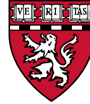
HOST FACOTRS IN HIV-1 INTEGRATION TARGETING

Dana-Farber
Cancer Institute



HARVARD
MEDICAL SCHOOL

University of
Applied Sciences

by

Nika Vujosevic

Dana-Farber Cancer Institute

Department: Cancer Immunology and Virology

IMC University of Applied Sciences

Bachelor program

Medical and Pharmaceutical Biotechnology

March 2024

Table of Contents

1 Description of the Problem & Scientific Background ... 3
1.1 Human Immunodeficiency Virus ... 3
1.2. Structure of HIV-1 Virion ... 4
1.3 HIV-1 entry into nucleus ... 4
1.4 Integration and its mechanisms ... 5
1.5 LEDGF/p75 overview ... 8
1.6 LEDGF/Integrase interaction ... 8
1.7 LEDGF interactions with IWS1, MLL1, PAF1 host factors ... 9
1.8 SPADs/LADs ... 15
2 Description of Aims and the Research Question ... 16
3 Presentation of the methods ... 17
Cell culture ... 17
CRISPR-Cas9 Transient Knockdown ... 17
Lentiviral Transductions ... 18
Cell lysis ... 19
SDS-PAGE and Western Blot ... 19
Blocking and antibody incubation ... 20
Transduction of cells with Edit-R All-in-one Lentiviral sgRNA particles - mechanism ... 21
4 Other key methods in overall project ... 22
p24 assay ... 22
Luciferase Assay ... 23
Sequencing – background and Next-generation sequencing ... 23
5 Workflow of the research ... 25
CRISPR mechanism and relevance ... 26
Western Blot mechanism and relevance ... 27
5 Results ... 28
6 Discussion ... 29
7 References ... 31

1 Description of the Problem & Scientific Background

1.1 Human Immunodeficiency Virus

HIV, or Human Immunodeficiency Virus, belongs to the family Retroviridae and is grouped to the Lentivirus genus. The primary classification of HIV is into HIV-1 and HIV-2 type, based on molecular characteristics and differences in viral antigens. It has been estimated, through epidemiology studies that, HIV first appeared in between the years of 1920 and 1940.[1] Approximately 39 million people are living with HIV as of 2022 and around 630,000 people dying from causes related to HIV and AIDS.[2] HIV-1 has been identified as the main cause of the acquired immunodeficiency syndrome (AIDS). The most frequent and common way of becoming HIV-1 positive is sexual transmission at the genital mucosa or needle sharing. Transmitted viruses are typically macrophage-tropic and do not possess the ability to induce multinucleated syncytia in cell culture.[3]
Before the development of antiretroviral therapy (ART), AIDS was undoubtably fatal for anyone contracting it. These antiretroviral drugs mostly targeted enzymes such as integrase, reverse transcriptase, and protease in the process of HIV-1 replication.[4]

ART has been used for more than 20 years as treatment for HIV infections. It is a very effective type of therapy, especially when used appropriately resulting in almost complete or complete suppression of the HIV replication. This allows for the improvement of the immune function of an infected individual and significantly decreases the risk of developing AIDS. It is important however that once ART is started it should not be stopped as the virus would undoubtably come back within a short time.[5]

1.2. Structure of HIV-1 Virion

Lentiviruses contains of a protein core withing which there is a diploid, single-stranded genome. The core of HIV-1 contains four nucleocapsid proteins, p24 (icosahedral capsid protein – CA), p17 (myristoylated matrix protein – MA), p9, and p7 (nucleocapsid proteins – NC), each of which is proteolytically cleaved from a 53-kD Gag precursor by the HIV-1 protease.[6]

The viral core, that is surrounded by lipid proteins consists of RT, integrase (IN), protease and nucleocapsid proteins (NC), as well as 2 copies of positive strand viral RNA. All of this is surrounded by icosahedral capsid protein (CA). Furthermore, the matrix protein (MA) is found below the surface of the lipid bilayer, surrounding the virion. Within the lipid bilayer, are the viral envelope glycoproteins, external surface glycoprotein (SU) and the transmembrane glycoprotein (TM).[1]

1.3 HIV-1 entry into nucleus

In order for HIV to effectively infect a cell target, it has to introduce its genetic material into the cytoplasm of the cell. In this process, the main prerequisite is the initial attachment of HIV-1 envelope protein (Env) to CD4+ receptors on the surface of target cells. The development of HIV-1 attachment inhibitors has been driven by structural studies which have explained the “lock and key“ structure of the trimeric gp120 Env spikes present on virions with CD4+ receptors existing on the surface of target cells. Furthermore, the introduction of a crucial phenylalanine (F) residues in the outer part of CD4+ protein into the sunken pocket in gp120 produces a high-affinity interaction between the two.[7,8]

There have been many studies which were able to identify a handful of monoclonal antibodies against HIV-1 that have the ability to neutralize primary isolates of HIV-1 *in vitro* and hence prevent infection in models such as non-human primates. However, even

though the monoclonal antibodies might prevent virus infection in animal models, the dose amounts of the antibody would have to be significantly higher to exhibit any clinical effects in HIV-infected individuals.[8–10] Primary isolates are virus isolates that have been minimally split in primary cells, and it is believes that they might more closely resemble the virus found in people infected with HIV-1, than the virus that is split.[11]

Following the initial phase of HIV entry, the second phase entails the engagement of HIV-1 coreceptors. The initial attachment of trimeric gp120 to CD4 initiates a conformational change in the envelope that acts as a promoter of attachment/binding of the virion to a specific subgroup of chemokine co-receptors. These receptors consist of seven domains, spanning the entire membrane, and they typically help hematopoietic cells (type of stem cells) to travel down specific chemokine gradients to inflammation sites. There are 12 different chemokine receptors which work as HIV-1 coreceptors in cultured cells – only two of them are generally used *in vivo* (CCR5 and CXCR4).[8,12]

Lastly, the third phase is the fusion of the virion. With the attachment of surface gp120, CD4 and chemokine coreceptors, a significant conformational change in gp41 (second HIV-1 env protein) is generated. The assembly of this protein is in a form of a trimer located on the virion membrane. Once this coiled protein springs open and project three peptide fusion domains that hit the lipid bilayer of the target cell. The trimers of gp41 subsequently fold back onto themselves thus forming a hairpin structure. An inhibitor called enfuvirtide, managed to prevent the formation of this hairpin structure which is crucial for a successful fusion. The fusion consequently leads to introduction of HIV-1 viral core into the cytoplasm.[8,13]

1.4 Integration and its mechanisms

Integration is an essential step in the process of retroviral replication cycle. Having passed the NPC (nuclear pore complex), the PIC (pre-integration complex) arrives to the complex nucleus periphery and can create a functional provirus. There it can find the environment which has anchoring sites for chromosomes.[14,15] The process of integration

is mediated by IN (integrase protein) – encoded by the virus, and alongside reverse transcriptase, viral RNA and other viral core proteins is inserted into the cells during the infection process. Following the synthesis of viral DNA by reverse transcription it is strongly bound to IN and other proteins as an NPC that is transported to nucleus for integration.[16]

Viral DNA entry into the chromosomes is catalyzed by an enzyme called (HIV) integrase which binds to viral DNA ends. Following the integration of proviral DNA, it is subsequently replicated parallel with cellular DNA in the process of cell division. Provirus is utilized as a transcription template of viral RNA. Majority of the full-length viral RNA is utilized as genomic RNA in progeny virions, while the rest can be translated with the aim of yielding viral proteins.[16]

There are two key enzymatic reactions which are carried out by IN – strand transfer reactions for stable insertion of HIV-1 genome into chromatin and 3’ end viral DNA processing. It is possible for both the processing and oligonucleotide matching HIV-1 DNA ends integration to be replicated *in vitro* when only IN is present. Furthermore, in studies that employed hybrid viruses with modified integration preferences, the main role of integrase (IN) in selection of integration site has been supported.[15,17]

Tetramers of integrase correlate with growing viral DNA for the formation of functional integrase-viral DNA complex called intasome. This complex is associated with many cellular factors that allow the correct alignment of pre-integration complex in the nuclear space and relative to cellular chromatin.[15,18] The C-terminal domain (CTD), among all the retroviral integrase proteins, has the least sequence preservation. It contains four lysine residues which are acetylated by histone acetyltransferases. This is a type of post-translational modification (PTM), which controls integrase strand transfer activity, HIV-1 integration and eventually the reduction in viral replication.[15]

