Epi-genomic determinants of HIV-1 integration in human primary CD4$^+$ T cells and macrophages

Dissertation
submitted to the
Combined Faculty of Natural Sciences and Mathematics
of the Ruperto Carola University Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

Presented by
M.Sc. Mia Stanić
born in: Sarajevo
Oral examination: 14.10.2019

Dissertation

submitted to the

Combined Faculty of Natural Sciences and Mathematics

of the Ruperto Carola University Heidelberg, Germany

for the degree of

Doctor of Natural Sciences

Presented by

M.Sc. Mia Stanić

born in: Sarajevo

Oral examination: 14.10.2019

Epi-genomic determinants of HIV-1 integration
in human primary CD4$^+$ T cells and macrophages

Referees: Prof. Dr. Hans-Georg Kräusslich
Prof. Dr. Oliver T. Fackler

Declaration

The applicant, Mia Stanić, declares that she is the sole author of the submitted dissertation and that no other sources of help, apart from those specifically indicated, have been used in writing this work.

In addition, the applicant declares that she has not applied for permission to enter the examination procedure at any other institution and that his dissertation has not been presented to other faculties or used in its current, or any other form, in another examination.

Date Signature

_____ _____

Acknowledgements

First of all, I have to say a big "Thank you" to Dr. Marina Lusic. Thanks for recruiting me to your Lab, giving me the opportunity to learn, to grow and to work on a truly exciting project. Without her great ideas and support the output of this project would not be as it is now, and the work related to this project would not be as interesting as it is. Also, she is not just a scientific supervisor, but a friend too!

I have to express special gratefulness towards my TAC members, Prof. Dr. Hans-Georg Kräusslich and Prof. Dr. Oliver Fackler, for giving great scientific input and contributing to very fruitful scientific discussions in every TAC-meeting, but also beside official TAC meetings just by meeting at CIID's corridors, and providing full support and understanding for me as a PhD candidate.

People that deserve special acknowledgements are all the members of the Lusic Lab, which are not just colleagues, but also turned into great friends. Working in such a friendly environment makes every day "lab battles" way easier to win. Special thanks to Bojana Lucic who was a great help and support, especially in the beginning of my work in the lab, and Ines Castro who was helping out every day in anything she could. As well, Iart Luca Shytaj cannot be missed in this regard.
I have to acknowledge people from both, the Kräusslich and the Fackler Lab who were always very helpful and cooperative whenever asked for help. Especially David Bejarano, Virginia Pierini, Thorsten Müller and Vojtech Zila were contributing to this work by sharing reagents, cells, experience, time and good advice. Apart from being good scientist, these people are great friends too.
Also, Dr. Vibor Laketa from the CIID Imaging Platform deserves special thanks for his time, trainings on the microscopes and great expertise.

A general "Thank you" to all CIID, because this is a great environment to work in, in which one is surrounded with great people with great expertise and scientific backgrounds, great facilities and all scientific possibilities.

The greatest "Thank you" of all is for my family, especially my parents. Without their support, understanding, care and love nothing would be possible. A special "Thank you" (which is bold, sparkling and shining) in this section is directed to Dule. This PhD journey and life would not be the same without his limitless support and love.
My friends and professors from Sarajevo represented a special kind of support during this time. Without them, science and friendship would not be the same. Thank you Maja, Lara, Andrijana, Anesa, Lejla, Lada and Naris.

Finally, even though it was not the greatest of all love stories, but thank you Heidelberg. I am thankful for all the great people I met here, people that became my friends.
I would do it again…

Summary

The infection with HIV-1 nowadays does not represent a condition with a deadly outcome. Due to current therapeutic approaches, the infection with HIV-1 represents a chronic condition in which viral load is kept at undetectable levels, but patients depend on a lifelong therapy without a chance of cure. The eradication of integrated viral DNA still remains the biggest challenge in curing HIV-1.

The aim of this work was to contribute to a better understanding and definition of genomic regions and epi-genomic features that HIV-1 targets for integration, and give a detailed description on the importance of chromatin accessibility, as well as the importance of certain genomic features in the process of HIV-1 integration.

The first part of this project deals with the importance of HMT G9a activity and H3K9me2 histone mark distribution and deposition in the context of HIV-1 integration in primary $CD4^+$ T cells, which was studied by the application of G9a inhibitor BIX0129, also known as a very potent latency reversing agent. The significance of G9a activity and facultative heterochromatin mark H3K9me2 deposition has previously been shown to affect T cell development and impact shaping of the nuclear architecture. In this work it was demonstrated that the chemical inhibition of G9a and depletion of H3K9me2 by BIX01294 has an increasing effect on HIV-1 integration. The increase in integration was also followed by increased viral transcriptional activity, as well as spatial repositioning of the provirus from the preferred nuclear periphery towards the nuclear center. Similar spatial repositioning has been demonstrated for genes highly and recurrently targeted by HIV-1 for integration (RIGs). However, genic nuclear repositioning upon BIX01294 treatment did not affect transcriptional profiles of HIV-1 RIGs, as demonstrated by RNA microarray analysis, but other groups of genes mainly involved in iron metabolism and inflammatory response were upregulated upon BIX01294 treatment. In addition, HIV-1 integration patterns were shown not to be affected by H3K9me2 depletion, and the virus was still targeting similar genic regions for integration. The analysis of chromatin mark distribution and chromatin binding elements upon BIX01294 treatment on RIGs revealed increased binding profiles of open chromatin mark H3K36me3 which is followed by increased LEDGF/p75 binding upon H3K9me2 depletion. The observed phenomenon might provide an explanation for the observed increased viral integration upon BIX01294 treatment, considering that LEDGF/p75 is a prominent host cell factor involved in the viral integration process.

Overall, the first part of this study clearly demonstrated that chromatin accessibility significantly affects HIV-1 integration levels which are directly proportional to viral expression levels and viral activity.

The second part of this study deals with the relevance of R-loops, as specific genomic structures, as sites selected for HIV-1 integration in primary $CD4^+$ T cells and macrophages. It was demonstrated that the GFP tagged IN enzyme of HIV-1, in a high occurrence, colocalizes with R-loops in cells, and that for the occurrence of this process a functionally active IN is required. This finding implicated that the observed colocalization is not randomly taking place and that HIV-1 is actively docked to R-loop forming genomic sites. In addition, biochemical as well as computational meta data analysis revealed that HIV-1 RIGs are enriched in R-loops and that R-loop forming sites can accommodate integrated viral DNA. Further on, it was demonstrated that HIV-1 IN has R-loop binding capacity and is also capable of performing the strand transfer reaction on R-loop containing DNA templates. It was also demonstrated that R-loop depletion by RNase H1 overexpression in several cell lines, as well as in primary cells, significantly impairs HIV-1 integration, indicating that R-loop presence is crucial for efficient HIV-1 integration. In line with this result was the finding that RIGs expression was not affected by R-loop removal, indicating that only the presence of R-loops, as structural genomic elements, is more affecting HIV-1 integration compared to gene expression levels. The final finding is also in line with previous work from our lab.

In summary, the second part of this study provides strong evidence that R-loops represent structural genomic elements targeted by HIV-1 for integration and also gives new insight into HIV-1 IN functional

features which have not been addressed before.

Zusammenfassung

Die Infektion mit HIV-1 ist heutzutage nicht mehr ein Zustand mit tödlichem Ausgang. Dank moderner therapeutischer Maßnahmen ist die Infektion ein chronischer Zustand in dem die Viruslast in Patienten auf nicht nachweisbaren Niveau gehalten wird. Jedoch sind Patienten auf lebenslange Therapie angewiesen, ohne eine Chance auf Heilung. Die Eradikation integrierter viraler DNA bleibt die größte Herausforderung in der Heilung einer HIV-1 Infektion.
Das Ziel dieser Arbeit war es, zu einem besseren Verständnis und einer detaillierten Definition von HIV-1 Integrationsstellen beizutragen, als auch einen detaillierten Überblick über die Bedeutung und Relevanz von Chromatin Struktur und spezifischer genomischer Elemente im Kontext des HIV-1 Integrations Prozesses zu geben.

Der erste Teil dieser Arbeit befasst sich mit der Aktivität von G9a und der Komposition von Chromatin Merkmal H3K9me2 im Zusammenhang mit HIV-1 Integration in primären $CD4^+$ T-Zellen. Um den Effekt von G9a Aktivität und H3K9me2 Komposition in Zusammenhang mit HIV-1 integration zu ermessen, wurde BIX01294, als prominenter chemischer Inhibitor von G9a angewendet. BIX01294 ist auch als sehr effizientes Latenzumkehrmittel bekannt. In vorherigen Studien wurde gezeigt das die Aktivität von G9a und der Aufbau von Heterochromatin Markierung H3K9me2 einen signifikanten Einfluss auf die Entwicklung von T-Zellen haben, als auch einen großen Einfluss auf die Chromatin Architektur des Zellkerns. Als Teil dieser Arbeit wurde gezeigt dass die chemische Hemmung von G9a und die darauf folgende gehemmte Ablagerung von H3K9me2, durch BIX01294, eine signifikante Wirkung auf HIV-1 Integration haben. Die durch BIX01294 ausgelöste erhöhte HIV-1 Integration hatte auch eine erhöhte virale Transkriptionsaktivität zur folge, sowie eine räumliche Neupositionierung des Provirus von der bevorzugten Kernperipherie zum Kernzentrum. Eine ähnliche räumliche Neupositionierung im Zellkern wurde auch für Gene gezeigt die HIV-1 wiederholt für seine Integration selektiert (RIGs). Jedoch hatte die Neupositionierung von RIGs keinen Einfluss auf ihre Transkription, dass via RNA microarray Analyse gezeigt wurde. Gruppen von Genen die durch die Behandlung mit BIX01294 eine erhöhte Transkription zeigten gehörten hauptsächlich zu Genen die in Prozessen wie Eisenstoffwechsel oder Entzündungsreaktionen involviert sind. Darüber hinaus wurde gezeigt, dass HIV-1 Integrationsmuster nicht durch den Schwund von H3K9me2 beeinflusst werden, und dass das Virus immer noch ähnliche Regionen zur Integration favorisiert. Die Analyse der Verteilung von Chromatin Markierungen und der Chromatin Bindungselemente nach H3K9me2 Schwund ergab dass HIV-1 RIGs eine erhöhte Bindung von Chromatin Markierungen spezifisch für offenes Chromatin hatten, was zur Folge auch eine erhöhte LEDGF/p75 Bindung aufgewiesen hat, ausgelöst durch H3K9me2 Schwund nach der Behandlung mit BIX01294. Dieses Resultat könnte eine Erklärung für die erhöhte HIV-1 Integration darstellen da LEDGF/p75 ein prominenter Wirtszellenfaktor ist der in die virale Integration involviert ist.
Insgesamt zeigte dieser Teil der Arbeit, dass die Zugänglichkeit von Chromatin das HIV-1 Integrationsniveau signifikant beeinflusst, und dass das virale Integrationsniveau direkt proportional zu viraler Genexpression und virale Aktivität ist.

Der zweite Teil dieser Studie beschäftigt sich mit der Bedeutung von R-Loops als Strukturen favorisiert für die HIV-1 Integration in primären $CD4^+$ T-Zellen und Makrophagen. Es wurde gezeigt, das die HIV-1 IN in hohem Maße mit R-Loops in Zellen kolokalisiert und dass für ein hohes Auftreten dieser Kolokalisierung eine aktive IN erforderlich ist. Dies implizierte, dass die beobachtete Kolokalisierung nicht zufällig erfolgt und dass HIV-1 aktiv an R-Loops angedockt wird. Darüber hinaus zeigten molekulare Ergebnisse, als auch Computeranalysen dass HIV-1 RIGs in R-Loops angereichert sind und dass R-Loops integrierte virale DNA enthalten können. Weiterhin wurde gezeigt, dass die HIV-1 IN eine R-Loop Bindungskapazität besitzt und auch in der Lage ist die Strang-Transfer-Reaktion an einem R-Loop durchführen kann. Es wurde auch gezeigt, dass die R-Loop-Depletion durch RNase H1 in mehreren Zelllinien sowie in primären Zellen die HIV-1-Integration signifikant beeinträchtigt, was darauf hinweist dass die Präsenz von R-Loops für eine effiziente HIV-1-Integration von entscheidender Bedeutung ist. In

Übereinstimmung mit diesem Ergebnis war auch die Feststellung, dass die RIG-Expression nicht durch die Entfernung der R-Loops beeinflusst wurde, was darauf hindeutete dass die Anwesenheit von R-Loops, als Strukturen im Genom, die Integration von HIV-1 stärker beeinflusst als das Expressionsniveau dieser Gene. Dieses Resultat übereinstimmt auch mit bisherigen Erkenntnissen aus unserem Labor.
Zusammenfassend liefert der zweite Teil dieser Studie eindeutige Beweise dafür, dass R-Loops strukturelle genomische Elemente darstellen die sich HIV-1 zur Integration aussuchen kann. Außerdem liefert dieser Teil der studie auch neue Erkenntnisse über bisher nicht erforschte Funktionsmerkmale der HIV-1 IN.

List of Contents

I. Introduction

I.I. Human immunodeficiency virus 1 (HIV-1)

Human immunodeficiency virus 1 (HIV-1) is a lentivirus that primary infects cells of the host immune system. HIV primary infects $CD4^+$ T cells, macrophages and dendritic cells causing a depletion of immune cell numbers (mainly $CD4^+$ T cells) potentially causing a disease assigned as Acquired Immune Deficiency Syndrome (AIDS).
HIV-1 remains a major global and public health challenge (Sharp & Hahn, 2011; UNAIDS. AIDS by the numbers (2015)). More than 36 million individuals live with HIV world-wide. And about 2 million new infections are occurring each year (UNAIDS. Fact Sheet 2018. 1–6 (2018).

HIV-1 origin and phylogenetic position

Strong phylogenetic evidence, which can be traced back to a single transmission event, suggests that HIV-1 originated by zoonotic cross-species transmission of the simian lentivirus; Simian immunodeficiency virus (SIV) from the chimpanzee subspecies *Pan troglodytes troglodytes* to humans (Gao et al, 1999; Keele et al, 2006; Sharp & Hahn, 2011).

HIV-1 is characterized by extensive and dynamic genetic diversity. This heterogeneity is driven by several factors. A major source of HIV-1 genetic heterogeneity is the lack of proofreading ability of the viral enzyme reverse transcriptase (RT) (Roberts et al, 1988; Op de Coul et al, 2001), the rapid turnover of HIV-1 *in vivo* (Ho et al, 1995), host selective immune pressure (Michael, 1999), as well as recombination events during replication (Temin, 1993). Mentioned phenomena contribute to HIV-1 being a virus with several distinct molecular subtypes and recombinant forms. Such a high diversity of the virus has significant implications on general understanding of viral transmission, pathogenesis and diagnosis (Buonaguro et al, 2007).

According to the Baltimore classification of viruses (1971) which is based on structural features of viral genomes (type of nucleic acid, strandedness, sense and method of replication), HIV-1 is

classified into virus group VI (single stranded (ss) RNA-RT viruses, + strand (sense), with a DNA intermediate in its life cycle), family *Retroviridae*, genus *Lentivirus*.

Due to its variability, HIV-1 variants can be further classified into three major phylogenetic groups: group M (main), group O (outlier) and group N (non-M/non-O) (Gürtler et al 1994; Simon et al, 1998; Ayouba et al, 2000). Viruses classified into group M are responsible for most of the infections worldwide. This group can be further subdivided into 10 recognized phylogenetic subtypes (A-K).

Classification of HIV-1 subtypes was originally based on subgenomic regions of individual genes. Nowadays, with sequencing method improvements, HIV-1 phylogenetic classification is based on nucleotide sequences derived from multiple subgenomic regions (*gag*, *pol* and *env* gene) of the same viral isolate or on full-length sequence analysis (Buonaguro et al, 2007) (Figure 1.1.).

Apart from HIV-1, in 1986 a related virus was isolated from African individuals and named HIV-2, which is known to be less pathogenic than HIV-1 (Clavel et al, 1986) and has two major subtypes, A and B. The lower infectivity of HIV-2 is assigned to its relatively poor capacity for transmission. HIV-2 is largely confined to West Africa (Reeves & Doms, 2002) (Figure 1.1.).

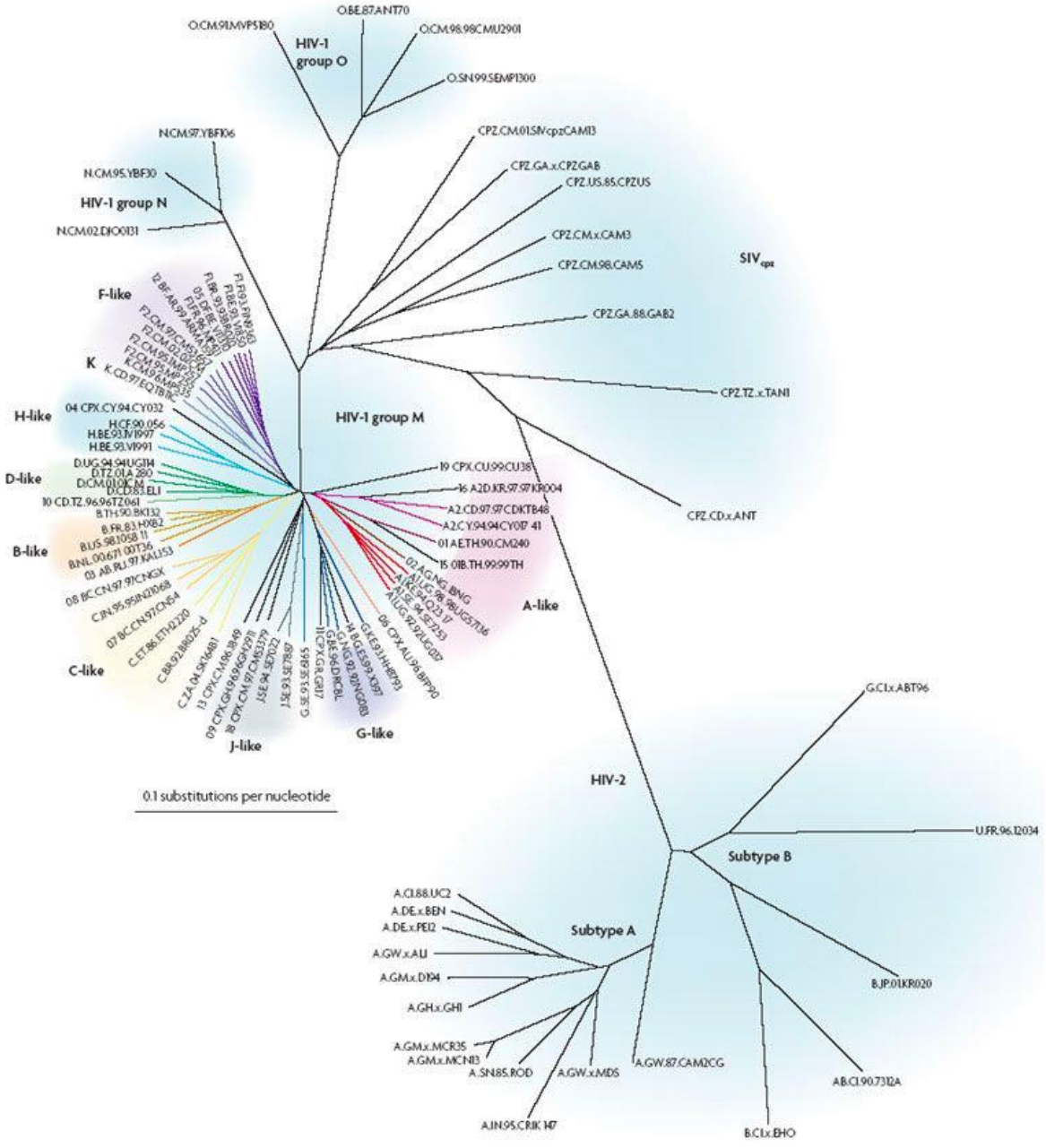


Figure 1.1. Phylogenetic tree of HIV subtypes and simian lentiviruses.

The phylogenetic tree shows the similarity between HIV and SIV generated by full genome alignment of 87 human and simian lentiviruses (from: Arien et al, 2007).

HIV-1 & AIDS

A large sero-epidemiological study conducted in 1984 clearly showed that HIV causes AIDS. In addition, this research was the basis of the development of the first diagnostic tests for HIV infection (Schupbach et al, 1984; Brun-Vezient et al, 1984).

From a historical point of view, an estimated number of 75 million people worldwide have been infected with HIV. Untreated HIV replication causes progressive CD4 T cell loss and a wide range of immunological abnormalities, which can also lead to an increased risk of infectious and oncological complications. Since the mid-1980s, when the first successful reverse transcriptase inhibitor was produced, antiretroviral drugs are highly effective in inhibiting HIV replication. Viral suppression enables immune recovery and the near elimination of the risk for developing AIDS (Deeks et al, 2015).

Still, the problem of not being able to fully eradicate the virus from infected individuals persists. Apart from mentioned reasons for HIV-1 heterogeneity which represent one of the difficulties for HIV-1 cure strategies and beside the viral immune escape, the major reason why HIV-1 infection can not be eradicated by current therapies is due to the property of the virus to establish latent infection in resting CD4 T cells. These latently infected cells do not express viral proteins and therefore remain undetectable for the immune system and refractory to the current antiretroviral therapies (Chun et al, 1995; Lassen et al, 2004; Williams & Greene, 2007). In addition, non-T cell derived viral reservoir represents another challenge in eradicating the virus from infected individuals. Several studies showed that the viral genome can also be found in different cell types, such as monocytes/macrophages (Le Douce et al, 2010; Kumar et al, 2014), other myeloid cells (Scadden et al, 1990; Churchil et al, 2006) and even epithelial cells (Veazey et al, 1998; Liu et al, 2008; Liu et al, 2013a).

I.II. HIV-1 structure

HIV-1 virion structure

The HIV-1 viral particle (virion) has a spherical shape with a diameter of 100 nm. HIV-1 is an enveloped virus in which the envelope functions as a protector for inner viral components and contains proteins that allow the virus to recognize host cells. Inside the virion viral structural proteins assemble into a shell, or the viral capsid (CA) which protects and organizes the viral genome. HIV-1 virions can adopt two distinct morphological states (Ganser-Pornillos et al, 2008; Briggs & Kräusslich 2011; Sundquist & Kräusslich 2012). First the virion buds from a producer cell in its immature form, having a spherical CA which is composed of the immature (precursor) structural protein Gag. In order to become infectious, the immature virion undergoes morphological changes into the mature form (Figure 1.2.). This process of HIV-1 maturation is triggered by site-specific proteolytic processing; cleavage of the precursor Gag protein by the viral protease. This cleavage creates new structural proteins among which the capsid protein or p24 is one of the most important ones and it resembles into a new cone-shaped CA. The mature CA contains the viral RNA genome which is associated to enzymes important for integration (integrase – IN) and reverse transcription (reverse transcriptase – RT). The mature CA and its contents together constitute the virion core. The core is the structural part which is introduced into the cytoplasm of an infected cell at the start of a new round of viral replication.

Apart from the core, which contains two copies of the single stranded viral RNA genome and is closely bound to nucleocapsid proteins (Figure 1.2., right), the virion contains viral enzymes IN, RT, proteases, ribonucleases, as well as an envelope.

The envelope is composed of a lipid bilayer which originates from the host cell plasma membrane in the process of budding. The envelope of a virion contains proteins inherited from the host cell and some copies of the viral envelope protein. The viral envelope protein is composed of three molecules of glycoprotein 120 (gp120) and three molecules of glycoprotein 41 (gp41) (Figure 2; legend on the right). The function of viral envelope proteins is the attachment to a host target cell, fusion of the lipid membranes of the viral particle and host cell membrane and thereby release the viral core into the infected cell (Chan et al, 1997).

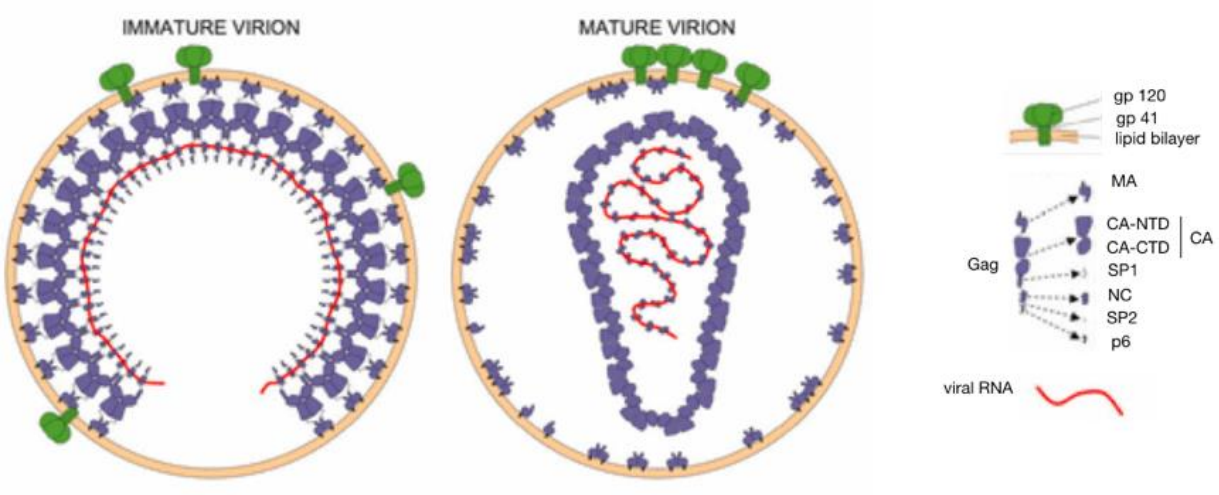


Figure 1.2. Schematic representation of HIV-1 virion structure.

Left: Structure of an immature HIV-1 virion. The viral capsid is not yet assembled. Gag molecules (purple) are concentrated under the viral envelope (yellow and green) and in complex with the viral RNA genome (red).

Right: Structure of a mature HIV-1 virion. The viral core is assembled upon Gag cleavage by the viral protease. Upon protease cleavage nuclear core (NC) proteins (purple dots) associate with the viral RNA genome (red) and mature matrix proteins (MA) are docked to the viral envelope (yellow and green).

A more detailed composition of HIV-1 virion building blocks is shown in the legend on the right (adapted from: Pornillos & Ganser-Pornilos, 2013).

Structure of the HIV-1 genome

The genome of HIV-1 is a coding RNA with nine open reading frames which code for fifteen different proteins (Coffin et al, 1997). The size of the HIV-1 HXB2 complete genome is 9719 bp (Genbank accession number K03455.1). The HIV-1 genome contains three major genes *gag*, *pol* and *env* which encode the main structural proteins and enzymes of the virus (Figure 1.3).

The 5' and 3' ends of the viral genome each contain a Long Terminal Repeat (LTR) sequence, which participate in the process of proviral integration into the host genome. Both LTRs are composed of identical nucleotide sequences, and divided into three identical regions (U3, R, U5). However, once the virus has been integrated, the 5' end LTR serves as a promoter for the viral genome, and the 3' LTR is important in the polyadenylation of the transcribed viral RNA (Krebs et al, 2001). In theory, both LTRs can function as transcriptional promoters, since they have the

same nucleotide sequence. Still, it has been suggested that the transcriptional activity of the 5' LTR is more potent in comparison to the 3' LTR, which is also very similar to the scenario occurring in other retroviruses (Klaver & Berkhout, 1994).

Viral proteins are translated as polyproteins which are further processed and proteolytically cleaved to produce the mature proteins. From the *gag* polyprotein precursor, the matrix protein (MA or p24), capsid (CA or p7), nucleocapsid (NC or p17) and p6 protein are generated. From the *gag-pol* polyprotein important viral enzymes as protease (PR), reverse transcriptase (RT) and integrase (IN) are generated. The *env* gene encodes a 30 amino acid signal peptide – SP and two glycoproteins assigned as gp120 and gp41. Apart from the three major HIV-1 genes, additional six genes (*tat*, *rev*, *nef*, *vpr*, *vif*, *vpu*) encode HIV-1 accessory proteins which are participating in the life cycle of HIV-1 (Figure 1.3). Inside of virions, the HIV-1 genomic RNA is found as a non-covalent dimer with a 5' cap, a 3' polyadenylated tale and annealed to a host $tRNA^{Lys3}$ molecule (Coffin et al, 1997; Watts et al, 2009). The $tRNA^{Lys3}$ acts as a primer for RT-catalyzed synthesis of viral DNA (vDNA) (Kleiman, 2002).

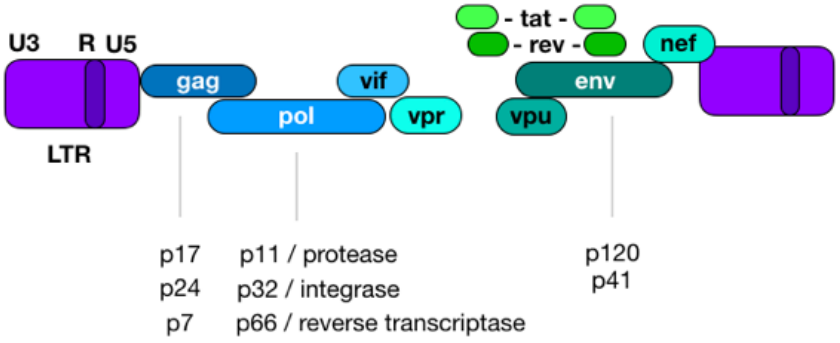


Figure 1.3. Scheme of HIV-1 genome structure and its functional products. HIV-1 genome is 9719 bp in size and contains 9 open reading frames. Viral LTRs are represented in gray rectangles, whereas the viral genes are depicted in colorful rectangles. The functional products of the three major viral genes (*gag*, *pol* and *env*) are listed below them.

I.III. HIV-1 life cycle

The life cycle of HIV-1 is a multistep process which can be divided into eight stages:

1) binding;
2) fusion;
3) reverse transcription;
4) integration;
5) replication;
6) assembly;
7) budding and
8) maturation (Figure 1.4).

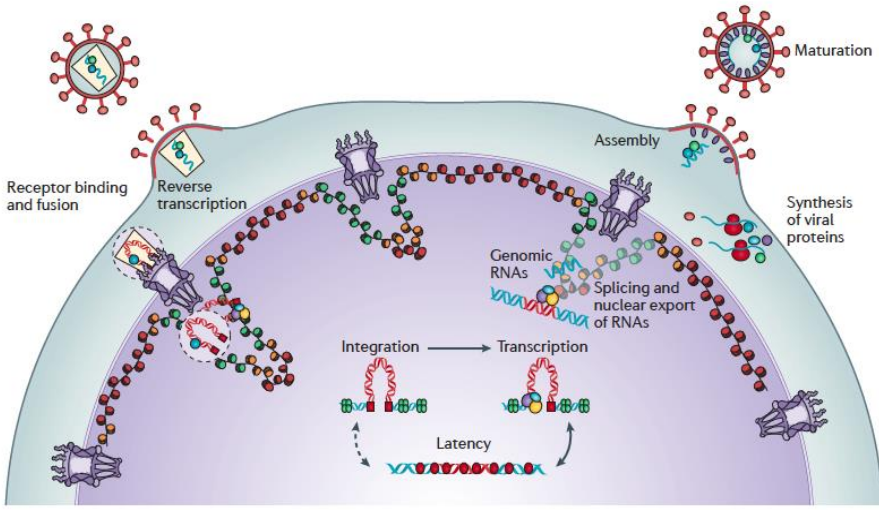


Figure 1.4. Overview of HIV-1 life cycle.
The illustration depicts the most important stages in the life cycle of HIV-1 from the moment of entering the target cell to possible scenarios of an active transcription of the virus and virion production, or reversion into latency (from: Lusic & Siliciano, 2017).

Binding & fusion

As already mentioned, HIV-1 is an enveloped virus. The viral envelope is formed during the budding process and contains both viral and cellular proteins. Viral proteins present in the envelope are gp120 or the surface subunit of the envelope and gp41, the transmembrane subunit. gp120 interacts with specific cellular receptors; the CD4 receptor and two chemokine receptors CCR5 and CXCR4 (Dalgleish et al, 1984; Klatzmann et al, 1986; Alkhatib et al, 1996; Choe et al, 1996; Endres et al, 1996; Deng et al, 1996, Feng et al, 1996). Based on its ability to bind these receptors HIV-1 can be divided into different strains; CCR5 (R5 strain), CXCR4 (X4) strain or both (R5X4).

Previously it has been assumed that HIV-1 directly fuses with the plasma membrane of the target cell (Stein et al, 1987; McClure et al, 1988; Melikyan, 2008). However, fusion occurs after a conformational change in gp41 (the transmembrane unit of the HIV-1 envelope) (Chan & Kim, 1998). But it was also shown that HIV-1 entry can occur via virion endocytosis (Miyauchi, 2009).

After HIV-1 virion fusion, uncoating steps take place, which means that the viral CA is disassembled and the viral genome is released into the cytoplasm.

Reverse transcription

Reverse transcription is the process in which the viral RNA genome is reversely transcribed into cDNA (complementary DNA). The produced cDNA represents the substrate for the integration process. The most important enzyme involved in the process of reverse transcription is the viral enzyme reverse transcriptase (RT) (Baltimore, 1970).

Two important features of RT are necessary to carry out reverse transcription;

1) it is a DNA polymerase that can copy either an RNA or DNA template and

2) it has RNase H properties - it can degrade RNA only if it is a part of a DNA-RNA duplex.

RT requires a primer and a template for its function. The viral genomic RNA is plus-stranded and the primer for the first minus-stranded DNA is a host cell $tRNA^{Lys3}$ whose 3' is complementary to a sequence close to the 5' end of the viral RNA, known as the primer binding site (pbs). The ends of the viral RNA are direct repeats, called "R". They function as a bridge that allows the newly synthesized

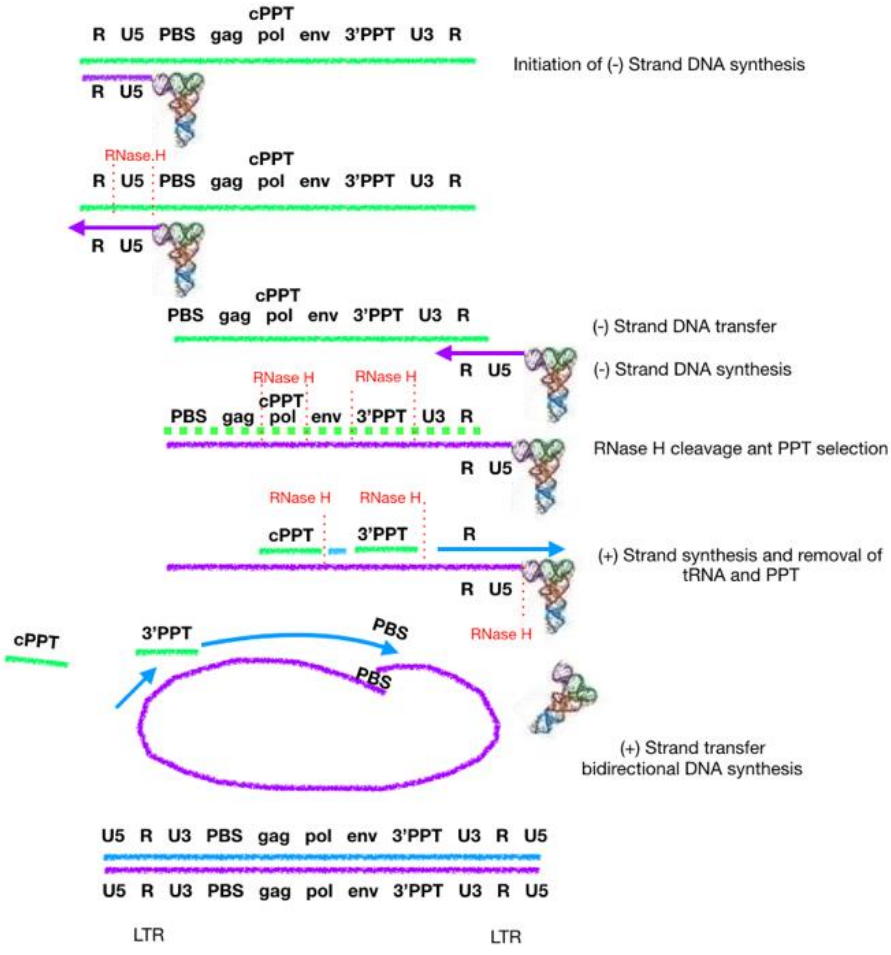


Figure 1.5. Schematic representation of HIV-1 reverse transcription process

minus-strand DNA to be transferred to the 3' end of the viral RNA, where the synthesis of the minus-strand can continue along the genome. The HIV-1 RNA genome contains two purine-rich regions (the so called polypurine tract – ppt) that are resistant to RNase H degradation by the RT. This ppt region functions as a primer for the initiation of the plus-strand DNA synthesis. After reverse transcription a DNA product is generated that is longer than the initial RNA genome. Both ends of the DNA contain sequences from each end of the RNA template (U3 from the 3' end and U5 from the 5' end), meaning that each end of the viral DNA has the same sequence; U3-R-U5. These are the Long Terminal Repeats (LTRs), that represent the ends of the integrated provirus (Hu & Hugens, 2012) (Figure 1.5).

Integration

After reverse transcription, the viral cDNA is actively transported from the cytoplasm to the nucleus, through the nuclear pore complex (NPC) in a large nucleoprotein complex known as the Preintegration Complex (PIC). The PIC is derived from the core of the infecting virion (Bowerman et al, 1989) and, apart from the viral DNA, contains the viral integrase (IN), the enzyme responsible for the process of integration.

In vitro studies showed that IN requires magnesium for its activity. The CCD (Catalytic Core Domain) of HIV-1 IN contains the DDE (aspartic acid - aspartic acid - glutamic acid) motif essential for the catalytic function of the enzyme (Kulkosky et al, 1992; Leavitt et al, 1993). The amino acid residues of the DDE motif require a pair of Mg^{2+} ions for its catalytic function in the process of integration and the presence of Mg^{2+} ions was shown to be very important in the stabilization of the intasome (Miri et al, 2014; Engelman & Cherepanov, 2014).

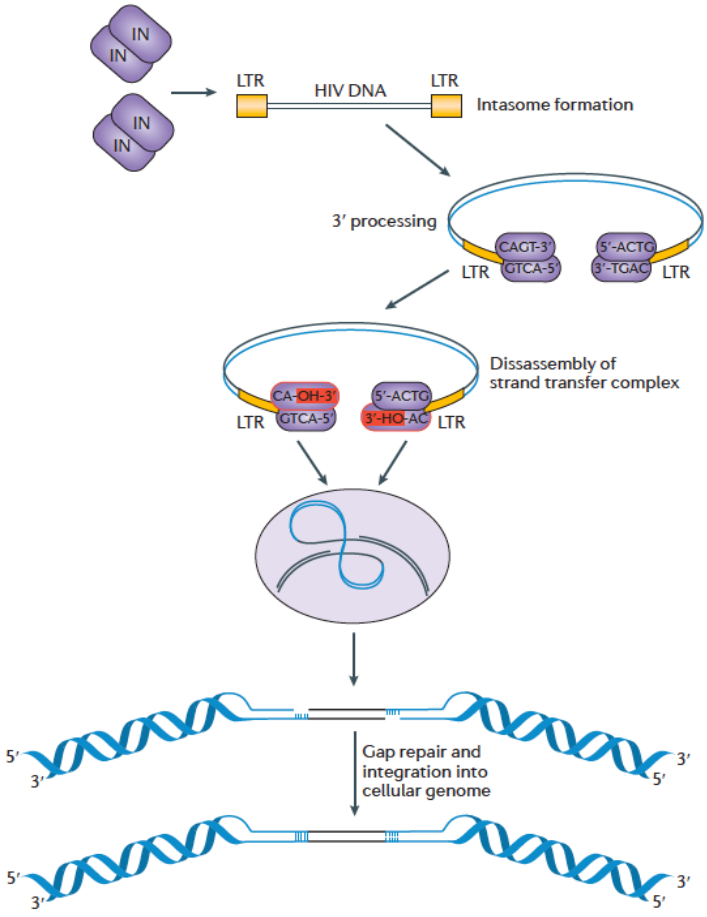


Figure 1.6. Scheme of HIV-1 integration (from: Lusic & Siliciano, 2017).

In addition, studying integration intermediates revealed DNA cutting and joining steps of the viral DNA during the process of integration (Fujiwara & Mizuuchi, 1988; Brown et al, 1989).

Before the start of any enzymatic activity of IN and the integration of the vDNA into the target host DNA, the intasome has to be formed. "The intasome comprises a homo-hexadecamer of IN with a tetramer-of-tetramers architecture featuring eight structurally distinct types of IN protomers supporting two catalytically competent subunits" (Ballandras-Colas et al, 2017), which stably bind and assemble on both ends of the vDNA (Lesbats et al, 2016).

Upon intasome formation, the first step of integration represents the 3' end processing and removal of two nucleotides from both 3' ends of the viral DNA. The resulting 3' ends of the vDNA in all cases terminate with the conserved CA-3' sequence. The second step of integration is the DNA strand transfer in which the "sticky" 3' ends of the vDNA attack a pair of phosphodiester bonds on opposite strands across the major groove of the host target DNA. At this

intermediate step, the 3'ends of the vDNA are covalently joined to the target DNA. The single strand gaps and the two-nucleotide overhang at the 5' end of the vDNA have to be repaired by cellular enzymes in order to complete the integration process. The joining sites of the two target DNA strands in the case of HIV-1 IN activity are five base pairs apart, resulting in a five base-pair duplication (Brown et al, 1987; Bushman et al, 1990; Engelman et al, 1991) (Figure 1.6).

Viral replication, assembly, budding & maturation

Upon successful integration, the viral genome can be actively transcribed and new infectious virions can be produced. During active transcription of HIV-1, the viral mRNA is transported out of the nucleus and viral proteins are translated in the cytoplasm. The produced viral proteins are involved in the process of virion assembly and budding from the host cell membrane.

Virion assembly and release of HIV-1 virions are processes that occur predominantly at the plasma membrane of infected cells (Sundquist & Krausslich, 2012; Freed 2015). The building "blocks" of an HIV-1 virion are composed of two copies of the viral RNA genome, which are associated to cellular $tRNA^{Lys,3}$ to prime later on cDNA synthesis, viral Env protein, the Gag polyprotein and the three major viral enzymes: IN, RT and PR. The Gag polyprotein is the main mediator in the process of virion assembly, which means that it binds the host cell plasma membrane and mediates protein-protein interactions that promote virion assembly. Via its binding to the host cell plasma membrane, the Gag polyprotein affects the docking and concentration of Env proteins and the packaging of the viral genome via its RNA packaging sequence (Ψ) (Sundquist & Kräusslich, 2012).

Even though the viral Gag polyprotein is the main actor in mediating virion assembly, in the process of virion budding the virus takes advantage of the host cell endosomal machinery, the so called endosomal sorting complex required for transport (ESCRT). The ESCRT pathway is to terminate Gag polymerization and catalyze virion release (Morita & Sundquist, 2004; Bieniasz, 2009, Peel et al, 2011).

Maturation of the viral particle starts simultaneously with the budding process or right after budding. The main mediator in this process is the viral PR. PR cleaves off the Gag and Gag-Pro-Pol polyprotein in order to produce fully processed viral MA, CA, NC, p6, PR, RT and IN (Hill et al, 2005). During the process of maturation, the mentioned viral proteins undergo significant rearrangements in order to shape the mature and infectious viral particle (Figure 1.2.).

I.IV. HIV-1 integration

Viral DNA variants

Successful integration of the viral genome into the host cell genome is a prerequisite for productive HIV-1 infection. However, the viral cDNA can also persist as non-integrated viral DNA, which is ubiquitously formed in infected cells (Hamid et al, 2017).

Non-integrated vDNA can be represented in the form of 2LTR or 1LTR circles (2 or 1 Long Terminal Repeat circles). Most 2LTR circles are formed due to the process of autointegration - meaning that the viral DNA integrates into itself creating modified forms of double-stranded DNA circles (Lee & Craigie, 1994). A minor fraction of 2LTRs is created in the process of the end-joining reaction of linear viral DNA, whereas 1LTR circles have been shown to represent a product of recombination between the 5' and 3' LTR of the vDNA (Farnet & Haseltine, 1991), but can also be formed due to aberrant reverse transcription by strand displacement failure (Hu & Hughes, 2012). Non-integrated forms of vDNA do not support viral replication and represent a dead-end of the viral life cycle (Farnet & Haseltine, 1991).

Certain cellular factors contribute to the formation of non-integrated vDNA forms. Cellular nucleases as RAD50, MRE11 and NBS1 have been shown to be involved in the formation of 1LTR circles (Sloan & Wainberg, 2011). The inactivation of factors of the DNA repair machinery as Ku70/80, ligase IV and XRCC4 causes an impairment of 2LTR circle formation indicating their importance in the circularization reaction necessary for the creation of those non-integrated viral DNA forms (Li et al, 2001; Jeanson et al, 2002). In addition, 2LTR circles can be formed as a result of cellular DNA repair mechanisms activated due to foreign DNA sensing (Shoemaker, et al 1980). Interestingly, a host cell protein, BAF, involved in the process of DNA condensation, was *in vitro* shown to block auto integration of vDNA (Yan et al, 2009).

The persistence of non-integrated DNA forms of HIV-1 in infected cells is debated and different studies demonstrate opposing results in that regard (Butler et al, 2002; Sharkey et al, 2005; Murray et al, 2012; Pace et al, 2013). However, an observed rapid drop of 2LTRs in T cells is possibly a consequence of cell division and not direct degradation of 2LTR circles (Stein et al, 2016). In contrast, macrophages, as non-dividing cells, represent a certain reservoir of 2LTR circles where persistent levels of 2LTR circles have been detected up to 21 days after infection (Gillim-Ross et al, 2005).

Non-integrated vDNA forms are usually studied as a marker for nuclear import (Bushman, 2003; Hamid et al, 2016), as well as a marker for on going viral replication in patients on combined Antiretroviral Therapy (cART) (Pace et al, 2013). In a functional context, non-integrated forms of vDNA represent a dead end of viral replication.

Factors involved in HIV-1 integration

The entire host genome is in theory available for HIV-1 integration. However, certain genomic features seem to be preferred for integration in comparison to others. Integration site selection of retroviruses does not seem to depend on the underlying sequences of the host genome, but it is also not a random process (Mitchell et al, 2004). Sites targeted by HIV-1 for integration are frequently actively transcribed genomic regions (Schröder et al, 2002; Ikeda et al, 2007; Brady et al, 2009; Maldarelli et al, 2014; Wagner et al, 2014) associated to marks of active transcription as histone H3/H4 acetylation (Wang et al, 2007; Marini et al, Lucic et al, 2019). Regions targeted by HIV-1 for integration are mainly composed of nucleosomal DNA (Pruss et al, 1994), contain repetitive genomic elements (Stevens & Griffith, 1994), and are mainly in introns of transcribed genes (Cohn et al, 2015).

In the context of structured elements of the genome, HIV-1 integration sites were shown to be outside of centromeric regions (Carteau et al, 1998) and lamina associated domains (LADs) (Guelen et al, 2008; Marini et al, 2015; Achuthan et al, 2018), and in line with that, away from repressive chromatin marks as H3K27me3 or DNA CpG methylation (Wang et al, 2007).

In addition, HIV-1 integration site selection seems to be affected by the organization of the nuclear architecture. Apart from integrating into transcribed genes, especially in their intron

regions (Han et al, 2004; Cohn et al, 2015), the virus also shows integration preference for genes harboring marks of structured elements as enhancers (Wang et al, 2007; Chen et al, 2017b). Spatially, HIV-1 preferentially occupies the outer shell of the host cell nucleus where it binds elements of the nuclear pore complex (Marini et al, 2015). In contrast, a recent study demonstrated that CPSF6 acts as the main nuclear viral trafficking factor, leading the virus towards gene rich regions away from the nuclear periphery. In addition, the loss of CPSF6 caused viral integration into lamina associated domains (LADs) (Achuthan et al, 2018). However, this study was performed in HEK293T cells (Human Embryonic Kidney cell line), which are not primary targets of HIV-1. The HIV-1 integration site dataset used in this study (Sowd et al, 2016), generated from HEK293Tcells, was cross-compared with DamID (DNA adenine methyltransferase identification) data (Meuleman et al, 2013), generated from human embryonic stem cell line SHEF2, in order to show the lack of association of viral integration sites with the NPC. In addition, in order to underlie the deterministic role of CPSF6 in nuclear positioning of the virus, CPSF6, LEDGF7p72 knockout (KO) cells, as well as double knockout cells of HEK293T and $CD4^+$ T cells were included into the integration site and FISH/imaging analysis. The results showed that integration sites retrieved from KO cells represent mainly cell type specific genes, and the FISH results demonstrated that the majority of those genes indeed have a perinuclear position (Achuthan et al, 2018). Accordingly, cell activation status as well as cell type specificity seem to be important factors that need to be taken into account in the context of HIV-1 integration. Those factors can significantly affect the viral life cycle, but also the fate of the host cell (Churchil et al, 2006; Liu et al, 2008; Le Douce et al, 2010). In the next section the most important viral factors and host cell factors included in the process of HIV-1 integration will be described.

HIV-1 integrase

HIV-1 integrase (IN) is a protein containing 288 amino acids (32 kDa) which is encoded by the end of the viral *pol* gene. It is synthesized as a part of the Gag-Pol polypeptide precursor by viral protease-mediated cleavage.

The enzyme contains three independent functional domains; (1) the N-terminal domain (NTD) that carries an HHCC motif which is analogous to a zinc finger and very effectively binds Zn^{2+}

(Zheng et al, 1996). This domain possibly favors protein multimerization which is a key step in viral integration (Zheng et al, 1996; Lee et al, 1997).

(2) The central domain or the catalytic domain (CCD) of the enzyme contains a D, D-35, E motif which is necessary for the catalytic activity and represents a domain which is highly conserved between retroviral IN enzymes and transposases. The CCD is essential in vDNA binding (residues Q148, K156 and K159) (Engelman et al, 1993; Jenkins et al, 1997; Heuer & Brown, 1997; Drake et al, 1998; Esposito & Craigie, 1998; Johnson et al, 2006). Structurally the CCD shows an RNase H folding pattern. The RNase H-like fold was shown to be one of the evolutionary oldest protein folds (Ma et al, 2008). The Ribonuclease H-like (RNHL) superfamily members (called also the retroviral integrase superfamily) share the position of active site residues, which typically include aspartic acid, glutamic acid and in some cases histidine (Yang et al, 1990; Majorek et al, 2014), which is also the case for HIV-1 IN, where the CCDs active site is composed of Asp64, Asp116 and Glu152. The active site contains a conserved spacing of 35 amino acids between Asp116 and Glu152 (D,D(35)E motif). During the integration process the CCD interacts with the vDNA and the target DNA. Mutations in the D,D(35)E motif cause inhibition of IN activity (Kulkosky et al, 1992; Leavitt et al, 1993). Importantly, all IN activities require the presence of a metallic cation cofactor which is coordinated by two residues of the catalytic triad (D64 and D116) (Goldgur et al, 1998; Maignan et al, 1998).

(3) The C-terminal domain (CTD) binds non-specifically to DNA and is mainly involved in stabilizing the complex of IN and the DNA. Residues 220-270 are responsible for the DNA binding function. The CTD contains three lysine residues, in particular, lysine residues at position 264, 266 and 273, which can be acetylated. These post-translational modifications regulate IN strand transfer activity and viral integration (Cereseto et al, 2005; Lusic & Siliciano, 2017). It has also been demonstrated that the same three lysine residues (K264, K266 and K273), have RNA binding properties and when mutated cause the production of non-infectious viral particles (Kessl et al, 2016).

The function of IN is crucial in the process of HIV-1 integration. However, the enzyme itself is only functional in its multimeric form (van Gent et al, 1993; van den Ent et al, 1999), structurally forming a homo-hexadecamer (Ballandras-Colas et al, 2017). The presence of Zn^{2+} ions was

shown to be very important in the process of IN multimerization, but also in facilitating the Mg^{2+} ion-dependent catalytic activity of the viral enzyme (Lee et al, 1997).
Integration starts upon intasome formation in which IN stably binds both ends of the vDNA (Lesbats et al, 2016). Also, apart from the crucial presence of Mg^{2+} ions for the catalytic function of IN, they have been shown to have an important role in intasome stabilization (Miri et al, 2014; Engelman & Cherepanov, 2014).

Other factors involved in HIV-1 integration

Apart from IN other viral and host cell factors are involved in enabling the integration process of HIV-1.

Viral factors

Most of the viral proteins which are known to participate in the integration process are involved in the nuclear entry of the PIC. Among these proteins are CA, MA, Vpr and the central polypurine tract (Suzuki & Craigie, 2007; Matreyek & Engelman, 2013).

Capsid

The HIV-1 CA is involved in early viral post entry events and participates in PIC moving all the way to the nucleus of the infected cell, while viral uncoating is taking place (Campbell & Hope, 2015). During that process, several proteins interact with CA, and very likely, affect the navigation process towards the nucleus (Di Nunzio et al, 2012; Hilditch & Towers, 2014; Chin et al, 2015).
The CA protein has an important role in the PIC nuclear entry, especially in non-dividing cells as macrophages (Yamashita & Emerman, 2004). Via its interaction with other cellular factors, CA was shown to affect HIV-1 integration profiles in infected cells (Schaller et al, 2011; Koh et al, 2013; Sowd et al, 2016).
Cyclophilin A (CypA) is a prominent example of a protein that via its direct interaction with CA affects HIV-1 infection (Sokolskaja et al, 2004; Hatziioannou et al, 2005). Interestingly, the

effects of Cyp-CA interactions seem to be cell type dependent (De laco & Luban, 2014). The knockout of CypA impairs HIV-1 infectivity in Jurkat cells, whereas viral infectivity is not affected in other cell types (Braaten & Luban, 2001). CypA can also have a stabilizing effect on the viral CA core by binding to it and thereby causing a delay in the uncoating process, but it can at the same time have a destabilizing effect too (Li et al, 2009; Shah et al, 2013; De laco & Luban, 2014). What kind of effect Cyp-CA interactions may have in the context of CA core stabilization seem to be regulated by dose-dependent presence of CypA (Li et al, 2009; Liu et al, 2016).

A colocalization between CA and CPSF6 (host cell factor involved in RNA binding and processing) has also been demonstrated in the cytoplasm (Peng et al, 2014). In addition, the depletion of CPSF6 causes a decrease in viral infectivity (Lee et al, 2010; Sowd et al, 2016), a reduction in vDNA and CA localization in the nucleus (Chin et al, 2015), as well as decreased viral integration (Sowd et al, 2016).

Moving from the cytoplasm closer to the nucleus, it was shown that CA directly interacts with NUP358, a component of the NPC, via the NUPs cyclophilin homology domain (CHD) (Schaller et al, 2011; Di Nunzio et al, 2012). CA does not contain a nuclear localization signal (NLS) and might localize to the nucleus by its interaction with NUP358 on the outer side of NPC (Lee et al, 2010; Schaller et al, 2011). In addition, in the context of nuclear entry, it has been demonstrated that CA directly binds another protein of the nuclear pore complex - NUP153 (Matreyek et al, 2013). According to demonstrated interactions of CA and nuclear pore complex proteins, as well as the fact that the HIV-1 integration process has to occur rapidly after nuclear entry and that it occurs in the outer shell of the nucleus (Di Primo et al, 2013; Marini et al, 2015), it is evident that HIV-1 integration is also, in part, regulated by CA and proteins interacting with CA (Matreyek & Engelman, 2013).

Further on, it has recently been shown that CA stays associated to nuclear PICs in primary macrophages and its presence also affects viral infectivity (Peng et al, 2014; Bejarano et al, 2019). Mentioned findings indicate that CA is not only playing a role in the docking of PICs to the nuclear envelope and NPC, but also exerts functions inside the nucleus. In line with that are recent findings showing that CFSF6 via its interaction with CA functions as the main regulator of PIC localization in the nucleus, leading it to transcriptionally active chromatin (Sowd et al,

2016), but also moving it away from heterochromatic regions at the nuclear periphery and more towards gene rich genomic regions in the nuclear interior (Achuthan et al, 2018).

Matrix

The MA protein was shown to be implicated in the nuclear import of HIV-1 cDNA, given the fact that it contains two nuclear localization signals (NLSs) (Burinsky et al, 1993; Nadler et al, 1997; Hearps & Jans, 2007). However, the role of MA in the nuclear import does not seem to be crucial since it was shown that HIV-1 can still infect macrophages even without having a functional MA (Reil et al, 1998).

Vpr

Vpr is a small virion-associated, nucleocytoplasmic shuttling regulatory protein, which is best known for its function of arresting the cell cycle in G2 phase that is completed by the induction of apoptosis (Rogel et al, 1995). The cause of cell cycle arrest activates the host cell DNA repair machinery, which plays an important role in completing the integration process (Votteler & Schubert, 2008; Yan et al, 2019). It is still under debate how Vpr can induce apoptosis, but it has also been demonstrated that uninfected bystander cells can be targeted by Vpr (Cummins & Badley, 2010; Abbas, 2013).
Vpr was one of the first viral proteins that were shown to be associated in the docking process of the PIC to the NPC and binding to NUPs (Popov et al, 1998; Fouchier et al, 1998; Lusic & Siliciano, 2017).

Host cell factors

LEDGF/p75 (lens epithelium-derived growth factor)

LEDGF/p75 is probably the most important host cell factor involved in HIV-1 integration. LEDGF/p75 was first identified as a transcriptional mediator protein that promotes activator-dependent transcription *in vitro* (Ge et al, 1998a; Craigie & Bushman, 2012). In a series

of different studies LEDGF/p75 was identified as a cellular protein that tightly binds HIV-1 IN (Cherepanov et al, 2003; Turlure et al, 2004; Emiliani et al, 2005).

LEDGF/p75 is ubiquitously expressed by the PSIP1 gene and represents a protein containing 530 amino acids (Ge et al, 1998a). Apart from LEDGF/p75, cells also express a shorter splice variant - LEDGF/p52. Both proteins share the first 532 amino acids of the N-terminus, but contain unique C-terminal ends (Ge et al, 1998a; Ge et al, 1998b). Only the longer LEDGF/p75 form was found to directly interact with HIV-1 IN (Cherepanov et al, 2003; Cherepanov et al, 2005), via its IN-binding domain located at the C-terminus of the protein, which is missing in the shorter p52 variant (Cherepanov et al, 2005; Busschots et al, 2007).

In a series of studies it was also demonstrated that the interaction between LEDGF/p75 and IN is important for the nuclear localization of IN (Cherepanov et al, 2003; Maertens et al, 2003; Llano et al, 2004a). In addition, IN mutants, designed to be unable to bind to LEDGF/p75, lose their ability to associate to chromatin (Emiliani et al, 2005). Imaging studies showed that LEDGF/p75 can be bound to condensed chromosomes during mitosis. Accordingly, it has been demonstrated that in the presence of LEDGF/p75, HIV-1 IN could accumulate on chromatin too (Maertens et al, 2003; Craigie & Bushman, 2012). In line with this finding is the described ability of LEDGF/p75 to specifically read the histone modification H3K36me3 via its PWWP domain (Pradeepa et al, 2012; Eidahl et al, 2013), which gives strong indication that LEDGF/p75 is responsible for the tethering of IN to chromatin.
In addition, LEDGF/p75 was shown to facilitate integration in gene bodies (Sowd et al, 2016), which are usually decorated with H3K36me3 - a mark of transcribed genomic regions associated with elongating RNA Pol2 (RNA Polymerase 2) (Venkatesh et al, 2012). LEDGF/p75 also, apart from its function of tethering IN to chromatin, protects IN from degradation and evidently affects HIV-1 integration profiles (Llano et al, 2004b; Poeschla, 2014).
Structural analysis of LEDGF/p75 showed that the PWWP chromatin-binding domain is located at the N-terminus of the protein, together with an NLS, as well as two copies of an A/T hook domain which is involved in DNA binding.

CPSF6 (Cleavage and polyadenylation specificity factor 6)

CPSF6 is a member of the Pre-messenger RNA cleavage factor I complex and takes place in mRNA polyadenylation site determination (Rüegsegger et al, 1996). CPSF6 is predominantly localized in the nucleus where it interacts with HIV-1 CA (Lee et al, 2010; Fricke et al, 2013), but has also been shown to interact with CA in the cytoplasm (Peng et al, 2014). The protein is composed of the N-terminal domain, which is involved in RNA recognition, a proline-rich central domain, and a C-terminal domain rich in arginine and serine residues (S/R). The C-terminal S/R-rich domain is involved in the binding of CA (Price et al, 2012). Apart from its interaction with HIV-1 CA it was shown that CPSF6 can be implicated in the nuclear import of the PIC (Lee et al, 2012; Bhattacharya et al, 2014; Rasheedi et al, 2016), and also seems to play a role in the targeting of HIV-1 integration sites (Achuthan et al, 2018). In addition, it has been shown that CPSF6 depletion decreases the nuclear localization of vDNA and CA (Chin et al, 2015), and has a decreasing effect on viral integration (Sowd et al, 2016). Certain amino acid substitutions in the viral CA, as N74D and A77V, which are located in the CPSF6 RNA binding domain, have been identified having a decreasing effect on the interaction between virus and host cell (Lee et al, 2010; Saito et al, 2016).
Although CPSF6 is not considered to be essential for viral integration, CPSF6-binding mutants of HIV-1, interestingly, seem to make the virus insensitive to the depletion of other cofactors, as TNPO3, NUP358 and NUP153, which are also involved in the process of nuclear entry and integration (Schaller et al, 2011; Lee et al, 2010; Matreyek & Engelman, 2011; Price et al, 2012).

Cyclophilin A

Cyclophilin A (Cyp A) is a product of the PPIA gene and represents a peptidyl-prolyl *cis-trans* isomerase, usually involved in the process of *cis-trans* isomerization of certain peptide bonds. The active site of Cyp A represents a hydrophobic pocket that binds to proline-rich peptides. It was shown that Cyp A affects HIV-1 infection via its direct interaction with CA (Sokolskaja et al, 2004; Luban 2007). The effect of Cyp A on HIV-1 infection seems to be cell type specific (De laco & Luban, 2014), considering that it was shown that Cyp A depletion in Jurkat cells has a decreasing effect on viral infectivity, but not in other analyzed cells (Braaten & Luban, 2001;

Saito et al, 2016). Interestingly, Cyp A can affect any post-entry step of the HIV-1 infection. It was shown to affect the viral uncoating process (Li et al, 2009; De laco & Luban, 2014), the efficiency of the RT process (De laco & Luban, 2014), the interaction and utilization of other factors involved in the nuclear entry of the PIC, and to affect integration site selection of the virus (Schaller et al, 2011; Yamashita & Engelman, 2017).

Other host cell factors participating in HIV-1 integration were identified in studies in which the technology of short-interfering RNA (siRNA) was applied. In such screens proteins involved in the nuclear import as NUP153, RANBPs, TNPO3 and RGPD8 were identified to be associated with the trafficking of HIV-1 (Varadarajan et al, 2005; Brass et al, 2008; Konig et al, 2008; Luban 2008; Christ et al, 2008).

RANBP2 (RAN binding protein 2) or NUP358

RANBP2 (or NUP358) is a protein located at the outer part of the NPC and has the capacity to bind RAN, which is a small GTP-binding protein involved in nuclear trafficking. RANBP2 is enriched in FG repeats, contains a cyclophilin-related nucleoporin and a domain that binds UBC9 important for SUMO1 transfer. It has been demonstrated that HIV-1 CA can directly bind to this protein, which affects the nuclear import of the viral PIC (Schaller et al, 2011).

NUP153

NUP153 is a nucleoporin of the inner nuclear basket showed to be involved in the PIC import via its interaction with CA (Di Nunzio et al, 2013; Matreyek, 2013), as well as in the export of HIV-1 Rev protein (Zolotukhin, 1999). It was also demonstrated that NUP153 can affect HIV-1 integration site distribution (Koh et al, 2013) and to bind directly the integrated viral genome (Marini et al, 2015).

TNPO3

TNPO3 was shown to be involved in the PIC import into the nucleus (Brass et al, 2008; Konig et al, 2008; Christ et al, 2008).

RGPD8 (RANBP2-like and GRIP domain containing 8)

RGPD8 is shown to accumulate at the NPC and is believed to assist in RNA and protein transport (Bushman et al, 2009)

Other host cell factors

Other proteins involved in HIV-1 integration that were mainly identified in siRNA screens were described as members of the cell microtubule system. Among them are MAP4 (a microtubule-associated protein), MID1IP1 (a regulator of microtubule organization) and CAV2 (involved in plasma membrane invaginations) (Bushman et al, 2008). Accordingly, other studies suggest that HIV particles may traffic along microtubules to reach the nucleus (McDonald et al, 2002). Also, in a yeast two-hybrid system, other members of the microtubule machinery, MAP1A and MAP1S have been identified to bind HIV-1 CA cores *in vitro*, colocalize with the virus, and when depleted, cause impaired infection as a result of impaired CA trafficking (Fernandez et al, 2015).

It has also been shown that Dynein is involved in the movement of the virus along microtubules and that its depletion causes an accumulation of viral particles in the periphery of the cell (McDonald et al, 2002; Arhel et al, 2006). In addition, Dynein was demonstrated to take function in viral uncoating too (Lukic et al, 2014; Pawlica et al, 2014).

Finally, Kinesins, as another group of proteins associated to the microtubule machinery, have been demonstrated to be involved in the process of HIV-1 infection. They have been shown to take action in viral trafficking, as well as the uncoating process (Lukic et al, 2014).

I.V. Cells targeted for HIV-1 infection

It has been reported that HIV-1 has the capability to infect a wide range of different cell types *in vitro*. For instance, it was demonstrated that cells from organs as the liver, lungs, salivary glands, eyes, prostate, testicles and adrenal glands could be infected. However, the only cell types that are *in vivo* regularly detected to be infected by HIV-1 are cells of the lymphoid and myeloid cell lineage, and among them, predominantly $CD4^+$ T lymphocytes and macrophages (Telesnitsky & Goff, 1997).

HIV-1 infection of $CD4^+$ T cells and macrophages

A critical point in the HIV-1 life cycle is the viral entry. It was shown that viral entry is not only dependent on the expression of the primary CD4 receptor on the cell surface, but also on expression of co-receptors as CCR5 or CXCR4, which are differentially expressed on $CD4^+$ T cells (Grivel et al, 2000; Doms, 2001). CXCR5 is highly expressed on the vast majority of peripheral T cells, whereas CCR5 is not expressed by all $CD4^+$ T cells. Its expression is upregulated as part of the late differentiation of effector and effector memory T cells (Veazey et al, 2000; Poles et al, 2001; Okoye & Picker, 2013). Still, $CD4^+$ T helper cells represent major target cells for HIV-1 in the blood, since they can express high levels of the CD4 receptor and are highly permissive for HIV-1 production (Maddon et al, 1986; Lifson et al, 1986).

Other immune cells also express CD4 and HIV-1 co-receptors and serve as viral targets too. Among these cells macrophages were described to carry markers of productive HIV-1 infection *in vivo* (Koenig et al, 1986), even though they express low levels of the CD4 receptor. Macrophages are infected by CCR5 tropic HIV-1 (R5 viruses – using CCR5 as co-receptor during infection). Together with dendritic cells and $CD4^+$ T memory cells macrophages are the first immune cells facing the virus in the mucosa (Shen et al, 2011). A significant portion of macrophages is directly productively infected in the mucosa with HIV-1 (Shen et al, 2009). In addition, macrophages secrete cytokines that attract and recruit T lymphocytes to sites of infection. They can even facilitate viral infection by enlarging the number of viral target cells at sites of infection (Swingler et al, 1999, Herbein et al, 2010).

Further on, macrophages are more resistant to viral cytopathic effects than activated $CD4^+$ T cells (Perelson et al, 1996; Reynoso et al, 2012), and in particular long-lived macrophages may harbor the virus for longer periods, which represents a major obstacle for virus eradication (Koppensteiner et al, 2012).

In summary, apart from $CD4^+$ T cells which represent the main targets for HIV-1 infection, macrophages represent long-term reservoirs of the virus and can help to establish viral infection at viral entry sites.

The importance of activation and chromatin status in primary $CD4^+$ T cells in HIV-1 infection

In the context of HIV-1 infection in primary $CD4^+$ T cells it is notable that HIV-1 preferentially infects and replicates in activated $CD4^+$ T cells (Margolick et al, 1987).

Although integrated HIV-1 DNA can be detected in resting $CD4^+$ T cells from infected individuals (Chun et al, 1995: Chun et al, 1997), it is still unclear how these cells are infected. According to one hypothesis, HIV-1 can directly infect resting $CD4^+$ T cells (Brady et al, 2009; Pace et al, 2012). The prevailing view is that resting $CD4^+$ T cells with integrated HIV-1 DNA are derived from activated infected $CD4^+$ T that reverted back to a resting memory state (Han et al, 2014; Sengupta & Siliciano, 2018).

Activated $CD4^+$ T cells are highly permissive for HIV-1 infection, but resting cells seem to resist infection. Blocks to infection in resting cells occur at different post-entry steps of the viral life cycle (Doitsch et al, 2010).

Widely accepted is that a deoxynucleoside triphosphate triphosphohydrolase (SAMHD1 – SAM domain and HD domain-containing protein), which is expressed in resting $CD4^+$ T cells, dendritic cells, monocytes and macrophages, but not in T cell lines, prevents the reverse transcription of HIV-1 RNA in resting $CD4^+$ T cells and macrophages by limiting the access of free dNTPs (Goldstone et al, 2011; Powell et al, 2011). By exposing this function, SAMHD1 represents a major HIV-1 restriction factor in resting $CD4^+$ T cells (Baldauf et al, 2012; Descours et al, 2012), and acts as a lineage-specific infection barrier for HIV-1 (Hrecka et al, 2011; Laguette et al, 2011).

The activation of $CD4^+$ T cells occurs through simultaneous engagement of the T-cell receptor (TCR) and co-receptors (mainly CD28) on the T cell surface by major histocompatibility

complex (MHCII) peptide and co-stimulatory molecules on the surface of an antigen presenting cell (APC).

One of the most important co-stimulatory molecules on the surface of an APC is the B7 protein (Figure 1.7.). The interplay of these two activation signals (coming from the MHC and B7) determines the outcome of a T cell's response to an antigen (Sharma et al, 2011).

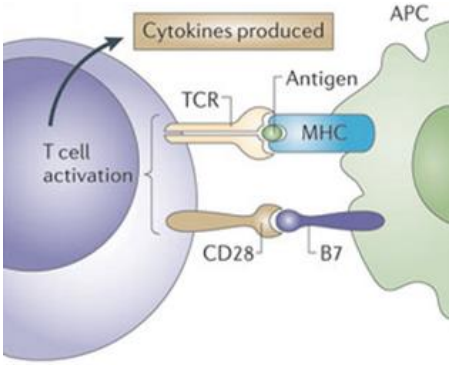


Figure 1.7. Schematic overview of the T cell activation process (from: Sharma et al, 2011).

Activation of $CD4^+$ T cells causes major changes in cell function, gene expression, metabolism as well as changes in chromatin organization. It has been shown that $CD4^+$ T cells upon activation undergo a global increase in chromatin accessibility, showing higher amounts of open chromatin features such as increased acetylation of lysine 27 of histone 3 (H3K27Ac), as compared to resting cells (Gate et al, 2018).

Target cell chromatin organization plays an important role in HIV-1 integration site selection as well as viral fate, especially regarding replication and virion assembly (Lusic & Siliciano, 2017). Chromatin composition as well as its importance and effects on HIV-1 life cycle will be discussed further on.

I.VI. HIV-1 integration site selection

As previously briefly mentioned in the chapter of HIV-1 integration, the integration of HIV-1 vDNA can hypothetically occur throughout the host cell genome, but it has been demonstrated that integration is rather a non-random process (Mitchell et al, 2004; Bushman et al, 2005).

Certain genomic factors affect the process of integration site selection among which the target DNA sequence plays a specified role, its chromatin composition, as well as their interplay with HIV-1 tethering factors.

Genomic regions targeted for HIV-1 integration

On a global sequence specific level, HIV-1 integration sites show features of DNase I hypersensitivity (Verdin, 1991), they are enriched in genic and GpC island rich regions and show high G/C skew (Brady et al, 2009).

Even though there is considerable variation among various retroviruses, each of them displays certain weak consensus sequence preferences. These consensus sequences extend at least over 20 nucleotides, which can be determined by aligning large numbers of integration sites (Figure 1.8.). Although the consensus sequence is palindromic, reflecting the average of all integration sites in both directions, individual integration sites are not necessarily symmetric. Since retroviruses have a high mutation rate, it is plausible that mutations in IN, which alter its interaction with the host DNA, can also alter the preferred sequence where the vDNA is integrated (Demeulemeester et al, 2014; Serrao et al, 2014; Hughes & Coffin, 2016). However, the consensus sequences for viral integration are always palindromic.

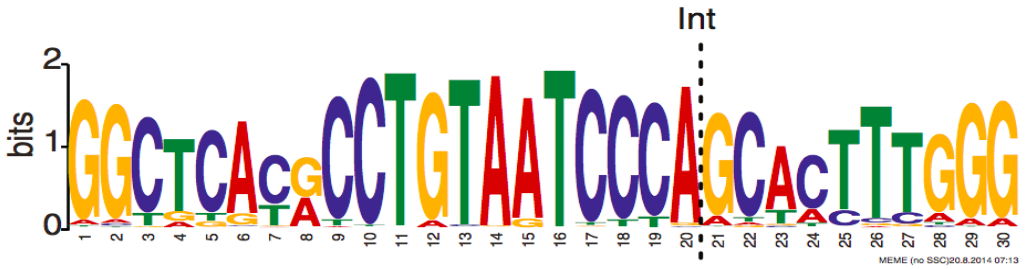


Figure 1.8. Consensus motif of HIV-1 integration site.

A 30 bp sequence consensus motif for HIV-1 integration. The preferred site of integration is marked with the dotted line. The sequence was generated by analyzing and comparing 100 bp around viral integration sites by MEME algorithm (Bailey & Elkan, 1994) (adapted from: Cohn et al, 2015).

About 80% of HIV integration events are in genes. For both HIV and MLV, integration site preferences reflect, at least in part, the chromatin distribution of host factors to which the respective INs bind (LEDGF/p75 and BET proteins). In the absence of LEDGF/p75, the efficiency of HIV integration is reduced (Llano et al, 2006; Engelman & Cherepanov, 2008), as well as its preference for integrating into actively expressed genes, although integration is not completely random (Busschots et al, 2005; Ciuffi et al, 2005). It is also likely, at least in the case

of HIV-1 and MLV, that the PICs interact with other host factors which affects integration site selection (as previously described).

DNA sequence specificity

In several independent studies which included the analysis of different cell lines, it was shown that HIV-1 preferentially integrates into gene rich regions, close to *Alu* elements, distant from gene transcription start sites (TSS) and rather into intronic than exonic regions of genes (Stevens & Griffith, 1994; Schroder et al, 2002; Lewinski et al, 2006; Barr et al, 2006; Singh et al, 2015). This preference was confirmed *in vivo* in resting $CD4^+$ T cells from infected patience (Demeulemeester et al, 2015) as well as by viral outgrowth assays performed from patients under retroviral therapy (Cohn et al, 2015). In addition, it was shown that HIV-1 in different cell types recurrently integrates in the same set of highly expressed genes which are not randomly distributed in the genome, but are rather positioned in clusters on targeted chromosomes. These genes were assigned as recurrent integration genes (RIGs) (Marini et al, 2015; Lucic et al, 2019). These genes are included in different lists of HIV-1 integration sites derived from either *in vitro* experiments or from patient material. One of these genes is BACH2 (Basic Leucine Zipper Transcription Factor 2). In a study of HIV-1 integration sites derived from patient data it was shown that this gene contained even 15 independent HIV-1 integration events in intron 4 (Maldareli et al, 2014). Some other interesting RIGs are STAT5B (Signal Transducer And Activator Of Transcription 5B) and MKL2 (MKL1/Myocardin Like 2) which are, as well, frequently appearing on different HIV-1 integration site lists (Sherrill-Mix et al, 2013; Maldarelli et al, 2014; Wagner et al, 2014; Cohn et al, 2015). An interesting fact is that all of the three mentioned genes encode proteins which are cellular growth factors or mediate the process of DNA transcription, and are directly or indirectly included in the cell cycle. Some studies suggest that HIV-1 integration into certain genes may lead to clonal expansion and contribute to the persistence of HIV-1 infected cells (Maldarelli et al, 2014; Cohn et al, 2015; Laskey et al, 2016). In line with this observation is the finding that HIV-1 integrations are frequently occurring in cancer-associated genes and can be found in the same patient at the same genomic spot in multiple cells that harbor the provirus. In the group of "cancer-associated genes" CREBBP, STAT5B, BACH2, C2CD3 and MKL2 are frequently found (Wagner et al, 2014). Again, these

genes are all involved in processes as cell growth, proliferation and mitosis (Rezaei & Cameron, 2015).

