Genome scale identification of HIV Dependency factors across multiple HIV-1 strains

Vanessa R. Montoya

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2022

Reading Committee:

Michael Emerman, Chair

Jennifer Friesen (Lund)

Stephen Tapscott

Program Authorized to Offer Degree:

Molecular and Cellular Biology

University of Washington

University of Washington

Abstract

Genome scale identification of HIV Dependency factors across multiple HIV-1 strains

Vanessa R. Montoya

Chair of the Supervisory Committee:

Michael Emerman

Department of Microbiology

In this thesis, I developed a method called HIV-CRISPR screening at genome-wide scale, and then at a smaller, targeted scale, to uncover novel genes that may act as HIV-1 dependency factors. By using a screening technique which assesses enrichment or depletion of guide based on HIV-1 release from infected cells, I was able to assess the role of host factors across the entire life cycle of HIV in a T lymphocyte cell line. Host dependency factors for HIV-1 were uncovered across multiple strains, from different clades and with different co-receptor tropisms. Nearly all of the host dependency factors were validated in subsequent follow-up studies and there was a direct relationship between the scoring of a dependency factor in the screen and its effect on HIV replication in validation studies. Work from this thesis thus identifies at least several dozen novel HIV dependency factors across multiple pathways.

Table of Contents

List of Figures

List of Tables

Acknowledgements

There are so many incredibly generous and kind people in my life, who without them, this PhD would not have been possible. I am truly grateful to each and every one of them for their unending support, words of encouragement, and wonderful friendships.

First and foremost, I must thank all the scientific mentors I've had that led me here today. As a Hispanic woman from a small town in New Mexico, the path to becoming a PhD hasn't always been clear to me. Thank you to each of my STEM teachers before college and my STEM professors and mentors in college for setting the foundation and enriching my interest in science. Dr. Kathy Hanley, my undergrad mentor at New Mexico State University, was the first mentor to put a pipette in my hand, teach me about the wonders of virology, gave me a really cool project in which I got to dissect mosquitos and measure how dengue virus evolved, and ignited the scientific flame inside me. She also pushed me to push myself out of my comfort zone and into multiple summer internships and eventually a postbacc position with Dr. Sonja Best at the Rocky Mountain Laboratories in Montana. Sonja was an incredible mentor who gave me the flexibility to learn about myself and what I wanted to do in the next steps of my career, while embracing an exciting new project, new techniques, new viruses to study, and asking interesting scientific questions at the host-pathogen interface. Both Kathy and Sonja set strong examples of the type of mentorship I wanted for graduate school, taught me a lot about myself as a scientist, and set an example as two inspiring, successful women scientists. In both labs, I had awesome postdoc mentors, Dr. Christy Andrade at NMSU, and Dr. Abhilash Chiramel at RML, and I owe much of my growth and the way I approached a scientific problem prior to grad school to them, as well.

I feel extremely fortunate to have been trained during my PhD by my mentor, Michael Emerman. Michael is an extremely gifted mentor in the way he is able to adapt to each of his mentees' styles at different stages of their academic careers. Throughout my time in his lab, he always encouraged me, reminded me of the excitement to come when we faced daunting challenges in the project, and cultivated an environment of growth and optimism that kept me and others in the lab moving forward and learning through the ups and the downs of science. There have been points in the PhD journey that have felt insurmountable, though Michael has helped me to reframe things in my mind to instead embrace the unknown—because of "how exciting it is to learn something new." I never expected to earn a PhD during a pandemic, but while we all worked from home, Michael also helped each of us move our projects forward in the ways we could, while staying connected with the outside world with regular zoom coffee hours. He is a great example of a mentor that is passionate about the science and about mentoring and I hope to take each of these lessons forward with me and pass them along to my future mentees.

Michael has created an environment of collaboration within the lab scientifically, and also through the extracurriculars we all embraced – the annual lab hike, the pumpkin carving contest, and our yearly lab Halloween costume. The memories of us all planning out our costumes and

making our pumpkin contest entries each year are all very dear to me. I am extremely grateful to my lab mates, past and present, for all the laughs, the lab lunches, the stimulating scientific conversations, the helpful scientific advice, and above all, the wonderful friendships. Emily Hsieh, I feel very lucky to have gone through this journey with such an amazing scientist and friend from almost day one. It's been a true honor. Mollie and Suzanne, who are no longer in the lab but left a lasting impression, thank you for bringing lightheartedness and laughter into our bays and keeping us on our toes with the practical jokes and belly laughs. Special thanks to Mollie and Erin, for your friendships and mentorship in the early stages of grad school and in the lab. Ferdi, thank you for setting the framework for genome-wide screening and your mentorship. Nick, Terry, Carley, Caroline, Jessie, Joy, and Rechel, it has been a great fun learning and growing with friends. Graduate school is a rollercoaster, but having lab mates to laugh with, commiserate with, and grow with has been a great joy. Thank you immensely to each of you. Huge thanks to Trine Ready, my partner in crime, who helped me wrap up the validation portion of the project. It's been a great privilege mentoring you, learning about you, your friendship, and teaching you-- I know you'll do great things, whatever you decide to do next. You all inspire me and I look forward to seeing what's next for you all!!

I also want to specially thank Dr. Molly OhAinle, an amazing mentor to me and many others in the Emerman lab while she was a senior staff scientist. Molly now has her own lab at UC Berkeley, but for all of my graduate career, she was another strong mentor for me. I'm forever grateful for the conversations, about science and diversity in STEM, the helpful feedback on my project throughout the years, the contagious excitement about the science, and for being a role model. Thank you.

Many thanks to Team Screen: Molly, Emily, Terry, Carley, and Daryl Humes and Hannah Itell of the Overbaugh lab. Thank you for the extremely helpful advice, camaraderie as we learned new things together, and for being a great squad to collaborate with.

I'm grateful for the incredible community I've found in my graduate program and at the Fred Hutch. Thank you to the Nina and Celeste, the MCB office, and the graduate student office at the Fred Hutch for all the logistical and personal support they've given along the way. I especially would like to thank Maia Low, Andrea Brocato, and Denise Barnes for going the extra mile for us and for those who dedicated their time to creating a more equitable and diverse environment in the MCB program and at UW. I will also be forever grateful for the graduate student community I've built with friends in my 2017 cohort, as well as friends in other cohorts and programs at UW.

Thank you to my committee—Julie Overbaugh, Aakanksha Singhvi, Stephen Tapscott, and Jennifer Lund—for your guidance and ideas for my project, encouragement, and your roles in

my growth as a scientist. Thanks especially to Stephen and Jenny, the members of my reading committee, for your helpful comments and feedback on this thesis.

I must also thank my ARCS benefactors, Kathy Fraser and Sally and David Wright, for giving me extra funding my first three years of graduate school for living expenses. I am extremely grateful for their sponsorship, which allowed me to focus more on my success in graduate school and a little bit less on the financial strain. Thank you for believing in me.

Thank you to my fellow SACNISTAs for building an amazing community with me, bringing a bit of my culture back into my life while living so far from home, doing K-12 STEM outreach, professional development workshops, and building gingerbread houses for social events. Participating in something that felt bigger than me and my project, such as building each other up and trying to bring up the next generation of scientists, gave me an extra bit of drive to keep going. Special thanks to my fellow officers, Arianna Gomez, Madeleine Hart, Brittany Ulloa, Monica Cesinger, Feli Ruiz, Cera Hassinan, Robin Kai, and Carlos Segura, as well as Anthony Salazar, David Kimelman, and Billie Swalla, our staff and faculty advisors, for maintaining this space for the advancement of chicanxs/Hispanics and Native Americans in science.

To my new friends in Seattle and my old friends from Montana and New Mexico, I am forever grateful for the companionship, support, and advice. Thank you to my friends in Seattle for teaching me so much in their disciplines, the practice talks, the celebrations, the commiseration, the concerts, the laughs, the adventures, the hikes, the meals, the friendship, and for joining me on this journey. To my friends afar, thank you for all the phone calls, the meetups, the picking up where we left off, the pep talks, and the lifelong friendships. Thank you all.

To my amazing family, thank you for being my rock, my biggest fans, my advocate, and for kicking me in the butt when I need to be reminded what my goals are. Huge thanks to my parents for instilling in me from the beginning that I am capable of doing whatever I put my mind to. My parents have always supported me in whatever form they could, whether it was a pep talk when things weren't working, to reminding me why I came to grad school. I could not have asked for more amazing parents and I'm forever grateful for them. Thank you to my brother for always having my back, the endless jokes and laughter, and for bringing me out of the stress for some fun. Thank you to my grandparents who always encouraged me to keep going and always gave a warm place to come home for the holidays to be with the whole familia. Thank you to my aunts, uncles, and cousins have all also been my biggest fans. My Nino, Nina, and Auntie all set the example for working hard in college and building a good life with your degree. The check-ins, words of encouragement, and interest in my work and science always meant a lot. Thank you all for the unconditional love.

Finally, I want to thank my partner Andrew. Andrew came into my life about two years ago and helped me immensely with the final stretches of grad school. He has always been super supportive of my work, helped me stay focused when I needed to write, fed me when I had no time to cook, ran many computational analyses for my project, helped me build figures in Tableau, and in general has been an amazing partner. Thank you for building me up and also going on many adventures with me, nearby—at the Skagit valley tulip fields, apple picking, a tree house birthday, and star gazing, and far away—hiking in Hawaii and meeting each other's families and learning each other's traditions. Thank you for all of your love and encouragement.