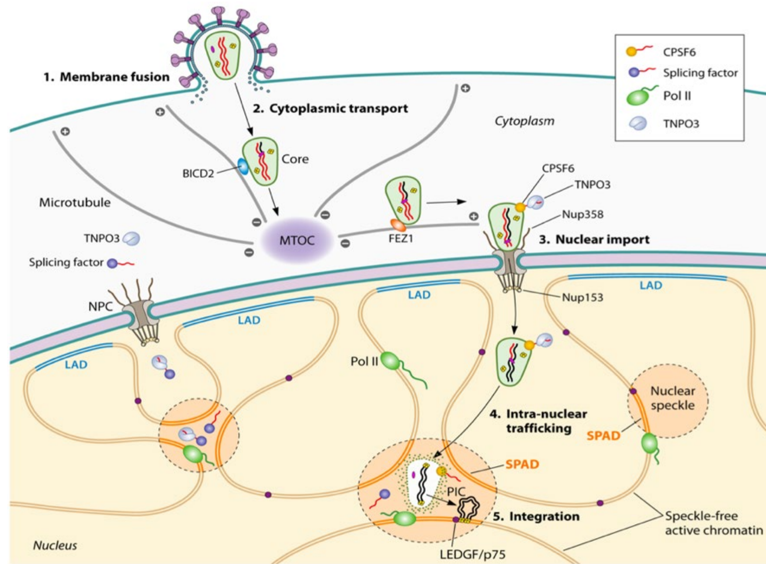


Figure 1: Capsid-Host Interactions for HIV-1 ingress[19]

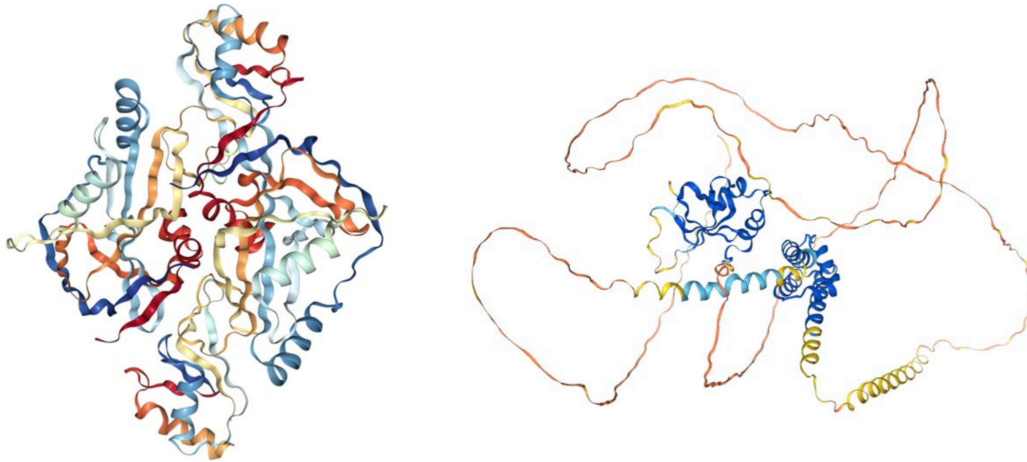


Figure 2: CPSF6 structure[20] **Figure 3**: LEDGF/p75 structure[21]

1.5 LEDGF/p75 overview

LEDGF/p75 is crucial in HIV-1 integration and has a role as a cellular cofactor. This activity involves its simultaneous involvement with the host chromatin and viral integrase enzyme. LEDGF/p75 mutants that are missing their chromatin or integrase-binding activity are acutely effective in their HIV-1 cofactor function. Replacement of the chromatin binding domain of LEDGF/p75 by heterologous chromatin binding domain leads to proteins supporting HIV-1 DNA integration.[22]

LEDGF/p75 is a growth-factor associated protein, that has roles in transcriptional regulation and cellular survival.[23]

However, the HIV-1 integration site distribution observed in LEDGF/p75-deficient cells expressing these figments is altered and defined by the specificity of the added chromatin binding domain. These findings indicate that the role of LEDGF/p75 chromatin-binding domain is to provide a tight interaction to the preintegration complex with the host chromatin.

Additionally, LEDGF/p75 perseveres tightly bound to the chromatin throughout all cell cycle phases.[22]

1.6 LEDGF/Integrase interaction

In the past decade, it has been concluded that host factors may play a significant role in HIV-integration targeting. A host factor called LEDGF/p75 (lens epithelium-derived growth factor/transcription cofactor p75) has been shown to have a crucial role in the process of integration.[23]

Initial research based on this interaction indicated that an interaction between HIV-1 IN and LEDGF/p75 might play an important role in retroviral integration. The following study has shown observations that the N-terminal zinc binding domain as well as the central core domains of IN have a role in its interaction with LEDGF/p75. Both domains are crucial for the process of IN nuclear localization and association of condensed chromosomes

and IN during mitosis. In the event of overexpression of LEDGF/p75, IN's core domain fragment was enlisted to the mitotic and nucleic chromosomes with a delivery pattern trait of a full-length protein. This implies that IN holds the determining factor for a LEDGF/p75 interaction. The outcomes determined that LEDGF/p75 is important in nuclear (creation of nuclear complex) and chromosomal targeting of HIV-1 IN.[24,25]

Furthermore, research done solely on LEDGF/p75 factor obtained data, which suggests that this factor can have an impact on the choice of target sites in cells during HIV-1 integration. LEDGF/p75-regulated genes recognized by transcriptional profiling have been shown to be the preferred integration targets, which tracks with the notion that LEDGF/p75 bound genes are favored during HIV-1 integration. This has been shown through a LEDGF/p75 knockdown which provided unsurprising results – HIV integration was increased in target DNA containing higher G-C content.[26,27]

Additionally, there are other studies which made similar conclusions, observing that there is a LEDGF/p75 reduction which has solid negative effects in viral replication.[27]

1.7 LEDGF interactions with IWS1, MLL1, PAF1 host factors

After the role of LEDGF was ascertained as important in HIV-1 integration and is a crucial cofactor of IN, further research showed that it particularly regulates HIV transcription in proviral reactivation and latency.[28] Latency is a process whereas virus transcription and production are activated to induce immune or virus-mediated cell apoptosis.[29]

LEDGF/p75 and IWS1 interaction

Host factor IWS1 has been identified as a novel LEDGF/p75 interaction associate.[30] By using several different methods and experimental approaches, a IWS1 and LEDGF/p75 interaction has been established and confirmed.

Another paper reported the role of LEDGF/p75 and IWS1 interaction in HIV latency. Furthermore, they identified a complex consisting of LEDGF/p75, IWS1 and Spt6 which can bind repressed HIV genomes which overall aids the process of HIV latency. This occurs through the inhibition of viral expression and management of histone occupancy on the long terminal repeat (LTR). The complex has also been shown to play a role in nonproductive infections in proliferating lymphocytes. During their return to the resting state, the latent infected could be beneficial to the viral reservoir. Through these findings it is clear that there is a crucial role IWS1 has in HIV-1 integration, but it is also important in post-integration silencing of HIV-1.[31]

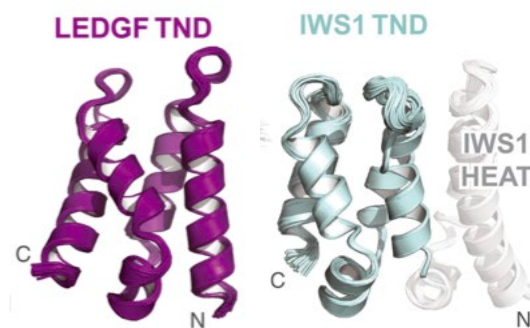


Figure 4: LEDGF AND IWS1 TND structures determined by NMR spectroscopy.[32]

One of the studies done on protein interactions has shown that the transcription elongation factor TFIIS N-terminal domains (TNDs) and conserved unstructured sequences - TND-interacting motifs (TIMs) are key elements in a network of interactions linking through the important members of the transcription elongation machinery.[32]

Nuclear magnetic resonance (NMR) spectroscopy is a useful technique used for determination of the molecular structure at the atomic level. Furthermore, it can detect conformational and phase changes, solubility, and diffusion potential.[33] A series of NMR studies conducted in 2021, have shown that IWS1 possesses the highest number of interactions with other TNDs and TIMs in comparison to other factors among which are both LEDGF and PAF1. It is important to mention, that IWS1 has both a TND and a part of TIMs which can possibly bring several TND and TIM-containing factors together all at once. Because of these observations, focus was primarily on IWS1 as a model for

establishing selectivity in TND-TIM interactions, and to focus on importance and sufficiency of these interactions within cells.

Motifs which are known to bind IWS1 and LEDGF homologs share many similarities and thus using this data from other species can help discover associated motifs in humans which bind TNDs – TIMs.[32]

In this study, a proteome-wide search provided findings such as – three TIMs in IWS1, whereas one of them is a motif that binds LEDGF and HRP2, and two members of the PAF1 complex – PAF1 and LEO1.