Considering that HIV-1 integrations are frequently found in those genes, it may be possible that certain genomic regions are preferred for viral integration over others. It has also been demonstrated that in $CD4^+$T cells HIV-1 integration genes are distributed in clusters over the human genome and may also move together in clusters during cellular activation (Marini et al, 2015; Lucic et al, 2019).

The analysis of HIV-1 patient samples revealed that BACH2 and MKL2 had a very peculiar pattern of viral integration. Upon prolonged cART exposure, integrated viral DNA could only be detected upstream of the transcription start site in intron 5 of BACH2, and intron 4 or 6 of MKL2. In addition, the analyzed proviral integration sites had the same transcriptional orientation as the host target gene, and again confirmed HIV-1 integration close to *Alu* genomic elements (Maldarelli et al, 2014; Cohn et al, 2015).

In a strictly host sequence specific context, apart from detecting HIV-1 integrations in active transcription units (Mitchell et al, 2004; Wang et al, 2007), it was demonstrated that viral integrations seem to be favored in GC genomic regions which overlap with regions of high gene density (Craigie & Bushman, 2012; Goncalves et al, 2016). HIV-1 integrations are also frequently found at sites showing DNaseI hypersensitivity and distribution of CpG islands over longer genomic stretches (Brady et al, 2009). It was also shown that HIV-1 integration frequently occurs at DNA sections that wrap around nucleosomes (Pruss et al, 1994; Wang et al, 2007). DNA that wraps around nucleosomes was shown to be enriched in AT repeating motifs (Segal et al, 2006). This finding is in contrast to the observed preference of the virus to integrate into GC rich genomic sites (Brady et al, 2009). However, the viral preference for integration into nucleosomal DNA is still in correlation with integration site data from $CD4^+$ T cells that show that when longer stretches of DNA (up to 10 Mb) are analyzed for HIV-1 integration sites, the overall integration is favored at GC rich genomic sites. Although the observed trend is less obvious when shorter genomic intervals (shorter than 2 kb) are analyzed (Brady et al, 2009).

Apart from the host cell DNA sequence features, chromatin marks as well as genome organization significantly impact the distribution of viral integration sites (Wang et al, 2007; Brady et al, 2009; Pradeepa et al, 2012; Sherrill-Mix et al, 2013; Marini et al, 2015; Lucic et al, 2019).

General chromatin features of HIV-1 integration sites

Chromatin is an important determinant of HIV-1 integration site selection. HIV-1 preferentially targets regions of open chromatin, as it has been shown that the viral PIC selectively targets open chromatin regions (Albanese et al, 2008). Upon integration into regions of open active chromatin (euchromatin) the provirus can be actively transcribed, whereas integration into condensed inactive chromatin regions (heterochromatin) would result in transcriptionally inactive proviruses (Maldarelli et al, 2014; Cohn et al, 2015; Laskey et al, 2016; Lusic & Siliciano, 2017).
Several studies showed a link between certain chromatin features and the frequency of HIV-1 integration sites (Ikeda et al, 2007; Wang, 2007; Marini et al, 2015). For example, it was shown that H3K9 acetylation (a mark of transcriptionally active genes) and trimethylated H3Lys36 (H3K36me3 – a mark associated with the elongating RNA Pol2 regions and actively transcribed genes) are related to frequently targeted HIV-1 integration sites (Marini et al, 2015), which is in line with the finding that LEDGF/p75, as the main IN binding partner, can specifically read the H3K36me3 chromatin mark via its PWWP domain (Pradeepa et al, 2012; Eidahl et al, 2013). In addition, HIV-1 integrations are frequently found in regions of open-accessible chromatin harboring marks as H3/H4 acetylation, but away from repressive chromatin marks as H3K27me3 or DNA CpG methylation (Wang et al, 2007; Goncalves et al, 2016).

I.VII. Chromatin organization and nuclear architecture

The nucleus is a highly dynamic and structured environment. The functional organization of the genome can be described in a hierarchical fashion, where the interplay of spatial and temporal factors act together to enable major nuclear functions (Misteli, 2007). The genomic DNA is compacted and organized into functional domains by the activity of several protein complexes and distinct DNA-binding inter-players inside of the nuclear space. The nucleus is separated from the cytoplasm by the nuclear envelope, which does not function only as a physical barrier, but significantly contributes to genome organization and regulation (Buchwalter et al, 2019). The nuclear envelope is perforated with nuclear pore complexes, which enable the communication between the inner nuclear space and the cytosol, while at the same time impacting genome organization and function (Akhtar & Gasser, 2007; Capelson & Hetzer, 2009; Capelson et al,

2010). The next section describes chromatin structure starting from the histone modifications, and then moving towards higher order genomic structures which shape the nuclear architecture, ending with the relation of HIV-1 to the spatial organization of the host cell genome.

Chromatin organization

The DNA fiber is compacted by histone proteins, and this DNA-histone complex is defined as chromatin. Chromatin is further compacted into chromosomes and the mitotic chromosome represents the final level of DNA compaction.

The basic repeating structural and functional unit of chromatin is the nucleosome, which contains eight histone proteins and about 146 pairs of (bp) DNA (Van Holde, 1988; Wolfe, 1999). Histones are a family of small, positively charged proteins termed H1, H2A, H2B, H3 and H4 (Van Holde, 1988). Pairs of histone proteins H2A, H2B, H3 and H4 form an octamer of histones to compose the nucleosome core.

Histone H1 is holding the nucleosome core and the wrapped DNA around it (146 bp), and additionally wraps another 20 bp long stretch of DNA around the histone octamer (Figure 1.9.).

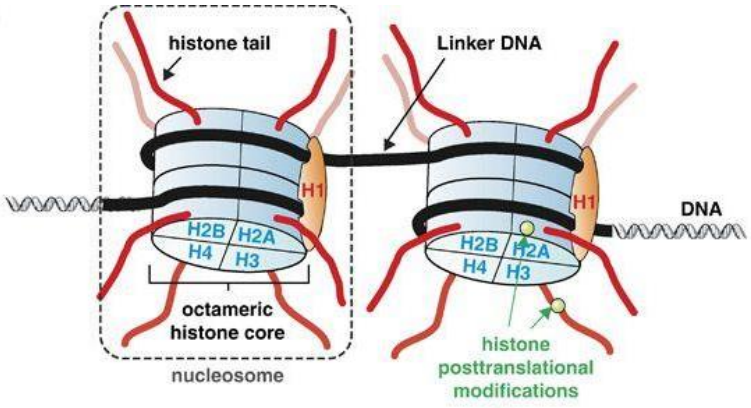


Figure 1.9. Schematic representation of a histone octamer (adapted from: Füllgrabe et al, 2011).

The compaction of DNA into nucleosomes has, generally, a repressive effect on gene expression. However, the positioning of nucleosomes at gene elements as promoters, has a significant impact on gene regulation in cells (Wolfe, 1999; Wyrick et al, 1999). Nucleosomes are not rigid structures. *In vitro* they can be shifted and positioned in a way that enables the access of RNA Polymerase and the transcription of nucleosomal DNA (Meersseman et al, 1992; Studitsky et al, 1997). *In vivo*, chromatin remodeling elements are directly interacting with nucleosomes and affect their movement along the DNA, even without DNA release from the nucleosome core (Elgin & Workman, 2000; Tsukiyama 2002). DNA

structure analysis in the nucleosome core revealed that the DNA structure in nucleosomes is significantly different when compared to "naked" DNA. Nucleosomal DNA was shown to have an unexpectedly high curvature, caused by the binding of histone proteins, which can affect the DNA sequence-dependent recognition of chromatin remodelers or transcription factors, as well as the movement of nucleosomes (Richmond & Davey, 2003).
HIV-1 IN shows preference in binding to nucleosomal DNA (Pryciak & Varmus 1992), and the integrated viral genome has been shown to have precisely positioned nucleosomes on the LTRs (Verdin, 1991; Verdin et al, 1993; Van Lint et al, 1996).

Apart from forming the chromatin fiber and compacting DNA, histones are important determinants in the process of gene expression, as they make the chromatin more or less accessible to transcription factors and other enzymes. Easier access for polymerases and other enzymes is either achieved by the displacement of histones upon the action of chromatin remodeling complexes (Smith & Peterson, 2005), or by enzymatic modifications of histone amino acid residues ("histone tails") (Figure 1.9) by the addition of phosphate, acetyl, methyl, crotonyl, ubiquitin or SUMO modifications, as well as modified by the activity of ADP-ribosyl transferases (Jason et al, 2002; Fischle et al, 2005; Gomez et al, 2006; Keppler & Archer, 2008).

In the process of histone acetylation, which is performed by the enzymatic activity of histone acetyl transferases (HATs), an acetyl group is added to the ∑-amino group of histone lysine side residues. The acetyl group is negatively charged, and by its addition to positively charged histone lysine residues, a neutralization of the histon's electric charge occurs, causing loosening of the tight association to the DNA (Bannister & Kouzarides, 2011). The opposite reaction, which is the removal of acetyl groups, is performed by histone deacetylases (HDACs) and causes compaction of the chromatin structure (Wang et al, 2009; Frank et al, 2016).
In the context of HIV-1, it has been described that during an acute infection or reactivation from latency HATs as CBP, GCN5 or P/CAF are recruited to the proviral LTR (Marzio et al, 1998; Benkirane et al, 1998; Lusic et al, 2003; Turner & Margolis, 2017) and thereby influencing proviral transcription. In addition, it has also been demonstrated that HATs can affect Tat activity by its direct acetylation (Ott et al, 1999; Col et al, 2001). Marks of acetylated histones frequently found close to transcriptionally active HIV-1 proviruses are H3K9ac, H3K14ac, H4K5ac,

H4K8ac, and H4K16ac (Benkirane et al, 1998; Lusic et al, 2003; Brady et al, 2009; Turner & Margolis, 2017).

On the other hand, the importance of HDACs in the HIV-1 life cycle has been noticed when HDAC inhibitors were shown to reactivate the virus from latency (Shirkawa et al, 2013; Archin et al, 2014). In addition, HDAC1 was one of the first proteins demonstrated to bind to the proviral LTR in complex with other transcriptional repressors (Romerio et al, 1997; Coull et al, 2000) and cause proviral silencing. Further work confirmed the binding of a variety of different proteins to the proviral LTR such as YY1, LSF, c-myc, Sp1 or NF-kB which cause the recruitment of HDACs and subsequent proviral silencing (Romerio et al, 1997; Coull et al, 2000; Williams et al, 2006; Jiang et al, 2007). Some HDAC inhibitors are currently even under clinical investigations as latency reversing agents (LRAs) (Turner & Margolis, 2017) .

Histone methylation predominantly takes place at histone lysine and arginine residues. The addition of the methyl group, by histone methyltransferases (HMTs) and their active SET domain, does not cause any change in the charge of histone proteins. The position of the added methyl group, as well as the number of added methyl groups (-mono, -di or -tri-methylation) (Murray, 1964; Paik & Kim, 1967; Hempel et al, 1968) can have different outcomes in the context of chromatin compaction and accessibility. Depending on the context and potential of histone modifications a variety of multi-protein complexes are recruited to the chromatin, allowing for tightly regulated transcriptional control (Greer & Shi, 2012). The removal of methyl groups is performed by histone demethylases (HDMs). Different groups of enzymes regulate the demethylation of lysine (Forneris et al, 2005; Tsukada et al, 2006) and arginine residues (Cuthbert et al, 2004; Wang et al, 2004).

H3K4me3 is a mark associated with active promoters and has been linked to both proviral activation and repression (Le Douce et al, 2012). H3K36me3 marks gene bodies of actively transcribed genes and is the mark most frequently associated with HIV-1 integration sites (Brady et al, 2009), probably due to the fact that host integration factor LEDGF/p75 specifically reads this chromatin mark (Pradeepa et al, 2012). Interestingly, recently it has been demonstrated that active genes which are frequently targeted for HIV-1 integration have a distinct distribution of H3K9me2 chromatin mark (Lucic et al, 2019), which is usually linked to facultative heterochromatin (Huisinga et al, 2006). However, the importance of H3K9me in the context of

HIV-1 has first been described in microglia cells by investigating the activity of transcriptional repressor CTIP2. The study found that CTIP2 negatively affects HIV-1 transcription by the simultaneous recruitment of Tat and HP1-α (Marban et al, 2005). HP1-α (heterochromatin protein 1-α) binds methylated lysine residues and causes transcriptional repression (James and Elgin, 1986). One of the main H3K9 HMTs is SUV39. However, in the context of HIV-1 latency it has been demonstrated that the direct knockdown of another HMT; G9a, or its chemical inhibition by BIX01294 has a more increasing effect on viral transcription and is even synergized by the addition of HDAC inhibitor SAHA (Imai et al, 2010). Recent work demonstrated that the combination of H3K9me and H3K27me machineries can specifically affect the regulation of HIV-1 latency and force the virus to engage even in a deeper state of latency (Matsuda et al, 2015).

H3K27me2/me3 are chromatin marks are mainly decorating repressive chromatin regions which are linked to reduced genetic activity (Wang et al, 2008). H3K27me3 and the presence of its HMT EZH2 have been shown to be linked to HIV-1 latency (Friedman et al, 2011; Matsuda et al, 2015). Different levels of H3K27me3 at the viral LTR have been demonstrated to affect the ability of TNF-α (a strong proinflammatory cytokine) to reactivate HIV-1 from latency (Friedman et al, 2011). In addition, EZH2 inhibitors in combination with HDAC inhibitor SAHA and the bromodomain inhibitor JQ1 have been demonstrated to have a more pronounced effect on reverting HIV-1 from latency (Tripathy et al, 2015), compared to the effects of single LRAs.

When focusing on histone phosphorylation, it has been shown that this chromatin modification mainly affects serine, threonine and tyrosine residues (Xhemalce et al, 2011). The phosphate group is added from an ATP molecule to a hydroxyl group of a targeted amino acid of the histone tail by the enzymatic activity of kinases. The phosphate group is negatively charged and its addition to a histone tail has a significant impact on chromatin structure, mainly causing its compaction (Bannister & Kouzarides, 2011). The best known function of histone phosphorylation occurs during cellular response to DNA damage, when phosphorylated histone H2A(X) marks large chromatin domains around the site of DNA breakage (Rosseto et al, 2012). It has been demonstrated that retroviral integration causes a transient formation of phosphorylated histone H2A(X) regions at viral integration sites (Daniel et al, 2004). In addition, in HIV-1 latently infected Jurkat cell clones, as well as infected primary T cells, a significant

increase in phosphorylated histone H2A(X) sites was detected upon treatment with DNA damage response (DDR) inducing agents or cancer chemotherapies, compared to non-infected cells. The authors of the study propose a new marker and strategy to detect and eliminate latently infected cells (Piekna-Przybylska et al, 2017).

Chromatin organization as well as the activity of chromatin modifying machineries significantly impact transcriptional control and, accordingly, many cellular processes. In line with that are results of several studies demonstrating that chromatin modifications and chromatin modifying enzymes impact HIV-1 transcription. Understanding the interplay between HIV-1 latency and its reversal, and the maintenance of normal host cell function could provide a more detailed insight and better understanding on general chromatin regulation, but also provide clues for a more targeted application of LRAs and a potential eradication of latent HIV-1 (Lusic & Siliciano, 2017; Turner & Margolis, 2017).

The 3D genome organization

The composition of genomic DNA including AT/GC skew, gene density, distribution of repetitive elements or virus derived elements represents a linear perspective of genome organization (Bickmore & van Steensel, 2013). The first level of genome structuring represents the compaction of DNA together in a complex with histone proteins forming the nucleosome, which represents the elementary repetitive unit of chromatin (Van Holde, 1988; Wolfe, 1999). The further organization of chromatin, including the formation of chromatin loops, contacts, chromatin domains, chromosomes and chromosomal territories represent higher orders of genome organization (van Steensel et al, 2001; Cremer et al, 2006; Bickmore & van Steensel, 2013).

Chromosomes and chromosome territories

Ever since the pioneering work on cell division, chromosomes have been proposed to reside in distinct chromosome territories (Rabl, 1885; Boveri, 1909). Nowadays it is established that chromosome territories represent the main feature of chromosome architecture (Dixon et al,

2016). More recent studies, applying genome capturing techniques (C-techniques), provide evidence that a single chromosome represents a specialized individual territory in which genomic segments that are closer to each other interact more with each other (cis-interactions), compared to genomic segments which are on the other arm of the same chromosome, or even on another chromosome (trans-interactions) (Lieberman-Aiden et al, 2009; Kalhor et al, 2012; Sexton et al, 2012; Zhang et al, 2012; Bickmore & van Steensel, 2013). However, it has also been demonstrated that certain trans-interactions can be very robust. Such trans interacting sequences are usually located in regions of high gene density and transcriptional activity, and also show DNase I hypersensitivity (Simonis et al, 2006; Lieberman-Aiden et al, 2009; Hou et al, 2012; Kalhor et al, 2012; Sexton et al, 2012). It has also been shown that colocalization of genomic loci in trans can have effects on gene expression (Noordermeer et al, 2011), emphasizing the relevance of genome organization, genomic contacts and spatial gene positioning in a functional context.

Chromatin domains

Series of recent studies provide evidence that single chromosomes are composed of distinct chromatin domains or topologically associated domains (TADs). TADs are comprised of long stretches of chromatin spreading over up to several million bases (Jackson and Pombo, 1998; Ma et al, 1998; Dixon et al, 2012; Nora et al, 2012; Sexton et al, 2012). TADs are defined as genomic compartments in which two genomic regions are observed to associate more frequently, compared to regions outside the TAD (Dixon et al, 2012; Nora et al, 2012; Sexton et al, 2012; Dixon et al, 2016). TADs have been demonstrated to remain stable during cell division and show high similarity among evolutionary related species. (Cremer & Cremer, 2010; Dekker & Heard, 2015; Sexton & Cavalli, 2015). TADs frequently contain "sub-TADs" (Phillips-Cremins et al, 2013; Rao et al, 2014), mainly composed of locally separated loops or insulated neighborhoods (Rao et al, 2014; Dowen et al, 2014; Ji et al, 2016). Sub-TADs usually represent features that can have a distinct organization in different cell types, depending on cell type specific functions (Phillips-Cremins et al, 2013; Dowen et al, 2014; Ji et al, 2016; Dixon et al, 2016). The organization of TADs significantly influences gene expression, DNA replication and recombination (Lucas et al, 2014; Pope et al, 2014; Hu et al, 2015). The organization of

chromatin domains and their effect on genome activity depends on the chromatin site which creates a domain (Boettiger et al, 2016). For instance, domains harboring polycomb proteins have a remarkably different organization compared to other chromatin domains. A hallmark of polycomb domains is the decoration of chromatin with H3K27me3 mark and the dense compaction of chromatin forming a mainly repressive chromatin environment (Morey & Helin, 2010; Boettiger et al, 2016; Dixon et al, 2016).

Also, chromatin marked with H3K9me2/me3 induces a special compartmentalization of the chromatin. H3K9me2 and H3K9me3 are frequently considered to be repressive marks, but many examples show their involvement in marking actively transcribed genes (Kwon & Workman, 2001; Bickmore & van Steensel, 2013). Those, usually long stretches of DNA covered with H3K9me2 and H3K9me3 which cover up to 45% of the genome are also referred to as large organized chromatin domains - LOCKs (Wen et al, 2009; Hathaway et al, 2012).

Beside the importance of certain chromatin stages and presence of histone modifications and protein complexes in TADs, the interaction between different types of TADs is an important element of genome organization. One of the most important proteins involved in the shaping of TAD interactions is the insulator protein CTCF, which is a zinc finger DNA binding protein (Phillips-Cremins & Corces, 2013; Ghirlando & Felsenfeld, 2016). In line with this is the finding that TAD borders of mammalian cells are significantly enriched in CTCF binding sites (Dixon et al, 2012). The binding of insulator proteins as CTCF is demonstrated to be a highly regulated process which finally affects the transcriptional activity of certain chromatin domains. For instance, the binding of isolator proteins to DNA can affect and regulate interactions between enhancers and gene promoters, as well as affect the genomic distribution of euchromatin and heterochromatin and thereby influence gene expression/repression (Phillips-Cremins & Corces, 2013; Ghirlando & Felsenfeld, 2016; Dixon et al, 2016).

The nuclear periphery and the nuclear pore complex

The nuclear periphery is composed of the nuclear envelope (NE), the nuclear lamina (NL) and nuclear pore complex (NPC). Apart from these structural elements the underlying chromatin and its composition is important in determining the functional aspects of the genome.

Parts of the chromatin fiber which are occupying the tight space under the NL in the nuclear periphery are often organizing nuclear territories assigned as Lamina-associated domains (LADs). LADs are regions of condensed chromatin enriched in repressive chromatin marks as histone 3 lysine 9 dimethylation (H3K9me2), which is a mark of heterochromatin and silenced genes (Pombo & Dillon, 2015). Regions of open chromatin and actively transcribed genes are rather excluded from LADs.

The NPC has been shown to associate with both silent and active chromatin domains, and can therefore represent an environment having an activating or repressing effect on gene expression (Raices & D'Angelo, 2017). First studies on the chromatin surrounding the NPCs showed an association of silent chromatin with the NPC (Brown et al, 2008; Jacinto et al, 2015). However, more recent studies showed that genes associated to the NPC can be either activated or silenced, and not exclusively repressed as previously thought (Ibarra et al, 2016; Pascual-Garcia et al, 2017; Toda et al, 2017; Raices et al, 2017). Considering that NPCs can have activating or repressing effects on genes closely associated to them, as well as the fact that have the capacity of binding active and silent chromatin it becomes evident that those structures also have the potential to create and shape the nuclear landscape around them. The chromatin surrounding NPCs is mostly having and open and decondensed structure (Schermelleh et al, 2008; Lemaitre & Bickmore, 2015). It has also been demonstrated that super enhancers can cluster in their proximity (Ibarra et al, 2016) and that certain groups of genes need the association with nuclear pore proteins for their expression (Raices et al, 2017). In addition, the "gene-gating" hypothesis proposes an explanation in which compact chromatin associates with the nuclear lamina, while transcribed genes associate with the NPC, which would facilitate the nuclear export of RNA. Also, the non-random distribution of NPCs in the nuclear envelope reflects the non-random organization of chromatin in the nuclear periphery (Blobel, 1985; Ciabrelli & Cavalli, 2015; Lemaitre & Bickmore, 2015). Mentioned findings indicate that NPCs take an active role in shaping the chromatin in the nuclear periphery, and influence gene expression (D'Angelo, 2017).

Lamina Associated Domains (LADs)

The nuclear lamina (NL) represents a scaffold element of the nucleus, located in its periphery and is also involved in chromatin organization (Goldberg et al, 1999; Gruenbaum & Medalia, 2015;

Ulianov et al, 2019), DNA repair and nuclear assembly (Burke & Stewart, 2013; de Leeuw et al, 2017). The NL is composed of lamin proteins which are organized into dense, net-like structures, and also interact with other lamin-binding proteins.

The NL of mammalian cells is composed of 4 lamin isoforms; two type A lamins (lamin A and C), and two type B lamins (lamin B1 and B2). Type A lamins are encoded by the LMNA gene, whereas type B lamins are encoded by the LMNB2 gene (Krohne et al, 2005; de Leeuw et al, 2017).

Mammalian cell LADs are usually large and gene-poor domains spanning from 0.1 to 10 Mb, covering about 40 % of the genome in a cell population (Guelen et al, 2008; Peric-Hupkes et al, 2010; Ciabrelli & Cavalli, 2015).

LADs that are present in all so far analyzed cell types are defined as constitutive LADs, whereas LADs that differ between cell types or change their association to the NL depending on the cell cycle are known as facultative LADs (Dixon et al, 2016; Bonev & Cavalli, 2016; Lusic & Silliciano, 2017). Apart from having a role in shaping the nuclear architecture, LADs are important in gene expression determination. Genes located inside LADs were shown to have low transcriptional activity, lacking any active histone mark. In addition, single cell studies in mammalian cells showed that less than one third of LADs are located less than 1 μm from the NL at any time in the cell cycle, indicating a dynamic character of lamin tethering (Kind et al, 2013). In addition, certain gene promoters consistently found in LADs have certain mechanisms of avoiding the repressive effect of the LAD environment, whereas other types of LAD-associated promoters can be activated when moved away from the LAD (Leemans et al, 2019).

The repressive nature of LADs is evolutionary conserved and is supported by genome-wide studies, showing that the average gene expression levels in LADs are lower than outside of them (Guelen et al, 2008; van Bemmel et al, 2010; Kind et al, 2013; Ciabrelli & Cavalli, 2015).

Mammalian LADs also show very sharp borders. These borders are genetically defined since they contain bidirectional transcription units where the active domains are mainly pointing outward from LADs and bearing CpG islands and CTCF binding sites (Guelen et al, 2008), where CTCF binding is important in shaping LAD borders and chromatin looping between its binding sites (Splinter et al, 2006; Handoko et al, 2011).

Also, more recent studies underlie the importance of lamins in shaping the interactions between TADs and the effect on genome regulation based on these interactions (Zheng et al 2018; Ulianov et al, 2019). The loss of lamins was shown to disrupt LAD organization, change the interaction between active and inactive chromatin domains and cause global changes in transcription (Zheng et al, 2018).

Considering specific chromatin marks, it has been demonstrated that mammalian LADs harbor specific features as enrichment in G9a-dependent H3K9me2 chromatin mark which is important fort LAD tethering to the NL (Guelen et al, 2008; Kind et al, 2013).

Chromatin composition in the nuclear periphery

The global nuclear architecture is shaped by the enzymatic activity of many different proteins, but a special subset of them is involved in the organization of the mainly transcriptionally repressive environment in the nuclear periphery. The interactions of the NL and chromatin is maintained due to the activity of several chromatin modifying enzymes and DNA binding proteins (LAP2B, EMD, BAF, LBR, HP1). Histone methyltransferases (HMTs) G9a/GLP and histone deacetylases (HDACs) are shown to interact with NL proteins (Somech et al, 2005; Holaska & Wilson, 2007; Montes de Oca et al, 2009) and may locally dimethlyate H3K9 and deacetylate histones of chromatin positioned in close proximity to the NL, as well as contributing to the repression of genetic activity close to the NL Figure 1.10.). Large domains of repressive chromatin marks as H3K27me3 (established by the activity of Polycomb Repressive Complex 2 – PRC2), H3K9me2 and H3K9me3 are present in this area of the nucleus, rather then clearly active marks as H3K4me3 or lysine acetylations (Guelen et al, 2008; Ciabrelli & Cavalli, 2015;

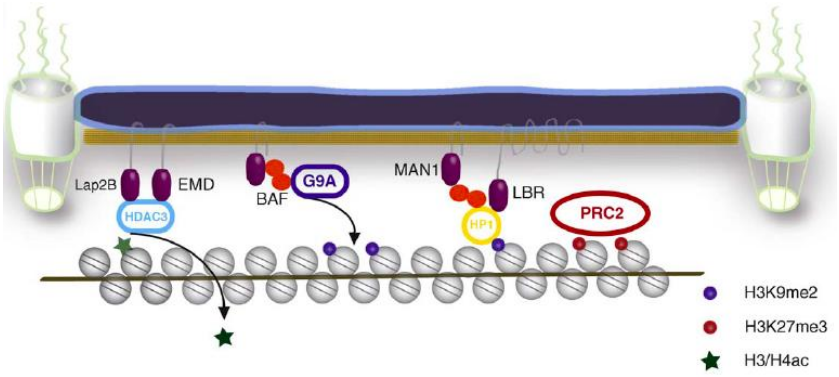


Figure 1.10. Representation of the repressive chromatin environment of LADs.
Chromatin associated with the nuclear lamina is mostly decorated with repressive marks (H3K9me2 and H3K27me3) and HDACs are very often present and ensure the establishment of a repressive chromatin environment. DNA binding proteins ensure the close association of DNA and the nuclear lamina (Lap2B, EMD, BAF) and the maintenance of a closed chromatin state (HP1). The presence of protein complexes as PRC2 helps keeping a closed chromatin structure and an environment of low gene expression. Cell receptors as LBR anchor heterochromatin at the nuclear lamina, whereas receptor MAN1 can release the chromatin from the nuclear lamina in case of growth factor signaling and induce, by a cascade of reactions, gene expression (the bigger red dots in the picture represent the interactions) (adapted from: Kind & van Steensel, 2010).

Briefly, in the context of higher order chromatin structures (chromosomes) and their relation to the nuclear periphery, it was shown that distinct chromosomes generally occupy a preferred position in the nucleus (Dekker et al, 2013). Chromosome positioning seems to correlate with gene density, where gene-poor chromosomes tend to localize to the nuclear periphery (e.g. human chromosome 18), whereas gene-dense chromosomes rather occupy inner nuclear positions (Rouquette et al, 2010; Ciabrelli & Cavalli, 2015).

H3K9me2 chromatin mark and G9a histone methyltransferase

H3K9me2 was shown to cover large genomic portions described as Large Organized Chromatin Domains (LOCKS) (Wen et al, 2009), and these domains where shown to mainly overlap with LADs (Kind & van Steensel, 2010). H3K9me1/me2 marks are important in anchoring the chromatin fiber to the nuclear periphery, whereas H3K9me3 is responsible for transcriptional repression (Towbin et al, 2012).

Chromatin decorated with H3K9me1/me2 marks, except for its role in shaping the chromatin landscape of the nuclear periphery, is also found in large euchromatic regions, suggesting a contribution of these chromatin marks in dynamic mechanisms of gene repression or activation (Tachibana et al, 2002; Black et al, 2012).

G9a histone methyltransferase (HMT) belongs to the Su(var)3-9 family mainly catalyzing H3K9me1 and H3K9me2. Structurally G9a is composed of a catalytic SET domain, a domain containing ankyrin repeats and a nuclear localization signal on the N-terminus (Esteve et al, 2006; Nakanishi et al, 2008; Collins & Cheng, 2010) (Figure 1.11.). The SET domain is responsible for the addition of methyl groups on histone H3, whereas the ankyrin repeats have been shown to represent mono- and dimethyl lysine (K) binding regions. Accordingly, G9a is not only able to methylate histone tails, but also able to recognize this modification and function as a scaffold for the recruitment of other molecules on the chromatin (Shahbazian et al, 2005; Casciello et al, 2015). G9a functions mainly as a heterodimer together in a complex with GLP (G9a Like Protein). It has been demonstrated that if one of the two proteins is depleted, H3K9 mono- and dimethylation levels get reduced in euchromatic regions (Tachibana et al, 2002). *In vitro* studies showed that both G9a and GLP can also form homodimers and perform H3K9 mono-, di-, and trimethylation (Collins et al, 2005; Kubicek et al, 2007).

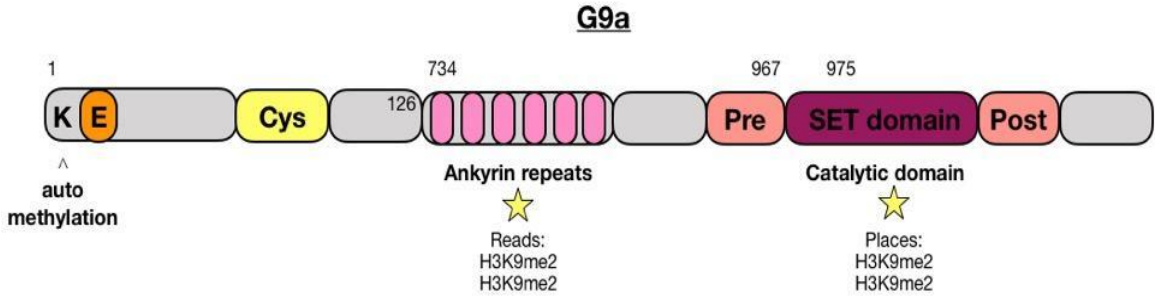


Figure 1.11. Scheme of G9a histone methyltransferase structure.

At the N-terminal end of the enzyme ankyrin repeats (orange) recognize (read) H3K9me1 or H3K9me2 chromatin mark. At the C-terminus the catalytic SET domain (red) deposits mono or di-methyl H3K9.

G9a mainly methylates H3K9 in euchromatic regions, which causes gene repression (Tachibana et al, 2002; Hyun et al, 2017). Considering that H3K9 methylation affects gene expression, the

activity of G9a is very important in the context of cell viability and development, respectively. In mammalian cells, the disruption or inhibition of G9a was shown to weaken the NL-LAD association (Bian et al, 2013; Chen et al, 2012; Kind et al 2013; Harr et al, 2015), which, as mentioned, can affect gene expression.

In studies with murine G9a-deficient embryonic stem cells (ESC) a global loss methylation in chromatin was observed but surprisingly not in heterochromatic regions. While HMT function can generally be associated with both heterochromatin and euchromatin organization (Tachibana, 2015), this example shows that G9a can specifically associate with euchromatin and be involved in the repression of active promoters (Peters et al, 2001; Rice et al, 2003; Tachibana et al, 2005). In addition, the repressive activity of G9a was proved to be essential in embryo-genesis in mice. G9a depletion resulted in embryonic lethality or severe differentiation defects in ESC, showing that G9a is crucial for the repression of developmental genes and necessary during development (Tachibana et al, 2002). Likewise, G9a was found to be involved in the development of cell specification as an important factor of *oct-3/4* gene inhibition, a homeobox gene important for the maintenance of pluripotency (Feldman et al, 2006).

In terms of T-cells (which are the main target cells for HIV-1 infection), it was shown that the normal functioning of G9a is important in hematopoietic cell lineage commitment, as well as T-cell development during inflammatory response (Lehnertzt et al, 2014; Antignano et al, 2014). Specifically, it was shown that G9a is a negative regulator of pathogenic T-cell differentiation as G9a depleted T-cells were shown to have an increased sensitivity to TGF-β1 (Transforming Growth Factor Beta 1) which is promoting naïve T-cell differentiation in pathogenic T-cells, Th17 (T-helper cells) and Treg (T-regulatory cells), in the absence of intestinal inflammation (Antignano et al, 2014). This finding showed that G9a is important in the maintenance of T-cell homeostasis. In addition, G9a was shown to be involved in T-cell differentiation, as $CD4^+$ T cells fail to differentiate into Th2 cells both *in vitro* and *in vivo* in the absence of G9a. Mice carrying a T-cell specific G9a deletion could not develop Th2 cells in response to infection in the absence of interferon γ (Lehnertz et al, 2014).

In summary, apart from mediating the establishment of H3K9me2 and shaping the chromatin conformation in the nuclear periphery of T cells, G9a is evidently very important for T cell development and normal cell functioning.

HIV-1 in the 3D nuclear space

HIV-1 tethering factors and proviral interactors

One of the most important factors that influences HIV-1 integration into certain chromatin regions is LEDGF/p75. Binding sites of LEDGF/p75 were mapped on chromosomes making reference to the highly annotated ENCODE (encyclopedia of DNA elements). LEDGF/p75 binding sites (also assigned as LEDGF islands) were described to be dominantly distributed at transcribed genomic regions which overlay with the observed trend for these regions to be favored for HIV-1 integration (De Rijck et al, 2010). In addition, LEDGF/p75 knock down was observed to cause increased integration in regions of higher G/C content and closer to CpG islands (Brady et al, 2009). The overall dataset defines a model in which LEDGF/p75 directs HIV-1 integration by binding to IN on one side, and on the other side to open chromatin regions (De Rijck et al, 2010; Craigie & Bushman, 2012). LEDGF/p75 has been described as an H3K36me3 chromatin mark reader (Venkatesh et al, 2012; Pradeepa et al, 2012; Demeulemester et al, 2015, Singh et al, 2015), and it was shown that LEDGF/p75 binds to H3K36me3 regions in the genome (Marini et al, 2015) supporting the notion that LEDGF/p75 binds to open chromatin.

Apart from LEDGF/p75 and its function in IN binding and directing the process of integration site selection, it was shown that the interaction of the PIC with proteins of the NPC is important in positioning of the virus inside of the nuclear space and determining its transcriptional activity (Marini et al, 2015; Lelek et al, 2015). It was shown that vDNA integration sites are associated to transcriptionally active histone marks and several nucleoporins; as NUP62, NUP96, NUP153 and the inner nuclear basket protein TPR. In addition, TPR depletion was shown to reduce LTR-driven gene expression in HIV-1 infected cells (Lelek et al, 2015). Although the knockdown of TPR might inhibit the levels of HIV-1 expression, it has a minor impact integration of the vDNA. Knockdown of NUP153 on the other hand almost completely abolishes HIV-1 integration, with HIV DNA signals, most probably corresponding to the non-integrated virus, can be observed randomly positioned in the nuclear space (Marini et al, 2015).

Chromatin composition underneath the NPC has been shown to be composed of “open-easy access-transcriptionally active” chromatin (Schermelleh et al, 2008), and depletion of TPR was shown to cause a reduced density of H3K36me3 in the nuclear periphery, creating a less

accessible chromatin environment (Lelek et al, 2015). It has also been demonstrated that the overexpression of TPR leads to a stabilization of LEDGF7p75 at the nuclear periphery (Lelek et al, 2015; Wong et al, 2015). These findings underlie the significant interplay between H3K36me3 chromatin mark, LEDGF/p75 as a reader of that mark (Pradeepa et al, 2012) and TPR in the context of creating an environment for HIV-1 integration (Lusic & Siliciano, 2017).

It has also been shown that the role of LEDGF/p75 in viral integration is more in the process of navigating the virus towards gene bodies, whereas the function of CA interacting factor CPSF6 (Lee et al, 2010; Price et al, 2012; Fricke et al, 2013) is the localization of the virus into euchromatic regions (Sowd et al, 2016).

On the other hand, it has also been demonstrated that CPSF6 via its interaction with CA localizes the virus towards euchromatic regions in the nuclear interior, away from the nuclear periphery and that LEDGF/p75 plays a more important role in directing the virus towards gene bodies, than determining its nuclear localization (Sowd et al, 2016; Achuthan et al, 2018). The consensus of mentioned studies is that the integrated virus preferentially resides in open chromatin regions, in the gene bodies of actively transcribed genes (Mitchell et al, 2004; Wang et al, 2007; Marini et al, 2015; Achuthan et al, 2018).

Spatial positioning of HIV-1 in the nucleus

In respect to the spatial positioning of the virus in the nucleus, structural cell specific features may potentially affect the discrete proviral positioning in the nucleus, considering that opposing data arose from studies which analyzed viral positioning in different cell types (Albanese et al, 2008; Di Nunzio et al, 2012; Marini et al, 2015; Lelek et al, 2015; Achuthan et al, 2018). However, upon knockout of HIV-1 tethering factors LEDGF/p75 and CPSF6, and HIV-1 infection in HeLa and $CD4^+$ T cells it was shown that genes targeted for integration by the virus were mostly cell type specific and had a perinuclear position in the different cell types (Achuthan et al, 2018). Accordingly, cell type specificity seems to be a significant determinant of HIV-1 integration site selection and proviral spatial positioning in the nucleus.

In addition, in the context of LEDGF/p75, knock down experiments performed in HeLa cells demonstrated that LEDGF/p75 does not play a crucial role in the macrolocalization of HIV-1 in

the nucleus, but rather contributes to the directing of the virus into gene-rich regions (Quercioli et al, 2016; Sowd et al, 2016).

However, in primary $CD4^+$ T cells, which are the main target of HIV-1, the proviral genome predominantly resides in the nuclear periphery, in open chromatin regions which are located in the vicinity of the NPC and excluded from LADs (Albanese et al, 2008; Di Nunzio et al, 2012; Marini et al, 2015).

Regarding chromatin features of the environment where HIV-1 positions, apart from knowing that the chromatin of these regions is decorated with open chromatin marks, which mark actively transcribed genes (Schröder et al, 2002; Wang et al, 2007; Marini et al, 2015), it was confirmed that the provirus occupies regions with low densities of heterochromatin specific markers (Brady et al, 2009; Friedman et al, 2011; Vranckx et al, 2016). Chromatin accessibility and chromatin features evidently play important roles in HIV-1 integration site selection and spatial positioning of the provirus.

On a genome-wide scale, chromatin composition and nucleosome positioning/shuffling throughout the genome have a significant impact on genome regulation, by modifying the *in vivo* availability of binding sites of TFs and the general transcription machinery. Mentioned phenomena affect DNA-dependent processes such as transcription, DNA repair, replication and recombination (Radman-Livaja & Rando, 2010; Tsompana & Buck, 2014), and can therefore also affect HIV-1 transcriptional activity (Easley et al, 2010; Matysiak et al, 2017).

In studies aiming to understand how nucleosomal positioning regulates gene expression it was shown that transcriptional activation coincides with nucleosome perturbation, whereas transcriptional regulation requires the repositioning of nucleosomes throughout the eukaryotic lineage (Wallrath et al, 1994; Boeger et al, 2003; Lee et al, 2004; Hogan et al, 2006; Buck & Lieb, 2006; Schones et al, 2008; Shivaswamy et al, 2008). In addition, open or accessible regions of the genome are very often regarded as primary positions for regulatory elements (John et al, 2011) and have been in the very beginning characterized by nuclease hypersensitivity assays *in vivo* (Gross & Garrard, 1988). Low-throughput experiments in *Drosophila* using DNase I and MNase treatment, provided the first demonstration that active chromatin coincides with nuclease hypersensitivity, which is chromatin accessibility (Wu et al, 1979, Wu, 1980; Keene & Elgin, 1981; Tsompana & Buck, 2014). Accordingly, several independent studies proved that HIV-1

integration sites coincide with regions of DNase I hypersensitivity (Verdin, 1991; Brady et al, 2009; Busschots et al, 2005; Ciuffi et al, 2005; Chen et al, 2017a).

In silico predictions have shown that HIV-1 preferentially integrates into nucleosomal, compared to naked DNA, as well as that HIV-1 integration sites are mainly found on outer portions of the major DNA groove which is also attributed to a preference in integrating into more flexible DNA regions (Wang et al, 2007; Michieletto et al, 2019).

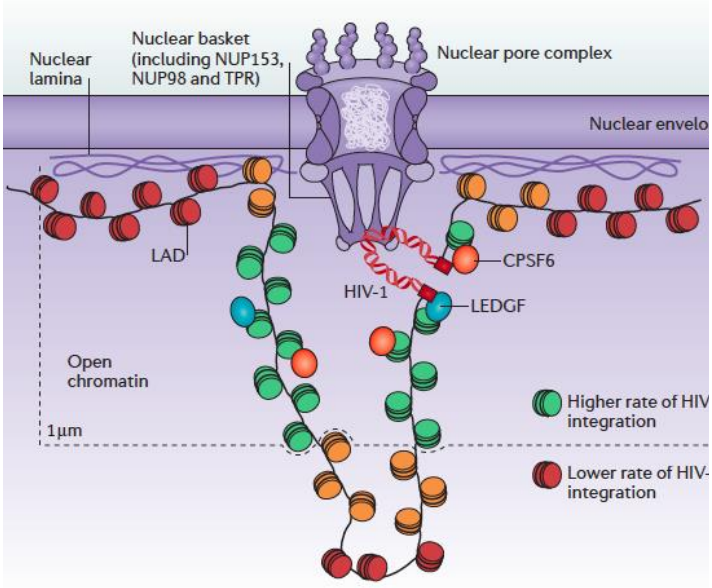


Figure 1.12. Scheme of HIV-1 integration.
HIV-1 integrating into open chromatin regions (outside of LADs) and interacting with elements of the NPC, as well as host cell factor LEDGF/p75 and CPSF6) (from: Lusic & Siliciano, 2017).

In summary, the interplay between host cell chromatin features (Marini et al, 2015; Lusic & Siliciano, 2017; Lucic et al, 2019), HIV-1 tethering factors, as well as host cell type specificity significantly impact the spatial positioning of HIV-1 provirus in the host cell nucleus (Sowd et al, 2016; Achutnan et al, 2018) (Figure1.12.).

I.IX. R-loops – DNA-RNA hybrids shaping the chromatin landscape

DNA-RNA duplexes that are mainly occurring co-transcriptionally, in DNA accessible and DNase I hypersensitive regions, are named R-loops (Sanz et al, 2016; Chen et al, 2017a). R-loops represent tri-stranded nucleic acid structures composed of a DNA-RNA hybrid and a displaced single stranded DNA (Figure 1.13.). These structures, first described in 1976 (Thomas et al, 1976) are predominantly found in G/C rich genomic regions (Ginno et al, 2012; 2013). where RNA polymerase II (Pol II) transcribes a C-rich template. The generated G-rich RNA transcript bounds back to C-rich DNA generating a hybrid of increased thermodynamic stability. Although it is still unclear how R-loops are really formed, the most probable seems the "thread back model", according to which the RNA/DNA hybrid is formed before the two strands of the DNA duplex can reanneal.

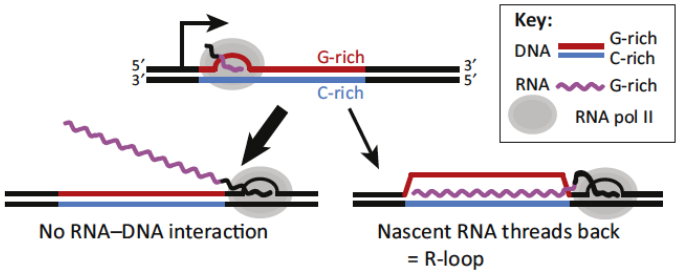


Figure 1.13. Scheme of an R-loop and the co-transcriptional mechanism of its formation.

Legend on the right indicates elements and factors included in the process of R-loop formation.

The picture on the left, below the thick arrow, shows a scenario in which the transcribed mRNA normally exits the transcriptional machinery.

The picture on the right, below the thin arrow, shows a situation in which an R-loop is formed as a consequence of high G/C content asymmetries between the leading and lagging DNA strands and the re-annealing of the nascent RNA (adapted from: Chedin, 2016).

R-loops have been shown to be ubiquitously present across different organisms from bacteria to mammals and they are believed to have numerous regulatory effects, which may also affect cell fitness (Skourti-Stathaki & Proudfoot, 2014).

I.IX.I. Mechanisms of R-loop formation and their genome-wide distribution

R-loop formation is supported by factors that provide a thermodynamic advantage to the DNA-RNA hybrid, which is formed during transcription, over the corresponding DNA strand

duplex.

Factors which are facilitating R-loop formation represent intrinsic R-loop prone DNA sequences which can form much more stable duplexes with RNA than DNA (Roy et al, 2008; Belotserovskii et al, 2010). In addition, R-loop formation is favored at genomic sites prone to form breaks in the non-template DNA strand, negative supercoils (positively affecting DNA unwinding) (Roy et al, 2010), or non-canonical DNA structures (as G4-quadruplexes and triplexes) (Duquette et al, 2004). R-loop formation also frequently occurs during transcription scenarios in which the non-template DNA strand is sequestered by a ligand (Belotserovskii & Hanawalt, 2015). Interestingly, factors sequestering the nascent RNA, as it is exiting the transcriptional complex, seem to suppress R-loop formation (Santos-Pereira et al, 2013).

The formation of an R-loop and the maintenance of an R-loop structure seem to be two distinct processes. *In vitro* studies demonstrated that for the R-loop formation initiation, within duplex DNA, an R-loop initiating sequence is required. Such a sequence contains clusters of guanines (very high G/C skew) which have also been demonstrated to form very stable R-loops. On the other hand, a higher thermo-dynamic stability of an R-loop, compared to a DNA-DNA duplex, might be sufficient for R-loop maintenance, but not for R-loop formation initiation (Roy & Lieber, 2009). However, *in vivo* R-loop formation depends not only on intrinsic R-loop forming properties of a certain DNA sequence, but also on DNA-protein interactions. It was shown that R-loop formation is strongly increased close to polyA stretches, even though they are not intrinsically R-loop prone sequences. An explanation for the observed phenomenon might be that polyA stretches disfavor nucleosome binding and offer more "space" for the occurrence of R-loops in terminal gene regions (Wahba et al, 2016; Sanz et al, 2016).

It was also shown that R-loop formation predominantly occurs in genomic regions transcribed by Pol II and that R-loop turnover is a very dynamic and controlled process. Co-transcriptionally formed R-loops are resolved with a half-life of ~10-20 minutes (Sanz et al, 2016). The observed phenomenon is in correlation with estimates of the paused promoter-proximal Pol II half-life (Jonkers et al, 2014), suggesting that R-loop formation is compatible with the normal dynamics of transcription. Accordingly, R-loop formation is a process most probably regulated in terms of frequency and residence time (Sanz et al, 2016).

Apart from high G/C content occurring in R-loop forming regions, especially in the non-template

DNA strand (Ginno et al, 2012, 2013) (Figure 1.13.), other R-loop features have recently been described. The size of an average R-loop ranges from 150 to 500 bp, as directly visualized by electron microscopy (EM) (Duguette et al, 2004), but their size can range up to 2000 bp as shown by molecular R-loop mapping approaches (reviewed in: Santos-Pereira & Aguilera, 2015). In addition, it was estimated that R-loops occupy 5-8 % of the genome (Sanz et al, 2016; Halasz et al, 2017).

The detection of R-loops is based on the use of the S9.6 monoclonal antibody which specifically detects DNA-RNA hybrids (Boguslawski et al, 1986). Most R-loop studies combine the use of the S9.6 antibody for R-loop pull-down, deep sequencing and genome-wide R-loop mapping in different species, from yeast (Chan et al, 2014; El Hage et al, 2014; Wahba et al, 2016) to plants (Xu et al, 2017) and mammals (Ginno et al, 2012; 2013; Nadel et al, 2015; Sanz et al, 2016; Stork et al, 2016; Halasz et al, 2017).

R-loop peaks have been mapped close to gene promoters (Chen et al, 2015), but they have also been found being more spread in gene bodies (Ginno et al, 2012; 2013; Stork et al, 2016) or downstream of transcription start sites (TSS) (Sanz et al, 2016). The observed differences my also possibly be attributed to slight differences in R-loop mapping approaches. Although most R-loop studies show a clear link between R-loops and G/C skew, in yeast R-loops, could be detected even an A/T skew (Wahba et al, 2016).

In a recent study where R-loops where captured in a method called R-ChIP (based on R-loop binding via a catalytically dead RNase H1 followed by strand-specific amplification of immunoprecipitated DNA) it was shown that R-loops have a predominant distribution in promoter proximal regions (59,3%), whereas additional R-loops were mapped across various locations within gene bodies (17,2%), close to gene terminal (6,6%), or in intergenic regions (16,9%). In addition, most detected R-loop peaks were overlapping with open chromatin, which suggests that active gene promoters are hotspots for R-loop formation in the genome (Chen et al, 2017a).

The interplay of R-loops and chromatin composition

R-loops are involved in all stages of gene expression, but are also implicated in almost all the other nuclear processes involving chromatin. Their formation is therefore tightly correlated with

chromatin composition, while on the other hand R-loops also influence chromatin patterning at genomic regions. Several studies, including those with genome-wide mapping of R-loops showed an association of R-loop peaks with open chromatin marks and an overall chromatin accessibility in R-loop rich regions occurring at gene promoters or gene bodies (Nadel et al, 2015; Zeller et al, 2016, Sanz et al, 2016, Chen et al, 2017).

At promoter regions, R-loops were proposed to negatively regulate DNA methylation levels and to facilitate transcription (Gino et al, 2012). R-loop destabilization at the human VIM promoter causes a shift from an open, non-methylated chromatin state to a closed and methylated chromatin (Boque-Sastre et al, 2015). R-loops in promoter regions were shown to associate with high levels of histone marks of active transcription, such as mono and trimethylation of lysine 4 of histone H3 (H3K4me1 and H3K4me3), histone acetylation as well as transcription elongation mark H3K36me3. H3K4me1 was also found at terminal R-loops, implying that this is a common mark of R-loops (Chen et al, 2015; Sanz et al, 2016; Chedin, 2016).

In addition, R-loops may affect the chromatin environment by recruiting chromatin-modifying complexes, Genes with R-loop(+) promoters show high amounts of H3K36me3 over the R-loop formation peak (Sanz et al, 2016). This might be further connected with recruitment of other chromatin factors such as SETD2. Elongation factor Spt6 responsible for Pol II escape from promoter pausing (Vos et al, 2018) and elongation plays a role in maintaining the transcription within protein-coding gene transcription units. In the absence of this factor, H3K36me3 redistributes on intergenic regions, causing increased long non-coding RNA (lncRNA) transcription. These aberrant lncRNAs transcripts can anneal to DNA templates and cause formation of damaging levels of R-loops (Nojima et al, 2018).

H3K4-methylation is also enriched over R-loop(+) promoters - H3K4me3 shows a predominant transcription start site (TSS) distribution, whereas H3K4me1 distributes 1-2 kb downstream of TSS (Sanz et al, 2016), Accordingly, it is proposed that R-loops may have a more direct role in recruitment of H3K4me1, but the mechanism still needs to be elucidated. Interestingly, PAF1 complex which contributes to both H3K4me1 and H3K36me3 methylations, is enriched over R-loop regions (Sanz et al, 2016; Chedin, 2016).

Loss of histone chaperone complex FACT, which functions in chromatin reassembly, leads to R-loop accumulation (Herrera-Moyano et al, 2014). This finding suggests that not only R-loops are inhibitory for nucleosome formation, but nucleosomes may help prevent R-loop

accumulation. Assuming that R-loops are similarly inhibitory for nucleosome formation through the genome, R-loops formation likely alters the kinetics of nucleosome turnover, and consequently contributes to regulation of transcription (Dion et al, 2007).

R-loops were also connected to H3S10P, a histone modification associated mainly with mitotic chromatin condensation. It has been shown that H3S10P levels could be suppressed by expression of RNase H1 enzyme, suggesting a link to R-loops (Castellano-Pozo et al, 2013) but it remains unclear in which context is this histone modification related to R-loops .

Interestingly, R-loop formation has also been linked to increased deposition of marks of repressive chromatin. In one such example, R-loop formation has been linked to H3K9me2/me3 and heterochromatin formation on expanded triplet repeats in Friedrichs's ataxia and Fragile X syndrome, (Colak et al, 2014; Groh et al, 2014).

In *C. elegans* knockout of MET-2 and SET-25, the only two HMTs placing H3K9-methylation results in the frequent accumulation of R-loops behind replication forks, causing genomic instability due to occurring torsional stress. DRIP-Seq (DNA-RNA immunoprecipitation followed by deep sequencing revealed that DNA transposons, usually repressed genomic sequences, show higher levels of R-loops compared to RNA transposons and that loss of H3K9 methylations resulted in de-repression of those sequences. It was thus proposed that H3K9me2 or H3K9me3 chromatin marks stabilize and protect repeat-rich genomes and sequences by suppressing transcription-induced replication stress (Zeller et al, 2016).

Accumulation of H3K9me2 marks and increased R-loop formation was observed in the transcription termination regions in a subset of genes in mammalian cells. This seems to be a consequence antisense transcription, localized dsRNA accumulation and RNAi factor recruitment, as well as the recruitment of heterochromatin protein 1 (HP1) (Skourti-Stathaki et al, 2014).

Interestingly, R-loops were also shown to be negatively correlated with transcription of several developmental genes in murine embryonic stem cells. There, formed R-loops are responsible for the increased binding of Polycomb Repressive Complex 1 and 2, strong epigenetic regulators of transcriptional repression. Conversely, R-loop removal resulted in decreased recruitment of PRC1 and PRC2 and increased Pol II activation (Skourti-Stathaki et al, 2019).

Still, what remains very important to note when studying R-loop biology is their context

dependent function. It needs to be taken into account that R-loops can both facilitate, or terminate DNA transcription depending on their genomic position and interaction with other factors. Hence, the effects of chromatin regulatory factor binding may vary at genes according to the extent of R-loop accumulation (Fazzio, 2016).

Factors interacting with R-loops

It is well established that R-loop forming regions can also be associated to DNA damage (Skourti-Stathaki et al, 2011; Chan et al, 2014; El Hage et al, 2014).

R-loops formed after the transcribing Pol II can block further elongation (Huertas & Aguilera, 2003), while the ones formed after replication forks can block the DNA replicating machinery (Tudori et al, 2009), induce torsional stress, transcription and replication conflicts and contribute to genomic instability (Zeller et al, 2016; Hamperl et al, 2017).

Consequently, cells have evolved multiple mechanisms to control R-loop accumulation and avoid DNA damage and genomic instability, which are crucial in monitoring and keeping cell homeostasis since molecular approaches showed a clear correlation between R-loops and certain diseases (Groh et al, 2014; Lim et al, 2015; Garcia-Rubio et al, 2015; Grunseich et al, 2018).

All regulators of R-loops have a negative effect on them, either resulting in their removal or by preventing their formation (Santos-Pereira & Aguilera, 2015) (Figure 1.14.). One prominent example is the DNA-RNA helicase SETX (Sen1 in yeast). SETX causes the degradation of the

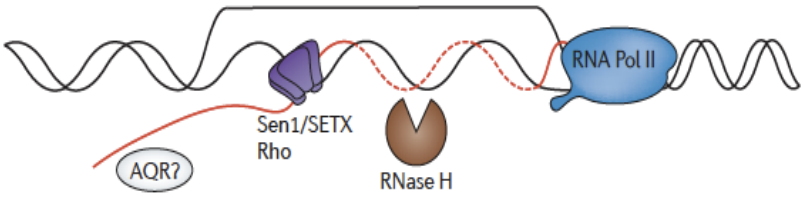


Figure 1.14. Schematic representation of the activity of certain R-loop removing factors.

The removal of co-transcriptionally formed R-loops can occur by the interplay of distinct enzymes.

The RNA moiety of an R-loop can be degraded by RNase H enzymatic activity. R-loops can also be removed by unwinding mechanisms of cellular helicases (for example Rho in bacteria, Sen1 in yeast and SETX in mammalian cells). The human helicase aquarius (AQR) is possibly also involved in the process of R-loop removal (adapted from: Santos-Pereira & Aguilera, 2015).

DNA-RNA duplex of R-loops (Kim et al, 1999), which can be followed by the degradation of the RNA transcript by the RNA exonuclease, Xrn2 (Skourti-Stathaki et al, 2011).

R-loop removal occurs also via the activity of RNase H enzymes (RNase H1 and RNase H2 in eukaryotes), which degrade the RNA strand of the R-loop through their endonuclease activity (Cerritelli & Crouch, 2009). RNase H enzymes represent a family of non-sequence-specific endonucleases that catalyze the cleavage of RNA in a DNA-RNA hybrid via a hydrolytic mechanism (Figure 1.15.). Two main types of RNase H enzymes exist and at least one of them is present in most organisms (Cerritelli & Crouch, 2009).

Early studies on RNase H1 enzymes demonstrated that the N-terminal domain of the protein is responsible for binding of RNA-RNA and DNA-RNA duplexes (Cerritelli et al, 2003; Gaidamakov et al, 2005). This domain is known as the Hybrid Binding Domain (HBD). The HBD shows a 25-fold higher preference for DNA-RNA hybrid binding compared to the same sequence of dsRNA (Nowotny et al, 2005). In addition, it was shown that type 1 RNase H enzymes require at least a four ribonucleotide substrate for cleavage to occur (Cerritelli & Crouch, 2009).

RNase H2 was shown to be clearly different from RNase H1 regarding DNA-RNA hybrid hydrolysis. RNase H2 displays higher affinity in recognizing and cleaving a single ribonucleotide embedded in a DNA duplex. This implies that the function of RNase H2 resembles that of a repair enzyme, since it is cleaving DNA at sites where DNA polymerases mistakenly incorporates a ribo- instead of a deoxyribonucleotide (Cerritelli & Crouch, 2009).

Apart from RNase H enzymes and helicases, Topoisomerase I (Top I) is also involved in the removal of R-loops. Top I is responsible for the relaxation of supercoiled DNA, especially during transcription and DNA replication. For example, depletion of Top I in HEK293T cells results in both R-loop loss and R-loop gain at thousands of transcribed loci. R-loop gains were observed in highly transcribed genes located in gene-poor regions associated to LADs, in close proximity to H3K9me3 enriched regions. On the other hand, R-loop loss in response to Top I depletion was observed in regions repressed by Polycomb group complexes and marked with H3K27me3 mark. In addition, Top I depletion and rearrangement of R-loops throughout the genome resulted in cell cycle arrest in G0/G1 phase, which indicates that Top I function and R-loop regulation is just another level of replication initiation control (Manzo et al, 2018).

Also, by applying a very broad approach in mapping R-loop interacting partners in a human B cell line, 803 new R-loop interacting partners have been identified. Most identified proteins can be classified into groups of proteins involved in RNA processing as splicing or RNA strand unwinding. Interestingly, among the identified R-loop interacting proteins, nuclear pore complex proteins have also been identified too (Wang et al, 2018). So far, it is not known whether there are proteins that specifically promote R-loop stability, either directly or by counteracting the functions of proteins that destabilize R-loops.

I.X. The rationale of studying HIV-1 integration site selection in a chromatin and R-loop dependent context

HIV-1 integration, as well as the completion of the viral life cycle and viral activity, significantly depend on host cell chromatin organization (Le Douce et al, 2012; Maldarelli et al, 2014; Cohn et al, 2015; Matsuda et al, 2015; Laskey et al, 2016; Lusic & Siliciano, 2017). Several studies have demonstrated an association between the distribution of certain chromatin marks and HIV-1 integration sites (Benkirane et al, 1998; Lusic et al, 2003; Ikeda et al, 2007; Wang, 2007; Brady et al, 2009; Marini et al, 2015; Goncalves et al, 2016; Turner & Margolis, 2017).

On the other hand, R-loops, interestingly, expose features that have already been described for HIV-1 integration sites, as DNase I hypersensitivity (Verdin, 1991; Sanz et al, 2016; Chen et al, 2017a) and occurrence in G/C rich genomic regions (Brady et al, 2009; Roy & Lieber, 2009; Ginno et al 2012; Sanz et al, 2016). In addition, R-loop rich regions are frequently associated to open chromatin marks (Sanz et al, 2016; Chedin, 2016) which are also distributed over HIV-1 integration sites (Brady et al, 2009; Marini et al, 2015; Turner & Margolis, 2017).

Finally, HIV-1 RT has a known RNase H function (Davies et al, 1991; Beilhartz & Götte, 2010) and HIV-1 IN exposes features that can be attributed to RNase H enzymes; enzymes that specifically recognize and resolve R-loops (Majorek et al, 2014). The following section will provide further insight into putative RNase H functions of HIV-1 IN.

I.X.I. HIV-1 IN is an RNase H-like enzyme

HIV-1 IN is composed of three independent functional and structural domains, as previously mentioned. In the context of putative RNase H properties of HIV-1 IN the CCD stands out. Also, as mentioned before, the CCD of IN shows an RNase H folding pattern and HIV-1 IN is accordingly classified into the group of Ribonuclease H-like (RNHL) superfamily of enzymes (Figure 1.15.) (Majorek et al, 2014).

Still, the two most important functions of the CCD, and IN in general, are performing the 3' processing and strand transfer reaction. IN order to successfully perform its function, IN activity highly depends on the efficiency of host cell DNA binding. In addition, of significant importance is the fact that IN activity is also highly regulated by the structure of the viral and host DNA substrates which can be influenced by protein interactions. It was shown that the structure of target host DNA greatly influences the site of viral integration as well as that DNA curvature and flexibility influence the frequency of integration (Pruss et al, 1994a; Pruss et al, 1994b; Wang et al, 1999).

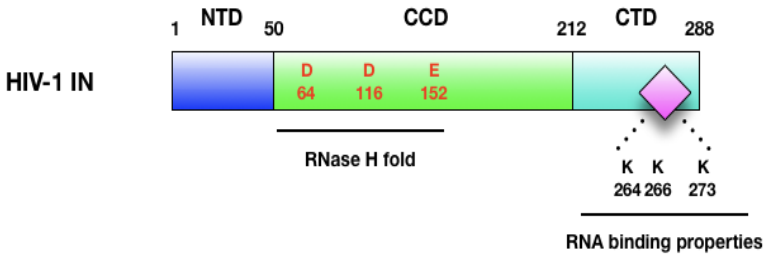


Figure 1.15. Scheme of HIV-1 IN structural domains.
The active residues of the CCD are indicated in red. This region of IN has a demonstrated RNase H conformation.
Three lysine residues that have been described to have RNA binding properties (K264, K266, K273) are highlighted in the CTD of IN.

Apart from DNA binding properties, it was also shown that IN exhibits properties of binding highly structured RNA elements. It was demonstrated that HIV-1 IN in virions shows RNA binding properties where 95% of the bound RNA was derived from viral RNAs, whereas the other 5% were of cellular origin. In addition, it was shown that IN shows preference for binding selected structural viral RNA elements (RNA-TAR). In this context, three lysine residues in the CTD of HIV-1 IN were identified to be important for IN RNA binding properties (K264,

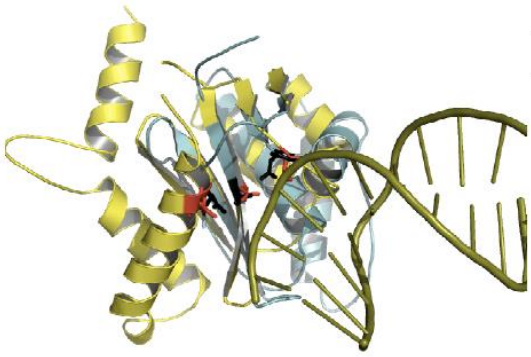


Figure 1.16. Structural superimposition of crystal structures of HIV-1 IN CCD and *B. halodurans* RNase H in complex with an RNADNA-RNA hybrid.
The CCD of HIV-1 IN is represented in yellow, whereas the bacterial RNase H enzyme is represented in turquoise. The red and black regions represent the structurally aligned protein residues of IN (red) and RNase H (black) (adapted from: Savarino, 2007).

K266, K273) (Kessl et al, 2016) (Figure 1.15.).

Supporting evidence for potential RNase H function, as well as R-loop binding properties of IN come from an *in silico* study in which the crystal structures of HIV-1 IN CCD (Goldgur et al, 1999) and an averaged crystal structure of RNase H from *Bacillus halodurans* in the complex with an R-loop (Nowotny et al, 2005) were superimposed (Figure 1.16.). The structural alignment involved 45 amino acids surrounding D71 and D132 of the bacterial RNase H and D64 and D116 of IN (Savarino, 2007).

Considering that the viral RT has a known RNase H function (Kati et al, 1992), and that both enzymes are derived by proteolytic cleavage of the Gag-Pol polyprotein, it seems plausible that the RNase H like domain in HIV-1 IN might have some remaining similar function. .

I.X.II. Questions and objectives

The integrated HIV-1 genome can still not be eradicated from infected cells in patients. Accordingly, it is very important to understand how HIV-1 selects sites in the genome where it integrates and what is the contribution of certain epi-genomic features that could possibly affect this process.

In respect to the fact that HIV-1 integration could potentially occur anywhere in the host cell genome, but still does not, it is worthwhile to further investigate and describe in more detail where and how HIV-1 selects certain genomic sites for integration.

Considering that the virus shows certain preferences for integration site selection and that these integration sites expose similar epi-genomic features, it would be interesting to assess HIV-1 integration patterns once the chromatin composition of HIV-1 target cells is changed.

Apart from understanding and analyzing HIV-1 integration profiles, it would also be important to further explore the 3D positioning of the virus in the nuclear environment and its consequences on viral activity, since previous studies (Albanese et al, 2008; Di Nunzio et al, 2013; Marini et al, 2015) indicate that the nuclear position of the virus is also not random and that the virus preferentially positions in open chromatin regions of the nuclear periphery which are underlying the NPC.

In the context of HIV-1 integration site selection and viral positioning in the genome, the aim would be to focus the attention on the importance and effects of H3K9me2 chromatin mark composition. Mainly because previous studies showed a significant influence of this chromatin mark in shaping the nuclear landscape (Wen et al, 2009; Kind et al, 2013), as well as the importance of H3K9me2 depositing HMT G9a in T cell life cycle and functioning (Lehnertz et al, 2014; Antignano et al, 2014), which represent the main cells targeted for HIV-1 infection. Another reason why the focus of this work is directed to this particular chromatin mark is the fact that it was shown that genes targeted for HIV-1 integration have a slightly different distribution of H3K9me2 chromatin mark throughout their gene body, compared to non-HIV-1 target genes (Lucic et al, 2019).

Taking into account that the host cell chromatin under the NPC is the first open chromatin structure the incoming virus encounters on its way to the nucleus, and that the surrounding closed chromatin is significantly decorated with H3K9me2 chromatin mark, it seems to be important to

further explore the dynamics of HIV-1 integration from a chromatin perspective.
Questions that have to be addressed in this context and that will be pursued in the scope of this work are:

1. Is the virus positioning in open chromatin regions in close proximity to the NPC, just because it is the first accessible chromatin domain on its way to the nucleus?
2. How does chromatin composition affect HIV-1 integration?
3. Are some other genomic features also affecting HIV-1 integration site selection and its positioning in the 3D nuclear environment?

In addition, in order to have a broader insight into HIV-1 integration site selection this work aims to implement questions on HIV-1 integration site selection from a viral perspective too, but still integrating the possible importance of R-loops (as a genomic feature) as targets for HIV-1 integration.
The main reasoning to implement such a perspective on HIV-1 integration site selection into this work arises from HIV-1 IN properties, as well as R-loop properties that can be associated to viral integration sites.

Respectively, HIV-1 IN is classified into the enzymatic family of Ribonuclease H-like enzymes and its CCD shows an RNase H folding pattern which can also *in silico* be superimposed to the structure of a bacterial RNase H enzyme in complex with a DNA-RNA hybrid (Savarino, 2007). This finding from a computational approach already demonstrates that HIV-1 IN has potential RNase H function and could potentially bind to a DNA-RNA hybrid.
Apart from computational approaches demonstrating DNA-RNA hybrid binding properties of HIV-1 IN, clear RNA binding properties of the CTD of HIV-1 IN have been shown (Kessl et al, 2016) which represents supporting evidence that HIV-1 IN could bind to an R-loop, and particularly to its RNA part.
In addition, when comparing certain features of HIV-1 integration sites with features of R-loop forming regions certain similarities seem to be evident;
Most R-loops are formed during transcription in G/C rich genomic regions (Ginno et al, 2012; 2013), whereas HIV-1 preferentially integrates into gene bodies of highly transcribed genes with high G/C content (Busschots et al, 2005; Ciuffi et al, 2005; Brady et al, 2009). In addition,

R-loop forming regions as well as HIV-1 integration sites show features of DNase I hypersensitivity (Verdin, 1991; Chen et al, 2017a), and both R-loops and HIV-1 integration sites are mostly associated to regions of active chromatin marks (Wang et al, 2007; Marini et al, 2015; Sanz et al, 2016).

Finally, with respect to all described features of HIV-1 IN, R-loops as specific genomic features and their similarities to sites targeted for HIV-1 integration, questions that can be raised and that will be addressed in this work are:

1. Do HIV-1 target genes contain R-loops?
2. Do R-loops represent genomic features recognized by HIV-1 IN?
3. Does HIV-1 integrate into or in the close proximity of R-loops?

The main objective of this work is to contribute to a better and more detailed understanding, and description of sites targeted for HIV-1 integration.
In addition, by trying to answer all set questions, this work will try to unreel potential novel features of HIV-1 IN, as well as try to contribute to a better understanding of the importance of H3K9me2 chromatin mark and R-loops in the context of HIV-1 integration.

II. Results

II. I. Chromatin structure disruption affects HIV-1 integration in primary CD4$^+$ T cells

Primary CD4$^+$ T cells, isolated from healthy blood donors, were pretreated with several drugs affecting the activity of different chromatin modifying enzymes and subsequently infected with HIV-1. Used drugs in this assay were SAHA, JQ1 and BIX01294 (Figure 2.1.).

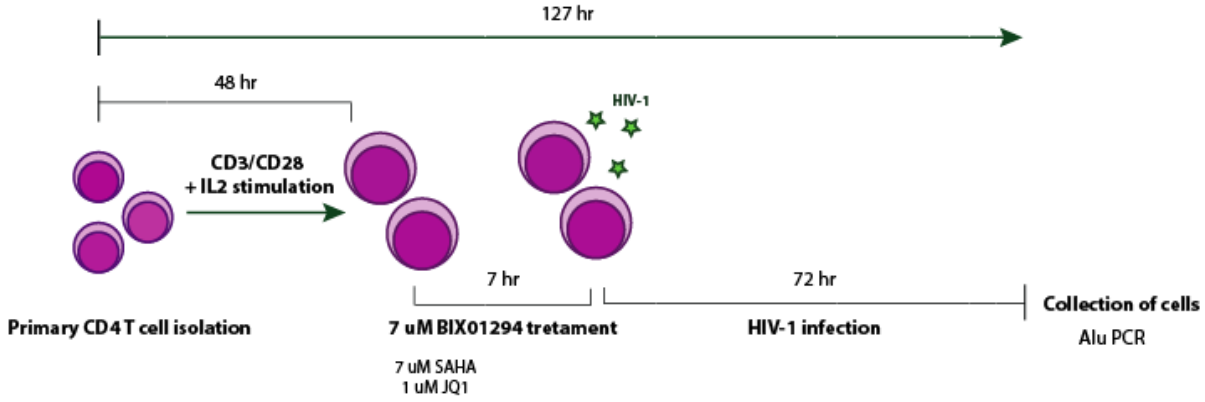


Figure 2.1. Scheme of the experimental procedure of primary CD4$^+$ T isolation, activation and infection.
Primary CD4$^+$ T cells were isolated from whole blood of healthy blood donors and activated for 48 h with CD3/CD28 activation beads and IL-2. After activation cells were treated with selected drugs for 7 h and subsequently infected with HIV-1. The cells were collected 72 h post infection and processed for Alu PCR analysis.

HDAC inhibitor SAHA (also known as Vorinostat) is a pan- histone deacetyltransferase (HDAC) inhibitor expected to increase global histone acetylation patterns (Bradner et al, 2010; Kim et al, 2013; Seto & Yoshida, 2014). BET inhibitor JQ1 was described to cause the unloading of super-enhancers and thereby cause an overall chromatin remodeling (Filippakopoulos et al, 2010). HMT inhibitor BIX01294 is known to specifically inhibit the activity of G9a, cause a reduction of H3K9me2 repressive chromatin mark and modify an overall chromatin structure (Kubicek et al, 2007). The drugs effects in changing the chromatin composition and their

potential effects on HIV-1 integration levels were assessed by Alu PCR (Figure 2.2. A). All tested drugs caused an increase in HIV-1 integration, but the effect of G9a inhibitor BIX01294 was consistently the most prominent, by causing an overall increase of integration by 2,8-5 fold compared to HIV-1 integration in untreated conditions.

In addition, the effect of UNC0642 on HIV-1 integration, which is another G9a inhibitor with improved biochemical properties (more specificity and improved Pharmacokinetic properties in comparison to other HMT inhibitors) (Liu et al, 2013b) was also tested. Still, the effect of BIX01294 causing an increase in HIV-1 integration by 2-4,6 fold was always more pronounced, since UNC0642 was only able to cause an increase in HIV-1 integration by maximum of 1,6 fold (Figure 2.2. B). Based on these results, we opted to use BIX01294 in all further experiments with a goal to determine the importance of chromatin composition on HIV-1 integration.