Dedication

This thesis work is dedicated to my parents.
Thank you for the sacrifices you made to help me get here today.

Chapter 1: Introduction

General Introduction

Viruses, as intracellular obligate parasites, have evolved to co-opt the cellular functions of the host cells they target. Many RNA virus genomes consist of few genes, requiring the ability to adapt to different host cell environments and host factors in order to successfully proceed through each step in the viral life cycle. One of the many RNA viruses that require cellular machinery to infect host cells is Human Immunodeficiency Virus (HIV), the causative agent of Acquired Immunodeficiency Disease Syndrome (AIDS). Although the development of antiretroviral therapy has vastly improved the outcomes of many people living with HIV, we do not yet have an FDA approved vaccine. HIV continues to place a large global burden with approximately 38.4 million people living with HIV in 2021, with an estimated 1.5 million new cases of HIV (UNAIDS 2021). One approach to counteract the ability of the virus to use our own cells is to map which cellular factors are required for viral infection for any of the steps of the viral life cycle, as well as which factors may aid in immune evasion or viral transmission. This knowledge can help us counteract the virus, by informing us of potential host factors that can be inhibited by The goal of this thesis was to use high-throughput genome-wide screening to obtain a comprehensive view of the host factors that are required for HIV to replicate in a human T cell line.

Impacts of RNA viruses on the human population

Viruses are intracellular obligate parasites, meaning they depend on the host they infect to successfully proliferate and lead to progeny production and spread. Viruses are ubiquitous throughout nature, infecting insects, animals, humans, and even other microorganisms such as bacteria. Of RNA viruses alone, there are at least 214 human-infecting viruses that have been discovered over the last ~120 years [1]. RNA viruses have had tremendous impact on the human population and have the potential to cause immeasurable tragedy as seen throughout history with 1918 influenza virus, Zika virus, Ebola virus, and most recently SARS coronavirus 2.

Human Immunodeficiency virus (HIV) has caused one of the most devastating pandemics of human history, killing approximately 40.1 million people since it was first recognized as the novel etiological agent causing Acquired Immunodeficiency Syndrome (AIDS) in 1981 [2-4]. Although HIV-1 was first discovered in the 1980s, HIV-1 DNA has been amplified from a 1959 African plasma sample found to be HIV-1 seropositive [5] and advancements in phylogenetic and molecular clock analyses date the most recent common ancestor of the pre-AIDS HIV-1 Group M (the principle group of the AIDS pandemic) between 1884-1924 [6, 7]. To date (by the end of 2021), approximately 38.4 million people [33.9-43.8 million] people are living with HIV, with 1.5 million of those cases being newly acquired in 2021 (WHO 2021). HIV continues to be an active public health threat as although there are effective prophylactic treatments and anti-retroviral therapies, there is still no FDA approved vaccine and no facile cure to HIV infection. Advancements in creation of new therapies, and potentially downstream a cure, rely on an understanding of the basic biology of HIV-1 infection in T cells, the primary target and reservoir of HIV.

HIV life cycle relies on host dependency factors

The RNA genome of HIV is approximately 9 kb, consisting of the genetic material needed to synthesize 15 proteins (Figure 1). Compared to the mammalian hosts that HIV infects, this is surprisingly small, though this is still enough to replicate and assemble into new virions. However, the HIV proteins themselves are not sufficient for the virus to complete its life cycle, but rather it relies on the ability of HIV to hijack multiple host factors within the infected cell. Host factors that viruses are able to hijack for the benefit of viral replication are called "dependency factors". Dependency factors are important for many viruses, especially RNA viruses that have limited genetic material.

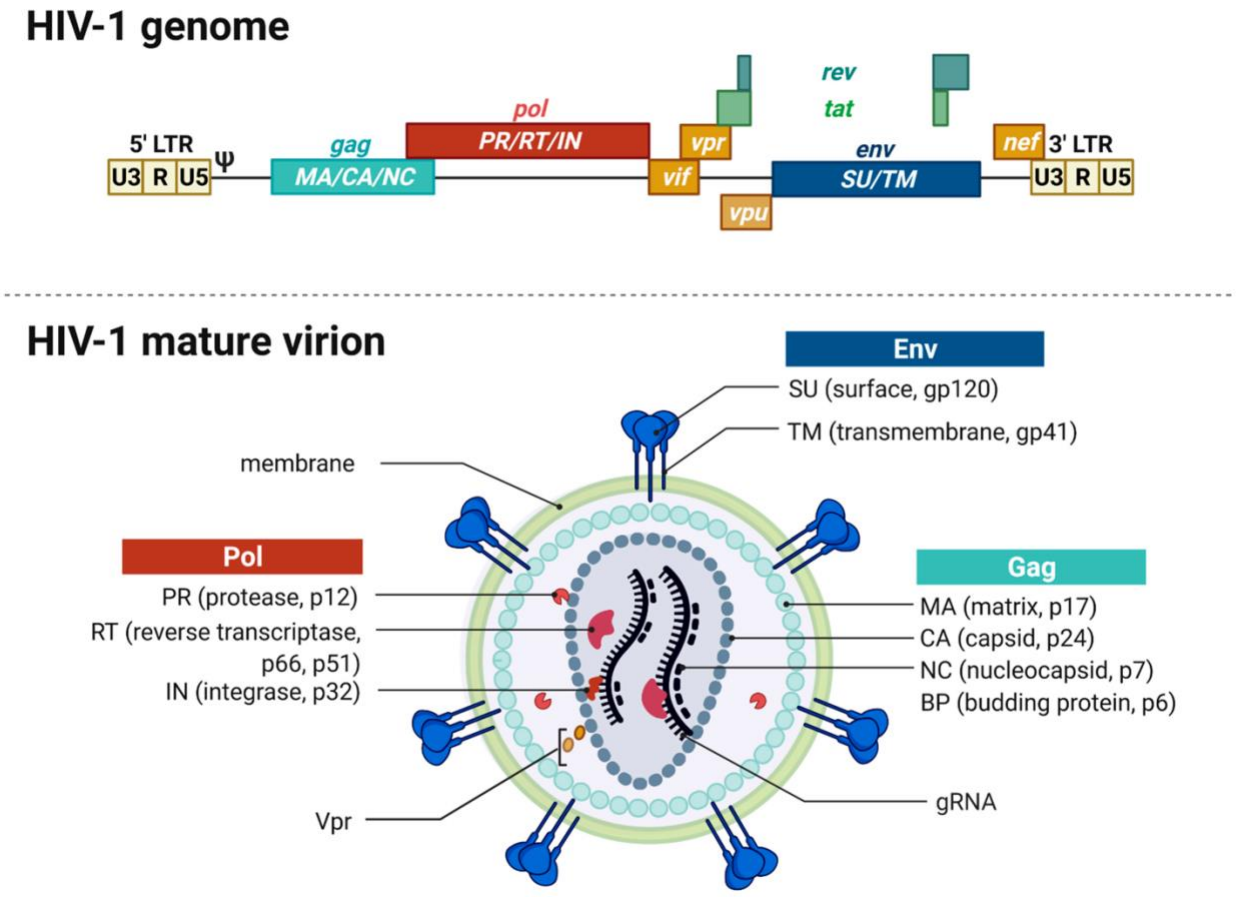


Figure 1. HIV-1 genome and virion.

This figure is from Ref. [8], used under Creative Commons CC-BY 4.0 license.

Importantly, HIV also must use many cellular host factors to complete the viral life cycle (Figure 2; [8]). The first described example of a host dependency factor for HIV-1 was the discovery of its receptor, CD4 [9, 10]. HIV-1 tropism is also dependent on the co-receptors C-X-C motif chemokine receptor type 4 (CXCR4) or chemokine receptor type 5 (CCR5), depending on the strain of HIV-1. The initial step of HIV-1 entry is when the Env protein, comprised of gp120 and gp41 subunits, binds to the cell surface, often described as via α4β7 integrin[11, 12], negatively charged heparin sulfate proteoglycans [13], or pattern recognition receptors like DC-SIGN [14]. This then increases the proximity of Env with CD4, the HIV-1 receptor. HIV-1 binding to CD4 induces a conformational change of Env that increases its affinity for the co-receptors, which in turn reveals the hydrophobic gp41 fusion peptide causing viral-cellular membrane fusion [15].