There are a handful of proteins in this network including IWS1, SPT6, LEO1 and PAF1 have multiple TIMs. This opens the possibility that TND- and TIM-containing factors may participate in dynamic higher-order structures, such as oligomers, or nucleate higher concentrations of their binding partners at biologically active sites.[32]

Furthermore, IWS1 gene is a Polymerase II (Pol II) elongation factor that was first discovered in *S. cerevisiae* yeast as a crucial protein interacting with Spt. The human IWS1 interacts with a homolog of Spt6 – SUPT6h and is key crucial for cell viability. It has been reported that IWS1 interacts with nuclear mRNA export factor – REF1/ALY, and that the depletion of IWS1 causes the nuclear retention of bulk mRNA in HeLa cell line. Hence, it seems that the interaction of SUPT6h with Pol II CTD Ser2p mediates co-transcriptional recruitment of IWS1 and numerous mRNA processing and export factors to their receptive genes.[34,35]

LEDGF/p75 and MLL1 interaction

It has been established that there is a solid connection between LEDGF/p75 and MLL1 as they both have rather comparable roles in provirus reactivation.[28]

MLL1 depletion has been shown to decrease proviral reactivation and hints at MLL1 has a role as predominant methyltransferase at the proviral long terminal repeat and might aid in its transcriptional activity in latency reversal process. In contrast to both LEDGF/p75 and PAF1, MLL1 is non-essential for proviral latency reservation. It has however been shown that activated provirus has accelerated reentry into the latent state as a result of

MLL1 depletion, which solidifies its function in proviral transcription activation during the reversal of latency.[28]

MLL has been known to encode a histone methyltransferase forming a ternary complex with LEDGF/p75 and tumor suppressor menin – MEN1. Furthermore, it is commonly targeted by chromosomal translocations. Various stable genetic rearrangements result in the formation of new fusion proteins linked through both adult and childhood *de novo* acute leukemias, and therapy-related acute myeloid and lymphoblastic leukemias.
The LEDGF/p75 binding part of MLL1 can be found in C-terminal fusions (MLL_N-fusion), which allows for tethering of these fusion complexes to MLL1 target genes, resulting in leukemic transformation.[36]
This LEDGF/p75-MLL1-menin complex has been structurally characterized; however, the published x-ray data only revealed a part of the ternary complex. Subsequently, it was identified that there is another menin-independent MLL1 binding site on LEDGF/p75. Both sites have been shown to be important for MLL fusion mediated transformation and signify valuable novel therapeutic targets.[37] It is important to note, that although LEDGF/p75 dependent chromatin tethering is crucial for MLL1 fusion driven leukemia and HIV-1 integration, a direct connection to H3K36me3 marks has not yet been ascertained. Besides HIV integrase and the MLL1-menin complex, LEDGF/p75 interacts with various cellular proteins via its integrase binding domain (IBD).[36]

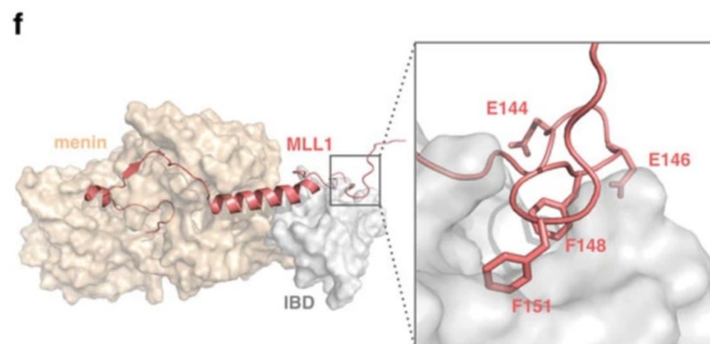


Figure 5: MLL1-menin-LEDGF/P75 ternary complex; red is MLL1, beige is menin and grey is LEDGF/p75 IBD.[36]

MLL1 gene and its main fusion partners include numerous protein-protein interaction (PPI), that have a critical role in gene expression regulation in normal physiology as well as in leukemia initiation and maintenance. Furthermore, wild-type MLL1 (WT MLL1) in another allele has been discovered to be key for MLL-r leukemia. Thorough biological, biochemical, and biophysical studies (namely X-ray crystallography and NMR) of alike PPIs have been executed and significant process has been accomplished in understanding their structures and structure-function relationships as well as mechanisms for leukemogenesis.[38]

Human MLL1 is a member of the MLL/KMT2A family of lysine methyltransferases (KMT), including MLL2 (KMT2D), MLL3 (KMT2C), MLL4 (KMT2B), MLL5 (KMT2E), SET1A (KMT2F), and SET1B (KMT2G). Other proteins that belong to this family also have roles as key regulators in gene transcription as well as in physiology and diseases.[39]

Although it is mainly known as a lysine methyltransferase (KMT), its predominant role is as a transcription factor; it has been discovered to bind thousands of gene promoters and play a role in their gene expression. Furthermore, it has an overall role in positive regulation of transcription of numerous essential genes in mammals (e.g. clustered Hox genes that mediate differentiation of multiple tissues as well as hematopoietic system during the process of embryogenesis.[38]

LEDGF/p75 and PAF1 interaction

The interaction between LEDGF/p75 and PAF1 has been established to be direct, whereas it occurs for both the LEDGF/p75 and PAF1 exogenously expressed in HEK293T cells and bacteria, and LEDGF/p75 and PAF1 endogenous complex in latently infected cells. Through additional truncation analysis conducted it has been shown that LEDGF/p75's integrase binding domain (IBD) and PAF1's C-terminal region are necessary to mediate their association.

Initially, PAF1 was identified to be an elongation factor in promotion and regulation of Pol II transcription in many organisms which led to further research on its role in HIV

transcription and latency, which has shown that PAF1 is non-essential for proviral transcription in reversal of HIV latency. Overall, it has been demonstrated that LEDGF/p75-associated PAF1 complex has a key role in establishment and perseveration of HIV latency by suppressing proviral transcription, and that LEDGF/p75 works upstream of PAF1 and might be useful in recruitment of PAF1 in order to suppress long terminal repeat activity in HIV latency. [28] Additionally, another study has concluded that PAF1 plays a definitive role in the defense of host cells in the process of viral infection, but it is likely that PAF1 is more active in this role in other viruses than in HIV-1.[40]

In an innate immune response, the organism depends on the primary detection of pathogens by host pattern recognition receptors. This can then initiate numerous signaling pathways that culminate in responsive gene expression aimed to prevent or fight the infection.[41]

PAF1 complex, or polymerase associated factor 1 complex is becoming progressively more implicated in a response similar to the mentioned one at the level of gene expression, which is a direct result from its crucial roles in transcriptional elongation, histone modification and chromatin remodeling, Alongside PAF1, other key members in this nuclear complex are LEO1, CTR9 and CDC73.[42]

PAF1C regulation of gene expression influences the ability of cells to exhibit an antimicrobial response. More specifically, it has been observed that the depletion of PAF1 disrupts interferon-related signaling and leads to increases in viral replication in some viruses. Furthermore, PAF1 and other complex members have been found to act as restriction factors for HIV-1.[41,43]

Research has shown that PAF1 displays significant negative correlations with levels of cell-associated HIV RNA and might be a contributing factor in controlling viral expression and ongoing replication during anti-retroviral therapy (ART). However, it is not included in the interferon-α treatment in vivo, in contrast to most of other recognized anti-HIV-1 restriction factors.[43]

Furthermore, it has been observed that PAF1C expression restricts HIV-1. Specifically, it restricts infection of a monocytic cell line and is expressed in primary CD4+ T cells,

monocytes, and macrophages. This results in less viral transcripts in early and in late stages of infection, as well as less integrated proviral DNA.
It has been shown despite the slight chance that it might be important in this process, PAF1's cell cycle activity was not the reason for the viral restriction.
One of two possibilities for PAF1C's role in viral restriction would have to start with having PAF1 interact with RNA polymerase II which leads to its involvement in transcription, elongation, and degradation. As there is a significant amount of PAF1C found in the nucleus it is possible that its activity in gene transcription may increase expression of anti-viral factors. PAF1 complex seems to inhibit early events of viral life cycle from reverse transcription to integration step.[44]

1.8 SPADs/LADs

During the process of HIV-1 integration transcriptionally active SPADs (speckle-associated domains) are preferred, in comparison to LADs (lamina-associated domains), a heterochromatin which is unpreferred.[45]

Across many cell types, it has been shown that HIV-1 significantly favors integration into SPADs, specifically around 30% of integrations happened within a SPAD while a big part of other sites was delivered to the nearby chromatin. Targeting of HIV-1 pre-integration complex to nuclear speckles (NS) clarifies the integration site preference it possesses, as an important preference for integration into SPADs of the human genome. This is a trafficking pathway shown to be dominant in several cell types and it has vital importance for preferred HIV-1 integration targeting into SPADs.[46]

The integration into SPADs hugely depends on CA-CPSF6 interaction however, which is crucial for the entire process of HIV-1 integration.[46] Furthermore, CPSF6 is key in enabling HIV-1 entry into the nuclear interior well past the nuclear periphery. In the case where this interaction does not occur, the pre-integration complex uncharacteristically localizes to the nuclear periphery and targets LADs for HIV-1 integration.[47,48]

HIV-1 does not favor integration into LADs chromatin areas, consisting of sparse gene regions and transcriptionally silent genes that work together with lamina at the nuclear periphery. It has been shown that LADs are discrete in their function and precisely defined chromatin units. There is a tiny number of genes within LADs that are transcribed by Pol II which suggests that transcription and localization within LADs are compatible.[47]

2 Description of Aims and the Research Question

The main aim of my project is to investigate and test the role of the host cofactor –IWS1, MLL1, PAF1 in HIV-1 integration targeting into the nucleus, and down the line see if they interact with LEDGF/p75 during that process. It has been assessed that most of these genes have high percentages to which they can be knocked out, which is useful for further experiments and determination if these genes/factors play an important role during HIV-1 integration targeting.

