opin. Cell Biol. *58*, 8–14.

Nishizawa, Y., Usukura, J., Singh, D.P., Chylack Jr., L.T., and Shinohara, T. (2001). Spatial and temporal dynamics of two alternatively spliced regulatory factors, lens

epithelium-derived growth factor (ledgf/p75) and p52, in the nucleus. Cell Tissue Res *305*, 107–114.

Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., Van Berkum, N.L., Meisig, J., Sedat, J., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature *485*, 381–385.

Nora, E.P., Goloborodko, A., Valton, A.-L., Gibcus, J.H., Uebersohn, A., Abdennur, N., Dekker, J., Mirny, L.A., and Bruneau, B.G. (2017). Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. Cell *169*, 930-944.e22.

Nora, E.P., Caccianini, L., Fudenberg, G., So, K., Kameswaran, V., Nagle, A., Uebersohn, A., Hajj, B., Saux, A. Le, Coulon, A., et al. (2020). Molecular basis of CTCF binding polarity in genome folding. Nat. Commun. *11*, 5612.

Nott, A., Holtman, I.R., Coufal, N.G., Schlachetzki, J.C.M., Yu, M., Hu, R., Han, C.Z., Pena, M., Xiao, J., Wu, Y., et al. (2019). Brain cell type–specific enhancer–promoter interactome maps and disease-risk association. Science (80-.). *366*, 1134.

Nuebler, J., Fudenberg, G., Imakaev, M., Abdennur, N., and Mirny, L.A. (2018). Chromatin organization by an interplay of loop extrusion and compartmental segregation. Proc. Natl. Acad. Sci. U. S. A. *115*, E6697–E6706.

van Nuland, R., van Schaik, F.M., Simonis, M., van Heesch, S., Cuppen, E., Boelens, R., Timmers, H.M., and van Ingen, H. (2013). Nucleosomal DNA binding drives the recognition of H3K36-methylated nucleosomes by the PSIP1-PWWP domain. Epigenetics Chromatin *6*, 12.

Di Nunzio, F., Fricke, T., Miccio, A., Valle-Casuso, J.C., Perez, P., Souque, P., Rizzi, E., Severgnini, M., Mavilio, F., Charneau, P., et al. (2013). Nup153 and Nup98 bind the HIV-1 core and contribute to the early steps of HIV-1 replication. Virology *440*, 8–18.

Nyborg, J.K., Egan, D., and Sharma, N. (2010). The HTLV-1 Tax protein: revealing mechanisms of transcriptional activation through histone acetylation and nucleosome disassembly. Biochim. Biophys. Acta *1799*, 266–274.

Oberlin, E., Amara, A., Bachelerie, F., Bessia, C., Virelizier, J.L., Arenzana-Seisdedos, F., Schwartz, O., Heard, J.M., Clark-Lewis, I., Legler, D.F., et al. (1996). The CXC chemokine SDF-1 is the ligand for LESTR/fusin and prevents infection by T-cell-line-adapted HIV-1. Nature *382*, 833–835.

Ocwieja, K.E., Sherrill-Mix, S., Mukherjee, R., Custers-Allen, R., David, P., Brown, M., Wang, S., Link, D.R., Olson, J., Travers, K., et al. (2012). Dynamic regulation of HIV-1 mRNA populations analyzed by single-molecule enrichment and long-read sequencing. Nucleic Acids Res. *40*, 10345–10355.

Omenn, G.S., Lane, L., Overall, C.M., Corrales, F.J., Schwenk, J.M., Paik, Y.-K., Van Eyk, J.E., Liu, S., Snyder, M., Baker, M.S., et al. (2018). Progress on Identifying and Characterizing the Human Proteome: 2018 Metrics from the HUPO Human Proteome Project. J. Proteome Res. *17*, 4031–4041.

Oomen, M.E., Hansen, A.S., Liu, Y., Darzacq, X., and Dekker, J. (2019). CTCF sites display cell cycle-dependent dynamics in factor binding and nucleosome positioning. Genome Res. *29*, 236–249.

Pace, M.J., Graf, E.H., Agosto, L.M., Mexas, A.M., Male, F., Brady, T., Bushman, F.D., and O'Doherty, U. (2012). Directly infected resting CD4+T cells can produce HIV Gag without spreading infection in a model of HIV latency. PLoS Pathog. *8*, e1002818.

Pai, A., and Weinberger, L.S. (2017). Fate-Regulating Circuits in Viruses: From Discovery to New Therapy Targets. Annu. Rev. Virol. *4*, 469–490.

Paik, W.K., and Kim, S. (1969). Enzymatic methylation of histones. Arch. Biochem. Biophys. *134*, 632–637.

Palecek, J.J. (2019). SMC5/6: Multifunctional Player in Replication. Genes (Basel). *10*.

Panet, A., and Cedar, H. (1977). Selective degradation of integrated murine leukemia proviral DNA by deoxyribonucleases. Cell *11*, 933–940.

Paris, C., Pentland, I., Groves, I., Roberts, D.C., Powis, S.J., Coleman, N., Roberts, S., and Parish, J.L. (2015). CCCTC-binding factor recruitment to the early region of the human papillomavirus 18 genome regulates viral oncogene expression. J. Virol. *89*, 4770–4785.

Park, R.J., Wang, T., Koundakjian, D., Hultquist, J.F., Lamothe-Molina, P., Monel, B., Schumann, K., Yu, H., Krupzcak, K.M., Garcia-Beltran, W., et al. (2017). A genome-wide CRISPR screen identifies a restricted set of HIV host dependency factors. Nat. Genet. *49*, 193–203.

Passos, D.O., Li, M., Yang, R., Rebensburg, S. V, Ghirlando, R., Jeon, Y., Shkriabai, N., Kvaratskhelia, M., Craigie, R., and Lyumkis, D. (2017). Cryo-EM structures and atomic model of the HIV-1 strand transfer complex intasome. Science *355*, 89–92.

Passos, D.O., Li, M., Jóźwik, I.K., Zhao, X.Z., Santos-Martins, D., Yang, R., Smith, S.J., Jeon, Y., Forli, S., Hughes, S.H., et al. (2020). Structural basis for strand-transfer

inhibitor binding to HIV intasomes. Science *367*, 810–814.

Paul, M.R., Hochwagen, A., and Ercan, S. (2019). Condensin action and compaction. Curr. Genet. *65*, 407–415.

Peeters, M., Honoré, C., Huet, T., Bedjabaga, L., Ossari, S., Bussi, P., Cooper, R.W., and Delaporte, E. (1989). Isolation and partial characterization of an HIV-related virus occurring naturally in chimpanzees in Gabon. AIDS *3*, 625–630.