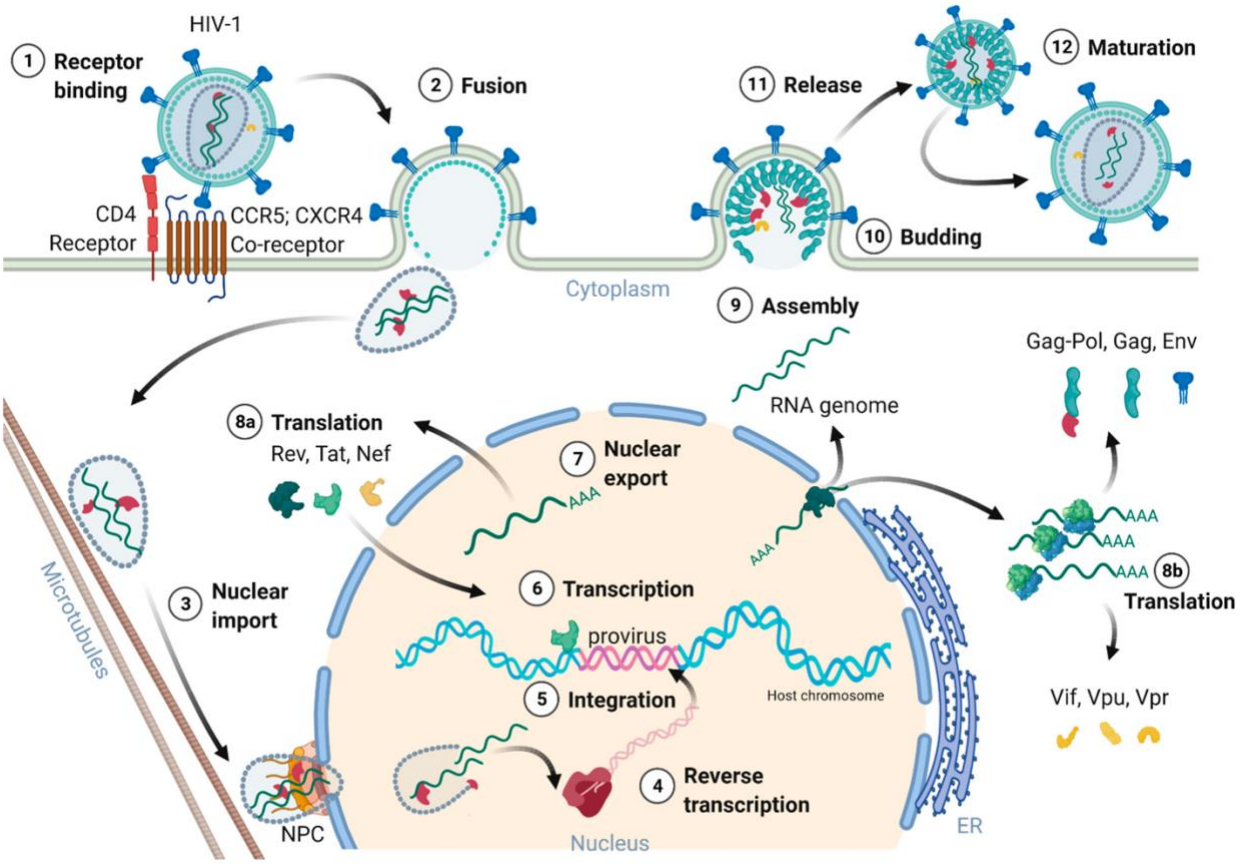


Figure 2. Schematic of the HIV-1 replication cycle.

1) HIV-1 virion binds to the CD4 leading to a conformational change allowing for binding to the co-receptors CCR5 or CXCR4. 2) The virus membrane fuses with the cellular membrane and the capsid transits via the microtubules to the nuclear envelope where 3) it utilizes host nucleoproteins, TNPO3 and CPSF6 to import into the nucleus. 4) The viral reverse transcriptase enzyme completes reverse transcription of RNA to DNA once imported in the nucleus and is now able to be 5) integrated into actively transcribed gene bodies in the cellular genome, as facilitated by the pre-integration complex and host proteins CPSF6 and LEDGF/p75. 6) The HIV-1 genome is then transcribed by host machinery to make transcripts that get multiply-, singly-, or unspliced that serve as new HIV-1 genomes. 8a) Multiply-spliced transcripts are exported and translated to make accessory proteins Rev, Tat, and Nef. Viral protein Rev and host protein CRM1 transport back into the nucleus, bind the remaining unspliced or singly-spliced transcripts that still contain a stem-loop region called the Rev Response Element, and export out of the nucleus to 8b) translate Gag-Pol, Gag, Env, Vif, Vpu, and Vpr. Two full length genomes are dimerized and packaged into nascent virions recruited to the viral membrane by the Gag precursor. At the plasma membrane, HIV assembles near lipid rafts. The Env polyprotein is cleaved by a cellular furin-like protease into surface proteins gp120-SU and gp41-TM. 9-10) Gag assembles at the membrane, Env

trimers are held at the assembly site, and Gag recruits the ESCRT proteins to achieve 11) membrane scission and release of the immature virion. 12) The virion matures as viral protease PR cleaves the Gag and Gag-Pol precursors into the enzymes PR, RT, IN, and the structural proteins. This figure is from Ref. [8], used under Creative Commons CC-BY 4.0 license.

Once inside the cell, the viral capsid/ core (containing capsid proteins, viral RNA genome, reverse transcriptase (RT), integrase (IN), and others proteins required for infection), also called the pre-integration complex, must translocate from the plasma membrane to the nucleus. There is evidence that Lck, a lymphoid specific Src Kinase important for T cell signaling downstream of CD4 engagement, plays a role in F-actin rearrangement during T cell activation to assist in intracellular transport [16]. Transport of the pre-integration complex (PIC) to from the cell surface, to the nuclear pore is also dependent on the microtubules, motor protein dynein, and dynein adaptor protein bicaudal D2 (BICD2) ([17-19].

Prior to the capsid disassembly stage, it is thought that the capsid acts as a protective barrier for the viral complexes against the immune system in addition to translocation to and into the nucleus [20]. Although initially controversial, newer data now shows that fully or nearly intact viral cores enter the nucleus through nuclear pores complexes (NPC), which are composed of approximately 30 nucleoporin proteins [21-23]. The conical shaped viral capsids dock to the nucleoporin, NUP358/RanBP2, on the cytoplasmic face of the NPCs [24, 25]. The capsid-containing pre-integration complex then inserts and moves through the NPC, interacting with the NUP62 complex nucleoporins. While it was previously assumed that the diameter of the NPC was smaller than the diameter of the viral capsid, it has recently been shown the NPC size *in cellulo* is large enough for intact capsids to enter the nucleus by imaging disrupted and empty fragments of capsid within the nucleus [26]. Thus, recent studies supports the hypothesis that capsid uncoating occurs after entering the nucleus. While the nuclear entry and uncoating

process is largely capsid-driven, additional host factors are important for both HIV-1 entry into the nucleus and subsequent integration specificity. These include CPSF6 (cleavage and polyadenylation specific factor 6) which interacts with the HIV-1 capsid and is important for both nuclear import of intact capsids [22], as well as a role for integration (see integration paragraph below), and TNPO3, a karyopherin protein that mediates nuclear entry of CPSF6 bound to HIV-1 capsid [27-29]

Reverse transcription is an essential step in the life cycle to convert the single-stranded viral RNA to double-stranded DNA which is then integrated into the host genome. Up until recently, it was also thought that reverse transcription (RT) begins in and is completed within the cytoplasm. However, there is now evidence that RT begins in the cytoplasm when dNTPs enter through the R18 pore of the capsid [30] but finishes in the nucleus [21, 22, 31] (reviewed in [32]) .The availability of dNTPs, along with the template (viral genomic RNA), and the primer (host tRNA that was within the packaged virion), are all limiting factors for reverse transcription to occur. The molecular RT process [33] occurs as follows: a primer derived from a host tRNA binds the primer binding site on the viral RNA just 3' to the long terminal repeat (LTR) and degrades the RNA template. The extended minus strand DNA then transfers to the 3'end of the viral RNA, allowing minus strand RT to continue. The polypurine tract encoded 5' to the LTR (ppt) resists RNase H cleavage and is the primer for subsequent plus strand DNA initiation. Plus strand DNA synthesis continues until the first 18 nucleotides are also synthesized. This allows RNase H to cleave and remove the tRNA primer, leading to the second plus strand transfer, and extension of both the plus and minus strands for a double-stranded linear viral DNA which can be integrated into the host genome.

Aside from the host derived tRNAs, there are other host factors that assist with the process of reverse transcription. SIP1/Gemin2 is a host factor that normally acts in the survival motor neuron complex that mediates assembly of spliceosomal snRNPs and snoRNPs, however in the context of HIV-1, has been shown to stabilize integrase multimerization which then enhances RT assembly on viral RNA [34]. Cyclin dependent kinase 2 (Cdk2) is another reported HIV dependency factor that phosphorylates RT at a conserved threonine residue, stabilizing and enhancing reverse transcriptase activity [35].

After completion of reverse transcription, this viral cDNA, now called the provirus, is integrated into the host cell chromosome. While the enzyme that completes this reaction, Integrase, is virally encoded, there are host proteins that determine the broad specificity of where integration occurs. Host protein CPSF6 (cleavage and polyadenylation specific factor 6) is understood to be a primary determinant for integration site selection in gene-rich euchromatin regions in the nucleus [36, 37]. CPSF6-capsid interactions allow for CPSF6 to guide the PIC beyond peripheral heterochromatin to the gene-dense nuclear interior[36]. Another host protein, LEDGF/p75, drives the positional targeting within gene bodies [38-40].