Another aim would be seeing if there is an interaction with integrase or both LEDGF/integrase. It is well established that an interaction between LEDGF and integrase exists – targeting intasome to the middle of the genes. Small molecule inhibitors of LEDGF/p75-integrase interaction reduce the activation potential of resulting integrated proviruses, indicating that there might be a clinical potential for drugs altering integration landscape.

By doing different types of experiments the role of host factors potentially interacting with LEDGF/p75, integrase or both can be evaluated.

Proposed methods for this research topic would be doing a lentiviral (lentiCRISPR) transduction with adherent cells and generating knockdown cell lines from this experiment. Furthermore, the cell lines are infected with the virus and after they start growing normally in selection media (medium with puromycin), the knockdown of each cell line can be validated through immunoblotting, and Illumina sequencing (Next-generation sequencing method) down the line.

3 Presentation of the methods

Cell culture

The cells which used for all experiments are HEK293T cells (mammalian derivative kidney cell line). Cells are cultured from passage 7 or 8 (P7 or P8) and are split every 2-3 days around a confluency of 70-80%, in 100 mm cell culture plates. HEK293T cells are washed with 1xPBS (Thermo Fisher Scientific), trypsinized (0.25% v/v) (Corning) and resuspended in complete DMEM. A complete DMEM (CDMEM), is DMEM (Gibco) supplemented with 4.5 g/L D-Glucose, L-Glutamine and Sodium Pyruvate as well as heat-inactivated Fetal Bovine Serum, FBS, (Gibco) and 100xPenicillin Streptomycin Solution (Corning) . Cells are maintained for up to 3 weeks and are kept in a 37ºC 5% CO_2 humidified incubator.

CRISPR-Cas9 Transient Knockdown

An sgRNA (Horizon Discovery, Synthego) is reconstituted in 1xTE buffer (IDT) and spinned down. The sgRNA is then used alongside Cas9 (Synthego) to make an sgRNP (single guide ribonucleoprotein) complex. Cells are heat treated in a 37ºC heating plate for 15 minutes. HEK293T cells are trypsinized and pelleted. The pellet is then reconstituted in the Nucleofector solution (Lonza). Fresh CDMEM (DMEM supplemented with FBS and antibiotics) is plated in a 6-well plate (Thermo Fisher Scientific) and pre-incubated in the 37ºC 5%CO_2 humidified incubator. sgRNP complex mix is then spinned down and entire contents is added to the HEK293T pellet and Nucleofector solution. The entire mix is transferred in a nucleofector cuvette (Lonza) making sure there is no bubble formation. Cuvette is put into the Amaxa Nucleofector device (Lonza); Q-01 program is selected and applied. After nucleofection, 500 uL of pre-warmed CDMEM is added into the cuvette and everything is transferred to a 6-well plate using single-use pipettes (Lonza). Plate is incubated at 37ºC in 5%CO_2 humidified incubator for 72h. Cells can then be harvested and editing efficiency can be analyzed through immunoblotting.

Lentiviral Transductions

The cells used for transfections are modified HEK293T cells – hHL6 cells. A 100 mm plate at a confluency of ~70% is used. A desired number of cells is pelleted in a multifuge (Thermo Fisher Scientific). A bottle of DMEM (Gibco) supplemented with 4.5 g/L D-Glucose, L-Glutamine and Sodium Pyruvate, is warmed up in a 37ºC water bath. By using the formula: V=MOI x CN / VT X 1000, volume of lentiviral particles can be calculated. Once the cells are trypsinized and pelleted, Edit-R All-in-one Lentiviral sgRNA particles (Horizon Discovery) are thawed on wet ice and brought into the hood. Calculated amounts are pipetted into 1.7 mL tubes (Eppendorf) using micropipettes (Thermo Fisher Scientific), transduction medium is added, and samples are resuspended gently. Tubes are put in the 37ºC 5% CO_2 humidified incubator with open caps for 1h. Every 15min tubes are closed, taken into the hood and lightly flicked. Following the incubation, CDMEM is plated in each well used, and is also added to each tube, from which after mixing is transferred into the well. Plate is incubated in 37ºC 5% CO_2 humidified incubator for 5-6h. After the incubation, medium is carefully removed, new medium is added and resume incubation at 37ºC 5% CO_2 humidified incubator for 48h. 48h post-transduction medium is replaced with selection medium – supplemented with 4.5 g/L D-Glucose, L-Glutamine and Sodium Pyruvate as well as 1.5 ug/mL puromycin (10 mg/ml) (Gibco). Selection medium is replaced every 48-72h and cells are split when necessary. Once cells are growing normally in the selection medium, cells are expanded to freeze a sufficient number of aliquots for future use. Freezing media used is a mix of Dimethyl Sulfoxide, DMSO, (Sigma-Aldrich) and ice-cold FBS (Gibco). Cell pellet is resuspended in the freezing media and aliquoted in a cryovial (Sigma-Aldrich).

Cell lysis

In order to determine how effective a knockdown truly was following a CRIPR-Cas9 transient knockdown, immunoblotting is conducted. Cells are first harvested, pelleted and put on wet ice immediately. Following, is the cell lysis with the RIPA buffer (Radioimmunoprecipitation assay buffer) which contains Tris-HCl pH 7.4 (0.5 M), NaCl (5 M), EDTA (0.5 M), IGEPAL CA-630 (100% v/v), SDS (10% w/v), 1 capsule of cOmplete Mini EDTA-free protease inhibitor (Sigma-Aldrich) and ddH2O – for a total volume of 10 mL. Lysates are then incubated on wet ice for 30 minutes and vortexed (Fisher Scientific) every 10 minutes. Following the incubation period lysates are centrifuged (Eppendorf) at maximum speed (21,000xg) for 10 minutes at 4ºC, after which the supernatants are transferred to new tubes on ice. A part of the supernatant is saved for BCA assay (Bicinchoninic acid assay) and remainder is stored at -80ºC. BCA Assay is performed to determine the total protein concentration of the samples. Protein standards are prepared (A-I) in various concentrations for the final volumes of 2000, 1000, 750, 500, 250, 125, 25 and 0 ug/mL, and stored at RT – Pierce™ BCA standards (Thermo Fisher Scientific) and deionized water. Standards and samples are pipetted in duplicates into a 96-well plate (Thermo Fisher Scientific). A Pierce™ BCA working reagent (Thermo Fisher Scientific) (reagent A : reagent B = 50:1) is added to all wells and the plate is incubated at 37ºC for 30 minutes in a tabletop incubator. Following incubation, absorbance is measured at 562 nm in a microplate reader.

SDS-PAGE and Western Blot

Using the results from the BCA assay analysis, volumes required for each sample for SDS-PAGE can be calculated. All the samples need to have 30 ug of total protein concentration. Based on that data mixtures containing the lysate, 4.25x sample buffer – consisting of 5 uL beta/2-mercaptoethanol, 10 uL 1M DTT and 85 uL 4x Laemmli Smaple Buffer (BioRad), and UltraPure water can be made (up to a total of 20 uL). Once made, samples are boiled in a heating block (Thermo Fisher Scientific) at 100ºC for 10 minutes. Samples are then centrifuged (Sigma-Aldrich) at 16,500xg for 30 seconds at RT. The

protein gel box is assembled next, using a pre-cast Bolt™ 4-12% Bis-Tris Plus gel (12-well) (Thermo Fisher Scientific) and filling the center of the box with 1xMES SDS running buffer, made from 20x MES SDS Running Buffer (Thermo Fisher Scientific) and water, and let it overfill to the sides approximately 5 cm. The samples are then loaded using the gel pipette tips, starting with the protein ladder – Precision Plus Protein Dual Color Standard (BioRad) . Once all samples are loaded the gel is first run at 100V until the samples reach the middle of their run and then switch to 200V. Using a pre-cut Trans-Blot Turbo Transfer Pack (BioRad) – with mini format 0.2 um PVDF. The gel is gently placed in between two pre-cut membranes and the blot is run in a Trans-Blot machine (BioRad) – Mini gel > High MW.