Peluso, M.J., Bacchetti, P., Ritter, K.D., Beg, S., Lai, J., Martin, J.N., Hunt, P.W., Henrich, T.J., Siliciano, J.D., Siliciano, R.F., et al. (2020). Differential decay of intact and defective proviral DNA in HIV-1-infected individuals on suppressive antiretroviral therapy. JCI Insight *5*.

Pentland, I., Campos-León, K., Cotic, M., Davies, K.-J., Wood, C.D., Groves, I.J., Burley, M., Coleman, N., Stockton, J.D., Noyvert, B., et al. (2018). Disruption of CTCF-YY1-dependent looping of the human papillomavirus genome activates differentiation-induced viral oncogene transcription. PLoS Biol. *16*, e2005752.

Perez, E.E., Wang, J., Miller, J.C., Jouvenot, Y., Kim, K.A., Liu, O., Wang, N., Lee, G., Bartsevich, V. V, Lee, Y.L., et al. (2008). Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. Nat Biotechnol *26*, 808–816.

Perkins, K.J., Lusic, M., Mitar, I., Giacca, M., and Proudfoot, N.J. (2008). Transcription-dependent gene looping of the HIV-1 provirus is dictated by recognition of pre-mRNA processing signals. Mol. Cell *29*, 56–68.

Peterson, J.J., Lewis, C.A., Burgos, S.D., Manickam, A., Xu, Y., Rowley, A.A., Clutton, G., Richardson, B., Zou, F., Simon, J.M., et al. (2022). A histone deacetylase network regulates epigenetic reprogramming and viral silencing in HIV infected cells. BioRxiv 2022.05.09.491199.

Philip, S., Zahoor, M.A., Zhi, H., Ho, Y.-K., and Giam, C.-Z. (2014). Regulation of Human T-Lymphotropic Virus Type I Latency and Reactivation by HBZ and Rex. PLOS Pathog. *10*, e1004040.

Phillips, J.E., and Corces, V.G. (2009). CTCF: master weaver of the genome. Cell *137*, 1194–1211.

Pierson, T.C., Zhou, Y., Kieffer, T.L., Ruff, C.T., Buck, C., and Siliciano, R.F. (2002). Molecular characterization of preintegration latency in human immunodeficiency virus type 1 infection. J. Virol. *76*, 8518–8531.

Pinzone, M.R., VanBelzen, D.J., Weissman, S., Bertuccio, M.P., Cannon, L., Venanzi-Rullo, E., Migueles, S., Jones, R.B., Mota, T., Joseph, S.B., et al. (2019). Longitudinal

HIV sequencing reveals reservoir expression leading to decay which is obscured by clonal expansion. Nat. Commun. *10*, 728.

Pizzato, M., Erlwein, O., Bonsall, D., Kaye, S., Muir, D., and McClure, M.O. (2009). A one-step SYBR Green I-based product-enhanced reverse transcriptase assay for the quantitation of retroviruses in cell culture supernatants. J. Virol. Methods *156*, 1–7.

Plaza-Jennings, A.L., Valada, A., O'Shea, C., Iskhakova, M., Hu, B., Javidfar, B., Hutta, G. Ben, Lambert, T., Murray, J., Kassim, B., et al. (2022). HIV integration in the human brain is linked to microglial activation and 3D genome remodeling. Mol. Cell *82*, 2022.05.03.490485.

Pluta, A., Jaworski, J.P., and Douville, R.N. (2020). Regulation of Expression and Latency in BLV and HTLV. Viruses *12*.

Poiesz, B.J., Ruscetti, F.W., Gazdar, A.F., Bunn, P.A., Minna, J.D., and Gallo, R.C. (1980). Detection and isolation of type C retrovirus particles from fresh and cultured lymphocytes of a patient with cutaneous T-cell lymphoma. Proc. Natl. Acad. Sci. U. S. A. *77*, 7415–7419.

Poirier, M.G., Bussiek, M., Langowski, J., and Widom, J. (2008). Spontaneous access to DNA target sites in folded chromatin fibers. J. Mol. Biol. *379*, 772–786.

Pombo, A., and Dillon, N. (2015). Three-dimensional genome architecture: players and mechanisms. Nat. Rev. Mol. Cell Biol. *16*, 245.

Poon, B., Chang, M.A., and Chen, I.S.Y. (2007). Vpr is required for efficient Nef expression from unintegrated human immunodeficiency virus type 1 DNA. J. Virol. *81*, 10515–10523.

Pradeepa, M.M., Sutherland, H.G., Ule, J., Grimes, G.R., and Bickmore, W.A. (2012). Psip1/Ledgf p52 binds methylated histone H3K36 and splicing factors and contributes to the regulation of alternative splicing. PLoS Genet *8*, e1002717.

Priller, J., and Prinz, M. (2019). Targeting microglia in brain disorders. Science (80-.). *365*, 32.

Probst, A. V, Dunleavy, E., and Almouzni, G. (2009). Epigenetic inheritance during the cell cycle. Nat. Rev. Mol. Cell Biol. *10*, 192–206.

Proietti, F.A., Carneiro-Proietti, A.B.F., Catalan-Soares, B.C., and Murphy, E.L. (2005). Global epidemiology of HTLV-I infection and associated diseases. Oncogene *24*, 6058–6068.

Pugacheva, E.M., Kubo, N., Loukinov, D., Tajmul, M., Kang, S., Kovalchuk, A.L., Strunnikov, A. V, Zentner, G.E., Ren, B., and Lobanenkov, V. V (2020). CTCF

mediates chromatin looping via N-terminal domain-dependent cohesin retention. Proc. Natl. Acad. Sci. U. S. A. *117*, 2020–2031.

Purcell, D.F., and Martin, M.A. (1993). Alternative splicing of human immunodeficiency virus type 1 mRNA modulates viral protein expression, replication, and infectivity. J. Virol. *67*, 6365–6378.

Qi, Q., Cheng, L., Tang, X., He, Y., Li, Y., Yee, T., Shrestha, D., Feng, R., Xu, P., Zhou, X., et al. (2021). Dynamic CTCF binding directly mediates interactions among cis-regulatory elements essential for hematopoiesis. Blood *137*, 1327–1339.

Quinodoz, S.A., Ollikainen, N., Tabak, B., Palla, A., Schmidt, J.M., Detmar, E., Lai, M.M., Shishkin, A.A., Bhat, P., Takei, Y., et al. (2018). Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus. Cell *174*, 744-757.e24.

Quintero, A., Hübschmann, D., Kurzawa, N., Steinhauser, S., Rentzsch, P., Krämer, S., Andresen, C., Park, J., Eils, R., Schlesner, M., et al. (2021). ShinyButchR: Interactive NMF-based decomposition workflow of genome-scale datasets. Biol. Methods Protoc. *5*, 1–7.

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. Nature *470*, 279–283.