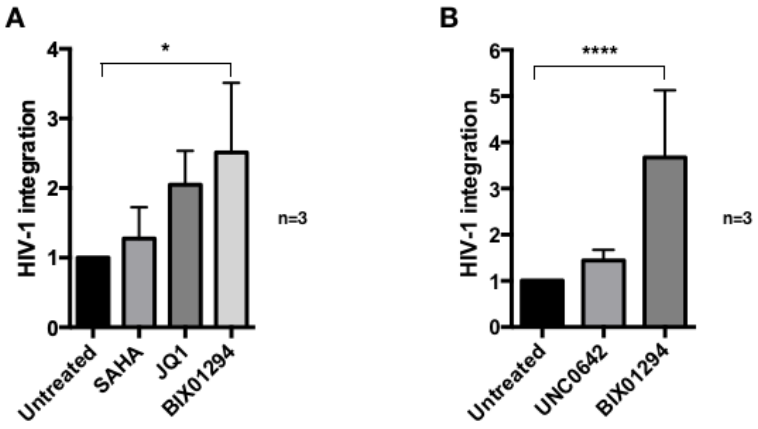


Figure 2.2. Alu PCR results of HIV-1 integration levels upon drug treatments.

A. HIV-1 integration levels were determined in at least 3 independent donors upon primary $CD4^+$ T cell pretreatment with SAHA, JQ1 and BIX0124. The untreated infected control is set to 1, and is used as a reference for the integration efficiency, as measured by Alu PCR, under different treatment conditions. Integrated vDNA (Alu Ct values) is normalized over total genomic DNA levels amplified with primers for the lamin B2 gene (B13 region). The graphs represent a summary of 3 experiments, **B.** HIV-1 integration levels determined in 3 independent donors upon primary $CD4^+$ T cell pretreatment with 2 different G9a inhibitors; UNC0642 and BIX0124. The untreated infected control is set to 1, and is used as a reference for the integration efficiency, as measured by Alu PCR, under different treatment conditions. Integrated vDNA (Alu Ct values) is normalized over total genomic DNA levels amplified with primers for the lamin B2 gene (B13 region).

II. II. Determining the kinetics and effects of BIX01294 in primary CD4$^+$ T cells

In the first instance, cells were treated 6-7 h with BIX01294. The effect of BIX01294 after 6-7 h treatment of primary CD4$^+$ T cells was first tested and confirmed by western blot, where a significant reduction of H3K9me2 chromatin mark was observed (Figure 2.3. A).

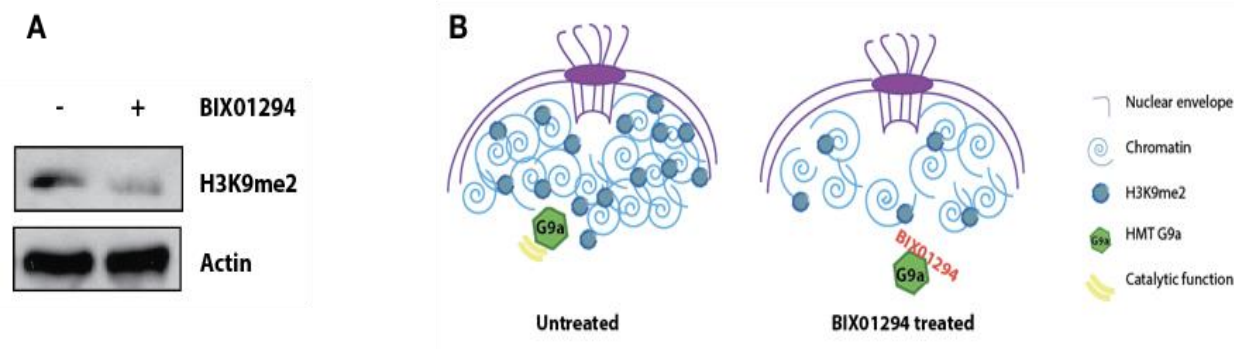


Figure 2.3. Effect of BIX01294 on H3K9me2 chromatin mark depletion.

A. Western blot results upon 6-7 h 7 µM BIX01294 treatment. As a readout for drug treatment efficiency H3K9me2 chromatin mark levels are shown, as BIX01294 inhibits G9a activity which deposits H3K9me2 chromatin mark. As protein loading control total levels of ß-Actin are presented. **B.** Schematic representation of the effect of BIX01294 on G9a activity and deposition of H3K9me2 chromatin mark. In untreated conditions G9a is depositing H3K9me2 chromatin mark, creating an environment of condensed chromatin primarily in the nuclear periphery. During BIX01294 treatment the catalytic activity of G9a is inhibited which causes a decrease in H3K9me2 chromatin mark deposition and a global opening of the chromatin.

In addition, the reduction of H3K9me2 could also be appreciated by IF (Immunofluorescence) staining and STED (Stimulated Emission Depleted) microscopy, where a clear global reduction of H3K9me2 chromatin mark was determined in cell nuclei (Figure 2.4.). Interestingly, two different H3K9me2 staining patterns have been detected among cells of the same donor (Figure 2.4. upper panels). One staining pattern shows a global nuclear distribution of H3K9me2 which upon BIX01294 treatment also gets reduced on a global nuclear level. The other observed staining pattern of H3K9me2 shows a more peripheral distribution of the chromatin mark which upon BIX01294 treatment also gets reduced in the nuclear periphery.

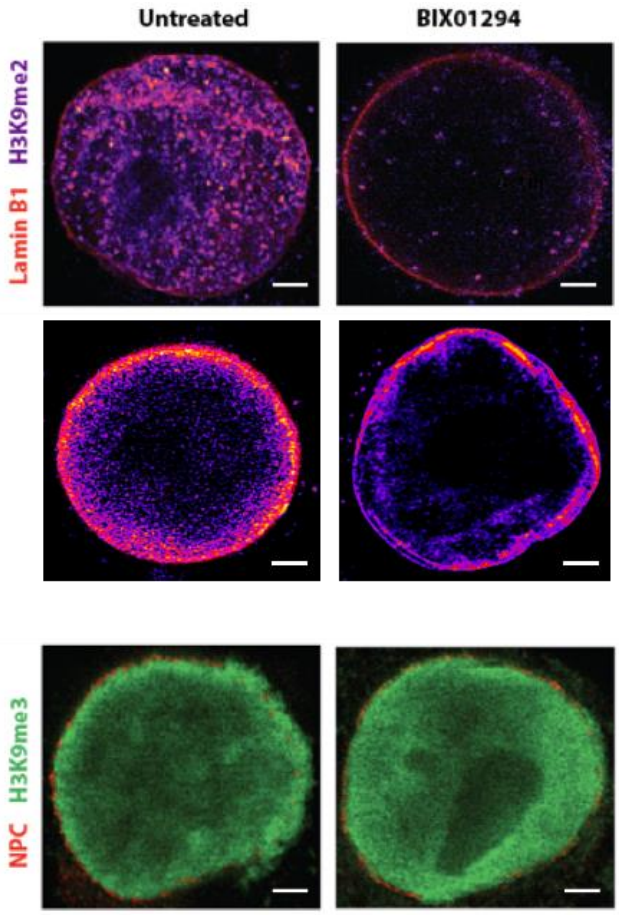


Figure 2.4. Specific reduction of H3K9me2 upon BIX01294 treatment.

IF staining for H3K9me2 chromatin mark followed by STED microscopy of control (untreated) or cells treated with BIX01294 is shown in the upper panels. Staining of related H3K9me3 chromatin mark under same conditions represents a control for the specificity of the drug treatment (lower panels) (scale bars: 1 μm) (imaged with the help of Dr. Vibor Laketa).

In addition, a control staining of related chromatin mark H3K9me3 was performed in untreated and BIX01294 treated conditions, since BIX01294 is expected to specifically inhibit the activity of HMT G9a and respectively only affect the deposition of H3K9me2 chromatin mark. As expected, and as seen by IF and immunoblot, H3K9me3 was not affected by BIX01294 treatment (Figure 2.4. and Figure 2.6. A).

Cell viability assessment upon BIX01294 treatment and HIV-1 infection

In order to assess cell viability upon BIX01294 treatment, HIV-1 infection as well as simultaneous drug treatment and infection, the MTT assay (3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromidefor assay) was performed.

The MTT assay is used to determine cell metabolic activity by assessing the presence of NAD(P)H dependent oxidoreductases in the sample. Oxidoreductases can reduce the tetrazolium dye MTT to its insoluble form - formazan which turns the sample into a purple color. The more oxidoreductases are present in the sample, the more formazan will be produced indicating that the sample contains less viable cells.

Upon BIX01294 treatment no significant difference in cell viability was observed. On average 92% of the cells are still viable in comparison to untreated conditions where assumed that 100% of cells are viable. Cells upon HIV-1 infection on average do not show any difference in viability compared to untreated conditions, whereas upon BIX01294 treatment and HIV-1 infection, on average, 64% of cells are still viable (Figure 2.5). However, some cell death upon HIV-1 infection was expected (Costin, 2007), and especially upon drug treatment and infection. Accordingly, the observed MTT results indicate that 7 µM BIX01294 treatment does not have any harmful effect on cell viability, which is also confirmed by one-way ANOVA statistical analysis which gave a p value of 0,073 (the value is higher than 0,0001 and considered as statistically not significant) .

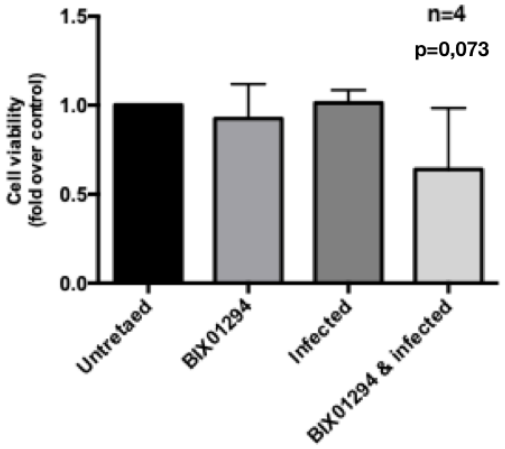


Figure 2.5. CD4$^+$ T cell viability test upon BIX01294 treatment and HIV-1 infection.

MTT assay results of 4 independent donors upon BIX01294 treatment, upon HIV-1 infection and simultaneous BIX01294 treatment and HIV-1 infection. The untreated control is set to 1 and used as a reference to assess cell viability in the analyzed conditions. The difference in cell viability between analyzed conditions is not statistically significant as assessed by one-way ANOVA analysis (p value is shown in the graph).

The timeline of BIX01294 effect on H3K9me2 chromatin mark

In order to further assess the kinetics of BIX01294 treatment and its effects, levels of H3K9me2 were assessed at different time points post drug treatment and drug release by western blot (Figure 2.6. A). Blot quantifications showed that after 6 h of treatment H3K9me2 levels were reduced to 62,33%; 12 h post drug release to 40,33%; 24 h post drug release to 33,67%; 48 h post drug release to 57,67% and 72 h post drug release to 69,67% of initial H3K9me2 levels. The strongest effect regarding H3K9me2 reduction was appreciated between 12 and 24 h post drug release. After 24 h post drug release H3K9me2 levels start getting reconstituted back to initial levels (Figure 2.6. B).

Observing the strongest reduction of H3K9me2 between 12 and 24 h post drug release is also in accordance with the timing of occurred HIV-1 integration (Mohammadi et al, 2015). Obtained results are compatible with increased levels of HIV-1 integration upon BIX01294 treatment observed by Alu PCR (Figure 2.2).

In addition, it can also be appreciated that H3K9me3 levels are not changing over time, which also shows that the drug treatment specifically targets only G9a and H3K9me2 chromatin mark (Figure 2.6. A).

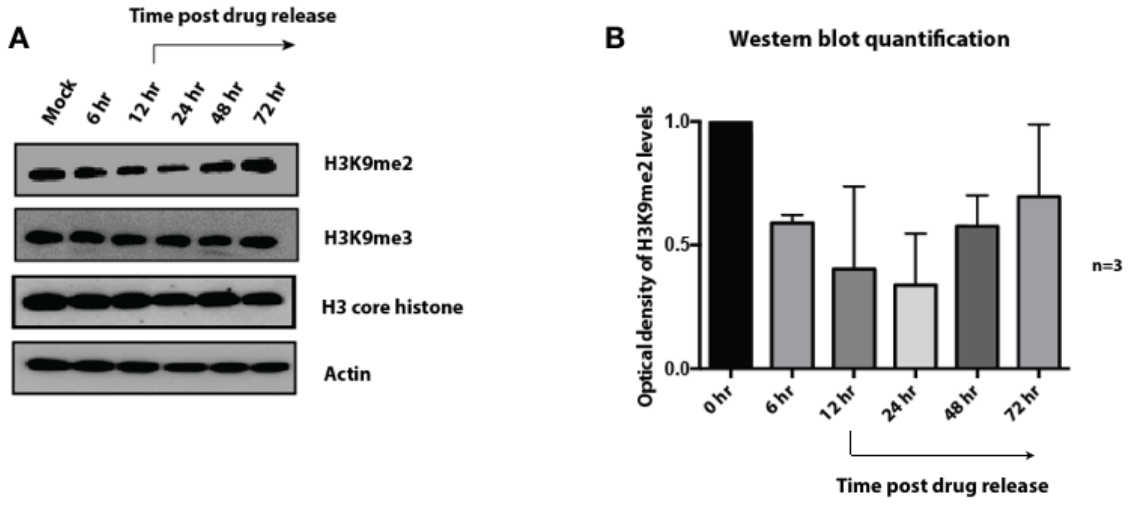


Figure 2.6. The kinetics of BIX01294 treatment followed by the assessment of H3K9me2 levels at different time points post drug treatment and drug release.

A. Levels of H3K9me2 chromatin mark at different time points post BIX01296 treatment and drug release were

assessed by immunoblot. At each time point, levels of H3 core histones and H3K9me3 control chromatin mark, which are not changed upon BIX01294 treatment, were also assessed. ß-Actin served as a loading control. **B.** Quantification of optical density of 3 independent immunoblots for H3K9me2 chromatin mark upon BIX01294 treatment and drug release. Optical density values of H3K9me2 chromatin mark at each time point are normalized over protein loading control ß-Actin values. Time point 0 h of the drug treatment is set to 1 and used as a reference to assess optical density values of H3K9me2 chromatin mark at different time points post drug treatment initiation and drug release.

Assessment of potential DNA damage upon BIX01294 treatment

In order to exclude the possibility that the observed increase in HIV-1 integration might be caused by increased levels of DNA damage upon drug treatment, immunofluorescence (IF) staining of Gamma H2AX, DNA damage marker, which stains DNA double strand breaks was performed in untreated and drug treated conditions.

By confocal microscopy it could be observed that primary $CD4^+$ T cells in untreated conditions and upon 6-7 h of 7 µM BIX01294 treatment have similar numbers of cells positive for Gamma H2AX signals. In untreated conditions 32% cells contained Gamma H2AX signals, whereas in BIX01294 treated conditions 28% cells contained Gamma H2AX signals (Figure 2.7. B).

In addition, the total amount of Gamma H2AX signals was also quantified by measuring total immunofluorescence intensity in both conditions. In both cases, no difference in total immunofluorescence intensity of Gamma H2AX could be detected. Accordingly, any DNA damage occurring upon BIX01294 treatment could be excluded (Figure 2.7. B and C).

The obtained results indicate that increased levels of HIV-1 integration upon BIX01294 treatment are not a consequence of possible DNA damage occurring due to drug treatment.

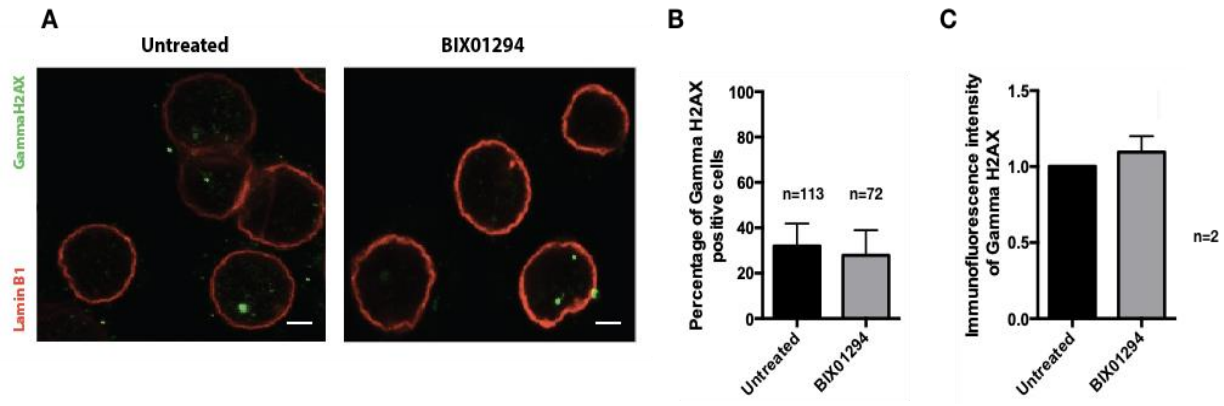


Figure 2.7. Assessment of Gamma H2AX DNA double-strand break marker occurrence upon BIX01294 treatment.

A. Confocal images of primary $CD4^+$ T cells stained for Gamma H2AX (green) and lamin B1 (red) in untreated conditions and upon BIX01294 treatment (scale bars: 2 µm). **B.** Graph shows results of absolute cell counts containing Gamma H2AX signals in untreated conditions and upon BIX01294 treatment, expressed in percentages. Numbers above the bars show total cell counts for 2 independent donors. **C.** Graph shows results of total IF intensity signals quantified in untreated conditions and upon BIX01294 treatment in 2 independent donors. The IF intensity values in untreated conditions are set to 1 and used as a reference to assess IF intensity values upon BIX01294 treatment.

Cell activation status upon BIX01294 treatment

Considering the observed increase in HIV-1 integration upon BIX01294 treatment and the fact that HIV-1 can only infect activated $CD4^+$ T cells, the activation status of resting primary CD4 T cells upon BIX01294 treatment was tested by staining for CD38 and CD69 cell surface marker and subsequent FACS (Fluorescence-activated cell sorting) assay. This experiment was performed in order to verify whether the observed increase in HIV-1 integration upon BIX01294 treatment could possibly be a consequence of increased cell activation status.

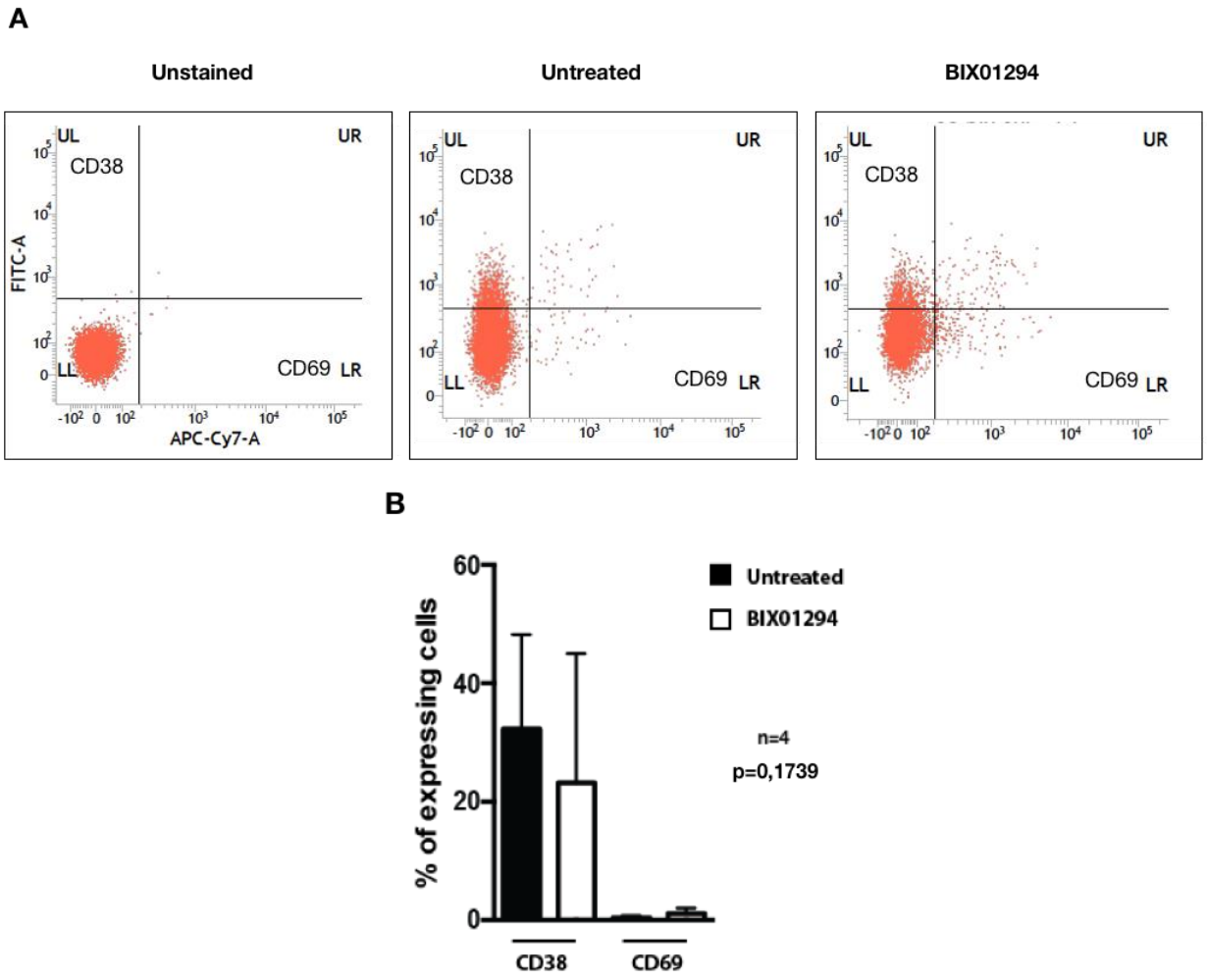


Figure 2.8. Determination of CD38 and CD69 cell surface marker expression upon BIX01294 treatment.
A. Representative FACS profiles of resting primary $CD4^+$ T cells for CD38 and CD69 cell surface marker in untreated conditions and upon BIX01294 treatment. The unstained condition is given as a negative control, and shows the gating of only live cells which are occupying forward and side scatter scales of up to 10^2. CD38 positive cells appear in the upper left quadrant of the FACS plot (FITC stained) (y-axis). CD69 positive cells appear in the lower right quadrant of the FACS plot (APC-Cy7 stained) (x-axis). The given p value of two-way ANOVA analysis is given in the graph. **B.** The graph shows CD38 and CD69 cell surface marker expression levels summarized for 4 independent donors expressed in percentages. The legend in the upper right corner indicates that the black bars correspond to cell surface marker expression results in untreated conditions, whereas the white bars correspond to results upon BIX01294 treatment.

Levels of CD38 expression vary during lymphocyte development, activation and differentiation, and is usually highly expressed in activated cells. CD69 is considered to be an early activation

marker. In untreated conditions 32,26% cells are expressing CD38, whereas upon BIX01294 treatment 29,37% cells express CD38. In case of CD69, in untreated conditions 0,42% cells express CD69, whereas upon BIX01294 treatment 1,11% cells express CD69 (Figure 2.8.).

The obtained FACS results, performed statistical analysis and given p value of p=0,1739 (p>0,0001) indicate that there is no significant change in cell activation status upon BIX01294 treatment of resting primary CD4 T cells. The results indicating that the observed increase in HIV-1 integration levels upon BIX01294 treatment are not due to increased cell activation status, but rather due to a general more open and permissive chromatin structure.

II.III. G9a knock down has the same effect on HIV-1 integration as BIX01294 treatment

To verify whether BIX01294 treatment and G9a depletion have the same increasing effect on HIV-1 integration, G9a was directly knocked down (KD) by using a lentivirus containing a siRNA (small interfering RNA) sequence against this histone methyltransferase in Jurkat T cell line.

The lentivirus additionally contained a puromycin resistance gene, which enabled selection of cells containing the successfully integrated viral vector harboring G9a shRNA.

Consistent with an observed reduction of G9a (Figure 2.9. A) and decreased H3K9me2 levels, a ~3-fold increase in HIV-1 integration upon infection of Jurkat cells was measured by Alu PCR (Figure 2.9. B).

Obtained results indicate that BIX01294 treatment, as well as direct KD of G9a induce a global reduction of H3K9me2 chromatin mark, which both have an increasing effect on HIV-1 integration.

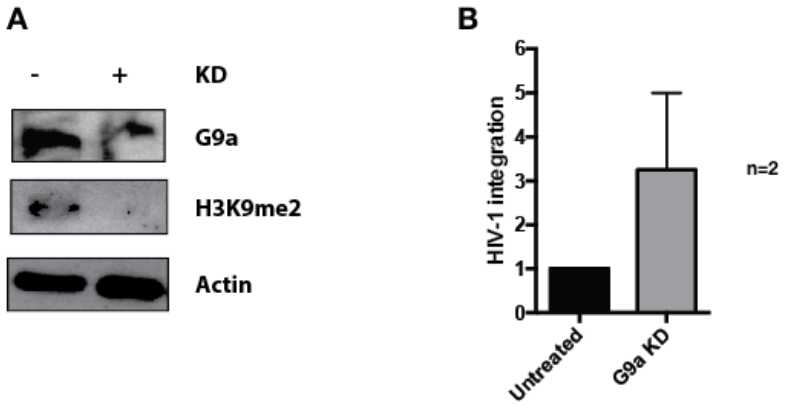


Figure 2.9. G9a knock down and its effect on HIV-1 integration in Jurkat T cell line.

A. Efficiency of G9a KD, as well as consequent H3K9me2 depletion measured by immunoblot. G9a levels were assessed upon lentiviral KD of the protein, followed by the assessment oh H3K9me2 levels upon G9a KD. As protein loading controls total levels of ß-Actin are presented. **B.** Alu PCR results showing HIV-1 integration levels upon G9a KD, summarized for 2 individual experiments. HIV-1 integration in untreated conditions is set to 1 and used as a reference to assess HIV-1 integration levels upon G9a KD. Integrated vDNA (Alu Ct values) is normalized over total genomic DNA levels amplified with primers for the lamin B2 gene (B13 region).

II. IV. H3K9me2 depletion has an increasing effect on HIV-1 integration, viral RNA expression and viral activity

Considering that BIX01294 treatment results showed increased HIV-1 integration (Figure 2.10 A), we asked if these integrations read to productive or aberrant viral transcription. We therefore assessed the levels of HIV-1 transcription and viral mRNA production. For this purpose, RNA was extracted upon BIX01294 treated and HIV-1 infected cells, and viral transcripts were quantified by RT-PCR followed by qPCR (quantitative PCR) using primer and probes for short, mostly (NUC1B) aberrant and long (U1A) viral transcripts from the *gag* gene (Lusic et al, 2003). Indeed, the increase in HIV-1 integration upon H3K9me2 depletion was followed by increased viral mRNA levels in both, early and late viral transcripts, 72 h post infection, compared to untreated and infected samples. Short aberrant viral transcripts were 1,57 fold increased, whereas long viral transcripts, corresponding to the mature forms of the *gag* gene were 2,54 fold increased, following the increase in HIV-1 integration of 2,5 fold in comparison to untreated conditions (Figure 2.10. B).

Viral activity upon BIX01294 treatment was assessed by FACS assay, by measuring the production of viral p24 protein. Obtained FACS results show an increase in p24 levels upon BIX01294 treatment and viral infection up to 2,59 fold compared to untreated conditions (Figure 2.10. C). The obtained FACS results are in correlation with the previously observed increase in HIV-1 integration upon BIX01294 treatment of 2,5 fold and the increase in detected mature forms of *gag* gene transcripts of 2,54 fold, compared to untreated conditions.

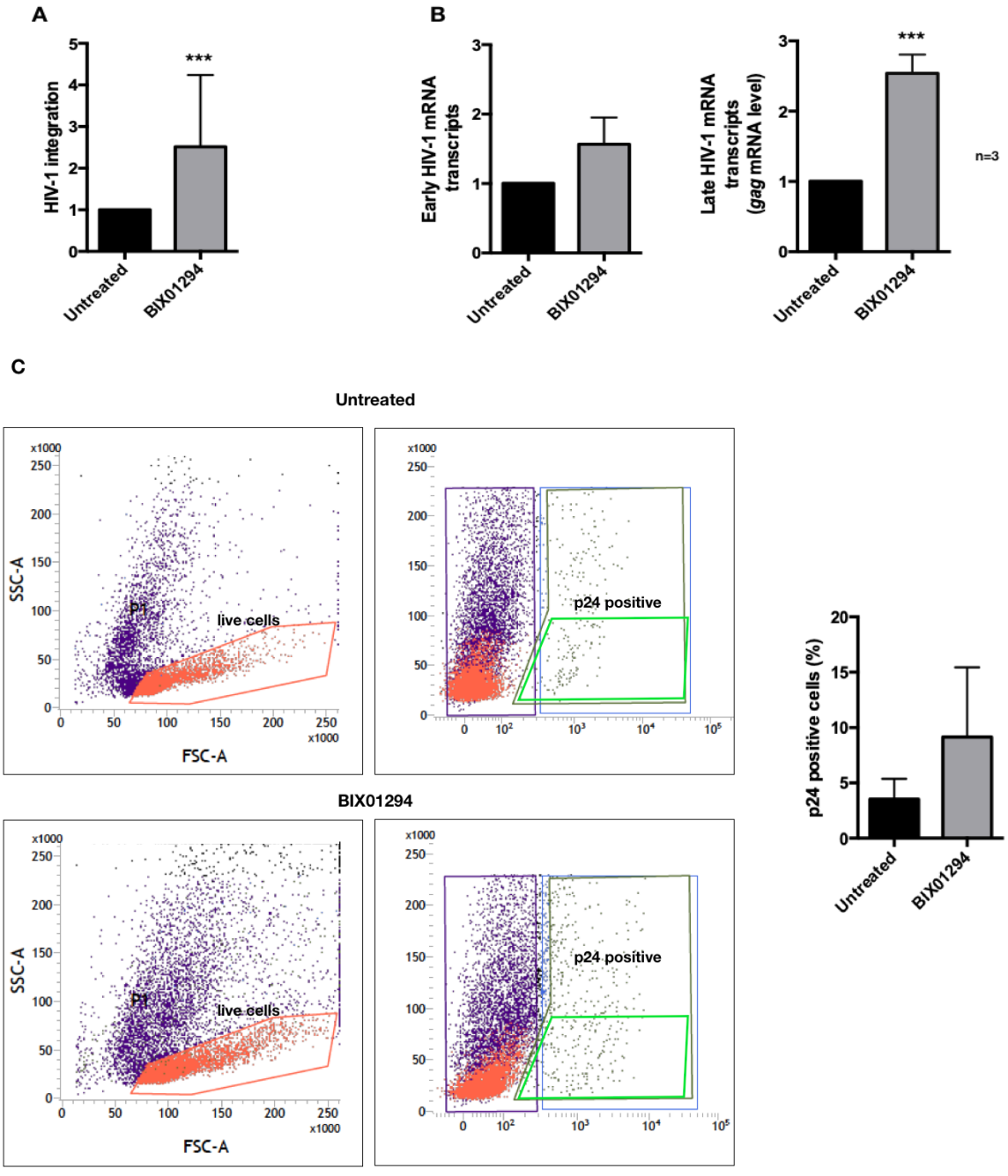


Figure 2.10. Assessment of HIV-1 integration levels, viral RNA expression and viral activity upon BIX01294 treatment and H3K9me2 depletion.

A. HIV-1 integration in untreated conditions and upon BIX01294 treatment, summarized for 3 independent donors, as measured by Alu PCR. HIV-1 integration in untreated conditions is set to 1 and used as a reference to assess HIV-1 integration upon BIX01294 treatment. Integrated vDNA (Alu Ct values) is normalized over total genomic DNA levels amplified with primers for the lamin B2 gene (B13 region). **B.** HIV-1 RNA expression

levels upon BIX01294 treatment, summarized for 3 independent donors. Graph on the left shows mRNA expression of early viral transcripts, where the values in untreated conditions are set to 1 and used as a reference to assess the expression of early viral transcripts upon BIX01294 treatment. The graph on the right shows mRNA expression levels of late viral transcripts, where the values in untreated conditions are set to 1 and used as a reference to assess the expression of early viral transcripts upon BIX01294 treatment. Viral gene expression levels are normalized over host cell housekeeping gene GAPDH. **C.** Representative FACS profiles of primary $CD4^+$ T cells infected with HIV-1 in untreated and BIX01294 treated conditions and stained for viral protein p24. Unstained cells in both conditions are represented as a negative control, and show the gating of only live cells which are represented in the orange window and occupy forward and side scatter scales of up to 10^2. Live p24 positive cells are marked in respect to the cells marked as live in the unstained control, but shifted to forward scatter values up to 10^5, caused by the presence of the p24 protein. The graph on the right represents a summary of p24 positive cells upon HIV-1 infection in untreated and BIX01294 treated conditions for 3 independent donors expressed in percentages. Results from 3 independent donors shown in panel A correspond to the same donors and results shown in panel B and C.

Apart from increased HIV-1 integration upon H3K9me2 depletion, which is followed by increased viral activity, assessed by measuring viral RNA and p24 production (Figure 2.10. A, B and C), we also observed a decrease in non-integrated viral DNA. This decrease, as measured by PCR for 2LTR circles (Figure 2.11. A and B), suggests that following nuclear entry all copies of HIV-1 viral DNA get integrated into the cellular genome, probably due to more open chromatin structure following BIX01294 treatment.

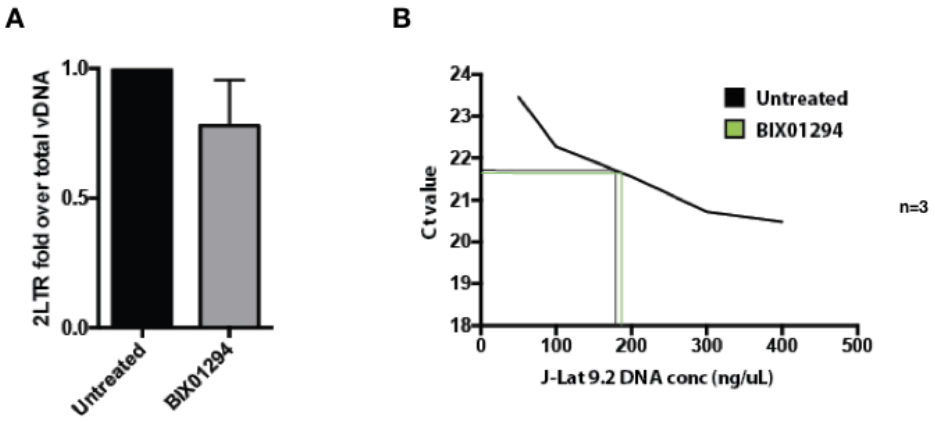


Figure 2.11. Quantification of non-integrated HIV-1 DNA (2LTR circles) upon H3K9me2 depletion.
A. Graph shows the quantification of 2LTR circles upon HIV-1 infection following BIX01294 treatment, as normalized over levels of total vDNA (graph in panel B). A summary of 3 independent donors is shown **B.** Quantification of total vDNA upon HIV-1 infection in untreated and BIX01294 treated conditions, summarized for 3 independent donors. Total vDNA was quantified by plotting the amounts of total vDNA qPCR Ct values (y-axis) over the values obtained by serial dilutions of DNA (represented as ng/µl) from latent HIV-1 clone J-Lat 9.2 cell DNA (x-axis). 3 independent donors, same as the ones shown in Figure 2.10., were used for quantification.

II.V. H3K9me2 depletion affects 3D nuclear position of HIV-1

The 3D positioning of HIV-1 in $CD4^+$ T cell nuclei was determined by immuno-DNA FISH (Fluorescent *in situ* Hybridization) followed by confocal microscopy in untreated and BIX01294 treated conditions. In untreated conditions, the virus preferentially positions in the outer shell of the nucleus (occupying nuclear zone 1), as previously described (Marini et al, 2015). Upon drug treatment, due to the modification of histone methylation patterns, more HIV-1 DNA FISH signals per nucleus (cell) were observed, consistent with the detected increase in HIV-1 integration upon BIX01294 treatment, as seen by Alu PCR.

Interestingly, upon BIX01294 treatment, we observed changes in the distribution patterns of the virus, as it is positioned in the interior zones of the nucleus, occupying predominantly nuclear zone 2 (Figure 2.12. A and B). In untreated conditions 68,87% HIV-1 DNA FISH signals occupy nuclear zone 1, 28,12% occupy nuclear zone 2 and 3,01% nuclear zone 3. In BIX01294 treated conditions 40,28% HIV-1 DNA FISH signals occupy nuclear zone 1, 52,78% occupy nuclear

zone 2 and 6,94% nuclear zone 3. This finding was quantified by radial measurements of nuclei and determination of relative distances of HIV-1 signals to the nuclear lamina as previously described in Marini et al, 2015.

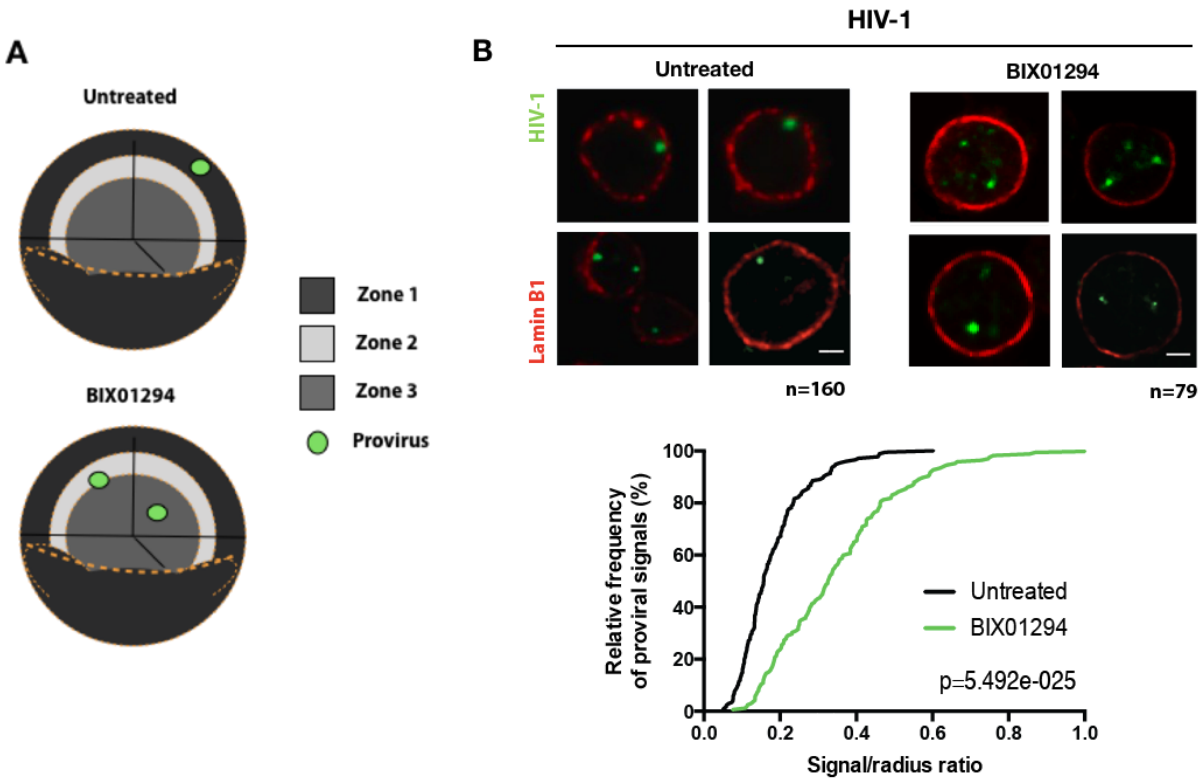


Figure 2.12. 3D nuclear position of HIV-1 upon H3K9me2 depletion and BIX01294 treatment.

A. Schematic representation of the nuclear zones and the 3D nuclear distribution of HIV-1 in untreated vs. BIX01294 treated conditions. Legend on the right indicates nuclear zones and division of the nucleus into areas of equal volume according to their distance to the nuclear lamina. **B.** FISH results of HIV-1 3D nuclear position in untreated conditions and upon BIX01294 treatment. Numbers under the images indicate the number of infected cells in which the distance of HIV-1 FISH signals to the nuclear lamina was measured (scale bars: 2 µm). Graph below shows the cumulative distribution of HIV-1 proviral signals by the Kolmogorov-Smirnov test in relation to their relative distance to the nuclear lamina, with the corresponding p value above the x-axis of the graph.

II.VI. HIV-1 RIGs follow the pattern of HIV-1 positioning in the 3D nuclear space

As for HIV-1, the nuclear position of selected HIV-1 RIGs and HIV-1 non-target genes in untreated and BIX01294 treated conditions was determined by immuno-DNA FISH and confocal microscopy.

In total, 15 genes were analyzed (BACH2, NPLOC4, STAT5B, MKL2, DNMT1, RPTOR, SMG1, GRB2, GNB1, NFATC3, PACS2, KDM2B, ACTN1, RNF157, CNTN4). Genes as BACH2, NPLOC4, STAT5B, MKL2, DNMT1, RPTOR, SMG1, GRB2, GNB1, RNF157 and NFATC3 were chosen among RIGs, whereas PACS2, KDM2B, ACTN1, CNTN4 where chosen as HIV-1 non-target genes. RIGs BACH2, NPLOC4, STAT5B, MKL2, DNMT1, RPTOR, SMG1 GRB2 and RNF157 clearly change their nuclear position upon BIX01294 treatment and H3K9me2 depletion (Figure 2.13.), whereas GNB1& NFATC3 did not change their 3D nuclear position upon BIX01294 treatment (Figure 2.14.).

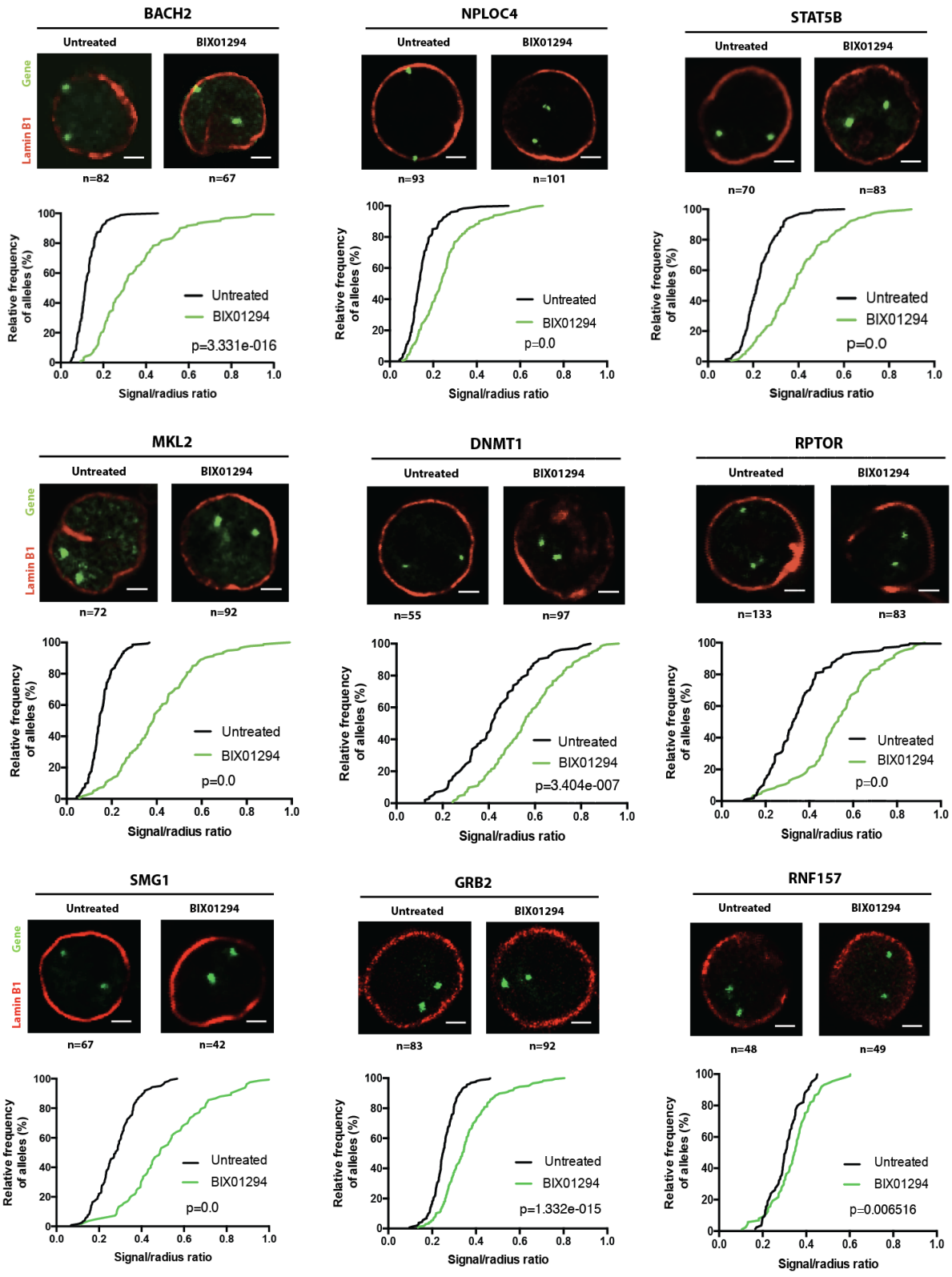


Figure 2.13. 3D nuclear position of RIGs that change their nuclear position following BIX01294

treatment.

Numbers under the images indicate the total cell count in which the distance of the gene FISH signals to the nuclear lamina was measured. Graphs below the images show the cumulative distribution of alleles by the Kolmogorov-Smirnov test in relation to their relative distance to the nuclear lamina, with the corresponding p value above the x-axis of the graph (scale bars: 2 µm).

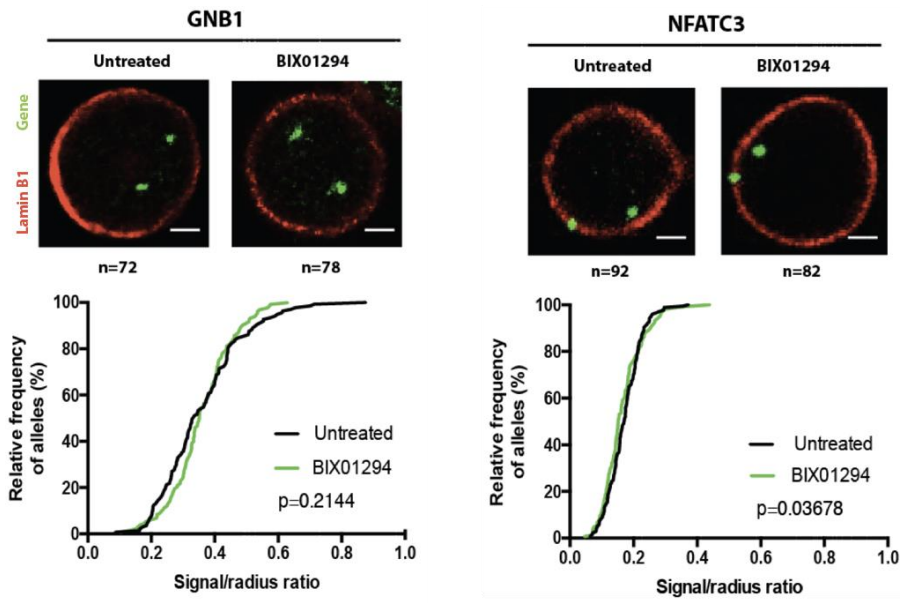


Figure 2.14. 3D nuclear position of RIGs that do not change their nuclear position upon H3K9me2 depletion and BIX01294 treatment.

Numbers under the images indicate the total cell counts in which the distance of the gene FISH signals to the nuclear lamina was measured. Graphs below the images show the cumulative distribution of alleles by the Kolmogorov-Smirnov test in relation to their relative distance to the nuclear lamina, with the corresponding p value above the x-axis of the graph (scale bars: 2 µm).

Chosen HIV-1 non-target genes PACS2, KDM2B, ACTN1, CNTN4 do not move from their initially occupied nuclear zone, which is zone 2 for PACS2 and KDM2B. For CNTN4 gene locus, which is inside a LAD, we observed a slight, statistically significant ($p<0,0001$) shift towards the nuclear periphery upon BIX0124 treatment. In the case of ACTN1 the opposite phenomenon can be appreciated, where a minor, although statistically significant ($p=0,0014$) shift towards the center was observed. (Figure 2.15.).

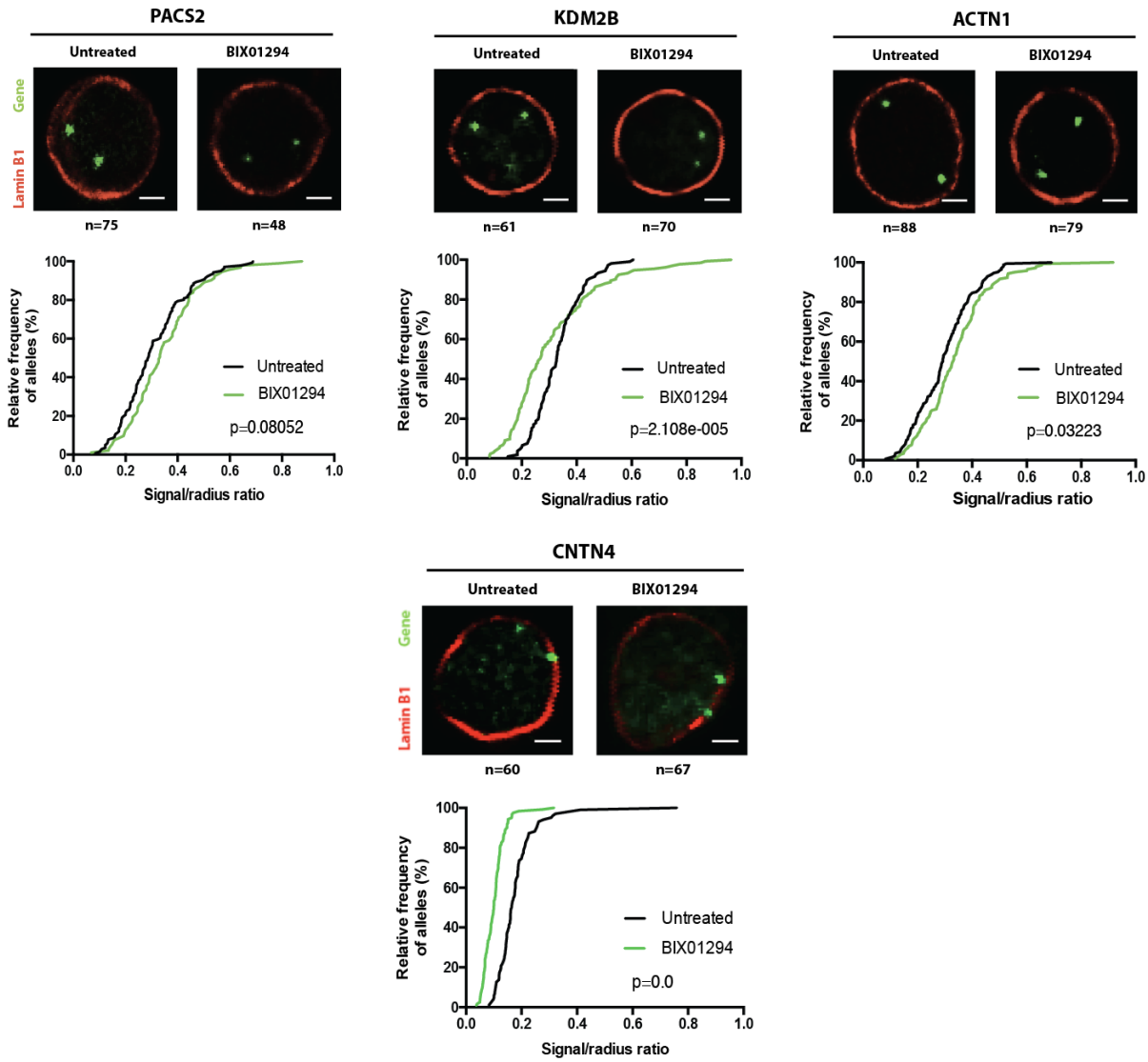


Figure 2.15. 3D nuclear position of HIV-1 non-target genes.

Numbers under the images indicate the total cell counts in which the distance of the gene FISH signals to the nuclear lamina was measured. Graphs below the images show the cumulative distribution of alleles by the Kolmogorov-Smirnov test in relation to their relative distance to the nuclear lamina, with the corresponding p value above the x-axis of the graph (scale bars: 2 µm).

In summary, out of 11 analyzed RIGs, 9 genes change their nuclear position and move more towards the nuclear center, following the example of HIV-1 upon BIX01294 treatment and H3K9me2 depletion. Out of 4 analyzed non-HIV-1 target genes, 2 do not change their nuclear position, but 2 are slightly moved more towards the nuclear periphery (CNTN4), or more towards

the nuclear center (ACTN1). In total, out of 15 analyzed genes 12 change their nuclear position upon BIX01294 treatment and H3K9me2 depletion, following the behavior of HIV-1 provirus (Figure 2.16).

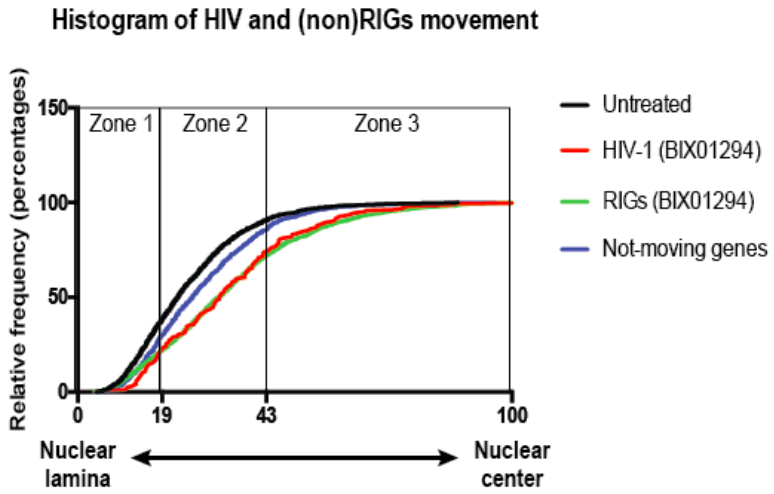


Figure 2.16. Summarized representation of 3D nuclear position of HIV-1, RIGs and non-RIGs in untreated and BIX01294 treated conditions.

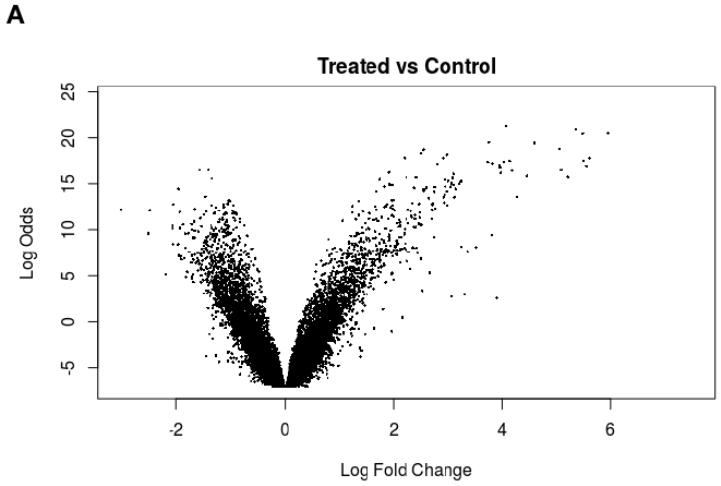
Graph shows the cumulative distribution of all analyzed alleles and vDNA signals by the Kolmogorov-Smirnov test in relation to their relative distance to the nuclear lamina. The black line on the histogram indicates the position of HIV-1 and RIGs in untreated conditions. The red line shows the repositioning of HIV-1 upon BIX01294 treatment. The green line shows the repositioning of RIGs upon BIX01294 treatment. The blue line shows the nuclear position of not moving RIGs and non-RIGs in the nucleus upon BIX01294 treatment.

II.VII. Changed nuclear position does not affect transcriptional activity of HIV-1 RIGs

Given that BIX01294 treatment causes a reduction of facultative heterochromatin mark H3K9me2 and that it also has an effect on gene positioning, we reasoned that there might be also a change of global gene expression levels.

In order to assess potential changes in gene expression levels, RNA microarray analysis, of untreated and BIX01294 treated primary $CD4^+$ T cells 6-7 hours post treatment was performed, via BeadArray microarray Technology (Illumina) by using the HumanHT-12 Expression

BeadChip kit. Interestingly, RNA expression profiles of most genes were not affected by the treatment. Still, we observed that certain portion of genes were either downregulated or upregulated (Figure 2.17. A). In addition, we performed the Gene Ontology (GO) term analysis and among the downregulated genes we find genes involved in membrane trafficking, protein and ion binding. In the group of upregulated genes, genes involved in iron metabolism and inflammatory response are dominating (Figure 2.17. B). It is known that G9a function is pivotal in cell proliferation and differentiation (Lehnertz et al, 2010; Rao et al, 2016). In addition, it has been shown that G9a impairment by BIX01294 treatment can reduce cellular proliferation rates or cause cell cycle arrest (Shankar et al, 2013). On the other hand, iron is vital for T cell proliferation and is involved in pathways protecting cells from cytotoxicity, oxidative stress and DNA damage (Bowlus, 2003; Ruttkay-Nedecky et al, 2013). Thus, it might be plausible that there is an ongoing compensatory effect, where iron and inflammatory pathways are activated in order to prevent cell cycle arrest.



B

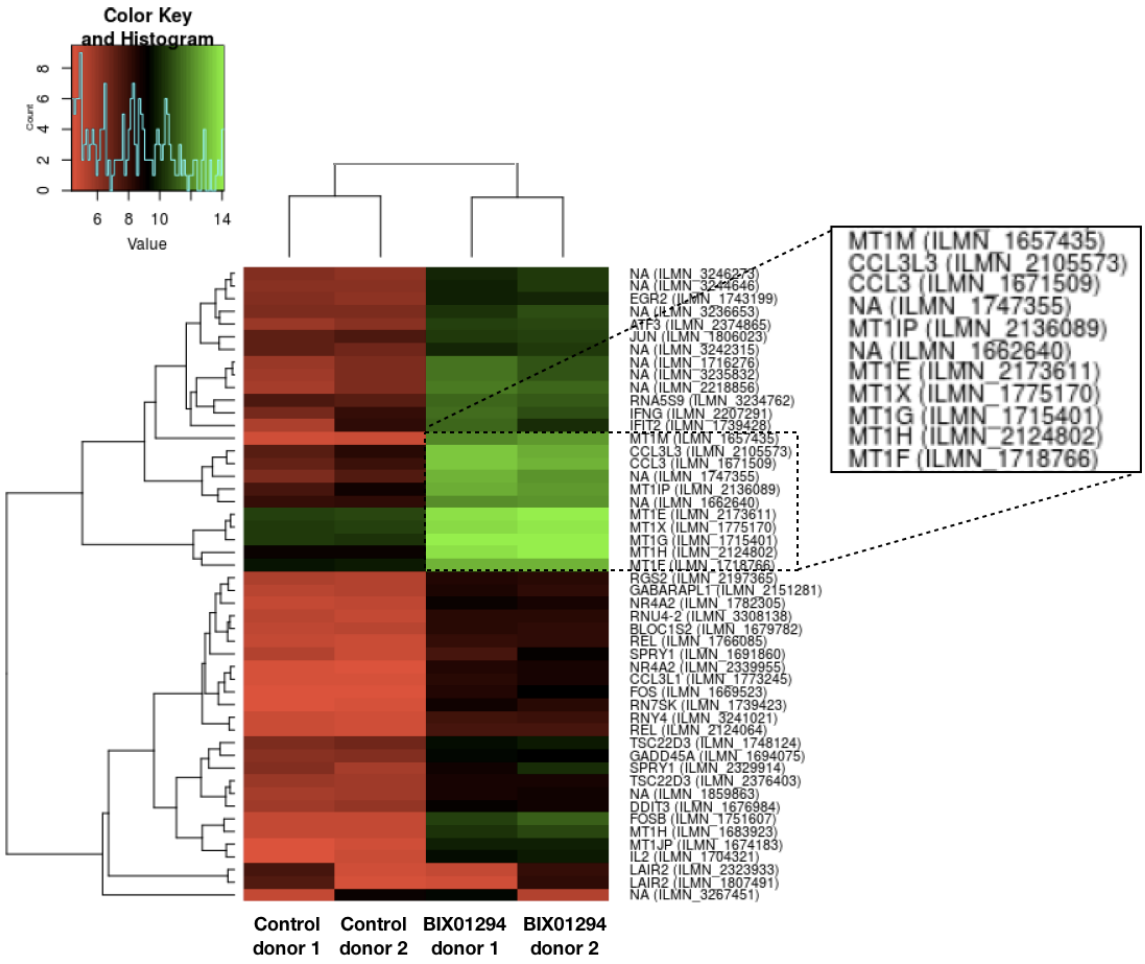


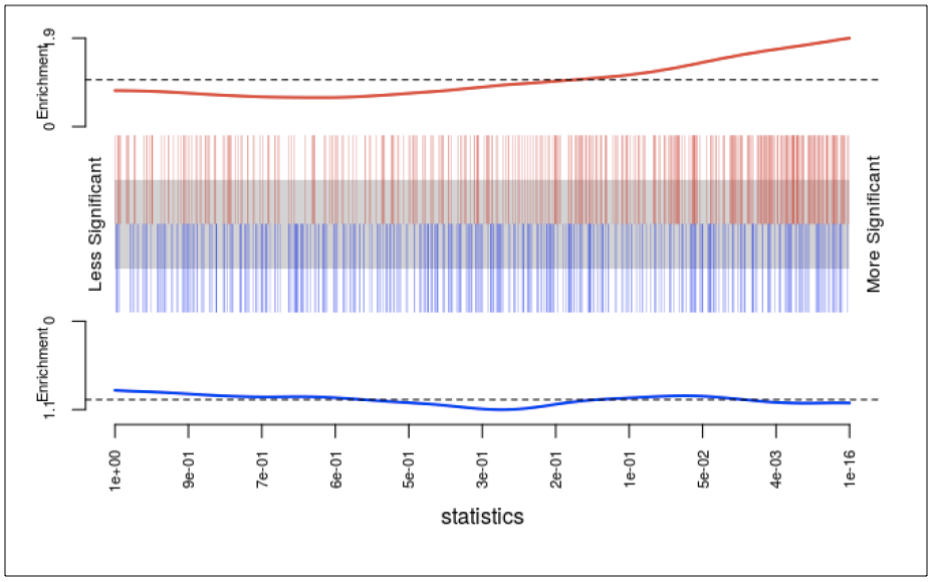
Figure 2.17. RNA microarray results of global gene expression profiles upon BIX01294 treatment.

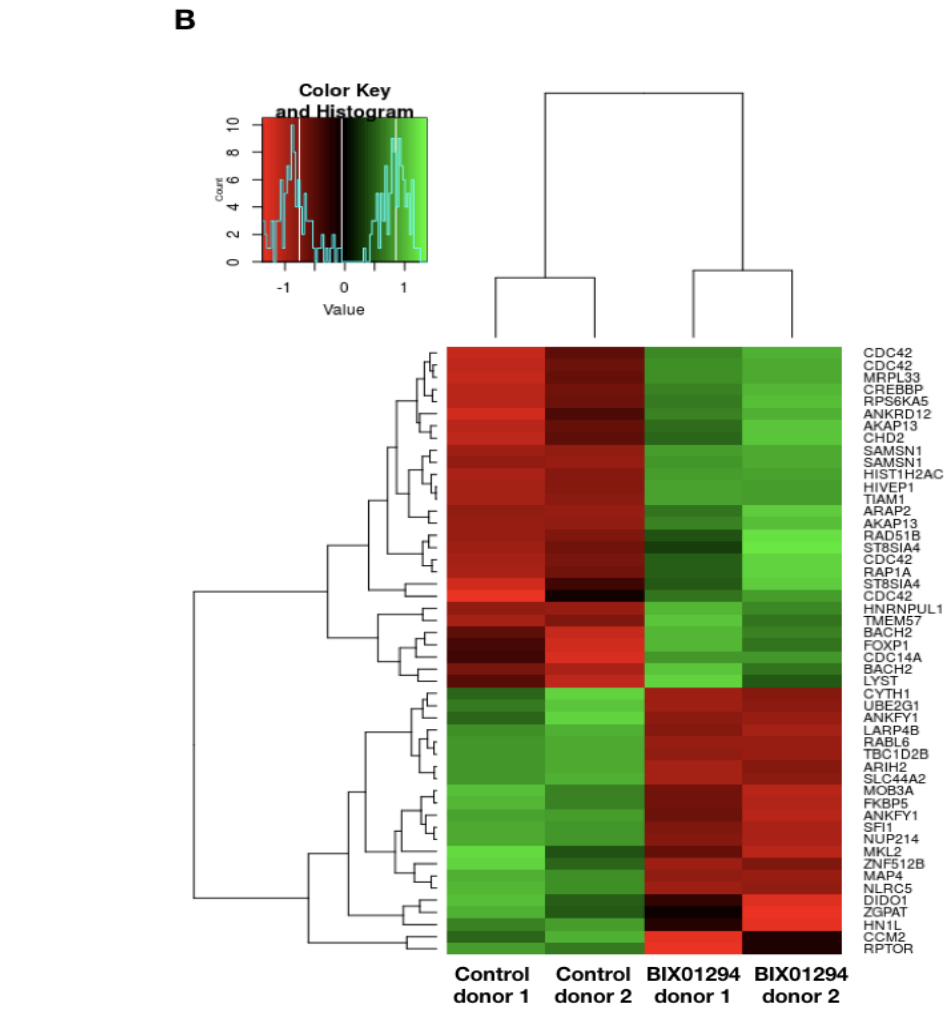
A. Volcano plot shows the global change in gene expression profiles upon BIX01294 treatment. The plot shows a clear diversification between downregulated genes (genes on the left of the x-axis, showing an expressional log fold change bellow the value 0) and upregulated genes (genes on the right of the x-axis, showing an expressional log fold change above the value 0) upon drug treatment, summarized for 2 independent donors (data analysis performed by Constantin Ahlmann-Eltze). **B.** Heatmap shows the top 50 genes with changed expression profiles upon BIX01294 treatment. The color code in the upper left corner indicates log fold changes of gene expression levels. The top upregulated genes are indicated on the box on the right. The top upregulated genes are members of metalloproteases and chemokine receptors involved in iron metabolism and inflammatory response (data analysis performed by Constantin Ahlmann-Eltze).

We then analyzed the expression profiles of HIV-1 targeted genes (RIGs) and observed again that some of them were downregulated whereas others were upregulated (Figure 2.18. A, B and C).

We compared f RIGs to a random set of 424 genes (Figure 2.18. panel A; blue lines) a clear shift towards lower p-values could be noticed, suggesting that overall, RIGs change their expression profiles in BIX01294 treatment (Figure 2.18. A), although this seems to be a rather mild change. The only significant change could be appreciated for BACH2, with a log fold change of 2, which corresponds to the same result we have observed by RT-qPCR (Figure 2.18. C).

A





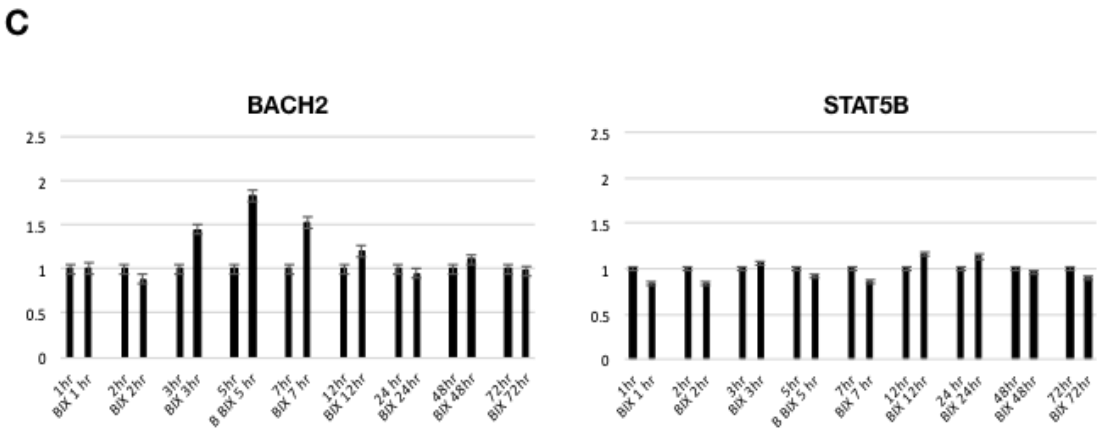


Figure 2.18. RIGs expression profiles upon BIX01294 treatment.

A. The barcode plot shows, in red, the distribution of p-values for each RIG expression change. A maximum p value of 7e-14 is detected (generated by using the geneSetTest from limma in R studio) (data analysis

performed by Constantin Ahlmann-Eltze). **B.** The heatmap shows the top 50 RIGs with changed expression profiles upon BIX01294 treatment, with BACH2 being the most significantly upregulated RIG. The color code in the upper left corner indicates log fold changes of gene expression levels (data analysis performed by Constantin Ahlmann-Eltze). **C.** Bar plots show RT-qPCR results of mRNA expression levels of BACH2 and STAT5B at different time points post BIX01294 treatment. The values in untreated conditions for each time point are set to 1 and used as a reference to assess mRNA expression changes upon BIX01294 treatment at each analyzed time point. Gene mRNA expression is normalized over expression profiles of housekeeping gene 18.

In conclusion, the obtained results show that BIX01294 treatment affects gene expression profiles. However, RIGs expression was not dramatically impacted by the treatment, whereas groups of genes involved in iron metabolism and inflammatory response showed clear overexpression. This result might also indicate that increased HIV-1 integration upon BIX01294 treatment is not due to increased RIGs expression, but rather due to a more loose and accessible chromatin composition.

Cell cycle profiles upon BIX01294 treatment

Considering that BIX01294 treatment causes a global reduction in H3K9me2 chromatin mark, opening of chromatin structure and change in gene expression levels, testing potential changes in cell cycle profiles of untreated activated primary $CD4^+$ T cells and BIX01294 treated cells was also performed.

The cell cycle analysis was performed by propidium iodide staining of cells, followed by FACS analysis 7 h post BIX01294 treatment. Upon BIX01294 treatment, no significant change in the cell cycle profile was observed. Cells were predominantly in G1 phase (on average 67,32% of untreated cells; 72,75% of BIX01294 treated cells), a portion was also found in G2 (on average 25,54 % of untreated cells; 22,11 % of BIX01294 treated cells) phase and late G2 or M phase (on average 5,06% of untreated cells; 314% of BIX01294 treated cells) (Figure 2.19. A and B).

In untreated conditions 2,08% of cells were not gated as live, whereas in BIX01294 treated conditions 2% of cells were not gated as live.

Obtained results indicate that BIX01294 treatment does not cause any changes in cell cycle profiles of primary $CD4^+$ T cells. Such a scenario has also previously been described in the case

of G9a KD, where cell cycle profiles and cell proliferation were not affected after G9a KD (Yokochia et al, 2009).

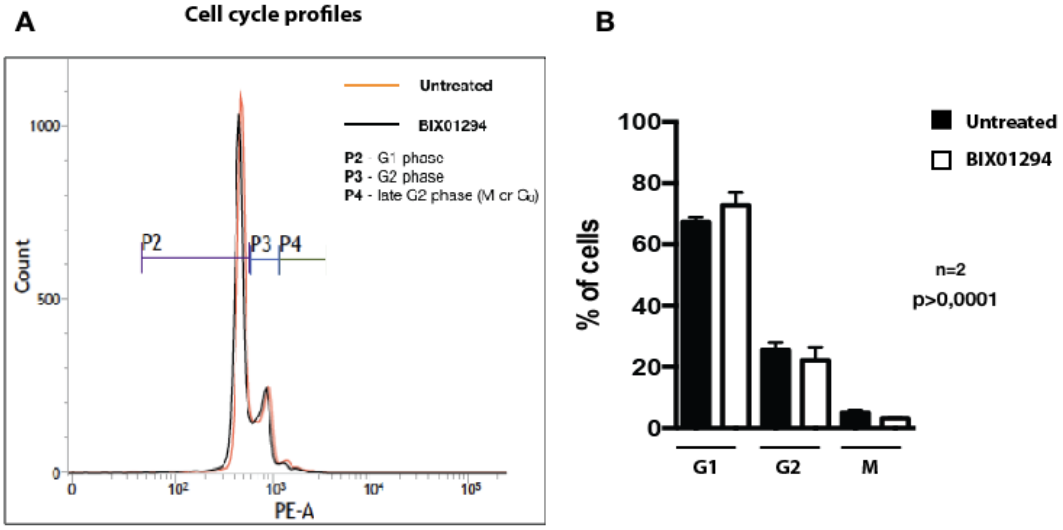


Figure 2.19. Cell cycle profiles upon BIX01294 treatment.

A. Histogram of cell FACS profiles upon BIX01294 treatment (orange line) compared to untreated conditions (black line). Results represent an output of the BD Biosciences FACS analysis software where peaks occurring before 10^3 PE-A density represent cells in G1 phase of the cell cycle, peaks occurring at 10^3 PE-A density represent cells in G2 phase of the cell cycle and peaks occurring after 10^3 PE-A density represent cells in the late G2 or M phase of the cell cycle. **B.** The graph shows a summary of total amounts of cells in G1, G2 and M phase of the cell cycle expressed in percentages for 2 independent donors.

II. VIII. Chromatin composition of HIV-1 target genes upon H3K9me2 depletion

Assessment of RIGs chromatin composition by ChIP-qPCR

BIX01294 treatment and H3K9me2 depletion did not cause a significant change in transcription profiles of HIV-1 target genes, but significantly affected their position in $CD4^+$ T cell nuclei. We therefore sought to understand how H3K9me2 depletion affects the surrounding chromatin composition and LEDGF/p75 binding at selected RIGs.

ChIP-qPCR (Chromatin Immunoprecipitation followed by quantitative PCR) was performed for selected chromatin marks (H3K9me2, H3K9me3, H3K36me3), as well as for host cell factor LEDGF/p75. In total, 9 genes were analyzed by ChIP-qPCR, 5 of which were RIGs (BACH2, STAT5B, MKL2, RPTOR, NPLOC4) and 4 were non-RIGs (KDM2B, PACS2, ACTN1, CNTN4). For all the genes, primers for the first intron region, gene body and exon/intron junction at the gene end were designed in order to follow possible changes along selected genes. H3K9me2 chromatin mark served as a readout for efficient BIX01294 treatment, and was expected to decrease in regions where it otherwise binds. H3K9me3 served as a control for the treatment, since the binding of this chromatin mark should not be significantly affected by BIX01294 treatment. H3K36me3 was probed in ChIP as it is expected that this chromatin mark is deposited on HIV-1 target genes (Marini et al, 2015; Lucic et al, 2019), and might also be affected by BIX01294 treatment. Host cell factor LEDGF/p75 was included into the analysis considering its important role in HIV-1 integration site selection (De Rijck et al, 2010; Craigie & Bushman, 2012; Achuthan et al, 2018) as well as its know feature to act as H3K36me3 chromatin mark reader (Venkatesh et al, 2012; Pradeepa et al, 2012 Demeulemester et al, 2015, Singh et al, 2015).