Blocking and antibody incubation

Fresh Milk A – Blotting-Grade Blocker (BioRad) is prepared (1g in 20 mL) and Milk B – Blocking-Grade Blocker (BioRad) (2.5g in 20 mL) – used for non-specific binding and store it in 4ºC fridge. The membrane from the transfer sandwich is removed and placed in a gel box in which Milk A is added. The blot is then incubated for 1h on a tilted shaker at RT. After the incubation is over, Milk A is removed and Milk B with the appropriate concentration of the antibody (1:10,000; 1:2,500) is added and incubated in the cold room overnight (on a tilted shaker). After 12-18 hours milk is removed and wash the membrane is washed 3x with 10 mL of PBST (Phosphate-Buffered Saline with Tween20). After the last PBST wash, and and add appropriate secondary antibody is added to 10 mL of Milk B in the gel box and everything is incubated for 1 hour. Following the incubation, 3x PBST wash is repeated. A 1:1 Pierce™ ECL Western blotting Substrate detection solution with reagents A and B (Thermo Fisher Scientific) is pipetted over the membrane right before imaging with a GelDoc (BioRad). For the imaging: Chemiluminescence (for samples) > manual exposure time > Colorimetric (for marker) > auto exposure time > gallery > select the sample sand marker photo > merge > print and save to USB. After imaging, 3x wash with PBST is done and the membrane is stripped using 10 mL of Restore™ PLUS Western Blot Stripping Buffer (Thermo Fisher Scientific) for 10 minutes on the tilted

shaker at RT. Washing steps are repeated and membrane is blocked again with Milk A for 1 hour. Following the Milk A incubation, Milk B and appropriate concentration of loading control antibody (beta-actin) is added. The membrane is stained overnight and imaging steps are repeated the next day. Analysis is done using ImageJ to see the depletion level of the proteins.

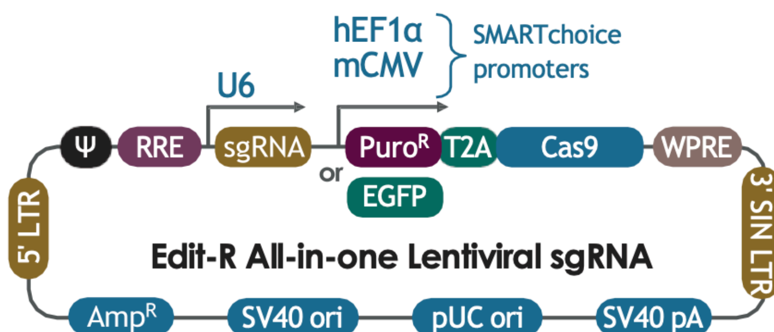


Figure 6: map of Edit-R All-in-one Lentiviral sgRNA vector.[49]

Transduction of cells with Edit-R All-in-one Lentiviral sgRNA particles - mechanism

The Edit-R All-in-one Lentiviral sgRNA (single-guide RNA) vector has a human codon-optimized form of *S. pyogenes cas9* (*csn1*) gene and the marker for puromycin resistance (Puro^R). Both of them are presented as a bicistronic transcript (encoding two independent proteins) with a 2A peptide sequence linker controlled by a single promoter (as seen on the graphic above). The location of Puro^R is purposefully upstream of Cas9 coding region and not on 3' end, as to prevent any additional amino acids being added to the C-terminus and hence preserving the complete Cas9 endonuclease activity. The expression of gene-specific sgRNA is controlled by the human U6 promoter, and the optimized sgRNA scaffold utilized in Edit-R All-in-one Lentiviral sgRNA can additionally improve the efficiency of the gene knockout. There are two existing promoter options (hEF1alpha and mCMV) for the selection of Cas9 nuclease vector with the most active

promoter for specific cells of interest. These sgRNAs are specific to the gene of interest or choice. Its crRNA (CRIPSR RNA) region contains 19-20 nucleotides which are identical to genomic DNA target site, followed by non-variable sRNA scaffold that consists of *S. pyogenes* tracrRNA (trans-activating CRISPR RNA) sequence. The genomic DNA target of interest has to be directly upstream if PAM (Protospacer Adjacent Motif). The principal PAM nucleotide sequence of *S. pyogenes* is NGG. [49]

4 Other key methods in overall project

p24 assay

HIV p24 antigen, mentioned in the subchapter *1.3 Structure of the HIV-1 virion*, is the most abundant HIV-1 protein and is key for capsid assembly (the capsid surrounds the genetic material of the virus). p24 antigen can be detected by an assay and the aim of this method it to diagnose early HIV-1 infection, in a clinical setting, during the period where HIV-specific antibodies are not yet detectable. It is important to mention however that these p24 assays have low sensitivity which has shown to not be useful in observation of HIV-1 throughout the course of infection.

Following this observation, the development of high sensitivity p24 assays began – one of which has been named digital enzyme-linked immunosorbent assay (ELISA), as it is based the microscopic well capture and digital camera platform. This high sensitivity assay has the ability to allow detection for very low-level concentrations of p24 antigen. The lowest limit of quantification being 0.01 pg/mL (picogram/milliliter) and lowest limit of detection 0.003 pg/mL. With the development of this assay and taking in consideration its high sensitivity, the detection of early HIV-1 infection has been made easier. Moreover, p24 assay is also being applied for the measurement of *in vitro* HIV-1 latency reactivation.[50]

Luciferase Assay

Luciferase assay has the goal of determining if a protein of interest can either activate or suppress the expression of the target gene. Unlike some other assays (e.g. ChIP, EMSA), luciferase is assay can establish a functional connection between the presence of a protein and the amount of gene product being produced. It is however unable to verify if the protein directly interacts with DNA itself – the protein could implicitly affect transcription through activation or suppression of another protein or protein complex that affects the transcription process.

This assay is used for bioluminescence by numerous organisms in nature. In the lab however, a construct in which the regulatory region of a target gene is fused with the DNA coding sequence for luciferase is produced. A separate DNA construct encodes the protein which is theorized to affect transcription. A transfection is conducted with e.g. HEK293T cells, using both constructs. If the protein can upregulate transcription of the target gene, cells will express luciferase; if there is downregulation, luciferase expression will be decreased. This is assessed ~2-3 days after the initial cell transfection. Subsequently, cells are lysed, and the cell contents are transferred into a reaction tube. When the appropriate substrate is added, luciferase will be able to catalyze a reaction which produces light, which will then be detected with a luminometer – device that precisely quantifies the amount of light produced in a reaction tube. The quantity of light produced offers a quantitative measure of the effect the protein has on expression of the target gene.[51]

Sequencing – background and Next-generation sequencing

Although the next-generation sequencing is the method to be used in this research, it is crucial to mention Sanger sequencing as the pre… method. It was developed in 1977, with the first mention of it in 1975. This method used specific chain-terminating nucleotides that do not contain a 3'-OH group; hence no phosphodiester bonds can be formed by DNA polymerase, leading to termination of the developing DMA chain at that

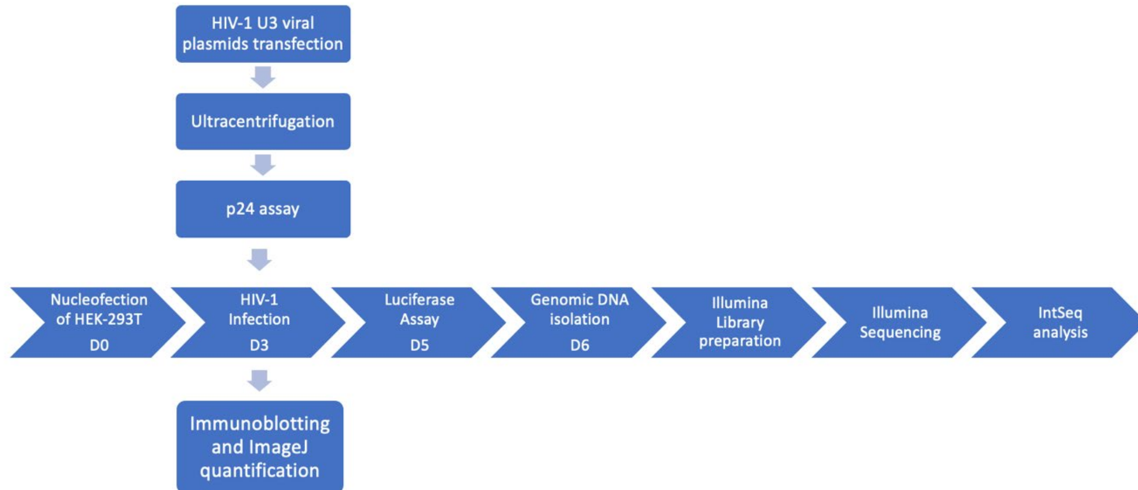
site. The dideoxynucleotides triphosphates (ddNTPs) are fluorescently or radioactively labeled for detection in automated sequencing machines or gels (for sequencing), respectively.
Furthermore, although it is a quite slow method in comparison to next-generation sequencing (NGS) many improvements have been made in its methodology and automation which allows it to still be utilized as one of the most suitable sequencing methods for numerous experiments especially on a smaller scale.[52]

When it comes to second generation sequencing or next-generation sequencing (NGS), one of the most popular technologies is Illumina sequencing. The basis of this technique is known under a name “bridge amplification” where DNA molecules that contain appropriate adapters ligated on each end are utilized as substrates for repeated amplification synthesis reaction on a solid support e.g. glass slide. This solid support contains oligonucleotide sequences which are complementary to a ligated adapter. Oligos on the slide are spread out so that the DNA – subjected to repeat rounds of amplification, generates clonal clusters containing approximately 1000 copies of each oligonucleotide fragment. Each individual glass slide can support millions of parallel cluster reactions.
Moreover, during the synthesis specifically modified nucleotides which correspond to each of the four bases, with their individual fluorescent label are then integrated and detected.
Nucleotides also work as synthesis terminators for every reaction and are released after detection for the next round of synthesis. The repetition of the reactions reaches around 300 or more rounds in total; the utilization of fluorescent detection amplifies the speed of detection as a result of direct imaging, in contrast to camera-based imaging.