Ramani, V., Deng, X., Qiu, R., Gunderson, K.L., Steemers, F.J., Disteche, C.M., Noble, W.S., Duan, Z., and Shendure, J. (2017). Massively multiplex single-cell Hi-C. Nat. Methods *14*, 263–266.

Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., et al. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell *159*, 1665–1680.

Rao, S.S.P., Huang, S.-C., Glenn St Hilaire, B., Engreitz, J.M., Perez, E.M., Kieffer-Kwon, K.-R., Sanborn, A.L., Johnstone, S.E., Bascom, G.D., Bochkov, I.D., et al. (2017). Cohesin Loss Eliminates All Loop Domains. Cell *171*, 305-320.e24.

Rappaport, J., and Volsky, D.J. (2015). Role of the macrophage in HIV-associated neurocognitive disorders and other comorbidities in patients on effective antiretroviral treatment. J. Neurovirol. *12*, 235–241.

De Ravin, S.S., Su, L., Theobald, N., Choi, U., Macpherson, J.L., Poidinger, M., Symonds, G., Pond, S.M., Ferris, A.L., Hughes, S.H., et al. (2014). Enhancers are major targets for murine leukemia virus vector integration. J. Virol. *88*, 4504–4513.

Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B.D., Sun, Z.W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C.P., Allis, C.D., et al. (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature *406*, 593–599.

Reines, D., Conaway, J.W., and Conaway, R.C. (1996). The RNA polymerase II general elongation factors. Trends Biochem. Sci. *21*, 351–355.

Dos Reis, R.S., Sant, S., Keeney, H., Wagner, M.C.E., and Ayyavoo, V. (2020). Modeling HIV-1 neuropathogenesis using three-dimensional human brain organoids (hBORGs) with HIV-1 infected microglia. Sci. Rep. *10*, 15209.

Dos Reis, R.S., Sant, S., and Ayyavoo, V. (2023). Three-Dimensional Human Brain Organoids to Model HIV-1 Neuropathogenesis. Methods Mol. Biol. *2610*, 167–178.

Rheinberger, M., Costa, A.L., Kampmann, M., Glavas, D., Shytaj, I.L., Sreeram, S., Penzo, C., Tibroni, N., Garcia-Mesa, Y., Leskov, K., et al. (2023). Genomic profiling of HIV-1 integration in microglia cells links viral integration to the topologically associated domains. Cell Rep. *42*, 112110.

Richetta, C., Thierry, S., Thierry, E., Lesbats, P., Lapaillerie, D., Munir, S., Subra, F., Leh, H., Deprez, E., Parissi, V., et al. (2019). Two-long terminal repeat (LTR) DNA circles are a substrate for HIV-1 integrase. J. Biol. Chem. *294*, 8286–8295.

De Rijck, J., Vandekerckhove, L., Gijsbers, R., Hombrouck, A., Hendrix, J., Vercammen, J., Engelborghs, Y., Christ, F., and Debyser, Z. (2006). Overexpression of the lens epithelium-derived growth factor/p75 integrase binding domain inhibits human immunodeficiency virus replication. J Virol *80*, 11498–11509.

Robertson, K., Landay, A., Miyahara, S., Vecchio, A., Masters, M.C., Brown, T.T., and Taiwo, B.O. (2020). Limited correlation between systemic biomarkers and neurocognitive performance before and during HIV treatment. J. Neurovirol. *26*, 107–113.

Rohdewohld, H., Weiher, H., Reik, W., Jaenisch, R., and Breindl, M. (1987). Retrovirus integration and chromatin structure: Moloney murine leukemia proviral integration sites map near DNase I-hypersensitive sites. J. Virol. *61*, 336–343.

Romerio, F., Gabriel, M.N., and Margolis, D.M. (1997). Repression of human immunodeficiency virus type 1 through the novel cooperation of human factors YY1 and LSF. J. Virol. *71*, 9375–9382.

Roth, S.L., Malani, N., and Bushman, F.D. (2011). Gammaretroviral integration into nucleosomal target DNA in vivo. J Virol *85*, 7393–7401.

Rouzine, I.M., Weinberger, A.D., and Weinberger, L.S. (2015). An evolutionary role for

HIV latency in enhancing viral transmission. Cell *160*, 1002–1012.

Sabo, A., Lusic, M., Cereseto, A., and Giacca, M. (2008). Acetylation of conserved lysines in the catalytic core of cyclin-dependent kinase 9 inhibits kinase activity and regulates transcription. Mol Cell Biol *28*, 2201–2212.

Sagai, T., Hosoya, M., Mizushina, Y., Tamura, M., and Shiroishi, T. (2005). Elimination of a long-range cis-regulatory module causes complete loss of limb-specific Shh expression and truncation of the mouse limb. Development *132*, 797–803.

Saito, A., Henning, M.S., Serrao, E., Dubose, B.N., Teng, S., Huang, J., Li, X., Saito, N., Roy, S.P., Siddiqui, M.A., et al. (2016). Capsid-CPSF6 Interaction Is Dispensable for HIV-1 Replication in Primary Cells but Is Selected during Virus Passage In Vivo. J Virol *90*, 6918–6935.

Saldaña-Meyer, R., González-Buendía, E., Guerrero, G., Narendra, V., Bonasio, R., Recillas-Targa, F., and Reinberg, D. (2014). CTCF regulates the human p53 gene through direct interaction with its natural antisense transcript, Wrap53. Genes Dev. *28*, 723–734.

Saldaña-Meyer, R., Rodriguez-Hernaez, J., Escobar, T., Nishana, M., Jácome-López, K., Nora, E.P., Bruneau, B.G., Tsirigos, A., Furlan-Magaril, M., Skok, J., et al. (2019). RNA Interactions Are Essential for CTCF-Mediated Genome Organization. Mol. Cell *76*, 412-422.e5.

Sanna, P.P., Fu, Y., Masliah, E., Lefebvre, C., and Repunte-Canonigo, V. (2021). Central nervous system (CNS) transcriptomic correlates of human immunodeficiency virus (HIV) brain RNA load in HIV-infected individuals. Sci. Rep. *11*, 12176.

Santos-Rosa, H., Schneider, R., Bannister, A.J., Sherriff, J., Bernstein, B.E., Emre, N.C.T., Schreiber, S.L., Mellor, J., and Kouzarides, T. (2002). Active genes are tri-methylated at K4 of histone H3. Nature *419*, 407–411.

Satou, Y., Yasunaga, J., Yoshida, M., and Matsuoka, M. (2006). HTLV-I basic leucine zipper factor gene mRNA supports proliferation of adult T cell leukemia cells. Proc. Natl. Acad. Sci. U. S. A. *103*, 720–725.

Satou, Y., Miyazato, P., Ishihara, K., Yaguchi, H., Melamed, A., Miura, M., Fukuda, A., Nosaka, K., Watanabe, T., Rowan, A.G., et al. (2016). The retrovirus HTLV-1 inserts an ectopic CTCF-binding site into the human genome. Proc. Natl. Acad. Sci. U. S. A. *113*, 3054–3059.