This integration step into the cellular genome is largely why we have yet to find a cure for HIV as the integration is irreversible, and a minority of the integrated proviruses become silent in a stage called “latency.” During both release from latency and during productive viral infection, many host factors are recruited to the HIV LTR and used to initiate and complete transcription of viral mRNAs and viral genomes. For example, the HIV-1 LTR as binding sites for NFkB, Sp1, Ap-1 and NFAT. The preferential use of each transcription factor varies by cell type. Additionally, many members of the Mediator complex have been shown to be important for transcriptional activation for HIV-1 transcription. Eighteen mediator genes have been reported as

dependency factors for HIV through siRNA knockdown and mechanistic studies [41-44] implying a role for the complex as a whole. These proteins and RNA polymerase II make up the preinitiation complex on the 5' long terminal repeat (LTR) promoter.

Transcriptional elongation is a particularly limiting step for transcription of HIV genomes and HIV mRNAs. In the absence of the viral protein Tat, HIV transcription is able to initiate efficiently and short non-polyadenylated transcripts (~60 nucleotides long) are made as RNA pol II clears the promoter; however, transcriptional elongation is inhibited [45]. The HIV transactivation response element (TAR) RNA stem loop is responsible for the polymerase pausing as negative transcription elongation factors (N-TEFs) including NELF/DSIF are recruited [46]. To overcome this block, Tat binds the 5' bulge of the TAR with the help of CyclinT1 [47]. pTEFb, composed of Cyclin T and CDK9, is therefore recruited to the TAR RNA of the viral transcript via Tat, and will phosphorylate Ser2 on the CTD of the RNA pol II tail [48] and RD of the DSIF complex [46], leading to the dissociation of negative factors and elongation of viral transcripts.

Transcription, mRNA splicing/ processing, and nuclear export is a dynamic process. Normally, splicing occurs co-transcriptionally or immediately after transcription, leading to 3' end processing (cleavage and addition of the poly A tail) that is essential for the stability of an mRNA transcript. However, because HIV uses alternative splicing to create multi-, singly-, or un-spliced mRNA from the same full-length transcripts encoding for different HIV proteins or the full-length genome, splicing is indirectly controlled by a differential RNA export that is mediated by a viral protein Rev and host cellular proteins. Fully spliced transcripts are able to be exported from the nucleus before complete splicing due to Rev protein recruitment of host protein CRM1 to a viral sequence called the Rev response element (RRE). This leads to CRM1-

mediated mRNA export for singly-spliced or unspliced HIV mRNA transcripts that encode for virion proteins and the viral genome. When Rev levels are high enough, Rev is able to co-opt CRM1 function, a protein used for nuclear-cytoplasmic transit, and Rev-CRM1 binds to the RRE that is preserved on singly- or unspliced HIV mRNA transcripts, but not multiply-spliced transcripts and facilitates the RanGTP-dependent nuclear export [49].

As fully-spliced transcripts are able to be exported without CRM1-Rev, there is an initial buildup of viral proteins Nef, Tat, and Rev. Similar to cellular mRNA transcripts, the HIV mRNA transcripts have also been capped with a 7-methylguanosine (m^7G) and polyadenylated at the 3' end. To synthesize viral proteins, HIV-1 co-opts the translational machinery by recruiting the 40S ribosome to the 5'UTR and undergoes Cap binding complex (CBP)-mediated pioneer round of translation [50]. To ensure translation is able to proceed through highly structured motifs in the HIV transcript, including the TAR element, HIV-1 requires the host RNA helicase DDX3 to directly bind to the 5'UTR and associate with the eIF4F complex via eIF4G-PABP to promote translation initiation of HIV transcripts [51]. For the first 24-48 hours of replication, HIV-1 uses cap-dependent translation initiation mechanisms [52] and after uses a combination of cap-dependent and IRES-dependent (cap-independent) translation [50, 52, 53].

Once the incompletely spliced or full-length mRNAs are exported, there is a shift to translation of structural proteins/ proteins important for assembly of new virions. The full length RNA acts as both the template for Gag and GagPol polyproteins or the genomes that are packaged into nascent virions [54]. Two full-length genomes are dimerized and packaged into these virions, which allows for strand transfer during reverse transcription and recombination that leads to greater genetic diversity of viral progeny [55]. Dimerization and packaging is highly dependent on the 5' leader (5'UTR through early parts of the gag gene), which contains the psi

packaging signal. Gag encodes matrix (MA), capsid (CA), and nucleoprotein (NC), each of which is highly important for assembly and maturation. The CA, MA, and p2 domains of the precursor Gag (pr55-Gag) are important for gag multimerization, the MA domain is required for gag recruitment to the cellular membrane. The NC domain is important for selectively binding the psi packaging signal and the Dimer Initiation Signal and oligomerization of the pr55-gag. KIF-4 is yet another hijacked protein used for HIV-1, this time for trafficking Gag to the cell membrane [56]. KIF-4 is a host cellular motor protein that has been shown to bind pr55-Gag and a matrix-capsid intermediate (Pr42), but not other processed Gag products [56]. In addition, Staufen1 is a host protein that has been shown to regulate virion assembly by preferentially associating with precursor Gag and may regulate selection of genomic viral RNA [57-59]. HIV capsid stability and virion maturation is dependent on host metabolite inositol hexakisphosphate (IP6) [60, 61].

At the plasma membrane, HIV assembles near lipid rafts. The Env polyprotein gp160 is cleaved by a cellular furin-like protease into surface proteins that make up the virion's lipid bilayer (along with host proteins) gp120-SU and gp41-TM. As Gag assembles at the membrane, Env trimers are held at the assembly site by the Gag MA domain. The Gag multimer continues to assemble on the plasma membrane and recruits the ESCRT proteins to assist in budding of the growing spherical particle [8, 62]. The HIV protein p6 recruits ESCRT-1 proteins EXRT-1 and ALIX to the membrane. As the sphere continues to grow, Gag gets ubiquitinated and binds the ESCRT-II proteins PTAP-TSG101 or LYPX$_n$L-ALIX. Eventually, the ESCRT-III subunits CHMP2A/B, CHMP3, and CHMP4A-C and ATPases VPS4A/B are recruited and facilitate the membrane scission and ultimately release of the immature viral particle from the membrane [62]. The viral particle matures as the viral protease PR, sequentially cleaves the Gag and Gag/Pol

precursors into the enzymes PR, RT, IN and structural proteins p17-MA, p24-CA, and p7-NC [8, 63, 64]. The mature virion is now infectious and can infect a new target cell [64].

Screening to identify dependency factors in a high-throughput fashion (RNA interference screens)

Historically, dependency factors for HIV have been identified and validated one-by-one, using a series of time and resource consuming assays to investigate a single host factor that may or may not play a role in HIV-1 biology. However, in more recent years, high throughput methods have been developed in attempts to more comprehensively identify host pathways important for HIV replication. The first of these to be applied was that of arrayed screens using small interfering RNA (siRNA) to knock-down expression of host proteins followed by HIV infection.

RNA interference is a biological mechanism first discovered in plants [65] and further mechanistically explored and exploited in C. elegans, an animal model widely used to study genetics following (siRNA) knockdown [66]. RNAi is a biological mechanism in which double stranded RNA (micro RNA or small interfering RNA) molecules (that are specific to a certain degree of mismatches) are used by cellular machinery to find matching mRNA strands for inhibition of expression, either through stalling translation or cutting of the strand by the Argonaute complex.

In the late 2000's, several genome-wide siRNA screens sought to identify novel dependency factors for HIV infection [42-44]. In 2008, Brass et al. used a two-part screen to survey early vs. late acting host factors in HeLa-derived cells expressing endogenous CXCR4, transgenic CD4, and CCR5, using a readout of an integrated β -galactosidase (β -gal) reporter gene whose

expression was dependent on the HIV-1 Tat protein. To assess early acting genes, this group infected siRNA knockdown cells with the HIV-1 IIIB strain and measured p24 (gag) 48 hours post infection. To interrogate late-acting factors, fresh reporter cells were incubated with the part one viral supernatants for 24 hours and Tat-dependent reporter β-gal expression was quantified. Of 21,121 genes targeted by 4 siRNAs/gene, 386 of these (1.8%) met their significance criteria. Importantly, this screen identified novel factors (described above in the HIV life cycle section) in the nuclear pore complex (Nup85, Nup107, Nup133, and Nup160) and components of the Mediator complex (Med4, Med6, Med7, Med14, and Med28). They also identified and validated genes encoding proteins involved in glycosylation, autophagy, and notably TNPO3, described above to be important for mediating nuclear entry of CPSF6 and the HIV capsid.

As many of the factors identified in Brass et al. were implicated for late infection, rather than early, Konig et al. sought to identify early acting dependency factors in 293T cells. Using a single-cycle HIV-1-VSVg psuedotyped reporter virus (pNL43R+E- luc [VSVg]), they infected siRNA knockdown cells in an arrayed fashion using a genome-wide library targeting approximately 20,000 genes (6 siRNAs/gene). 24 hours post infection, infection was monitored by luciferase activity. This early analysis was intended to focus the results towards early steps of the life cycle, e.g. entry, RT, integration. Using mRNA expression profiles, protein-protein interactions, and gene ontology, 294 genes matched their selection criteria (1.47%). Again, in this screen, nuclear import factors were identified, including RANBP2, TNPO3, and nucleoporins NUP153 and NUP98. Genes involved in facilitating reverse transcription, integration, transcription/splicing, and the ubiquitin proteasome system were also identified.