This type of sequencing – Illumina, aids numerous protocols from genomic sequencing and targeted sequencing to RNA sequencing, and CHIP-seq and methylome methods. Different sequencing machines offer different levels of throughput, some of which include MiniSeq, NextSeq and HiSeq models. The most commonly used one, in the experiments such as the ones performed in this paper, is NextSeq, which can provide 120GB with 400 million reads at 2x150bp read length.[52]

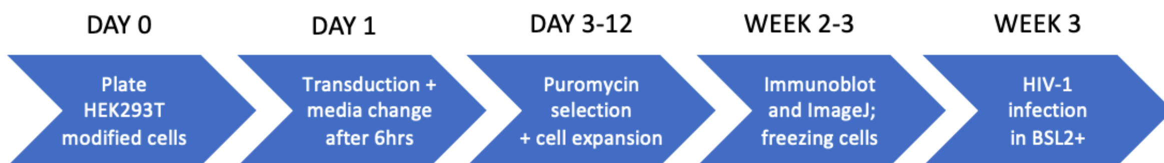
5 Workflow of the research

Project workflow:



This entire part of the experiment has been done/will have been done by my colleague. Essentially both his nucleofected cells and my transduced KO cells are to be used for both the HIV-1 infection and in other steps that follow in this project.

Workflow of my work:



The cells used in the lentiviral transduction experiment were modified HEK293T cells (human embryonic kidney cells which express the SV40 large T antigen). Since this modification was conducted by collaborators who have not published their results yet, I am unable to disclose what the exact edit is. However, the morphology and overall behavior of the cells is almost identical to HEK293T.

CRISPR mechanism and relevance

CRISPR (clustered regularly interspaced palindromic repeats)-Cas9 (CRIPSR associated protein) system is a bacterial and archaeal adaptive defense mechanism that works to identify and silence incoming foreign nucleic acids. At the point of infection by a bacteriophage or any other foreign elements, a host organism can integrate short sequences from the genetic invaders, names protospacers, into a specified region of the genome (locus) among short palindromic DNA repeats of various lengths. Several spacer-repeat units are grouped at the CRISPR locus to form the CRISPR array. The entirety of the locus including the array is transcribed by RNA polymerase into primary transcript (pre-crRNA), followed by processing into a small mature crRNAs that include sequences corresponding to the foreign, invading DNA. The crRNAs guide a multifunctional protein or a protein complex with a goal to cleave corresponding target DNA adjacent to PAMs. This way the organism acquires a way to shield itself from any later infections. [53]

When it comes to engineering a CRISPR-Cas9 platform for mammalian gene editing, *Streptococcus pyogenes* has been widely studied. There are only three steps in total which are necessary for targeted DNA cleavage at specific target sites adjacent to PAM – 1) endonuclease Cas9, 2) mature crRNA which programs Cas9, processed from transcription of CRIPSR locus and array that complexes with 3) additionally CRISPR locus-encoded RNA (tracrRNA). Another option would be crRNA fusing to tracrRNA and thus creating a chimeric structure called sgRNA.[54,55]

Following the site-specific dsDNA cleavage, the cell can repair this break in one of two ways – either through a non-homologous end joining (NHEJ) or homology directed repair (HDR). The first is often flawed as it can result in small insertions and deletions (indels) that can result in nonsense mutations leading to gene disruptions that produce gene knockouts. This endogenous DNA break repair process combined with the highly controllable *S. pyogenes* CRISPR-Cas9 system, which permits a readily engineered platform to permanently disturb gene function.[56,57]

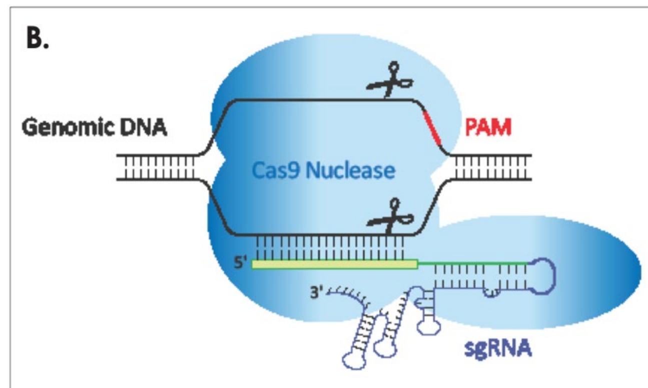


Figure 7: CRISPR-Cas9 system, where Cas9 nuclease is programmed by crRNA (green): sgRNA to cut both genomic DNA strands 5' of the PAM.[49]

Western Blot mechanism and relevance

Western blotting is an essential technique in molecular biology which is regularly used in various areas of research. The main role of a western blot is identification of specific proteins from an intricate mixture of proteins extracted from cells. The main steps in this technique are the following – 1) separation by size, 2) membrane transfer, and 3) staining the target protein with appropriate primary and secondary antibody for visualization.[58]

A western blot functions by separating a mix of proteins through their molecular weights (measured in kiloDaltons or kDa) and type through gel electrophoresis (SDS-PAGE). The gel is then transferred to a membrane on which the specific bands for each protein can be observed, followed by antibody incubation with appropriate specific antibodies which are supposed to detect the protein of interest. This antibody can then be detected through imaging with a GelDoc machine. It is expected that only a single band is visible for the specific protein of interest. It is important to always include a control in case the protein of interest might not be visible in a specific sample – this is especially relevant for knockout cell lines. In case the knockout was successful, a band will not be visible.

5 Results

To choose a specific guide RNA, or more specifically single-guide RNA, a transient CRISPR-Cas9 nucleofection was conducted using a set of 3 sgRNAs. Following the nucleofection, on day 3 immunoblotting was conducted and the best knockdown was chosen as the sgRNA which would be used in the lentiviral transduction – for this a LentiCRISPRv2 sgRNA was used.

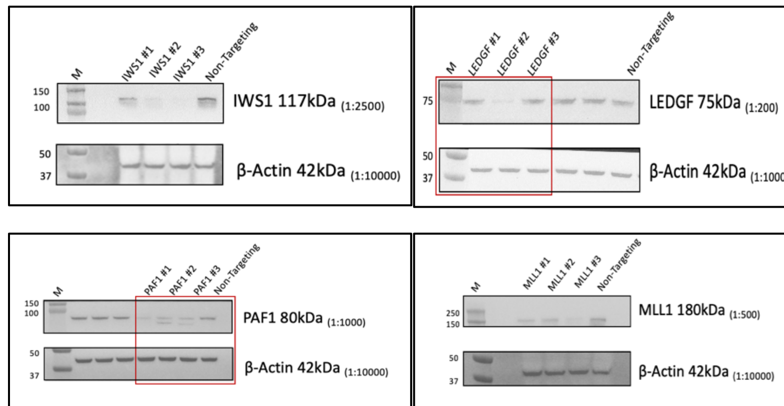


Figure 8: Immunoblotting of sgRNAs was conducted following a transient nucleofection using a variant of HEK293T cells. The genes of interest were IWS1, LEDGF, PAF1 and MLL1. All the blots include a marker, sgRNA and non-targeting control.

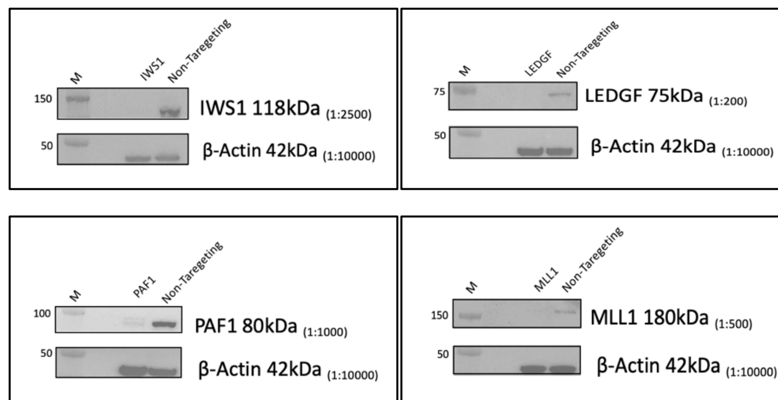


Figure 9: Immunoblotting of LentiCRISPR sgRNAs was conducted following a lentiviral transduction using a variant of HEK293T cells. The genes of interest were the same as in the nucleofection. All the blots include a marker, sgRNA and non-targeting control.