Schaller, T., Ocwieja, K.E., Rasaiyaah, J., Price, A.J., Brady, T.L., Roth, S.L., Hué, S., Fletcher, A.J., Lee, K., KewalRamani, V.N., et al. (2011). HIV-1 Capsid-Cyclophilin

Interactions Determine Nuclear Import Pathway, Integration Targeting and Replication Efficiency. PLOS Pathog. *7*, e1002439.

Schifferdecker, S., Zila, V., Müller, T.G., Sakin, V., Anders-Össwein, M., Laketa, V., Kräusslich, H.-G., and Müller, B. (2022). Direct Capsid Labeling of Infectious HIV-1 by Genetic Code Expansion Allows Detection of Largely Complete Nuclear Capsids and Suggests Nuclear Entry of HIV-1 Complexes via Common Routes. MBio *13*, e0195922.

Schmidt, D., Schwalie, P.C., Wilson, M.D., Ballester, B., Gonçalves, A., Kutter, C., Brown, G.D., Marshall, A., Flicek, P., and Odom, D.T. (2012). Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. Cell *148*, 335–348.

Schmitges, F.W., Prusty, A.B., Faty, M., Stützer, A., Lingaraju, G.M., Aiwazian, J., Sack, R., Hess, D., Li, L., Zhou, S., et al. (2011). Histone methylation by PRC2 is inhibited by active chromatin marks. Mol. Cell *42*, 330–341.

Schmitt, A.D., Hu, M., Jung, I., Xu, Z., Qiu, Y., Tan, C.L., Li, Y., Lin, S., Lin, Y., Barr, C.L., et al. (2016). A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome. Cell Rep. *17*, 2042–2059.

Scholz, S.J., Fronza, R., Bartholoma, C.C., Cesana, D., Montini, E., von Kalle, C., Gil-Farina, I., and Schmidt, M. (2017). Lentiviral Vector Promoter is Decisive for Aberrant Transcript Formation. Hum Gene Ther *28*, 875–885.

Schrijvers, R., De Rijck, J., Demeulemeester, J., Adachi, N., Vets, S., Ronen, K., Christ, F., Bushman, F.D., Debyser, Z., and Gijsbers, R. (2012). LEDGF/p75-independent HIV-1 replication demonstrates a role for HRP-2 and remains sensitive to inhibition by LEDGINs. PLoS Pathog *8*, e1002558.

Schroder, A.R., Shinn, P., Chen, H., Berry, C., Ecker, J.R., and Bushman, F. (2002). HIV-1 integration in the human genome favors active genes and local hotspots. Cell *110*, 521–529.

Schwarzer, W., Abdennur, N., Goloborodko, A., Pekowska, A., Fudenberg, G., Loe-Mie, Y., Fonseca, N.A., Huber, W., Haering, C.H., Mirny, L., et al. (2017). Two independent modes of chromatin organization revealed by cohesin removal. Nature *551*, 51–56.

Selyutina, A., Persaud, M., Lee, K., KewalRamani, V., and Diaz-Griffero, F. (2020). Nuclear Import of the HIV-1 Core Precedes Reverse Transcription and Uncoating. Cell Rep. *32*, 108201.

Sengupta, S., and Siliciano, R.F. (2018). Targeting the Latent Reservoir for HIV-1. Immunity *48*, 872–895.

Serrao, E., Krishnan, L., Shun, M.C., Li, X., Cherepanov, P., Engelman, A., and Maertens, G.N. (2014). Integrase residues that determine nucleotide preferences at sites of HIV-1 integration: implications for the mechanism of target DNA binding. Nucleic Acids Res *42*, 5164–5176.

Serrao, E., Cherepanov, P., and Engelman, A.N. (2016). Amplification, Next-generation Sequencing, and Genomic DNA Mapping of Retroviral Integration Sites. J. Vis. Exp. 53840.

Shahbazian, M.D., and Grunstein, M. (2007). Functions of site-specific histone acetylation and deacetylation. Annu. Rev. Biochem. *76*, 75–100.

Shan, L., Deng, K., Shroff, N.S., Durand, C.M., Rabi, S.A., Yang, H.-C., Zhang, H., Margolick, J.B., Blankson, J.N., and Siliciano, R.F. (2012). Stimulation of HIV-1-Specific Cytolytic T Lymphocytes Facilitates Elimination of Latent Viral Reservoir after Virus Reactivation. Immunity *36*, 491–501.

Sharkey, M., Triques, K., Kuritzkes, D.R., and Stevenson, M. (2005). In vivo evidence for instability of episomal human immunodeficiency virus type 1 cDNA. J. Virol. *79*, 5203–5210.

Sharp, P.M., and Hahn, B.H. (2011). Origins of HIV and the AIDS pandemic. Cold Spring Harb Perspect Med *1*, a006841.

Shen, C., Feng, X., Mao, T., Yang, D., Zou, J., Zao, X., Deng, Q., Chen, X., and Lu, F. (2020). Yin-Yang 1 and HBx protein activate HBV transcription by mediating the spatial interaction of cccDNA minichromosome with cellular chromosome 19p13.11. Emerg. Microbes Infect. *9*, 2455–2464.

Shen, Q., Wu, C., Freniere, C., Tripler, T.N., and Xiong, Y. (2021). Nuclear Import of HIV-1. Viruses *13*.

Sherrill-Mix, S., Lewinski, M.K., Famiglietti, M., Bosque, A., Malani, N., Ocwieja, K.E., Berry, C.C., Looney, D., Shan, L., Agosto, L.M., et al. (2013). HIV latency and integration site placement in five cell-based models. Retrovirology *10*, 90.

Shi, X., Hong, T., Walter, K.L., Ewalt, M., Michishita, E., Hung, T., Carney, D., Peña, P., Lan, F., Kaadige, M.R., et al. (2006). ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. Nature *442*, 96–99.

Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J.R., Cole, P.A., Casero, R.A., and Shi, Y. (2004). Histone demethylation mediated by the nuclear amine oxidase

homolog LSD1. Cell *119*, 941–953.

Shukla, S., Kavak, E., Gregory, M., Imashimizu, M., Shutinoski, B., Kashlev, M., Oberdoerffer, P., Sandberg, R., and Oberdoerffer, S. (2011). CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. Nature *479*, 74–79.

Shun, M.C., Raghavendra, N.K., Vandegraaff, N., Daigle, J.E., Hughes, S., Kellam, P., Cherepanov, P., and Engelman, A. (2007). LEDGF/p75 functions downstream from preintegration complex formation to effect gene-specific HIV-1 integration. Genes Dev *21*, 1767–1778.