Zhou et al. took a third approach for genome-wide siRNA screening for HIV dependency factors. Here, they infected siRNA knockdown HeLa P4/P5 cells (which express CD4 and

contain an integrated HIV LTR-driven β -galactosidase reporter [67]) with the HIV-1 HBX2 strain, and measured infection either 24 hours later (for early stages of infection) or 48 or 96 hours later (for all stages of infection) [43]. Of 19,079 genes targeted by 3 siRNAs/gene, 390 genes met the cutoff determined for their first screen. They then screened *in silico* for identification of which hits were expressed in multiple HIV relevant tissues (i.e. activated T cells or macrophages). 232 genes were then rescreened and validated to have inhibited infection (1.58%). Of these, again members of the mediator complex (MED4, MED7, and MED28) were identified, though additionally, MED6, MED8, MED11, MED17, MED19, MED20, MED26, MED27, and MED31. Additionally, they also showed an enrichment for factors involved in adipokine signaling, and oxidative phosphorylation.

To screen cellular proteins that affect HIV replication after sustained knockdown without causing cytotoxic cellular defects, Yeung et al. took a different approach for dependency factor screening. Using an shRNA library (targeting 54,509 transcripts / 3 to 5 shRNAs per transcript) cloned into a lentiviral vector, they transduced a more biologically relevant T lymphocytic cell line, Jurkat cells, to knockdown a single gene per cell. After puromycin selection to kill off cells without an integrated shRNA, they infected with HIV-1 NL4-3 and following infection-induced cell lysis, small RNAs were isolated from surviving cells, amplified, and analyzed using microarray hybridization. 252 genes fit their criteria of significance. Of these genes, there was no overlap with the Brass et al. screen, three genes (EPAS1, IBTK, and NUP98) overlapped with Konig et al., and three genes (NFKB1, RPL3, and RAB11A) overlapped with Zhou et al.. Additionally, they further validated the mRNAs STXBP2, NCOA3, PRDM2, and EXOSC5. This screening approach addressed some of the limitations of the siRNA screens, such as using more relevant cells, Jurkats, and using stable knockdowns.

While these screens identified genes important for nuclear entry, integration, transcription, and nuclear export, the problem with this series of screens is that there was very poor overlap across the three published siRNA screens, with only 34 genes out of 842 genes identified that overlapped between 2 of 3 screens, and only 3 genes overlapped between all 3 screens, including MED6, MED7, and RELA [68]. There are several complications that could explain the lack of overlap. Incomplete and poor knock-down is a problem with both siRNA and shRNA techniques, as well as known off-target effects. In addition, the screens were all done in non-relevant cells, and none of them used assays that examined the entire lifecycle of HIV. Additionally, some of the siRNA screens were complicated by using pseudotyped virus rather than authentic HIV, and using a readout that depends on Tat related transactivation which would bias the readout towards VSVg entry or Tat/ transcription related factors.

Screening to identify dependency factors in a high-throughput fashion (CRISPR/Cas9 screens)

More recent advancements in screening technology led us to CRISPR/Cas9 gene editing. This technology stems from the discovery that bacteria and archaea have both evolved a line of defense which uses an RNA-mediated adaptive immune response called Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/Cas (CRISPR-associated) to defend against viruses and plasmids [69]. Similar to RNAi, short RNAs are used to detect a specific sequence of invading nucleic acids, but in this case the DNA locus is cleaved in a permanent fashion rather than acting on mRNA or translation levels directly.

The advantage of CRISPR/Cas9 over the siRNA or shRNA screens are that CRISPR/Cas9 leads to a knockout rather than a knockdown, so that it could identify host dependency factors that are needed at only low levels. In addition, the selection of guides for CRISPR/Cas9 is

generally more reliable than that of siRNA/shRNA leading to fewer false negative hits and the off-target effects of CRISPR/Cas9 is generally lower than that of siRNA/shRNA leading to fewer false positive hits.

The first genome-wide CRISPR screen performed to identify HIV dependency factors used a CD4+ T cell line model engineered by stably transducing CCRF-CEM cells with CCR5-hygR and HIV-1 LTR-GFP (referred to as GXR). These cells were transduced with a lentiviral construct containing Cas9 and following 1 week of antibiotic selection, single cells were sorted and selected for a clonal line with high receptor / co-receptor expression, low basal GFP expression, and high infectivity (now referred to as GXR-Cas9 cells). [70]. These cells were transduced with a lentiviral sgRNA library, targeting 18,543 genes and selected for 1 week. The pooled mutant knockouts were then serially infected with CCR5-tropic HIV-1 strain JR-CSF, and the surviving mutant population was compared to the uninfected cell sgRNA abundance. Of 18,543 genes surveyed in this pooled fashion, 5 hits were identified and experimentally validated (0.03%), including CD4, CCR5, TPST2, SLC35B2, and ALCAM. This study did identify host dependency factors, however, as these genes involve binding and entry (CD4 and CCR5), modulate CCR5 sulfonation (SLC35B2 and TPST2), and mediates cell aggregation (ALCAM), may have been biased towards proteins necessary for virus attachment, entry, and/or syncytia formation. Furthermore, they failed to identify many of the known HIV dependency factors that had been previously well-characterized. It is not clear why this screen was so biased towards entry factors, but that could have been an artifact of the selection by flow cytometry for GFP expression which would have been more influenced by the initial events.

As an alternative to genetic screens, Jager et al. sought to identify cellular proteins that interact with all 18 HIV proteins, in a high-throughput approach using affinity

tagging/purification mass spectrometry (AP-MS), in both 293Ts and Jurkat cells [71]. They either transiently transfected 293T or stably transfected tetracycline inducible Jurkat T cells with the affinity tagged HIV protein clones, harvested 48h or 16h later, respectively, immunoprecipitated, and analyzed by mass spectrometry. Using a scoring system specially designed to identify AP-MS derived host-pathogen protein-protein interactions, termed MiST, they identified 435 proteins that interact with HIV with approximately 40 percent of these identified in both cell types. Additionally, 55 of these hits overlapped with the previously performed RNAi screens described above. Several hits identified in this protein-interaction screen were subsequently validated experimentally as being important for HIV infection including CBFbeta[72] and UBE2O[73] as dependency factors, and AMBRA1 as a restriction factor[74].

The entire list of HIV interactions was then used in a subsequent large-scale validation experiment using CRISPR/Cas9 knockouts in primary CD4+ T cells with a readout of HIV spreading infection [75] . In this work, in vitro-assembled CRISPR-Cas9 RNPs were electroporated into primary T cells in a 96-well arrayed format to target 430 genes (3 crRNA per gene) and activated and expanded each polyclonal knockout or control well. After 6 days, genomic DNA was extracted for deep sequencing of the target site (knockout efficiency analysis) and the following day, the expanded cells were infected in triplicate with and HIV strain, NL4-3, that was engineered to express GFP as well. Cells were sampled by flow cytometry at days 3, 5, and 7 to determine a role in HIV biology at early or late stages of infection. Of the 430 genes that encoded proteins originally identified as interacting with HIV proteins, 90 genes, including positive controls CD4, CXCR4, LEDGF, and HEXIM1 were implicated, of which 40 of these were significant across all donors, though 50 had some donor-to-donor variation or dependency.

Of the 86 novel factors, 40 had been previously been linked to HIV biology in the literature (described above). Upon a second screen, they showed that 32 of the 62 previously called dependency factors reproducibly decreased in infection across three blood donors (note: 11 of the 62 (15% of the 86) gene targets had viability defects). 55% of the genes surveyed showed an effect on HIV biology, either as a dependency factor or restriction factor, and some of these may act through multi-cycle infections only. One example reported is AMBRA1, a substrate previously identified in the AP-MS screen [71] that had been described as a restriction factor [74], though did not recapitulate in primary cells when subsequently retested in the subsequent study [75]. Although this approach has many strengths, including the use of relevant cells (primary CD4+ T cells), confirming knock-out of each gene by deep sequencing, and the ability to pair infection with knockout efficiency, there are also drawbacks to consider. The CRISPR library designed and utilized in this study was based only on factors previously identified in AP-MS studies, so this library would not include guides targeting genes that encode for proteins that are in more fragile complexes which would not be detected using immunoprecipitation, nor would it include HIV dependency factors that did not directly interact at the protein-protein level. While the use of primary T cells in the assays was admirable since it is more relevant than cell lines for HIV infection, the use of primary T cells introduced a large amount of variability into the results. For example, donor to donor differences in variation in knockout efficiency, susceptibility to infection, and magnitude of phenotype following knockout caused variation in results. Even with the steps to account for and normalize for these factors, there were 30% of the initial candidates that failed to recapitulate the initial phenotype in a follow-up validation experiment [75]. Nevertheless, this approach is one of the most advanced high-throughput screening approach that couples proteomics to genetics in more relevant cells.

Fu et al. takes yet another approach, leveraging what has previously been reported as dependency factors through screens and validation studies, as well as GIANT (human tissue-specific gene functional network) and using this to inform a machine learning model to predict novel host factors. In short, they predicted 1,578 novel genes and several of the top 20 enriched pathways had common themes of splicing, RNA polymerase/ transcription, and proteasomal degradation [76].