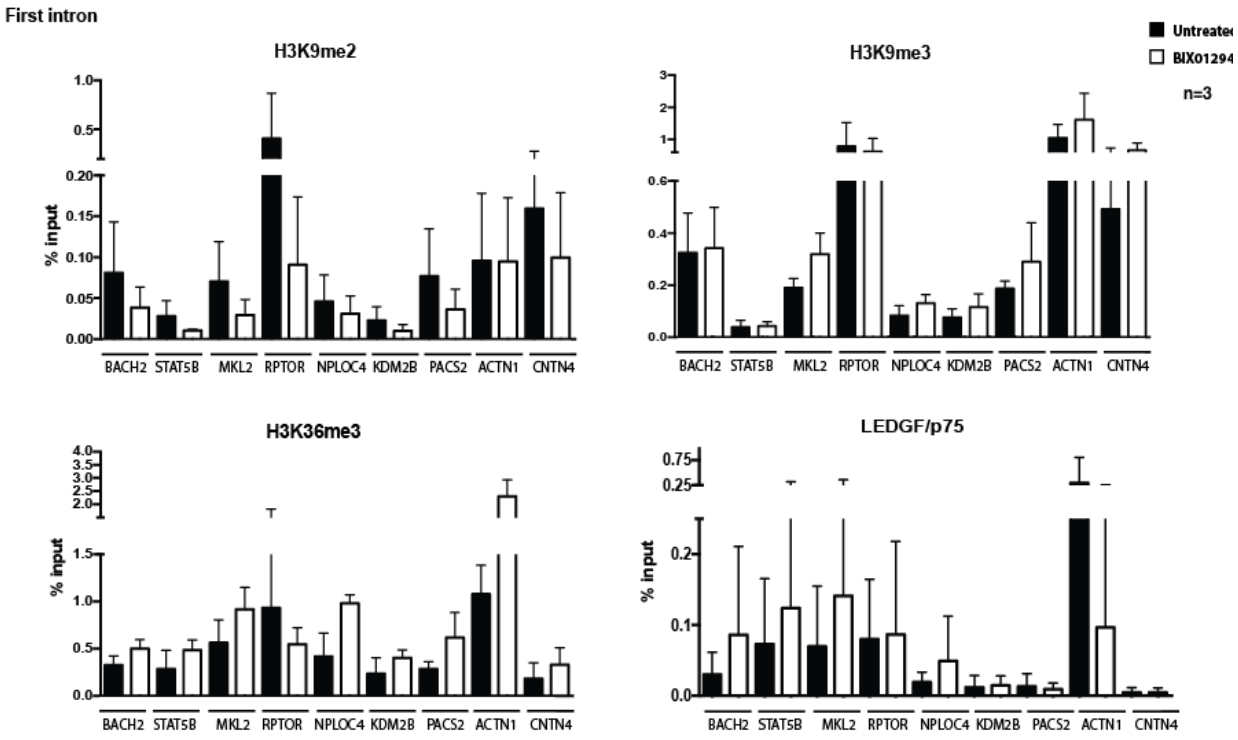


Figure 2.20. ChIP-qPCR results of the first intron region of selected genes.

Distribution of H3K9me2, H3K9me3 and H3K36me3 chromatin marks, as well as of LEDGF/p75 in the first intron region of selected RIGs and non-RIGs in untreated conditions and upon BIX01294 treatment. The graphs represent summarized results for 3 independent donors, normalized over input DNA levels. The legend in the upper right corner indicates which bars of the graphs correspond to untreated and BIX01294 treated samples.

H3K9me2 distribution in untreated and BIX01294 treated conditions showed an overall decrease in H3K9me2 levels upon BIX01294 treatment. This trend could be observed in the first intron region in almost all of the analyzed genes (except for ACTN1) (Figure 2.20.), as well as in gene bodies and gene ends (Figure 2.21. and 2.22.). When the enrichment of H3K9me3 was analyzed, as expected, and as previously seen by immunoblot and IF (Figure 2.4. and 2.6.), the levels of this chromatin mark mainly remained unchanged upon BIX01294 treatment. H3K9me3 levels do not change in the first intron region and gene ends of analyzed genes (Figure 2.20. and 2.22.). However, a slight decrease of H3K9me3 binding can be appreciated in the gene bodies of analyzed genes (Figure 2.21.), which is not surprising considering the very strong effect of BIX01294 on related H3K9me2 chromatin mark and the high sensitivity of the used antibodies (Rothbart et al, 2015).

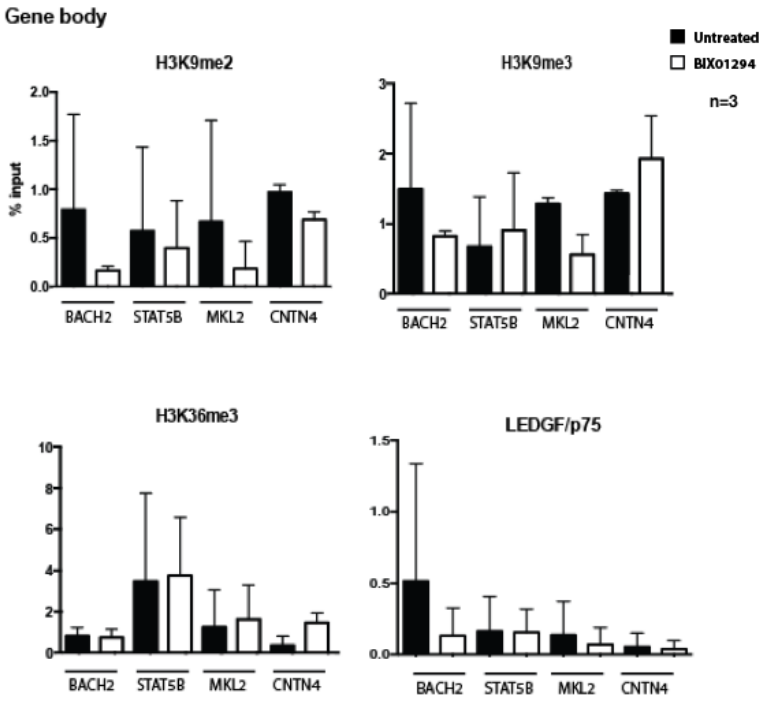


Figure 2.21. ChIP-qPCR of the gene body region of selected genes. Same as in Figure 2.20., assays in PCR with primers designed for the gene body (coding regions) of the respective genes.

H3K36me3 distribution showed an overall increase in the first intron region of all analyzed genes upon BIX01294 treatment (Figure 2.20.), whereas it's levels in gene bodies seemed not affected by the treatment. Interestingly, in the gene ends of MKL2 and CNTN4 an increased binding of H3K36me3 could be detected. As a mark of elongating Pol II transcription, H3K36me3 is usually found in gene bodies and transcribed

genomic regions (Venkatesh et al, 2012), but it can also be linked to alternative splicing, dosage compensation, DNA replication and repair (Wagner & Carpenter, 2012; Suzuki et al, 2016). We did not observe an increased deposition of H3K36me3 in gene bodies of analyzed genes (Figure 2.21.) upon BIX01294 treatment, consistent with the unchanged expression profiles of RIGs upon H3K9me2 depletion and BIX01294 treatment. Interestingly, increased deposition of H3K36me3 was observed in gene ends of MKL2 and CNTN4, and might possibly be associated to some alternative implications of this chromatin mark.

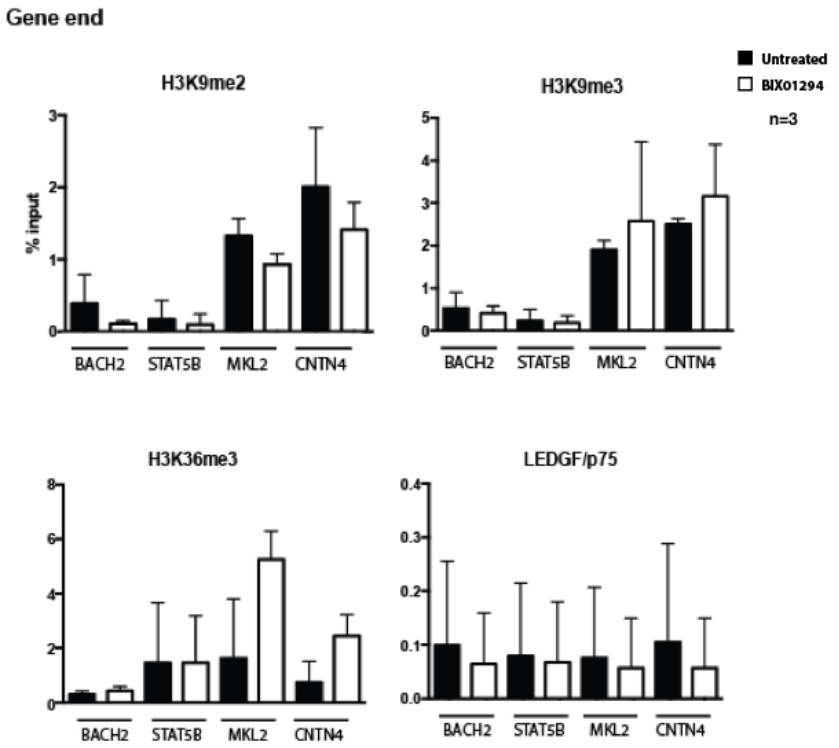


Figure 2.22. ChIP-qPCR results of the gene end region of selected genes. Same as in Figure 2.20., assays in PCR with primers designed for the gene end (exon/intron junction) of the respective genes.

Analyzing ChIP-qPCR profiles of LEDGF/p75, which is also a known H3K36me3 chromatin reader (Pradeepa et al, 2012), revealed that LEDGF/p75 binding was enriched only in the first intron of RIGs upon BIX01294 treatment (Figure 2.20.), whereas it remained unchanged on non-RIGs, similarly to the observed H3K36me3 distribution. When looking into gene bodies and gene ends, levels of LEDGF/p75 did not change upon BIX01294 treatment (Figure 2.21. and 2.22.). Only in the gene bodies of BACH2 and MKL2 a certain reduction of LEDGF/p75 binding could be observed, which might be justified by almost 50% more binding in the first intron region of these genes and the repositioning of LEDGF/p75 binding along these genes. In addition, the observed enrichment of H3K36me3 and LEDGF/p75 in the first intron regions of RIGs could be associated to the observed increased integration of HIV-1 upon BIX0124 treatment and the fact

that HIV-1 preferentially integrates into introns of expressed genes (Craigie & Bushman, 2012; Han et al, 2014; Cohn et al, 2015).

Global H3K36me3 and LEDGF/p75 protein levels remain unchanged upon H3K9me2 depletion

Given that H3K36me3 and LEDGF/p75 showed some differences in their deposition profiles along analyzed genes upon BIX01294 treatment, we assessed their overall levels in BIX01294 treatment. Immunoblot analysis showed that global levels of H3K36me3 remain unchanged upon BIX01294 treatment (2.23. A and B). This result indicates that, while the global levels of H3K36me3 do not change, some probably local changes in its distribution occur upon BIX01294 treatment (Figure 2.20., 2.21. and 2.22).

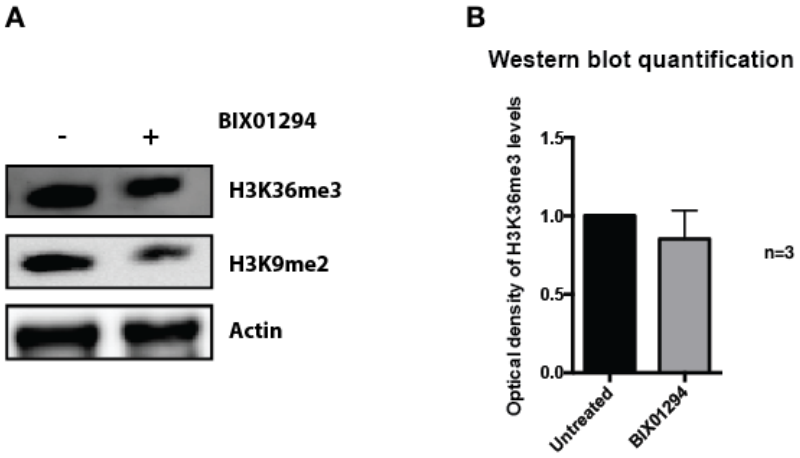


Figure 2.23. Quantification of H3K36me3 levels upon BIX01294 treatment.

A. Immunoblot assessment of H3K36me3 levels upon BIX01294 treatment. As a readout for the treatment efficiency, H3K9me2 levels were probed. ß-Actin served as a loading control **B.** Optical density quantification of H3K36me3 chromatin mark of western blot results of 3 independent donors upon BIX01294 treatment. Optical density values of H3K36me3 are normalized over optical density values of ß-Actin loading control and in untreated conditions set to 1 and used as a reference to assess optical density changes in H3K36me3 chromatin mark upon BIX01294 treatment.

Global levels of LEDGF/p75 were assessed by western blot and IF staining in two additional conditions; untreated, BIX01294 treated, HIV-1 infected only and BIX01294 treated and HIV-1 infected. Levels of LEDGF/p75 remain unchanged upon BIX01294 treatment, HIV-1 infection or under both, treated and infected conditions, as seen and quantified by western blot as well as quantified by microscopy and total immunofluorescence intensity (2.24. A, B, C and D).

Similarly, as in the case of H3K36me3, this result indicates that only the binding profile of LEDGF/p75 is affected upon BIX01294 treatment and H3K9me2 depletion, whereas total protein level remain unchanged.

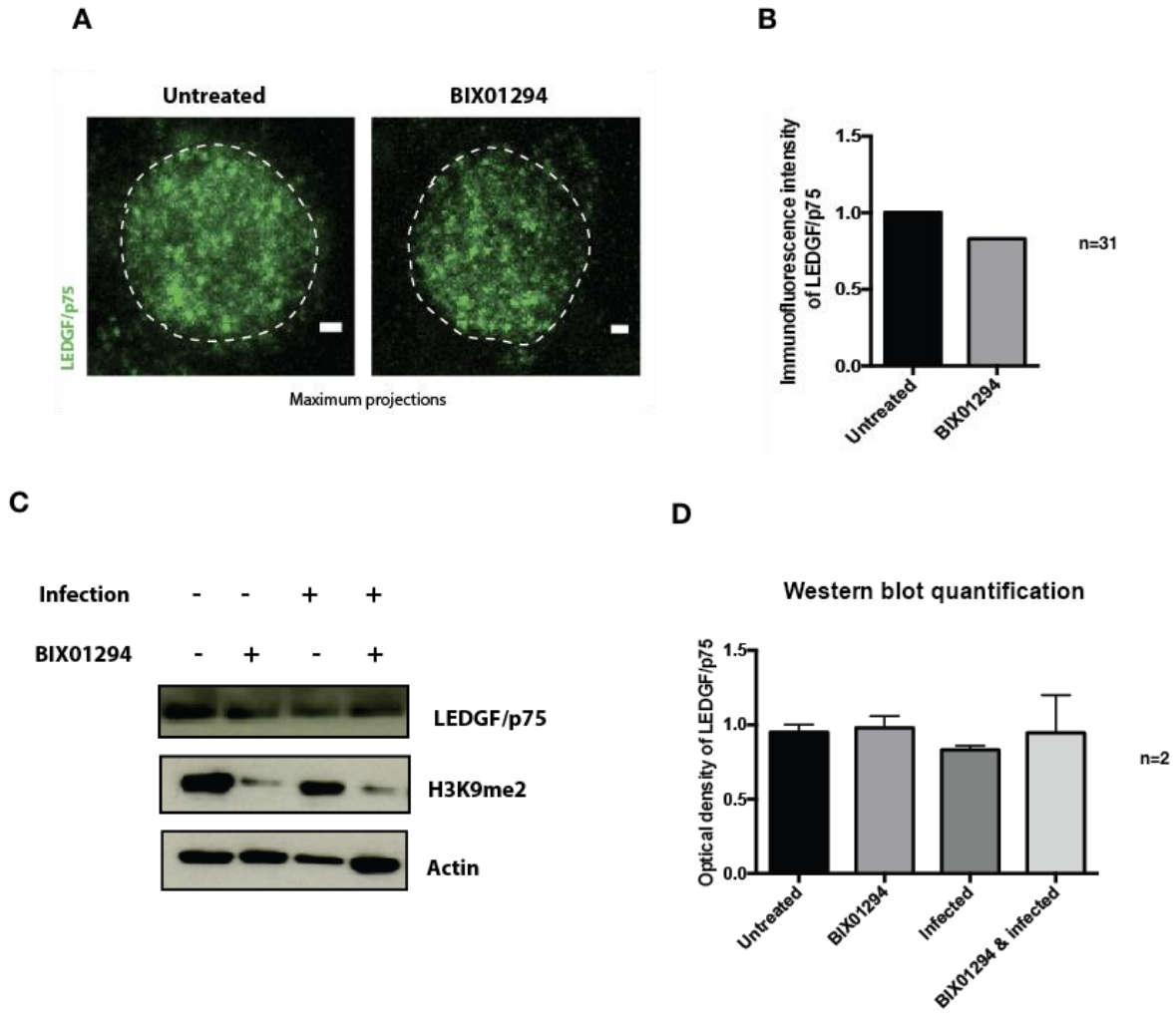


Figure 2.24. Quantification of total LEDGF/p75 levels upon BIX01294 treatment.

A. IF staining results of LEDGF/p75 upon BIX01294 treatment. The images show maximum projection of taken Z-stacks. The dashed line indicates the nuclear rim, and was marked in relation to Hoechst nuclear staining (scale bars: 1 µm). **B.** Total IF intensity quantification of LEDGF/p75 upon BIX01294 treatment (31 cells in total). IF intensity values in untreated conditions are set to 1 and used as a reference to assess IF intensity changes upon BIX01294 treatment. **C.** Western blot results show the assessment of LEDGF/p75 levels upon BIX01294 treatment, HIV-1 infection and BIX01294 treatment and infection. The efficiency of BIX01294 treatment is demonstrated by the assessment of H3K9me2 chromatin mark levels. As protein loading controls total levels of ß-Actin are presented. **D.** Optical density quantification of LEDGF/p75 levels of western blot results of 2 independent donors upon BIX01294 treatment and HIV-1 infection. Optical density values of LEDGF/p75 are normalized over optical density values of the ß-Actin loading controls and in untreated conditions set to 1 and used as a reference to assess optical density changes in the analyzed conditions.

II. IX. Determining HIV-1 integration sites upon H3K9me2 depletion

Finally, we wanted to address HIV-1 integration profiles by directly sequencing the integration sites. As BIX01294 treatment and H3K9me2 depletion resulted in alterations of spatial distribution of HIV-1 in the nuclear space (Figure 2.12. B), we wanted to test if HIV-1 integration profiles changed as well. Therefore, we sought to directly sequence HIV-1 integration upon drug treatment.

HIV-1 integration sites have been sequenced in untreated and BIX01294 treated conditions by LAM-PCR (Schmidt et al, 2007). In addition, a comparison of HIV-1 integration profiles in both conditions has been conducted, as well as a comparison of our dataset to other HIV-1 integration sites lists (Ikeda et al, 2007; Brady et al, 2009; Wagner et al, 2014; Maldarelli et al, 2014; Kok et al, 2016) (Figure 2.25.). This was done in collaboration with the group of Manfred Schmidt (previously NCT, now Genewerk) and the data analysis was performed by Dr. Raffaele Fronza.

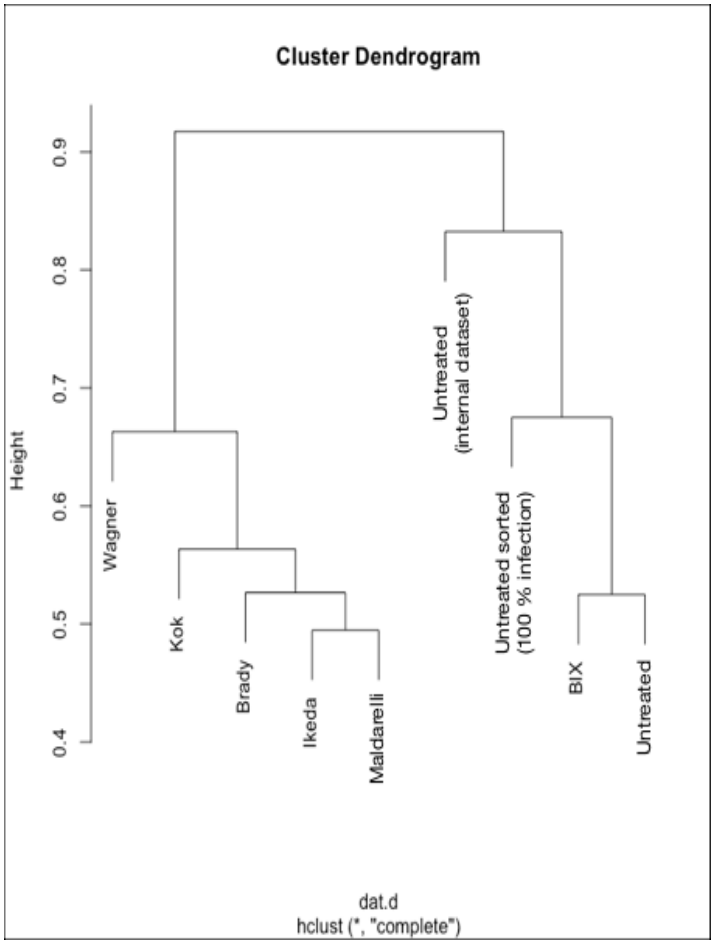


Figure 2.25. Cluster dendrogram of HIV-1 integration sites from untreated and BIX1294 treated conditions compared to other HIV-1 integration site lists.

The obtained cluster dendrogram shows that HIV-1 integration sites obtained upon BIX01294 treatment produce a distribution pattern similar to integration sites obtained in untreated conditions, and similar to integration sites from other studies (indicated on the height scale of the dendrogram) (data analysis performed by Dr. Raffaele Fronza).

The obtained results show that HIV-1 integration patterns in untreated and BIX01294 treated conditions cluster together, and are clearly separated from other HIV-1 integration site datasets. This indicates that a high degree of similarity between the two conditions we analyzed is probably due to the sequencing method used.

The obtained result shows that even with a small dataset (in total 910 integration sites; 413 in untreated conditions and 497 in BIX01294 treated conditions from 2 donors) an increase of 1,2 fold more integration sites upon BIX01294 treatment can be appreciated (Table 2.1.).

Table 2.1. Summary statistics of retrieved HIV-1 integration sites in untreated and BIX01294 treated conditions.

Total No of IS	**910**
Untreated	413
BIX01294	497
Chromosomes with multiple integrations	**22**
Untreated	9
BIX01294	13
Genes with multiple integrations	**34**
Untreated	12
BIX01294	22

However, the increase in HIV-1 integration upon BIX01294 treatment observed by Alu PCR, which was always in the range of a 2,8-5 fold increase (Figure 2.2), does not correspond to the increase observed by integration site sequencing. This discrepancy might be explained by underestimating capacities of integration site sequencing approaches (Cui et al, 2016; Hughes & Coffin, 2017).

In addition, the results indicate that H3K9me2 depletion does not cause a change in HIV-1 integration site selection, but rather causes an increase in integration into the proximal genomic regions due to increased chromatin accessibility. This possibility will be tested in the future experiments, where the same donors in which HIV-1 integration sites were sequenced, will be tested for chromatin profiles of H3K9me2m H3K36me3 and LEDGF/p75 under BIX01294 treated conditions.

Our results so far show that chromatin composition, in particular that related to H3K9me2 deposition, has an effect on HIV-1 integration levels. The obtained results indicate that a global increase in chromatin accessibility induced by G9a inhibitor BIX01294 increases HIV-1 integration and changes the localization of the viral genome in the 3D nuclear space.

In addition, the results show that the drug treatment specifically targets H3K9me2 chromatin mark, without affecting related H3K9me3 chromatin mark. It can also be excluded that the observed increase in HIV-1 integration occurs due to any DNA damage, changes in cell cycle profiles or activation status of the cells caused by drug treatment, indicating that the observed increase in HIV-1 integration is related to increased chromatin accessibility.

Analyzed gene expression profiles upon BIX01294 treatment show that RIGs expression levels are not significantly impacted by the drug treatment and that potentially increased HIV-1 target gene expression profiles can not be correlated to the observed increase in HIV-1 integration upon BIX01294 treatment. On the other hand, gene expression profiles of genes involved in iron metabolism, inflammatory response or membrane trafficking were significantly affected by the drug treatment.

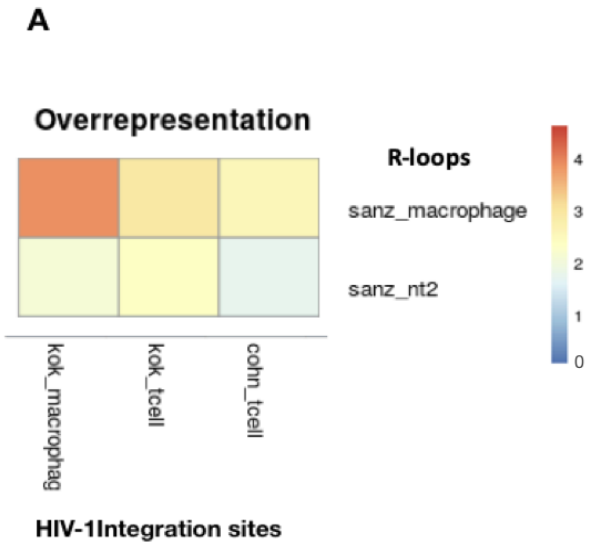
Finally, the detailed analysis of chromatin composition of RIGs and non-RIGs upon BIX01294 treatment and H3K9me2 depletion revealed increased binding of H3K36me3 chromatin mark in the first intron region of analyzed genes, which was followed by increased binding of host cell factor LEDGF/p75 only on RIGs. This results might give a certain explanation for the observed increase in HIV-1 integration upon BIX01294 treatment, which might be connected to a global increase in chromatin accessibility.

The observed increase in HIV-1 integration could additionally be observed by the analysis of viral integration sites upon BIX01294 treatment and H3K9me2 depletion. The integration site data demonstrated no change in the distribution of HIV-1 integration sites upon H3K9me2 depletion, indicating that still the same genomic sites are targeted for HIV-1 integration, but they become even more targeted due to a global increase in chromatin accessibility upon BIX01294 treatment. This finding also indicated that the observed nuclear re-distribution of the proviral genome upon BIX01294 treatment, as seen by FISH analysis, could not be connected to a redistribution in viral integration sites.

In the following results section, a more viral perspective on HIV-1 integration will be given, integrating the interplay of viral IN and its catalytic and chromatin binding properties with R-loops as special genomic features.

II. X. R-loops show strong association with HIV-1 RIGs

Our results so far show that chromatin composition related to H3K9me2 levels has a significant impact on HIV-1 integration levels. The global increase in chromatin accessibility induced by G9a inhibitor BIX01294 increases significantly HIV-1 integration and changes the localization of the viral genome in the 3D nuclear space. In order to understand if there are other genomic features related to H3K9me2 chromatin mark and to viral integration, we performed a computational analysis in which we analyzed the presence of DNA-RNA hybrids or R-loops on HIV-1 integration sites (genes). For the analysis, we used HIV-1 integration sites from two studies (Cohn et al, 2015; Kok et al, 2016) reporting patient derived integration sites from the two major HIV-1 target cell types (T cells and macrophages). The R-loop profiles were derived from the study by the Chedin lab, where K562 myelogenous leukemia cell line and Nt2 human embryonic carcinoma cell line were assayed by DRIP-Seq (DNA-RNA Immunoprecipitation followed by genome wide sequencing (Sanz et al, 2016). By using a cut-off of at least 3 hybrids per gene, we observed that genes targeted by HIV-1 have an increased incidence of R-loop forming structures. (Figure 2.26. A). We then analyzed the presence of R-loops (as mapped in K562 myelogenous leukemia cell line by Sanz et al, 2016) in HIV-1 RIGs (Lucic et al, 2019) and observed that a striking 83% of HIV-1 RIGs have 3 or more R-loops (Figure 2.26. B). The same graph shows the control analysis of HIV-1 RIGs with the LAD regions, retrieved from human fibroblasts and Jurkat T cell line (Guelen et al, 2008; Robson et al, 2017), from which HIV-1 RIGs are normally excluded (Marini et al, 2015, Achuthan et al, 2018, Bativelli et al, 2018).



B

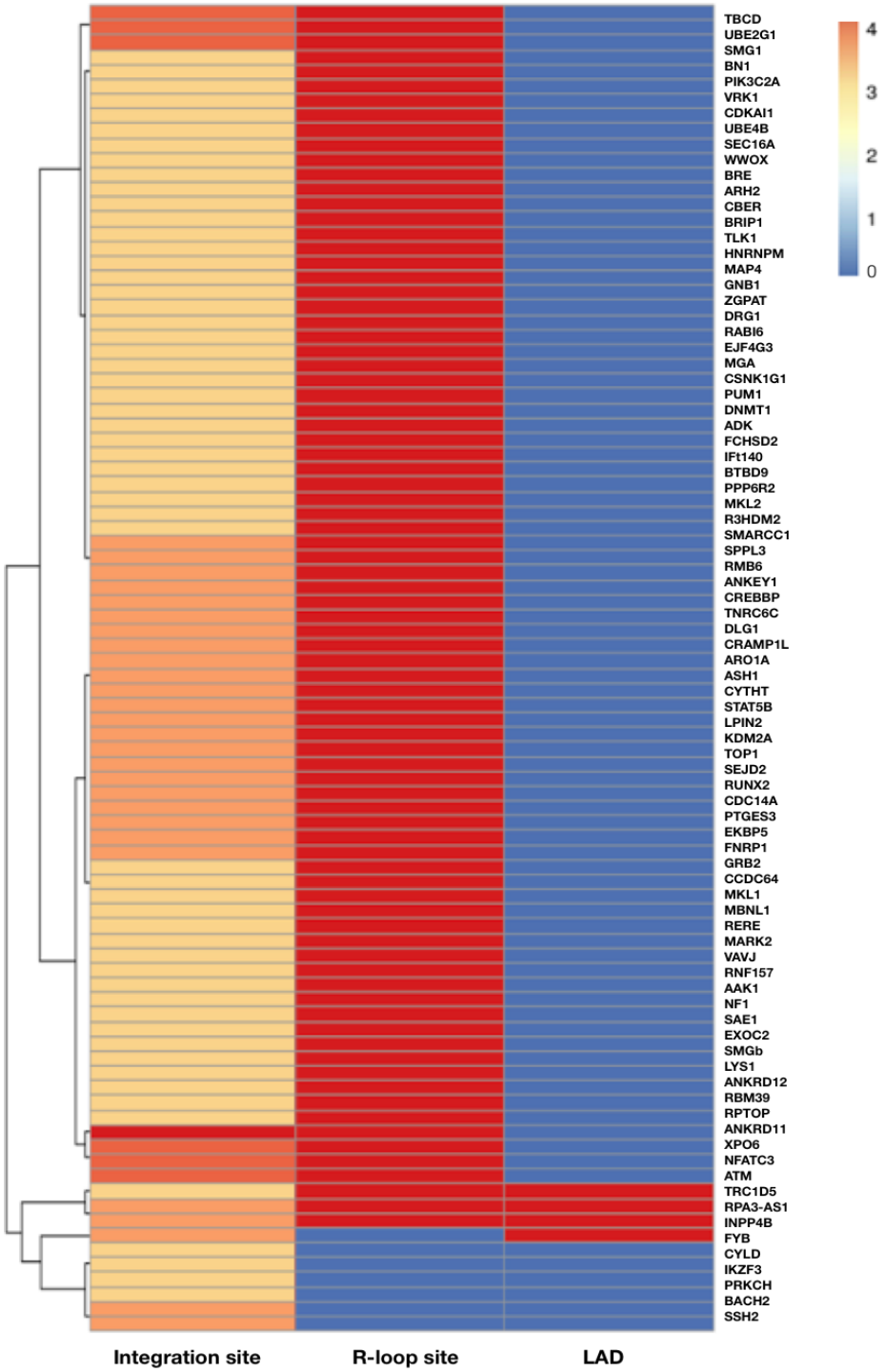

TBCD
UBE2G1
SMG1
BN1
PIK3C2A
VRK1
CDKAI1
UBE4B
SEC16A
WWOX
BRE
ARH2
CBER
BRIP1
TLK1
HNRNPM
MAP4
GNB1
ZGPAT
DRG1
RABI6
EJF4G3
MGA
CSNK1G1
PUM1
DNMT1
ADK
FCHSD2
IFt140
BTBD9
PPP6R2
MKL2
R3HDM2
SMARCC1
SPPL3
RMB6
ANKEY1
CREBBP
TNRC6C
DLG1
CRAMP1L
ARO1A
ASH1
CYTHT
STAT5B
LPIN2
KDM2A
TOP1
SEJD2
RUNX2
CDC14A
PTGES3
EKBP5
FNRP1
GRB2
CCDC64
MKL1
MBNL1
RERE
MARK2
VAVJ
RNF157
AAK1
NF1
SAE1
EXOC2
SMGb
LYS1
ANKRD12
RBM39
RPTOP
ANKRD11
XPO6
NFATC3
ATM
TRC1D5
RPA3-AS1
INPP4B
FYB
CYLD
IKZF3
PRKCH
BACH2
SSH2
4
3
2
1
0
Integration site
R-loop site
LAD

Figure 2.26. Computational overlap of HIV-1 integration site lists and R-loop maps.
A. Heatmap shows an overlap of HIV-1 integration sites (Cohn et al, 2015; Kok et al, 2016) and R-loop maps (Sanz et al, 2016). The color code on the right indicates levels the overrepresentation of R-loop forming sites in HIV-1 integration sites (data analysis performed by Constantin Ahlmann-Eltze). **B.** Heatmap shows an overlap of RIGs (listed in the upper part of the heatmap) and an R-loop map from K562 cells (Sanz et al, 2016), by including genomic LAD regions retrieved from human fibroblasts and Jurkat T cell line (Guelen et al, 2008; Robson et al, 2017), as outliers, into the analysis (data analysis performed by Maja Kuzman).

This *in silico* retrieved dataset represented the first indication that R-loop forming sites might be associated to HIV-1 integration sites.

II. XI. HIV-1 Preintegration Complexes (PICs) colocalize with R-loops

In order to gain the experimental support for our hypothesis that R-loops represent the necessary genomic feature for HIV-1 integration, we first set up to visualize at the cellular level a possible colocalization of HIV-1 PICs and R-loops in primary macrophages as well as primary $CD4^+$ T cells was performed.

Cells were infected with an engineered wt HIV-1 virus containing a GFP-tag on the viral IN protein, produced by the Vpr-trans incorporation strategy (Albanese et al, 2008). The cells were collected at early time points post infection (36 h in macrophages and 18 h in $CD4^+$ T cells) in order to capture the PICs before the integration step is completed. IF staining of R-loops was performed using the S9.6 antibody, followed by microscopy and PIC/R-loop colocalization analysis.

A striking colocalization pattern between R-loop structures and HIV-1 PICs could be observed at early time points post HIV-1 infection in primary human $CD4^+$ T cells and macrophages, in both confocal and super-resolution microscopy. In $CD4^+$ T cells 75% of HIV-1 PICs in the nucleus colocalized with R-loops (Figure 2.27. A). In macrophages a similar trend of 72% of colocalizing R-loop signals and HIV-1 PICs was observed (Figure 2.27. B).

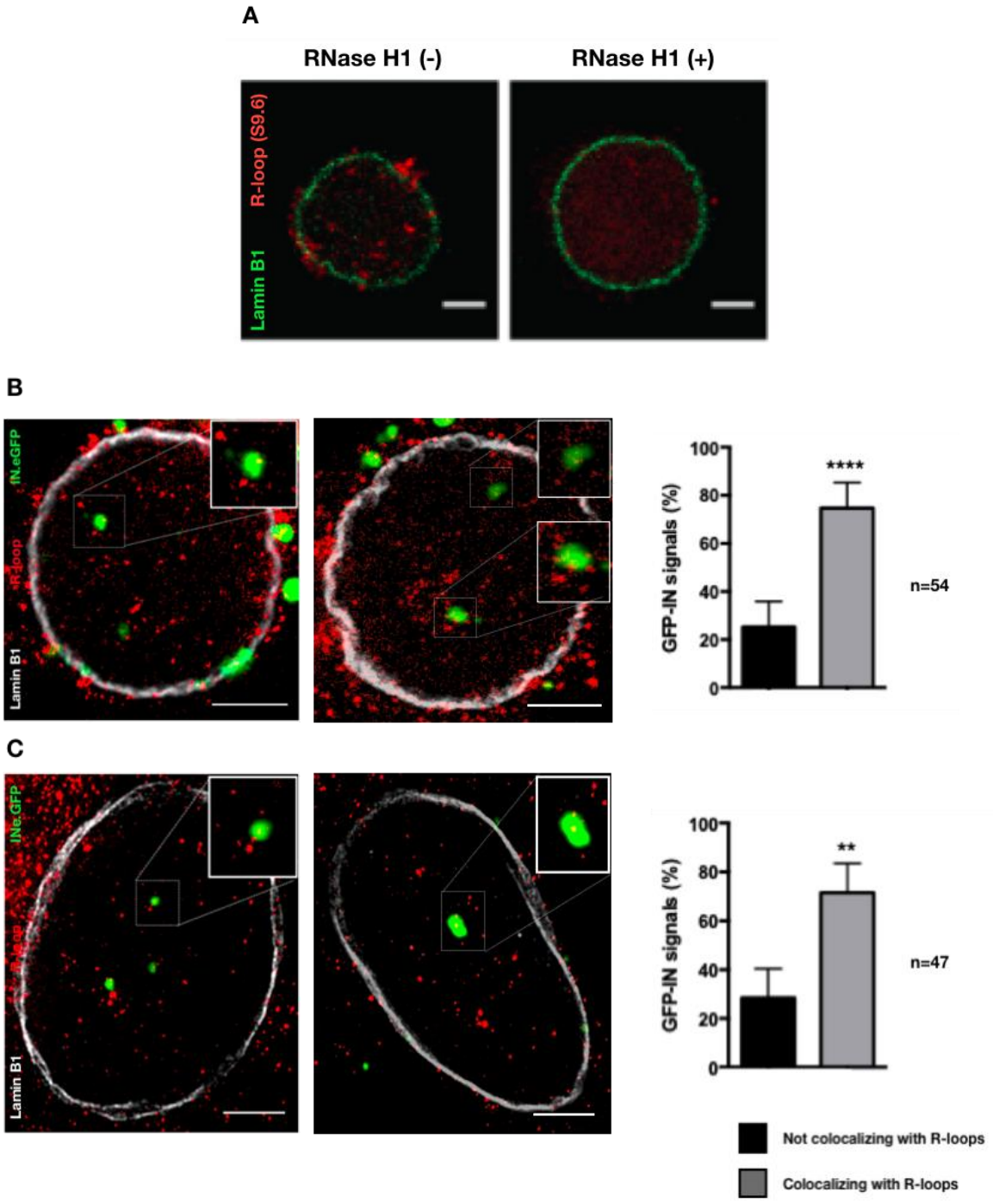


Figure 2.27. HIV-1 IN.eGFP colocalization with R-loops.

A. IF R-loop staining with S9.6 antibody in primary $CD4^+$ T cells upon RNase H1 pretreatment of fixed cells, followed by confocal microscopy. The cell on the left shows R-loop staining patterns in untreated conditions. The cell on the right shows the reduction of R-loop signals upon RNase H1 pretreatment of the cells, confirming the specificity of the S9.6 antibody for IF. The images show maximum projections of acquired z-stacks (staining

performed by the help of Dr. Bojana Lucic) (scale bars: 2 µm). **B.** Primary human $CD4^+$ T cells infected with HIV-1 INeGFP (18 h post infection) stained with S9.6 antibody - representative STED images, with the quantification of colocalizing PICs and R-loops, shown in the graph below (expressed in percentages) - 75% of PICs colocalize with R-loops. The number next to the graph indicates the number of analyzed cells in 6 independent donors. **C.** Macrophages infected with HIV-1 IN.eGFP (36 h post infection) stained with S9.6 antibody - a representative STED image, with the quantification of colocalizing PICs and R-loops, shown in the graph below (expressed in percentages) - 72% of PICs colocalize with R-loops. The number next to the graph indicates the number of analyzed cells in 4 independent donors (scale bars: 2 µm).

These obtained results could be confirmed by HIV-1 DNA FISH in primary macrophages at different time points post infection (36 h, 48 h & 60 h) (Figure 2.28. A and B).

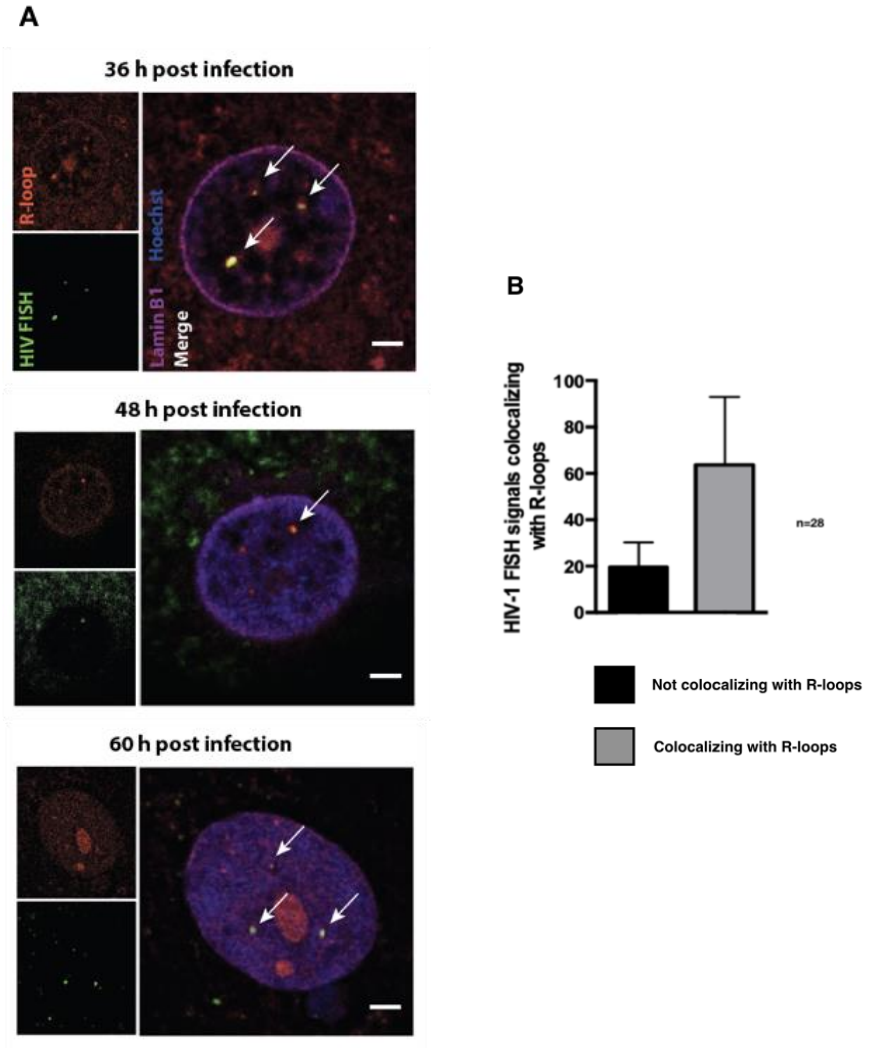


Figure 2.28. HIV-1 DNA immuno-FISH and R-loop staining in macrophages.

A. The images show viral DNA colocalizing with R-loop signals at different time points post infection assessed by IF staining of R-loops (S9.6 antibody) and performed HIV-1 DNA FISH. Macrophages were infected with HIV-1 IN.eGFP and collected at early time points post infection for DNA immuno-FISH, followed by confocal microscopy. The red channel represents the R-loop IF signal. The green channel represents the HIV-1 DNA FISH signal. In the merged view the blue channel represents the DNA-Hoechst staining. The white arrows indicate a colocalizing HIV-1 FISH and R-loop signal (scale bars: 2 µm). **B.** Graph shows the quantification of HIV-1 FISH signals colocalizing with R-loop signals summarized for all analyzed time points, expressed in percentages (total number of analyzed infected cells n=28).

Obtained results show that HIV-1 PICs colocalize with R-loops, thus suggesting that HIV-1 PICs by docking to R-loop structures, could integrate the viral genome into these genomic structures (as seen by HIV-1 DNA FISH).

II. XII. Functionally impaired HIV-1 IN contributes to less PIC and R-loop colocalization in macrophages

In order to verify that the initially observed colocalization between HIV-1 PICs and R-loops does not represent an artifact or a phenomenon occurring just by chance, the same set of experiments was performed on macrophages, as already described, but with PICs containing a functionally impaired IN (CCD mutant D116N). As an additional control, we used Raltegravir (Markowitz et al, 2007), which impairs HIV-1 integration by acting as an Integrase strand transfer inhibitor. All viruses used in this set of experiments were produced by the Vpr-trans incorporation strategy (Albanese et al, 2008) and contained a IN.eGFP.

HIV-1 IN CCD mutant

We used a CCD mutant containing a point mutation at position 116 (D116N) of HIV-1 IN, which renders the virus catalytically inactive for integration (Engelman et al, 1995). With this mutant we infected macrophages to test for R-loop colocalization in the absence of integration.
Upon infection of primary macrophages a significant reduction of PICs colocalizing with R-loops could be observed. Only 30,1% of the PICs were detected to colocalize with R-loops, which represents a drop of 41,45 % compared to the wt virus (Figure 2.29.).

HIV-1 wt IN in the presence of Raltegravir

Upon the assessment of a drop in HIV-1 D116N IN PICs and R-loop colocalization, the following question to answer was whether a chemical inhibitor of wt IN would contribute to a similar drop in PIC and R-loop colocalization. The colocalization of wt IN PICs and R-loops was quantified in the presence of IN strand transfer inhibitor Raltegravir.

Raltegravir contains Mg^{2+} chelating groups and is supposed to bind the two metal ions in the catalytic site of IN and block the binding of host DNA (Mouscadet & Tchertanov, 2009).

Upon infection of macrophages with HIV-1 and the addition of Raltegravir at the moment of infection and after 24 h of infection, a drop in the number of PICs colocalizing with R-loops at 60 h post infection could be appreciated. Only 21,45% of the PICs were colocalizing with R-loops, which represents a drop in colocalization of 50,1%, compared to the colocalization of wt IN without the addition of any IN inhibitor (Figure 2.29.).

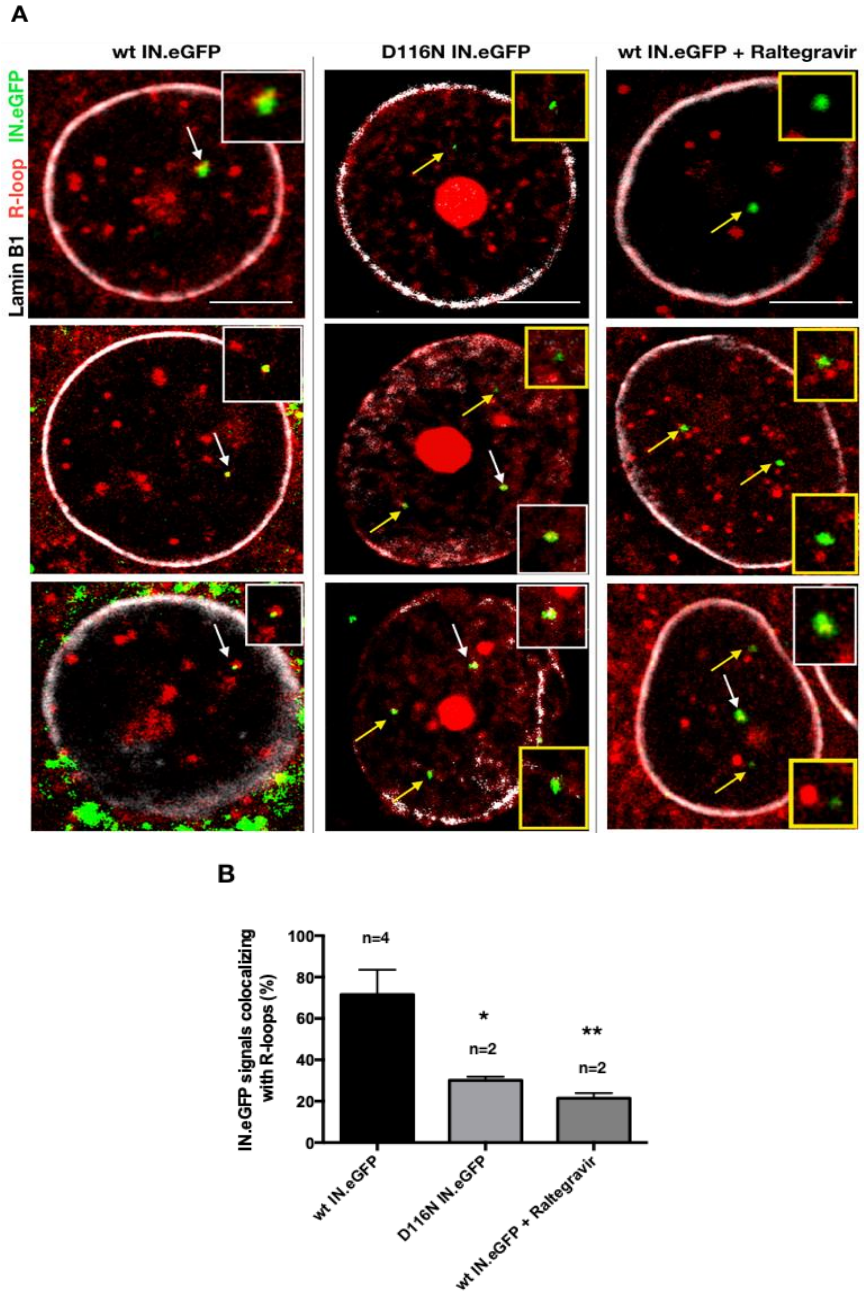


Figure 2.29. Comparison of wt IN.eGFP HIV-1, D116N IN.eGFP HIV-1 and wt IN.eGFP HIV-1 in the presence of Raltegravir and their colocalization with R-loops.

A. Representative confocal images of HIV-1 PIC colocalization with R-loops in all analyzed conditions. White arrows indicate PICs colocalizing with R-loops, whereas yellow arrows indicate PICs not colocalizing with R-loops. The blue channel represents IF staining of Lamin B1. The red channel represents IF staining of R-loops. The green signals represent the GFP signals from the IN.eGFP. The very pronounced R-loop signals correspond to cellular nucleoli which are composed of ribosomal genes and nucleolar organizer regions known to be very R-loop rich (Wahba et al, 2016; Kuznetsov et al, 2018) (scale bars: 2 µm). **B.** Graph shows a summary of quantified PIC colocalization with R-loops for all analyzed conditions. The numbers above the bars indicate numbers of analyzed donors. Showed results were generated 60 h post infection.

The obtained results suggest that catalytically active HIV-1 IN is required for high occurrence of PIC and R-loop colocalization. In addition, the results point to a scenario in which PICs containing wt IN do not just by chance colocalize with R-loops, but that this colocalization represents a non-random process that is observed in the case of 75% of PIC signals in primary $CD4^+$ T cells and 72% of PICs in macrophages.

II.XIII. Assessment of IN C-terminal mutant and replication deficient HIV-1 colocalization with R-loops

We further reasoned that 3 residues in the C-terminal region of HIV-1 IN, which have previously been shown to decrease the DNA binding activity of the viral enzyme (Cereseto et al, 2005) could contribute to the R-loop binding capacity of HIV-1 PIC. The same 3 residues (K264,K266 and K273) were also shown to have RNA binding properties (Kessl et al, 2016), and could therefore indeed be important for the recognition of RNA-DNA hybrids. Accordingly, cells were infected with a virus containing mutations in those 3 lysine residues (K264,K266,K273) in the C-terminal domain of IN.

We also used a *tat* deficient, replication incompetent HIV-1, in order to verify if the colocalization between PICs and R-loops is due to the recognition and of viral docking to R-loop sites, or if it is a consquence of of R-loop formation at the replicating viral genome.

All virus variants used in this set of experiments were produced by the Vpr-trans incorporation strategy (Albanese et al, 2008) and contained a IN.eGFP.

HIV-1 IN C-terminal mutant (K264,266,273Q)

Viral infectivity of the C-terminus IN mutant was assessed by infecting primary $CD4^+$ T cells and Jurkat cells. The virus showed a severe drop in integration capacity, which is followed by the absence of viral RNA production (Figure 2.30. B and C). In addition, the assessment of early and late viral cDNA transcripts was performed, in order to exclude any other deficiency in the viral life cycle and to prove that only integration is affected by the C-terminal domain mutations of IN (K264,266,273Q) (Figure 3.30. B and D). The results show that the reverse transcription is not affected by the C-terminal domain mutations of IN, and that viral DNA is generated.

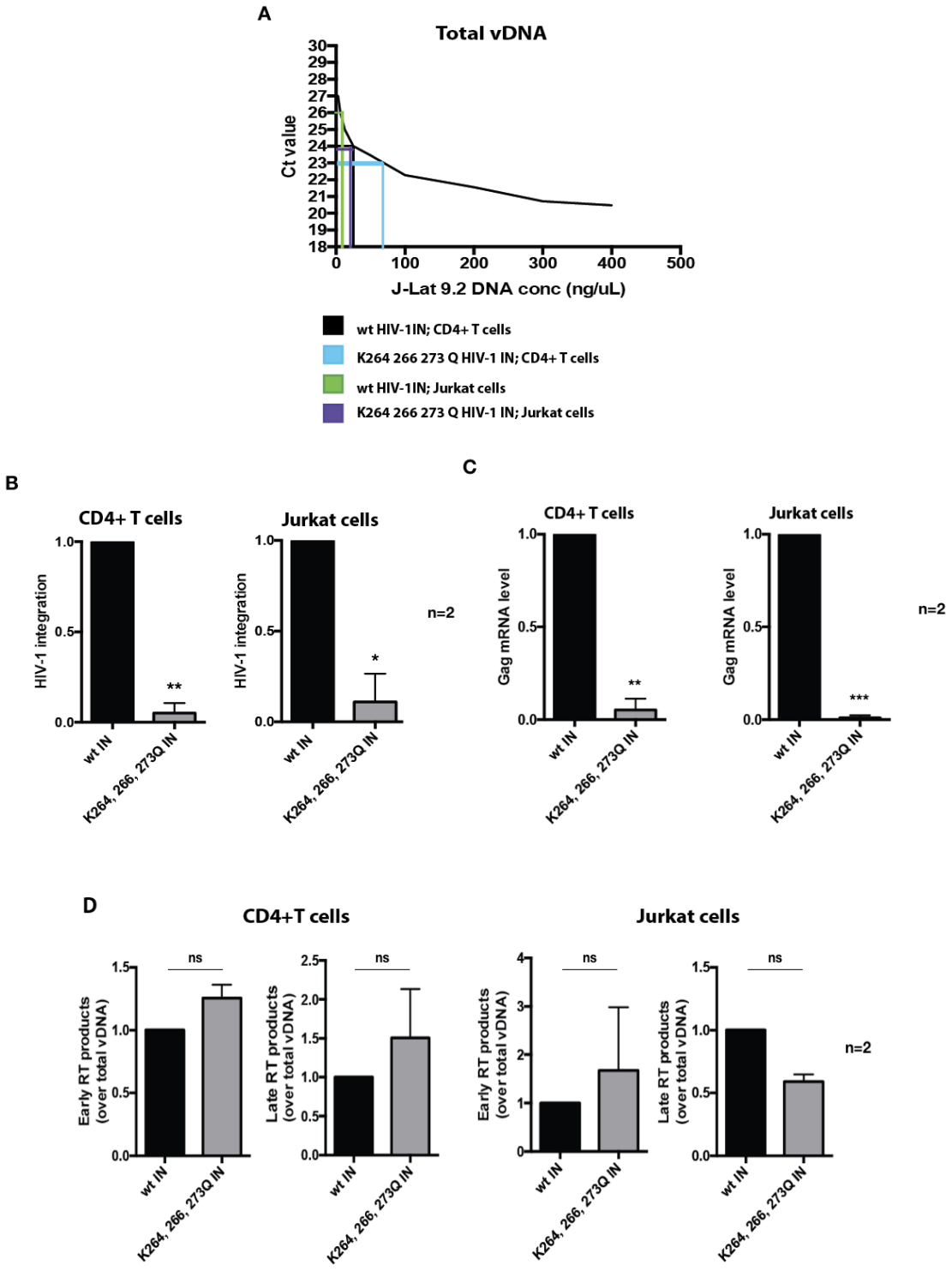

A
Total vDNA
Ct value
J-Lat 9.2 DNA conc (ng/uL)
wt HIV-1IN; CD4+ T cells
K264 266 273 Q HIV-1 IN; CD4+ T cells
wt HIV-1IN; Jurkat cells
K264 266 273 Q HIV-1 IN; Jurkat cells
B
CD4+ T cells
Jurkat cells
HIV-1 integration
n=2
C
CD4+ T cells
Jurkat cells
Gag mRNA level
n=2
wt IN
K264, 266, 273Q IN
D
CD4+T cells
Jurkat cells
Early RT products (over total vDNA)
Late RT products (over total vDNA)
ns
n=2

Figure 2.30. Features of HIV-1 IN (K264,266,273Q) C-terminal mutant.

A. Quantification of total vDNA levels upon $CD4^+$ T cell and Jurkat cell infection with HIV-1 K264,266,273Q IN. Total vDNA was quantified by plotting sample total vDNA qPCR Ct values over a graph of total vDNA qPCR Ct values (y-axis) of serial dilutions of latent HIV-1 clone J-Lat9.2 cell DNA (x-axis). **B.** Graphs show integration levels of HIV-1 K264,266,273Q IN compared to wt HIV-1 levels in $CD4^+$ T cells and Jurkat cells. HIV-1 integration levels in wt HIV-1 infection are set to 1 and used as a reference to assess integration levels upon cell infection with HIV-1 K264,266,273Q IN mutant. Integrated vDNA (Alu Ct values) is normalized over total genomic DNA levels amplified with primers for the lamin B2 gene (B13 region). **C.** Graph shows RNA expression levels of the *gag* gene upon HIV-1 K264,266,273Q IN infection. Viral *gag* mRNA expression levels in wt HIV-1 infection are set to 1 and used as a reference to assess *gag* mRNA expression levels upon cell infection with HIV-1 K264,266,273Q IN mutant. Viral *gag* mRNA expression is normalized over expression profiles of housekeeping gene GAPDH. **D.** Graphs show the production of early and late viral DNA transcripts of wt HIV-1 and HIV-1 K264,266,273Q IN, normalized over the levels of total vDNA. The levels of early and late viral DNA transcripts of wt HIV-1 are set to 1 and used as a reference to assess the amount of early and late viral DNA transcripts upon cell infection with HIV-1 K264,266,273Q IN mutant. Panels A-D represent the results of the same material from 2 independent donors and 2 independent repetitions.

Further on, the colocalization between PICs of HIV-1 K264,266,273Q IN.eGFP and R-loops was assessed in primary $CD4^+$ T cells and macrophages.

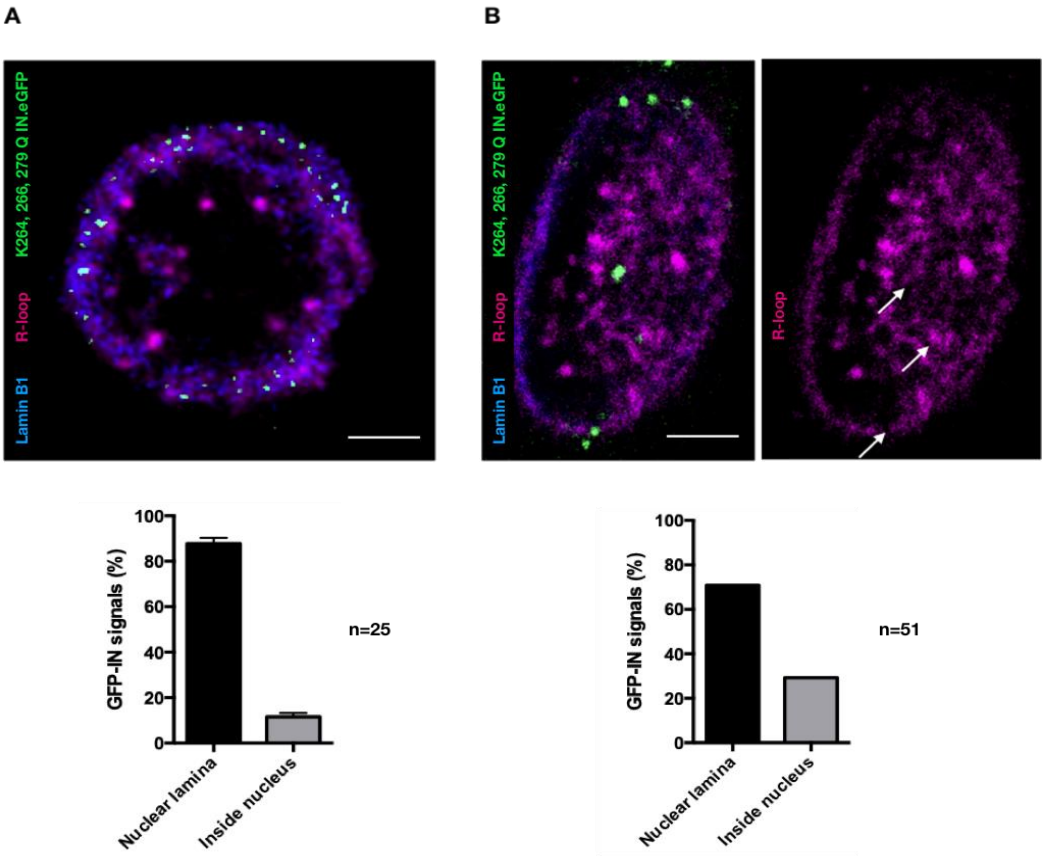


Figure 2.31. HIV-1 IN (K264,266,273Q) C-terminal mutant colocalization with R-loops.

A. Representative confocal images of HIV-1 K264,266,273Q IN,eGFP PICs colocalization with R-loops in primary $CD4^+$ T cells. Images were acquired 72 h post infection. The blue channel represents IF staining of Lamin B1. The magenta channel represents IF staining of R-loops. The green signals represent the GFP signals from the IN.eGFP. Graph below corresponds to the analysis of the nuclear position of IN.eGFP signals in relation to the nuclear lamina or nuclear interior, analyzed in 25 cells of 2 donors. **B.** Representative confocal images of HIV-1 K264,266,273Q IN.eGFP PICs colocalization with R-loops in primary macrophages. The blue channel represents IF staining of Lamin B1. The magenta channel represents IF staining of R-loops. The green signals represent the GFP signals from the IN.eGFP. The R-loop channel is shown separately and the white arrows are pointing to the spots where HIV-1 PICs are positioned. Images were acquired 60 h post infection. Graph below corresponds to the analysis of the nuclear position of IN.eGFP signals in relation to the nuclear lamina or nuclear interior, analyzed in 51 cells of one donor (scale bars: 2 µm).

The obtained results show that HIV-1 K264,266,273Q IN.eGFP PICs either remain trapped in the nuclear lamina, or once inside the nucleus, do not colocalize with R-loops. Based on the analysis performed on primary $CD4^+$ T cells, 88% of HIV-1 K264,266,273Q IN.eGFP PICs are found in the nuclear lamina, whereas 12% of the signals can be detected inside of the nucleus (Figure 2.31. A). A very similar scenario was observed for macrophages in which 71% of HIV-1 K264,266,273Q IN.eGFP PICs were detected in the nuclear lamina, and 29% inside the nucleus (Figure 2.33. B). It remains elusive at this point why HIV-1 K264,266,273Q IN.eGFP PICs are being trapped in the nuclear lamina. However, the PICs of the analyzed C-terminal mutant of IN are not docked to R-loop forming genomic regions, which represents another indication that only wt HIV-1 PICs are successfully docked to R-loops.

II. XIV. R-loops are enriched in HIV-1 RIGs

The initial bioinformatic analysis of HIV-1 RIGs and R-loops (Sanz et al, 2016) (Figure 2.26. B) indicated that RIGs are enriched in R-loops. To experimentally verify the presence of DNA-RNA hybrids on selected RIGs, immunoprecipitation followed by qPCR (DRIP-qPCR) was performed in primary $CD4^+$ T cells, according to the method from Chedin's lab (Ginno et al, 2012, 2013; Sanz et al, 2016 and personal communication). DRIP-qPCR represents an approach in which R-loops can be mapped by using the R-loop specific antibody S9.6 (Boguslawski et al, 1986) (Figure 2.32.).

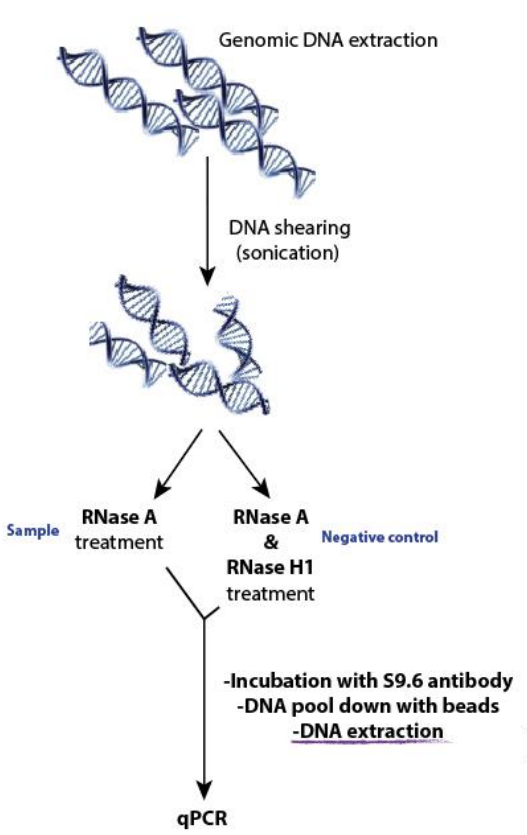


Figure 2.32. Schematic representation of DRIP-qPCR workflow.

Genomic DNA, isolated from primary $CD4^+$ T cells, was sheared to fragment sizes around or below 500 bp to enable efficient S9.6 antibody binding and final DNA pull down. After shearing, the initial material was split in two different tubes, which were treated as a positive (only RNase A treated) and negative control sample (RNase A and RNase H1 treated). Upon shearing and enzymatic digestion both samples were incubated with the S9.6 antibody overnight which was followed by DNA pull down with agarose beads and DNA extraction. After DNA extraction, qPCR was performed for selected control genes known to be either R-loop rich or R-loop poor genomic sites (Gino et al, 2012; Sanz et al, 2016), as well as for selected RIGs (BACH2 and STAT5B) (Figure 2.33.).

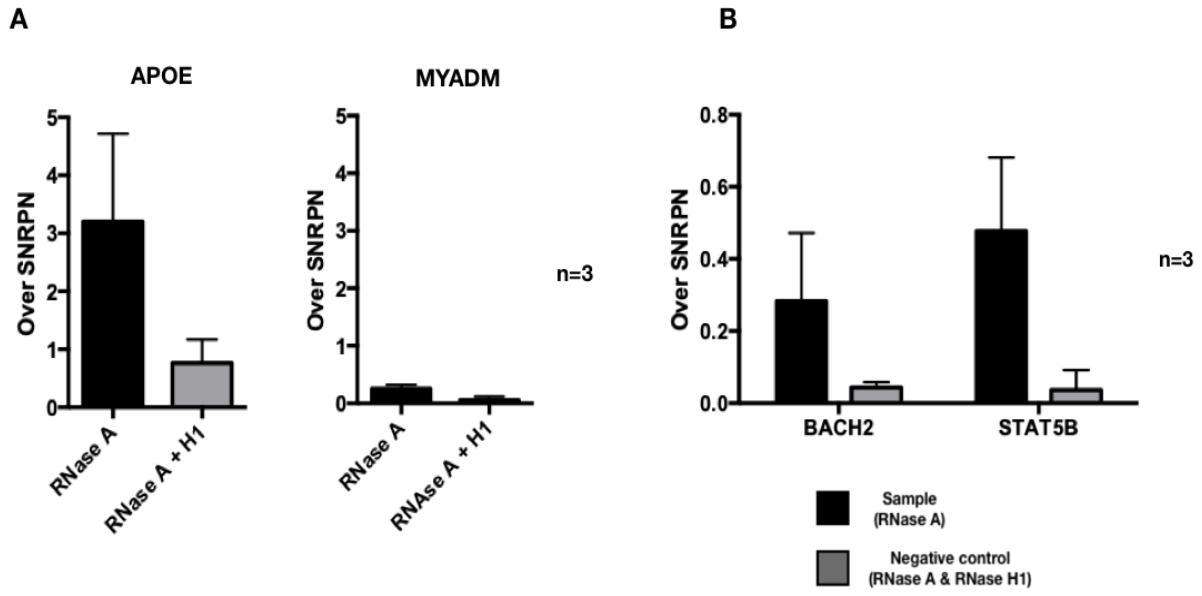


Figure 2.33. DRIP-qPCR results of control genes and selected RIGs.

A. S9.6 immunoprecipitation of DNA-RNA hybrids on selected positive control gene APOE and negative control gene MYADM in the presence of RNase A (control treatment) or RNase A+ RNase H1 treatment, which removes the formed RNA-DNA Hybrids. All samples were normalized over R-loop negative gene SNRPN. The qPCR results of positive and negative R-loop genes represent a control and readout for the efficiency of the DRIP procedure. **B.** Graph shows the enrichment of R-loops in BACH2 and STAT5B gene in primary $CD4^+$ T cells, normalized over R-loop negative gene SNRPN. The black bars in the graphs correspond to the positive portion of the samples (only RNase A treated) which should contain R-loops, whereas the gray bars corresponds to the negative control samples (RNase A and RNAse H1 treated) in which R-loops should be depleted. The graphs represent a summarized results of the same 3 independent donors.

According to the results of DRIP-qPCR on primary $CD4^+$ T cells, a clear enrichment in R-loop signals could be appreciated on two most prominent RIGs, BACH2 and STAT5B (Figure 2.33. B). The results represent a confirmation of the metadata analysis (Figure 2.26.) which indicated that RIGs are enriched in R-loops. The detection of R-loops in RIGs was also somewhere expected, considering that HIV-1 integration genes usually have high G/C skew and are highly expressed (Schröder et al, 2002; Brady et al, 2009; Marini et al, 2015) and that R-loops mainly occur co-transcriptionally in G/C rich genomic regions (Gino et al, 2012; Sanz et al, 2016; Chedin, 2016).

II. XV. Integrated HIV-1 is detected in R-loop rich regions

We aimed to confirm that the integrated HIV-1 genome can be found in genomic regions where DNA-RNA hybrids are formed. For this purpose, the DRIP procedure was performed on HIV-1 infected primary CD4$^+$ T cells, and was modified to specifically amplify the genomic regions containing HIV-1 by applying the Alu PCR protocol to detect the integrated viral DNA in the precipitated fraction (Figure 2.34. A).

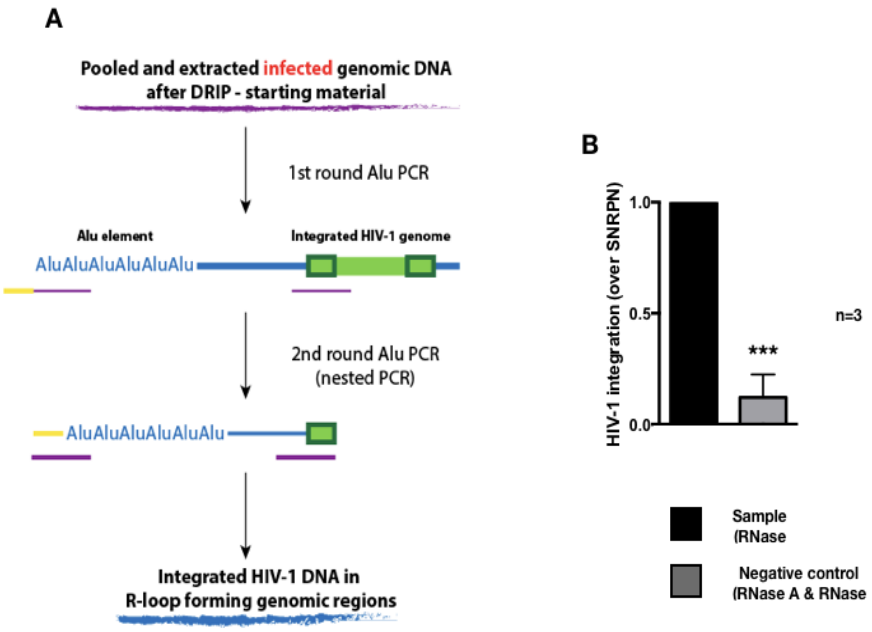


Figure 2.34. Assessment of the presence of integrated vDNA in R-loop rich genomic DNA.

A. Schematic representation of the DRIP-Alu PCR in which upon the classical DRIP approach on infected primary CD4$^+$ T cells Alu PCR is performed. **B.** Graph shows results of the DRIP-Alu PCR in HIV-1 infected primary CD4$^+$ T cells. HIV-1 integration levels in the actual sample are set to 1 and used as a reference to assess HIV-1 integration levels in the RNase H1 treated, negative control sample. DRIP-Alu PCR are normalized over R-loop negative gene SNRPN. The graph represents a summary of 3 independent donors.

Upon performing this modified approach on infected primary CD4$^+$ T cells, we observed that the integrated HIV-1 DNA is enriched in R-loop rich genomic sites (Figure 2.34. B). This result represents a direct proof that HIV-1 integration can occur in R-loop regions, thus confirming the meta data analysis (Figure 2.26.), as well as the colocalizing events of HIV-1 PICs and R-loops (Figure 2.27. B).

II. XVI. *In vitro* assessment of the interaction between HIV-1 IN and R-loops

Upon making the observation that HIV-1 PICs colocalize with R-loops with high occurrence (more than 72% of PIC signals) in primary $CD4^+$ T cells and macrophages, and that a functionally active HIV-1 IN is required for this colocalization, the following questions to answer were whether HIV-1 IN indeed has R-loop binding properties, whether it can perform the strand transfer reaction on an R-loop, and whether it eventually has RNase H1 properties and can thus resolve an R-loop. The need to answer those questions was even more strengthened by the observation that according to the DRIP-Alu PCR in HIV-1 infected primary $CD4^+$ T cells the integrated viral DNA can be detected at R-loop forming sites.

In order to get a more detailed insight into the catalytic features of HIV-1 IN in the context of R-loops, *in vitro* binding and strand transfer assays were performed in the presence of HIV-1 IN and R-loop and non-R-loop nucleic acid substrates.

The main part of the following results were generated by our collaborator from the University of Bordeaux, Dr. Vincent Parissi.

HIV-1 IN binds R-loops

The assessment of R-loop binding by HIV-1 IN was performed in an assay in which an *in vitro* formed R-loop on the pFC53 plasmid (provided by the Chedin Lab), as well as the pFC53 plasmid only (without R-loop) were offered to HIV-1 IN for binding.

The pFC53 plasmid contains a G/C rich *Airn* region flanked by a T3 and T7 promoter. By *in vitro* transcription initiation from the plasmids T3 promoter, by T3 RNA polymerase, R-loop formation can be induced on the plasmid. R-loop formation can be monitored by agarose gel electrophoresis and appreciated as a shift on the gel (Figure 2.35.). In addition, the R-loop forming region of the plasmid is flanked by recognition sites of restriction enzymes and can be cut from the plasmid, which enables the monitoring of the R-loop and non-R-loop fragments of the plasmid in further experimental steps.

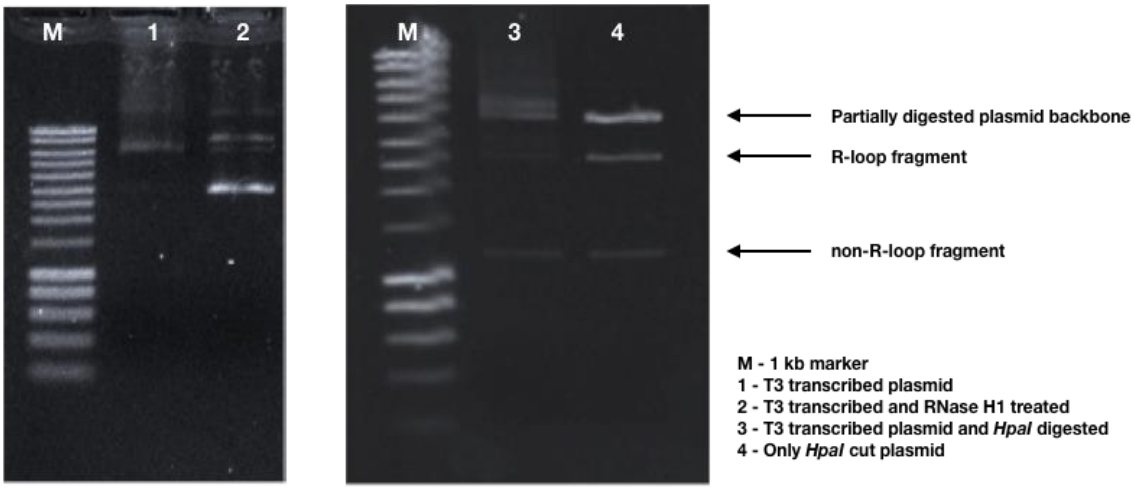


Figure 2.35. Monitoring of *in vitro* R-loop formation on the pFC53 plasmid.