Shytaj, I.L., Lucic, B., Forcato, M., Penzo, C., Billingsley, J., Laketa, V., Bosinger, S., Stanic, M., Gregoretti, F., Antonelli, L., et al. (2020). Alterations of redox and iron metabolism accompany the development of HIV latency. EMBO J. *39*, e102209.

Simonetti, F.R., Sobolewski, M.D., Fyne, E., Shao, W., Spindler, J., Hattori, J., Anderson, E.M., Watters, S.A., Hill, S., Wu, X., et al. (2016). Clonally expanded CD4+ T cells can produce infectious HIV-1 in vivo. Proc Natl Acad Sci U S A *113*, 1883–1888.

Simonetti, F.R., Zhang, H., Soroosh, G.P., Duan, J., Rhodehouse, K., Hill, A.L., Beg, S.A., McCormick, K., Raymond, H.E., Nobles, C.L., et al. (2021). Antigen-driven clonal selection shapes the persistence of HIV-1-infected CD4+ T cells in vivo. J. Clin. Invest. *131*.

Singh, P.K., Plumb, M.R., Ferris, A.L., Iben, J.R., Wu, X., Fadel, H.J., Luke, B.T., Esnault, C., Poeschla, E.M., Hughes, S.H., et al. (2015). LEDGF/p75 interacts with mRNA splicing factors and targets HIV-1 integration to highly spliced genes. Genes Dev *29*, 2287–2297.

Singh, P.K., Bedwell, G.J., and Engelman, A.N. (2022). Spatial and Genomic Correlates of HIV-1 Integration Site Targeting. Cells *11*.

Sloan, R.D., and Wainberg, M.A. (2011). The role of unintegrated DNA in HIV infection. Retrovirology *8*, 1–15.

Sobhian, B., Laguette, N., Yatim, A., Nakamura, M., Levy, Y., Kiernan, R., and Benkirane, M. (2010). HIV-1 Tat assembles a multifunctional transcription elongation complex and stably associates with the 7SK snRNP. Mol. Cell *38*, 439–451.

Sousa, C., Golebiewska, A., Poovathingal, S.K., Kaoma, T., Pires-Afonso, Y., Martina, S., Coowar, D., Azuaje, F., Skupin, A., Balling, R., et al. (2018). Single-cell transcriptomics reveals distinct inflammation-induced microglia signatures. EMBO Rep. *19*, 1–17.

Sousa, R., Chung, Y.J., Rose, J.P., and Wang, B.C. (1993). Crystal structure of bacteriophage T7 RNA polymerase at 3.3 A resolution. Nature *364*, 593–599.

Sowd, G.A., Serrao, E., Wang, H., Wang, W., Fadel, H.J., Poeschla, E.M., and Engelman, A.N. (2016). A critical role for alternative polyadenylation factor CPSF6 in targeting HIV-1 integration to transcriptionally active chromatin. Proc. Natl. Acad. Sci. U. S. A. *113*, E1054–E1063.

Spicuglia, S., and Vanhille, L. (2012). Chromatin signatures of active enhancers. Nucleus *3*, 126–131.

Spittau, B. (2017). Aging Microglia-Phenotypes, Functions and Implications for Age-Related Neurodegenerative Diseases. Front. Aging Neurosci. *9*, 194.

Spitz, F., and Furlong, E.E.M. (2012). Transcription factors: from enhancer binding to developmental control. Nat. Rev. Genet. *13*, 613–626.

Spudich, S., Robertson, K.R., Bosch, R.J., Gandhi, R.T., Cyktor, J.C., Mar, H., Macatangay, B.J., Lalama, C.M., Rinaldo, C., Collier, A.C., et al. (2019). Persistent HIV-infected cells in cerebrospinal fluid are associated with poorer neurocognitive performance. J. Clin. Invest. *129*, 3339–3346.

Sreeram, S., Ye, F., Garcia-mesa, Y., Nguyen, K., Sayed, A. El, Leskov, K., and Karn, J. (2022). Trends in Immunology The potential role of HIV-1 latency in promoting neuroin fl ammation and HIV-1-associated neurocognitive disorder. Trends Immunol. *xx*, 1–10.

Stein, B.S., Gowda, S.D., Lifson, J.D., Penhallow, R.C., Bensch, K.G., and Engleman, E.G. (1987). pH-independent HIV entry into CD4-positive T cells via virus envelope fusion to the plasma membrane. Cell *49*, 659–668.

Sterner, D.E., and Berger, S.L. (2000). Acetylation of histones and transcription-related factors. Microbiol. Mol. Biol. Rev. *64*, 435–459.

Stevens, T.J., Lando, D., Basu, S., Atkinson, L.P., Cao, Y., Lee, S.F., Leeb, M., Wohlfahrt, K.J., Boucher, W., O'Shaughnessy-Kirwan, A., et al. (2017). 3D structures of individual mammalian genomes studied by single-cell Hi-C. Nature *544*, 59–64.

Stik, G., Vidal, E., Barrero, M., Cuartero, S., Vila-Casadesús, M., Mendieta-Esteban, J., Tian, T. V, Choi, J., Berenguer, C., Abad, A., et al. (2020). CTCF is dispensable for immune cell transdifferentiation but facilitates an acute inflammatory response. Nat. Genet. *52*, 655–661.

Stoler, M.H., Eskin, T.A., Benn, S., Angerer, R.C., and Angerer, L.M. (1986). Human T-cell lymphotropic virus type III infection of the central nervous system. A preliminary

in situ analysis. JAMA *256*, 2360–2364.

Stroud, H., Yang, M.G., Tsitohay, Y.N., Davis, C.P., Sherman, M.A., Hrvatin, S., Ling, E., and Greenberg, M.E. (2020). An Activity-Mediated Transition in Transcription in Early Postnatal Neurons. Neuron *107*, 874-890.e8.

Sunshine, S., Kirchner, R., Amr, S.S., Mansur, L., Shakhbatyan, R., Kim, M., Bosque, A., Siliciano, R.F., Planelles, V., Hofmann, O., et al. (2016). HIV Integration Site Analysis of Cellular Models of HIV Latency with a Probe-Enriched Next-Generation Sequencing Assay. J Virol *90*, 4511–4519.

Swanta, N., Aryal, S., Nejtek, V., Shenoy, S., Ghorpade, A., and Borgmann, K. (2020). Blood-based inflammation biomarkers of neurocognitive impairment in people living with HIV. J. Neurovirol. *26*, 358–370.

Swiggard, W.J., Baytop, C., Yu, J.J., Dai, J., Li, C., Schretzenmair, R., Theodosopoulos, T., and O'Doherty, U. (2005). Human immunodeficiency virus type 1 can establish latent infection in resting CD4+ T cells in the absence of activating stimuli. J. Virol. *79*, 14179–14188.