HIV-CRISPR screening: a new high-throughput method for identifying dependency factors

In an effort to identify host factors involved at each step of the viral life cycle in a non-biased fashion, that would be high throughput and not rely on arrayed assays, the Emerman Lab developed a novel CRISPR-based screen, called HIV-CRISPR, in which virus replication itself serves as a readout for host genes whose knockout affects HIV [77]. The HIV-CRISPR vector is a lentiviral vector that contains the sequences for a guide RNA (from an sgRNA library) targeting a single gene, the Cas9 enzyme, a viral packaging signal (psi), and repaired long terminal repeats which allow transcription of packageable genomic-length viral RNA. When cells are transduced with HIV-CRISPR, the vector is transcribed, allowing production of the mRNA for Cas9 production and subsequent sgRNA targeting. In addition, they produce full length HIV-CRISPR genomes that can be packaged into nascent virions. Upon infection with HIV-1, newly-produced virions will encapsidate the HIV-CRISPR genome in trans with HIV genomes. These packaged HIV-CRISPR genomes will then serve as a barcode which we use to quantify the impact that knockout had on HIV infection, i.e. was there an increase or decrease of HIV infection in these individual knockout cells, as determined by the abundance or depletion of the barcode in the gDNA vs. the supernatant. This tool is able to detect both enrichment and

depletion of guideRNAs in a pooled fashioned, which allows for the detection of novel restriction factors as well as novel dependency factors, respectively (Figure 3).

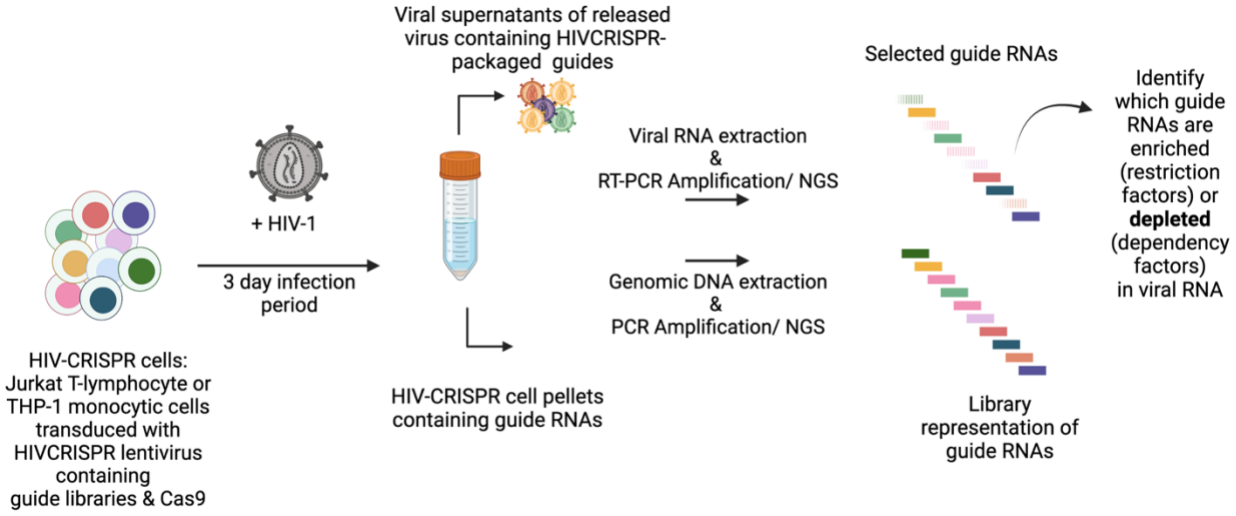


Figure 3. HIV-CRISPR screening process.
ZapKO-Jurkat-CCR5 or ZapKO-THP-1 cells containing the CRISPR knockout libraries were infected with HIV-1 in duplicate infections. Viral RNA and genomic DNA were collected 3 days post infection and sgRNA sequences present in virions (vRNA) and genomic DNA (gDNA) were quantified through RT-PCR/PCR and deep sequencing. This figure was made in Biorender and is a modified version of [78], used under Creative Commons CC-BY 4.0 license.

The HIV-CRISPR screening method was initially used in our lab to identify restriction factors in THP-1 cells (monocytic cell line) that were involved in the interferon response against HIV using a focused CRISPR guide library of interferon-stimulated genes (ISGs) and a variety of HIV-1 strains [77]. This library, termed the Packageable ISG Knockout Assembly (PIKA) library, was designed based on microarray and RNA-seq datasets from multiple HIV permissive cell types (PBMCs, THP-1, primary CD4+ T cells, monocyte derived macrophages (MDMs), and monocytes, amounting to 1905 ISGs. By screening with the PIKA library, OhAinle et al. was able identify a small panel of restriction factors including MxB, TRIM5alpha, IFITM1, and

Tetherin as important for HIV-1 restriction in a strain-dependent manner in a monocyte-derived macrophage cell line (THP-1 cells).

Importantly, although this library was biased to target human ISGs in HIV-1 target cells with the intention of finding restriction factors, this screen also identified known dependency factors, CXCR4 and NFkB, as well as novel dependency factors CD169, SEC62, and TLR2 [77]. Prior to this study, Siglec/CD169 had been characterized to enhance trans dendritic cell-mediated CD4+ T cell infection by binding the HIV virion through sialylated glycosphingolipids [79]. Though as HIVCRISPR screening assesses cell autonomous effects, OhAinle et al. showed a novel effect of interferon-induced expression of Siglec/CD169 in THP-1 cells. They also showed a novel pro-viral role for anti-bacterial Toll-like receptor 2 that was independent of the mechanism of entry. Lastly, they showed that knockdown of SEC62, a member of the SEC complex that facilitates translocation of transmembrane proteins to the ER for migration to the cell surface, inhibits CD4 translocation to the cell membrane, thus inhibiting wild type, but not VSVg psuedotyped LAI infection. Since this library was heavily biased towards restriction factor identification and the screen was comprehensive enough to still identify dependency factors that were in the library, this is strong evidence that HIV-CRISPR screening is a powerful tool to uncover host factors that act as dependency factors for HIV at each stage of the life cycle.

Thus, we still have an incomplete understanding of which cellular genes and proteins play essential roles, how they aid infection, and which ones are specific to HIV versus other related lentiviruses, which can provide insights for additional therapeutic or cure design strategies. In my dissertation, I will discuss how I leveraged the technology of HIVCRISPR, combined with genome-wide and targeted CRISPR guide sub-libraries, to create a pan-proviral list of HIV-1 dependency factors across multiple strains of HIV.

Chapter 2. Materials and Methods

Cell Culture

The Jurkat (ATCC) cell line was cultured in RPMI-1640 medium (Thermo Fisher Scientific) supplemented with 10% Fetal Bovine Serum (FBS), Penicillin-Streptomycin (Pen/Strep), and 10 mM HEPES. 293T and TZM-bl cells (ATCC) were cultured in DMEM (Thermo Fisher Scientific) with 10% FBS and Pen/Strep. Mycoplasma testing was done prior to screens and validation studies by the Fred Hutch Specimen Processing/Research Cell Bank Shared Resource and was not detected in any cell lines used in these experiments.

Plasmids

The HIV-CRISPR vector was previously described [78]. HIV-CRISPR constructs targeting genes of interest were cloned by annealing complementary oligos with overhangs that allow directional cloning into HIV-CRISPR using the BsmBI restriction sites (Supplemental Table 2). lentiCRISPRv2 plasmid was obtained from Feng Zhang via Addgene #52961. pMD2.G and psPAX2 were obtained from Didier Trono via Addgene #12259/12260. sgRNAs designed to target genes of interest were cloned into lentiCRISPRv2 using methods previously published [78] . The wild type (HIV-1_{LAI}) and *env*-deleted (HIV-1_{LAI}VSV-G) HIV-1_{LAI} proviruses were previously described [80, 81]. The Clade A HIV-$1_{Q23.BG505}$ molecular clone was previously described [82, 83]. CH470TF was provided by Beatrice Hahn [84]. The lentiviral pHIV-dTomato (Addgene plasmid; 21374; RRID:Addgene_21374) and pHIV-Zsgreen (Addgene plasmid; 18121; RRID:Addgene_18121) expression vectors were deposited by Bryan Welm. The pHIV-zsGreen/CCR5 and pHIV-dTomato/CCR5 constructs were created by cloning the human CCR5

CDS into pHIV-zsGreen and pHIV-dTomato backbones using the BamHI and NotI sites. The Toronto human knockout pooled library (TKOv3) was obtained from Jason Moffat via Addgene #90294 [85].