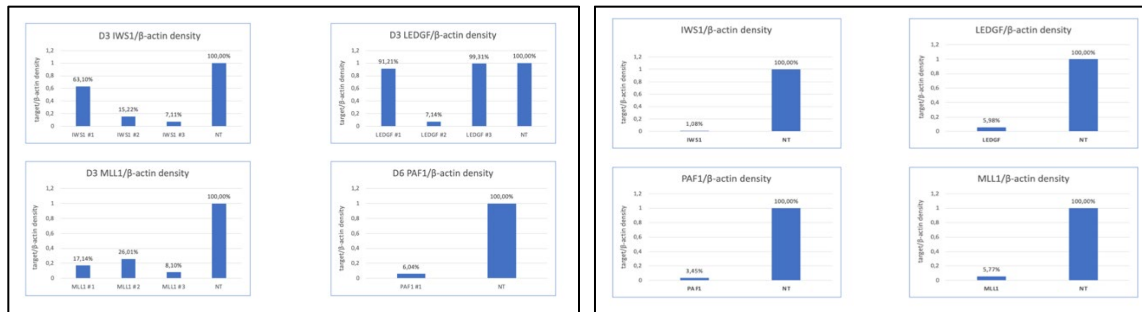


Figure 10: ImageJ analysis of western blots for the following genes of interest – IWS1, PAF1, LEDGF and MLL1, including a non-targeting control for comparison –results are from the transient nucleofection. **Figure 11**: ImageJ analysis of western blots for the genes of interest–results are from the lentiviral transduction. The graph shows significant depletion levels in all genes of interest, particularly in IWS1.

From these results it can be observed that in the nucleofection there was always one out of three which was depleted the most, which can then be optimized for use in other experiments. The levels between the nucleofection and transduction of the same sgRNAs in all host factors are overall comparable and the depletion levels are low enough that all generated cell lines can be used for the HIV infection and other steps leading up to Illumina sequencing.

6 Discussion

From repeated measurements, the results indicate that the knockout cell lines from all host factors of interest have been effectively knocked out, especially IWS1. There is an exception of the factor PAF1, where in several subsequent immunoblotting the levels went up indicating that it was not stable at 3.45%. In order for results to be reliable it is important to have a stable knockout cell line, and thus optimization is important.

Identifying HIV-1 host factors and characterizing them is a very important objective in HIV-1 biology. One method that has made and continues to make a difference is CRISPR-Cas9 gene editing technology[59] – in this paper utilized through a CRICPR-Cas9 nucleofection and LentiCRISPR transduction. This method is especially useful in studying function and mechanisms to learn more about the host factors and their potential future

roles. IWS1, PAF1 and MLL1 knockout cell lines that were generated could be useful in further studies of the mechanism of HIV-1 integration and in establishing a relationship between a genotype and phenotype.

Host factors have key roles in viral pathogenesis and replication, which makes them promising targets for next-generation therapeutics development.[60] Although in this paper the focus is not on the work with antiretroviral drugs, it is important to mention one example. One of the HIV-1 host factors – CCR5 is a binding site for the FDA-approved antiretroviral drug Maraviroc which works to prevent the binding of the viral envelope and thus inhibiting viral entry.[61,62]

To be able to reach this step at one point, it is important to conduct additional experiments including HIV-1 infection of the knockout cell lines to observe how stable they are when the virus is present and subsequently mapping integration sites to ensure if there was any change in integration site targeting by having knocked out these specific genes. In case there is a significant change this could help in development of next-generation therapeutics that inhibit viral entry.

Since the knockouts were successful as seen in very significant depletion levels from ImageJ analysis, my overall conclusion is that they should be utilized for further analysis to see if they play any important roles during HIV-1 integration targeting, by mapping of integration sites to observe if there is any change in integration site targeting by having knocked out any of these host factors.

7 References

1. German Advisory Committee Blood (Arbeitskreis Blut): Assessment of Pathogens Transmissible by Blood. Assessment of Pathogens Transmissible by Blood: Human Immunodeficiency Virus (HIV). . *Transfus Med Hemother* . 2016;3(43):203-222.
2. World Health Organization. HIV statistics, globally and by WHO region, 2023 - Epidemiological Fact Sheet.
3. Kahn JO, Walker BD. Acute Human Immunodeficiency Virus Type 1 Infection. *New England Journal of Medicine*. 1998;339(1). doi:10.1056/nejm199807023390107
4. Schwetz TA, Fauci AS. The extended impact of human immunodeficiency virus/AIDS research. *Journal of Infectious Diseases*. 2019;219(1). doi:10.1093/infdis/jiy441
5. Deeks SG, Overbaugh J, Phillips A, Buchbinder S. HIV infection. *Nat Rev Dis Primers*. 2015;1(1):15035. doi:10.1038/nrdp.2015.35
6. Epstein FH, Greene WC. The Molecular Biology of Human Immunodeficiency Virus Type 1 Infection. *New England Journal of Medicine*. 1991;324(5):308-317. doi:10.1056/NEJM199101313240506
7. Kwong PD, Wyatt R, Robinson J, Sweet RW, Sodroski J, Hendrickson WA. Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature*. 1998;393(6686):648-659. doi:10.1038/31405
8. Greene WC, Matija Peterlin B, Stremlau MH. Molecular biology of HIV. In: *Sande's HIV/AIDS Medicine*. Elsevier; 2012:25-43. doi:10.1016/B978-1-4557-0695-2.00003-1
9. Pantophlet R, Burton DR. GP120: Target for Neutralizing HIV-1 Antibodies. *Annu Rev Immunol*. 2006;24(1):739-769. doi:10.1146/annurev.immunol.24.021605.090557
10. Wu X, Yang ZY, Li Y, et al. Rational Design of Envelope Identifies Broadly Neutralizing Human Monoclonal Antibodies to HIV-1. *Science (1979)*. 2010;329(5993):856-861. doi:10.1126/science.1187659
11. Takefman DM, Sullivan BL, Sha BE, Spear GT. Mechanisms of Resistance of HIV-1 Primary Isolates to Complement-Mediated Lysis. *Virology*. 1998;246(2):370-378. doi:10.1006/viro.1998.9205
12. Doms RW, Trono D. The plasma membrane as a combat zone in the HIV battlefield. *Genes Dev*. 2000;14(21):2677-2688. doi:10.1101/gad.833300
13. Chan DC, Kim PS. HIV Entry and Its Inhibition. *Cell*. 1998;93(5):681-684. doi:10.1016/S0092-8674(00)81430-0
14. Volberding P, Greene W, Lange JMA, Gallant JE, Sewankambo N. *Sande's HIV/AIDS Medicine: Medical Management of AIDS 2013: Second Edition*.; 2012. doi:10.1016/B978-1-4557-0695-2.00048-1

15. Lusic M, Siliciano RF. Nuclear landscape of HIV-1 infection and integration. *Nat Rev Microbiol*. 2017;15(2). doi:10.1038/nrmicro.2016.162

16. Robert Craigie, Frederic D Bushman. HIV DNA integration. *Cold Spring Harb Perspect Med* . Published online July 2012.

17. Bushman FD, Fujiwara T, Craigie R. Retroviral DNA Integration Directed by HIV Integration Protein in Vitro. *Science (1979)*. 1990;249(4976). doi:10.1126/science.2171144

18. Krishnan L, Li X, Naraharisetty HL, Hare S, Cherepanov P, Engelman A. Structure-based modeling of the functional HIV-1 intasome and its inhibition. *Proc Natl Acad Sci U S A*. 2010;107(36). doi:10.1073/pnas.1002346107

19. Jang S, Engelman AN. Capsid–Host Interactions for HIV-1 Ingress. *Microbiology and Molecular Biology Reviews*. Published online September 2023. doi:10.1128/mmbr.00048-22

20. CPSF6 Protein Overview: Sequence, Structure, and Function.

21. PSIP1 Gene - PC4 And SRSF1 Interacting Protein 1.

22. Astiazaran P, Bueno MT, Morales E, Kugelman JR, Garcia-Rivera JA, Llano M. HIV-1 integrase modulates the interaction of the HIV-1 cellular cofactor LEDGF/p75 with chromatin. *Retrovirology*. 2011;8(1):27. doi:10.1186/1742-4690-8-27

23. Schrijvers R, Debyser Z. Role of LEDGF/p75 in Cell Biology and Disease Pathogenesis. In: *Encyclopedia of AIDS*. ; 2018. doi:10.1007/978-1-4939-7101-5_72

24. Cherepanov P, Maertens G, Proost P, et al. HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells. *Journal of Biological Chemistry*. 2003;278(1). doi:10.1074/jbc.M209278200