Tan, M., Luo, H., Lee, S., Jin, F., Yang, J.S., Montellier, E., Buchou, T., Cheng, Z., Rousseaux, S., Rajagopal, N., et al. (2011). Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. Cell *146*, 1016–1028.

Tan, W., Dong, Z., Wilkinson, T.A., Barbas, C.F., and Chow, S.A. (2006). Human Immunodeficiency Virus Type 1 Incorporated with Fusion Proteins Consisting of Integrase and the Designed Polydactyl Zinc Finger Protein E2C Can Bias Integration of Viral DNA into a Predetermined Chromosomal Region in Human Cells. J. Virol. *80*, 1939–1948.

Tang, D., Zhao, H., Wu, Y., Peng, B., Gao, Z., Sun, Y., Duan, J., Qi, Y., Li, Y., Zhou, Z., et al. (2021). Transcriptionally inactive hepatitis B virus episome DNA preferentially resides in the vicinity of chromosome 19 in 3D host genome upon infection. Cell Rep. *35*, 109288.

Tang, Y., Margolis, D.M., Jiang, G., Tang, Y., Chaillon, A., Gianella, S., Wong, L.M., Li, D., Simermeyer, T.L., Porrachia, M., et al. (2023). Brain microglia serve as a persistent HIV reservoir despite durable antiretroviral therapy Brain microglia serve as a persistent HIV reservoir despite durable antiretroviral therapy. *133*.

Tansey, K.E., Cameron, D., and Hill, M.J. (2018). Genetic risk for Alzheimer's disease is concentrated in specific macrophage and microglial transcriptional networks.

Genome Med. *10*, 1–10.

Taverna, S.D., Li, H., Ruthenburg, A.J., Allis, C.D., and Patel, D.J. (2007). How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. Nat. Struct. Mol. Biol. *14*, 1025–1040.

Tay, T.L., Savage, J.C., Hui, C.W., Bisht, K., and Tremblay, M.È. (2017). Microglia across the lifespan: from origin to function in brain development, plasticity and cognition. J. Physiol. *595*, 1929–1945.

Temin, H.M., and Mizutani, S. (1970). RNA-dependent DNA polymerase in virions of Rous sarcoma virus. Nature *226*, 1211–1213.

Tempera, I., Wiedmer, A., Dheekollu, J., and Lieberman, P.M. (2010). CTCF prevents the epigenetic drift of EBV latency promoter Qp. PLoS Pathog. *6*, e1001048.

Tesseur, I., Zou, K., Esposito, L., Bard, F., Berber, E., Can, J. Van, Lin, A.H., Crews, L., Tremblay, P., Mathews, P., et al. (2006). Deficiency in neuronal TGF-β signaling promotes neurodegeneration and Alzheimer's pathology. J. Clin. Invest. *116*, 3060–3069.

Thompson, K.A., Cherry, C.L., Bell, J.E., and McLean, C.A. (2011). Brain cell reservoirs of latent virus in presymptomatic HIV-infected individuals. Am. J. Pathol. *179*, 1623–1629.

Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., Vernot, B., et al. (2012). The accessible chromatin landscape of the human genome. Nature *489*, 75–82.

Tian, R., Huang, Z., Li, L., Yuan, J., Zhang, Q., Meng, L., Lang, B., Hong, Y., Zhong, C., Tian, X., et al. (2023). HPV integration generates a cellular super-enhancer which functions as ecDNA to regulate genome-wide transcription. Nucleic Acids Res. *51*, 4237–4251.

Tong-Starksen, S.E., Luciw, P.A., and Peterlin, B.M. (1987). Human immunodeficiency virus long terminal repeat responds to T-cell activation signals. Proc. Natl. Acad. Sci. U. S. A. *84*, 6845–6849.

Torrano, V., Navascués, J., Docquier, F., Zhang, R., Burke, L.J., Chernukhin, I., Farrar, D., León, J., Berciano, M.T., Renkawitz, R., et al. (2006). Targeting of CTCF to the nucleolus inhibits nucleolar transcription through a poly(ADP-ribosyl)ation-dependent mechanism. J. Cell Sci. *119*, 1746–1759.

Tsompana, M., and Buck, M.J. (2014). Chromatin accessibility: a window into the genome. Epigenetics Chromatin *7*, 33.

Tsukada, Y., Fang, J., Erdjument-Bromage, H., Warren, M.E., Borchers, C.H., Tempst, P., and Zhang, Y. (2006). Histone demethylation by a family of JmjC domain-containing proteins. Nature *439*, 811–816.

Tsukiyama, T. (2002). The in vivo functions of ATP-dependent chromatin-remodelling factors. Nat. Rev. Mol. Cell Biol. *3*, 422–429.

Turner, B.G., and Summers, M.F. (1999). Structural biology of HIV11Edited by P. E. Wright. J. Mol. Biol. *285*, 1–32.

Ueberham, U., Hilbrich, I., Ueberham, E., Rohn, S., Glöckner, P., Dietrich, K., Brückner, M.K., and Arendt, T. (2012). Transcriptional control of cell cycle-dependent kinase 4 by Smad proteins—implications for Alzheimer's disease. Neurobiol. Aging *33*, 2827–2840.

UNAIDS UNAIDS.

Vandergeeten, C., Fromentin, R., and Chomont, N. (2012). The role of cytokines in the establishment, persistence and eradication of the HIV reservoir. Cytokine Growth Factor Rev. *23*, 143–149.

Vansant, G., Chen, H.-C., Zorita, E., Trejbalová, K., Miklík, D., Filion, G., and Debyser, Z. (2020). The chromatin landscape at the HIV-1 provirus integration site determines viral expression. Nucleic Acids Res. *48*, 7801–7817.

Varghese, C.S., Parish, J.L., and Ferguson, J. (2022). Lying low-chromatin insulation in persistent DNA virus infection. Curr. Opin. Virol. *55*, 101257.

Veazey, R.S., DeMaria, M., Chalifoux, L. V, Shvetz, D.E., Pauley, D.R., Knight, H.L., Rosenzweig, M., Johnson, R.P., Desrosiers, R.C., and Lackner, A.A. (1998). Gastrointestinal tract as a major site of CD4+ T cell depletion and viral replication in SIV infection. Science *280*, 427–431.

Veenstra, M., Leon-Rivera, R., Li, M., Gama, L., Clements, J.E., and Berman, J.W. (2017). Mechanisms of CNS Viral Seeding by HIV(+) CD14(+) CD16(+) Monocytes: Establishment and Reseeding of Viral Reservoirs Contributing to HIV-Associated Neurocognitive Disorders. MBio *8*.

Vera, J.H., Guo, Q., Cole, J.H., Boasso, A., Greathead, L., Kelleher, P., Rabiner, E.A., Kalk, N., Bishop, C., Gunn, R.N., et al. (2016). Neuroinflammation in treated HIV-positive individuals: A TSPO PET study. Neurology *86*, 1425–1432.