HIV Dependency Factor (HIVDEP) CRISPR/Cas9 sgRNA Library Construction

The HIVDEP sgRNA library is composed of 525 genes (4,191 sgRNAs). The top scoring 368 genes (as determined by -log10 MAGeCK score) of the genome-wide screen were included in the HIVDEP library. Additional unique top scoring genes with <10%FDR from HuEpi and PIKA screens were added to the HIVDEP library that contained an additional 131 genes. NR5A1, NHLRC4, SFTPA2, ZNF768, MYL10, GIMAP5, SPG21, CHSY3, ZNF25, REG1A, and ATXN3 were manually selected as non-essential genes [86] that were neither enriched nor depleted in the genome-wide screen. Six new sgRNAs were designed using two algorithms, GUIDES [87] and CHOPCHOP [88]. We also manually included other known dependency factors, including CCR5 and PAPSS1. Two hundred and twelve Non-targeting controls were designed using GUIDES and also included. The HIVDEP sgRNA library was synthesized (Twist Biosciences) and cloned into HIV-CRISPR. Oligo pools were amplified using Herculase II Fusion DNA Polymerase (Agilent; 600677) combined with 1 ng of pooled oligo template, primers ArrayF and ArrayR (ArrayF primer: TAACTTGAAAGTATTTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACA CCG and ArrayR primer: ACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTC TAGCTCTAAAAC), an annealing temperature of 59°C, an extension time of 20 seconds, and 25 cycles. Following PCR amplification, a 140 bp amplicon was gel-purified and cloned into

BsmBI (NEB; R0580) digested HIVCRISPR using Gibson Assembly (NEB; E2611S). Each Gibson reaction was carried out at 50°C for 60 min. Drop dialysis was performed on each Gibson reaction according to the manufacturer's protocol using a Type-VS Millipore membrane (VSWP 02500). 5 µl of the reaction was used to transform 25 µl of Endura electrocompetent cells (Lucigen; 60242-2) according to the manufacturer's protocol using a Gene Pulser (BioRad). To ensure adequate representation, sufficient parallel transformations were performed and plated onto carbenicillin containing LB agarose 245 mm x 245 mm plates (Thermo Fisher) at 492-times the total number of oligos of each library pool. After overnight growth at 37°C, colonies were scraped off, pelleted, and used for plasmid DNA preps using the Endotoxin-Free Nucleobond Plasmid Midiprep kit (Takara Bio; 740422.10). The HIVDEP library was sequenced and contains all 4,191 sgRNAs included in the synthesis (GEO Dataset, submission in progress).

Virus and Lentivirus Production

293 T cells (ATCC; CRL-3216) were plated at $1.5x10^5$ cells/mL in 6 well plates one day prior to transfection. 3 ul of TransIT-LT1 reagent (Mirus Bio LLC) transfection agent was used per ug of DNA. For lentiviral preps, 293Ts were transfected with 667 ng lentiviral plasmid, 500 ng psPAX2 and 333 ng MD2G. For HIV-1 production, 293Ts were transfected with 1 ug/well proviral DNA. One day post-transfection media was replaced. Two- or three- days post-transfection lentiviral supernatants of the same type were combined and filtered through a 0.2 µm filter (Thermo Scientific; 720-1320). For lentiviral preps used for the creation of CRISPR/Cas9 knockout lines, supernatants were harvested and clarified from three wells per each lentiviral prep and were concentrated in microcentrifuge tubes for 1 hr at 4°C at 16,100 x g. The volume was reduced to equal 4x concentration before vortexing vigorously and resuspending at 4°C for 2

days. Each tube was combined per lentiviral stock before using for transduction. For HIVDEP library preps, supernatants from 40 x 6 well plates were pooled and concentrated by ultracentrifugation as described in [89]. Concentrated lentivirus was used immediately or aliquots were made and stored at -80°C. To increase infectivity of the CH470 TF stock, we concentrated ~75 mL of virus to ~3 mL (25x concentration) using an Amicon Ultra-15 Centrifugal Filter Unit (Millipore Sigma; UFC905008). All viral and lentiviral infections and transductions were done in the presence of 20 μg/mL DEAE-Dextran (Sigma; D9885).

HIV-CRISPR screening

Prior to the HIV-CRISPR screening or generation of knockout cell lines for validation studies, we performed CRISPR/Cas9-mediated knockout of Zinc Antiviral Protein ZC3HAV1 (ZAP) to increase efficiency of the HIV-CRISPR vector [78]. We used gene Knockout v2 kit (GKOv2) for ZAP, including the following sgRNA sequences: GTGGTGTTGGAGACCGG, CCTGGAGCAGCGCGTCC, and TGAAGCAGCACACCTCC (Synthego, Redwood City, CA) complexed with 1 μL of 20 μM Cas9-NLS (UC Berkeley Macro Lab) and single cell sorted into a 96-well U-bottom plate to make clonal knockouts. Clonal KO lines were identified and selected using the ICE editing analysis software (Synthego). An individual clone with biallelic knockouts were used to create both ZsGreen/CCR5 and dTomato/CCR5 stably expressing lines by transduction with pHIV-ZsGreen/CCR5 or pHIV-dTomato/CCR5 lentiviruses followed by cell cytometry sorting for expression of CCR5 via ZsGreen or dTomato fluorescence using the BD FACSCANTO II or Sony MA900 (Fred Hutch Flow Cytometry Core) and analyzed with FlowJo software. These Jurkat cells were transduced with HIV-CRISPR vectors containing the TKOv3, HuEpi, PIKA, or HIVDEP library at an MOI of <1 and selected for puromycin

resistance at 0.4 ug/mL. Ten days post puromycin selection, cells were infected with HIV-1 strains and levels of infection were measured by intracellular p24gag staining (Table 1) to obtain at least an MOI of 0.1. To maintain ~500X coverage of sgRNAs for the HIV-1 Q23BG505 screen which was more difficult to achieve an MOI of 0.1, we increased the number of cells infected proportionally to the lower MOI (Table 1). Genomic DNA and viral RNA was harvested and amplified 3 days post infection, sequenced, and analyzed using the MAGeCK-FLUTE algorithm [90].

Statistical analysis of HIV-CRISPR screen data

Library pools were demultiplexed, reads were assigned of libraries to assign sequences to each sample (allowing no mismatches), trimmed, and aligned to the TKOv3 or HIVDEP sgRNA libraries using Bowtie [91]. An artificial or "Synthetic" NTC sgRNA set the same size of TKOv3 or HIVDEP was created by iteratively binning NTC sgRNA sequences (4 NTC sgRNAs/ gene or 8 NTC sgRNAs/gene for TKOv3 or HIVDEP, respectively). Relative enrichment of sgRNAs and genes were analyzed using the MAGeCK-Flute statistical package [90]. Enriched gene ontologies of screen data were determined by Gene Set Enrichment Analyses [92] as a part of the MAGeCK-Flute package. For each biological HIVDEP screen replicate, z scores were calculated for each gene for comparison across screens performed with different viruses at different points in time. These were calculated as is described in [93]. These z scores and enriched pathway data of the TKOv3 screen were used to generate pathway-focused heatmaps across each HIVDEP screen using the Morpheus (https://software.broadinstitute.org/morpheus).

Dependency factor validation through spreading infections and luciferase quantification

Twenty-four candidate genes were knocked out in Jurkat- CCR5 cells using the two most efficient guides per gene from the sgRNA library (as calculated by $\log_2$ fold change sgRNA enrichment). We also generated non-targeting control (NTC), CD19-, AAVS1- , and CD4-CRISPR/Cas9 knockout Jurkat-CCR5 cells as negative and positive controls. Knockout cell pools were created via transduction with lentiCRISPRv2 containing gene targeting constructs. Less than 24 hours post transduction, media was replaced with RPMI containing 0.4 ug/mL Puromycin. Transduced cell pools were selected with puromycin for 10 days prior to HIV infection. gDNA was harvested for editing analysis after 10-12 days under selection. Knock out cells were maintained as pools rather than individual clones to remove artifacts of clone to clone heterogeneity in infection. We measured infection of the knockout pools by both luciferase luminescence at two days post infection and by measuring virus release three days after infection by measuring reverse transcriptase activity in viral supernatants as described [94]. A stock of HIV-1 LAI was titered, aliquoted at −80°C and used as the standard curve in all assays.

Genomic Editing Analysis

Knockout cells were harvested and genomic DNA was extracted using the QIAamp DNA Blood Mini Kit (Qiagen; 51185). Sites of editing were amplified using primers specific to each targeted locus (Supplemental Table 3). Primers were designed to amplify a 500 or 1000 base pair amplicon of the targeted locus using either Q5 High-Fidelity DNA Polymerase (NEB; M0491S) or Platinum Taq DNA Polymerase High Fidelity (ThermoFisher Scientific; 11304011). PCR amplicons were sequenced (Fred Hutch Shared Resources Genomics Core – sanger sequencing)

and analyzed by ICE (Synthego) to determine the percent of alleles edited at each locus in the cell population [95].

Flow Cytometry/p24gag Analysis

Intracellular p24gag staining was conducted on cells 3 days post infection to determine viral titer of each stock before using these for CRISPR screens or infection experiments with LAI, LAI-VSV-G, Q23BG505, and CH470TF. Cells were harvested and fixed in 4% paraformaldehyde for 10 min and diluted to 1% in DPBS. Cells were permeabilized in 0.5% Triton-X for 10 min and stained with 1:300 KC57-FITC (Beckman Coulter 6604665; RRID: AB_1575987) or 1:300 KC57-RD1 (numbers). Cells were analyzed on a Celesta or Fortessa Flow Cytometer (Fred Hutch Flow Cytometry Core). To assess infectivity of viral stocks used for the spreading infections, cells were fixed in BD Fix/Perm for 20 minutes, washed, and permeabilized in in BD Perm/Wash Buffer. The cells were stained with 1:300 KC57-FITC (Beckman Coulter 6604665; RRID: AB_1575987) or 1:300 KC57-RD1 (Beckman Coulter 6604667 in BD Perm/Wash Buffer for 30 minutes, washed, and then resuspended in PBS. Cells were read on a Celesta 2 or 3 (Fred Hutch Flow Cytometry Core) and analyzed in FlowJo. For CD4 cell surface marker staining, cells were washed twice in PBS, stained in PBS/1% BSA, incubated at 4°C for 1 hr, washed twice in PBS, and resuspended in PBS. The cells were stained with 1:50 CD4-APC (BD Biosciences 555349; AB_398593) and analyzed on Celesta 2 flow cytometer (Fred Hutch Flow Cytometry Core).