25. Maertens G, Cherepanov P, Pluymers W, et al. LEDGF/p75 is essential for nuclear and chromosomal targeting of HIV-1 integrase in human cells. *Journal of Biological Chemistry*. 2003;278(35). doi:10.1074/jbc.M303594200

26. Ciuffi A, Llano M, Poeschla E, et al. A role for LEDGF/p75 in targeting HIV DNA integration. *Nat Med*. 2005;11(12). doi:10.1038/nm1329

27. Shun MC, Raghavendra NK, Vandegraaff N, et al. LEDGF/p75 functions downstream from preintegration complex formation to effect gene-specific HIV-1 integration. *Genes Dev*. 2007;21(14). doi:10.1101/gad.1565107

28. Gao R, Bao J, Yan H, et al. Competition between PAF1 and MLL1/COMPASS confers the opposing function of LEDGF/p75 in HIV latency and proviral reactivation. *Sci Adv*. 2020;6(20). doi:10.1126/sciadv.aaz8411

29. Tanaka K, Kim Y, Roche M, Lewin SR. The role of latency reversal in HIV cure strategies. *J Med Primatol*. 2022;51(5). doi:10.1111/jmp.12613

30. Tesina P, Cermáková K, Horejší M, et al. Multiple cellular proteins interact with LEDGF/p75 through a conserved unstructured consensus motif. *Nat Commun*. 2015;6. doi:10.1038/ncomms8968

31. Gérard A, Ségéral E, Naughtin M, et al. The Integrase Cofactor LEDGF/p75 Associates with Iws1 and Spt6 for Postintegration Silencing of HIV-1 Gene Expression in Latently Infected Cells. *Cell Host Microbe*. 2015;17(1). doi:10.1016/j.chom.2014.12.002

32. Cermakova K, Demeulemeester J, Lux V, et al. A ubiquitous disordered protein interaction module orchestrates transcription elongation. *Science (1979)*. 2021;374(6571):1113-1121. doi:10.1126/science.abe2913

33. Krishnan V. Molecular Thermodynamics Using Nuclear Magnetic Resonance (NMR) Spectroscopy. *Inventions*. 2019;4(1):13. doi:10.3390/inventions4010013

34. Yoh SM, Cho H, Pickle L, Evans RM, Jones KA. The Spt6 SH2 domain binds Ser2-P RNAPII to direct Iws1-dependent mRNA splicing and export. *Genes Dev*. 2007;21(2):160-174. doi:10.1101/gad.1503107

35. Oqani RK, Lin T, Lee JE, Kang JW, Shin HY, Il Jin D. Iws1 and Spt6 Regulate Trimethylation of Histone H3 on Lysine 36 through Akt Signaling and are Essential for Mouse Embryonic Genome Activation. *Sci Rep*. 2019;9(1):3831. doi:10.1038/s41598-019-40358-3

36. Tesina P, Čermáková K, Hořejší M, et al. Multiple cellular proteins interact with LEDGF/p75 through a conserved unstructured consensus motif. *Nat Commun*. 2015;6(1):7968. doi:10.1038/ncomms8968

37. Murai MJ, Pollock J, He S, et al. The same site on the integrase-binding domain of lens epithelium–derived growth factor is a therapeutic target for MLL leukemia and HIV. *Blood*. 2014;124(25):3730-3737. doi:10.1182/blood-2014-01-550079

38. Li X, Song Y. Structure, function and inhibition of critical protein–protein interactions involving mixed lineage leukemia 1 and its fusion oncoproteins. *J Hematol Oncol*. 2021;14(1):56. doi:10.1186/s13045-021-01057-7

39. Rao RC, Dou Y. Hijacked in cancer: the KMT2 (MLL) family of methyltransferases. *Nat Rev Cancer*. 2015;15(6):334-346. doi:10.1038/nrc3929

40. Liu L, Oliveira NMM, Cheney KM, et al. A whole genome screen for HIV restriction factors. *Retrovirology*. 2011;8. doi:10.1186/1742-4690-8-94

41. Kenaston MW, Pham OH, Petit MJ, Shah PS. Transcriptomic profiling implicates PAF1 in both active and repressive immune regulatory networks. *BMC Genomics*. 2022;23(1):787. doi:10.1186/s12864-022-09013-6

42. Mueller CL, Jaehning JA. Ctr9, Rtf1, and Leo1 Are Components of the Paf1/RNA Polymerase II Complex. *Mol Cell Biol*. 2002;22(7):1971-1980. doi:10.1128/MCB.22.7.1971-1980.2002

43. Abdel-Mohsen M, Wang C, Strain MC, et al. Select host restriction factors are associated with HIV persistence during antiretroviral therapy. *AIDS*. 2015;29(4):411-420. doi:10.1097/QAD.0000000000000572

44. Liu L, Oliveira NM, Cheney KM, et al. A whole genome screen for HIV restriction factors. *Retrovirology*. 2011;8(1):94. doi:10.1186/1742-4690-8-94

45. Bedwell GJ, Jang S, Li W, Singh PK, Engelman AN. Rigrag: High-resolution mapping of genic targeting preferences during HIV-1 integration in vitro and in vivo. *Nucleic Acids Res*. 2021;49(13). doi:10.1093/nar/gkab514

46. Francis AC, Marin M, Singh PK, et al. HIV-1 replication complexes accumulate in nuclear speckles and integrate into speckle-associated genomic domains. *Nat Commun*. 2020;11(1). doi:10.1038/s41467-020-17256-8

47. Achuthan V, Perreira JM, Sowd GA, et al. Capsid-CPSF6 Interaction Licenses Nuclear HIV-1 Trafficking to Sites of Viral DNA Integration. *Cell Host Microbe*. 2018;24(3). doi:10.1016/j.chom.2018.08.002

48. Chin CR, Perreira JM, Savidis G, et al. Direct Visualization of HIV-1 Replication Intermediates Shows that Capsid and CPSF6 Modulate HIV-1 Intra-nuclear Invasion and Integration. *Cell Rep*. 2015;13(8). doi:10.1016/j.celrep.2015.10.036

49. Horizon Discovery. Dharmacon Edit-R CRISPR-Cas9 Gene Engineering with All-in-One Lentiviral Cas9 and sgRNA.

50. Anderson AM, Tyor WR, Mulligan MJ, Waldrop-Valverde D, Lennox JL, Letendre SL. Measurement of Human Immunodeficiency Virus p24 Antigen in Human Cerebrospinal Fluid With Digital Enzyme-Linked Immunosorbent Assay and Association With Decreased Neuropsychological Performance. *Clinical Infectious Diseases*. 2018;67(1):137-140. doi:10.1093/cid/ciy056

51. Carter M, Shieh J. Biochemical Assays and Intracellular Signaling. In: *Guide to Research Techniques in Neuroscience*. Elsevier; 2015:311-343. doi:10.1016/B978-0-12-800511-8.00015-0

52. Slatko BE, Gardner AF, Ausubel FM. Overview of Next-Generation Sequencing Technologies. *Curr Protoc Mol Biol*. 2018;122(1). doi:10.1002/cpmb.59

53. Bhaya D, Davison M, Barrangou R. CRISPR-Cas Systems in Bacteria and Archaea: Versatile Small RNAs for Adaptive Defense and Regulation. *Annu Rev Genet*. 2011;45(1):273-297. doi:10.1146/annurev-genet-110410-132430

54. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science (1979)*. 2012;337(6096):816-821. doi:10.1126/science.1225829

55. Deltcheva E, Chylinski K, Sharma CM, et al. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*. 2011;471(7340):602-607. doi:10.1038/nature09886

56. Mali P, Yang L, Esvelt KM, et al. RNA-Guided Human Genome Engineering via Cas9. *Science (1979)*. 2013;339(6121):823-826. doi:10.1126/science.1232033

57. Sampson TR, Weiss DS. Exploiting CRISPR/Cas9 as systems for biotechnology. *BioEssays*. 2014;36(1):34-38. doi:10.1002/bies.201300135

58. Yang PC, Mahmood T. Western blot: Technique, theory, and trouble shooting. *N Am J Med Sci*. 2012;4(9):429. doi:10.4103/1947-2714.100998

59. Hsu PD, Lander ES, Zhang F. Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell*. 2014;157(6):1262-1278. doi:10.1016/j.cell.2014.05.010

60. Goff SP. Host factors exploited by retroviruses. *Nat Rev Microbiol*. 2007;5(4):253-263. doi:10.1038/nrmicro1541

61. Wood A, Armour D. The Discovery of the CCR5 Receptor Antagonist, UK-427,857, A New Agent for the Treatment of HIV Infection and AIDS. In: ; 2005:239-271. doi:10.1016/S0079-6468(05)43007-6

62. Cisneros WJ, Cornish D, Hultquist JF. Application of CRISPR-Cas9 Gene Editing for HIV Host Factor Discovery and Validation. *Pathogens*. 2022;11(8):891. doi:10.3390/pathogens11080891