In lane 1 a shift and a smear on the gel can be detected upon transcription initiation on the pFC53 plasmid, indicating successful R-loop formation. In lane 2 the recognizable profile of the transcribed and RNase H1 treated plasmid can be appreciated. In line 3 a clear shift on the gel can be appreciated, indicating successful R-loop formation, as well as a release of all plasmid fragments upon *HpaI* digestion. Lane 4 represents a control digestion of the pFC53 plasmid with *HpaI*, confirming that the plasmid can be successfully cut into its 2 main fragments; the R-loop fragment and non-R-loop fragment.

Upon *in vitro* R-loop formation and preparation of nucleic acid substrates, HIV-1 IN was added to the reaction. The binding of IN to the different substrates was monitored by gel shift assays. IN was also added in different concentrations to the substrates, ranging from 0,2-1,6 µM, in order to monitor binding affinity and efficiency (Figure 2.36. A). The results show that HIV-1 IN can bind to R-loops (shifted high molecular complexes that can be detected in the gel pockets). In addition, upon this initial observation, IN binding to R-loops was further monitored by using lower concentrations of IN (below 0,4 µM) to assess binding efficiency (Figure 2.36. B). The results show that IN can efficiently bind to R-loops, even when low concentrations of IN are exposed to nucleic acid substrates containing formed R-loops.

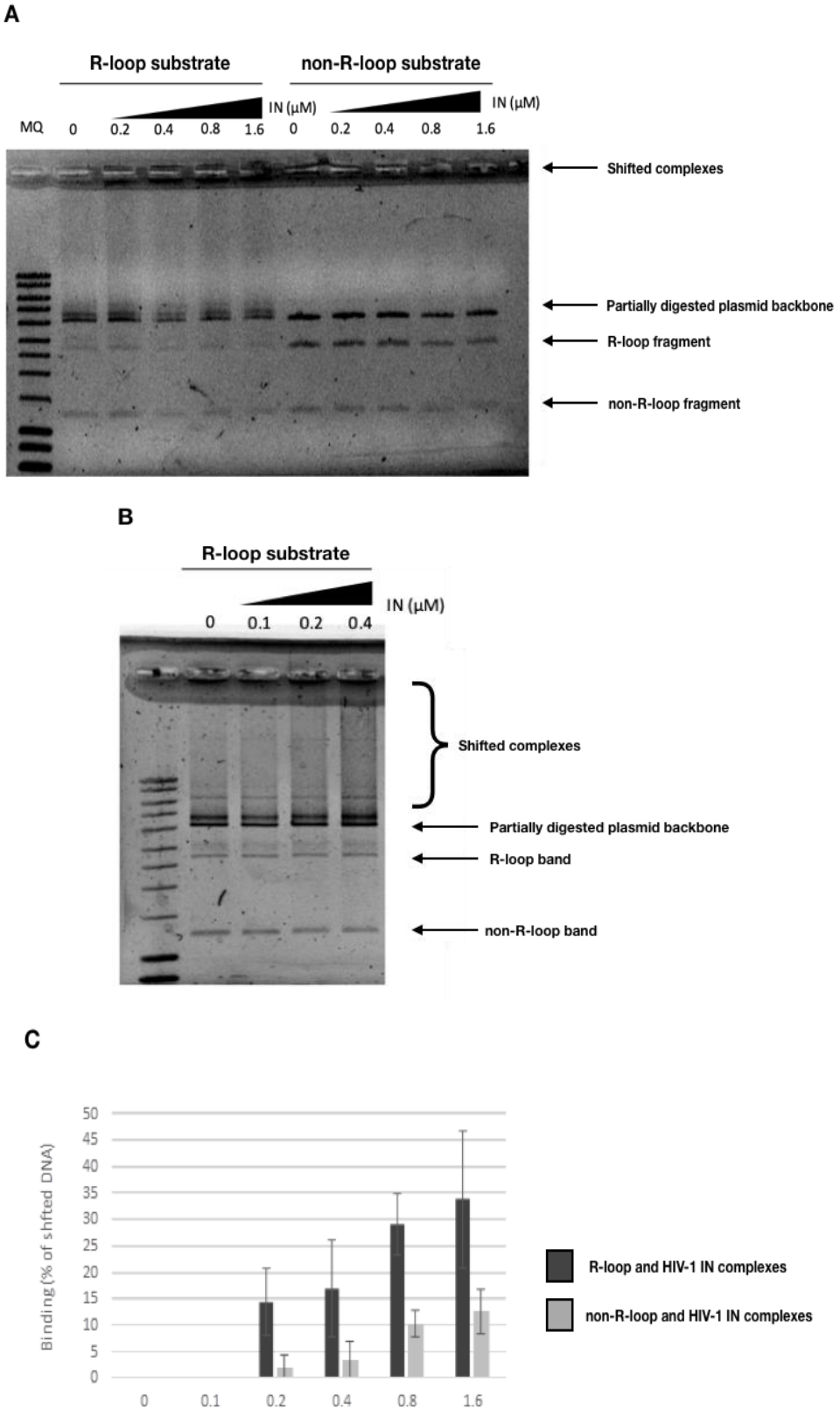

A
R-loop substrate
non-R-loop substrate
IN (μM)
IN (μM)
MQ
0
0.2
0.4
0.8
1.6
0
0.2
0.4
0.8
1.6
Shifted complexes
Partially digested plasmid backbone
R-loop fragment
non-R-loop fragment
B
R-loop substrate
IN (μM)
0
0.1
0.2
0.4
Shifted complexes
Partially digested plasmid backbone
R-loop band
non-R-loop band
C
Binding (% of shfted DNA)
50
45
40
35
30
25
20
15
10
5
0
0
0.1
0.2
0.4
0.8
1.6
R-loop and HIV-1 IN complexes
non-R-loop and HIV-1 IN complexes

Figure 2.36. *In vitro* binding of HIV-1 IN to nucleic acid substrates which contain or not contain formed R-loop structures.

A. Comparison of HIV-1 IN binding affinity to nucleic acid substrates containing R-loops and not containing R-loops via gel shift assay and by the addition of increasing amounts of recombinant IN enzyme (indicated in the upper part of the gel). R-loop formation was performed via transcription initiation on the pFC53 plasmid which was used as the nucleic acid of the binding assay. The marks on the right part of the gel indicate the generated plasmid fragments upon *HpaI* enzymatic digestion of the pFC53 plasmid which either contained formed R-loops or did not. **B**. Assessment of HIV-1 IN binding efficiency on R-loops by providing low concentrations of IN to the nucleic acid substrate (pFC53 plasmid) via gel shift assay (indicated in the upper part of the gel). The marks on the right part of the gel indicate the generated plasmid fragments upon *HpaI* enzymatic digestion of the offered nucleic acid substrate. **C.** Graph shows quantified optical density values of shifted R-loop/IN and non-R-loop/IN complexes from the gel shift assays (data obtained by Dr. Vincent Parissi)**.**

In addition, optical density quantification of shifted R-loop/IN and non-R-loop/IN complexes from the gel shift assays indicated that HIV-1 IN showed more affinity for the binding of R-loops in comparison to nucleic acid substrates without containing an R-loop structure (Figure 2.36. C). The obtained results represent proof that HIV-1 IN can efficiently bind R-loops.

HIV-1 IN performs the strand transfer on R-loop templates

Upon defining that HIV-1 IN has the property of binding R-loop containing nucleic acid substrates, the assessment of IN strand transfer reaction on R-loop templates was performed. As previously mentioned, nucleic acid substrates were generated from the pFC53 plasmid and offered to HIV-1 IN in the context of a strand transfer reaction. After the integration assay was performed, the generated integration products were treated with either RNase A or RNase A and RNase H1 and *HpaI* in order to release the integration products. Upon enzymatic digestion, the products of the strand transfer assay were monitored by gel shift assays (Figure 2.37.).

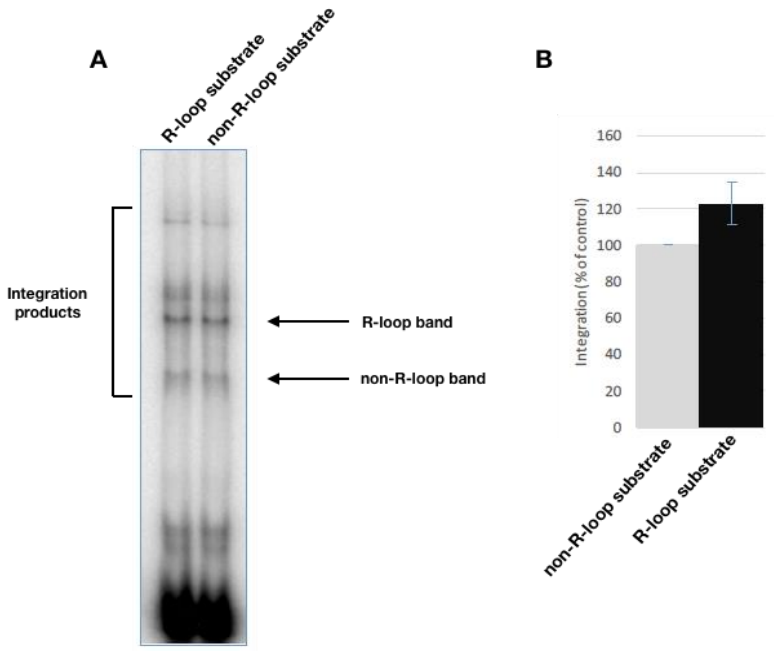


Figure 2.37. Monitoring the HIV-1 IN strand transfer reaction on R-loop containing and non-R-loop containing nucleic acid substrates.

A. Gel shift assay results of the performed strand transfer reaction on R-loop and non-R-loop containing DNA templates generated on the pFC53 plasmid. Different smears on the gel represent the generated and shifted integration products. The arrows on the right indicate the positions of the R-loop forming and non-R-loop forming part of the pFC53 plasmid upon *HpaI* enzymatic digestion, which was used as the nucleic acid template for the strand transfer reaction. The first lane shows the result of the integration assay performed on the pFC53 plasmid containing an *in vitro* formed R-loop, whereas the second line shows the result of the integration assay performed on the naked pFC53 plasmid. **B.** Graph shows quantified optical density values of integration products generated R-loop and non-R-loop containing DNA templates (data obtained by Dr. Vincent Parissi).

The obtained results show that HIV-1 IN can perform the strand transfer reaction with equal efficiency on nucleic acid substrates containing R-loops or not containing R-loops. A slightly higher efficiency towards the R-loop containing substrate was, however, not statistically significant (Figure 2.37.).

In addition, in a very similar experimental setup in which RNase H1 was added to the strand transfer reaction in the presence of HIV-1 IN and nucleic acid substrates, it was observed that the strand transfer reaction is performed more efficiently on R-loop templates once the R-loop is resolved by RNase H1 (Figure 2.38.). This result implies that HIV-1 IN does not have RNase H1

functions, and is not capable to solve an R-loop independently. The obtained result suggests that HIV-1 IN might require a binding partner with helicase activity, possibly similar to RNase H1, in order to efficiently perform integration of the viral genome into R-loop forming regions.

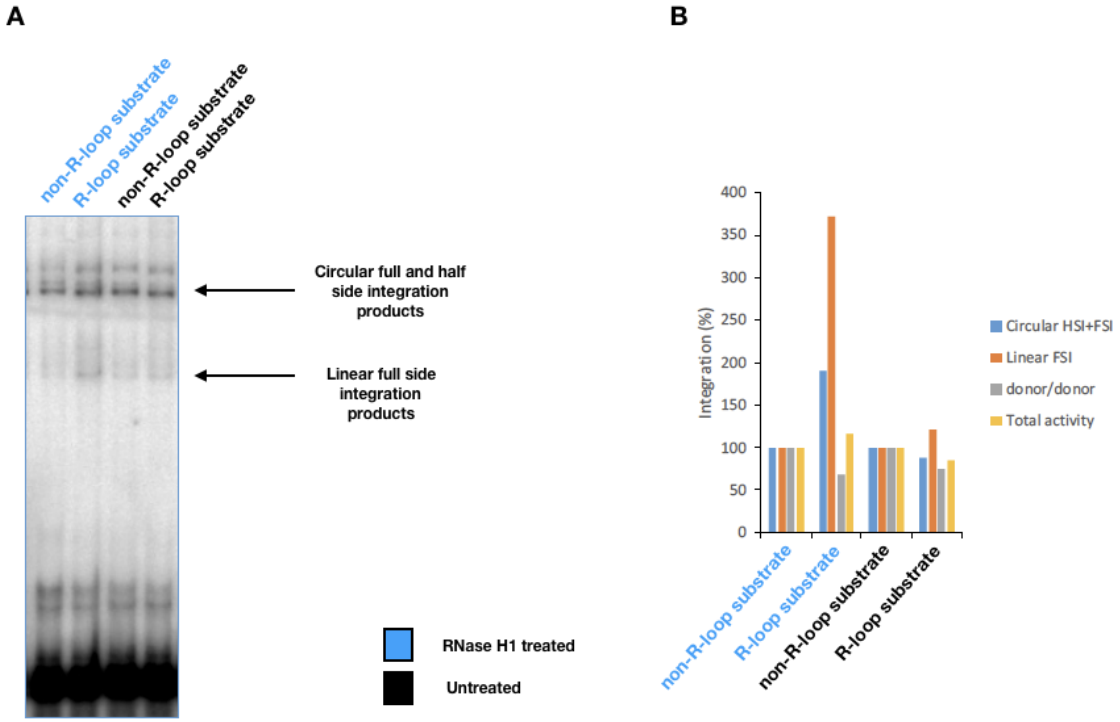


Figure 2.38. HIV-1 IN strand transfer reaction results with nucleic acid substrates with and without R-loop structures in the presence of RNase H1.

A. Gel shift assay results of the performed strand transfer reaction on R-loop and non-R-loop containing DNA templates in the presence of RNase H1. The pFC53 plasmid was used as the nucleic acid template for the strand transfer reaction. **B.** Graph shows quantified optical density values of integration products generated R-loop and non-R-loop containing DNA templates in the presence of RNase H1. The legend on the right shows the comparison of integration product amounts between nucleic acid templates and RNase H1 treated and untreated templates. Blue bars indicate the amount of circular half and full side integration products, the red bar shows the amount of linear full side integration products, the gray bar shows the amount of present donor DNA (viral genome) and the yellow bar represents the total activity of the recombinant IN enzyme measured by the presence of integration products (data obtained by Dr. Vincent Parissi).

II. XVII. R-loop depletion affects HIV-1 integration

Considering the datasets obtained so far, which strongly supports HIV-1 integration into R-loops, it was important to understand how and if HIV-1 integration is affected once R-loops are depleted in cells. In order to test whether HIV-1 integration is affected by R-loop removal, different cell lines (HEK293T, HeLa and Jurkat cells) as well as primary cells ($CD4^+$ T cells and macrophages) were transfected with the pMMM-FLST plasmid construct coding for human RNase H1 and containing a M27 nuclear localization signal (to ensure RNase H1 expression and R-loop depletion in the cell nucleus), as well as a GFP tag (to control for plasmid transfection efficiency) (Suzuki et al, 2010). Upon transfection and RNase H1 overexpression, cells were infected with HIV-1, and subsequently HIV-1 integration was assessed by Alu PCR (Figure 2.39.).

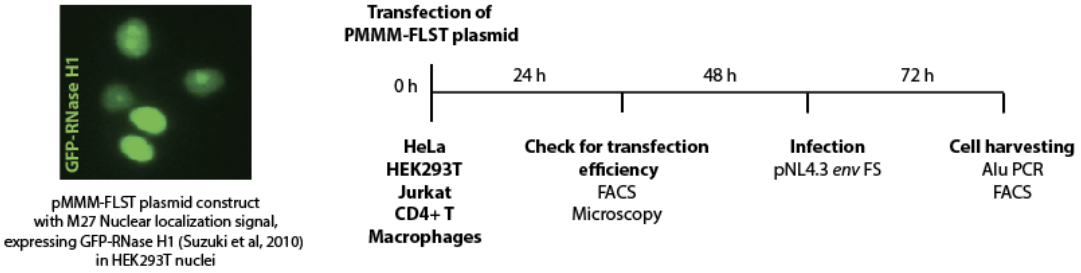
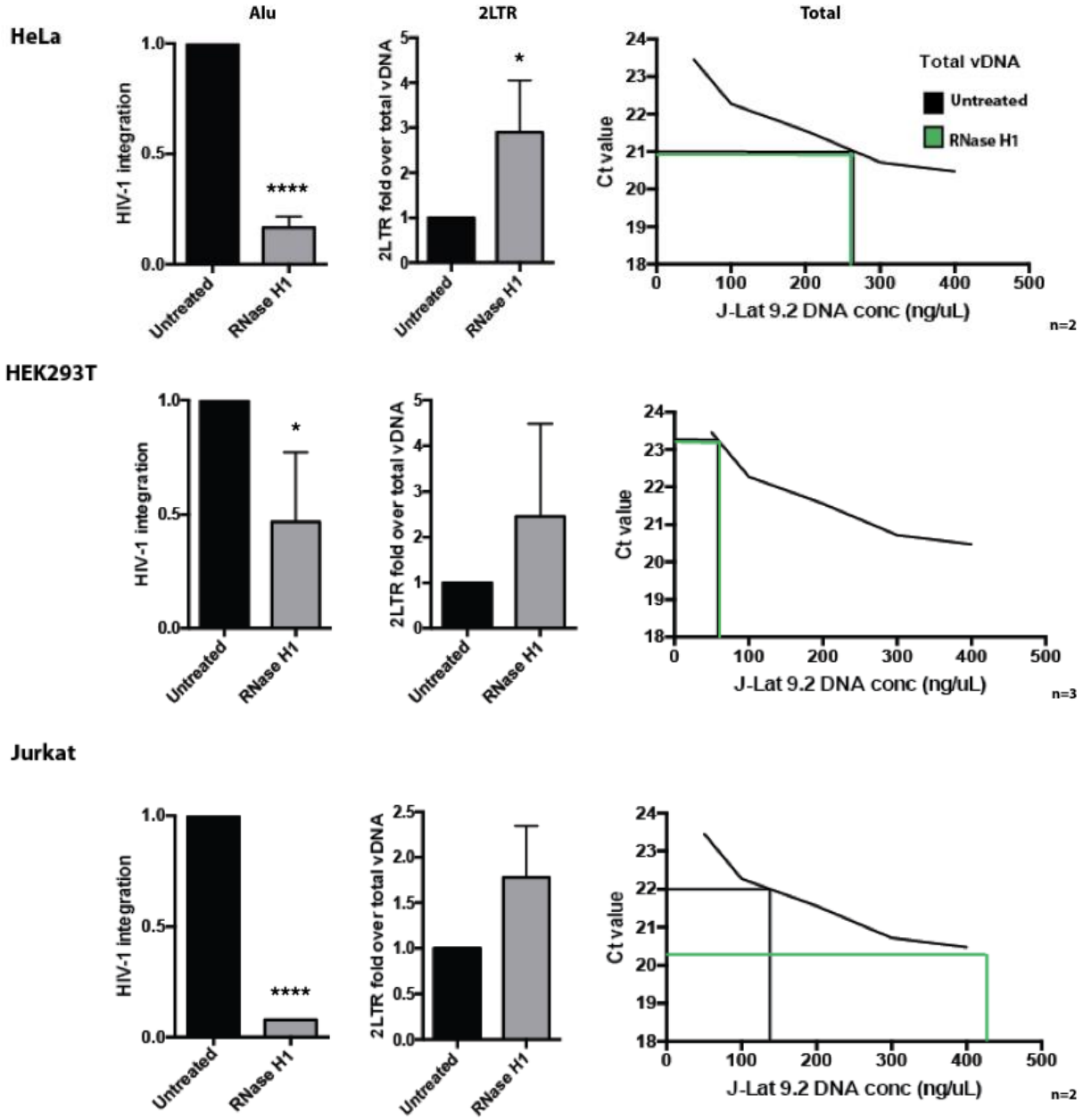


Figure 2.39. Scheme of experimental workflow of R-loop depletion and HIV-1 infection.

The image on the left shows successful transfection and expression of the pMMM-FLST construct in HEK293T cell nuclei. The timeline on the right shows the experimental procedure at certain time points. First, cells are transfected with the pMMM-FLST plasmid construct containing an NLS and expressing an GFP-RNase H1. 24 h post transfection, transfection efficiency is assessed by FACS analysis or microscopy approaches. Cells are left for another 24 h in culture to reach a peak of GFP expression (estimated to occur 48 h post plasmid construct transfection). 48 h post transfection cells are infected with HIV-1 and collected 72 h post infection for Alu PCR analysis.

Apart from Alu PCR, PCRs in which we controlled for the presence of total vDNA and the formation of 2LTR circles were performed (Figure 2.40.).



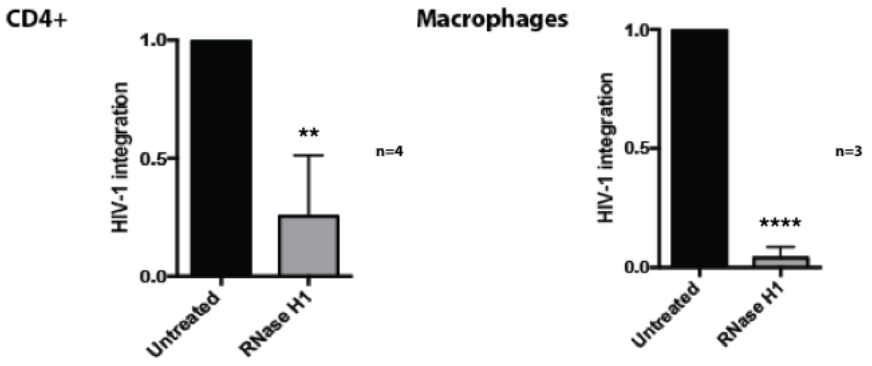


Figure 2.40. Assessment of HIV-1 integration upon RNase H1 overexpression in human cell lines and primary cells.

Alu PCR results of HIV-1 integration upon RNase H1 overexpression are shown, as well as the assessment of 2LTR circles upon RNase H1 overexpression and HIV-1 infection. Alu PCR values in untreated conditions are set to 1 and used as a reference to assess HIV-1 integration levels upon RNase H1 overexpression. Integrated vDNA (Alu Ct values) is normalized over total genomic DNA levels amplified with primers for the lamin B2 gene (B13 region). The amounts of 2LTR circles were normalized over total vDNA values and 2LTR circle levels in untreated conditions are set to 1 and used as a reference to assess 2LTR circle levels upon RNase H1 overexpression. Total vDNA was quantified by plotting sample total vDNA qPCR Ct values over a graph of total vDNA qPCR Ct values (y-axis) of serial dilutions of latent HIV-1 clone J-Lat9.2 cell DNA (x-axis). The numbers below or next to the graphs indicate the number of independently performed repetitions.

According to the Alu PCR results, HIV-1 integration is significantly decreased after overexpression of RNase H1 and depletion of R-loops in different human cell lines (HeLa, HEK293T, Jurkat), as well as primary $CD4^+$ T cells and macrophages.

The remaining HIV-1 integration in HeLa cells compared to untreated conditions upon infection was 16,67%, in HEK293T cells 46,67% and in Jurkat cells almost fully abrogated (0,79%).

In primary $CD4^+$ T cells integration levels of 25,5% could still be observed, whereas in macrophages integration was almost fully abrogated (0,4%), as seen in Jurkat cells.

The observed decrease in HIV-1 integration upon RNase H1 overexpression, was followed by increased levels of 2LTR circles measured in cell lines. In HeLa cells an increase in 2LTR circles of 2,89 fold was observed, in HEK293T cells 2,45 and in Jurkat cells 1,76.

In addition, total vDNA levels were assessed by direct comparison of real time PCR Ct values of amplified total vDNA (Apolonia et al, 2007) and real time PCR Ct values of serial dilutions of latent HIV-1 clone J-Lat 9.2 DNA.

In HeLa and HEK293T cells in untreated conditions and upon RNase H1 overexpression identical amounts of total vDNA were detected, whereas in Jurkat cells in RNase H1 treated conditions 3 times more total vDNA was detected, and still less integrated HIV-1 DNA.
Obtained results suggest that the depletion of R-loops in cell lines, as well as primary cells affects efficient HIV-1 integration.

In order to claim that the observed decrease in HIV-1 integration upon RNase H1 overexpression is a consequence of R-loop removal, and not due to potential degradation of viral RT products by the presence of RNase H1, we controlled for the presence of early (short DNA reverse transcripts from the 5' LTR) and late (reverse transcripts from the *gag* gene) viral RT-products, as well as 2LTR circle formation upon RNase H1 overexpression and infection at different time points post infection in primary $CD4^+$ T cells. For this purpose the digital droplet PCR (ddPCR) was performed, which enables an absolute quantification of selected targets in the sample, by using specifically designed primer probes for early and late HIV-1 RT products, as previously described by Bejarano et al, 2018. In addition, we controlled for successful R-loop removal upon RNase H1 overexpression by dot blot assay as previously described by Wahba et al, 2013 (Figure 2.41. A).

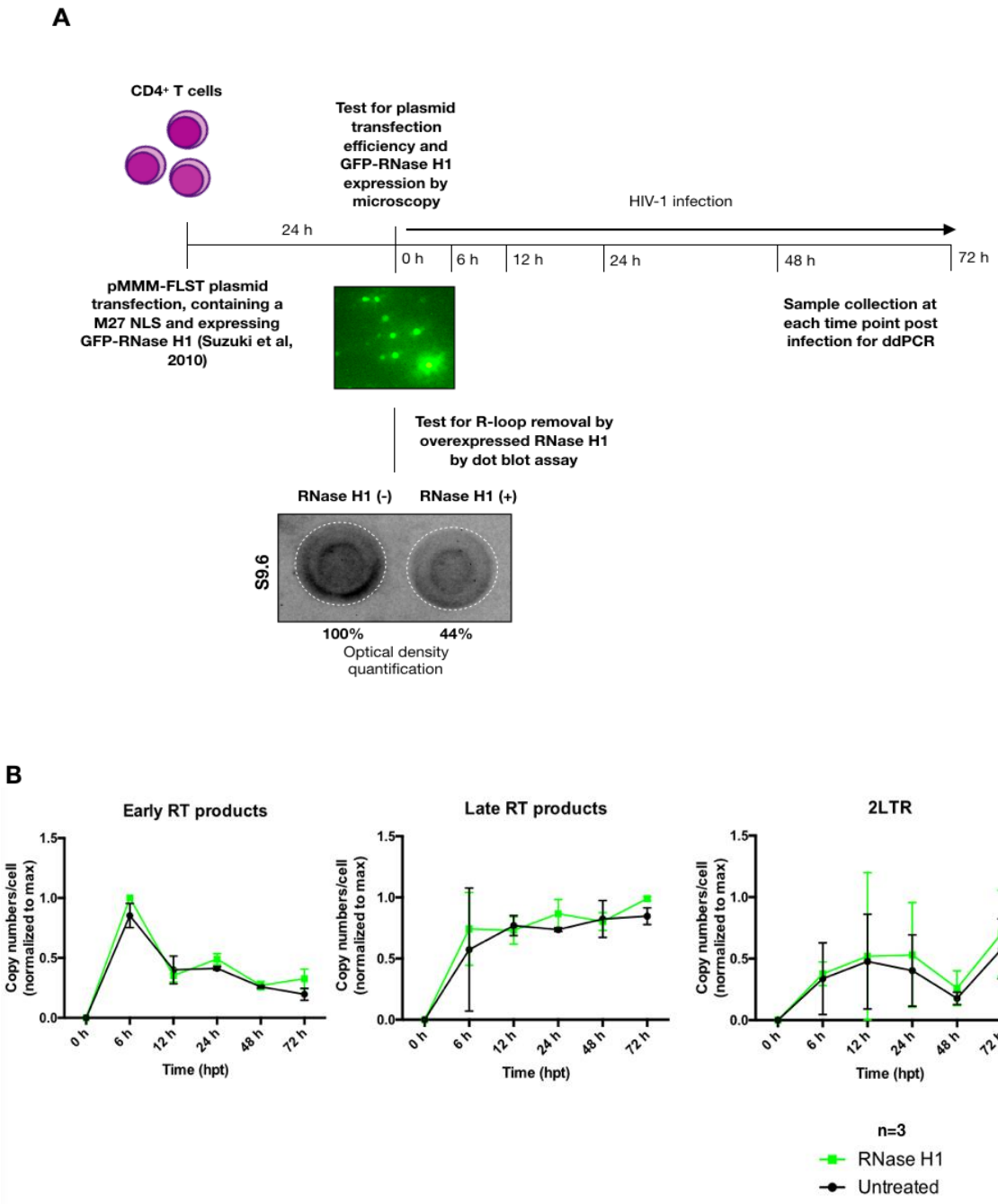


Figure 2.41. RNase H1 overexpression in primary CD4⁺ T cells and quantification of early and late HIV-1 RT-products, as well as 2LTR circles at different time points post infection.

A. Overexpression of RNase H1 in primary CD4⁺ T cells was performed by transfecting the pMMM-FLST plasmid construct containing a GFP-RNase H1 gene sequence. Plasmid transfection efficiency as well as GFP-RNase H1 expression was verified by microscopy approaches 24 h post transfection. The R-loop quantification 24 h post plasmid transfection is demonstrated by showing a representative dot blot where levels of

R-loops were assessed by the incubation of 500 µg loaded total genomic DNA with the S9.6 antibody. Optical density quantification of R-loop levels are shown below the dot blot. 24 h upon plasmid transfection, cells were infected with HIV-1 and collected at different time points post infection for ddPCR analysis of early and late viral RT products, as well as the presence of 2LTR circles. **B.** The quantification of early and late viral RT products, as well as 2LTR circles was performed by ddPCR. The obtained values for early and late viral RT products, as well as 2LTR circles were normalized over housekeeping gene RPP30 values. The obtained values were subsequently normalized to the highest detected absolute number. The black lines correspond to untreated condition, whereas the green lines correspond to RNase H1 overexpression. The graphs represent summarized results of 3 independent donors.

The obtained results show that the production of early and late viral RT-products is not affected by RNase H1 overexpression, indicating that the previously observed decrease in HIV-1 integration is a consequence of R-loop removal (Figure 2.41. B). In addition, the quantification of 2LTR circles showed an increasing trend in samples where RNase H1 was overexpressed (Figure 2.42. B), which is in correlation with the previously observed decrease in successful vDNA integration upon RNase H1 overexpression.

RNase H1 mutants do not affect HIV-1 integration

The observation that RNase H1 overexpression strongly reduces HIV-1 integration in cell lines and primary cells prompted us to better characterize this effect of the transfected enzyme. In particular, we wanted to understand if the observed effect is indeed due to the R-loop removal, or if it can be caused by the inability of HIV-1 IN to bind the R-loops in the presence of the functional RNase H1 helicase.

In order to answer this question, HEK293T and HeLa cells were transfected with plasmid constructs defective in R-loop binding (D210N), or RNase H1 mutants defective in both R-loop binding, as well as in the catalytic activity (WKKD). The cells were then infected with HIV-1, as further processed by Alu PCR to measure HIV-1 integration (Figure 2.42.).

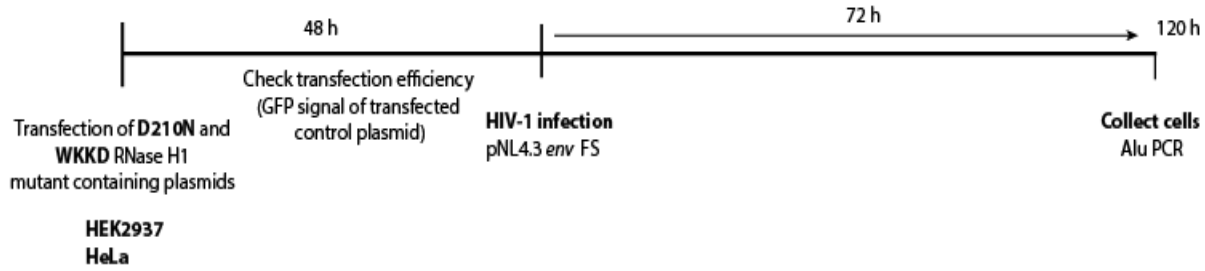


Figure 2.42. Scheme of experimental workflow of RNase H1 mutant D210N and WKKD overexpression and HIV-1 infection.

Cells transfected with either D210N or WKKD RNase H1 mutant plasmid constructs, both containing a GFP reporter gene (used to assess the transfection efficiency) were infected with HIV-1 48 h post transfection, and collected 72 h post infection to determine HIV-1 integration levels via Alu PCR.

Measuring HIV-1 integration upon cell transfection with RNase H1 mutants revealed that HIV-1 integration levels are not affected by the overexpression of these RNase H1 variants (Figure 2.43.).

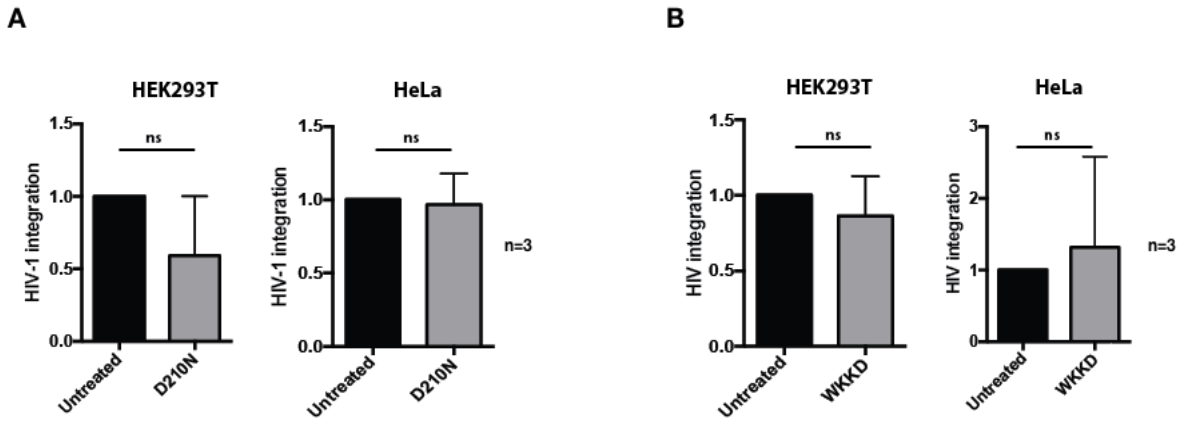


Figure 2.43. Estimation of HIV-1 integration levels upon overexpression of RNAse H1 mutants D210 and WKKD in HEK293T and HeLa cells.

A. HIV-1 Alu PCR results of HEK293T and HeLa cells upon RNase H1 binding mutant D210N overexpression. Alu PCR values in untreated conditions are set to 1 and used as a reference to assess HIV-1 integration upon D210N RNase H1 mutant overexpression. **B.** HIV-1 Alu PCR results of HEK293T and HeLa cells upon RNase H1 binding and catalytic mutant WKKD overexpression. Alu PCR values in untreated conditions are set to 1 and used as a reference to assess HIV-1 integration upon WKKD RNase H1 mutant overexpression. Integrated vDNA (Alu Ct values) is normalized over total genomic DNA levels amplified with primers for the lamin B2 gene (B13 region). The numbers next to the graphs represent numbers of performed repetitions.

The results obtained with the mutant forms of RNase H1 together with the results with the wild type enzyme strongly support the findings that HIV-1 integration depends on the presence of R-loops and that their removal, mediated by the catalytic activity of RNase H1, has a very significant effect on HIV-1 integration efficiency.

II. XVIII. HIV-1 RIGs expression is not affected by R-loop depletion

Taking into account that most R-loops are formed co-transcriptionally (Sanz et al, 2016, Chedin, 2016), and that HIV-1 RIGs have been described as transcriptionally highly active genes (Marini et al, 2015, Lucic et al, 2019), the next question to ask was R-loop depletion by RNase H1 can affect expression profiles of HIV-1 RIGs. In addition, the assessment of RIGs expression profiles upon RNase H1 overexpression might provide an additional explanation to the observed decrease in HIV-1 integration upon R-loop depletion. Mainly because a potential decrease in RIGs expression levels would argue for a scenario in which HIV-1 integration is decreased because of decreased gene expression activity of HIV-1 target genes and not because of the removal of R-loops.

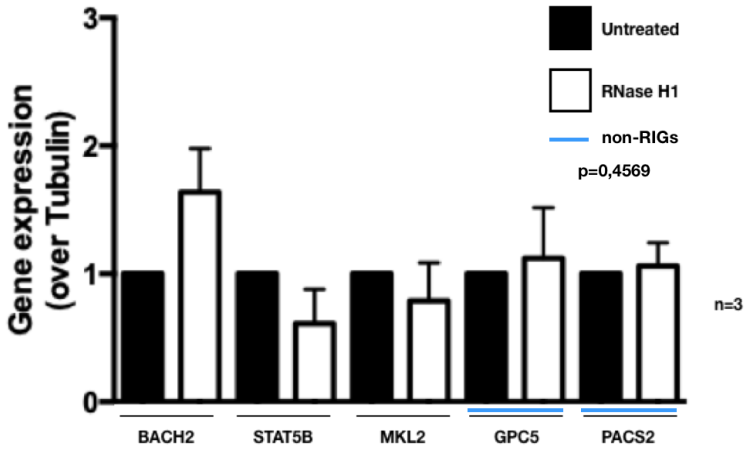


Figure 2.44. Gene expression profiles of selected HIV-1 RIGs and non-RIGs upon RNase H1 overexpression.

Gene expression profiles are generated by RNA extraction upon RNase H1 overexpression, following RT-qPCR analysis. Gene expression levels in untreated conditions are set to 1 and used as a reference to assess gene expression levels upon RNase H1 overexpression. Gene expression levels are normalized over expression profiles of housekeeping gene Tubulin. The slight differences observed in the height of the bars, which are indicating RNA expression levels, are not statistically significant as determined by one-way ANOVA statistical analysis (p value is given in the graph legend). The number next to the graph represents the number of independent donors. The legend in the right upper part of the graph indicates bars corresponding to untreated and RNase H1 overexpressed conditions, as well as which analyzed genes represent RIGs and non-RIGs.

To tackle these questions, RNase H1 was overexpressed in primary $CD4^+$ T cells and the RNA expression profiles of selected HIV-1 RIGs (BACH2, STAT5B and MKL2) and HIV-1 non-RIGs (GPC5 and PACS2) were analyzed by RT-qPCR.

The obtained results show that HIV-1 RIGs, as well as non-RIGs RNA expression profiles are not affected by RNase H1 overexpression and R-loop removal (Figure 2.44.).

The result might be in line with previously reported data stating that high R-loop occurrence does not necessarily mean high transcriptional output (Stork et al, 2016), considering that R-loops also have many regulatory functions and are not exclusively associated to gene expression (Ginno et al, 2012; Chen et al, 2015; Zeller et al, 2016).

However, the obtained result argues for a scenario in which HIV-1 integration is more affected by the presence of R-loops on genes which are highly targeted for HIV-1 integration, than by their expression levels. Such a result is in correlation with, yet unpublished data from our lab, which show that the spatial organization of super-enhancers is more important in the determination of HIV-1 integration sites, than gene expression activity of genes selected for HIV-1 integration (Lucic et al, 2019). These data indicate that the presence of a structure as R-loops, or the spatial conformation of T-cell nuclei has significant impact on HIV-1 integration site selection.

III. Discussion

III. I. Chromatin accessibility and HIV-1 integration

Integration of HIV-1 vDNA into the host genome represents an important step in the life cycle of the virus. Without the successful completion of this process, the virus is not able to replicate and produce viral progeny. At the same time, the viral integration site selection and underlying genomic features of those selected sites can contribute to our understanding on how HIV-1 integration sites contribute to viral transcriptional activity or silencing, which results in latent infection. In addition, a more detailed definition and understanding of the genomic features HIV-1 integration sites can provide a potential platform for future drug treatment strategies contributing to viral eradication from infected cells.

HIV-1 integration can potentially occur anywhere in the host genome. However, the virus shows certain preferences in integration site selection. HIV-1 preferentially integrates into gene bodies of actively transcribed genes (Schröder et al, 2002), especially into introns (Cohn et al, 2015), which are closely associated to Alu elements of the host genome (Stevens & Griffith, 1994). In addition, it has been shown that HIV-1 integration sites are decorated with marks of open/active chromatin (Wang et al, 2007) and delineated with super-enhancers (Lucic et al, 2019), which is in accordance with HIV-1 integration into actively transcribed genomic regions. It has also been shown that the integrated viral genome positions in the nuclear space in a non-random fashion, predominantly in the outer shell of the nucleus, in open chromatin regions, where it also interacts with elements of the NPC, as well as host cell factor LEDGF/p75 (Albanese et al, 2008; Di nunzio et al, 2013; Marini et al, 2015). Considering that the chromatin organization in HIV-1 target cell nuclei has an evident effect on viral integration site selection and viral replication (Lusic & Siliciano, 2017), the first aim of this study was to understand how chromatin composition affects HIV-1 integration? Does integration into open chromatin regions represent a phenomenon occurring just by chance, considering that this is the first chromatin environment the incoming virus encounters when entering the nucleus through the nuclear pore (Schermelleh et al, 2008; Lemaitre & Bickmore, 2015)? Alternatively, can there be an additional (epi)genomic

feature involved in HIV-1 integration site selection in such a nuclear and chromatin surrounding?

To start tackling these questions, primary $CD4^+$ T cells were first treated with drugs affecting the activity of different types of chromatin modifying enzymes, infected with HIV-1 and further on HIV-1 integration levels were assessed by Alu PCR in order to measure eventual changes in HIV-1 integration after changing the chromatin structure. The results showed that all used drugs, SAHA, JQ1 and BIX01294, had an increasing effect on HIV-1 integration levels.
SAHA, as an HDAC inhibitor was expected to cause a global opening of the chromatin structure (Kim et al, 2013; Frank et al, 2016), BET inhibitor JQ1 was expected to cause the unloading of super-enhancers and thereby cause a general chromatin reshuffling (Filippakopoulos et al, 2010; Kumar Bid & Kerk, 2016; Brown et al, 2018; Lucic et al, 2019), and BIX01294 as HMT G9a inhibitor was expected to cause a drop in H3K9me2 facultative heterochromatin mark deposition and increase global chromatin accessibility (Kubicek et al, 2007; Ciechomska et al, 2016).
The results of this experiment already gave an indication that chromatin structure, especially chromatin accessibility, can affect HIV-1 integration. The effect on HIV-1integration of HDAC inhibitor SAHA was very modest and could almost be neglected, and so was that of BET inhibitor JQ1. That JQ1 has no effect on HIV-1 integration profiles was also shown in our recent study, where we sequenced the integration sites using this drug (Lucic et al, 2019). On the other hand, G9a inhibitor BIX01294 had consistently in all the experiments an increasing effect on HIV-1 integration. HMT G9a and with it H3K9me2 chromatin mark, have a well-documented role in T-cell development and differentiation, as they regulate expression of genes important for T cell metabolism (Tachibana et al, 2002; Lehnertz et al, 2014). In addition, the previously described impact of H3K9me2 in chromatin structure and organization (Wen et al, 2009; Kind & van Steensel, 2010; Kind et al, 2013) could be confirmed here through the observed effect of BIX01294 on HIV-1 integration.

According to the assessed kinetics of BIX01294 in primary $CD4^+$ T cells after 6-7 h of 7µM BIX01294 treatment, primary $CD4^+$ T cells are still viable, which is in accordance to previously reported IC_{50} values of BIX01294 (Dhayalan et al, 2009; Ho et al, 2017), and the decrease in H3K9me2 chromatin mark can be appreciated in both immunoblot and immunofluorescence followed by microscopy analysis. Interestingly, STED microscopy revealed the presence of two

different staining patterns of H3K9me2 chromatin mark in the same population of primary $CD4^+$ T cells from the same donor. One staining pattern showed a global nuclear distribution of the chromatin mark, whereas the other staining pattern showed a clear peripheral distribution of the chromatin mark. The peripheral staining pattern of H3K9me2 is more in line with previous findings showing that this chromatin mark is mainly deposited in the nuclear periphery where it also contributes to the organization of LOCKs and LADs (Wen et al, 2009; Kim & van Steensel, 2010; Harr et al, 2015). However, the observation of different chromatin states or patterns in single cells of the same population is not a novel phenomenon (Lieberman-Aiden et al, 2009; Nagano et al, 2013). In addition, it is also well known that different cellular phenotypes or chromatin patterns in the same population of cells represent a consequence of distinct gene expression differences (Elowitz et al, 2002; Blake et al, 2003; Golkaram et al, 2017).

H3K9me2 levels, which showed the strongest drop 12-24 hours post BIX01294 release (after initial 6-7 hours of drug treatment), correlated well with the timing of HIV-1 integration post viral infection in T cells (Mohammadi et al, 2015) when also the increase in HIV-1 integration was observed. 24 hours post drug release H3K9me2 levels start reconstituting back to initial levels, which is underlining the importance of G9a activity in T cells, as well as the establishment of H3K9me2 chromatin mark (Tachibana et al, 2002; Lehnertz et al, 2014). Finally, BIX01294 treatment does not have any detectable effect on related H3K9me3 chromatin mark, as seen by immunoblot and immunofluorescence staining. This result indicates that BIX01294 treatment specifically affects G9a and the deposition of H3K9me2 chromatin mark.

Additionally, cell viability and proliferative potential, assessed by MTT test upon BIX01294 treatment only, or followed by HIV-1 infection showed that upon BIX01294 treatment no significant change in cell viability could be observed, whereas cells upon BIX01294 treatment and HIV-1 infection on average showed a drop of 36% in cell viability. This increased cell death, as detected with the MTT test (determines cellular metabolic activity through the presence of NAD(P)H dependent oxidoreductases) was expected upon HIV-1 infection (Costin, 2007), as well as upon drug treatment and infection (Ivanov et al, 2017; Shytaj et al, 2019).

G9a inhibitor UNC0642, which is supposed to have improved biochemical properties as compared to BIX01294 (Liu et al, 2013b) did not have a reproducibly significant effect on HIV-1 integration. We therefore continued using BIX01294 in the context of HIV-1 integration and

chromatin.

To confirm that the observed increase in HIV-1 integration is a consequence of the drug treatment and its effect on H3K9me2 chromatin mark only, we tested if the drug has any damaging effect on the cells. The assessment of DNA double strand break marker Gamma H2AX upon BIX01294 treatment by IF staining revealed that the drug treatment does not have DNA damaging effects in primary $CD4^+$ T cells, and gave more supporting evidence to the fact that the observed increase in HIV-1 integration upon BIX01294 treatment can be attributed to the changes in chromatin structure and organization mediated my G9a.

The activation status of cells represents an important aspect in HIV-1 infection. As HIV-1 preferentially infects and replicates in activated $CD4^+$ T cells (Margolick et al, 1987), even though integrated HIV-1 DNA can also be detected in resting $CD4^+$ T cells (Chun et al, 1995; Chun et al, 1997). However, activated $CD4^+$ T cells are highly permissive for HIV-1 infection. The process of cell activation has major repercussions on cellular function, gene expression, metabolism, affecting also chromatin organization. Activated $CD4^+$ T cells, as compared to the resting ones, undergo a global increase in chromatin accessibility, such as an increase in H3K27ac levels (Gate et al, 2018; Lucic et al, 2019). Considering the importance of cellular activation on chromatin organization and on HIV-1 infection, it was important to assess whether BIX01294 treatment can eventually influence the activation status of cells, which would consequently affect HIV-1 integration levels. The expression levels of cell surface markers CD38 and CD69 were assessed in primary resting CD4 T cellswhich were treated with BIX01294. Cell surface marker CD38 has a known tight regulation during T cell activation and differentiation (Sandoval-Montes & Santos-Argumedo, 2005; Quarona et al, 2013); CD69 is known as another marker of T cell activation status (Simms & Ellis, 1996). The obtained results show that BIX01294 treatment does not cause a change in cell activation status of resting T cells. Thus, the observed increase in HIV-1 integration upon BIX01204 treatment can not be attributed to enhanced cellular activation status of the cells. High permissiveness to HIV-1 integration can solely be attributed to the changes in chromatin structure. Finally, to control the effect of BIX01294 and exclude possible pleiotropic effects of the drug treatment, a direct knock down of G9a, histone methyltransferase that deposits H3K9me2 was performed followed by HIV-1 infection in Jurkat T cell line. The results demonstrated that direct G9a knock down significantly

depleted H3K9me2 chromatin mark, as expected. Similarly, an increasing effect on HIV-1 integration was observed, providing an important control and supporting the use of BIX01294 in the following study.

In conclusion, all performed controls regarding the effect of BIX01294 treatment on primary $CD4^+$ T cells contribute to a very detailed understanding of the drug's effect on $CD4^+$ T cells. The results provide strong evidence that the observed phenomenon of increased HIV-1 integration levels upon drug treatment can only be attributed to the changes in H3K9me2 chromatin levels and represent a consequence of increased global chromatin accessibility.

The effect of BIX01294 on increased HIV-1 integration was also followed by increased viral activity, as assessed by the production of viral RNAs and viral p24 protein. This finding was not surprising considering that more viral DNA was successfully integrated into the host genome due to a generally more open chromatin structure upon BIX01294 treatment, which might also have a stimulating effect on viral RNA production. It is in fact well established that the chromatin environment can influence basal HIV gene transcription (Jordan et al, 2001; Lusic & Giacca, 2014; Turner & Margolis, 2017). In addition, the assessment of total vDNA levels upon BIX01294 treatment and HIV-1 infection showed that viral entry into cells is not affected by the treatment and that identical levels of total viral DNA are found in cells upon drug treatment. Moreover, a decrease in 2LTR circles, episomal viral forms formed due to inefficient integration, and a concomitant increase in HIV-1 integration suggest that nuclear entry is not affected by BIX01294 treatment. Still, more copies of the viral genome integrated into the cellular chromatin due to its increased accessibility.

HIV-1 provirus positions non-randomly in primary $CD4^+$ T cell nuclei in respect to structural elements of the cellular chromatin and the NPC (Albanese et al, 2008; Guelen et al, 2008: Di Nunzio et al, 2012; Di Nunzio et al, 2013; Marini et al, 2015). It was therefore important to assess the effect of BIX01294 treatment and H3K9me2 depletion on HIV-1 provirus spatial positioning in primary $CD4^+$ T cell nuclei, especially considering that H3K9me2 represents a chromatin mark which mainly decorates chromatin in the nuclear periphery (Kind & van Steensel, 2010). The performed immuno-DNA FISH results in this context showed a clear repositioning of the provirus from the nuclear periphery towards the nuclear center. Apart from the fact that the integrated virus occupies a new nuclear environment upon BIX01294 treatment,

the increase in HIV-1 integration levels upon drug treatment could also be visually appreciated (Figure 2.12.).

Considering that HIV-1 integration genes occupy certain nuclear areas which are excluded from LADs, but still in the outer shell of the nucleus (Marini et al, 2015; Lucic et al, 2019), it became relevant to assess their nuclear position upon BIX01294 treatment. Obtained immuno-DNA FISH results of HIV-1 RIGs upon BIX01294 treatment clearly indicate that the majority of analyzed genes follow the behavior of the provirus and change their nuclear position, from peripheral towards inner nuclear zones. In addition, considering chromosomal positions of RIGs, this result indicates that RIGs are not just clustering together on chromosomes (Marini et al, 2015), but could also potentially move in clusters upon H3K9me2 depletion, as recently observed in the case of transition from resting to activated T cell status (Lucic et al, 2019).

Focusing only on HIV-1 provirus spatial positioning in untreated conditions, results from previous studies could be confirmed, demonstrating that the provirus predominantly positions in the nuclear periphery (Albanese et al, 2008; Di Nunzio et al, 2012; Di Nunzio et al, 2013; Marini et al, 2015; Lucic et al, 2019). However, a recent study suggested a different scenario, according to which integrated HIV-1 preferably occupies central areas of cell nuclei (Achuthan et al, 2018). The study however differs in several points from here presented results. First of all, the amounts of the virus the authors used in their study exceeds significantly the more physiological conditions which we aimed in maintaining in this and in a parallel study by Lucic et al, 2019. Second, most of their analysis was performed on HeLa and U20S cells, which are not primary targets of HIV-1. This could account for the observed differences in integration patterning and gene position, as HIV-1 integration sites are delineated with super-enhancers which mainly control the expression of cell identity genes (Whyte et al, 2013; Lucic et al, 2019). Mentioned findings indicate that cell type specificity seems to be another significant determinant of HIV-1 integration site selection and proviral spatial positioning in the nucleus apart from chromatin composition, as demonstrated by the results of this work.

The analysis of RIGs expression levels upon BIX01294 treatment was performed in order to test if changed chromatin composition and changed nuclear position could affect their expression levels. A potential increase in RIGs expression could give an additional explanation for increased HIV-1 integration upon drug treatment considering also previous studies which demonstrated that

HIV-1 predominantly integrates into active transcription units (Han et al, 2014; Cohn et al, 2015).

However, RNA microarray profiling upon BIX01294 treatment showed that RIGs transcriptional profiles are not significantly affected upon drug treatment. Possibly because these genes are already highly expressed in T cells (Marini et al, 2015; Lucic et al, 2019). However, two groups, one of downregulated and the other of upregulated genes could be distinguished upon BIX01294 treatment. According to the performed GO term analysis, in the group of downregulated genes, genes involved in membrane trafficking, protein and ion binding are dominating, whereas in the group of upregulated genes, genes involved in iron metabolism and inflammatory response are overrepresented. It is known that G9a activity (which is inhibited by BIX01294 treatment) is pivotal in cell proliferation and differentiation (Lehnertz et al, 2010; Rao et al, 2016). In addition, it has been shown that G9a impairment by BIX01294 treatment can reduce cellular proliferation rates or cause cell cycle arrest (Shankar et al, 2013), whereas iron is vital for cell proliferation in T cells and is involved in pathways protecting cells from cytotoxicity, oxidative stress and DNA damage (Bowlus, 2003; Ruttkay-Nedecky et al, 2013). The results might indicate an ongoing compensatory effect, where iron and inflammatory pathways are activated in order to prevent cell cycle arrest.

In order to prove whether BIX01294 treatment causes cell cycle arrest in primary $CD4^+$ T cells, cell cycle analysis was performed. The obtained FACS profiles upon propidium iodide staining of cells show that BIX01294 treatment does not cause any changes in cell cycle profiles of primary $CD4^+$ T cells. This result is in accordance with previous findings in which H3K9me2 depletion caused by direct G9a knock down did not affect cell cycle profiles and cell proliferation (Yokochia et al, 2009).

In respect to the fact that BIX01294 is a drug that affects the activity of HMT G9a, and consequently changes the deposition of H3K9me2 chromatin mark, which might alter the distribution of other histone marks or chromatin binding factors, ChIP-qPCR was performed in order to verify whether any type of changes can be appreciated in that regard. The results show that upon BIX01294 treatment the most prominent changes, in the context of chromatin mark binding, can be detected in the first intron region of analyzed genes. However, a general drop throughout analyzed RIGs and non-RIGs in H3K9me2 can be detected, which is not surprising

considering the known effect of BIX01294. Also, H3K9me3 chromatin mark distribution was not significantly affected by the treatment, which is in line with previously obtained immunoblot and IF results. However, a slight decrease of H3K9me3 binding was appreciated in the gene bodies of analyzed genes, which is not surprising considering the very strong effect of BIX01294 on related H3K9me2 chromatin mark and the high sensitivity of the used antibodies (Rothbart et al, 2015). The distribution of H3K36me3 and LEDGF/p75 upon BIX01294 treatment was not significantly affected in regions of gene bodies or gene ends. Main differences were observed for H3K36me3 and LEDGF/p75 binding in the first intron region of analyzed genes, but this requires further confirmation at the global genome-wide level. H3K36me3 is a chromatin mark which is mainly associated to gene bodies and represents a signal of elongating Pol II transcription (Venkatesh et al, 2012). However, a slight increase in H3K36me3 binding was detected in gene ends of MKL2 and CNTN4. This finding might be connected to some alternative implications of this chromatin mark, considering that it can also be linked alternative splicing, dosage compensation, DNA replication and repair (Wagner & Carpenter, 2012; Suzuki et al, 2016). Interestingly, an increased LEDGF/p75 binding was detected only on RIGs upon BIX01294 treatment, and could not be detected on non-RIGs. One can presume that LEDGF/p75 cannot bind to non-RIGs in the first intron region due to the presence of H3K9me3, which was detected in first introns of analyzed genes. Considering that H3K9me3 is a repressive chromatin mark that induces the compaction of chromatin it might also prevent LEDGF/p75 binding. LEDGF/p75 is a transcriptional mediator protein that promotes activator-dependent transcription and is binding to active transcription units (Ge et al, 1998a; De Rijck et al, 2010; Craigie & Bushman, 2012). However, the fact that LEDGF/p75 binding is following H3K36me3 enrichment on RIGs is not very surprising considering that LEDGF/p75 is a known reader of this particular chromatin mark (Pradeepa et al, 2012).

Furthermore, the observed enrichment of H3K36me3 and LEDGF/p75 in the first intron regions of RIGs can possibly be associated to the observed increased integration of HIV-1 upon BIX0124 treatment and the fact that HIV-1 preferentially integrates into introns of expressed genes (Craigie & Bushman, 2012; Han et al, 2014; Cohn et al, 2015).

In addition, taking into account the obtained RNA microarray results which suggest that RIGs gene expressions are not affected by the drug treatment, meaning that RIGs are under all circumstances highly expressed in T cells (Marini et al, 2015), argues for a scenario in which

LEDGF/p75 binding is always facilitated on RIGs. This possibility might be explained by the fact that LEDGF/p75 is known to bind active transcription units, located downstream of transcription start sites (De Rick et al, 2010).
Interpretation of increased HIV-1 integration in the context of the obtained ChIP-qPCR results points further to the necessity of performing a global, genome-wide analysis of chromatin features upon BIX01294 treatment.

HIV-1 immuno-DNA FISH results upon BIX01294 treatment revealed a changed proviral nuclear position and gave a visual proof of increased HIV-1 integration upon drug treatment. These data also pointed to the possibility that HIV-1 might target alternative genomic sites for integration upon H3K9me2 removal. Consequently, HIV-1 integration site sequencing was performed. The initial statistical analysis of retrieved HIV-1 integration sites upon BIX01294 treatment, compared to integrations obtained after infection of $CD4^+$ T cells from the same donor under non-treated conditions confirmed a slight increase in HIV-1 integration. However, the number of retrieved integration sites did not fully correlate to HIV-1 integration levels measured by Alu PCR. This discrepancy might be explained by general underestimating capacities of integration site sequencing approaches (Cui et al, 2016; Hughes & Coffin, 2017).
The analysis of HIV-1 integration patterns showed that HIV-1 still maintains its integration patterns into active genes, probably marked by H3K36me3 and bound by LEDGF/p75. In addition, the analysis revealed that the *in vitro* generated dataset of HIV-1 integration sites from this study can be separated from other HIV-1 integration site datasets where different sequencing methods were used (Ikeda et al, 2007; Brady et al, 2009; Wagner et al, 2014; Maldarelli et al, 2014; Kok et al, 2016). This could be due to different library preparation methods, as well as different sequencing methods used in our and different studies to retrieve HIV-1 integration sites. However, based on a prediction that was recently made, HIV-1 integrations seem to be non saturating, meaning that as many times as integration site sequencing is performed, new integrations can always be detected (Lucic et al, 2019). It is of course worth noting that some genomic regions, which can be defined as hot spots or recurrent integration genes do appear in both, untreated and treated conditions.
Although we did not detect a clear change in the integration patterns, it has to be pointed out that a more detailed chromatin patterning under BIX01294 conditions is required. Only by matching

BIX01294 integration sites with the respective chromatin profiles can one be absolutely sure that there are no differences in integration patterns between control and treated conditions. Similarly, the mentioned analysis will be crucial in making a final conclusion about the relevance of chromatin for HIV-1 integration. One can also take into account the possibility that there are other, yet undefined genomic features that contribute to the constant targeting of identical integration sites by the virus. This possibility was investigated in the second part of this study.

Overall, the results of the first part of this study show that chromatin composition and accessibility significantly affect HIV-1 integration levels. In the obtained dataset, G9a inhibition by BIX01294, and the consequent H3K9me2 depletion had the most prominent effect on HIV-1 integration levels. How the manipulation of other chromatin patterns might influence HIV-1 integration and integration site selection remains to be explored. Still, this part of the work underlines the importance of the chromatin state in target cells determining the fate of the invading virus, as well as infected cells, as already proven by other studies (Archin et al, 2009; du Chene et al, 2007; Gijsbers et al, 2011; Wagner et al, 2014; Cohn et al, 2015; Marini et al, 2015; Singh et al, 2015).

III. II. R-loops as genomic sites of HIV-1 integration

The investigation of HIV-1 integration from a chromatin related perspective, especially determining the effects of HMT G9a activity and deposition of H3K9me2 chromatin mark, revealed that chromatin accessibility significantly affects HIV-1 integration levels.
However, in this work, HIV-1 integration, or more precisely, HIV-1 integration site selection was also investigated from a different perspective, taking into account the intrinsic and poorly explored property of HIV-1 integrase, which is responsible for HIV-1 integration into the cellular chromatin. Hence, the second part of this thesis focuses on the viral IN and its catalytic and chromatin binding properties to R-loops, as special genomic features selected for HIV-1 integration.

The reasoning for this perspective on HIV-1 integration site selection arises from the fact that the CCD of HIV-1 IN exposes an RNase H folding pattern, and when *in silico* superimposed to the RNase H1 enzyme of *Bacillus halodurans* in complex with a DNA-RNA hybrid, a perfect match of the active residues of both enzymes can be appreciated (Goldgur et al, 1999; Nowotny et al, 2005; Savarino, 2007). Respectively, RNase H enzymes specifically resolve DNA-RNA hybrid structures (Eder et al, 1993; Cerritelli & Crouch, 2009). In addition, it has been shown that the CTD of HIV-1 has the ability of binding structured RNA elements (Kessl, 2016). Moreover, when comparing the known features defining HIV-1 integration sites with R-loop forming regions, a clear overlap can be observed. HIV-1 integration sites have been described as sites of transcriptionally active genes with increased DNaseI hypersensitivity (Verdin, 1991; Brady et al, 2009; Busschots et al, 2005; Ciuffi et al, 2005; Chen et al, 2017a), and with high G/C content and. On the other hand, R-loop regions represent sites of DNaseI hypersensitivity (Ginno et al, 2012, 2013; Sanz et al, 2016) and are formed mainly co-transcriptionally in highly transcribed G/C rich genomic regions (Ginno et al, 2012, 2013; Sanz et al, 2016). Intriguingly, another finding which might possibly connect R-loops to the first part of the study, is the fact that the distribution of H3K9me2 also seems to affect the formation of R-loops (Zeller et al, 2015; Skourti-Stathaki et al, 2014). However, in the second part of this work, the focus was exclusively directed to R-loops, as genomic structures, and their possible importance in HIV-1 integration.
First supporting evidence that R-loops could possibly represent sites for HIV-1 integration, arose

from computational metadata analysis in which R-loop maps (Sanz et al, 2016) were overlapped with lists of HIV-1 integration sites (Cohn et al, 2015; Kok et al, 2016), showing that R-loop forming regions tend to be overrepresented at sites of viral integration. Further analysis, in which R-loop maps from K562 cells (Sanz et al, 2016) and HIV-1 RIGs (Lucic et al, 2019) were overlapped, showed that 83% of HIV-1 integration genes (RIGs) had 3 or more R-loops on them. This feature clearly distinguishes HIV-1 integration genes from HIV-1 non-target regions. According to these preliminary analysis, we reasoned that studying R-loops as potential target sites for HIV-1 integration could provide a more detailed insight into HIV-1 integration site selection and lead to a more detailed description of HIV-1 integration sites.

In the first experimental analysis, we tested the colocalization between HIV-1 PICs and R-loops using an engineered viral construct carrying an eGFP tag on the C-terminus of viral IN, which enabled the detection of viral PICs (Albanese et al, 2008). Cells were collected at early time points post infection to increase the possibility to observe PIC docking to certain nuclear areas. According to this analysis in primary $CD4^+$ T cells 75% of analyzed PICs colocalized with R-loops, which was followed by the observation of a similar trend in macrophages where 72% of analyzed PICs colocalized with R-loops. This finding suggests that the vast majority of PICs in HIV-1 target cells are docked to R-loop forming sites. This colocalization was also confirmed by HIV-1 DNA immuno-FISH in macrophages at different time points post infection.

To further address the observed colocalization and to verify whether the high occurrence in PICs colocalizing with R-loops represents an active process and not just a stochastic observation, macrophages were either infected with the same viral construct in the presence of IN strand transfer inhibitor Raltegravir, or with a construct containing a mutation at the active residue D116N of the CCD of IN. The obtained colocalization results show that in the case of HIV-1 D116N mutant, only 31,1% of PICs colocalize with R-loops, which represents a drop of over 41%. In the presence of Raltegravir the effect is even more pronounced, and only 21,45% of PICs colocalize with R-loops representing a drop of more than 50%. Raltegravir is known to inhibit the strand transfer activity of HIV-1 integrase. The observed drop in colocalization with R-loops after Raltegravir treatment points to the fact that strand transfer reaction might be important to keep HIV-1 IN docked to R-loops. The results indicate that for a high occurrence of PIC and R-loop colocalization the presence of a functionally active IN is required. In addition, the results

provide evidence that colocalization between HIV-1PICs and R-loops is not a randomly occurring process.

Considering previous reports that the CTD of HIV-1 has RNA binding properties and that lysine residues 264, 266 and 273 have been defined as key players in that process (Kessl et al, 2016), assessing colocalization between R-loops and a HIV-1 variant containing mutations in those residues would possibly strengthen the hypothesis of HIV-1 IN binding to R-loops. Before performing the colocalization assessment, the activity and the behavior of such a virus had to be determined, considering that such a viral construct represented a newly obtained tool in the lab. Viral cDNA production, integration and RNA production capacities were determined in primary $CD4^+$ T cells and Jurkat cells. The obtained results show that the HIV-1 K264,266,273Q virus is able to produce viral DNA, but incapable of integrating its genome into the host cell genome, and accordingly does not produce viral RNA, as seen upon infection of both cell types. The results regarding viral integration are in accordance with previous findings showing that the acetylation of these three lysine residues by p300 histone acetyltransferase contributes to more efficient viral integration, and once these residues are mutated there is an evident drop in viral integration efficiency (Cereseto et al, 2005). The obtained results show that viral reverse transcription, as well as PIC assembly, are not affected by the presence of the three mutated residues, which enabled us to continue with a colocalization analysis of HIV-1 K264,266,273Q viral PICs and R-loops. Primary $CD4^+$ T cells and macrophages were infected with the virus and the colocalization with R-loops was assessed. The obtained results show that viral PICs either remain trapped in the nuclear lamina, or once inside the nucleus, do not colocalize with R-loops. In primary $CD4^+$ T cells, 88% of PICs are found in the nuclear lamina, whereas 12% of the signals can be detected inside of the nucleus, but outside of R-loop forming sites. A very similar scenario was observed for macrophages in which 71% of PICs were detected in the nuclear lamina, and 29% inside the nucleus, but also away from R-loop forming regions. The results clearly show that the PICs of the analyzed C-terminal mutant of HIV-1 IN are not docked to R-loop forming genomic regions, which represents another indication that only wt HIV-1 PICs, without the impairment of IN activity, are successfully docked to R-loops. On the other hand, what contributes to the phenomenon of HIV-1 K264,266,273Q PICs remaining trapped in the nuclear lamina, remains elusive at this point and would require further investigation.

After obtaining microscopic evidence that HIV-1 IN and R-loops are colocalizing in HIV-1 target cells, that for the occurrence of this process a catalytically active IN is required, indicating that PICs are actively docked to genomic R-loops, further attention was directed to another result that arose from the initial metadata analysis. Apart from a high correlation between HIV-1 integration sites and genomic R-loop sites, which was predicted by the analysis, the *in silico* analysis also indicated that most HIV-1 RIGs contain more than one R-loop forming region. In order to prove whether RIGs are enriched in R-loops, DRIP-qPCR (Gino et al, 2012; 2013) was performed on primary $CD4^+$ T cells. The results of this analysis demonstrate that selected RIGs, BACH2 and STAT5B, are enriched in R-loops, which confirms the *in silico* prediction. In addition, considering that BACH2 and STAT5B are highly expressed in T cells, as well as genes of high G/C content (Brady et al, 2009; Marini et al, 2015), such an outcome might not be surprising. However, those genes are found on the vast majority of HIV-1 integration site list (Maldarelli et al, 2014; Wagner et al, 2014; Cohn et al, 2015; Kok et al, 2016; Lucic et al, 2019), which supports the hypothesis of R-loops representing genomic sites of HIV-1 integration.

The obtained DRIP-qPCR result raised the question whether integrated viral DNA can be situated in R-loop forming sites. In order to tackle this question, a modified approach of the initial DRIP-qPCR procedure was applied on HIV-1 infected primary $CD4^+$ T cells. Instead of directly performing the qPCR upon the DRIP procedure, Alu PCR was performed in order to try detecting integrated viral DNA (Tan et al, 2006) in R-loop rich genomic regions that have previously been pulled down in the DRIP procedure. The obtained DRIP-Alu PCR results show that integrated viral DNA can be detected in R-loop rich genomic regions. Apart from representing another confirmation of the initial computational prediction, this result shows that R-loops represent genomic sites that can accommodate the integrated viral genome. Considering evident similarities between HIV-1 integration sites (Verdin, 1991; Brady et al, 2009; Busschots et al, 2005; Ciuffi et al, 2005; Chen et al, 2017a) and R-loop forming sites (Ginno et al, 2012, 2013; Sanz et al, 2016), this finding provides supporting evidence to the hypothesis that R-loops represent sites of HIV-1 integration and potential HIV-1 IN binding.

Attempting to further strengthen this hypothesis, and prove whether HIV-1 IN can directly interact with R-loops, *in vitro* binding and strand transfer assays of HIV-1 IN in complex with R-loop templates were performed (with the help of Dr. Vincent Parissi from the University of

Bordeaux). The performed *in vitro* binding assays show that HIV-1 IN can bind to nucleic acid substrates containing R-loops, even more efficiently compared to nucleic acid substrates without R-loops, as well as exposing higher binding affinity to nucleic acid substrates containing R-loops compared to non-R-loop containing DNA substrates. The obtained results might primarily not be very surprising considering the RNase H fold of the CCD of HIV-1 (Yang et al, 1990; Majorek et al, 2014). However, it represents a finding that is clearly in line with the obtained microscopy data showing that HIV-1 PICs are docked to R-loops and provides additional prove to the concept of defining R-loops as sites of HIV-1 integration.

The results of the *in vitro* strand transfer reaction performed on R-loop and non-R-loop containing nucleic acid templates showed that HIV-1 IN can perform the strand transfer reaction on R-loops, even slightly more efficient than on non-R-loop containing DNA templates. The results of this set of experiments also revealed that the putative RNase H domain, still present in HIV-1 IN, is not enzymatically functional, and that the strand transfer reaction is even more efficiently completed on R-loop containing DNA templates once the R-loop is resolved in the presence of an RNA helicase. The finding that HIV-1 IN does not have RNase H1 function, although its CCD has an RNase H-like fold is not very surprising considering that a certain protein fold, or folding pattern of a protein domain does not necessarily imply certain enzymatic function. In accordance with that is the fact that HIV-1 IN is classified into the RNHL (ribonuclease H-like) superfamily of proteins, which is composed of a large group of evolutionary related, but functionally diverse group of proteins (Majorek et al, 2014). For instance, when performing protein sequence alignment and similarity analysis, by using NCBI's BLASTp tool (Altschul et al, 1997), between HIV-1 RT, with known RNase H function, and the CCD of HIV-1 no sequence similarity can be found. On the other hand, the finding that HIV-1 IN can perform the strand transfer reaction on DNA templates containing R- loops is in accordance with previous findings showing that the structure of target host DNA greatly influences the site of viral integration as well as that DNA curvature and flexibility influences the frequency of integration (Pruss et al, 1994a; Pruss et al, 1994b; Wang et al, 1999; Michieletto et al, 2019). Also, the result of successful strand transfer on R-loop containing DNA templates is in line with previous findings showing that HIV-1 IN shows preference for binding structured RNA elements (Kessl et al, 2016), considering that prior to the strand transfer reaction on the R-loop, binding of

the R-loop has to occur.

The finding that the strand transfer reaction is even more efficiently completed on R-loop containing nucleic acid templates once the R-loop is resolved can possibly be explained by the fact that HIV-1 IN and RNase H1 are enzymes which are interacting, and most probably competing for the same substrate (Lusic Lab, unpublished data). In such a scenario, most probably both, HIV-1 IN and RNase H1 bind to the R-loop; and while RNase H1 resolves the R- loop structure, IN performs the strand transfer reaction.

Further on, upon obtaining mentioned series of molecular data indicating that RIGs are enriched in R-loops, that integrated viral DNA can be situated in R-loop forming regions and that HIV-1 IN can *in vitro* bind to R-loops and perform the strand transfer reaction on them, the focus was again directed to HIV-1 integration in an R-loop dependent context. In order to investigate how the presence of R-loops affects HIV-1 integration, viral integration was assessed upon RNase H1 overexpression and R-loop depletion in cell lines (HeLa, HEK293T, Jurkat), as well as primary cells ($CD4^+$ T and macrophage). The results of this experiment show that HIV-1 integration is significantly impacted in different cell lines and primary cells. Alu PCR results show that HIV-1 integration is decreased upon RNase H1 overexpression, throughout all analyzed groups of cells, and almost abrogated in Jurkat cells and primary macrophages. The observed decrease in integration is followed by an increasing trend of 2LTR circle formation, which is expected considering that integration is less efficient, or even abrogated in some cases. In addition, the final prove that R-loop removal is the main reason for decreased viral integration upon RNase H1 overexpression was provided by absolute quantifications of early and late HIV-1 RT-products, as well as 2LTR circles by ddPCR (Bejarano et al, 2018) in primary $CD4^+$ T cells. The obtained results showed that the production of both early and late RT-products is not affected by RNase H1 overexpression. Accordingly, the results indicate that viral reverse transcription process is not impaired by RNase H1 overexpression, and that the observed decrease in HIV-1 integration is an exclusive consequence of R-loop depletion. In addition, the absolute quantification of 2LTR circles in untreated conditions and upon RNase H1 overexpression showed a slight increase in the numbers of formed 2LTR circles upon RNase H1 overexpression, which is in correlation with the observed decrease in viral integration and previous quantification of 2LTR circles by qPCR. This result indicates that more non-integrated viral forms are present as a consequence of less efficient

integration upon R-loop depletion. Further on, this results also represents a confirmation that RNase H1 only gets expressed in the nucleus of the cell, considering the M27 NLS of the plasmid construct used for protein overexpression (Suzuki et al, 2010), and that viral RT-products cannot be exposed to the effect of RNase H1 since the protein seems to have a nuclear localization, an RT is considered to be initiated in the cytoplasm (Francis et al, 2016; Rankovic et al, 2017; Lusic & Siliciano, 2017).