Verdin, E., Paras, P., and Van Lint, C. (1993). Chromatin disruption in the promoter of human immunodeficiency virus type 1 during transcriptional activation. EMBO J. *12*, 3249–3259.

Vermunt, M.W., Zhang, D., and Blobel, G.A. (2019). The interdependence of gene-regulatory elements and the 3D genome. J. Cell Biol. *218*, 12–26.

Vijaya, S., Steffen, D.L., and Robinson, H.L. (1986). Acceptor sites for retroviral integrations map near DNase I-hypersensitive sites in chromatin. J. Virol. *60*, 683–692.

Vink, C., and Plasterk, R.H. (1993). The human immunodeficiency virus integrase protein. Trends Genet *9*, 433–438.

Visel, A., Blow, M.J., Li, Z., Zhang, T., Akiyama, J.A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F., et al. (2009). ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature *457*, 854–858.

Vostrov, A.A., Taheny, M.J., and Quitschke, W.W. (2002). A region to the N-terminal side of the CTCF zinc finger domain is essential for activating transcription from the amyloid precursor protein promoter. J. Biol. Chem. *277*, 1619–1627.

Vranckx, L.S., Demeulemeester, J., Saleh, S., Boll, A., Vansant, G., Schrijvers, R., Weydert, C., Battivelli, E., Verdin, E., Cereseto, A., et al. (2016). LEDGIN-mediated Inhibition of Integrase-LEDGF/p75 Interaction Reduces Reactivation of Residual Latent HIV. EBioMedicine *8*, 248–264.

Wagner, T.A., McLaughlin, S., Garg, K., Cheung, C.Y., Larsen, B.B., Styrchak, S., Huang, H.C., Edlefsen, P.T., Mullins, J.I., and Frenkel, L.M. (2014). HIV latency. Proliferation of cells with HIV integrated into cancer genes contributes to persistent infection. Science (80-.). *345*, 570–573.

Wallet, C., De Rovere, M., Van Assche, J., Daouad, F., De Wit, S., Gautier, V., Mallon, P.W.G., Marcello, A., Van Lint, C., Rohr, O., et al. (2019). Microglial Cells: The Main HIV-1 Reservoir in the Brain. Front. Cell. Infect. Microbiol. *9*, 362.

Wang, D.C., Wang, W., Zhang, L., and Wang, X. (2019). A tour of 3D genome with a focus on CTCF. Semin. Cell Dev. Biol. *90*, 4–11.

Wang, G.P., Ciuffi, A., Leipzig, J., Berry, C.C., and Bushman, F.D. (2007). HIV integration site selection: analysis by massively parallel pyrosequencing reveals association with epigenetic modifications. Genome Res *17*, 1186–1194.

Wang, G.Z., Wang, Y., and Goff, S.P. (2016a). Histones Are Rapidly Loaded onto Unintegrated Retroviral DNAs Soon after Nuclear Entry. Cell Host Microbe *20*, 798–809.

Wang, J., Wang, Y., and Lu, L. (2012). De-SUMOylation of CCCTC binding factor (CTCF) in hypoxic stress-induced human corneal epithelial cells. J. Biol. Chem. *287*,

12469–12479.

Wang, R., Lee, J.-H., Kim, J., Xiong, F., Hasani, L. Al, Shi, Y., Simpson, E.N., Zhu, X., Chen, Y.-T., Shivshankar, P., et al. (2023). SARS-CoV-2 restructures host chromatin architecture. Nat. Microbiol. *8*, 679–694.

Wang, S., Su, J.-H., Beliveau, B.J., Bintu, B., Moffitt, J.R., Wu, C., and Zhuang, X. (2016b). Spatial organization of chromatin domains and compartments in single chromosomes. Science *353*, 598–602.

Wang, W., Fasolino, M., Cattau, B., Goldman, N., Kong, W., Frederick, M.A., McCright, S.J., Kiani, K., Fraietta, J.A., and Vahedi, G. (2020). Joint profiling of chromatin accessibility and CAR-T integration site analysis at population and single-cell levels. Proc. Natl. Acad. Sci. U. S. A. *117*, 5442–5452.

Wang, Y., Reddy, B., Thompson, J., Wang, H., Noma, K., Yates, J.R. 3rd, and Jia, S. (2009). Regulation of Set9-mediated H4K20 methylation by a PWWP domain protein. Mol. Cell *33*, 428–437.

Washington, S.D., Musarrat, F., Ertel, M.K., Backes, G.L., and Neumann, D.M. (2018a). CTCF Binding Sites in the Herpes Simplex Virus 1 Genome Display Site-Specific CTCF Occupation, Protein Recruitment, and Insulator Function. J. Virol. *92*.

Washington, S.D., Edenfield, S.I., Lieux, C., Watson, Z.L., Taasan, S.M., Dhummakupt, A., Bloom, D.C., and Neumann, D.M. (2018b). Depletion of the Insulator Protein CTCF Results in Herpes Simplex Virus 1 Reactivation In Vivo. J. Virol. *92*.

Waszak, S.M., Delaneau, O., Gschwind, A.R., Kilpinen, H., Raghav, S.K., Witwicki, R.M., Orioli, A., Wiederkehr, M., Panousis, N.I., Yurovsky, A., et al. (2015). Population Variation and Genetic Control of Modular Chromatin Architecture in Humans. Cell *162*, 1039–1050.

Wei, P., Garber, M.E., Fang, S.M., Fischer, W.H., and Jones, K.A. (1998). 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 *92*, 451–462.

Wei, X., Decker, J.M., Liu, H., Zhang, Z., Arani, R.B., Kilby, J.M., Saag, M.S., Wu, X., Shaw, G.M., and Kappes, J.C. (2002). Emergence of resistant human immunodeficiency virus type 1 in patients receiving fusion inhibitor (T-20) monotherapy. Antimicrob. Agents Chemother. *46*, 1896–1905.

Weinberger, L.S., Burnett, J.C., Toettcher, J.E., Arkin, A.P., and Schaffer, D. V (2005). Stochastic gene expression in a lentiviral positive-feedback loop: HIV-1 Tat fluctuations drive phenotypic diversity. Cell *122*, 169–182.

Weintraub, A.S., Li, C.H., Zamudio, A. V, Sigova, A.A., Hannett, N.M., Day, D.S., Abraham, B.J., Cohen, M.A., Nabet, B., Buckley, D.L., et al. (2017). YY1 Is a Structural Regulator of Enhancer-Promoter Loops. Cell *171*, 1573-1588.e28.

Wells, D.W., Guo, S., Shao, W., Bale, M.J., Coffin, J.M., Hughes, S.H., and Wu, X. (2020). Erratum: An analytical pipeline for identifying and mapping the integration sites of HIV and other retroviruses (BMC Genomics (2020) 21 (216) DOI: 10.1186/s12864-020-6647-4). BMC Genomics *21*, 1–20.