In order to further validate that only R-loop removal represents the cause of the observed decrease in HIV-1 integration, the same experimental setup was performed on selected cell lines (HeLa and HEK293T), but using mutated variants of RNase H1. An RNase H1 binding mutant (D210N), as well as an RNase H1 binding and catalytic mutant (WKKD) were used for this purpose, which have also been used in previous studies for the purpose of R-loop pull down approaches (Chen et al, 2017a). Obtained Alu PCR results upon overexpression of the two mentioned RNase H1 mutants in cell lines showed that HIV-1 integration is not affected by their presence, which provided proof that the previously observed decrease in HIV-1 integration represents a consequence of R-loop depletion.
The results from this experiment indicate that R-loop presence is required for efficient HIV-1 integration, which is in accordance with the *in vitro* data of IN binding and strand transfer activity on R-loops, as well as the finding of integrated viral DNA in R-loop forming regions. These results together argue for the possibility of describing R-loops as sites of HIV-1 integration.

However, the observation that R-loop depletion causes a decrease in HIV-1 integration led to a question if gene expression levels are affected by RNase H1 overexpression and whether the decrease in viral integration could be due to a general drop in gene expression levels rather than a consequence of R-loop removal. These questions were especially important to respond considering that R-loops are mainly formed co-transcriptionally (Ginno et al, 2102; 2013) and that HIV-1 preferentially integrates into actively transcribed genes (Schröder et al, 2002, Lucic et al, 2019). In order to test for this option, RNA expression profiles of selected RIGs and non-RIGs were analyzed by RT-qPCR in primary $CD4^+$ T cells upon RNAse H1 overexpression. Obtained results show that both RIGs (BACH2, STAT5B, MKL2) and non-RIGs (GPC5, PACS2) expression profiles are not affected by RNase H1 overexpression. In the context of non-RIGs,

this result might not be very surprising considering that those are genes mainly poorly expressed in T cells (Marini et al, 2015). In the context of RIGs, this result argues for a scenario in which transcriptional activity of HIV-1 target genes is a less favored feature in the context of HIV-1 integration, compared to the presence of an R-loop structure. Such an outcome would conceptually be in line with the finding that the nuclear organization of T cells is more important in dictating HIV-1 integration site selection, compared to only gene expression levels (Lucic et al, 2019). In this context, the obtained results indicate that the structural organization of the genome, together with the presence of certain genomic structural elements and their interplay with host cell factors impact HIV-1 integration and integration site selection more than any other factor described to influence HIV-1 integration.

III. III. Perspectives and concluding remarks

The first part of this study clearly shows that HIV-1 integration is significantly affected by chromatin accessibility. Series of experimental data show that the use of chemical G9a inhibitor BIX01294 and the depletion of H3K9me2 chromatin mark in primary $CD4^+$ T cells have an increasing effect on HIV-1 integration, indicating that the process of HIV-1 integration and integration efficiency can not be separated from (epi)genomic factors such as chromatin composition.

In addition, the results from this part of the study also suggest that when manipulating with the chromatin composition of a certain chromatin mark, in order to appreciate discrete changes in the chromatin distribution of other chromatin binding factors at certain genomic sites, a deeper and more targeted analysis of selected genes or genomic sites will provide more information, compared to general global scale analysis.

Of note, the finding that chromatin accessibility in the context of H3K9me2 histone mark significantly affects HIV-1 integration becomes especially important considering that BIX01294 is a known and very strong HIV-1 latency reversing agent (Imai et al, 2010; Bouchat et al, 2012; Nguyen et al, 2017). Accordingly, future studies on HIV-1 latency reversing agents which are manipulating chromatin structure composition and accessibility should be conducted carefully and include the factor of possible increase in HIV-1 integration efficiency.

The final conclusion of the second part of this study is that the presence of R-loops, as structural genomic elements, in HIV-1 integration genes, is of more importance for efficient viral integration compared to high transcriptional activity of HIV-1 target genes.

Nevertheless, it would still be very important to assess how RNase H1 overexpression affects gene expression on a global scale. Also, it would be of significant importance to understand and distinguish the influence of co-transcriptional R- loops (Ginno et al, 2012; 2013) and regulatory R-loops (Yu et al, 2013; Santos-Pereira & Aguilera, 2015; Grunseich et al, 2018) in the context of HIV-1 integration. Still, in order to further strengthen the concept of R-loop structures as sites for HIV-1 integration it would be of significant importance to uncouple the importance and

presence of R-loops in the process of transcription (Skourti-Stathaki et al, 2011; Skourti-Stathaki et al, 2014), from the fact that R-loops are also structural genomic elements that evidently influence HIV-1 integration. In line with the concept of R-loops as structural genomic elements possibly affecting a plethora of other biological processes, are findings indicating that transcriptional output does not seem to be correlated to R-loop levels on individual genes (Stork et al, 2016; Wahba et al, 2016).

Overall, the results of the second part of this work provide strong evidence that R-loops, as structural genomic features, represent sites of HIV-1 IN binding and, accordingly, HIV-1 integration. The initially obtained computational results provided an accurate prediction on HIV-1 RIGs as genomic sites enriched in R-loops, which was further on confirmed by molecular approaches in cells. In addition, R-loops were shown to be genomic sites having the capacity to accommodate integrated HIV-1 DNA. In addition, microscopic data on HIV-1 PICs colocalization with R-loops provided first indication that HIV-1 IN is docked to R-loop forming genomic sites. In respect to the entire dataset on microscopy, it was demonstrated that for high occurrence of PIC and R-loop colocalization a functionally active viral IN is required, confirming that the observed colocalization does not represent a randomly occurring event. Further on, HIV-1 IN binding on R-loop nucleic acid templates was confirmed *in vitro*, as well as the capability of HIV-1 IN to perform the strand transfer reaction on R-loop containing DNA templates. In line with that are obtained results showing that R-loop depletion in cells has a decreasing effect on HIV-1 integration. And finally, R-loop depletion was shown not to affect transcriptional activity of selected RIGs, indicating that the presence of R-loops, as structural genomic elements, on these genes is more important in affecting HIV-1 integration, compared to their transcriptional activity.

However, the *in vitro* dataset also demonstrated that HIV-1 IN strand transfer is more efficient at R-loop forming sites once the R-loop is resolved by RNase H1. Studying the interplay between IN and other R-loop resolving enzymes (as Aquarius, Senataxin, Topoisomerases, *etc*) in the cell seems to be crucial in understanding HIV-1 integration in an R-loop dependent context, and gives rise to the need of further investigations.

Considering that the focus of the second part of this study was directed towards R-loops as structural genomic elements, without implying their regulatory aspects, it is important to mention the interplay between R-loops and chromatin, considering that the first part of this work predominantly deals with the importance of chromatin composition related to HIV-1 integration. Several studies showed the association of R-loop peaks and open chromatin marks, as well as overall increased chromatin accessibility in R-loop rich regions, mainly occurring at gene promoters or gene bodies (Zeller et al, 2016, Sanz et al, 2016, Chen et al, 2017a). Presumably, such findings could represent another connection between R-loops and HIV-1 integration sites considering that HIV-1 preferably integrates into gene bodies of actively transcribing genes, carrying open chromatin marks (Schröder et al, 2002; Wang et al, 2007; Cohn et al, 2015). However, gene terminal R-loops where shown to show features of transcription terminators (Sanz et al, 2016). Also, in a variety of different organisms it was shown that R-loop formation can be connected to RNA Pol II termination, as well as the establishment of heterochromatin (Castellano-Pozo et al, 2013; Skourti-Stathaki et al, 2014, Boque-Sastre et al, 2015; Sanz et al, 2016; Skourti-Stathaki et al, 2019). In summary, what appears evident from mentioned studies is that the interplay between chromatin and R-loops, as well as its outcome in a context of chromatin accessibility and genomic activity significantly depends on the region of R-loop formation. Respectively, if R-loop formation occurs at gene TSS, TTS, gene bodies, before or after replication forks significantly affects chromatin accessibility and *vice versa* (Santos-Pereira & Aguilera, 2015; Chedin, 2016).

In conclusion, both parts of this study provide strong evidence that chromatin composition, chromatin accessibility mediated by the deposition of H3K9me2 mark, as well as structural genomic elements - R-loops, significantly impact HIV-1 integration. Both aspects of this study provide more detailed description of HIV-1 integration sites, HIV-1 integration efficiency, as well as HIV-1 integration site selection.

IV. Materials and methods

IV.I. Cell lines and cell culture

Human lymphoblastiod cancer Jurkat cell line and latent HIV-1 clone J-Lat9.2 was kept in RPMI medium supplemented with heat inactivated 10% FBS and 1% Penicillin/Streptomycin.
Human epithelial HEK293T and HeLa cell lines were kept in DMEM medium supplemented with 10% FBS and 1% Pen/Strep antibiotics.

IV.II. Primary cell isolation

$CD4^+$ T cells

Primary $CD4^+$ T cells were purified from whole blood from healthy donors with RosetteSep™ Human $CD4^+$ T Cell Enrichment Cocktail (STEMCELL TECHNOLOGIES). Cells were activated with Dynabeads Human T-Activator CD3/CD28 (ThermoFisher) and 5 ng/mL IL-2 for 48 h and kept in RPMI medium supplemented with heat inactivated 10% FBS and 1% Pen/Strep antibiotics.

Monocyte derived macrophages (MDMs)

Human peripheral blood mononuclear cells (PBMCs) were isolated from buffy coats of healthy blood donors by Ficoll density gradient centrifugation. PBMCs were seeded in 8-well LabTek chamber slides (Thermo Fisher Scientific, Waltham, MA) in DMEM containing 10% heat inactivated FBS and incubated at 37°C for 6 h. Subsequently, floating lymphocytes were removed and adherent monocytes were washed followed by cultivation in DMEM containing 10% heat inactivated FBS and 10% human AB serum (Sigma-Aldrich, St. Louis, MO) for 7 days to allow differentiation into macrophages.

IV.III. Drug treatments

Primary $CD4^+$ T cells were treated for 6-7 h with 7 µM BIX10294 (Sigma-Aldrich), 1 µM JQ1 (Sigma-Aldrich), 7 µM SAHA (Sigma-Aldrich) and 7 µM UNC0642 (Sigma-Aldrich).

IV. IV. Cell viability test

Cell viability was determined by performing live cell staining and flow cytometry. For the purpose of this assay $2*10^5$ cells were collected for each analyzed time point. Living cells were washed in PBS and stained for 30 minutes on ice with Fixable Viability Dye eFluor 450 (eBioscience). After the staining cells were washed in PBS and fixed for 10 minutes in 4% PFA/PBS. Subsequently the cells were washed and resuspended in PBS. Flow cytometry was performed on BD FACSVerse flow cytometer and flow cytometry data was analyzed in the BD FACSuite software (BD Biosciences). Live cells were gated according to a negative control sample in which all cells were dead. The negative control sample was generated by keeping cells for 10 minutes in 100% ethanol, which would kill most cells in the sample. The cells were subsequently fixed and stained as previously described.

IV. V. Cell cycle analysis

Cell cycle analysis was performed on primary $CD4^+$ T cells after 7 µM BIX01294 treatment in order to determine potential changes in cellular cycles upon drug treatment. For the purpose of this analysis $1*10^6$ cells were collected after 7 h of drug treatment (as well as untreated control cells), washed in PBS and fixed in icecold methanol (methanol was added dropwise to cells while vortexing them). After fixation the cells were washed with PBS and treated with 1 mg/mL RNase A for 30 minutes at 37°C. After RNase A treatment cells were washed with PBS and stained with Propidium Iodide (Roche) in 1:20 ratio. Propidium Iodide is a DNA binding and intercalating dye and according to the amount of bound Propidium Iodide in the cells, cell cycle profiles can be determined, since cells contain certain DNA amounts in specific phases of the cell cycle.

Flow cytometry was performed on BD FACSVerse flow cytometer and flow cytometry data was analyzed in the BD FACSuite software (BD Biosciences).

IV. VI. Cell proliferation assay (MTT test)

The viability and proliferative potential of cells upon BIX01294 treatment was assessed through the CellTiter 96® Non-Radioactive Cell Proliferation Assay (MTT) assay (Promega; Madison, WI, USA), as previously described by Shytaj et al, 2015.

For the MTT assay $300*10^5$ cells of each analyzed condition were resuspended in 100 µL RPMI/10% FBS and transferred into a 96-well plate. To each well the MTT solution was added (15 µL) and, after 2-4 h, the reaction was stopped by the addition of 100 µL of the Solubilization/Stop Solution. Absorbance values at 570nm were acquired with an Infinite 200 PRO (Tecan, Männedorf, Switzerland) multimode plate reader. Reactions were conducted in triplicate and the averages of the triplicates were normalized over the matched untreated controls and expressed as a percentage.

IV. VII. G9a knockdown

Jurkat cells were stably transduced with pGZIP-G9a lentiviral vector containing a puromycin resistance sequence. 48 h after transduction, cells were grown in selection medium with increasing amounts of puromycin reaching a final concentration of 4 µg/mL.

IV. VIII. Total protein amount quantification by western blot

Cell pellets were lysed 10 minutes on ice in RIPA buffer (150mM NaCl, 20mM Tris-HCl pH 7.4, 1mM EDTA, 0.5% NP-40, 0.5% DOC, 0.1%SDS) and sonicated for 5 minutes (30 s ON, 30 s OFF) at 4°C in Bioruptor (diagenode). The total protein concentration of cellular extracts was measured by Bradford assay (BioRad). Western blot gels were loaded with 8 µg of protein lysate. Gels were running on 120 V and 400 mA (western blot system from Biometra) in 1X SDS running buffer (25 mM Tris, 192 mM Glycine, 0,1% SDS, pH 8.3). Proteins were transferred to a

nitrocellulose membrane with a Trans-Blot device for semi-dry transfer (BioRad). The nitrocellulose membrane was blocked for 1 h at room temperature in 1% skim milk in 0,1% PBS/Tween. Upon blocking the membrane was incubated with primary antibodies, which were also diluted in 1% skim milk in 0,1% PBS/Tween, rotating overnight at 4°C. Western blots were developed after washings in 0,1% PBS-Tween and incubation with secondary antibodies for 1 h at room temperature with Amersham ECL Prime Western Blotting Detection Reagent (GE Healthcare Life Sciences).

Used primary antibodies: mouse monoclonal anti-β-actin (Sigma-Aldrich), mouse monoclonal Anti-dimethyl-Histone H3 (Lys 9) 05-1249 (Millipore), rabbit-polyclonal Anti-Histone H3 06-755 (Millipore), rabbit polyclonal LEDGF/p75 A300-847A (Bethyl Laboratories), rabbit monoclonal G9a/EMT2 (C6H3) 3306 (Cell Signaling Technology).

Used secondary antibodies: anti- mouse and anti-rabbit (Jackson immunoresearch).

IV. IX. Virus production

Viral particle production procedure

Production of viral particles was performed as previously described in Marini et al, 2015.

Producer HEK293T cells were seeded to reach 70% confluency in plastic round 15 cm dishes. The Ca-Phosphate approach for cell transfection was used (list of reagents is listed in Table 4.1.). 6-8 h after plasmid DNA transfection into HEK293T cells, the medium was changed and cells were kept for another 48 h post transfection at 37°C to produce viral particles which were released into the medium (supernatant). After 48 h the cell medium was collected, filtered through 0,45 μm (Roth GmBH) filter and prepared for ultracentrifugation on a SW40 rotor (Beckman Coulter ultracentrifuge) through a 20% sucrose gradient at 28 000 rpm for 1,5 h. After ultracentrifugation the viral pellet was resuspended in sterile 1X PBS and aliquoted and stored at -80°C.

Single round infection HIV-1 viral stocks

For the production of single round infection HIV-1 viral stocks a plasmid obtained from the Env^- molecular clone pNL4-3/E^-R^- was used. This molecular clone contains a frameshift mutation introduces close to the 5' end of the *env* gene (Connor et al, 1995) and performs a single-round infection once pseudotyped with the glycoprotein envelope from vesicular stomatitis virus (VSV-G). This disables the virus from performing multiple rounds of infections and causing massive cell death.

This virus was used in the context of monitoring levels of viral integration by Alu PCR in either primary cells or cell lines, as well for FISH analysis in order to monitor the exact positioning of single proviruses and avoid superinfection.

C-terminal IN mutant HIV-1 viral stocks

For the production of viral particles containing K264,266,273R mutations in the viral IN the construct $HIV\text{-}1_{BRU(K264,266,273R)}$ was used (Cereseto et al, 2005), which contains the full viral genome.

This virus was used to test for its integration capacity, transcription capacity and the capacity to form retro-transcribed viral DNA products.

IN.eGFP HIV-1 viral stocks

Production of the engineered HIV-1 wt virus, HIV-1 virus with the K264,266,273R mutations in the C-terminus of the viral IN, the D116N CCD HIV-1 mutant, as well as the *tat* deficient HIV-1 mutant having an eGFP tagged integrase was performed as previously described in Albanese et al, 2008, through the "trans-incorporation" technique.

The trans-incorporation technique takes advantage of the viral protein Vpr being able to shuttle fused exogenous proteins inside the viral particle. In order to produce HIV-1 viruses containing a tagged IN protein pVpr-IN-eCFP was constructed by cloning Vpr (PCR amplified from pNL4.3) in frame with the codon optimized IN (Limon et al, 2002) into the peCFP-N1 vector (Clontech Laboratories, Inc., Saint-Germainen-Laye, France). Also, an HIV-1 protease cleavage site

(IRKVL), flanked at both C- and N- terminus by a flexible linker (KRIQST) which was introduced between Vpr and IN. pVpr-IN-eGFP was constructed by substituting the eCFP cDNA with eGFP.

For viral particle production the constructs containing either the full viral genome, a mutation in the CCD of IN, mutations in the C-terminus of IN, or a *tat* deletion, as well as the belonging pVpr-IN-eGFP construct were transfected via the Ca-Phosphate approach into HEK293T producer cells as previously described.

Creation of adequate Vpr constructs

The K264,266,273R mutations in the C-terminus of the viral IN, as well as the CCD mutant D116N were constructed by a quick change PCR and introduction point mutations in the IN domain of the created pVpr-IN-eGFP construct.

The *tat* deficient HIV-1 viral construct was made by deleting a 34 bp segment in the big *tat* exon of the pNL4-3 HIV-1 original molecular clone. This virus has also been tagged with the IN.eGFP fusion construct.

The mentioned eGFP tagged viral constructs were used mainly for microscopy approaches, but were also successfully used to verify their integration capacity Alu PCR.

Table 4.1. List of reagents for calcium phosphate transfection of HEK 293T cells.	
Reagent	**Amount per 15 cm dish of cells**
2.5 M CaCl2	300 µL
2X HBS (50 mM HEPES, 280 mM NaCl, 1.5 mM Na2HPO4, pH 7.5)	2.5 mL
pNL4-3/E^-R^- plasmid	25 µg
VSV-G plasmid	5 µg
H2O	2.5 mL
***Procedure:** Mix DNA and H2O. Add CaCl2 dropwise (incubate 10 min) Add mixture to 2X HBS by vortexing (incubate 20 min). Add to cells dropwise.	

IV. X. SG-PERT assay

The SG-PERT assay was performed for virion number quantification in viral stocks. SG-PERT assay was performed as previously described in Pizzato et al, 2009.
A 5 µL aliquot of the viral stock was lysed in 5 µL 2X lysis buffer (0.25% Triton X-100, 50 mM KCl, 100 mM TrisHCl pH7.4, 40% glycerol). RNase inhibitor (0.4 U/µL) was added to the volume of buffer needed for virus disruption immediately before use. 90 µL 1X dilution (diluted to 1X from 10 X stock (50 mM $(NH_4)_2SO_4$, 200 mM KCl and 200 mM Tris–HCl, pH 8.3) buffer was added to the lysed sample after 10 minutes of lysis. 10 µL of the diluted virus lysate samples were mixed immediately with 10 µL of 2X 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,1 µM reverse primer, 1.2 µg/ml BMV RNA). Hotstart Taq was added (0.2 U/reaction) immediately before use.

Used thermal cycler conditions used were: 30 min RT reaction at 37 °C, 5 min hot-start Taq activation at 95 °C and 45 cycles of amplification. Each amplification cycle was composed of 5 s denaturation at 95 °C, 5 s annealing at 55 °C, 15 s extension at 72 °C, 7 s acquisition at 83 °C, using the lightcycler; 5 s denaturation at 95 °C, 5 s annealing at 55 °C, 20 s extension at 72 °C, 11 s acquisition at 83 °C, using the CFX96 Touch platform (Bio Rad).
Data analysis was performed by using the CFX Maestro Software (Bio Rad) by comparison of real time Ct values of 10 fold serial dilutions of a recombinant HIV-1 RT standard which was prepared in parallel to the analyzed samples (activity of undiluted standard is 5,088 * 10^9 pUnits RT/µL) and the analyzed sample.

*Used primers:
forward: 5′-TAGTTGTTGGGCTTCGCTTT-3′;
reverse:5′-TTGTCGGCTTTACCTGCTTT3′

IV. XI. Infections

Cell infections for FISH and integration analysis

Cell lines and primary cells were infected with 100 ng p24 per 10^5 cells pNL4-3/E^-R^- HIV-1, and subsequently kept in culture and collected 72 h post infection. For the purpose of FISH analysis, only MDMs were fixed and collected at different time points post infection (36, 48 and 60 h post infection).

Cell infections for colocalization analysis

Primary cells were infected with pNL4-3/E^-R^- HIV-1, IN.eGFP HIV-1 wt, IN.eGFP D116N, IN.eGFP K264, 266, 273R and IN.eGFP **Δ**tat in a concentration of 100 ng p24 per 10^5 cells.
For microscopy and colocalization analysis primary $CD4^+$ T and macrophages were collected at different time points post infection (18, 36, 48, 60 or 96 h).

5 uM Raltegravir (HIV-1 strand transfer inhibitor) was added to macrophages at the moment of infection wit IN.eGFP HIV-1 wt and sequentially added every 24 h as long as the cells were kept in culture.

IV. XII. Genomic DNA extraction

DNA was extracted with DNeasy Blood & Tissue kit (Qiagen) by following manufacturer instructions. DNA concentration was measured via spectrophotometry on P-class P 300 NanoPhotometer (IMPLEN).

IV. XIII. Measurement of integrated vDNA (Alu PCR)

The presence of integrated viral DNA was evaluated by Alu PCR as previously described (Tan et al, 2006). Alu LTR sequences were amplified from 250 ng of genomic DNA in the first round PCR. In the second round Real Time PCR, a 1:50 dilution of the first round PCR product was

used as a template together with the I-specific primer λT, the internal LTR primer LR and the probe ZXF-P.

As a normalizer gene for the Real Time PCR the amounts of housekeeping gene lamin B2 (B13 region) were quantified from 10 ng genomic DNA.

Alu PCR was normalized using the 2(-Delta Delta C(T)) method for analysis of real time quantitative PCR (Livak & Schmittgen, 2001).

Table 4.2. List of used primers for Alu PCR.

Primer	**Sequence**
LM667	ATGCCACGTAAGCGAAACTCTGGCTAACTAGGGAACCCACTG
LR	TCCACACTGACTAAAAGGGCTTGA
ZXF-P probe	TGTGACTCTGGTAACTAGAGATCCCTCAGACCC
Alu1	TCCCAGCTACTGGGGAGGCTGAGG
λ-primer	ATGCCACGTAAGCGAAACT
B13 Fwd primer	CCCCAGGGAGTAGGTTGTGA
B13 Rev primer	TGTTATTTGAGAAAAAGCCCAA
B13 probe	CAGCAGGAAAGGAC

IV. XIV. Quantification of 2LTR circles and total viral DNA

Quantification of 2 LTR circles and total viral DNA was performed as previously described by Apolonia et al, 2007. Total DNA from infected cells was extracted with DNeasy Blood & Tissue kit (Qiagen) as previously described. qPCR was performed by using an input of 250 ng genomic DNA, primers that were designed for the 5' and 3' LTR junction (for 2 LTR quantification) and primers annealing to the 5' LTR and Gag gene (for total viral DNA quantification), as well as fluorescently labeled probes and TaqMan chemistry (BioRad) (Table 4.3.). 2LTR circles were quantified by normalization over the total viral DNA values by using the 2(-Delta Delta C(T)) method for analysis of real time quantitative PCR (Livak & Schmittgen, 2001).

Table 4.3. List of used primers and probes for 2 LTR circle and total viral DNA quantification.	
Primer	**Sequence**
2 LTR circles Fwd primer	AACTAGAGATCCCTCAGACCCTTTT
2 LTR circles Rev primer	CTTGTCTTCGTTGGGAGTGAATT
2 LTR circles probe	CTAGAGATTTTCCACACTGAC
Total viral DNA Fwd primer	TGTGTGCCCGTCTGTTGTGT
Total viral DNA Rev primer	GAGTCCTGCGTCGAGAGAGC
Total viral DNA probe	CGCCCGAACAGGGACTTGAA

IV. XV. Infectivity test

The amount of infected primary $CD4^+$ T cells was determined via staining of p24 marker on cellular surfaces cells and subsequent flow cytometry analysis. For the purpose of this assay $2*10^5$ cells were collected 72 h post infection, washed in PBS and fixed for 1,5 h in 4% PFA/PBS. Cells were stained for 30 minutes with anti-p24 antibody Coulter Clone KC57-RD1 (Beckman Coulter) and subsequently washed and resuspended in PBS. Flow cytometry was performed on BD FACSVerse flow cytometer and flow cytometry data was analyzed in the BD FACSuite software (BD Biosciences).

IV. XVI. HIV-1 integration site sequencing

HIV-1 integration site analysis was performed as previously described in Schmidt et al, 2007, by performing linear amplification-mediated PCR (LAM-PCR).

Upon DNA extraction of previously infected primary $CD4^+$ T cells the material was prepared for the linear PCR initiated from 5′-biotinylated vector-specific primer(s) as described in the table below (Table 4.4. and Table 4.10.). A starting material of 500 ng was used for the initial linear PCR reaction.

Table 4.4. Linear PCR setup.	
Reagent	**Final amount (1X reaction)**
10X Taq polymerase reaction buffer	5 µL
dNTPs	1 µL
Primer LTR I (0.5 pmol/µL)	0.5 µL
Template DNA (1 ug)	X µL
Taq polymerase (2.5 U/µL)	0.5 µL
ddH2O	Up to 50 µL

Linear PCR parameters are described in the table below (Table 4.5.).

Table 4.5. Linear PCR.		
Temperature	**Time**	**Cycle number**
95C	5 min	1
95°C 60°C 72°C	1 min 45 s 90 s	52
72°C	10 min	
4°C	∞	

After performing linear PCR, magnetic Dynabeads (Invitrogen) were washed twice in 0.1% BSA/PBS and 3 M LiCl buffer (10 mM Tris-HCl, pH 7.5, 1 mM EDTA, 3M LiCl) and resuspended in 6 M LiCl buffer (10 mM Tris-HCl (pH 7.5), 1 mM EDTA, 6 M LiCl). Beads were added in a 1:1 ratio to the PCR product. To covalently link the biotinylated PCR product to the streptavidin-coupled magnetic beads the samples were overnight incubated at room temperature on a horizontal shaker at 300 rpm.

To generate double-stranded DNA fragments a hexanucleotide priming reaction was carried out on the bead-DNA complexes for 1 h at 37°C (reaction setup Table 4.6.).

Table 4.6. Hexanucleotide priming reaction.	
Reagent	**Amount (1X reaction)**
10X hexanucleotide mixture (Roche)	2 μL
dNTPs (200 uM)	0.5 μL
Klenow polymerase (2 U/μL)	1 μL
ddH2O	Up to 20 μL

After the hexanucleotide priming reaction, the bead-DNA complexes were washed twice with ddH2O on a magnetic stand and divided in 2 separate tubes and prepared for enzymatic digestion with MseI and MluCI restriction enzymes for 1-2 hr at 37°C to generate restriction sites suitable for the annealing of DNA linker cassettes. The linker cassette fulfills the purpose of a 12 bp unique barcode for each sample which can be later on tracked after the sequencing is performed. The linker cassette ligation reaction was performed after the enzymatic digestion (Table 4.7.).

Table 4.7. DNA linker cassette ligation reaction.	
Reagent	**Amount (1X reaction)**
*10X incubation buffer	1 μL
*ATP (10 mM)	1 μL
Linker cassette	2 μL
*Fast Link DNA ligase (2 U/μL)	1 μL
ddH2O	Up to 10 μL
*Fast Link DNA ligation kit (Biozym)	

After the linker cassette ligation reaction, the samples were washed twice with ddH2O on a magnetic stand. The bead-DNA complexes were denatured by a 10-60 min 0.1 M NaOH wash on a horizontal shaker at 300 rpm at room temperature (increasing denaturation times increases the purity of DNA). Beads were removed with a magnetic stand and the released DNA was prepared for the First Exponential PCR in order to amplify the amount of successfully linker-ligated DNA (Table 4.8. and Table 4.9.).

Table 4.8. First Exponential PCR setup.	
Reagent	**Final amount (1X reaction)**
10X Taq polymerase reaction buffer	5 µL
dNTPs	1 µL
Primer LTR II (50 pmol/µL)	0.5 µL
Primer LC I (50 pmol/ µL)	0.5 µL
Template DNA (1 ug)	2 µL
Taq polymerase (2.5 U/µL)	1 µL
ddH2O	Up to 50 µL

Table 4.9. Linear PCR I.		
Temperature	**Time**	**Cycle number**
95C	2 min	1
95°C 60°C 72°C	45 s 45 s 1 min	35
72°C	5 min	
4°C	∞	

PCR products of the first exponential PCR were captured with streptavidin-coupled magnetic beads as previously described (same procedure as after the Linear PCR). Captured DNA was prepared for the Second Exponential PCR. The Second Exponential PCR has the same temperature and cycle regime as the First Exponential PCR, as well as the same chemical regime. The only difference is that in this PCR setup another primer pair was used. Primer LTR III and LC II, in order to ensure the amplification of only successfully linker-ligated DNA on the 3' end, and DNA fragments containing the viral LTR on the 5' DNA end.

After performing the Second Exponential PCR a control 2% agarose gel was run with the obtained PCR products in order to verify that the LAM-PCR procedure was successfully performed and to control for eventual contaminations in the H2O negative control sample. Upon the Second Exponential PCR, DNA purification of the PCR products was performed.

DNA products were purified from the PCR with AMPure beads (Beckman Coulter). The beads were added to the PCR product in a 1:1,1 ratio and incubated 5 min at room temperature. The supernatant was removed by using a magnetic stand and the beads were washed twice with 80% EtOH. After the washing steps, the beads were resuspended in 20 µL water and the purified DNA

was transferred into fresh tubes by using a magnetic stand. This purified DNA was used as an input for the sequencing library preparation.

Table 4.10. List of used primer pairs for LAM-PCR procedure.	
Primer	**Sequence**
LTR I	GAGCTCTCTGGCTAACTAGG
LTR II	AGCTTGCCTTGAGTGCTTCA
LTR III	AGTAGTGTGTGCCCGTCTGT
LC I	GACCCGGGAGATCTGAATTC
LC II	AGTGGCACAGCAGTTAGG

Sequencing library preparation

For the sequencing library preparation, a 40 ng DNA input of the purified DNA from the LAM-PCR procedure was used and a PCR was performed to ligate the sequencing linker to the purified DNA (Table 4.11. and Table 4.12.).

Table 4.11. Mega PCR – Library preparation.	
Reagent	**Final amount (1X reaction)**
10X Taq polymerase reaction buffer	5 µL
dNTPs	1 µL
Mega primer (5 pmol/µL)	1 µL
*Mega Linker (CCTAACTGCTGTGCCACT)	1 µL
Template DNA (40 ng)	2 µL
Taq polymerase (2.5 U/µL)	1 µL
ddH2O	Up to 50 µL
*Always the same for all samples	

Table 4.12. Linear PCR II.		
Temperature	**Time**	**Cycle number**
95C	2 min	1
95°C 60°C 72°C	45 s 45 s 1 min	15
72°C	5 min	
4°C	∞	

The Mega primer sequence was individually chosen for each sample and was dependent on the ligated linker cassette. It is important that the Mega primer anneals to the unique linker cassette. After performing the PCR another control 2% agarose gel was run to check for successful DNA amplification and the amplified DNA was further purified with AMPure beads as previously described. The only difference in this DNA purification step is that the beads are added in a lower ratio to the PCR product (1:0.7) in order to amplify only PCR products above the size of 100 bp, since PCR products of this size most probably contain the target sequence and successfully bound linker cassettes and sequencing linker.

Next generation sequencing

The prepared sequencing libraries were sequenced on an Illumina MiSeq platform by 250 bp Paired End sequencing method.

Integration site data analysis and comparison

The pairwise comparison of the common integration site (CIS) profile was performed to evaluate the similarity of the samples. The CIS definition is based on an approach based on graph and described in Fronza et al, 2015. Each combination of the insertion site set of samples were analyzed in order to obtain a mixed CIS profile. Normalized entropy N between 0 and 1 was then computed for each CIS >2th order in the mixed profile. The normalized entropy value for the CIS in the mixed profile of the samples is defined as Nijl= (Sk pijlk)/log 2, where pijlk is the IS fraction belonging to the sample k, which represents the dataset of IS from other studies (Ikeda et

al, 2007; Brady et al, 2009; Wagner et al, 2014, Maldarelli et al, 2014; Kok et al, 2016). Then, all those CIS with Nijl > 0 were defined as shared. The fraction of shared CIS with respect to the total number of CIS was used as a measure of association between the two samples. The values together form a 9 x 9 association matrix, with values ranging from 0 (no shared CIS) to 1 (all CIS are shared). The matrix was then loaded into R suite and converted to a dissimilarity table. Briefly, for each element eij in the matrix, with i = 1, 2, … 9, a dissimilarity table was constructed where the dissimilarity values di,j = 1 – eij are used as a distance in order to perform a hierarchical clustering. The tree structure is then obtained by plotting the results returned from the R function hclust().

IV. XVII. RNA extraction

RNA was extracted with Invitrap spin universal RNA mini kit (Stratec biomedical) by following manufacturer instructions. RNA concentration was measured via spectrophotometry on P-class P 300 NanoPhotometer (IMPLEN).

IV. XVIII. Quantitative reverse transcription PCR

The quantitative reverse transcription PCR was performed as previously described in Marini et al, 2015.

For the quantification of HIV transcript levels, RNA was purified from the cells with the Invitrap spin universal RNA mini kit (Stratec biomedical), as previously described. The messenger RNA (mRNA) levels were quantified by TaqMan quantitative reverse transcription PCR (qRT-PCR) using HIV-1 primers and probe, and housekeeping gene 18S and GAPDH (both containing VIC™/TAMRA™ fluorescent probe) (Applied Biosystems) as controls.

The RT-PCR reaction was performed using M-MLV reverse transcriptase and reagents from Invitrogen and the qPCR reaction was performed using TaqMan technology from Biorad (Table 4.13.).

Table 4.13. RT-PCR setup.		
Reagent	**1X reaction reagent final amount**	**RT-PCR program**
Mix 1 RNA input Random primers dNTPs	 - 500 ng - 3 ⌈g - 10 mM	Incubate reaction 65 °C – 5 min 4 °C - ∞
Mix 2 5X First strand buffer buffer DTT RNA inhibitor	 - 1X - 3 ⌈M - 40 U	Add mix 2 to reaction and incubate 37 °C – 2 min 4 °C - ∞
Addition of M-MLV reverse transcriptase	- 200 U	Add enzyme to reaction and incubate 25 °C – 10 min 37 °C – 50 min 70 °C – 15 min 4 °C - ∞ } 1X

Table 4.14. List of used primers and probes for HIV mRNA expression levels quantification.		
Primer	**Sequence**	**Position**
U1a Fwd primer	ACATCAAGCAGCCATGCAAAA	543
U1a Rev primer	CAGAATGGGATAGATTGCATCCA	629
U1a probe	AAGAGACCATCAATGAGGAA	605
Nuc1b Fwd primer	CGTCTGTTGTGTGACTCTGGTAACT	111
Nuc1b DNA Rev primer	CACTGCTAGACATTTTCCACACTGA	158
Nuc1b probe	ATCCCTCAGACCCTTT	140

For the quantification of cell gene transcription levels, the same procedure as described for HIV transcript levels was performed.

Table 4.15. List of used gene expression assays (Applied Biosystems).	
Gene	**Expression assay code**
BACH2	Hs00222364_m1
STAT5B	Hs0273500_m1
MKL2	Hs00539161_s1
Tubulin 1a	Hs03045184_g1

IV. XIX. RNA expression analysis by microarray

RNA expression levels were determined via BeadArray microarray Technology (Illumina) by using the HumanHT-12 Expression BeadChip kit.
BeadArray data analysis was performed in GenomeStudio software (Illumina) and R statistical software.

IV. XX. Bacterial Artificial Chromosome (BAC) DNA purification

BAC DNA was purified with NucleoBond Xtra Maxi kit (Macherey-Nagel) according to manufacturer instructions. BAC DNA constructs containing sequences of selected RIGs and non-RIGs were used for the generation of fluorescent DNA FISH probes.

Table 4.16. List of used BAC constructs.

Gene name	BAC reference code
BACH2	RP11-59717
NPLOC4	RP11-765O14
MKL2	RP11-1072B15
STAT5B	CTD-3124P7
DNMT1	CTD-2240E14
RPTOR	RP11-28G8
SMG1	RP11-1035H13
GRB2	RP11-16C1
GNB1	RP11-798H13
NFATC3	RP11-67A1
KDM2B	RP11-44F24
ACTN1	RP11-226F19
PACS2	RP11-521B24
CNTN4	RP11-63O1
RNF157	RP11-449J21

IV. XXI. Immunofluorescence (IF)

Cell preparation and permeabilization

Suspension cells were seeded and fixed with 4% PFA/PBS on glass cover slips which were previously coated with 0,5 mg/mL PEI (Polyethyleneimine), whereas adherent cells were directly grown on glass coverslips.
Only for R-loop staining cells were fixed 10 minutes in methanol at -20°C and subsequently 1 minute in Acetone at -20°C. Cells were washed with PBS and permeabilized with 0,5% Triton/0,1% PBS/Tween for 10 minutes at room temperature, and subsequently washed with 0,1% PBS/Tween.

Staining

After washing and 30 minutes blocking in 4% BSA/PBS cells were incubated with primary antibodies diluted in 1% BSA/PBS overnight at 4°C. After incubation with primary antibodies cells were washed with 0,1% PBS/Tween and incubated with secondary Alexa Fluor antibodies (ThermoFisher) for 1 h at room temperature. After 0,1% PBS/Tween washing cells were counterstained with Hoechst and mounted on glass slides with mowiol solution.
For STED nanoscopy Anti-Rabbit IgG ATTO 647N and Anti-Mouse IgG ATTO 594 (Sigma-Aldrich) secondary antibodies were used.
Used primary antibodies: rabbit polyclonal lamin B1 16048 (abcam), mouse monoclonal mAb 414 24609 (abcam) mouse monoclonal Anti-Histone H3 (di methyl K9) 1220 (abcam), rabbit polyclonal Anti-Histone H3 (tri methyl K9) 8898 (abcam), rabbit polyclonal LEDGF/p75 A300-847A (Bethyl Laboratories), mouse monoclonal (9F3) to gamma H2A.X (phospho S139) (abcam).

R-loop staining

For the R-loop staining in primary $CD4^+$T cells and macrophages different staining approaches had to be applied. The R-loop antibody epitopes had to be exposed via different fixation and

permeabilization strategies in the two cell types.

Primary $CD4^+$T cells were fixed with 4% PFA/1X PBS/0,1%Tween for 10 minutes at room temperature. During the last two minutes of fixation a few drops of PBS/0,5% TritonX-100 were added. Cells were subsequently washed 3 times with 0,05%TritonX-100/PBS. Cells were permeabilized with 0,5% TritonX-100/PBS for 10 minutes at room temperature. Permeabilization was followed by 3 washes of 0,05%TritonX-100/PBS. Cells were blocked for 30 minutes at room temperature with 4% BSA/PBS/0,1% Tween. Cells were incubated overnight at 4°C with the primary antibody - S9.6 (Kerafast) and other tested antibodies (lamin B1 16048 - abcam) in a 1:500 dilution in 1% BSA/PBS/0,1% Tween. The cells were washed 3 times with PBS/0,1% Tween and incubated for 1 hour at room temperature with the secondary antibody Alexa Fluor antibodies (ThermoFisher) in a 1:1000 dilution in 1% BSA/PBS/0,1% Tween. After 3 times washing in PBS/0,1% Tween, cells were stained with Hoechst for 5 minutes at room temperature and once rinsed in PBS. Coverslips were mounted with mowiol on glass microscopy slides.

For R-loop staining in primary macrophages cells were fixed 10 minutes in methanol at -20ºC and 1 minute Acetone at -20ºC. Cells were subsequently washed 3 times with PBS and permeabilized with 0,5% Triton in 0,1% PBS/Tween for 10 minutes at room temperature. After 3 times washing with PBS/0.1% Tween and 30 minutes blocking in 4% BSA/PBS cells were incubated with primary antibodies (S9.6 and lain B1) diluted 1:500 in 1 % BSA/PBS overnight at 4C. Cells were afterwards washed 3 times with PBS/0.1% Tween and incubated for 1 hour at room temperature with the secondary antibody Alexa Fluor antibodies (ThermoFisher) in a 1:1000 dilution in 1% BSA/PBS/0,1% Tween. After 3 times washing in PBS/0,1% Tween washing cells were stained with Hoechst for 5 minutes, rinsed in PBS and coverslips were mounted on glass microscopy slides with mowiol solution.

IV. XXII. DNA fluorescent *in situ* hybridization (FISH)

Cell preparation

Cells were first prepared and permeabilized as already described in the IF section. Upon cell preparation, cells were first immunostained (as previously described) either with lamin B1 or

nuclear pore complex antibodies, which served as a landmark for the nuclear periphery. Upon incubation of the secondary antibodies, secondary antibodies were 10 minutes fixed with EGS/PBS in order to preserve their presence until completion of the entire FISH procedure. Cells were further permeabilized with 0.5% triton/0.5% saponin/PBS and 0.1 M HCl for 10 minutes, which was followed by 0,1% PBS/Tween washings between the incubation steps. Further on cells were equilibrated in 2X SSC buffer (Table 4.17.) and left in hybridization buffer overnight (Table 4.18.). Only for HIV FISH before cell equilibration in 2X SSC buffer and incubation in hybridization buffer, cells were treated with with 100 µg/ml RNase A (PureLink RNase A (Invitrogen)) in 2x SSC for 30 min at 37°C, in order to avoid possible HIV-1 FISH probe hybridization to viral RNA transcripts.

Table 4.17. Preparation of saline-sodium citrate buffer (SSC).	
20X SSC buffer	NaCl 175,3 g (In stock solution 3M)
	Na_3Citrate x 2 H_2O 88,2 g (in stock solution 300 mM)
	Dissolve chemicals in 800 ml ddH_20
	Adjust pH to 7.0 with 14 N (or similar) HCl

Table 4.18. Preparation of FISH hybridization buffer.	
FISH hybridization buffer	50 ml deionized Formamide
	10 ml 20x SSC
	20 ml sterile water; adjust pH to 7.0,
	20 ml of 50% dextran sulfate
	Final volume = 100 ml
	Store at 4°C

FISH probe preparation

Biotinylated HIV-1 FISH probes were prepared with the Nick Translation Kit (Roche) and Dioxigenin labeled BAC FISH probes were prepared with the DIG-Nick Translation Mix (Roche) according to manufacturer instructions. Probe size was controlled by loading 1 µL of probe labeling reaction on a 1% agarose gel. If the prepared DNA FISH probes had the desired size between 200-500 bp, they were cleaned up with Illustra Microspin G-25 columns (GE-Healthcare) according to manufacturer instructions, and precipitated in the presence of human

Cot-1 DNA (Roche) and DNA from herring sperm (Sigma). The presence of highly repetitive human sequences and fish sperm DNA in the target DNA FISH probe should facilitate specific binding of the FISH probe to the genomic target. Finally, after ethanol precipitation, the probes were resuspended in 10 µL formamide, incubated at 37°C for 15-20 min and 10 µL of 20% dextran in 4X SSC was added to a final volume of 20 µL.

FISH probe hybridization and detection

1-10 µL of the FISH probes were loaded onto glass slides which were covered with coverslips containing the cells, followed by sealing in a metal chamber, and heat-denatured in a water bath at 80°C for 7 min. Hybridization was performed for 48 h at 37°C in a water bath. Upon hybridization, coverslips containing the cells were washed in 2X SSC buffer (10 min each wash), which was followed with three washings in 0.5X SSC buffer at 56°C in order to wash away potential unspecific bound DNA FISH probe fragments. FISH detection, for Dig-labelled BAC probes, was performed by using FITC-labelled anti-digoxigenin antibody (Roche), whereas biotin-labelled HIV-1 probes were detected by a TSA Plus system from Perkin Elmer, allowing signal amplification, by using an anti-biotin antibody (SA-HRP) and a secondary antibody with a fluorescent dye (usually FITC for HIV).
Upon FISH probe detection, the cells were counterstained with Hoechst and coverslips were mounted with moviol solution on glass slides

IV. XXIII. Microscopy

Confocal microscopy

Confocal microscopy was performed on the confocal laser-scanning microscope Leica TCS SP8, by taking Z-stacks from each analyzed sample.
Images were analyzed in ImageJ Fiji image processing package and Volocity imaging software (Perkin Elmer).

Measurements and FISH data analysis

Distance measurements between the FISH signals and the nuclear envelope were measured using ImageJ Fiji image processing package and Volocity imaging software (Perkin Elmer). Distance measurements were normalized over nuclear radius (defined as half of the middle of the mAb414-TRITC or lamin B1 ring), and then binned into three classes of equal surface area (FISH signals positioned in the area of 1-19% of nuclear radius length occupied nuclear zone 1, signals in the area of 20-43% of nuclear radius length occupied nuclear zone 2 and signals in the area of 44-100% of nuclear radius length occupied nuclear zone 3).
In addition, movement and redistribution of alleles upon distance measurements was assessed by performing the Kolmogorov-Smirnov test and p value definition in Prism Graphpad software.
Images of representative cells containing FISH signals were adjusted in ImageJ Fiji image processing package by adjusting brightness and contrast ratio in order to increase signal to noise ratio and highlight the visibility of the FISH signal. Gaussian Blur filter was applied to the images by setting the gamma value to 1.

Colocalization analysis

The colocalization analysis between R-loop signals and HIV-1 IN.eGFP signals was conducted via visual appreciation of the analyzed signals. A colocalization between R-loop signals and HIV-1 IN.eGFP signals was counted as positive only in case of a total match/overlap of the analyzed signals (appreciated by a color-change in the colocalizing signals upon channel merge in Fiji image analysis software - R-loop channel set to magenta and IN.eGFP set to green; a merge of the channels and colocalization between analyzed signals appears as a white/yellow signal). Signals positioned next to each other were not counted as colocalizing signals.

The position assessment of HIV-1 IN.eGFP (K264,266,273Q) mutant in primary $CD4^+$ T cells and macrophages was performed via visual appreciation of the IN.eGFP signals in respect to their position to the nuclear lamina - stained with rabbit polyclonal lamin B1 16048 (abcam) antibody.

STED (Stimulated Emission Depletion) nanoscopy

STED nanoscopy was performed on a gated λ=775 nm STED system containing an easy 3D optics module (Abberior Instruments) and the λ=640 nm excitation laser line. Additional confocal images of eGFP signals excited at λ=488 nm were acquired.
Deconvolution of STED images was performed in Imspector software (Abberior Instruments) via the linear deconvolution tool.
Images were further processed in ImageJ Fiji image processing package and Gaussian Blur filter was applied to the images by setting the gamma value to 1, in order to increase signal to noise ratio.

IV. XXIV. Plasmid DNA purification

Plasmid DNA was extracted with Plasmid Midi Kit (Qiagen) according to manufacturer instructions.

IV. XXV. Plasmid transfections

pMMM-FLST plasmid, containing RNase H1 coding sequence and NLS, as well as GFP reporter sequence, was transfected into Jurkat cells and primary $CD4^+$T cells via electroporation using Amaxa Cell Line Nucleofector Kit V (Lonza). Again, 0,5 µg of plasmid DNA was used for transfection of 10^6 cells.
The same plasmid was transfected into MDMs with the PromoFectin kit (PromoKine). 0,5 µg of plasmid DNA was used for transfection of 10^6 cells.

pMMM-FLST, RNase H WKKD (binding and catalytic mutant) and D210N (binding mutant) plasmid constructs were delivered into HEK293T cells and HeLa wt cells via lipofection. $2*10^6$ cells were lipofected with 0,5 µg plasmid per $1*10^6$ cells with Lipofectamine LTX with Plus Reagent kit (Thermo Fisher).
The efficiency of plasmid delivery was assessed via fluorescence microscopy and detection of GFP expression on Olympus IX81 microscope (Olympus).

IV. XXVI. Chromatin immunoprecipitation followed by quantitative PCR (ChIP-qPCR)

ChIP was performed as described in Chen et al, 2012 with smaller modifications.

$CD4^+$T cells ($10x10^6$) were washed twice in PBS before crosslinking with 1% final formaldehyde for 10 min at room temperature, followed by termination of the reaction with 125 mM glycine on ice. The cell pellet was washed twice with PBS and was lysed in 0.5% NP-40 buffer (10 mM Tris-Cl pH7.4, 10 mM NaCl, 3 mM MgCl2, 1 mM PMSF and protease inhibitors). The nuclei obtained were washed once in the same buffer without NP-40. Lysis of the nuclei was performed using a nuclear lysis buffer (1% SDS, 10 mM EDTA, 50 mM Tris-HCl pH 8) and keeping the cells for 30 minutes on ice. DNA was sheared by sonication (on Bioruptor sonicator Diagenode) to obtain fragment size of 400-200 bp. Extracts were pre-cleared by incubation with immunoglobulin-γ and agarose beads, followed by centrifugation at 1,200g for 5-10min. The lysate was then diluted to 1X in a 5X ChIP binding buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1 % NP40, 0.25 % Na-DOC, 1 mM EDTA pH 8 and protease inhibitors) and incubated with 2-4 µg of the indicated antibody overnight at 4°C, followed by incubation for 2 h with MagnaChIP Protein A/GMagnetic Beads (Millipore). Beads were then washed with ChIP binding buffer (1X PBS, 0,1% BSA, 2 mM EDTA, pH7.4), LiCl-containing buffer (10 mM Tris-HCl, pH8, 250 mM LiCl, 1% IGEPAL, 1% NaDoc, 1 mM EDTA) and with TE buffer (10 mM Tris-HCl, pH8, 50 mM NaCl, 1mM EDTA). ChIP samples were subsequently treated with RNase A for at least 30min at 37°C, and then treated with proteinase K for at least 2 h at 56°C. De-crosslinking of protein-DNA complexes was performed by an overnight incubation at 65°C. Afterwards the DNA was pulled down by a 2 h incubation with AMPure XP magnetic beads (Beckman Coulter) at room temperature. The beads containing bound DNA fragments were washed twice with 80% ethanol and afterwards air dried for 45 min. Upon drying the beads were resuspended in water, and the magnetic beads were removed with a magnet, which released the immunoprecipitated DNA.

Target enrichment was determined by quantitative PCR. Enrichment was calculated by using the 2(-Delta Delta C(T)) method for analysis of real time quantitative PCR (Livak & Schmittgen, 2001) and the results were normalized over ChIP input values.

Used antibodies: mouse monoclonal Anti-Histone H3 (di methyl K9) 1220 (abcam), rabbit polyclonal Anti-Histone H3 (tri methyl K9) 8898 (abcam), rabbit polyclonal LEDGF/p75 A300-847A Bethyl Laboratories), rabbit polyclonal Anti-Histone H3 (tri methyl 36) 9050 (abcam).

Table 4.19. List of used primers for ChIP-qPCR analysis.

Gene	Region	Sequence
BACH2	First intron	F: GAAGAGTGCCTAGCACAGTAAG R: GAGCACACAGGTTTAGGAGATG
	Gene body	F: CAACTCCAAGGGCCAGAG R: AGAGGCTGAACCAACAAGAG
	Exon/intron junction (gene end)	F: TCTGCCTCTGATCCCTATGT R: GGTAGCAAGTAAGGGTCCAAA
STAT5B	First intron	F: TCAAACGTGGTGGAAGTGTTAG R: CACCCAGCCTACCATCTTATTT
	Gene body	F: GGTGATGGTCATGGTTTCCT R: GAGACACCTGCTTCTGCTG
	Exon/intron junction (gene end)	F: CTTCCTCTGAAGGGATGGTATTG R: TCACGCTGATGACTTGAAACT
MKL2	First intron	F: GACCCTTTGCGTACCTTCTT R: GTGGAGCTCAGAGTCTACATTTC
	Gene body	F: TGAAACACAAGATTCGGATCG R: GGGAATAAACAGAAATGGTGGAAAT
	Exon/intron junction (gene end)	F: AGAGGAAAGACCACTCCCTAT R: GGCTTTCTGTTGTGAGTTGTATG
RPTOR	First intron	F: TGCTGGCAGATTGTGGTAAA R: GGGAAGGAAGGAAGGAAGAAAG
	Gene body	-
	Exon/intron junction (gene end)	-
NPLOC4	First intron	F: CCAAGGCCAACTCCAGTATT R: GGTCTGTCTCTTCCTGTGTATTT
	Gene body	-
	Exon/intron junction (gene end)	-
KDM2B	First intron	F: GTTGCTTGGGTTGCAATGG R: AGAGAGCCCGGGACATTAT
	Gene body	-

	Exon/intron junction (gene end)	-
PACS2	First intron	F: CCCAAATCTCATAGCCAGTAGG R: GTGAGGGAGCACAGGAAATA
	Gene body	-
	Exon/intron junction (gene end)	-
ACTN1	First intron	F: AGGTAGAGAGGAACCAGAAA R: AATGCCAGGGAAAGGTAAGG
	Gene body	-
	Exon/intron junction (gene end)	-
CNTN4	First intron	F: CTTACACACCCTTCACCACTAC R: CGCAAGACTCCCACAATAGAA
	Gene body	F: CAGGAGTGATGGGTGAATATGAG R: TAAAGCAAAGCATTCCAGCTTC
	Exon/intron junction (gene end)	F: TACTTGTCTTTCGGGAGCATTT R: GCTCTGGGTCATCTAGATTTGG

IV. XXVII. DNA-RNA hybrid immunoprecipitation (DRIP)

DRIP was performed as previously described in Ginno et al, 2012.

10^7 primary CD4+ T cells were used as starting material. After collection, cells were washed with PBS and used for DNA extraction with DNeasy Blood & Tissue kit (Qiagen). 4 μg of DNA was sheared by sonication (Bioruptor, Diagenode). After sonication the sample was split, where one part of the sample was treated only with RNase A (to remove possible residual RNA) and the other part of the sample was treated with RNAse A (PureLink RNase A, Invitrogen) and RNAse H1 (NEB M0297S) (to remove R-loops) overnight at 37°C. The RNAse H1 treated sample was used as a negative control for the R-loop pull down procedure. After RNase A and RNase H1 digestion the samples were additionally digested with proteinase K (Proteinase K Solution, ThermoFisher) for 30 minutes at 37°C. After proteinase K digestion, the DNA was extracted from the samples by phenol/chloroform extraction. Samples are resuspended in 450 μL of TE buffer and 50 μL of binding buffer (10X binding buffer: 100 mM $NaPO_4$ pH 7, 1,4 M NaCl, 0,5% Triton X100,) was added. To each sample 10 μL of S9.6 antibody (Kerafast) was added and the samples are incubated at 4ºC overnight. Agarose A/G beads (Pierce) were pre-washed 3 times

with 1X binding buffer (diluted in TE) before addition to the samples. The beads were added to DNA-S9.6 antibody complexes and left incubating on a rotisserie shaker for 2 h at 4ºC. After incubation, bead-antibody complexes were washed 3 times with 1X binding buffer and resuspended in 250 µL elution buffer (50 mM Tris, pH 8; 10 mM EDTA, 0,5% SDS). Afterwards another proteinase K digestion was performed for 1 h at 55ºC, by the addition of 7 µL of the enzyme. After digestion, the DNA was cleaned up with QIAquick PCR purification kit (Qiagen).

The enrichment of R-loops was determined by qPCR using control primers for R-loop rich and R-loop poor genes (Table 4.20.). R-loop enrichment was also assessed for selected RIGs, and the results were analyzed by the 2(-Delta Delta C(T)) method for analysis of real time quantitative PCR (Livak & Schmittgen, 2001), and normalized over the R-loop poor region of the SNRPN gene, which turned out to be the best normalizer for primary $CD4^+$ T cells, and has previously been demonstrated to be an adequate DRIP-qPCR normalizer gene (Garcia-Rubio et al, 2015) .

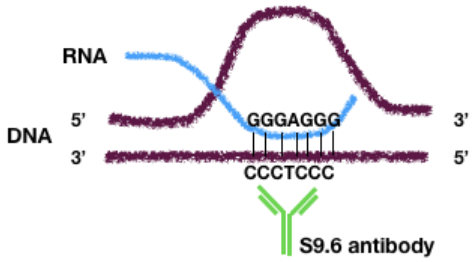


Figure 4.1. Scheme of the binding region of the R-loop specific antibody S9.6.
The epitope of the R-loop specific S9.6 antibody is demonstrated to be 6 bp in size (Phillips et al, 2013).

To control for the efficiency of the DRIP procedure and successful R-loop pull-down, the DRIP was in parallel performed on an R-loop containing plasmid construct (kindly provided by the lab of Frederick Chedin), which was also used as a spike in control in the actual DRIP samples.
The pFC53 plasmid, containing an R-loop forming sequence and a T3 promoter first had to be *in vitro* transcribed, which would initiate R-loop formation. For this reaction 3 µg of plasmid DNA was used, which was mixed with 1 M DTT, 2,5% Tween, 2,5 mM rNTPs, 5X reaction buffer (Promega) and T3 RNA Polymerase (Promega). The reaction was incubated for 30 minutes at 37ºC, and the T3 RNA Polymerase was subsequently heat inactivated for 10 minutes at 65ºC. The reaction was afterwards split for either only RNase A treatment or RNase A and RNase H

treatment, as previously described for the actual DRIP samples. Upon subsequent Proteinase K digestion, the samples were ran on a 2% agarose gel to check for R-loop formation. The formed R-loop fragment can be appreciated on the gel in form of a smear and slight shift of the plasmid on the gel. In order to separate the R-loop fragment from the rest of the plasmid and use both fragments as spike in controls, only the RNase A treated sample was further digested with ApaLI (NEB) restriction enzyme by following manufacturer instructions. Upon digestion, the reaction was cleaned up with the QIAquick PCR purification kit (Qiagen).

Table 4.20. List of used primers for DRIP-qPCR.	
Positive R-loop primer	**Sequence**
APOE	F: CCGGTGAGAAGCGCAGTCGG R: CCCAAGCCCGACCCCGAGTA
MYADM	F: CGTAGGTGCCCTAGTTGGAG R: TCCATTCTCATTCCCAAACC
RPL13A	F: AATGTGGCATTTCCTTCTCG R: CCAATTCGGCCAAGACTCTA
EGR1	F: CCAATTCGGCCAAGACTCTA R: CTTGTGGTGAGGGGTCACTT
BTBD19	F: GGCTGCTCAGGAGAGCTAGA R: ACCAGACTGTGACCCCAAAG
Negative R-loop primer	**Sequence**
SNRPN	F: GCCAAATGAGTGAGGATGGT R: TCCTCTCTGCCTGACTCCAT
EGR1 negative region	F: GAACGTTCAGCCTCGTTCTC R:GGAAGGTGGAAGGAAACACA
MYADM negative region	F: TGCATCTACATCCGCAAAAG R: AGAGTGGACGCTGCAGAAAT
pFC53 primer	**Sequence**
R-loop fragment	F: TTTAGAGCTTGACGGGGAAA R: CAACAGTTGCGTAGCCTGAA
Non-R-loop fragment	F: TTGCCGGGAAGCTAGAGTAA R: GCTGCCATAAGCATGAGTGA
RIGs primer	**Sequence**
BACH2	F: GAAGAGTGCCTAGCACAGTAAG R: GAGCACACAGGTTTAGGAGATG
STAT5B	F: TCAAACGTGGTGGAAGTGTTAG R: CACCCAGCCTACCATCTTATTT

IV. XXVIII. DNA-RNA hybrid immunoprecipitation (DRIP) combined with Alu PCR

In order to verify whether integrated HIV-1 DNA can be situated in R-loop rich regions a combined approach of DRIP and Alu PCR was performed.

DRIP was performed on infected material - pNL4-3/E^-R^- HIV-1 infected CD4+ T cells as previously described. DRIP was performed to the last step of DNA extraction, which was followed by using the extracted genomic DNA as an input for an Alu PCR reaction. Alu PCR was performed in 2 steps as previously described. The results were analyzed by the 2(-Delta Delta C(T)) method for analysis of real time quantitative PCR (Livak & Schmittgen, 2001), and normalized both over both, the B13 region of the lamin B2 gene or R-loop poor region of the SNRPN gene.

IV. XXIX. Absolute quantification of early and late HIV-1 RT-products and 2LTR circles

The absolute quantification of early and late HIV-1 RT-products and 2LTR circles was performed by digital droplet PCR (ddPCR) as previously described by Bejarano et al, 2018. Early and late RT-products were detected with primers and probes annealing to the long terminal repeat (LTR) or to the *gag* gene of the HIV-1 genome. A specifically designed set of primers and probe were used for the detection of 2LTR circles (Table 4.21.). As a housekeeping control gene of the host genome ribonuclease P protein subunit p30 (RPP30) was also quantified and used for the normalization of really and late RT-product values, as well as 2LTR circles. In addition, 20 µL reactions, containing 2-4 µL of the diluted DNA, 900 nM of each primer, 200 nM probe, 1x ddPCR Supermix for probes (no dUTP) (Bio Rad) and water were prepared for each sample and analyzed target. The droplets were generated by using the droplet generation oil for probes (BioRad). The droplets were immediately transferred to a 96-well microplate. The PCR amplification was performed by applying the following temperature regime: initial denaturation and stabilization at 95°C for 10 min, 40 cycles of denaturation at 94°C for 30 s, and annealing/extension at 57°C for 60 s, followed by 10 min at 98°C. Following PCR amplification, the droplets were sorted and analyzed in a QX200 droplet reader (BioRad) by using the QuantaSoft v1.6 software (BioRad) and the settings for absolute quantification. Results were

analyzed using the same software, and copy numbers were normalized to the copy numbers of the housekeeping gene.

Table 4.21. List of used primers/probes for ddPCR amplification of early and late HIV-1 RT-products and 2LTR circles (Bejarano et al, 2018).

Target	Sequence
5' LTR Fwd	TTAAGCCTCAATAAAGCTTGCC
5' LTR Rev	GTTCGGGCGCCACTGCTAG
5' LTR probe	FAM–CCAGAGTCACACAACAGACGGGCA–BHQ1
Gag Fwd	CATGTTTTCAGCATTATCAGAAGGA
Gag Rev	TGCTTGATGTCCCCCCACT
Gag probe	HEX—CCACCCCACAAGATTTAAACACCATGCTAA–BHQ1
2LTR Fwd	CTAACTAGGGAACCCACTGCT
2LTR Rev	GTAGTTCTGCCAATCAGGGAA
2LTR probe	FAM—AGCCTCAATAAAGCTTGCCTTGAGTGC–BHQ1
RPP30 Fwd	GATTTGGACCTGCGAGCG
RPP30 Rev	GCGGCTGTCTCCACAAGT
RPP probe	FAM–CTGACCTGAAGGCTCT–BHQ1

IV. XXX. Total quantification cellular R-loop levels by dot blot assay

The dot blot assay was performed as previously described by Wahba et al, 2013.
500 ng of genomic DNA was directly spotted on a nitrocellulose membrane (GE Healthcare Life Sciences) and left to dry. Afterwards, the nitrocellulose membrane was UV-crosslinked and blocked for 30 minutes in 5% FBS/1% Tween/1xPBS at room temperature. Further on the nitrocellulose membrane was incubated with the primary S9.6 antibody in a 1:500 dilution in 1% FBS/1% Tween/1xPBS overnight at 4°C. Afterwards the membrane was washed in 0,1% Tween/1xPBS and incubated with mouse secondary antibody (Jackson immunoresearch) for 1 h at room temperature. After secondary antibody incubation, the membrane was again washed in 0,1% Tween/1xPBS and R-loop signals were detected with the Amersham ECL Prime Western Blotting Detection Reagent (GE Healthcare Life Sciences) and exposing the membrane on a western blot chemiluminescence imaging device Alliance Q9 Touch (Mandel).

IV. XXXI. *In vitro* integration assays

Integration assays were performed as previously described by Benleulmi et al, 2015. For the purpose of assay performance, recombinant purified IN was used. IN/v DNA complexes were pre-assembled as previously described by Lesbats et al, 2008 and Benleulmi et al, 2015. 10 ng of donor DNA containing the U5 viral ends was used to assemble the IN/vDNA complexes. Pre-assembled complexes were further on incubated with 50 ng of, either transcribed (containing R-loop) or non-transcribed pFC53 plasmid constructs (not containing R-loop) in 20 mM HEPES pH7, 15% DMSO, 8% PEG, 10 mM MgCl2, 20 µM ZnCl2, 100 mM NaCl, 10 mM DTT final concentration. After the integration reaction was performed, the reaction was further treated for 1 h at 37ºC with *HpaI* restriction enzyme in order to cut the pFC53 plasmid construct into its R-loop and non-R-loop fragment, as well as to release the formed integration products. Afterwards the resultant integration products were deproteinized by Proteinase K treatment and phenol/chloroform/isoamyl alcohol (25/24/1 v/v/v) treatment before loading onto a 1% agarose gel. The obtained gel was further dried and submitted to autoradiography. The bands corresponding to free substrate (S), donor/donor, linear FSI (FSI) and circular HSI + FSI (HSI + FSI) products were quantified.

This part of the study was completed by the help and work of Dr. Vincent Parissi (University Bordeaux).

IV. XXXII. *In silico* analysis of R-loops and HIV-1 integration sites

The initial screen of the overlap between R-loop forming sites and HIV-1 integration sites was performed using the R statistical software via the application of the sequence alignment and analysis plugin Bioconductor, by using a cut-off window for sequence alignment of 50 kb. HIV-1 integration site maps were used from data generated from MDMs and CD4+ T cells (Cohn et al, 2015; Kok et al, 2016) and R-loop maps were used from data generated from the NT2 cell line and K562 cell line (Sanz et al, 2016).

This part of the study was completed by the help and work of Constantin Ahlmann-Eltze (Heidelberg University).

The second part of the screen was performed by using the extended list of RIGs from the Lusic lab (containing 11.537 genes) (Lucic et al, 2019) which was overlapped with R-loop maps from K562 cells (Sanz et al, 2016) using the R statistical software via the application of the sequence alignment and analysis plugin Bioconductor, by using a cut-off window for sequence alignment of 50 kb. As an internal control and as an outlier of the analysis, LAD maps from human fibroblasts (Guelen et al, 2008) and human Jurkat T cell line (Robson et al, 2017) were included into the analysis.

This part of the study was completed by the help and work of Maja Kuzman (University of Zagreb).

V. References

Abbas, W. (2013). T-Cell Signaling in HIV-1 Infection. The *Open Virology Journal*, 7: 57–71.

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., Hughes, S. H. Sarafianos, S. G., Brass, A. L. & Engelman, A. N. (2018). Capsid-CPSF6 Interaction Licences Nuclear HIV-1 Trafficking to Sites of Viral DNA Integration. *Cell Host and Microbe*, 24(3): 392-404.

Akhtar, A. & Gasser, S. M. (2007). The nuclear envelope and transcriptional control. *Nature Reviews Genetics*, 8: 507-517.

Albanese, A., Arosio, D., Terreni, M. & Cereseto, A. (2008). HIV-1 pre-integration complexes selectively target decondensed chromatin in the nuclear periphery. *PloS ONE*, 3: e2413.

Alkhatib, G., Combadiere C., Broder, C. C., Feng, Y., Kennedy, P. E., Murphy, P. M. & Berger, E. A. (1996). CC CKR5: a RANTES, MIP-1alpha, MIP-1beta receptor as a fusion cofactor for macrophage-tropic HIV-1. *Science*, 272(5270): 1955-1958.

Altschul, S. F., Madden,T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, 25: 3389-3402.

Antignano, F., Burrows, K., Hughes, M. R., Han, J. M., Kron, K. J., Penrod, N. M., Oudhoff, M. J., Wang, S. K. H., Min, P. H., Gold, M. J., Chenery, A. L., Braam, M. J. S., Fung, T. C., Rossi, F. M. V., McNagny, K. M., Arrowsmith, C. H., Lupien, M, Levings, M. K. & Zaph, C. (2014). Methyltransferase G9A regulates T cell differentiation during murine intestinal inflammation. *Journal of Clinical Investigation*, 124(5): 1945–1955.

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., Richman, D. D., Robertson, K. R., Kashuba, A. D., Bosch, R. J., Hazuda, D. J., Kuruc, J. D., Eron, J. J. & Margolis, D. M. (2014). HIV-1 expression within resting CD4+ T cells after multiple doses of vorinostat. *The Journal of Infectious Diseases*, 210(5): 728-735.

Arhel, N., Genovesio, A., Kim, K. A., Miko, S., Perret, E., Olivo-Marin, J. C., Shorte, S. & Charneau, P. (2006). Quantitative four-dimensional tracking of cytoplasmic and nuclear HIV-1 complexes. *Nature Methods*, 3: 817-824.

Arien, K. K., Vanham, G. & Arts, E. J. (2007). Is HIV-1 evolving to a less virulent form in humans?. *Nature Reviews Microbiology*, 5:141-151.

Ayouba, A., Souquieres, S., Njinku, B., Martin, P. M., Muller-Trutwin, M. C., Roques, P., Barre-Sinoussi, F., Mauclere, P., Simon, F. & Nerrienet., E. (2000). HIV-1 group N among HIV-1-seropositive individuals in Cameroon. *AIDS*, 14**:**2623-2625.

Bailey, T. L. & Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proceedings of the International Conference on Intelligent Systems for Molecular Biology*, 2: 28-36.

Baldauf, H., Pan, X., Erikson, E., Schmidt, S., Daddacha, D:, Burggraf, M., Schenkova, K., Ambiel, I., Wabnitz, G., Gramberg, T., Panitz, S:, Flory, E. Landau, N. R., Sertel, S., Rutsch, F., Lasitschka, F., Kim, B., König, R., Fackler, O. T. & Keppler, O. T. (2012). *Nature Medicine*, 18(11): 10.1038/nm.2964.

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., Andresdottir, V., Engelman, A. N., Costa, A. & Cherepanov, P. (2017). A supramolecular assembly mediates lentivirus DNA integration. *Science*, 355(6320): 93-95.

Baltimore, D. (1970). RNA-dependent DNA polymerase in virions of RNA tumour viruses. *Nature*, 226: 1209-1121.

Baltimore, D. (1971). Expression of Animal Virus Genomes. *Bacteriological Reviews*, 35(9): 235-241.

Bannister, A. J. & Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell Research*, 21: 381-395.

Barr. S. D., Ciuffi. A., Leipzig. J., Shinn. P., Ecker. J. R. & Bushman. F. D. (2006). HIV Integration Site Selection: Targeting in Macrophages and the Effects of Different Routes of Viral Entry. *MOLECULAR THERAPY*, 14(2): 218-225.

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

Beilhartz, G. L. & Götte, M. (2010). HIV-1 Ribonuclease H: Structure, Catalytic Mechanism and Inhibitors. *Viruses*, 2(4): 900-926.

Bejarano, D. A., Peng, K., Laketa, V, Börner, K., Jost, K. L., Lucic, B., Glass, B., Lusic, M., Müller, B. & Kräusslich, H. G. (2019). HIV-1 nuclear import in macrophages is regulated by CPSF6-capsid interactions at the Nuclear Pore Complex. *eLife*, 8:e41800.

Belotserkovskii, B. P., Liu, R., Tornaletti, S., Krasilnikova, M. M., Mirkin, S. M. & Hanawalt, P. C. (2010). Mechanisms and implications of transcription blockage by guanine-rich DNA sequences, *Proceedings of the National Academy of Sciences of the United States of America*, 107: 12816–12821.

Belotserkovskii, B. P. & Hanawalt, P. C. (2015). PNA binding to the non-template DNA strand interferes with transcription, suggesting a blockage mechanism mediated by Rloop formation. *Molecular Carcinogenesis*, 54: 1508-1512.

Benkirane, M., Chun, R. F., Xiao, H., Ogryzko, V. V., Howard, B. H., Nakatani, Y. & Jeang, K. T. (1998). Activation of integrated provirus requires histone acetyltransferase. p300 and P/CAF are coactivators for HIV-1 Tat. *Journal of Biological Chemistry*, 273(38): 24898-24905.

Benleulmi, M. S., Matysiak, J., Henriquez, D. R., Vaillant, C., Lesbats, P., Calmels, C., Naughtin, M., Leon, O., Skalka, A. M., Ruff, M., Lavigne, M., Andreola, M. L. & Parissi, V. (2015). Intasome architecture and chromatin density modulate retroviral integration into nucleosome. *Retrovirology*, 12: 13.

Bian, Q., Khanna, N., Alvikas, J. & Belmont, A. S. (2013). β-Globin cis-elements determine differential nuclear targeting through epigenetic modifications. *The Journal of Cell Biology*, 203(5): 767-783.

Bickmore, W. A. & van Steensel, B. (2013). Genome Architecture: Domain Organization of Interphase Chromosomes. *Cell*, 152: 1270-1284.

Bieniasz, P.D. (2009). The cell biology of HIV-1 virion genesis. *Cell Host and Microbe*, 5: 550–558.

Bhattacharya, A, Alam, S. L., Fricke, T., Zadrozny, K., Sedzicki, J., Taylor, A. B., Demeler, D., Pornillos, O., Ganser-Pornillos, B. K., Diaz-Griffero, F., Ivanov, D. N. & Yeager, M. (2014). Structural basis of HIV-1 capsid recognition by PF74 and CPSF6. *Proceedings of the National Academy of Sciences of the United States of America*, 111(52): 18625-18630.

Black, J. C., Van Rechem, C., Whetstine, J. R. (2012). Histone lysine methylation dynamics: establishment, regulation, and biological impact. *Molecular Cell*, 48: 491-507.

Blake, W. J., KAErn, M., Cantor, C. R. & Collins, J. J. (2003). Noise in eukaryotic gene expression. *Nature*, 422(6932): 633-637.

Blobel, G. (1985). Gene gating: a hypothesis. *Proceedings of the National Academy of Sciences of the United States of America*, 82(24): 8527-8529.

Boeger, H., Griesenbeck, J., Strattan, J. S., Kornberg, R. D. (2003). Nucleosomes unfold completely at a transcriptionally active promoter. *Molecular Cell*, 11(6): 1587-1598.

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

Boguslawski, S. J., Smith, D. E., Michalak, M. A., Mickelson, K. E., Yehle, C. O., Patterson, W. L. & Carrico, R. J.(1986). Characterization of monoclonal antibody to DNA.RNA and its application to immunodetection of hybrids. *Journal of Immunological Methods*, 1986;89: 123-130.

Bonev, J. & Cavalli, G. (2016). Organization and function of the 3D genome. *Nature Reviews Genetics*, 17: 661-678.

Boque-Sastre, R., Soler, M., Oliveira-Mateos, C., Portela, A., Moutinho, C., Sayols, S., Villanueva, A., Esteller, M. & Guil, S. (2015). Head-to-head antisense transcription and R-loop formation promotes transcriptional activation. *Proceedings of the National Academy of Sciences of the United States of America*, 5; 112(18): 5785-5790.

Bouchat, S., Gatot, J. S., abaya, K., Cardona, C., Collin, L., Herbein, G., De Wit, S, Clumeck, N., Lambotte, O., Rouzioux, C., Roh, O. & Van Lint, C. (2012). Histone methyltransferase inhibitors induce HIV-1 recovery in resting CD4(+) T cells from HIV-1-infected HAART-treated patients. *AIDS*, 26(12): 1473-1482.