Wiley, C.A., Schrier, R.D., Nelson, J.A., Lampert, P.W., and Oldstone, M.B. (1986). Cellular localization of human immunodeficiency virus infection within the brains of acquired immune deficiency syndrome patients. Proc. Natl. Acad. Sci. U. S. A. *83*, 7089–7093.

Williams, M.E., Stein, D.J., Joska, J.A., and Naudé, P.J.W. (2021). Cerebrospinal fluid immune markers and HIV-associated neurocognitive impairments: A systematic review. J. Neuroimmunol. *358*, 577649.

Williams, S.A., Kwon, H., Chen, L.F., and Greene, W.C. (2007). Sustained induction of NF-kappa B is required for efficient expression of latent human immunodeficiency virus type 1. J Virol *81*, 6043–6056.

Wong, M.E., Jaworowski, A., and Hearps, A.C. (2019). The HIV Reservoir in Monocytes and Macrophages. Front. Immunol. *10*, 1435.

Wu, V.H., Nordin, J.M.L., Nguyen, S., Joy, J., Mampe, F., del Rio Estrada, P.M., Torres-Ruiz, F., González-Navarro, M., Luna-Villalobos, Y.A., Ávila-Ríos, S., et al. (2023). Profound phenotypic and epigenetic heterogeneity of the HIV-1-infected CD4+ T cell reservoir. Nat. Immunol. *24*, 359–370.

Wu, X., Li, Y., Crise, B., and Burgess, S.M. (2003). Transcription start regions in the human genome are favored targets for MLV integration. Science (80-.). *300*, 1749–1751.

Wutz, G., Várnai, C., Nagasaka, K., Cisneros, D.A., Stocsits, R.R., Tang, W., Schoenfelder, S., Jessberger, G., Muhar, M., Hossain, M.J., et al. (2017). Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. EMBO J. *36*, 3573–3599.

Xiao, T., Hall, H., Kizer, K.O., Shibata, Y., Hall, M.C., Borchers, C.H., and Strahl, B.D. (2003). Phosphorylation of RNA polymerase II CTD regulates H3 methylation in yeast. Genes Dev. *17*, 654–663.

Yadav, A., and Collman, R.G. (2009). CNS inflammation and macrophage/microglial

biology associated with HIV-1 infection. J Neuroimmune Pharmacol *4*, 430–447.

Yang, B., Li, B., Jia, L., Jiang, Y., Wang, X., Jiang, S., Du, S., Ji, X., and Yang, P. (2020). 3D landscape of Hepatitis B virus interactions with human chromatins. Cell Discov. *6*, 95.

Yang, S., Gao, L., Lu, F., Wang, B., Gao, F., Zhu, G., Cai, Z., Lai, J., and Yang, Q. (2015). Transcription factor myocyte enhancer factor 2D regulates interleukin-10 production in microglia to protect neuronal cells from inflammation-induced death. J. Neuroinflammation *12*, 1–10.

Yap, K.L., and Zhou, M.-M. (2010). Keeping it in the family: diverse histone recognition by conserved structural folds. Crit. Rev. Biochem. Mol. Biol. *45*, 488–505.

Ye, F., Alvarez-Carbonell, D., Nguyen, K., Leskov, K., Garcia-Mesa, Y., Sreeram, S., Valadkhan, S., and Karn, J. (2022). Recruitment of the CoREST transcription repressor complexes by Nerve Growth factor IB-like receptor (Nurr1/NR4A2) mediates silencing of HIV in microglial cells. PLoS Pathog. *18*, e1010110.

Young, A.M.H., Kumasaka, N., Calvert, F., Hammond, T.R., Knights, A., Panousis, N., Park, J.S., Schwartzentruber, J., Liu, J., Kundu, K., et al. (2021). A map of transcriptional heterogeneity and regulatory variation in human microglia. Nat. Genet. *53*.

Yu, Q., Liu, X., Fang, J., Wu, H., Guo, C., Zhang, W., Liu, N., Jiang, C., Sha, Q., Yuan, X., et al. (2023). Dynamics and regulation of mitotic chromatin accessibility bookmarking at single-cell resolution. Sci. Adv. *9*, eadd2175.

Yu, S., Li, J., Ji, G., Ng, Z.L., Siew, J., Lo, W.N., Ye, Y., Chew, Y.Y., Long, Y.C., Zhang, W., et al. (2022). Npac Is A Co-factor of Histone H3K36me3 and Regulates Transcriptional Elongation in Mouse Embryonic Stem Cells. Genomics. Proteomics Bioinformatics *20*, 110–128.

Yuan, W., Xu, M., Huang, C., Liu, N., Chen, S., and Zhu, B. (2011). H3K36 methylation antagonizes PRC2-mediated H3K27 methylation. J. Biol. Chem. *286*, 7983–7989.

Yukl, S.A., Gianella, S., Sinclair, E., Epling, L., Li, Q., Duan, L., Choi, A.L.M., Girling, V., Ho, T., Li, P., et al. (2010). Differences in HIV burden and immune activation within the gut of HIV-positive patients receiving suppressive antiretroviral therapy. J. Infect. Dis. *202*, 1553–1561.

Zentner, G.E., Tesar, P.J., and Scacheri, P.C. (2011). Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. Genome Res. *21*, 1273–1283.

Zhang, T., Cooper, S., and Brockdorff, N. (2015). The interplay of histone modifications - writers that read. EMBO Rep *16*, 1467–1481.

Zhou, M., Huang, K., Jung, K.-J., Cho, W.-K., Klase, Z., Kashanchi, F., Pise-Masison, C.A., and Brady, J.N. (2009). Bromodomain protein Brd4 regulates human immunodeficiency virus transcription through phosphorylation of CDK9 at threonine 29. J. Virol. *83*, 1036–1044.

Zila, V., Margiotta, E., Turoňová, B., Müller, T.G., Zimmerli, C.E., Mattei, S., Allegretti, M., Börner, K., Rada, J., Müller, B., et al. (2021). Cone-shaped HIV-1 capsids are transported through intact nuclear pores. Cell *184*, 1032-1046.e18.

Zuin, J., Dixon, J.R., van der Reijden, M.I.J.A., Ye, Z., Kolovos, P., Brouwer, R.W.W., van de Corput, M.P.C., van de Werken, H.J.G., Knoch, T.A., van IJcken, W.F.J., et al. (2014). Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. Proc. Natl. Acad. Sci. U. S. A. *111*, 996–1001.

Zuin, J., Roth, G., Zhan, Y., Cramard, J., Redolfi, J., Piskadlo, E., Mach, P., Kryzhanovska, M., Tihanyi, G., Kohler, H., et al. (2022). Nonlinear control of transcription through enhancer–promoter interactions. Nature *604*, 571–577.