Boveri T. (1090). Die Blastomerenkerne von Ascaris megalocephala und die Theorie der Chromosomenindividualität. *Archive für Zellforschung*, 3:181-268.

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

Bowlus, C. L. (2003). The role of iron in T cell development and autoimmunity. *Autoimmunity Reviews*, 2: 73-78.

Bradner, J. E., West, N. Grachan, M. L., Greenberg, E. F., Haggarty, S. J., Warnow, T. & Mazitschek, R. (2010). Chemical phylogenetics of histone deacetylases. *Nature Chemical Biology*, 6(3): 238-243.

Brady, T., Agosto, L. M., Malani, N., Berry, C. C., O'Doherty, U. & Bushman, F. (2009). HIV integration site distributions in resting and activated $CD4^+$ T cells infected in culture. *AIDS*, 23(12): 1461-1471.

Brass, A. L., Dykxhoorn, D. M., Benita, Y., Yan, N., Engelman, A.,Xavier, R. J., Lieberman, J. & Elledge, S. J. (2008). Identification of host proteins required for HIV infection through a functional genomic screen. *Science*, 319(5865): 921-926.

Braaten, D. & Luban, J. (2001). Cyclophilin A regulates HIV-1 infectivity, as demonstrated by gene targeting in human T cells. *EMBO JOURNAL*, 20: 1300-1309.

Briggs, J. A. & Kräusslich, H. G. (2011). The molecular architecture of HIV. *Journal of Molecular Biology*, 410(4): 491-500.

Brown, P. O. Bowerman, B, Varmus, H. E. & Bishop, J. M. (1987). Correct integration of retroviral DNA in vitro. *Cell*, 49(3): 347-356.

Brown, P. O., Bowerman, B., Varmus, H. E. & Bishop, J. M. (1989). Retroviral integration: Structure of the initial covalent complex and its precursor, and a role for the viral IN protein. *Proceedings of the National Academy of Sciences of the United States of America*, 86: 2525-2529.

Brown, C. R., Kennedy, C. J., Delmar, V. A., Forbes, D. J., Silver, P. A. (2008). Global histone acetylation induces functional genomic reorganization at mammalian nuclear pore complexes. *Genes and Development*, 22(5): 627-639.

Brown, J. D., Feldman, Z. B., Doherty, S. P., Reyes, J. M., Rahl, P. B., Lin, C. Y., Sheng, Q., Duan, Q., Federation, A. J., Kung, A. L., Haldar, S. M., Young, R. A., Plutzky, J. & Bradner, J. E. (2018). BET bromodomain proteins regulate enhancer function during adipogenesis. *Proceedings of the National Academy of Sciences of the United States of America*, 115(9): 2144-2149.

Brun-Vezinet, F., Barre-Sinoussi, F., Saimot, A. G., Christol, D., Montagnier, L., Rouzioux, C., Klatzmann, D., Rozenbaum, W., Gluckmann, J. C. & Chermann, J. C. (1984). Detection of IgG antibodies to lymphadenopathy-associated virus in patients with AIDS or lymphadenopathy syndrome. *Lancet*, 323: 1253-1256.

Buchwalter, A., Kaneshiro, J. M. & Hetzer, M. (2019). Coaching from the sidelines: the nuclear periphery in genome regulation. *Nature Reviews Genetics*, 20: 39-50.

Buck, M. J. & Lieb, J. D. (2006). A chromatin-mediated mechanism for specification of conditional transcription factor targets. *Nature Genetics*, 38(12): 1446-1451.

Buonaguro, L., Tornesello, M. L. & Buonaguro, F. M. (2007). Human Immunodeficiency Virus Type 1 Subtype Distribution in the Worldwide Epidemic: Pathogenetic and Therapeutic Implications. *Journal of Virology*, 81(19): 10209-10219.

Burinsky, M. I., Haggerty, S., Dempsey, M. P., Sharova, N., Adzhubel, A., Spitz, L., Lewis, P., Goldfarb, D., Emerman, M. & Stevenson, M. (1993). A nuclear localisation signal within HIV-1 matrix protein that governs infection of non-dividing cells. *Nature*, 365: 666-669.

Burke, B. & Stewart, C. L. (2013). The nuclear lamins: flexibility in function. *Nature Reviews Molecular Cell Biology*, 14: 13-24.

Bushman, F., D., Fujiwara, T. & Craigie, R. (1990). Retroviral DNA integration directed by HIV integration protein in vitro. *Science*, 28(4976): 1555-1558.

Bushman, F. (2003). Measuring covert HIV replication during HAART: the abundance of 2-LTR circles is not a reliable marker. *AIDS*, 17: 749-750.

Bushman, F., Lewinski, M., Ciuffi, A., Barr, S., Leipzig, J., Hannenhalli, S. & Hoffmann, C. (2005). Genome-wide analysis of retroviral DNA integration. *Nature Reviews Microbiology*, 3: 848-858.

Bushman, F. D., Malani1, N., Fernandes, J., D'Orso, I., Cagney, G., Diamond, T. L., Zhou, H., Hazuda, D. J., Espeseth, A. S., Konig, R., Bandyopadhyay, S., Ideker, T., Goff, S. P., Krogan, N. J., Frankel, A. D., Young. J. A. T. & Chanda, S. K. (2008). Cell Factors in HIV Replication: Meta-Analysis of Genome-Wide Studies. *PLoS Pathogens*, 5(5): e1000437.

Bushman, F. D., Malani, N., Fernandes, I., D'Orso, I., Cagney, G., Diamond, T. L., Zhou, H., Hazuda, D. J., Espeseth, A. S., König, R., Bandyopadhyay, S., Ideker, T., Goff, S. P., Krogan, N. J., Frankel, A. D., Young, J. A. T. & Chanda, S. K. (2009). Host cell factors in HIV replication: meta-analysis of genome-wide studies. *PLoS Pathogens*, 5: p. e1000437.

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

Busschots, K., Voet, A., De Maeyer, M., Rain, J.C., Emiliani, S., Benarous, R., Desender, L., Debyser, Z. & Christ, F. (2007). Identification of the LEDGF/p75 binding site in HIV-1 integrase. *Journal of Molecular Biology*, 365(5): 1480-1492.

Butler, S. L., Johnson, E. P. & Bushman, F. D. (2002). Human immunodeficiency virus cDNA metabolism: notable stability of two-long terminal repeat circles. *Journal of Virology*, 76: 3739-3747.

Campbell, E. M. & Hope, T. J. HIV-1 capsid: the multifaceted key player in HIV-1 infection. (2015). *Nature Reviews Microbiology*, 13: 471-483.

Capelson, M. & Hetzer M.W. (2009). The role of nuclear pores in gene regulation, development and disease. *EMBO Reports*, 10: 697-705.

Capelson, M., Doucet, C. & Hetzer, M. W. (2010). Nuclear pore complexes: guardians the nuclear genome. *Cold Spring Harbor Symposia on Quantitative Biology*, 75: 585-597.

Carteau, S., Hoffman, C. & Bushman, F. (1998). Chromosome structure and human immunodeficiency virus type 1 cDNA integration: centromeric alphoid repeats are a disfavored target. *Journal of Virology*, 72(5): 4005-4014.

Casciello, F., Windloch, K., Gannon, F. & Lee, J. S. (2015) Functional role of G9a histone methyltransferase in cancer. *Frontiers in Immunology*, 6: 487.

Castellano-Pozo, M., Santos-Pereira, J. M., Rondon, A. G., Barroso, S., Andujar, E., Perez-Alegre, M., Garcia-Muse, T. & Aguilera, A. (2013). R loops are linked to histone H3 S10 phosphorylation and chromatin condensation. *Molecular Cell*, 21;52(4): 583-590.

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

Cerritelli, S. M., Frolova, E. G., Feng, C. G., Grinberg, A., Love, P. E. & Crouch, R. J. (2003). Failure to produce mitochondrial DNA results in embryonic lethality in Rnaseh1 null mice. *Molecular Cell*, 11: 807-815.

Cerritelli, S. M. & Crouch, R. J. (2009). Ribonuclease H: the enzymes in eukaryotes. *FEBS Journal*, 276:1494-1505.

Chan, D. C., Berger, J. M. & Kim, P. S. (1997). Core structure of gp41 from the HIV envelope glycoprotein. *Cell*, 89(2): 263-273.

Chan, D. C. & Kim P. S. (1998). HIV entry and its inhibition. *Cell*, 93(5): 681-684.

Chan, Y. A., Hieter, P. & Stirling, P. C. (2014). Mechanisms of genome instability induced by RNA processing defects. *Trends in Genetics*, 30(6):245-253.

Chedin, F. (2016). Nascent Connections: R-loops and Chromatin Patterning. *Trends in Genetics*, 32(12): 828-838.

Chen, X., Skutt-Kakaria, K., Davison, J., Ou, Y., Choi, E:, Malik, P., Loeb, K., Wood, B, Georges, G:, Torok-Storb, B. & Paddison, P. J. (2012). G9a/GLP dependent histone H3K9me2 patterning during human hematopoietic stem cell lineage commitment. *Genes and Development*, 26: 2499-2511.

Chen, P. B., Chen, H. V., Acharya, D., Rando, O. J. & Fazzio, T. G. (2015). R loops regulate promoter-proximal chromatin architecture and cellular differentiation. *Nature Structural & Molecular Biology*, 22(12): 999-1007.

Chen, L., Chen, J. Y., Zhang, X., Gu, Y., Xiao, R., Shao, C., Tang, P., Qian, H., Luo, D., Li, H., Zhou, Y., Zhang, D. E. & Fu, X. D. (2017a). R-ChIP Using Inactive RNase H Reveals Dynamic Coupling of R-loops with Transcriptional Pausing at Gene Promoters. *Molecular Cell*, 16; 68(4): 745-757.

Chen, H.-C., Martinez, J. P., Zorita, E., Meyerhans, A. & Filion, G. J. Position effects influence HIV latency reversal. (2017b). *Nature Structural & Molecular Biology*, 24: 47-54.

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

Cherepanov, P., Sun, Z. Y., Rahman, S., Maertens, G., Wagner, G. & Engelman, A. (2005). Solution structure of the HIV-1 integrase-binding domain in LEDGF/p75. *Nature Structural & Molecular Biology*, 12(6): 526-532.

Chin, C. R., Perreira, L. M., Savidis, G., Portmann, J. M., Aker, A. M., Feeley, E. M., Smith, M. C. & Brass, A. L. (2015). Direct visualization of HIV-1 replication intermediates shows that capsid and CPSF6 modulate HIV-1 intra-nuclear invasion and integration. *Cell Reports*, 13: 1717-1731.

Choe, H., Farzan M., Sun, Y., Sullivan, N., Rollins, B., Ponath, P. D., Wu, L., Mackay, C. R., LaRosa, G., Newman, W., Gerard, N., Gerard, C. & Sodroski, J. (1996). The beta-chemokine receptors CCR3 and CCR5 facilitate infection by primary HIV-1 isolates. *Cell*, 85(7): 1135-1148.

Christ, F., Thys, W., De Rijck, J., Gijsbers, R., Albanese, A., Arosio, D., Emiliani, S., Rain, J. C., Benarous, R., Cereseto, A. & Debyer, Z. (2008). Transportin-SR2 imports HIV into the nucleus. *Current Biology* 18(16): 1192-1202.

Chun, T. W., Finzi D.,Margolick, J., Chadwich, K., Schwartz, D. & Siliciano, R. F. (1995). In vivo fate of HIV-1-infected T cells: quantitative analysis of the transition to stable latency. *Nature Medicine*, 1(12): 1284-1290.

Chun, T. W., Carruth, L., Finzi, D., Shen, X., Digiuseppe, J. A., Taylor, H., Hermankova, M., Chadwick, K., Margolick, J., Quinn, T. C., Kuo, Y. H., Brookmeyer, R., Zeiger, M. A., Barditch-Crovo, P., & Siliciano, R. F. (1997). Quantitation of latent tissue reservoirs and total body load in HIV-1 infection. *Nature*, 387**:**183-188.

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

Ciabrelli, B. & Cavalli, G. (2015). Chromatin-Driven Behavior of Topologically Associating Domains. *Journal of Molecular Biology*, 427: 608–625.

Ciechomska, I. A., Przanowski, P., Jackl, J., Wojtas, B. & Kaminska, B. (2016). BIX01294, an inhibitor of histone methyltransferase, induces autophagy-dependent differentiation of glioma stem-like cells. *Scientific Reports*, 6: 38723, doi: 10.1038/srep38732.

Ciuffi, A., Llano, M., Poeschla, E., Hoffmann, C., Leipzig, J., Shinn, P., Ecker, J.R. & Bushman, F. (2005). A role for LEDGF/p75 in targeting HIV DNA integration. *Nature Medicine*, 11: 1287-1289.

Clavel, F., Guyader, M., Guetard, D., Salle, M., Montagnier, L. & Alizon, M. (1986). Molecular cloning and polymorphism of the human immune deficiency virus type 2. *Nature*, 324(6098): 691-695.

Coffin, J. M, Hughes, S. H., Varmus, H. E. (eds). *Retroviruses* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1997).

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., Klein, F., Caskey, M., Walker, B. D., Siliciano, J. D:, Siliciano, R. F., Jankovic, M. & Nussenzweig, M. C. (2015). HIV-1 integration landscape during latent and active infection. *Cell, 160* (3): 420-432.

Col, E., Caron, C., Seigneurin-Berny, D., Gracia, J., Favier, A. & Khochbin, S. (2001). The histone acetyltransferase, hGCN5, interacts with and acetylates the HIV transactivator, Tat. *Journal of Biological Chemistry*, 276(30): 28179-28184.

Colak, D., Zaninovic, N., Cohen, M. S., Rosenwaks, Z., Yang, W. Y., Gerhardt, J., Disney, M. D., Jaffrey, S. R. (2014). Promoter-bound trinucleotide repeat mRNA drives epigenetic silencing in fragile X syndrome. *Science*, 343(6174): 1002-1005.

Collins, R. E., Tachibana, M., Tamaru, H., Smith, K. M., Jia, D., Zhang, X., Selker, E. U., Shinkai, Y. & heng, X. (2005). *In vitro* and *in vivo* analyses of a Phe/Tyr switch controlling product specificity of histone lysine methyltransferases. *Journal of Biological Chemistry,* 280: 5563–5570.

Collins, R. & Cheng, X. (2010). A case study in cross-talk: the histone lysine methyltrans-ferases G9a and GLP. *Nucleic Acids Research,* 38(11): 3503–3511.

Connor, R.I., Chen, B. K., Choe, S., Landau, N. R. (1995). Vpr is required for efficient replication of human immunodeficiency virus type-1 in mononuclear phagocytes. *Virology,* 206: 935-944.

Costin, J. M. (2007). Cytopathy Mechanisms of HIV-1. *Virology Journal,* 4:100, doi:10.1186/1743-422X-4-100.

Coull, J. J., Romerio, F., Sun, J. M., Volker, J. L., Galvin, K. M., Davie, J. R., Shi, Y., Hansen, U., Margolis, D. M. (2000). The human factors YY1 and LSF repress the human immunodeficiency virus type 1 long terminal repeat via recruitment of histone deacetylase 1. *Journal of Virology*, 74(15): 6790-6709.

Craigie, R. & Bushman, F. (2012). HIV DNA Integration. *Cold Spring Harbor Perspectives in Medicine*, 2(7): a006890.doi: 10.1101/cshperspect.a006890.

Cremer, T., Cremer, M., Dietzel, S., Müller, S., Solovei, I. & Fankan, S. (2006). Chromosome territories—a functional nuclear landscape. *Current Opinion in Cell Biology*, 18: 307-316.

Cremer, T. & Cremer, M. (2010). Chromosome territories. *Cold Spring Harbor Perspectives in Biology*, 2: a003889.

Cui, P., Löber, U., Alquezar-Planas, D. E., Ishida, Y., Courtiol, A., Timms, P., Johnson, R. N., Lenz, D., Helgen, K. M., Roca, A. L., Hartman, S. & Greenwood, A. D. (2016). Comprehensive profiling of retroviral integration sites using target enrichment methods from historical koala samples without an assembly reference genome. *Peer Journal*, 4: e1847.

Cummins, N. W. & Badley, A. D. (2010). Mechanisms of HIV-associated lymphocyte apoptosis: 2010. *Cell Death and Disease,* 1: e99. doi: 10.1038/cddis.2010.77.

Cuthbert,G. L., Daujat, S., Snowden, A. W., Erdjument-Bromage, H., Hagiwara, T., Yamada, M., Schneider, R., Gregory, P. D., Tempst, P., Bannister, A. J. & Kouzarides, T. (2004). Histone deimination antagonizes arginine methylation. *Cell*, 118(5): 545–553.

Dalgleish, A. G., Beverley P. C., Clapham, P. R., Crawford, D. H., Greaves, M. F. & Weiss, R. A. (1984). The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. *Nature*, 312(5996): 763-767.

D'Angelo, M. (2018). Nuclear pore complex as hubs for gene regulation. *Nucleus*, 9(1): 142-148.

Daniel, R., Ramcharan, J., Rogakou, E., Taganov, K. D., Greger, J. G., Bonner, W., Nussenzweig, A., Katz, R. A. & Skalka, A. M. (2004). Histone H2AX Is Phosphorylated at Sites of Retroviral DNA Integration but Is Dispensable for Postintegration Repair. *The Journal of Biological Chemistry*, 279: 45810-45814.

Davies, J. F., Hostomska, Z., Hostomsky, Z., Jordan, S. R. & Matthews, D. A. (1991). Crystal structure of the ribonuclease H domain of HIV-1 reverse transcriptase. *Science*, 252: 88–95.

De laco, A. & 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.

de Leeuw, R., Gruenbaum, Y. & Medalia, O. (2017). Nuclear lamins: thin filaments with major functions. *Trends in Cell Biology*, 28: 34-45.

De Rijck, J., Bartholomeeusen, K., Ceulemans, H., Debyser, Z. & Gijsbers, R. (2010). High-resolution profiling of the LEDGF/p75 chromatin interaction in the ENCODE region. *Nucleic Acids Research*, 38(18): 6135-6147.

Deeks, S. G., Overbaugh, J., Phillips, A. & Buchbinder, S. (2015): HIV infection. *Nature Reviews Disease Primers*, 1(1): 15035.doi:10.1038/nrdp.2015.35.

Dekker, J., Marti-Renom, M. A. & Mirny, L. A. (2013). Exploring the threedimensional organization of genomes: interpreting chromatin interaction data. *Nature Reviews in Genetics*, 14: 390–403.

Dekker, J. & Heard, E. (2015). Structural and functional diversity of Topologically Associating Domains. *FEBS Letters*, 589 (20 Pt A): 2877-2884.

Dekker, J. & Mirny, L. (2016). The 3D genome as moderator of chromosomal communication. *Cell*, 164: 1110-1121.

Demeulemeester, J., Vets, S., Schrijvers, R., Madlala, P., De Maeyer, M., De Rijck, J., Ndung'u, T., Debyser, Z. & Gijsbers, R. (2014). HIV-1 integrase variants retarget viral integration and are associated with disease progression in a chronic infection cohort. *Cell Host and Microbe*, 16:651–662.

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

Deng, H., Liu, R., Ellmeier, W., Choe, S., Unutmaz, D., Burkhart, M., Di Marzio, P., Marmon, S., Sutton, R. E., Hill, C. M., Davis, C. B., Peiper, S. C., Schall, T. J., Littman, D. R. & Landau, N. R. (1996). Identification of a major co-receptor for primary isolates of HIV-1. *Nature*, 381(6584): 661-666.

Descours, B., Cribier, A., Chable-Bessia, C., Ayinde, D., Rice, G., Crow, Y., Yatim, A., Schwartz, O., Laguette, N. & Benkirane, M. (2012). SAMHD1 restricts HIV-1 reverse transcription in quiescent CD4(+) T-cells. *Retrovirology*, 9(87): doi: 10.1186/1742-4690-9-87.

Dhayalan, A., Dimitrova, E., Rathert, P. & Jeltsch, A. (2009). A Continuous Protein Methyltransferase (G9a) Assay for Enzyme Activity Measurement and Inhibitor Screening. *Journal of Biomolecular Screening*, 14(9): 1129-1133.

Dion, M. F., Kaplan, T., Kim, M., Buratowski, S., Friedman, N. & Rando, O. J. (2007). Dynamics of replication-independent histone turnover in budding yeast. *Science*, 315: 1405-1408.

Di Nunzio, F., Danckaert, A., Fricke, T:, Perez, P., Fernandez, J., Perret, E., Roux, R., Shorte, S., Charneau, P., Diaz-Griffero, F. & Arhel, N. J. (2012). Human nucleoporins promote HIV-1 docking at the nuclear pore, nuclear import and integration. *PLoS One*; 7: e46037.

Di Nunzio, F., Fricke, T., Miccio, A., Valle-Casuso, J. C., Perez, P., Souque, P., Rizzi, E., Severgnini, M., Mavilio, F., Charneau, P. & Diaz-Griffero, F. (2013). Nup 153 and Nup 98 bind the HIV-1 core and contribute to the early steps of HIV-1 replication. *Virology*, 440: 8-18.

Di Primio, C., Quercioli, V., Allouch, A., Gijsbers, R., Christ, F., Debyser, Z., Arosio, D. & Cereseto, A. (2013). Single-cell imaging of HIV-1 provirus (SCIP). *Proceedings of the National Academy of Sciences of the United States of America*, 110: 5636-5641.

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

Dixon, L. R., Gorkin, D. U. & Ren, B. (2016). Chromatin domain: the unit of chromosome organization. *Molecular Cell*, 62: 668-680.

Doitsh G, Cavrois, M., Lassen, K. G., Zepeda, O., Yang, Z., Santiago, M. L., Hebbeler, A. M. & Greene, W. C. (2010). Abortive HIV infection mediates CD4 T cell depletion and inflammation in human lymphoid tissue. *Cell*, 143(5): 789-801.

Doms, R. W. (2001). Chemokine receptors and HIV entry. *AIDS*, 15(Suppl): S34-S35.

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. & Young, R. A. (2014). Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell*, 159: 374-387.

Drake, R. R., Neamati, N., Hong, H, Pilon, A. A., Sunthankar, P., Hume, S. D., Milne, G. W. A. & Pommier, Y. (1998). Identification of a nucleotide binding site in HIV-1 integrase. *Proceedings of the National Academy of Sciences of the United States of America*, 95(8): 4170-4175.

Duquette, M. L., Handa, P., Vincent, J. A., Taylor, A. F. & Maizels, N. (2004). Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. *Genes and Development*, 18: 1618–1629.

Easley, R., Carpio, L., Dannenberg, L., Choi, S., Alani, D., Van Duyne, R., Guendel, I., Klase, Z., Agbottah, E., Kehn-Hall, K. & Kashanchi, F. (2010). Transcription through the HIV-1 nucleosomes: effects of the PBAF complex in Tat activated transcription. *Virology*, 405(2): 322-333.

Eder, P. S., Walder, R. Y. & Walder, J. A. (1993). Substrate specificity of Human RNase H1 and its Role in Excision Repair of Ribose Residues Misincorporated in DNA. *Biochimie*, 75: 123-126.

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., Poirier, M. G., Foster, M. P. & Kvaratskhelia, M. (2013). Structural basis for high-affinity binding of LEDGF PWWP to mononucleosomes. *Nucleic Acids Research*, 41(6): 3924-3936.

Elgin, S. C. R. & Workman, J. L. (2000). *Chromatin Structure and Gene Expression*. Oxford University. Press, Oxford.

El Hage, A., Webb, S., Kerr, A. & Tollervey, D. (2014). Genome-Wide Distribution of RNA-DNA Hybrids Identifies RNase H Targets in tRNA Genes, Retrotransposons and Mitochondria. *PLoS Genetics*, 10(10): e1004716. doi:10.1371/journal.pgen.1004716

Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. (2002). Stochastic gene expression in a single cell. *Science*, 297(5584): 1183-1186.

Emiliani, S., Mousnier, A., Busschots, K., Maroun, M., Van Maele, B., Tempe, D., Vandekerckhove, L., Moisant, F., Ben-Slama, L., Witvrouw, M.,Christ, F., Rain, J. C., Dargemont, C., Debyzer, Z. & Benarous, R. (2005). Integrase mutants defective for interaction with LEDGF/p75 are impaired in chromosome tethering and HIV-1 replication. *Journal of Biological Chemistry*, 280: 25517-25523.

Endres, M. J., Clapham, P. R., Marsh, M., Ahuja, M., Turner, J. D., McKnight, A., Thomas, J. F., Stoebenau-Haggarty, B., Choe, S., Vance, P. J., Wells, T. N., Power, C. A., Sutterwala, S. S., Doms, R. W., Landau, N. R. & Hoxie, J. A. (1996). CD4-independent infection by HIV-2 is mediated by fusin/CXCR4. *Cell*, 87(4): 745-756.

Engelman, A., Mizuuchi, K. & Craigie, R. (1991). HIV-1 DNA integration: Mechanism of viral DNA cleavage and DNA strand transfer. *Cell*, 67(6): 2711-2715.

Engelman, A., Bushman, F. D. & Craigie, R. (1993). Identification of discrete functional domains of HIV-1 integrase and their organization within an active multimeric complex. *EMBO JOURNAL*, 12: 3269–3275.

Engelman, A., G. Englund, J. Orenstein, M. Martin, & R. Craigie. (1995). Multiple effects of mutations in human immunodeficiency virus type 1 integrase on viral replication. *Journal of Virology*, 69**:** 2729-2736.

Engelman, A. & Cherepanov, P. (2008). The lentiviral integrase binding protein LEDGF/p75 and HIV-1 replication. *PLoS Pathogens*, 4, e1000046.

Engelman, A. & Cherepanov, P. (2014). Retroviral integrase structure and DNA recombination mechanism. *Microbiology Spectrum*, 2(6): 1-22.

Esposito, D. & Craigie, R. (1998). Sequence specificity of viral end DNA binding by HIV-1 integrase reveals critical regions for protein–DNA interaction. *EMBO JOURNAL*, 17: 5832-5843.

Esteve, P.O., Chin, H. G., Smallwood, A., Feehery, G. R., Gangisetty, O., Karpf, A. R., Carey, M. F. & Pradhan, S. (2006). Direct interaction between DNMT1 and G9a coordinates DNA and histone methylation during replication. *Genes and Development*, 20(22): 3089–3103.

Farnet, C. M. & Haseltine, W. A. (1991). Circularization of human immunodeficiency virus type 1 DNA in vitro. *Journal of virology*, 65(1): 6942-6952.

Fazzio, T. G. (2016). Regulation of chromatin structure and cell fate by R-loops. *Transcription*, 7(4): 121-126.

Feldman, N., Gerson, A., Fang, J., Li, E., Zhang, Y., Shinkai, Y., Cedar, H. & Bergman, Y. (2006). G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis. *Nature Cell Biology*, 8(2): 188-194.

Feng, Y., Broder C. C., Kennedy, P. E. & Berger, E. A. (1996). HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. *Science*, 272(5263): 872-877.

Fernandez, J., Portilho, D. M., Danckaert, A., Munier, S., Becker, A., Roux, P., Zambo, A., Shorte, S., Jacob, Y., Vidalain, P.-O., Charneau, P., Clavel, F. & Arhel, N. J. (2015). Microtubule-associated proteins 1 (MAP1) promote human immunodeficiency virus type I (HIV-1) intracytoplasmic routing to the nucleus. *Journal of Biological Chemistry*, 290: 4631-4646.

Filippakopoulos, P., Qi, J., Picaud, S., Shen, Y., Smith, W. B., Fedorov, O., Morse, E. M., Keates, T., Hickman, T. T., Felletar, I., Philopott, M., Munro, S., McKeown, M.R., Wang, Y., Christie, A. L., West, N., Cameron, M. J., Schwartz, B., Heightman, T. D., La Thangue, N., French, C. A., West, A. L., Knapp, S. & Bradner, J. E. (2010). Selective inhibition of BET bromodomains. *Nature*, 468(7327): 1067-1073.

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

Fiserova, J., Efenberkova, M., Sieger, T., Maninova, M. Uhlirova, J. & Hozal, P. (2017). Chromatin organization at the nuclear periphery as revealed by image analysis of structured illumination microscopy data. *Journal of Cell Science*, 130: 2066-2077.

Floyd, S. R., Pacold, M. E., Huang, Q., Clarke, S. M., Lam, F. C., Cannell, I. G., Bryson, B. D:, Rameseder, J., Lee, M. J., Blake, A. J., Fydrych, A., Ho, R:, Greenberger, B. A., Chen, G. C., Maffa, A., Del Rosario, A. M., Root, D. E., Carpenter, A. E., Hahn, W. C., Sabatini, D. M., Chen, C. C., White, F. W., Bradner, J. E. & Yaffe, M. B. (2013). The Bromodomain Protein Brd4 Insulates Chromatin from DNA Damage Signaling. *Nature*, 498(7453): 246-250.

Forneris, F., Binda, C., Vanoni, M. A., Mattevi, A. & Battaglioli, E. (2005). Histone demethylation catalysed by LSD1 is a flavin-dependent oxidative process. *FEBS Letters*, 579(10): 2203–2207.

Fouchier, R. A., Meyer, B. E., Simon, J. H. M., Fischer, U., Albright, A. V., Gonzalez-Scarano, F. & Malim, M. H. (1998). Interaction of the human immunodeficiency virus type 1 Vpr protein with the nuclear pore complex. *Journal of Virology*, 72: 6004-6013.

Frank, C. L., Manandhar, D., Gordan, R. & Crawfors, G. E. (2016). HDAC inhibitors cause site-specific chromatin remodeling at PU.1-bound enhancers in K562 cells. *Epigenetics and Chromatin*, 15(9), doi:10.1186/s13072-016-0065-5.

Francis, A. C., Marin, M., Shi, J., Aiken, C. & Melikya, G. B. (2016). Time-Resolved Imaging of Single HIV-1 Uncoating *In Vitro*and in Living Cells. *PLoS Pathogens*, 12(6): e1005709.

Fricke, T., Valle-Casuso, J. C., White, T. E., Brandariz-Nunez, A., Bosche, W. J., Reszka, N., Gorelick, R. & Diaz-Griffero, F. (2013). The ability of TNPO3-depleted cells to inhibit HIV-1 infection requires CPSF6. *Retrovirology*, 10: 46.

Friedman, J., Cho, W.-K., Chu, C. K. & Keedy, K. S. (2011). Epigenetic Silencing of HIV-1 by the Histone H3 Lysine 27 Methyltransferase Enhancer of Zeste 2. *Journal of Virology*, 85(17): 9078-9089.

Freed, E. O. (2015). HIV-1 assembly, release and maturation. *Nature Reviews Microbiology*, 13(8): 484-496.

Fronza, R., Vasciaveo, A., Besno, A. & Schmidt, M. (2015). A Graph Base Framework to Model Virus Integration Sites. *Computational and Structural Biotechnology Journal*, 14: 69-77.

Fujiwara, T. & Mizuuchi, K. (1988). Retroviral DNA integration: Structure of an integration intermediate. *Cell*, 54: 497-504.

Füllgrabe, J., Kavanagh, E. & Joseph, B. (2011). Histone onco-modifications. *Oncogene*, 30:3391-3403.

Gaidamakov, S. A., Gorshkova, I. I., Schuck, P., Steinbach, P. J., Yamada, H., Crouch, R. J. & Cerritelli, S. M. (2005). Eukaryotic RNases H1 act processively by interactions through the duplex RNA-binding domain. *Nucleic Acids Research*, 33: 2166-2175.

Ganser-Pornillos, B. K., Yeager, M. & Sundquist, W. I. (2008). The structural biology of HIV assembly. *Current Opinion in Structural Biology*, 18(2): 203-217.

Garcia-Rubio, M. L., Perez-Calero, C., Barroso, S. I., Tumini, E., Herrera-Moyano, E., Rosado, I. V. & Aguilera, A. (2015). The Fanconi Anemia Pathway Protects Genome Integrity from R-loops. *PLoS Genetics*, 11, e1005674.

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., Sharp, P. M. & Hahn; B. H. (1999). Origin of HIV-1 in the chimpanzee *Pan troglodytes troglodytes*. *Nature*, 397**:** 436-441.

Gate, R. E., Cheng, C. S., Aiden, A. P., Siba, A., Tabaka, M., Lituiev, D., Machol, I., Gordon, G. I., Subramaniam, M., Shamim, M., Hougen, K. L., Wortman, I., Huang, S. C., Durand, N. C., Feng, T., De Jager, P. L., Chang, H. Y., Lieberman Aiden, E., Benoist, C., Beer, M. A., Ye, C. J. & Regev, A. (2018). Genetic determinants of co-accessible chromatin regions in activated T cells across humans. *Nature Genetics*, 50:1140-1150

Ge, H., Si, Y. & Reeder, R. G. (1998a). Isolation of cDNAs encoding novel transcription coactivators p52 and p75 reveals an alternate regulatory mechanism of transcriptional activation. *EMBO JOURNAL*, 17(22): 6723-6729.

Ge, H., Si, Y. & Wolffe, A. P. (1998b). A novel transcriptional coactivator, p52, functionally interacts with the essential splicing factor ASF/SF2. *Molecular Cell*, 2(6): 751-759.

Ghirlando, R. & Felsenfeld, G. (2016). CTCF: making the right connections. *Genes and Development*, 30: 881-891.

Gillim-Ross, L., Cara, A. & Klotman, M. E. (2005). HIV-1 extrachromosomal 2-LTR circular DNA is long-lived in human macrophages. *Viral Immunology*, 18: 190-196.

Ginno, P. A., Lott, P. L., Christensen, H. C., Korf, I. & Chedin, F. (2012). R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Molecular Cell*, 45: 814-825.

Ginno, P. A., Lim, Y. W., Lott, P. L., Korf, I. & Chedin, F. (2013). GC skew at the 5′ and 3′ ends of human genes links R-loop formation to epigenetic regulation and transcription termination. *Genome Research*, 23:1590-1600**.**

Goldberg, M., Harel, A. & Gruenbaum, T. (1999). The nuclear lamina: molecular organization and interaction with chromatin. *Critical Reviews in Eukaryotic Gene Expression*, 9(3-4): 285-293.

Goldgur, Y., Dyda, F., Hickman, A. B., Jenkins, T. M., Craigie, R. & Davies, D. R. (1998). Three new structures of the core domain of HIV-1 integrase: an active site that binds magnesium. *Proceedings of the National Academy of Sciences of the United States of America*, 95(16): 9150-9154.

Goldgur, Y., Craigie, R., Cohen, G. H., Fujiwara, T., Yoshinaga, T., Fujishita, T., Sugimoto, H., Endo, T., Murai, H. & Davies, D. R. (1999). Structure of the HIV-1 integrase catalytic domain complexed with an inhibitor: a platform for antiviral drug design. *Proceedings of the National Academy of Sciences of the United States of America*, 23:13040-13043.

Goldstone D. C., Ennis-Adeniran, V., Hedden, J. J., Groom, H. C., Rice, G. I., Christodoulou, E., Walker, P. A., Kelly, G., Haire, L. F., Yap, M. W., de Carvalho, L. P., Stoye, J. P., Croq, Y. J., Taylor, I. A. & Webb, M. (2011). HIV-1 restriction factor SAMHD1 is a deoxynucleoside triphosphate triphosphohydrolase. *Nature*, 480:379-382.

Golkaram, M., Jang, J. Hellander, S., Kosik, K. S. & Petzold, L. R. (2017). The Role of Chromatin Density in Cell Population Heterogeneity during Stem Cell Differentiation. *Scientific Reports*, 7: 133307.

Gomez, M., Wu, J., Schreiber, V., Dunlap, J., Dantzer, F., Wang, Y. & Liu, Y. (2006) PARP1 Is a TRF2-associated poly(ADP-ribose) polymerase and protects eroded telomeres. *Molecular Biology of the Cell*, 17(4): 1686-1696.

Goncalves, J., Moreira, E., Sequeira, I. J., Rodrigues, A. S., Rueff, J. & Bras, A. (2016). Integration of HIV in the Human Genome: Which Sites Are Preferential? A Genetic and Statistical Assessment. *International Journal of Genomics*, 2016: ID: 2168590, 6.

Greer, E. L. & Shi, Y. (2012). Histone methylation: a dynamic mark in health, disease and inheritance. *Nature Reviews Genetics*, 13(5): 343-357.

Grivel, J. C., Penn, M. L., Eckstein, D. A., Schramm, B., Speck, R. F., Abbey, N. W., Herndier, B., Margolis, L. & Goldsmith, M. A. (2000). Human immunodeficiency virus type 1 coreceptor preferences determine target T-cell depletion and cellular tropism in human lymphoid tissue. *Journal of Virology*, 74: 5347-5351.

Groh, M., Lufino, M. M., Wade-Martins, R. & Gromak, N. (2014). R-loops associated with triplet repeat expansions promote gene silencing in Friedreich ataxia and fragile X syndrome. *PLoS Genetics*. 10, e1004318.
hypersensitive sites in chromatin. *Annual Review of Biochemistry*, 57: 159-197.

Gruenbaum, Y. & Medalia, O. (2015). Lamins: the structure and protein complexes. *Current Opinion in Cell Biology*, 32: 7-12.

Grunseich, C., Wang, I. X., Watts, J. A., Burdick, J. T., Guber, R. D., Zhu, Z., Bruzel, A., Lanman, T., Chen, K., Schindler, A. B., Edwards, N., Ray-Chaudhury, A., Yao, J., Lehky, T., Piszczek, G., Crain, B., Fischbeck, K. H. & Cheung, V. G. (2018). Senataxin Mutation Reveals How R-Loops Promote Transcription by Blocking DNA Methylation at Gene Promoters. *Molecular Cell*, 1; 69(3): 426-437.

Guelen, L., Pagie, L., Brasset, E., Meuleman, W., Faza, M. B., Talhout, W., Eussen, B. H., de Klein, A., Wessels, L.,

de Laat, W. & van Steensel, B. (2008). Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature*, 453: 948-951.

Gurtler, L., Eberle, J., von Brunn, A., Knapp, S., Hauser, H. P., Zekeng, L., Tsague, J. M., Elegny, E. & Kaptue**.** L. (1994). A new subtype of human immunodeficiency virus type 1 (MVP-5180) from Cameroon. *Journal of Virology*, 68**:**1581-1585.

Halász, L., Karányi, Z., Boros-Oláh, B., Kuik-Rózsa, T., Sipos, E., Nagy, E., Mosolygó-L, A., Mázló, A., Rajnavölgyi, E., Halmos, G. & Székvölgyi1, L. (2017). RNA-DNA hybrid (R-loop) immunoprecipitation mapping: an analytical workflow to evaluate inherent biases. *Genome Research*, 27: 1063-1073.

Hamid, F. B., Kim, J. & Shin, C. G. (2016). Cellular and viral determinants of retroviral nuclear entry. *Canadian Journal of Microbiology*, 62: 1-15.

Hamid, F. B., Kim, J. & Shin, C.-G. (2017). Distribution and fate of HIV-1 unintegrated DNA species: a comprehensive update. *AIDS Research and Therapy*, 14:9 doi.org/10.1186/s12981-016-0127-6.

Hamperl, S., Bocek, M. J., Saldivar, J. C., Swigut, T. & Cimprich, K. A. (2017). Transcription-Replication Conflict Orientation Modulates R-Loop Levels and Activates Distinct DNA Damage Responses. *Cell*, 10; 170(4): 774-786.

Han, Y., Lassen, K., Monie, D., Sedaghat, A. R., Shimoji, S., Liu, X., Pierson, T. C., Margolick, J. B., Siliciano, R. F. & Siliciano, J. D. (2014). Resting $CD4^+$ T Cells from Human Immunodeficiency Virus Type 1 (HIV-1)-Infected Individuals Carry Integrated HIV-1 Genomes within Actively Transcribed Host Genes. *Journal of Virology*, 78(12): 6122-6133.

Handoko, L., Xu, H., Li, G., Ngan, C. Y., Chew, E., Schnapp, M., Lee, C. W., Ye, C., Ping, J. L., Mulawadi, F., Wong, E., Sheng, J., Zhang, Y., Poh, T., Chan, C. S., Kunarso, G., Shahab, A., Bourque, G., Cacheux-Rataboul, V., Sung, W. K., Ruan, Y. & Wei, C. L. (2011). CTCF-mediated functional chromatin interactome in pluripotent cells. *Nature Genetics*, 43: 630-638.

Harr, J. C., Luperchio, T. R., Wong, X., Cohen, E., Wheelan, S. J. & Reddy, K. L. (2015). Directed targeting of chromatin to the nuclear lamina is mediated by chromatin state and A-type lamins. *The Journal of Cell Biology*, 208(1): 33-52.

Hathaway, N. A., Bell, O., Hodges, C., Miller, E. L., Neel, D. S. & Crabtree, G. R. (2012). Dynamics and memory of heterochromatin in living cells. *Cell*, 149: 1447-1460.

Hatziioannou, T., Perez-Caballero, D., Cowan, S. & Bieniasz, P. D. (2005). Cyclophilin interactions with incoming human immunodeficiency virus type 1 capsids with opposing effects on infectivity in human cells. *Journal of Virology*, 79: 176-183.

Hearps, A. C. & Jans, D. A. (2007). Regulating the functions of HIV-1 matrix protein. *AIDS Research and Human Retroviruses*. 23(3): 341-346.

Hempel, K., Lange, H. W. & Birkofer, L. (1968). Epsilon-N-trimethyllysine, a new amino acid in histones. *Naturwissenschaften,* 55(1): 37.

Herbein, G. & Varin, A. (2010). The macrophage in HIV-1 infection: from activation to deactivation. *Retrovirology*, 7: 33.

Herrera-Moyano, E., Mergui, X., Garcia-Rubio, M. L., Barroso, S. & Aguilera, A. (2014). The yeast and human FACT chromatin-reorganizing complexes solve R-loop-mediated transcription-replication conflicts. *Genes and Development*, 1;28(7): 735-48.

Heuer, T. S. & Brown, P. O. (1997). Mapping features of HIV-1 integrase near selected sites on viral and target DNA molecules in an active enzyme-DNA complex by photo-cross-linking. *Biochemistry*, 36(35): 10655-10665.

Hilditch, L. & Towers, G. J. (2014). A model for cofactor use during HIV-1 reverse transcription and nuclear entry. *Current Opinion in Virology*, 4: 32-36.

Hill, M., Tachedjian, G. & Mak, J. (2005). The packaging and maturation of the HIV-1 Pol proteins. *Current HIV Research*, 3: 73-85.

Ho, D. D., Neumann, A. U., Perelson, A. S., Chen, J., Leonard, J. M. & Markowitz, M. (1995). Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infections. *Nature*, 373**:** 123-126.

Ho, J. C., Abdullah, L. N., Pang, Q. Y, Jha, S., Chow, EK-H, Yang, H., Kato, H., Poellinger, L., Ueda, J. & Lee, K. L. (2017) Inhibition of the H3K9 methyltransferase G9A attenuates oncogenicity and activates the hypoxia signaling pathway. *PLoS ONE*, 12(11): e018805.

Hogan, G. J., Lee, C. K. & Lieb, J. D. (2006). Cell cycle-specified fluctuation of nucleosome occupancy at gene promoters. *PLoS Genetics*, 2(9): e158.

Holaska, J. M. & Wilson, K. L. (2007). An emerin ‘‘proteome’’: purification of distinct emerin-containing complexes from HeLa cells suggests molecular basis for diverse roles including gene regulation, mRNA splicing, signaling, mechanosensing, and nuclear architecture. *Biochemistry*, 46: 8897-8908.

Hou, C., Li, L., Qin, Z. S. & Corces, V. G. (2012). Gene density, transcription, and insulators contribute to the partition of the *Drosophila* genome into physical domains. *Molecular Cell*, 48: 471-484.

Hrecka K, Hao, C., Gierszewska, M., Swanson, S. K., Kesik-Brodacka, M., Srivastava, S., Florens, L. Washburn, M. P. & Skowronski, J. (2011). Vpx relieves inhibition of HIV-1 infection of macrophages mediated by the SAMHD1 protein. *Nature*, 474: 658-661.

Hu, W. S. & Hughes, S. H. (2012). HIV-1 Reverse Transcription. *Cold Spring Harbor Perspectives in Medicine*, 2(10): a006882.doi: 10.1101/cshperspect.a006882. PMCID: PMC3475395.

Hu, J., Zhang, Y., Zhao, L., Frock, R. L., Du, Z., Meyers, R. M., Meng, F. L., Schatz, D. G. & Alt, F. W. (2015). Chromosomal Loop Domains Direct the Recombination of Antigen Receptor Genes. *Cell*, 163: 947-959.

Huertas, P. & Aguilera, A. (2003). Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. *Molecular Cell*, 12: 711-721.

Hughes, H., S. & Coffin, J., M. (2016). What Integration Sites Tell Us about HIV Persistence. *Cell Host and Microbe*, 19: 588-598.

Huisinga, K. L., Brower-Toland, B. & Elgin, S. C. (2006). The contradictory definitions of heterochromatin: transcription and silencing. *Chromosoma*, 15(2): 110-122.

Hyun, K., Jeon, J., Park, K. & Kim, J. (2017). Writing, erasing and reading histone lysine methylations. *Experimental and Molecular Medicine*, 49: e324.

Ibarra, A., Benner, C., Tyagi, S., Cool, J., Hetzer MW. (2016). Nucleoporin-mediated regulation of cell identity genes. *Genes and Development*, (20): 2253-2258.

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

Imai, K., Togami, H. & Okamoto, T. (2010). Involvement of histone H3 lysine 9 (H3K9) methyltransferase G9a in the maintenance of HIV-1 latency and its reactivation by BIX01294. *Journal of Biological Chemistry*, 285(22): 16538-16545.

Ito, T., Teo, Y. V., Evans, S. A., Neretti, N. & Sedivy, J. M., (2018). Regulation of Cellular Senescence by Polycomb Chromatin Modifiers through Distinct DNA Damage- and Histone Methylation-Dependent Pathways. *Cell Reports*, 22(13): 348-3492.

Ivanov, A. V., Bartosch, B. & Isaguliants, M. G. (2017). Oxidative Stress in Infection and Consequent Disease. *Oxidative Medicine and Cellular Longevity*, doi: 10.1155/2017/3496043.

Jacinto, F. V., Benner, C. & Hetzer, M. (2015). The nucleoporin Nup153 regulates embryonic stem cell pluripotency through gene silencing. *Genes and Development*, 29(12): 1224-1238.

Jackson, D.A. & Pombo, A. (1998). Replicon clusters are stable units of chromosome structure: evidence that nuclear organization contributes to the efficient activation and propagation of S phase in human cells. *Journal of Cell Biology*, 140: 1285-1295.

James, T. C. & Elgin, S. C. (1986). Identification of a nonhistone chromosomal protein associated with heterochromatin in Drosophila melanogaster and its gene. *Molecular and Cellular Biology*, 6(11): 3862-3872.

Jason, L. J., Moore, S. C., Lewis, J. D., Lindsey, G. & Ausio, J. (2002). Histone ubiquitination: A tagging tail unfolds? *BioEssays*, 24:166-174.

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

Jenkins, T. M., Esposito, D., Engelman, A. & Craigie, R. (1997). Critical contacts between HIV-1 integrase and viral DNA identified by structure-based analysis and photo-crosslinking. *EMBO JOURNAL*, 16(22): 6849-6859.

Ji, X., Dadon, D. B., Powell, B. E., Fan, Z. P., Borges-Rivera, D., Shachar, S., Weintraub, A. S., Hnisz, D., Pegoraro, G., Lee, T. I., Misteli, T., Jaenisch, R. & Young, R. A. (2016). 3D Chromosome Regulatory Landscape of Human Pluripotent Cells. Cell Stem Cell, 18: 262-275.

Jiang, G., Espeseth, A., Hazuda, D. J. & Margolis, D. M. (2007). c-Myc and Sp1 contribute to proviral latency by recruiting histone deacetylase 1 to the human immunodeficiency virus type 1 promoter. *Journal of Virology*, 81(20): 10914-1023.

John, S., Sabo, P. J., Thurman, R. E., Sung, M. H., Biddie, S. C., Johnson, T. A., Hager, G. L. & Stamatoyannopoulos, J. A. (2011). Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. *Nature Genetics*, 43(3): 264-268.

Johnson, A. A., Santos, W., Pais, G. C. G., Marchand, C., Amin, R., Burke Jr., T. R., Verdine, G. & Pommier, Y. (2006). Integration Requires a Specific Interaction of the Donor DNA Terminal 5′-Cytosine with Glutamine 148 of the HIV-1 Integrase Flexible Loop. *Journal of Biological Chemistry*, 281: 461-467.

Jonkers, I., Kwak, H., Lis, J. T. (2014). Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. *eLife*, 3: e02407.

Jordan, A., Defechereux, P. & Verdin, E. (2001). The site of HIV-1 integration in the human genome determines basal transcriptional activity and response to Tat transactivation. *EMBO JOURNAL*, 20(7): 1726-1738.

Kalhor, R., Tjong, H., Jayathilaka, N., Alber, F., & Chen, L. (2012). Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. *Nature Biotechnology*, 30: 90-98.

Kati, W. M., Johnson, K. A., Jerva, L. F. & Anderson, K. S. (1992). Mechanism and fidelity of HIV reverse transcriptase. *Journal of Biological Chemistry*, 25;267(36): 25988-25997.

Keele, B. F., Van Heuverswyin, F., Yingying, L., Bailes, E., Takehisa, J., Santiago, M., L., Bibollet-Ruche, F., Chen, Y., Wain, L. V., Liegeois, F., Loul, S., Ngole, E., M., Bienvenue, Y., Delaporte, E., Brookfield, J. F. Y., Sharp, P. M., Shaw, G. M., Peeters, M. & Hahn, B. H. (2006). Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. *Science*, 313(5786): 523-526.

Keene, M. A. & Elgin, S. C (1981). Micrococcal nuclease as a probe of DNA sequence organization and chromatin structure. *Cell*, 27(2):57–64.

Keppler, B. R. & Archer, T. (2008). Chromatin-modifying enzymes as therapeutic targets – Part 1. *Expert Opinion on Therapeutic Targets*, 12(10): 1301-1312.

Kessl, J. J., Kutluay, S. B., Townsend, D., Rebensburg, S., Slaughter, A., Larue, R., C., Shkriabai, N., Bakouche, N., Fuchs, J. R., Bieniasy, P. D. & Kvaratskhelia, M. (2016). HIV-1 Integrase Binds the Viral RNA Genome and Is Essential during Virion Morphogenesis. *Cell*, 166(5): 1257-1268.

Kim, H. D., Choe, J. & Seo, Y. S. (1999). The sen1(+) gene of Schizosaccharomyces pombe, a homologue of budding yeast SEN1, encodes an RNA and DNA helicase. *Biochemistry*. 38(44):14697-14710.

Kim, M. S., Blake, M., Baek, J. H., Kohlhagen, G., Pommier, Y. & Carrier, F. (2003). Inhibition of histone deacetylase increases cytotoxicity to cancer drugs targeting DNA. *Cancer Research*, 63: 7291-7300.

Kind, J. & van Steensel, B. (2010). Genome-nuclear lamina interactions and gene regulation. *Current Opinion in Cell Biology*, 22:320-325.

Kind, J., Pagie, L., Ortabozkoyun, H., Boyle, S., de Vries, S. S., Janssen, H., Amendola, M., Nolen, L. D., Bickmore, W. A. & van Steensel, B. (2013). Single-cell dynamics of genome-nuclear lamina interactions. *Cell*, 153:178-192.

Klatzmann, C., Laporte J.P., Achour, A., Brisson, E., Gruest, J., Montagnier, L. & Gluckman, J. C. (1986). Functional inhibition by cyclosporin A of the lymphocyte receptor for the AIDS virus (HIV). *Comptes rendus de l'Académie des Sciences*, 303(9): 343-348.

Klaver, B. & Berkhout, B. (1994). Comparison of 5' and 3' long terminal repeat promoter function in human immunodeficiency virus. *Journal of Virology*, 68(6): 3830-3840.

Kleiman, L. (2002). tRNA(Lys3): the primer tRNA for reverse transcription in HIV-1. *IUBMB Life*, 53(2):107-114.

Koenig, S., Gendelman, H. E., Orenstein, J. M., Dal Canto, M. C., Pezeshkpour, G. H., Yungbluth, M., Janotta, F., Aksamit, A., Martin, M. A. & Fauci, A. S. (1986). Detection of AIDS virus in macrophages in brain tissue from AIDS patients with encephalopathy. *Science*, 233: 1089-1093.

Koh, Y., Wu, X., Ferris, A. L., Matreyek, K. A., Smith, A. J., Lee, K., KewalRamani, V. N., Hughes, S. H. & 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. *Journal of Virology*, 87: 648-658.

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

Konig, R., Zhou, Y., Elleder, D., Diamond, T. L., Bonamy, G. M., Irelan, J. T., Chiang, C. Y., Tu, B. P., De Jesus, P. D., Lilley, C. E., Seidel, S., Opaluch, A. M., Caldwell, J. S., Weitzman, M. D., Kuhen, K. L., Bandyopadhyay, S., Ideker, T., Orth, A. P., Miraglia, L. J., Bushman, F. D., Young, J. A. & Chanda, S. K. (2008). Global analysis of host-pathogen interactions that regulate early-stage HIV-1 replication. *Cell*, 135(1): 49-60.

Koppensteiner, H. Brack-Werner, R. & Schinder, M. (2012). Macrophages and their relevance in Human Immunodeficiency Virus Type I infection. *Retrovirology*, 9(1): 82.

Krebs, F. C., Hogan, T. H., Quiterio, S., Gartner, S. & Wigdahl, B. (2001). Lentiviral LTR expression sequence variation, and disease pathogenesis. In Kuiken, C., Foley, B. B., Marx, P., McCutchan, F., Mellors, J. W., Wolinsky, S., Korber, B. (Ed), *HIV Sequence Compendium* (pp.29-70). Los Alamos, NM, Theoretical Biology and Biophysics Group, Los Alamos National Laboratory.

Krohne, G., Benavente, R., Scheer, U. & Dabauvalle, M. C. (2005). The nuclear lamina in Heidelberg and Wurzburg: A personal view. *European Journal of Cell Biology*, 84: 163-179.

Kubicek, S., O'Sullivan, R., August, E. M., Hickey, E. R., Zhang, Q., Teodoro, M. L., Rea, S., Mechtler, K., Kowalski, J. A., Homon, C. A., Kelly, T. A. & Jenuwein, T. (2007). Reversal of H3K9me2 by a Small-Molecule Inhibitor for the G9a Histone Methyltransferase. *Molecular Cell*, 25: 473-481.

Kulkosky, J., Jones, K. S., Katz, R. A., Mack, J. P. & 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. *Molecular and Cellular Biology*, 12: 2331–2338.

Kumar, A., Abbas, W. & Herbein, G. (2014). HIV-1 latency in monocytes/macrophages. *Viruses*, 6: 1837-1860.

Kumar Bid, H. & Kerk, S. (2016). BET bromodomain inhibitor (JQ1) and tumor angiogenesis. *Oncoscience*, 3(11-12): 316-317.

Kuznetsov, V. A., Bondarenko, V., Wongsurawat, T., Yenamandra, S. P. & Jenjaroenpun. P. (2018). Toward predictive R-loop computational biology: genome-scale prediction of R-loops reveals their association with complex promoter structures, G-quadruplexes and transcriptionally active enhancers. *Nucleic Acid Research*, 46(15): 7566-7585.

Kwon, S. H. & Workman, J. L. (2011). The changing faces of HP1: From heterochromatin formation and gene silencing to euchromatic gene expression: HP1 acts as a positive regulator of transcription. *Bioessays*, 33: 280-289.

Laguette N, Sobhina, B., Casartelli, N., Ringeard, M., Chable-Bessia, C., Segeral, E., Yatim, A., Emiliani, S., Schwartz, O. & Benkirane, M. (2011). SAMHD1 is the dendritic- and myeloid-cell–specific HIV-1 restriction factor counteracted by Vpx. *Nature*, 474: 654–657.

Laskey, S. B., Pohlmeyer, C. W., Bruner, K. M. & Siliciano, R. F. (2016). Evaluating Clonal Expansion of HIV-Infected Cells: Optimization of PCR Strategies to Predict Clonality. *PLoS Pathogens*, 5:12(8): e1005689.

Lassen, K., Han Y., Zhou, Y., Siliciano, J. & Siliciano, R. F. (2004). The multifactorial nature of HIV-1 latency. *Trends in Molecular Medicine*, 10(11): 525-531.

Leavitt, A. D., Shiue, L. & Varmus, H. E. (1993). Site directed mutagenesis of HIV-1 integrase demonstrates differential effects on integrase functions in vitro. *Journal of Biological Chemistry*, 268: 2113-2119.

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

Le Douce, V., Colin, L., Redel, L., Cherrier, T., Herbein, G., Aunis, D., Rohr, O., Van Lint, C. & Schwartz, C. (2012). LSD1 cooperates with CTIP2 to promote HIV-1 transcriptional silencing. *Nucleic Acids Research*, 40(5): 1904-1915.

Lee, M. S. & Craigie, R. (1994). Protection of retroviral DNA from autointegration: Involvement of a cellular factor. *Proceedings of the National Academy of Sciences of the United States of America*, 91: 9823-9827.

Lee, K., Ambrose, Z., Martin, T. D., Oztop, I., Mulky, A., Julias, J. G., Vandegraaff, N., Baumann, J. G., Wang, R., Yuen, W., Takemura, T., Shelton, K., Taniuchi, I., Li, Y., Sodroski, J., Littman, D. R., Coffin, J. M., Hughes, S. H.,

Unutmaz, D., Engelman, A. & KewalRamani, V. N. (2010). Flexible use of nuclear import pathways by HIV-1. *Cell host and Microbe*, 7: 221-233.

Lee, S. P., Xiao, J., Knutson, J. R., Lewis, M. S. & Han, M. K. (1997) Zn2+ promotes the self-association of human immunodeficiency virus type-1 integrase in vitro. *Biochemistry*, 36: 173-180.

Lee, C. K., Shibata, Y., Rao, B., Strahl, B. D. & Lieb, J. D. (2004). Evidence for nucleosome depletion at active regulatory regions genome-wide. *Nature Genetics*, 36(8): 900-905.

Lee, K., Mulky, A., Yuen, W., Martin, T. D., Meyerson, N. R., Choi, L., Yu, H., Sawyer, S. L. & KewalRamani, V. N. (2012). HIV-1 capsid-targeting domain of cleavage and polyadenylation specificity factor 6. *Journal of Virology*, 86: 3851-3860.

Lehnertz, B., Pabst, C., Su, L., Miller, M., Liu, F., Yi, L., Zhang, R., Krosl, J., Yung, E., Kirschner, J., Rosten, P., Underhill, T. M., Jin, J., Hebert, J., Sauvageau, G., Humphries, R. K. & Rossi, F. M., (2014). The methyltransferase G9a regulates HoxA9-dependent transcription in AML. *Genes and Development*, 28(4): 317-327.

Lelek, M., Casartelli, N., Pellin, D, Rizzi, E., Souque, P., Severgnini, M., Di Serio, C., Fricke, T., Diaz-Griffero, F., Zimmer, C., Charneau, P. & Di Nunzio, F. (2015). Chromatin organization at the nuclear pore favors HIV replication. *Nature Communications*, 6: doi:10.1038/ncommc7483.

Lemaitre, C. & Bickmore, W. A. (2015). Chromatin at the nuclear periphery and the regulation of genome functions. *Histochemistry and Cell Biology*, 144(2): 111-122.

Leemans, C., van der Zwalm, M. C. H., Brueckner, L., Comoglio, F., van Schaik, T., Pagie, L., van Arensbergen, J. & van Steensel, B. (2019). Promoter-Intrinsic and Local Chromatin Features Determine Gene Repression in LADs. *Cell*, 177: 852-864.

Lesbats, P., Métifiot, M., Calmels, C., Baranova, S., Nevinsky, G., Andreola, M. L. & Parissi, V. (2008) In vitro initial attachment of HIV-1 integrase to viral ends: control of the DNA specific interaction by the oligomerization state. *Nucleic Acids Research*, 36(22): 7043-7058.

Lesbats, P., Engelman, A. N. & Cherepanov, P. (2016). Retroviral DNA integration. *Chemical Reviews*, 116(20): 12730-12757.

Lewinski, M. K., Yamashita, M., Emerman, M., Ciuffi, A., Marshall, H., Crawford, G., Collins, F., Shinn, P., Leipzig, J., Hannenhalli, S., Berry, C. C., Ecker, J. R. & Bishman, F. D. (2006). Retroviral DNA integration: viral and cellular determinants of target-site selection. *PLoS Pathogens*, 2: e60.

Li, L., Olvera, J., M. Yoder, K. E., Mitchell, R. S., Buttler, S. L., Lieber, M., Martin, S. L. & Bushman, F. D. (2001). Role of the non-homologous DNA end joining pathway in the early steps of retroviral infection. *EMBO JOURNAL*, 20: 3271-3281.

Li, Y., Kar, A. K. & Sodroski, J. (2009). Target cell type-dependent modulation of human immunodeficiency virus type 1 capsid disassembly by cyclophilin A. *Journal of Virology*, 83: 10951-10962.

Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B. R., Sabo, P. J., Dorschner, M. O., Sandstrom, R., Bernstein, B., Bender, M. A., Groudine, M., Gnirke, A., Stamatoyannopoulos, J., Mirny, L. A., Lander, E. S. & Dekker, J. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, 326: 289-293.

Lifson, J. D, Feinberg, M. B., Reyes, G. R., Rabin, L., Banapour, B., Chakrabarti, S., Moss, B., Wong-Staal, F., Steimer, K. S. & Engleman, E. G. (1986). Induction of CD4-dependent cell fusion by the HTLV-III/LAV envelope glycoprotein. *Nature*, 323: 725-728.

Lim, Y. W., Sanz, L. A., Xu, X., Hartono, S. R. and Chedin, F. (2015). Genomwide DNA hypomethylation and RNA:DNA hybrid accumulation in Aicardi-Goutieres syndrome. *eLife*, 4, e08007.

Liu, J., Bartesaghi, A., Borgnia, M. J., Sapiro, G. & Subramanaim, S. (2008). Molecular architecture of native HIV-1 gp120 trimers. *Nature*, 455(7209): 109-113.

Liu, R., Huang, L., Li, J., Zhou, X., Zhang, H., Zhang, T., Lei, Y., Wang, K., Xie, N., Zheng, Y., Wang, F., Nice, E. C., Rong, L., Huang, C. & Wei, Y. (2013a). HIV Infection in Gastric Epithelial Cells. *The Journal of Infectious Diseases*, 208(8): 1221-1230.

Liu, F., Barsyte-Lovejoy, D., Li, F., Xiong, Y., Korboukh, V., Huang, X.-P., Allali-Hassani, A., Janzen, W. P., Roth, B. L., Frye, S. V., Arrowsmith, C. H., Brown, P. J., Vedadi, M. & Jin, J. (2013b). Discovery of an in vivo Chemical Probe of the Lysine Methyltransferases G9a and GLP. *Journal of Medical Chemistry*, 56(21): doi: 10.1021/jm401480r.

Liu, C., Perilla, J. R., Ning, J., Lu, M., Hou, G., Ramalho, R., Himes, B. A., Zhao, G., Bedwell, G. J., Byeon, I.-J., Ahn, J., Gronenborn, A. M., Prevelige, P. E., Rousso, I., Aiken, C., Polenova, T. Schulten, K. & Zhang, P. (2016). Cyclophilin A stabilizes the HIV-1 capsid through a novel non-canonical binding site. *Nature Communications*, 7: 10714.

Livak, K. J. & 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(4): 402-408.

Llano, M., Vanegas, M., Fregoso, O., Saenz, D., Chung, S., Peretz, M. & Poeschla, E. M. (2004a). LEDGF/p75 determines cellular trafficking of diverse lentiviral but not murine oncoretroviral integrase proteins and is a component of functional lentiviral preintegration complexes. *Journal of Virology*, 8(17): 9524-9537.

Llano, M., Delgado, S., Vanegas, M. & Poeschla, E. M. (2004b). Lens epithelium-derived growth factor/p75 prevents proteasomal degradation of HIV-1 integrase. *Journal of Biological Chemistry*, 279(53): 55570-55577.

Llano, M., Saenz, D. T., Meehan, A., Wongthida, P., Peretz, M., Walker, W. H., Teo, W. & Poeschla, E.M. (2006). An essential role for LEDGF/p75 in HIV integration. *Science*, 314: 461-464.

Luban, J. (2007) Cyclophilin A, TRIM5, and resistance to human immunodeficiency virus type 1 infection. *Jounal of Virology*, 81:1054-1061.

Luban, J. (2008). HIV-1 infection: Going nuclear with TNPO3/Transportin-SR2 and integrase. *Current Biology*, 18(16): R710–R713.

Lucas, J. S., Zhang, Y., Dudko, O. K. & Murre, C. (2014). 3D trajectories adopted by coding and regulatory DNA elements: first-passage times for genomic interactions. *Cell*, 158: 339-352.

Lucic, B., Chen, H.-C., Kuzman, M., Zorita, E., Wegner, J., Minneker, V., Roukos, V., Wang, W., Fronza, R., Schmidt, M., Benkirane, M., Stadhouders, R., Vlahovicek, K., Filion, G. J. & Lusic, M. (2019). Spatially clustered loci with multiple enhancers are frequent targets of HIV-1. *Manuscript submitted for publication.*

Lukic, Z., Dharan, A., Fricke, T., Diaz-Giffero, F. & Campbell, E. M. (2014). HIV-1 uncoating is facilitated by dynein and kinesin 1. *Journal of Virology*, 88: 13613-13625.

Lusic, M., Marcello, A., Cereseto, A. & Giacca, M. (2003). Regulation of HIV-1 gene expression by histone acetylation and factor recruitment at the LTR promoter. *EMBO JOURNAL*, 22(24): 6550-6561.

Lusic, M. & Giacca, M. (2014). Regulation of HIV-1 latency by chromatin structure and nuclear architecture. *Journal of Molecular Biology*, 427(3): 688-694.

Lusic, M. & Siliciano, R. F. (2017). Nuclear landscape of HIV-1 infection and integration. *Nature Reviews in Microbiology*, 15(2): 69-82.

Ma, H., Samarabandu, J., Devdhar, R. S., Acharya, R., Cheng, P. C., Meng, C. & Berezney, R. (1998). Spatial and temporal dynamics of DNA replication sites in mammalian cells. *Journal of Cell Biology*, 143: 1415-1425.

Ma, B. G., Chen, L., Ji, H. F, Chen, Z. H., Yang, F. R., Wang, L., Qu, G., Jiang, Y. Y., Ji, C. & Zhang, H. Y. (2008). Characters of very ancient proteins. *Biochemical and Biophysical Research Communications*, 366(3): 607-611.

Maertens, G., Cherepanov, P., Pluymers, W., Busschots, K., De Clercq, E., Debyser, Z. & Engelborghs, Y. (2003). LEDGF/p75 is essential for nuclear and chromosomal targeting of HIV-1 integrase in human cells. *Journal of Biological Chemistry*, 278: 33528–33539.

Maddon, P. J., Dalgleish, A. G., McDougal, J. S., Clapham, P. R., Weiss, R. A. & Axel, R. (1986). The T4 gene encodes the AIDS virus receptor and is expressed in the immune system and the brain. *Cell*, 47: 333–348.

Maignan, S., Guilloteau, J. P., Zhou-Liu, Q., Clement-Mella, C. & Mikol, V. (1998). Crystal structures of the catalytic domain of HIV-1 integrase free and complexed with its metal cofactor: high level of similarity of the active site with other viral integrases. *Journal of Molecular Biology*, 282(2): 359-368.

Majorek, K. A., Dunin-Horkawicz, S., Steczkiewicz, K., Muszewska, A., Nowotny, M., Ginalski, K. & Bujnicki, J. M. (2014). The RNase H-like superfamily: new members, comparative structural analysis and evolutionary classification. *Nucleic Acids Research*, 42(7): 4160-4179.

Maldarelli, F., Wu, X., Su, L., Simonetti, F. R., Shao, W., Hill, S., Spindler, J., Ferris, A. L., Mellors, J. W., Kearney, M. F., Coffin, J. M. & Hughes, S. H. (2014). HIV latency. Specific HIV integration sites are linked to clonal expansion and persistence of infected cells. *Science, 345*(6193): 179-183.

Maldarelli, F. (2016). The role of HIV integration in viral persistence: no more whistling past the proviral graveyard. *Journal of Clinical Investigation*, 126(2): 438-447.

Manzo, S. G., Hartono, S. R., Sanz, L. A., Marinello, J., De Biasi, S., Cossarizza, A., Capranico, G, Chedin, F. (2018). DNA Topoisomerase I differentially modulates R-loops across the human genome. *Genome Biology*, 19:100.

Margolick, J. B., Volkman, D. J.,Folks, T. M. & Fauci, A. S. (1987). Amplification of HTLV-III/LAV infection by antigen-induced activation of T cells and direct suppression by virus of lymphocyte blastogenic responses. *Journal of Immunology*, 138**:** 1719-1723.

Marban, C., Redel, L., Suzanne, S., Van Lint, C., Lecestre, D., Chasserot-Golaz, S., Leid, M., Aunis, D., Schaeffer, E. & Rohr, O. (2005). COUP-TF interacting protein 2 represses the initial phase of HIV-1 gene transcription in human microglial cells. *Nucleic Acids Research*, 33(7): 2318-2331.

Marini, B., Kertesz-Farkas, A., Ali, H., Lucic, B., Lisek, K., Manganaro, L., Pongor, S., Luzzati, R., Recchua, A., Mavilio F., Giacca, M & Lusic, M. (2015). Nuclear architecture dictates HIV-1 integration site selection. *Nature*, 521(7551): 227-231.

Markowitz, M., Nguyen, B. Y., Gotuzzo, E., Mendo, F., Ratanasuwan, W., Kovacs, C., Prada, G., Morales-Ramirez, J. O., Crumpacker, C. S., Osaacs, R. D., Gilde, L. R., Wan, H., Miller, M. D., Wenning, L. A., Teppler, H. & Protocol 004 Part II Study Team. (2007). Rapid and durable antiretroviral effect of the HIV-1 Integrase inhibitor raltegravir as part of combination therapy in treatment-naive patients with HIV-1 infection: results of a 48-week controlled study. *Journal of Acquired Immune Deficiency Syndromes*, 46(2): 125-33.

Marzio, G., Tyagi, M., Gutierrez, M. I. & Giacca, M. (1998). HIV-1 tat transactivator recruits p300 and CREB-binding protein histone acetyltransferases to the viral promoter. *Proceedings of the National Academy of Sciences of the United States of America*, 95(23): 13519-13524.

Matreyek, K. A. & Engelman, A. (2011). The requirement for nucleoporin NUP153 during human immunodeficiency virus type 1 infection is determined by the viral capsid. *Journal of Virology*, 85(15): 7818-7827.

Matreyek, K. A. & Engelman, A. (2013). Viral and cellular requirements for the nuclear entry of retroviral preintegration nucleoprotein complexes. *Viruses*, 5: 2483-2511.

Matreyek, K. A., Yücel, S. S. & Li, X. & Engelman, A. (2013). Nucleoporin NUP153 phenylalanine-glycine motifs engage a common binding pocket within the HIV-1 capsid protein to mediate lentiviral infectivity. *PLoS Pathogens*, 9: e1003693.

Matsuda, Y., Kobayashi-Ishihara, M., Fujikawa, D., Ishida, T., Watanabe, T. & Yamagishi, M. (2015). Epigenetic heterogeneity in HIV-1 latency establishment. *Scientific Reports*, 5: 7701.

Matysiak, J., Lesbats, P., Mauro, E., Lapaillerie, D., Dupuy, J.-W., Lopez, A. P., Benleulmi, M. S., Calmels, C., Andreola, M.-L., Ruff, M., Llano, M., Delelis, O., Lavigne, M. & Parissi, V. (2017). *Retrovirology*, 14: 39.

McClure, M.O., Marsh M. & Weiss, R. A. (1988). Human immunodeficiency virus infection of CD4-bearing cells occurs by a pH-independent mechanism. *EMBO JOURNAL*, 7(2): 513-518.

McDonald, D., Vodicka, M. A., Lucero, G., Svitkina, T. M., Borisy, G. G, Emerman, M. & Hope, T. J. (2002). Visualization of the intracellular behavior of HIV in living cells. *Journal of Cell Biology*, 159(3): 441-452.

Meersseman, G., Pennings, S. & Bradbury, E. M. (1992). Mobile nucleosomes-a general behaviour. *EMBO JOURNAL*, 11: 2951–2959.

Melikyan, G.B. (2008). Common principles and intermediates of viral protein-mediated fusion: the HIV-1 paradigm. *Retrovirology*, 5: 111.

Meuleman , W., Peric-Hupkes, D., Kind, J., Beaudry, J. B., Magie, L., Kellis, M., Reinders, M., Wessels, L., & van Steensel, B. (2013). Constitutive nuclear lamina-genome interactions are highly conserved and associated with A/T rich sequence. *Genome Research*, 23: 270-280.

Michael, N. L. (1999). Host genetic influences on HIV-1 pathogenesis. *Current Opinion in Immunology*, 11**:** 466-474.

Michieletto, D., Lusic, M. & Marenduzzo, D. (2019). Physical principles of retroviral integration in the human genome. *Nature Communications*, 10: 575.

Miri, L., Bouvier, G., Kettani, A., Mike, A., Wakrim, L., Nilges, M. & Malliavin, T. E. (2014). Stabilization of the integrase-DNA complex by Mg2+ ions and prediction of key residues for binding HIV-1 integrase inhibitors. *Proteins*, 82(3): 466-478.

Misteli, T. (2007). Beyond the Sequence: Cellular Organization of Genome Function. *Cell*, 128: 787-800.

Mitchell, R. S., Beitzel, B. F., Schröder, A. R., Shinn P, Chen, H., Berry, C. C., Ecker, J. R. & Bushman, F. D. (2004). Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. *PLoS Biology*, 2: E234-10.1371/journal.pbio.0020234.

Miyauchi, K., Kim Y., Latinovic, O., Morozo, V. & Melikyan, G. B. (2009). HIV enters cells via endocytosis and dynamin-dependent fusion with endosomes. *Cell*, 137(3): 433-444.

Montes de Oca, R., Shoemaker, C. J., Gucek, M., Cole, R. N. & Wilson, K. L. (2009). Barrier-to-autointegration factor proteome reveals chromatin-regulatory partners. *PLoS One*, 4: e7050.

Morita, E. & Sundquist, W. I. (2004). Retrovirus budding. *Annual Review of Cell and Developmental Biology*, 20: 395-425.

Morey, L. & Helin, K. (2010). Polycomb group protein-mediated repression of transcription. *Trends in Biochemical Sciences*, 35: 323-332.

Mouscadet, J. & Tchertanov, L. (2009). Raltegravir: molecular basis of its mechanism of action. *European Journal of Medical Research*, 14(Suppl 3): 5-16.

Murray, K. (1964). THE OCCURRENCE OF EPSILON-N-METHYL LYSINE IN HISTONES. *Biochemistry*, 3: 10-15.

Murray, K. (1994). The Occurrence of iε-N-Methyl Lysine in Histones. *Biochemistry*, 3(1): 10-15.

Murray, J. M., McBride, K., Boesecke, C., Bailey, M., Amin, J., Suzuki, K., Baker, D., Zaunders, J. J., Emery, S., Cooper, D. A., Koelsck, K., K., Kelleher, A. D., PINT STUDY TEAM. (2012). Integrated HIV DNA accumulates prior to treatment while episomal HIV DNA records ongoing transmission afterwards. *AIDS*, 26(5): 543-550.

Nadel, J., Athanasiadou, R., , Lemetre, C., Wijetunga, A., Ó Broin, P., Sato, H., Zhang, Z., Jeddeloh, J., Montagna, C., Golden, A., Seoighe, C. & Greally, J. M. (2015). RNA:DNA hybrids in the human genome have distinctive nucleotide characteristics, chromatin composition, and transcriptional relationships. *Epigenetics and Chromatin*, 8: 46.

Nadler, S. G., Tritschler, D., Haffar, O. K., Blake, J., Bruce, A. G. & Cleaveland, K. S. (1997). Differential expression and sequence-specific interaction of karyopherin alpha with nuclear localization sequences. *Journal of Biological Chemistry*, 272: 4310-4315.

Nagano, T., Lubling, Y., Stevens, T. J., Schoenfelder, S., Yaffe, E., Dean, W., Laue, E. D., Tanay, A. & Fraser, P. (2013). Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature*, 502(7469): 59-64.

Nakanishi, S., Sanderson, B. W., Delventhal, K. M., Bradford, W. D., Staehling Hampton, K. & Shilatifard, A. (2008). A comprehensive library of histone mutants identifies nucleosomal residues required for H3K4 methylation. *Nature Structural Molecular Biology*, 15(8): 881-888.

Nojima, T., Tellier, M., Foxwell, J., Ribeiro de Almeida, C., Tan-Wong, S. M., Dhir, S., Dujardin, G., Dhir, A., Murphy, S. & Proudfoot, N. J. (2018). Deregulated Expression of Mammalian lncRNA through Loss of SPT6 Induces R-Loop Formation, Replication Stress, and Cellular Senescence. *Molecular Cell*, 72(6): 970-984.

Noordermeer, D., de Wit, E., Klous, P., van de Werken, H., Simonis, M., Lopez-Jones, M., Eussen, B., de Klein, A., Singer, R. H. & de Laat, W. (2011). Variegated gene expression caused by cell-specific long-range DNA interactions. *Nature Cell Biology*, 13: 944-951.

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., Gribnau, J., Barillot, E., Blüthgen, N., Dekker, J. & Heard, E. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*, 485: 381-385.

Nowotny, M., Gaidamakov, S. A., Crouch, R. J. &Yang, W. (2005). Crystal structures of RNase H bound to an RNA/DNA hybrid: substrate specificity and metal-dependent catalysis. *Cell*, 121:1005–1016.

Nguyen, K., Das, B., Dobrowolski, C. & Karn, J. (2017). Multiple histone lysine methyltransferases are required for the establishment and maintenance of HIV-1 latency. mBio 8: e00133-17.

Okoye, A. A. & Picker, L. J. (2013). $CD4^+$ T cell depletion in HIV infection: mechanisms of immunological failure. *Immunological Reviews*, 254(1): 54-64.

Op de Coul, E. L. M., Prins, M., Cornelissen, M., van der Schoot, A., Boufassa, F., Brettle, R. P., Hernandez-Aguado, I., Schiffer, V., McMenamin, J., Rezza, G., Robertson, R., Zangerle, R., Goudsmit, J., Coutinho, R. A. & Lukashov, V. (2001). Using phylogenetic analysis to trace HIV-1 migration among western European injecting drug users seroconverting from 1984 to 1997. *AIDS*, 15**:** 257-266.

Pace, M. J., Graf, E. H., Agosto, L. M., Mexas, A. M., Male, F., Brady, T., Bushman, F. D. & O'Doherty, U. (2012). Directly Infected Resting CD4+T Cells Can Produce HIV Gag without Spreading Infection in a Model of HIV Latency. *PLoS Pathogens*, 8(7): e1002818.

Ott, M., Schnölzer, M., Garnica, J., Fischle, W., Emiliani, S., Rackwitz, H. R. & Verdin, E. (1999). Acetylation of the HIV-1 Tat protein by p300 is important for its transcriptional activity. *Current Biology*, 9(24): 1489-1492.

Pace, M. J., Graf, E. H., Agosto, L. M., Mexas, A. M., Male, F., Brady, F., Bushman, F. D. & O'Doherty, U. (2012). Directly Infected Resting CD4+T Cell Can Produce HIV Gag without Spreading Infection in a Model of HIV Latency. *Plos Pathogens*: doi.org/10.1371/journal.ppat.1002818

Pace, M. J., Graf, E. H. & O'Doherty, U. (2013). HIV 2-long terminal repeat circular DNA is stable in primary CD4+T Cells. *Virology*, 441(1): 18-21.

Paik, W. K. & Kim, S. (1967). E-N-dimethyllysine in histones. *Biochemical and biophysical research communications*, 27: 479-483.

Pascual-Garcia, P., Debo, B., Aleman, J. R., Talamas, J. A., Lan, Y., Nguyen, N. H., Won, K. J. & Capelson, M. (2017). Metazoan Nuclear Pores Provide a Scaffold for Poised Genes and Mediate Induced Enhancer-Promoter Contacts. *Molecular Cell*, 68(1): 63-76.

Pawlica, P., Le Sage, V., Poccardi, N., Tremblay, M. J., Mouland, A. J. & Berthoux, L. (2014). Functional evidence for the involvement of microtubules and dynein motor complexes in TRIM5alpha-mediated restriction of retroviruses. *Journal of Virology*, 88: 5661-5676.

Peel, S., Macheboeuf, P., Martinelli, N. & Weissenhorn, W. (2011). Divergent pathways lead to ESCRT-III-catalyzed membrane fission. *Trends in Biochemical Science*, 36: 199-210.

Peng, K., Muranyi, W., Glass, B., Laketa, V. Yant, S. R., Tsai, L., Cihlar, T., Müller, B. & Kräusslich, H. G. (2014). Quantitative microscopy of functional HIV post-entry complexes reveals association of replication with the viral capsid. *eLife*, 3: e04114.

Perelson, A. S., Neumann, A. U., Markowitz, M., Leonard, J. M. & Ho, D. D. (1996). HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. *Science*, 271: 1582-1586.

Peric-Hupkes, D., Meuleman, W., Pagie, L., Bruggeman, S. W., Solovei, I., Brugman, W., Gräf, S., Flicek, P., Kerkhovrn, R. M., van Lohuizen, M., Reinders, M., Wessels, L. & van Steensel, B. (2010). Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. *Molecular Cell*, 38: 603-613.

Peters, A. H., O'Carroll, D., Scherthan, H., Mechtler, K., Sauer, S., Schofer, C., Weipoltshammer, K., Pagani, M., Lachner, M., Kohlmaier, A., Opravil, S., Doyle, M., Sibilia, M. & Jenuwein, T. (2001). Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. *Cell*, 107(3): 323-337.

Phillips, D. D., Garboczi, D. N., Singh, K., Hu, Z., Leppla, S. H. & Leysath, C. E. (2013). The sub-nanomolar binding of DNA- RNA hybrids by the single-chain Fv fragment of antibody S9.6. *Journal of Molecular Recognition*, 26: 376-381.

Phillips-Cremins, J. E. & Corces, V. G. (2013). Chromatin insulators: linking genome organization to cellular function. *Molecular Cell*, 50: 461-474.

Phillips-Cremins, J. E., Sauria, M. E., Sanyal, A., Gerasimova, T. I., Lajoie, B. R., Bell, J. S., Ong, C. T., Hookway, T. A., Guo, C., Sun, Y., Bland, M. J., Wagstaff, W., Dalton, S., McDevitt, T. C., Sen, R., Dekker, J., Taylor, J. & Corces, V. G. (2013). Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell*, 153: 1281-1295.

Piekna-Przybylska, D., Sharma, G., Maggirwar, S. B. & Bambara, R. A. (2017). Deficiency in DNA damage response, a new characteristic of cells infected with latent HIV-1. *Cell Cycle*, 16(10): 968-978.

Poeschla, E. M. (2014). Integrase, LEDGF/p75 and HIV Replication. *Cellular and Molecular Life Sciences*, 65(9): 1403-1424.

Poles, M. A., Elliott, J., Taing, P., Anton, P. A., Chen, I. S. (2001). A preponderance of CCR5(+) CXCR4(+) mononuclear cells enhances gastrointestinal mucosal susceptibility to human immunodeficiency virus type 1 infection. *Journal of Virology*, 75: 8390-8399.

Pombo, A. & Dillon, N. (2015). The nuclear envelope affects genome organization and function. *Nature Reviews Molecular Cell Biology*, 16:245–257.

Pope, B. D., Ryba, T., Dileep, V., Yue, F., Wu, W., Denas, O., Vera, D. L., Wang, Y., Hansen, R. S., Canfield, T. K., Thurman, R. E., Cheng, Y., Gülsoy, G., Dennis, J. H., Snyder, M. P., Stamatoyannopoulos, J. A., Taylor, J., Hardison, R. C., Kahveci, T., Ren, B. & Gilbert, D. M. (2014). Topologically associating domains are stable units of replication-timing regulation. *Nature*, 515: 402-405.

Popov, S., Rexach, M., Zybarth, G., Reiling, N., Lee, M. A., Ratner, L., Lane, C. M., Moore, M. S., Blober, G. & Bukrinsky, M. (1998). Viral protein R regulates nuclear import of HIV-1 preintegration complex. *EMBO JOURNAL*, 17: 909-917.

Pornillos, O. & Ganser-Pornillos, B. K. (2013). HIV-1 Virion Structure. *Encyclopedia of AIDS*. doi:10.1007/978-1-4614-9610-6_56-1.

Powell, R. D., Holland, P. J., Hollis, T. & Perrino, F.W. (2011). Aicardi-Goutieres syndrome gene and HIV-1 restriction factor SAMHD1 is a dGTP-regulated deoxynucleotide triphosphohydrolase. *Journal of Biological Chemistry*, 286: 43596-43600.

Pradeepa, M. M., Sutherland, H. G., Ule, J., Grimes, G. R. & Bickmore, W. A. (2012). Psip1/Ledgf p52 Binds Methylated Histone H3K36 and Splicing Factors and Contributes to the Regulation of Alternative Splicing. *PLoS Genetics*, 8(5), e1002717.

Price, A. J., Fletcher, A. J., Schaller, T., Elliott, T., Lee, K., KewalRamani, V. N., Chin, J. W., Towers, G. J. & James, L. C. (2012). CPSF6 defines a conserved capsid interface that modulates HIV-1 replication. *PLoS Pathogens*, 8: e1002896.

Pruss, D., Reeves, R., Bushman, F. D. & Wolffe, A. P. (1994a). The influence of DNA and nucleosome structure on integration events directed by HIV integrase. *Journal of Biological Chemistry*, 269: 25031-25041.

Pruss, D., Bushman, F. D. & Wolffe, A. P. (1994b). Human immunodeficiency virus integrase directs integration to sites of severe DNA distortion within the nucleosome core. *Proceedings of the National Academy of Sciences of the United States of America*, 91: 5913-5917.

Pryciak, P. M. & Varmus, H. E. (1992). Nucleosomes, DNA-binding proteins, and DNA sequence modulate retroviral integration target site selection. *Cell*, 69: 769-780.

Quarona, V., Zaccarello, G., Chillemi, A., Brunetti, E., Singh, V. K., Ferrero, E., Funaro, A., Horenstein A. L. & Malavasi, F. (2013). CD38 and CD157: a long journey from activation markers to multifunctional molecules. *Cytometry Part B: Clinical Cytometry*, 84(4), doi: 10.1002/cyto.b.21092.

Quercioli, V., Di Primio, C., Casini, A., Mulder, L. C. F., Vranckx, L. S., Borrenberghs, D., Gijsbers, R., Debyser, Z. & Cereseto, A. (2016). Comparative Analysis of HIV-1 and Murine Leukemia Virus Three-Dimensional Nuclear Distributions. *Journal of Virology*, 90(10): 5205-5209.

Rabl, C. (1885). Über Zelltheilung. *Morphologisches Jahrbuch*, 10: 214.

Radman-Livaja M. & Rando, O. J. (2010). Nucleosome positioning: how is it established, and why does it matter? *Developmental Biology*, 339(2): 258-266.

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. & Aiden, E. L. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, 159: 1665-1680.

Rao, V. K., , Ow, J. R., , Shankar, S. R., Bharathy, N., Manikandan, J., Wang, Y., & Taneja, R. (2016). G9a promotes proliferation and inhibits cell cycle exit during myogenic differentiation. *Nucleic Acids Research*, 44(17): 8129-8143.

Raices, M., Bukata, L., Sakuma, S., Borlido, J., Hernandez, L. S., Hart, D. O. & D'Angelo, M. A. (2017). Nuclear pores regulate muscle development and maintenance by assembling a localized mef2C complex. *Developmental Cell*, 41(5): 540-554.

Raices, M. & D'Angelo, M. A. (2017). Nuclear pore complexes and regulation of gene expression. *Current Opinion in Cell Biology*, 46: 26-32.

Rankovic, S., Varadarajan, J., Ramalho, R., Aiken, C. & Roussoa, I. (2017). Reverse Transcription Mechanically Initiates HIV-1 Capsid Disassembly. *Journal of Virology*, 91(12): e00289-17.

Rasheedi, S., Shun, M., Serrao, E., Sowd, G. A, Qian, J., Hao, C., Dasgupta, T., Engelman, A. N. & Skowronski, J. (2016). The Cleavage and Polyadenylation Specificity Factor 6 (CPSF6) Subunit of the Capsid-recruited Pre-messenger RNA Cleavage Factor I (CFIm) Complex Mediates HIV-1 Integration into Genes. *Journal of Biological Chemistry*, 291(22): 11809-11819.

Reeves, J. D. & Doms, R. W. (2002). Human Immunodeficiency Virus Type 2. *Journal of General Virology*, 83(6): 1253-1265.

Reil, H. Bukovsky, A. A., Gelderblom, H., R. & Gottlinger, H. (1998). Efficient HIV-1 replication can occur in the absence of the viral matrix protein. *EMBO JOURNAL*, 17: 2699-2708.

Reynoso, R., Wieser, M., Ojeda, D., Bonisch, M., Kuhnel, H., Bolcic, F., Quendler, H., Grillari, J., Grillari-Voglauer, R. & Quarleri, J. (2012). HIV-1 induces telomerase activity in monocyte-derived macrophages - safeguarding one of its reservoirs. *Journal of Virology*, 86(19): 10327-10337.

Rezaei, S. D. & Cameron, P. U. (2015). Human Immunodeficiency Virus (HIV-1) Integration Sites in Viral Latency. *Current HIV/AIDS Reports*, 12(1): 88-96.

Rice, J. C., Briggs, S. D., Ueberheide, B., Barber, C. M., Shabanowitz, J., Hunt, D. F., Shinkai, Y. & Allis, C. D. (2003) Histone methyltransferases direct different degrees of methylation to define distinct chromatin domains. *Molecular Cell*, 12(6): 1591-1598.

Richmond, T. J. & Davey, C A. (2003). The structure of DNA in the nucleosome core. *Nature*, 423: 145-150.

Roberts, J. D., Bebenek, K. & Kunkel**.**, T. A. (1988). The accuracy of reverse transcriptase from HIV-1. *Science*, 242**:** 1171-1173.

Robson, M. I., de las Heras, J. I., Czapiewski, R., Sivakumar, A., Kerr, A. R. W. & Schirmer, E. C. (2017). Constrained release of lamina-associated enhancers and genes from the nuclear envelope during T-cell activation facilitates their association in chromosome compartments. *Genome Research*, 27: 1-13.

Rogel, M. E., Wu, L. I. &Emerman, M. (1995). The human immunodeficiency virus type 1 vpr gene prevents cell proliferation during chronic infection. *Journal of Virology*, 69: 882–888.

Romerio, F., Gabriel, M. N., Margolis, D. M. (1997). Repression of human immunodeficiency virus type 1 through the novel cooperation of human factors YY1 and LSF. *Journal of Virology*, 71(12): 9375-9382.

Rossetto, D., Avvakumov, N. & Côté, J. (2012). Histone phosphorylation**;** A chromatin modification involved in diverse nuclear events**.** *Epigenetics*, 7(10): 1098-1108.

Rothbart, S. B., Dickson, B. M., Raab, J. R., Grzybowski, A. T., Krajewski, K., Guo, A. H., Shanle, E. K., Josefowicz, S. Z., Fuchst, S. M., Allis, C. D., Magnuson, T. R., Ruthenburg, A. J. & Strahl, B. D. (2015). An interactive database for the assessment of histone antibody specificity. *Molecular Cell*, 59(3): 502-511.

Rouquette, J., Cremer, C., Cremer, T. & Fakan, S. (2010). Functional nuclear architecture studied by microscopy: present and future. *International Review of Cell and Molecular Biology*, 282: 1-90.

Roy, D., Yu, K. & Lieber, M., R. (2008). Mechanism of R-loop formation at immunoglobulin class switch sequences. *Molecular Cell Biology*, 28: 50-60.

Roy, D. & Lieber, M. R. (2009). G clustering is important for the initiation of transcription induced R-loops *in vitro*, whereas high G density without clustering is sufficient thereafter, *Molecular Cell Biology*, 29: 3124-3133.

Roy, D., Zhang, Z., Lu, Z., Hsieh, C. L. & Lieber, M., R. (2010). Competition between the RNA transcript and the nontemplate DNA strand during R-loop formation in vitro: a nick can serve as a strong R-loop initiation site. *Molecular Cell Biology*, 30:146-159.

Rüegsegger, U., Beyer, K. & Keller, W. (1996). Purification and characterization of human cleavage factor Im involved in the 3' end processing of messenger RNA precursors. *Journal of Biological Chemistry*, 271(11): 6107-6113.

Ruttkay-Nedecky, B., Nejdl, L., Gumulec, J., Zitka, O., Masarik, M., Eckschlager, T., Stiborova, M., Adam, V., & Kizek, R. (2013). The Role of Metallothionein in Oxidative Stress. *International Journal of Molecular Science*,14: 6044-6066.

Saito, A., Henning, M. S., Serrao, E., Dubose, B. N., Teng, S., Huang, J., Li, X., Saito, N., Prasad Roy, S., Siddiqui, M. A., Ahn, J., Tsuji, M., Hatziioannou, T., Engelman, A. N. & Yamashitaa, M. (2016). The capsid-CPSF6 interaction is dispensable for HIV-1 replication in primary cells but is selected during virus passage *in vivo*. *Journal of Virology*, 90(15): 6981-6935.

Sandoval-Montes, C. & Santos-Argumedo, L. (2005). CD38 expressed selectively during the activation of a subset of mature T cells with reduced proliferation but improved potential to produce cytokines. *Journal of Leukocyte Biology*, 77(4): 513-521.

Santos-Pereira, J. M., Herrero, A. B., Garcia-Rubio, M. L., Marin, A., Moreno, S. & Aguilera, A. (2013). The Npl3 hnRNP prevents R-loop-mediated transcription-replication conflicts and genome instability. *Genes and Development*, 27(22): 2445-2458.

Santos-Pereira, J. M. & Aguilera, A. (2015). R-loops: new modulators of genome dynamics and function. *Nature Reviews Genetics*, 16, 583-597.

Sanz, L. A., Hartono, S. R., Lim, Y. W., Steyaert, S., Rajpurkar, A., Ginno, P. A., Xu, X. & Chedin, F. (2016). Prevalent, Dynamic, and Conserved R-Loop Structures Associate with Specific Epigenomic Signatures in Mammals. *Molecular Cell*, 63(1): 167-178.

Savarino, A. (2007). *In-Silico* docking of HIV-1 integrase inhibitors reveals a novel drug type acting on an enzyme/DNA reaction intermediate. *Retrovirology*, 4(21). dos:10.1186/1742-4690-4-21.

Scadden, D. T., Zeira, M., Woon, A., Wang, Z., Schieve, L., Ikeuchi, K., Lim, B. & Groopman, J. E. (1990). Human immunodeficiency virus infection of human bone marrow stromal fibroblasts. *Blood*, 76: 317-322.

Schaller, T., Ocwieja, K. E., Rasaiyaah, J., Price, A. J., Brady, T. L., Roth, S. L., Hue, S., Fletcher, A. J., Lee, K., KewalRamani, V. N., Noursadeghi, M., Jenner, R. G., Jamer, L. C., Bushman, F. D. & Towers, G. J. (2011). HIV-1 capsid-cyclophilin interactions determine nuclear import pathway, integration targeting and replication efficiency. *PLoS Pathogens*, 7: e1002439

Schermelleh, L. Carlton, P. M., Haase, S., Shao, L., Winoto, L., Kner, P., Burke, B., Cardoso, M. C., Agard, D. A., Gustafsson, M. G., Leonhardt, H. & Sedat, J. W.c(2008). Subdiffraction multicolor imaging of the nuclear periphery with 3D structured illumination microscopy. *Science*, 320: 1332-1336.

Schmidt, M., Schwarzwaelder, K., Bartholomae, C., Zaoui, K., Ball, C., Pilz, I., Braun, S., Glimm, H. & von Kalle, C. (2007). High-resolution insertion-site analysis by linear amplification-mediated PCR (LAM-PCR). *Nature Methods*, *4*(12): 1051-1057.

Schones, D. E, Cui, K., Cuddapah, S., Roh, T. Y., Barski, A., Wang, Z., Wei, G. & Zhao, K. (2008). Dynamic regulation of nucleosome positioning in the human genome. *Cell*, 132(5): 887-898.

Schupbach, J., Popovic, M., Gilden, R. V., Gonda, M. A., Sarngadharan, M. G. & Gallo, R. C. (1984). Serological analysis of a subgroup of human T-lymphotropic retroviruses (HTLV-III) associated with AIDS. *Science*, 224: 503-505.

Schröder, A. R., Shinn, P., Chen, H., Berry, C., Ecker, J. R. & Bushman, F. (2002). HIV-1 integration in the human genome favors active genes and local hotspots. *Cell*, 110: 521-529.

Segal, E., Fondufe-Mittendorf, Y., Chen, L., Thastrom, A., Field, Y., Moore, I. K., Wang, J. P. & Widom, J. (2006). A genomic code for nucleosome positioning. *Nature*, 442: 772-778.

Sengupta, S. & Siliciano, R. F. (2018). Targeting the Latent Reservoir for HIV-1. *Immunity*, 48(5): 872-895.

Serrao, E., Krishnan, L., Shun, M.C., Li, X., Cherepanov, P., Engelman, A. & 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 Research*, 42: 5164-5176.

Seto, E. & Yoshida, M. (2014). Erasers of histone acetylation: the histone deacetylase enzymes. *Cold Spring Harbor Perspectives in Biology*, 6(4): a018713.

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A. & Cavalli, G. (2012). Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell*, 148: 458-472.

Sexton, T. & Cavalli, G. (2015). The role of chromosome domains in shaping the functional genome. *Cell*, 160: 1049-1059.

Shah, V. B., Shi, J., Hout, D. R., Oztop, I., Krishnan, L., Ahn, J., Shotwell, M. S., Engelman, A. & Aiken, C. (2013). The host proteins transportin SR2/TNPO3 and cyclophilin A exert opposing effects on HIV-1 uncoating. *Journal of Virology*, 87: 422-432.

Shankar, S. R., Bahirvani, A. G., Rao, V. K., Bharathy, N., Ow, J. R. & Taneja, R. (2013). G9a, a multipotent regulator of gene expression. *Epigenetics*, 8(1): 16-22.

Sharkey, M., Triques, K., Kurtzkes, D. R. & Stevenson, M. (2005). In vivo evidence for instability of episomal human immunodeficiency virus type 1 cDNA. *Journal of Virology*, 79(8): 5203-5210.

Sharma, P., Wagner, K., Wolchok, J.D. & Allison, J.P. (2011). Novel cancer immunotherapy agents with survival benefit: recent successes and next steps. *Nature Reviews Cancer*, 11(11): 805-812.

Sharp, P. M. & Hahn, B. H. (2011). Origin of HIV and the AIDS pandemic. *Cold Spring Harbor Perspectives in Medicine*, 1(1): a006841.

Shen, R., Richter, H. E., Clements, R. H., Novak, L., Huff, K., Bimczok, D., Sankaran-Walters, S., Dandekar, S., Clapham, P. R., Smythies, L. E. & Smith, P. D. (2009). Macrophages in vaginal but not intestinal mucosa are monocyte-like and permissive to human immunodeficiency virus type 1 infection. *Journal of Virology*, 83: 3258-3267.

Shen, R., Richter, H. E. & Smith, P. D. (2011). Early HIV-1 target cells in human vaginal and ectocervical mucosa. *American Journal of Reproductive Immunology*, 65: 261-267.

Shahbazian, M. D., Zhang, K. & Grunstein, M. (2005). Histone H2B ubiquitylation con-trols processive methylation but not monomethylation by Dot1 and Set1. *Molecular Cell*, 19(2): 271-277.

Sharp, P. M. & Hahn, B. H. (2011). Origins of HIV and the AIDS pandemic. *Cold Spring Harbor Perspectives in Medicine*, 1: a006841.

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., Pace, M. J., Siliciano, R. F., O'Doherty, U., Guatelli, J., Panelles, V. & Bushman, F. D. (2013). HIV latency and integration site placement in five cell-based models. *Retrovirology*, 10: 90.

Shirakawa, K., Chavez, L., Hakre, S., Calvanese, V. & Verdin, E. (2013). Reactivation of latent HIV by histone deacetylase inhibitors. *Trends in Microbiology*, 21(6): 277-285.

Shivaswamy, S., Bhinge, A., Zhao, Y., Jones, S., Hirst, M., Iyer & V. R. (2008). Dynamic remodeling of individual nucleosomes across a eukaryotic genome in response to transcriptional perturbation. *PLoS Biology*, 6(3): e65.

Shoemaker, C., Goff, S., Gilboa, E., Paskind, M., Mitra, S. W. & Baltimore, D. (1980). Structure of a cloned circular Moloney murine leukemia virus DNA molecule containing an inverted segment: implications for retrovirus integration. *Proceedings of the National Academy of Sciences of the United States of America*, 77: 3932-3936.

Shytaj, I. L., Nickel, G., Arts, E., Farrell, N., Biffoni, M., Pal, R., Chung, H. K., LaBranche, C., Montefiori, D., Vargas-Inchaustegui, D., Robert-Guroff, M., Lewis, M. G., Sacha, J. B., Palamara, A. T. & Savarino, A. (2015). Two-Year Follow-Up of Macaques Developing Intermittent Control of the Human Immunodeficiency Virus Homolog Simian Immunodeficiency Virus SICmac251 in the Chronic Phase of infection. *Journal of Virology*, 89(15): 7521-7535.

Shytaj, I. L., Lucic, B., Forcato, M., Billingsley, J. M., Bosinger, S., Stanic, M., Gregoretti, F., Antonelli, L., Oliva, G., Frese, C. K., Trifunovic, A., Galy, B., Eibl, C., Silvestri, G., Bicciato, S., Savarino, A. & Lusic, M. (2019). Alterations of redox and iron metabolism accompany development of HIV-1 latency. *Unpublished manuscript*.

Simon, F., Mauclère, P., Roques, P., Loussert-Ajaka, I., Müller-Trutwin, M. C., Saragosti, S., Georges-Courbot, M. C., Barré-Sinoussi, F. & Brun-Vézinet, F. (1998). Identification of a new human immunodeficiency virus type 1 distinct from group M and group O. *Nature Medicine*, 4**:** 1032-1037.

Simonis, M., Klous, P., Splinter, E., Moshkin, Y., Willemsen, R., de Wit, E., van Steensel, B. & de Laat, W. (2006). Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nature Genetics*, 38: 1348-1354.

Simms, P. E. & Ellis, T. M. (1996). Utility of flow cut-metric detection of CD69 expression as a rapid method for determining poly- and oligoclonal lymphocyte activation. *Clinical and Vaccine Immunology*, 3(3): 301-304.

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., Kvaratskhelia, M. & Levin, H. L. (2015). LEDGF/p75 interacts with mRNA splicing factors and targets HIV-1 integration to highly spliced genes. *Genes and Development*, 29: 2287-2297.

Skourti-Stathaki, K., Proudfoot, N. J. & Gromak, N. (2011). Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Molecular Cell*, 42(6): 794-805.

Skourti-Stathaki, K. & Proudfoot, N. J. (2014). A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. *Genes and Development*, 28(13): 1384-13896.

Skourti-Stathaki, K., Torlai Triglia, E., Warburton, M., Voig, P., Bird, A. & Pombo, A. (2019). R-Loops Enhance Polycomb Repression at a Subset of Developmental Regulator Genes. *Molecular Cell*, 73: 930–945.

Smith, C. L., & Peterson, C. L. (2005). ATP-dependent chromatin remodeling. *Current Topics in Developmental Biology*, 65: 115-148.

Sloan, R. D. & Wainberg, M. A. (2011). The role of unintegrated DNA in HIV infection. *Retrovirology*, 8:52.

Somech, R., Shaklai, S., Geller, O., Amariglio, N., Simon, A. J., Rechavi, G. & Gal-Yam, E. N. (2005). The nuclear-envelope protein and transcriptional repressor LAP2beta interacts with HDAC3 at the nuclear periphery, and induces histone H4 deacetylation. *Journal of Cell Science*, 118: 4017-4025.

Sokolskaja, E., Sayah, D. M. & Luban, J. (2004). Target cell cyclophilin A modulates human immunodeficiency virus type 1 infectivity. *Journal of Virology*, 78: 12800-12808.

Sowd, G. A., Serrao, E., Wang, H., Wang, W:, Fadel, H. J. Poeschla, E. M. & Engelman, A. N. (2016). A critical role for alternative polyadenylation factor CPSF6 in targeting HIV-1 integration to transcriptionally active chromatin. *Proceedings of the National Academy of Sciences of the United States of America*, 113: E1054-E1063.

Splinter, E., Heath, H., Kooren, J., Palstra, R. J., Klous, P., Grosveld, F., Galjart, N., Laat de, W. (2006). CTCF mediates long-range chromatin looping and local histone modification in the β-globin locus. *Genes and Development*, 20: 2349-2354.

Stein, B.S., Gowda, S.D., Lifson, J. D., Penhallow, R. C., Bensch, K. G. & Engelman, E. G. (1987). pH-independent HIV entry into CD4-positive T cells via virus envelope fusion to the plasma membrane. *Cell*, 49(5): 659-668.

Stevens, S. W. & Griffith, J. D. (1994). Human immunodeficiency virus type 1 may preferentially integrate into chromatin occupied by LlHs repetitive elements (retroviral integration/repetitive DNA elements/chromatin structure). *Proceedings of the National Academy of Sciences of the United States of America*, 91: 5557-5561.

Stork, C. T., Bocek, M., Crossley, M. P., Sollier, J., Sanz, L. A., Chedin, F., Swigut. T. & Cimprich, K. A. (2016). Co-transcriptional R-loops are the main cause of estrogen-induced DNA damage. *e Life*, 5: e17548.

Studitsky, V. M., Kassavetis, G. A., Geiduschek, E. P. & Felsenfeld, G. (1997). Mechanism of transcription through the nucleosome by eukaryotic RNA polymerase. *Science*, 278**:** 1960-1963.

Sundquist, W. I. & Kräusslich, H. G. (2012). HIV-1 assembly, budding, and maturation. *Cold Spring Harbor Perspectives in Medicine*, 2(7): a006924.

Suzuky, Y. & Craigie, R. (2007). The road of chromatin – nuclear entry of retroviruses. *Nature Reviews Microbiology*, 5: 187-196.

Swingler, S., Mann, A., Jacque, J., Brichacek, B., Sasseville, V. G., Williams, K., Lackner, A. A., Janoff, E. N., Wang, R., Fisher, D. & Stevenson, M. (1999). HIV-1 Nef mediates lymphocyte chemotaxis and activation by infected macrophages. *Nature Medicine*, 5: 997-1003.

Tachibana, M., Sugimoto, K., Nozaki, M., Ueda, J., Ohta, T., Ohki, M., Fukuda, M., Takeda, N., Niida, H., Kato, H. & Shinkai, Y. (2002). G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. Genes and Development, 16: 1779 -1791.

Tachibana, M., Ueda, J., Fukuda, M., Takeda, N., Ohta, T., Iwanari, H., Sakihama, T., Kodama, T., Hamakubo, T. & Shinkai, Y. (2005). Histone methyltransferases G9a and GLP form heteromeric complexes and are both crucial for methylation of euchromatin at H3-K9. *Genes and Development*, 19(7): 815-826,

Tachibana, M. (2015). Epigenetic regulation of mammalian sex determination. *Journal of Medical Investigation*, 62(1-2). 19-23.

Tan, W., Dong, Z., Wilkinson, T. A., Barbas, C. F. & 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*ournal of Virology*, 80(4): 1939-1948.

Telesnitsky, A. & Goff, S. P. (1997). Reverse Transcriptase and the Generation of Retroviral DNA. In Coffin, J. M., Hughes, S. H. & Varmus, H. E. (Ed), *Retroviruses* (pp.121-160). Cold Spring Harbor Laboratory Press; Cold Spring Harbor, NY.

Temin, H. M. (1993). Retrovirus variation and reverse transcription: abnormal strand transfers result in retrovirus genetic variation. *Proceedings of the National Academy of Sciences of the United States of America*, 90**:** 6900-6903.

Thomas, M., White, R. L. & Davis, R. W. (1976). Hybridization of RNA to double-stranded DNA: formation of R-loops. Proceedings of the *National Academy of Sciences of the United States of America*, 73(7): 2294-2298.

Toda, T., Hsu, J. Y., Linker, S. B., Hu, L., Schafer, S. T., Mertens, J., Jacinto, F. V., Hetzer, M. & Gage, F. H. (2017). Nup153 Interacts with Sox2 to Enable Bimodal Gene Regulation and Maintenance of Neural Progenitor Cells. *Cell Stem Cell*, 21(5): 618-634.

Towbin, B. D., Gonzalez-Aguilera, C., Sack, R., Gaidatzis, D., Kalck, V., Meister, P., Askjaer, P. & Gasser, S. (2012). Step-wise methylation of histone H3K9 positions heterochromatin at the nuclear periphery. *Cell*, 150: 934-947.

Tripathy, M. K., McManamy, M. E., Burch, B. D., Archin, N. M. & Margolis, D. M. (2015). H3K27 Demethylation at the Proviral Promoter Sensitizes Latent HIV to the Effects of Vorinostat in Ex Vivo Cultures of Resting CD4+ T Cells. *Journal of Virology*, 89(16): 8392-8405.

Tsompana, M. & Buck, M. J. (2014). Chromatin accessibility: a window into the genome. *Epigenetics and Chromatin*, 7(1): 33. doi: 10.1186/1756-8935-7-33.eCollection2014.

Tsukada, Y., Fang, J., Erdjument-Bromage, H., Warren, M. E., Borchers, C. H., Tempst, P. & Zhang, T. (2006). Histone demethylation by a family of JmjC domain-containing proteins. *Nature*, 439(7078): 811–816.

Tsukiyama, T. (2002). The *in vivo* functions of ATP-dependent chromatin-remodelling factors. *Nature Reviews Molecular Cell Biology,* 3**:** 422-429.

Turlure, F., Devroe, E., Silver, P. A. & Engelman, A. (2004). Human cell proteins and human immunodeficiency virus DNA integration. *Frontiers in Bioscience*, 9: 3187-3208.

Turner, A.-M. W. & Margolis, D. M. (2017). Chromatin Regulation and the Histone Code in HIV Latency. *Yale Journal of Biology and Medicine*, 90(2): 229-243.

Ulianov, S. V., Doronin, S. A., Khrameeva, E. E., Kos, P, I, Luzhin, A. V., Starikov, S. S., Galitsyna, A. A., Nenasheva, V. V., Ilyin, A. A., Flyamer, I. M., Mikhaleva, E. A., Logacheva, M. D., Gelfand, M. S., Chertovich, A. V., Gavrilov, A. A:, Razin, S. V. & Shevelyov, Y. Y. (2019). Nuclear lamina integrity is required for proper spatial organization of chromatin in *Drosophila*. *Nature Communications*, 10: 1176.

van Bemmel, J. G., Pagie, L., Braunschweig, U., Brugman, W., Meuleman, W., Kerkhoven, R. M. & van Steensel, B. (2010). The insulator proteinSU(HW) fine-tunes nuclear lamina interactions of the Drosophila genome. *PLoS One*, 5:e15013.

van den Ent, F. M. I., Vos, A., and Plasterk, R. H. A. (1999). Dissecting the role of the N-terminal domain of human immunodeficiency virus integrase by trans-complementation analysis. *Journal of Virology*, 73: 3176-3183.

van Gent, D. C., Vink, C., Groeneger, A. A. & Plasterk, R. H. (1993). Complementation between HIV integrase proteins mutated in different domains. *EMBO JOURNAL*, 12(8): 3216-3267.

Van Holde, K. E. (1999). *Chromatin: Springer Series in Molecular Biology* (New York, Springer-Verlag, 1988).

Van Lint, C., Emiliani, S., Ott, M. & Verdin, E. (1996). Transcriptional activation and chromatin remodeling of the HIV-1 promoter in response to histone acetylation. *EMBO JOURNAL*, 15(5): 1112-1120.

van Steensel, B., Delrow, J. & Henikoff, S. (2001). Chromatin profiling using targeted DNA adenine methyltransferase. *Nature Genetics*, 27: 304-308.

Varadarajan, P., Mahalingam, S., Liu, P., Ng, S. B., Gandotra, S., Dorairajoo, D. S. & Balasundaram, D. (2005). The functionally conserved nucleoporins Nup124p from fission yeast and the human Nup153 mediate nuclear import and activity of the Tf1 retrotransposon and HIV-1 vpr. *Molecular Biology of the Cell*, 16(4): 1823-1838.

Veazey, R. S., DeMaria, M. A., Chalifoux, L. V., Shvetz, D. E., Pauley, D. R., Knight, H. L., Rosenzweig, M., Johnson, R. P., Desrosiers, R. C. & Lackner, A. A. (1998). Gastrointestinal tracts as a major site of $CD4^+$ T cell depletion and viral replication in SIV infection. *Science*, 280: 427-431.

Veazey, R. S, Mansfield, K. G., Than, I. C., Carville, A. C., Shvetz, D. E., Forand, A. E. & Lackner, A. A. (2000). Dynamics of CCR5 expression by CD4(+) T cells in lymphoid tissues during simian immunodeficiency virus infection. *Journal of Virology*, 74: 11001-11007.

Venkatesh, S., Smolle, M., Li, H., Gogol, M. M., Saint, M., Kumar, S., Natarajan, K. & Workman, J.L. (2012). Set2 methylation of histone H3 lysine 36 suppresses histone exchange on transcribed genes. *Nature*, 489: 452-455.

Verdin, E. (1991). DNase I-Hypersensitive Sites Are Associated with Both Long Terminal Repeats and with the Intragenic Enhancer of Integrated Human Immunodeficiency Virus Type 1. *Journal of Virology*, 65(12): 6790-6799.

Verdin, E., Paras ,P., J. & Van Lint, C. (1993) Chromatin disruption in the promoter of human immunodeficiency virus type 1 during transcriptional activation. *EMBO JOURNAL*, 2: 3249-3259.

Vos, S. M., Farnung, L., Boehning, M., Wigge, C., Linden, A., Urlaub, H. & Cramer, P. (2108). Structure of activated transcription complex Pol II-DSIF-PAF-SPT6. *Nature*, 560(7720): 607-612.

Votteler, J. & Schubert, U. (2008) Human Immunodeficiency Viruses: Molecular Biology. *Encyclopedia of Virology*, 3rd ed.: 517-525.

Vranckx, L. S., Demeulemeester, J., Saleh, S., Boll, A., Vansant, G., Schrijvers, R., Weydert, C., Battivelli, E., Verdin, E., Cereseto, A., Christ, F., Gijsbers, R. & Debyser, Z. (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. & Frenkel, L. M. (2014). HIV latency. Proliferation of cells with HIV integrated into cancer genes contributes to persistent infection. *Science*, *345*(6196), 570-573.

Wahba L, Gore, S. K. & Koshland, D. (2013). The homologous recombination machinery modulates the formation of RNA-DNA hybrids and associated chromosome instability. e*life*, 2: e00505.

Wallrath, L. L., Lu, Q., Granok, H. & Elgin, S. C. (1994). Architectural variations of inducible eukaryotic promoters: preset and remodeling chromatin structures. *Bioessays*, 16(3): 165-170.

Wang, T., Balakrishnan, M. & Jonsson, C. B. (1999). Major and minor groove contacts in retroviral integrase-LTR interactions. *Biochemistry*, 38: 3624-3632.

Wang, Y., Wysocka, J., Sayegh, J., Lee, Y. H., Perlin, J. R., Leonelli, L., Sonbuchner, L. S., McDonald, C. H., Cook, R. G., Dou, Y., Roeder, R. G., Clarke, S., Stallcup, S., Allis, C. D. & Coonrod, S. A. (2004). Human PAD4 regulates histone arginine methylation levels via demethylimination. *Science*, 306(5694): 279-283.

Wang, G. P., Ciuffi, A., Leipzig, J., Berry, C. C. & Bushman, F. D. (2007). HIV integration site selection: analysis by massively parallel pyrosequencing reveals association with epigenetic modifications. *Genome Research*, 17: 1186-1194.

Wang, Z., Zang, C., Rosenfeld, J.A., Schones, D.E., Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Peng, W., Zhang, M. Q. & Zhao, K. (2008). Combinatorial patterns of histone acetylations and methylations in the human genome. *Nature Genetics*, 40(7): 897-903.

Wang, Z., Zang, C., Cui, K., Schones, D. E., Barski, A., Peng, W. & Zhao, K. (2009). Genome-wide mapping of HATs and HDACs reveals distinct functions in active and inactive genes. *Cell*, 138(5): 1019-1031.

Wang, I. X., Grunseich, C., Fox, J. Burdick, J., Zhu, Z., Ravazian, N., Hafner, M. & Cheung, W. G. (2018). Human proteins that interact with RNA/DNA hybrids. *Genome Research*, 28: 1-10.

Wahba, L., Constantino, L., Tan, F. J., Zimmer, A. & Koshland, D. (2016). S1-DRIP-seq identifies high expression and polyA tracts as major contributors to R-loop formation. *Genes and Development*, 1; 30(11): 1327-1338.

Watts, J. M., Dang, K. K., Gorelick, R. J. Leonard, C. W., Bess, J. W., Swanstrom, R., Burch, C. L. & Weeks, K. M. (2009). Architecture and Secondary Structure of an Entire HIV-1 RNA Genome. *Nature*, 6: 460(7256): 711-716.

Wen, B., Wu, H., Shinkai, Y., Irizarry, R. A. & Feinberg, A. P. (2009). Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells. *Nature Genetics*, 41: 246-50.

Whyte, W. A., Orlando, D. A., Hnisz, D., Abraham, B. J., Lin, C. Y., Kagey, M. H., Rahl, P. B., Lee, T. I. & Young, R. A. (2013). Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes. *Cell*, 153(2): 307-319.

Williams, S. A., Chen, L. F., Kwon, H., Ruiz-Jarabo, C. M., Verdin, E. & Greene, W. C. (2006). NF-kappaB p50 promotes HIV latency through HDAC recruitment and repression of transcriptional initiation. *EMBO JOURNAL*, 25(1): 139-149.

Williams, S. A. & Greene, W. C. (2007). Regulation of HIV-1 latency by T-cell activation. *Cytokine*, 39(1): 63-74.

Wolffe, A. P. (1999). *Chromatin: Structure and Function*, 3rd ed. San Diego, CA, San Diego Academic.

Wong, R. W., Mamede, J. I. & Hope, T. J. (2015). Impact of Nucleoporin-Mediated Chromatin Localization and Nuclear Architecture on HIV Integration Site Selection. *Journal* of *Virology*, 89: 9702-9705.

Wu, C., Wong, Y. C., Elgin, S. C. (1979). The chromatin structure of specific genes: II. Disruption of chromatin structure during gene activity. *Cell*, 16(4): 807-814.

Wu, C. (1980). The 5′ ends of Drosophila heat shock genes in chromatin are hypersensitive to DNase I. *Nature*, 286(5776): 854–860.

Wyrick, J. J., Holstege, F. C., Jennings, E. G., Causton, H. C., Shore, D., Grunstein, M., Lander, E. S. & Young, R. A. (1999). Chromosomal landscape of nucleosome-dependent gene expression and silencing in yeast. *Nature*, 402: 418–421.

Xhemalce, B., Dawson, M. A. & Bannister, A. J. (2011). Histone modifications. In Meyers R. (Ed). *Encyclopedia of Molecular Cell Biology and Molecular Medicine.* Retrieved from: https://www.onlinelibrary.wiley.com/.

Xu, W., Xu, H., Li, K, Fan, Y., Liu, Y., Yang, X. & Sun, Q. (2017). The R-loop is a common chromatin feature of the Arabidopsis genome. *Nature Plants*, 3(9): 704-714.

Yamashita, M. & Emerman, M. (2004). Capsid is a dominant determinant of retrovirus infectivity in nondividing cells. *Journal of Virology*, 78: 5670-5678.

Yamashita, M. & Engelman, A., N. (2017). Capsid-Dependent Host Factors in HIV-1 Infection. *Trends in Microbiology*, 25(9): 741-755.

Yan, J., Shun, M.-C., Zhang, Y., Hao, C. & Skowronski, J. (2019). HIV-1 Vpr counteracts HLTF-mediated restriction of HIV-1 infection in T cells. *Proceedings of the National Academy of Sciences of the United States of America*, 116 (19): 9568-9577.

Yang W., Hendrickson, W. A., Crouch, R. J. & Satow, Y. (1990). Structure of ribonuclease H phased at 2 A resolution by MAD analysis of the selenomethionyl protein. *Science*, 249(4975): 1398-1405.

Yokochia, T., Poducha, K., Rybaa, T., Lua, J., Hiratania, I., Tachibana, M., Shinkaib, Y. & Gilberta, D., M. (2009). G9a selectively represses a class of late-replicating genes at the nuclear periphery. *Proceedings of the National Academy of Sciences of the United States of America*, 106(46): 19363-19368.

Yu, K., Chedin, F., Hsieh, C. L., Wilson, T. E., Lieber, M. R. (2003). R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. *Nature Immunology*, 4: 442-451.

Yan, N., Cherepanov, P., Daigle, J. E., Engelman, A. & Lieberman, J. (2009). The SET Complex Acts as a Barrier to Autointegration of HIV-1. *Plos Pathogens*, 5(3): e1000327.

Zeller, P., Padeken, J., van Schendel, R., Kalk, V, Tijsterman, M. & Gasser, S. M. (2016). Histone H3K9 methylation is dispensable for Caenorhabditis elegans development but suppresses RNA:DNA hybrid-associated repeat instability. *Nature Genetics*, 48(11): 1385-1395.

Zheng, R., Jenkins, T. M. & Craigie, R. (1996). Zinc folds the N-terminal domain of HIV-1 integrase, promotes multimerization, and enhances catalytic activity. *Proceedings of the National Academy of Sciences of the United States of America*, 93(24): 13659-13664.

Zhang, Y., McCord, R. P., Ho, Y. L., Lajoie, B. R., Hildebrand, D. G., Simon, A. C., Becker, M. S., Alt, F. W. & Dekker, J. (2012). Spatial organization of the mouse genome and its role in recurrent chromosomal translocations. *Cell*, 148: 908-921.

Zheng, X., Hu, J., Yue, S., Kristiani, L., Kim, M., Sauria, M., Taylor, J., Kim, Y. & Zheng, Y. (2018). Lamins Organize the Global Three-Dimesnional Genome from the Nuclear Periphery. *Molecular Cell*, 71(5): 802-815.

Zolotukhin, A. S. & Felber, B. K. (1999). Nucleoporins nup98 and nup214 participate in nuclear export of human immunodeficiency virus type 1 rev. *Journal of Virology*, 73(1): 120-127.

UNAIDS (2015). AIDS by the numbers [Fact sheet]. Retrieved from: http://www.unaids.com/

UNAIDS (2018). World AIDS day 2018 [Fact sheet]. Retrieved from: http://www.unaids.com/

VI. Appendix

VI.I. List of Abbreviations

Abbreviation	Elaboration
°C	degree Celsius
Δ	Delta
µg	microgram
µL	microliter
µm	micrometer
µM	micromolar
3 D	Tridimensional
A	Alanine
ACTN1	Alpha actinin
ADP	Adenosindiphosphat
AIDS	Acquired Immune Deficiency Syndrome
Airn	Antisense Igf2r RNA
Alu	*Arthrobacter luteus* (repetitive genomic sequence)
ApaLI	*Acetobacter pateurianus LI* (enzyme)
APC	Antigen presenting cell
APOE	Apolipoprotein E
AT	Adenin Thymine
B cell	Bursa Fabricius lymphocyte
BAC	Bacterial artificial chromosome
BACH2	Basic leucine zipper transcription factor 2
BAF	Barier to autoregulation protein
bp	Base pair
BTBD19	BTB Domain Containing 19
C2CD3	C2 Calcium Dependent Domain Containing 3
Ca	Calcium

Abbreviation	Elaboration	*Continuation*
CA	Capsid protein	
CaCl2	Calcium chloride	
cART	Combined antiretroviral therapy	
CCD	Catalytic core domain	
CCR5	C-C Motif Chemokine Receptor 5	
CD3	Cluster of Differentiation 3	
CD4	Cluster of Differentiation 4	
CD28	Cluster of Differentiation 28	
CD38	Cluster of Differentiation 38	
CD69	Cluster of Differentiation 69	
cDNA	Complementary deoxyribonucleic acid	
C. elegans	*Caenorhabditis elegans*	
ChIP	Chromatin immunoprecipitation	
CNTN4	Contactin 4	
CPSF6	Cleavage and polyadenilation specificity factor 6	
CREBBP	CREB Binding Protein	
Ct	Cycle threshold	
CTCF	CCCTC-binding factor	
CTD	C-terminal domain	
CXCR4	C-X-C Motif Chemokine Receptor 4	
Cyp A	Cyclophilin A	
D	Aspartic acid	
DamID	Deoxyribonucleic acid adenine methyltransferase identification	
DDR	Deoxyribonucleic acid damage response	
Dig	Digoxigenin	
DMEM	Dulbecco’s modified eagle medium	
DNA	Deoxyribonucleic acid	
DNase	Deoxyribonuclease	
DNMT1	DNA Methyltransferase 1	
dNTP	Deoxyribonucleotide	
DOC	Deoxycholic acid	

Abbreviation	Elaboration	*Continuation*
DRIP	DNA-RNA immunoprecipitation	
dsRNA	Double stranded ribonucleic acid	
DTT	DL-Dithiothreitol	
E	Glutamic acid	
ECL	Enhanced chemiluminescence	
eGFP	Enhanced green fluorescent protein	
EGR1	Early Growth Response 1	
EGTA	Ethylene glycol-bis(ßaminoethyl ether) tetraacetic acid	
EMD	Emerin	
env	Envelope	
ESC	Embryonic stem cell	
ESCRT	Endosomal sorting complex required for transport	
F	Forward	
FA	Formaldehyde	
FACS	Fluorescence activated cell sorting	
FACT	Facilitates transcription complex	
FBS	Fetal bovine serum	
FISH	Fluorescent in situ hybridization	
FG	Phenylalanine-Glycine	
FSI	Full site integration	
Fwd	Forward	
G0	Growth 0 phase	
G1	Growth 1 phase	
G2	Growth 2 phase	
GO	Gene ontology	
GC	Guanine Cytosine	
GLP	G9a-like protein	
GNB1	Guanine Nucleotide-Binding Protein G(I)/G(S)/G(T) Subunit Beta-1	
GRB2	Growth Factor Receptor Bound Protein 2	
h	hours	
H1	Histone 1	

Abbreviation	Elaboration	*Continuation*
H2A	Histone 2 A	
H2A(X)	Phosphorylated histone 2	
H2B	Histone 2 B	
H2O	Water	
H3	Histone 3	
H3K4me3	Histone 3 Lysine 4 trimethylation	
H3K9ac	Histone 3 Lysine 9 acetylation	
H3K9me2	Histone 3 Lysine 9 dimethylation	
H3K9me3	Histone 3 Lysine 9 trimethylation	
H3K14ac	Histone 3 Lysine 14 acetylation	
H3K27ac	Histone 3 Lysine 27 acetylation	
H3K27me3	Histone 3 Lysine 27 trimethylation	
H3K36me3	Histone 3 Lysine 36 trimethylation	
H3S10P	Histone 3 serine 10 phosphorylation	
H4	Histone 4	
H4K5ac	Histone 4 Lysine 5 acetylation	
H4K8ac	Histone 4 Lysine 8 acetylation	
HAT	Histone acetyl transferase	
HBD	Hybrid binding domain	
HBS	HEPES buffered saline	
HDAC	Histone deacetylase	
HDAC1	Histone deacetylase 1	
HDM	Histone demethylase	
HEK293T	Human embryonic kidney 293T cells	
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid	
HIV-1	Human immunodeficiency virus type 1	
HIV-1	Human immunodeficiency virus type 2	
HMT	Histone methyltransferase	
HP1	Heterochromatin protein 1	
HpaI	*Haemophilus parainfluenza* I (enzyme)	
HSI	Half site integration	

Abbreviation	Elaboration	*Continuation*
IBD	Integrase binding domain	
IF	Immunofluorescence	
IN	Integrase protein	
IN.eGFP	Vpr.Integrase.enhanced Green Fluorescent Protein-fusion	
K	Lysine	
kb	kilobase	
KCl	Kalium chloride	
kDa	kilodalton	
KDM2B	Lysine Demethylase 2B protein	
KO	Knockout	
L	Liter	
LAD	Lamina associated domain	
LAM PCR	Linear amplification mediated polymerase chain reaction	
LAP2B	Lamina associated polypeptide 2	
LBR	Lamin B receptor	
LC	Linker cassette	
LEDGF	Lens epithelium derived growth factor	
LiCl	Lithium chloride	
LOCK	Large organized chromatin domain	
LRA	Latency reversing agent	
LTR	Long terminal repeat	
Lys	Lysine	
m	mili (10^{-3})	
M	Mitosis	
mA	miliamper	
Ma	Matrix protein	
Mb	Megabase	
MDM	Monocyte derived macrophage	
Mg	Manesium	
MgCl	Magnesium chloride	
MHC	Major histocompatibility complex	

Abbreviation	Elaboration	*Continuation*
min	Minute	
MluCI	*Micrococcus luteus* CI (enzyme)	
MLV	Murine leukemia virus	
MKL2	Myocardin like 2	
mM	milimolar	
MNase	Micrococcal nuclease	
mRNA	Messenger ribonucleic acid	
MseI	*Micrococcus* species (R. Morgan) (enzyme)	
MTT assay	3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromidefor assay	
MYADM	Myeloid Associated Differentiation Marker	
N	Asparagine	
NaCl	Sodium chloride	
NC	Nuclear core protein	
NE	Nuclear envelope	
nef	Negative factor	
NFATC3	Nuclear Factor Of Activated T-Cells 3	
ng	nanogram	
NL	Nuclear lamina	
NLS	Nuclear localization signal	
nm	nanometer	
NPC	Nuclear pore complex	
NPLOC4	NPL4 Homolog, Ubiquitin Recognition Factor	
NTD	N-terminal domain	
NUP	Nuclear pore protein	
ON	Overnight	
p	piko (10^{-12})	
PACS2	Phosphofurin Acidic Cluster Sorting Protein 2	
PBMC	Purified Blood Mononuclear Cell	
PBS	Phosphate-buffered saline	
pbs	Primer binding site	

Abbreviation	Elaboration	*Continuation*
PCR	Polymerase chain reaction	
PIC	Preintegration complex	
PEI	Polyethylenimine	
Pen/Strep	Penicillin/Streptomycine	
PFA	Paraformaldehyde	
PMSF	Phenylmethylsulfonyl fluoride	
pol	Polymerase	
polyA	Polyadenylated	
ppt	Polypurine tract	
PR	Protease protein	
PRC2	Polycomb repressive complex 2	
PWWP	Proline-Tryptophan-Tryptophan-Proline	
Q	Glutamine	
qPCR	Quantitative polymerase chain reaction	
R	Reverse	
RANBP2	Ran binding protein 2	
Rev	Reverse	
RGPD8	RANBP2-like and GRIP domain containing 8	
RIG	Recurrent integration gene	
RIPA	Radioimmunoprecipitation assay buffer	
RNA	Ribonucleic acid	
RNAi	Ribonucleic acid interferin	
RNA Pol2	Ribonucleic acid polymerase 2	
RNase	Ribonuclease	
RNF157	Ring Finger Protein 157	
RNHL	Ribonuclease H-like	
rNTP	Ribonucleotide	
RPL13A	Ribosomal Protein L13a	
RPMI	Roswell Park Memorial Institute medium	
RPTOR	Regulatory Associated Protein Of MTOR Complex 1	

Abbreviation	Elaboration	*Continuation*
RT	Reverse Transcriptase (enzyme), Reverse Transcription (process) or Room Temperature	
RT-PCR	Reverse transcriptase polymerase chain reaction	
s	Second	
SA-HRP	Streptavidin-horse radish peroxidase	
SAMHD1	SAM domain and HD domain-containing protein	
SDS	Sodium dodecyl sulfate	
SETX	Senataxin	
SG-PERT	SYBR Green PCR-enhanced reverse transcriptase assay	
siRNA	Small interfering ribonucleic acid	
SIV	Simian immunodeficiency virus	
SMG1	Serine/threonine-protein kinase 1	
SNRPN	mall nuclear ribonucleoprotein polypeptide N	
SSC	Saline-sodium citrate	
STAT5B	Signal transducer and activator of transcription 5B	
STED	Stimulated emission depletion (microscopy)	
TAD	Topologically associated domain	
T cell	Thymus derived lymphocyte	
tat	Transactivator of transcription	
TF	Transcription factor	
TGF-β1	Transforming Growth Factor Beta 1	
Th	T-helper cell	
Top I	Topoisomerase 1	
TRC	T cell receptor	
Treg	T-regulatory cells	
Tris	Tris Buffered Saline	
tRNA	Transport ribonucleic acid	
TSS	Transcription start site	
U	Unit	
V	Valine (aminoacid) or volt (electrical unit of voltage)	
vDNA	Viral Deoxyribonucleic acid	
vif	Viral infectivity factor	

Abbreviation	Elaboration	*Continuation*
vpr	Viral protein R	
vpu	Viral protein U	
VSV-G	Vesicular stomatitis virus G protein	
WKKD	Tryptophan-Lysine-Lysine-Aspartic acid	
wt	Wilde type	
YY1	Ying yang 1 protein	
Zn	Zink	
ZnCl	Zink chloride	

VI.II. List of Figures and Tables

Section **Page**

I. Introduction

II. Results

Section **Page**

Section **Page**

VI.III. List of Reagents and Material

Cells and culture medium

Cell type	Medium	Medium composition	
Primary $CD4^+$ T cells	Supplemented RRPI	RPMI 10% FBS 1% Pen/Strep	*For $CD4^+$ T activation: CD3/CD28 Dynabeads Human T-Activator CD3/CD28 (ThermoFisher) and 5 ng/mL IL-2
MDMs			
Jurkat T cell line			
HEK293T cell line	Supplemented DMEM	DMEM 10% FBS 1% Pen/Strep	
HeLa cell line			

Drug compounds

Drug name	Supplier
SAHA	Sigma-Aldrich
JQ1	Sigma-Aldrich
BIX01294	Sigma-Aldrich
UNC0642	Sigma-Aldrich

Buffers and solutions

Buffer/solution	Composition
RIPA	150mM NaCl 20mM Tris-HCl pH 7.4 1mM EDTA 0.5% NP-40 0.5% DOC 0.1%SDS
	25 mM Tris

Buffer/solution	Composition
1X SDS running buffer	192 mM Glycine 0,1% SDS, pH 8.3
2X HBS	50 mM HEPES 280 mM NaCl 1.5 mM Na2HPO4, pH 7.5
SG-PERT cell lysis buffer	0.25% Triton X-100 50 mM KCl 100 mM TrisHCl pH7.4 40% glycerol
10X SG-PERT dilution buffer	50 mM (NH4)2SO4 200 mM KCl 200 mM Tris–HCl, pH 8.3
2X SG-PERT reaction mix	10 mM (NH4)2SO4 40 mM KCl 40 mM Tris–Cl pH 0.3 10 mM MgCl2 0.2 mg/ml BSA 1/10,000 SYBR Green I 400 µM dNTPs 1 µM forward primer 1 µM reverse primer 1.2 µg/ml BMV RNA
3M LiCl buffer	10 mM Tris-HCl, pH 7.5 1 mM EDTA 3M LiCl
6M LiCl buffer	10 mM Tris-HCl, pH 7.5 1 mM EDTA 6 M LiCl
IF blocking buffer	4% PFA in 1X PBS
Cell permeabilization buffer	0,5% Triton 0,1% Tween in 1X PBS
IF wash buffer	0,1% Tween in 1X PBS
FISH permeabilization buffer	0.5% Triton 0.5% Saponin in 1X PBS
20X SSC buffer	3M NaCl 300mM Na_3Citrate x 2 H_2O pH 7
FISH hybridization buffer	50% Formamide 2X SSC buffer 10% Dextran sulfate pH 7

Buffer/solution	**Composition**
NP-40 buffer	0.5% NP-40 10 mM Tris-Cl pH7.4 10 mM NaCl2 3 mM MgCl 1 mM PMSF and protease inhibitors
Nuclear lysis buffer	1% SDS 10 mM EDTA 50 mM Tris-HCl, pH 8
5X ChIP buffer	50 mM Tris-HCl, pH 7.4 150 mM NaCl 1 % NP40 0.25 % Na-DOC 1 mM EDTA, pH 8 and protease inhibitors
ChIP binding buffer	1X PBS 0,1% BSA 2 mM EDTA, pH7.4
LiCl buffer	10 mM Tris-HCl, pH8 250 mM LiCl 1% IGEPAL 1% NaDoc 1 mM EDTA
TE buffer	10 mM Tris-HCl, pH8 50 mM NaCl 1mM EDTA
DRIP binding buffer	100 mM NaPO4, pH 7 1,4 M NaCl 0,5% Triton X100
DRIP elution buffer	50 mM Tris, pH 8 10 mM EDTA 0,5% SDS
In vitro integration assay buffer	20 mM HEPES pH 7 15% DMSO 8% PEG 10 mM MgCl2 20 µM ZnCl2 100 mM NaCl 10 mM DTT
Mowiol mounting medium	77 µM Mowiol 4-88 30% Glycerol 10mM Tris-HCl, pH 8.5

Chemicals and reagents

Chemical/reagent	Supplier
Amersham ECL Prime Western Blotting Detection Reagent	GE Healthcare Life Sciences
AMPure XP magnetic beads	Beckman Coulter
Cot-1 DNA	Roche
Fixable Viability Dye eFluor 450	eBioscience
Hering sperm DNA	Sigma-Aldrich
Immunoglobulin-γ and agarose beads	Sigma-Aldrich
MagnaChIP Protein A/G Magnetic Beads	Millipore
PBS	Sigma-Aldrich
PEI	Sigma-Aldrich
Propidium Iodide	Roche
Puromycin	InvivoGen

Enzyme	Supplier
ApaLi	NEB
DNA Polymerase	Invitrogen
HpaI	NEB
MluCI	NEB
M-MLV Reverse transcriptase	Invitrogen
MseI	NEB
Proteinase K	Invitrogen
PureLink RNase A	Invitrogen
RNase H1	NEB
T3 RNA Polymerase	Promega

Antibody/ target protein	Species	Supplier
β-actin	mouse	Sigma-Aldrich
dimethyl-Histone H3 (Lys 9)	mouse	Millipore
G9a/EMT2	rabbit	Cell Signaling Technology
Histone H3	rabbit	Millipore
Lamin B1	rabbit	Abcam
LEDGF/p75	rabbit	Bethyl Laboratories
mAb 414	mouse	Abcam
p24	-	Beckman Coulter
S9.6	mouse	Kerafast
trimethyl-Histone H3 (Lys 9)	rabbit	Abcan
trimethyl-Histone H3 (Lys 36)	rabbit	Abcam
gamma H2A.X (phospho S139)	mouse	Abcam

Secondary antibodies	Species	Supplier
anti-digoxigenin	-	Roche
Alexa 488	mouse	ThermoFisher
Alexa 568	mouse	ThermoFisher
Alexa 647	rabbit	ThermoFisher
ATTO 594	mouse	Sigma-Aldrich
ATTO 647N	rabbit	Sigma-Aldrich
Western blot anti-mouse	mouse	Jackson immunoresearch
Western blot anti-rabbit	rabbit	Jackson immunoresearch

Plasmid constructs

Plasmid name	Description
pFC53	DNA template for R-loop formation *in vitro*
pGZIP-G9a	Lentiviral vector for G9a KO
pMD2.G	Expression vector for viral VSV-G presudotyping
pMMM-FLST	Expression vector for RNase H1 overexpression
pMMM-FLST, RNase H1 D210N	Expression vector for RNase H1 D210N mutant overexpression
pMMM-FLST, RNase H1 WKKD	Expression vector for RNase H1 WKKD mutant overexpression
pNL4.3	Full-length proviral HIV-1 group M sequence containing plasmid for virus particle production
pNL4-3/E^-R^-	Full-length proviral HIV-1 group M sequence with a frameshift mutation in *env* gene containing plasmid for virus particle production
HIV-1BRU(K264,266,273R)	Full-length proviral HIV-1 group M sequence with K264, K266 and R273 mutations in IN C-terminus for virus particle production
vpr.IN.eGFP	Expression vector for fluorescent virus particle labeling

Commercial kits

Kit name	Supplier
10X hexanucleotide mixture	Roche
Amaxa Cell Line Nucleofector Kit V	Lonza
Bradford assy	BioRad
CellTiter 96® Non-Radioactive Cell Proliferation Assay (MTT) assay	Promega
DIG-Nick Translation kit	Roche
DNeasy Blood & Tissue kit	Qiagen
Fast Link DNA ligation kit	Biozym

Kit name	Supplier
Invitrap spin universal RNA mini kit	Stratec biomedical
Lipofectamine LTX with Plus Reagent kit	ThermoFisher
Nick Translation Kit	Roche
NucleoBond Xtra Maxi kit	Macherey-Nagel
Plasmid Midi Kit	Qiagen
PromoFectin kit	PromoKine
QIAquick PCR purification kit	Qiagen
RosetteSep™ Human $CD4^+$ T Cell Enrichment Cocktail	STEMCELL TECHNOLOGIES
TSA Plus system	Perkin Elmer

Devices and instruments

Device/instrument	Supplier
BD FACSVerse flow cytometer	BD Biosciences
BeadArray microarray Technology / HumanHT-12 Expression BeadChip kit.	Illumina
Bioruptor (sonication device)	diagnose
ddPCR automated droplet generator QX200	BioRad
ddPCR droplet reader QX200	BioRad
DNA gel electrophoresis device Biometra Compact	Biometra
Infinite 200 PRO multimode plate reader.	Tecan
Leica TCS SP8 Laser Confocal Leica Microsystems Scanning Microscope	Leica
MiSeq platform	Illumina
Olympus IX81 microscope	Olympus

Device/instrument	Supplier
P-class P 300 NanoPhotometer	IMPLEN
PCR Biorad CFX96 Touch platform	BioRad
SimpliAmp Thermal Cycler	ThermoFisher
STED system containing an easy 3D optics module	Abberior Instruments
Trans –Blot device for semi-dry transfer	BioRad
Transiluminator (GEL-V-U-04-07)	Zenith
Ultracentrifuge L8-M Beckman	Beckman Coulter
Western blot gel running system	Biometra
Western blot chemiluminescence imaging device Alliance Q9 Touch	Mandel

Analytical softwares

Software name	Application	Supplier
Adobe Illustrator	Data graphs and image arranging	Adobe
Adobe Photoshop	Image analysis	Adobe
BD FACSuite software	FACS data analysis	BD Biosciences
CFX Manager	qPCR data analysis	BioRad
GraphPad Prism	Data organization and statistic analysis	GraphPad Software Inc.
ImageJ Fiji	Image analysis	Open source
Imspector software	STED image data deconvolution	Abberior Instruments
R software	*In silico* data analysis; microarray data analysis	R Core Team
QuantaSoft Software	ddPCR droplet analysis	BioRad
Volocity imaging software	Image analysis; FISH distance measurements	Perkin Elmer

Primer and primer probe sequences

Primer name	Primer sequence (5'-3' direction)	Application
SG-PERT F:	TAGTTGTTGGGCTTCGCTTT	SG-PERT assay
SG-PERT R:	TTGTCGGCTTTACCTGCTTT3	
LM667	ATGCCACGTAAGCGAAACTCTGGCTAACTAGGGAACCCACTG	Alu PCR
LR	TCCACACTGACTAAAAGGGCTTGA	
ZXF-P probe	TGTGACTCTGGTAACTAGAGATCCCTCAGACCC	
Alu1	TCCCAGCTACTGGGGAGGCTGAGG	
λ-primer	ATGCCACGTAAGCGAAACT	
B13 Fwd primer	CCCCAGGGAGTAGGTTGTGA	
B13 Rev primer	TGTTATTTGAGAAAAAGCCCAA	
B13 probe	CAGCAGGAAAGGAC	
2 LTR circles Fwd primer	AACTAGAGATCCCTCAGACCCTTTT	2LT circle quantification
2 LTR circles Rev primer	CTTGTCTTCGTTGGGAGTGAATT	
2 LTR circles probe	CTAGAGATTTTCCACACTGAC	
Total viral DNA Fwd primer	TGTGTGCCCGTCTGTTGTGT	Total viral DNA quantification
Total viral DNA Rev primer	GAGTCCTGCGTCGAGAGAGC	
Total viral DNA probe	CGCCCGAACAGGGACTTGAA	
LTR I	GAGCTCTCTGGCTAACTAGG	LAM-PCR primer
LTR II	AGCTTGCCTTGAGTGCTTCA	
LTR III	AGTAGTGTGTGCCCGTCTGT	

Primer name	Primer sequence (5'-3' direction)	Application
LC I	GACCCGGGAGATCTGAATTC	
LC II	AGTGGCACAGCAGTTAGG	
U1a Fwd primer	ACATCAAGCAGCCATGCAAAA	Quantification of late HIV-1 mRNA products
U1a Rev primer	CAGAATGGGATAGATTGCATCCA	
U1a probe	AAGAGACCATCAATGAGGAA	
Nuc1b Fwd primer	CGTCTGTTGTGTGACTCTGGTAACT	Quantification of early HIV-1 mRNA products
Nuc1b DNA Rev primer	CACTGCTAGACATTTTCCACACTGA	
Nuc1b probe	ATCCCTCAGACCCTTT	
BACH2 First intron F	GAAGAGTGCCTAGCACAGTAAG	ChIP-qPCR primer
BACH2 First intron R	GAGCACACAGGTTTAGGAGATG	
BACH2 Gene body F	CAACTCCAAGGGCCAGAG	
BACH2 Gene body R	AGAGGCTGAACCAACAAGAG	
BACH2 Gene end F	TCTGCCTCTGATCCCTATGT	
BACH2 Gene end R	GGTAGCAAGTAAGGGTCCAAA	
STAT5B First intron F	TCAAACGTGGTGGAAGTGTTAG	
STAT5B First intron R	CACCCAGCCTACCATCTTATTT	
STAT5B Gene body F	GGTGATGGTCATGGTTTCCT	
STAT5B Gene body R	GAGACACCTGCTTCTGCTG	
STAT5B Gene end F	CTTCCTCTGAAGGGATGGTATTG	
STAT5B Gene end R	TCACGCTGATGACTTGAAACT	
MKL2 First intron F	GACCCTTTGCGTACCTTCTT	
MKL2 First intron R	GTGGAGCTCAGAGTCTACATTTC	
MKL2 Gene body F	TGAAACACAAGATTCGGATCG	

Primer name	Primer sequence (5'-3' direction)	Application
MKL2 Gene body R	GGGAATAAACAGAAATGGTGGAAAT	
MKL2 Gene end F	AGAGGAAAGACCACTCCCTAT	
MKL2 Gene end R	GGCTTTCTGTTGTGAGTTGTATG	
CNTN4 First intron F	CTTACACACCCTTCACCACTAC	
CNTN4 First intron R	CGCAAGACTCCCACAATAGAA	
CNTN4 Gene body F	CAGGAGTGATGGGTGAATATGAG	
CNTN4 Gene body R	TAAAGCAAAGCATTCCAGCTTC	
CNTN4 Gene end F	TACTTGTCTTTCGGGAGCATTT	
CNTN4 Gene end R	GCTCTGGGTCATCTAGATTTGG	
RPTOR First intron F	TGCTGGCAGATTGTGGTAAA	
RPTOR First intron R	GGGAAGGAAGGAAGGAAGAAAG	
NPLOC4 First intron F	CCAAGGCCAACTCCAGTATT	
NPLOC4 First intron R	GGTCTGTCTCTTCCTGTGTATTT	
KDM2B First intron F	GTTGCTTGGGTTGCAATGG	
KDM2B First intron R	AGAGAGCCCGGGACATTAT	
PACS2 First intron F	CCCAAATCTCATAGCCAGTAGG	
PACS2 First intron R	GTGAGGGAGCACAGGAAATA	
ACTN1 First intron F	AGGTAGAGAGGAACCAGAAA	
ACTN1 First intron R	AATGCCAGGGAAAGGTAAGG	
APOE F	CCGGTGAGAAGCGCAGTCGG	DRIP-qPCR primer
APOE R	CCCAAGCCCGACCCCGAGTA	
MYADM positive F	CGTAGGTGCCCTAGTTGGAG	
MYADM positive R	TCCATTCTCATTCCCAAACC	

Primer name	Primer sequence (5'-3' direction)	Application
RPL13A F	AATGTGGCATTTCCTTCTCG	
RPL13A R	CCAATTCGGCCAAGACTCTA	
EGR1 F	CCAATTCGGCCAAGACTCTA	
EGR1 R	CTTGTGGTGAGGGGTCACTT	
BTBD19 F	GGCTGCTCAGGAGAGCTAGA	
BTBD19 R	ACCAGACTGTGACCCCAAAG	
SNRPN F	GCCAAATGAGTGAGGATGGT	
SNRPN R	TCCTCTCTGCCTGACTCCAT	
EGR1 negative F	GAACGTTCAGCCTCGTTCTC	
EGR1 negative R	GGAAGGTGGAAGGAAACACA	
MYADM negative F	TGCATCTACATCCGCAAAAG	
MYADM negative R	AGAGTGGACGCTGCAGAAAT	
R-loop fragment F	TTTAGAGCTTGACGGGGAAA	
R-loop fragment R	CAACAGTTGCGTAGCCTGAA	
Non-R-loop fragment F	TTGCCGGGAAGCTAGAGTAA	
Non-R-loop fragment R	GCTGCCATAAGCATGAGTGA	
BACH2 F	GAAGAGTGCCTAGCACAGTAAG	
BACH2 R	GAGCACACAGGTTTAGGAGATG	
STAT5B F	TCAAACGTGGTGGAAGTGTTAG	
STAT5B R	CACCCAGCCTACCATCTTATTT	
Early RT products F	TTAAGCCTCAATAAAGCTTGCC	Quantification of early RT products
Early RT products R	GTTCGGGCGCCACTGCTAG	
Early RT products probe	CCAGAGTCACACAACAGACGGGCA	

Primer name	Primer sequence (5'-3' direction)	Application
Late RT products F	CATGTTTTCAGCATTATCAGAAGGA	Quantification of late RT products
Late RT products R	TGCTTGATGTCCCCCCACT	
Late RT products probe	CCACCCCACAAGATTTAAACACCATGCTAA	
2LTR circles ddPCR F	CTAACTAGGGAACCCACTGCT	Quantification of 2LTR circles
2LTR circles ddPCR R	GTAGTTCTGCCAATCAGGGAA	
2 LTR circles ddPCR probe	AGCCTCAATAAAGCTTGCCTTGAGTGC	
RPP30 F	GATTTGGACCTGCGAGCG	Normalizer gene for absolute quantification of early and late RT products and 2LTR circles
RPP30 R	GCGGCTGTCTCCACAAGT	
RPP30 probe	CTGACCTGAAGGCTCT	

Gene expression assays

Target gene	Expression assay code (Applied Biosystems)
BACH2	Hs00222364_m1
STAT5B	Hs0273500_m1
MKL2	Hs00539161_s1
Tubulin 1a	Hs03045184_g1

BAC constructs

Gene name	BAC reference code
BACH2	RP11-59717
NPLOC4	RP11-765O14
MKL2	RP11-1072B15
STAT5B	CTD-3124P7
DNMT1	CTD-2240E14
RPTOR	RP11-28G8
SMG1	RP11-1035H13
GRB2	RP11-16C1
GNB1	RP11-798H13
NFATC3	RP11-67A1
KDM2B	RP11-44F24
ACTN1	RP11-226F19
PACS2	RP11-521B24
CNTN4	RP11-63O1
RNF157	RP11-449J21