Chapter 3.

A Virus-Packageable CRISPR System Identifies HIV Dependency Factors Co-opted by Multiple Strains.

Abstract

At each stage of the HIV life cycle, host cellular proteins are hijacked by the virus to establish and enhance infection. We adapted the virus packageable HIV-CRISPR screening technology at a genome-wide scale to comprehensively identify host factors that affect HIV replication in a human T cell line. Using a smaller, targeted HIV Dependency Factor (HIVDEP) sub-library, we then performed screens across HIV strains representing different clades and with different biological properties to define which T cell host factors are important across multiple HIV strains. Nearly 90% of genes selected across various host pathways validated in subsequent assays as bona fide host dependency factors including numerous proteins not previously reported to play role in HIV biology such as UBE2M, MBNL1, FBXW7, PELP1, SLC39A7, and others. Our ranked list of screen hits across diverse HIV-1 strains form a resource of HIV dependency factors for future investigation of host proteins involved in HIV biology.

Introduction

With a small genome of ~9.2 kb that encodes 14 major proteins, HIV must hijack host cellular machinery to successfully establish infection. These host proteins, called dependency factors, are needed for entry, transit into the nucleus, uncoating, viral integration, as well as subsequent steps of transcription of viral RNA, RNA splicing and export, protein translation, and virion assembly and budding.

While many of these important host factors have been identified one by one, high throughput methods have also attempted to create comprehensive maps of host dependency and/or restriction factors. For example, four genome-wide RNA-interference screens [42-44, 96] as well as a genome-wide CRISPR/Cas9 screen [70] sought to identify dependency factors for HIV infection. While these screens identified genes important for cellular attachment and entry, nuclear entry, integration, transcription, and nuclear export [42-44, 70, 96], there was very poor overlap across screens [68]. Other approaches have sought to identify novel dependency factors through protein-protein interaction screens and identify hundreds human proteins that are physical interactors with the 18 HIV-1 proteins and polyproteins in Jurkat T-lymphocyte and 293T cells [71]. A subset of these interacting proteins were functionally validated in primary CD4+ T cells as HIV dependency or restriction factors [75]. In addition, interaction based and gene network approaches have been used to computationally predict dependency factor genes [76] and to develop a viral-host dependency epistasis map (Ve-MAP) [97]. However, none of these approaches have used an assay with a functional readout to examine multiple viral strains across the entire HIV lifecycle in T cells.

We had previously developed a high throughput CRISPR screening method, called HIV-CRISPR, in which lentiviral genomes encoding sgRNAs are incorporated into budding virions

serving as a readout for genes important for HIV infection [78, 89, 98]. Although this work focused on finding restriction factors against HIV-1, we also demonstrated that this method could also identify dependency factors, although as the guide library used in those studies specifically targeted interferon-stimulated genes, the number of dependency factors identified was relatively limited. Here, we use a whole genome guide library to look more globally at HIV dependency factors in a T cell line. We used these data to inform the design of a smaller, custom HIV Dependency Factor CRISPR sublibrary (HIVDEP) to identify HIV dependency factors across multiple HIV strains. Our screens identified many previously reported HIV dependency factors across multiple parts of the viral life cycle, including the HIV-1 receptor CD4, co-receptor CXCR4, LEDGF/p75, NFkB, and many genes encoding components of the Mediator complex. Further, we identify genes not previously reported to play a role in HIV biology involved in transcriptional regulation, protein degradation, RNA regulation, as well as epigenetic factors affecting both early and late events of viral replication. Genes not previously identified as being important for HIV replication that were validated here include genes involved in Cullin-ring ligase mediated protein degradation, such as UBE2M and FBXW7, pre-mRNA alternative splicing regulator MBNL1, transcription factor PELP1, and zinc transporter/ tyrosine kinase activator SLC39A7. Our expanded catalog of ~200 host dependency factors required at different stages of the life cycle has potential to inform therapeutic and cure design.

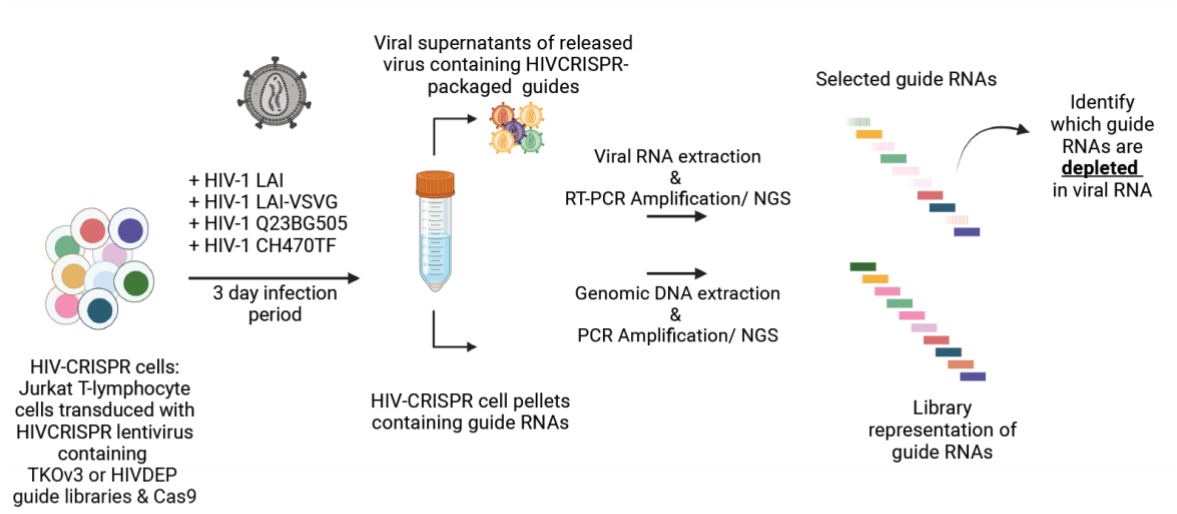
Results

A whole genome CRISPR screen for HIV dependency genes

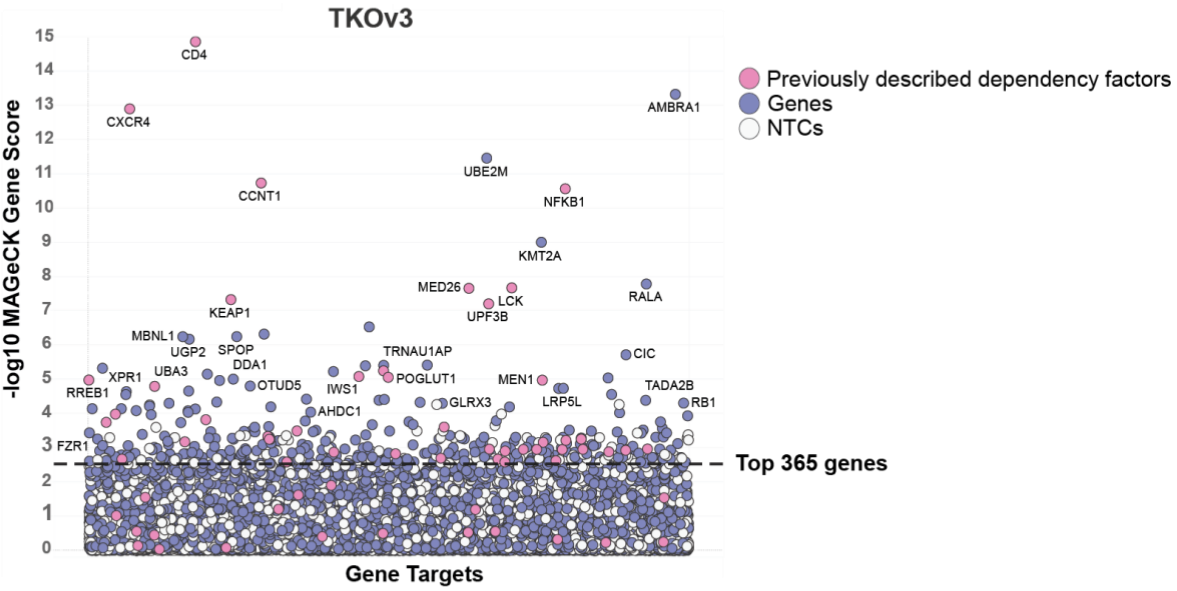
The HIV-CRISPR screening method has previously identified restriction factors in THP-1 cells (a CD4+ monocytic leukemia cell line) that were involved in the interferon response against HIV using a focused CRISPR guide library of interferon-stimulated genes (ISGs) [78]. In order to more comprehensively identify HIV dependency factors, we adapted this approach to a whole genome strategy utilizing the Toronto Knockout version 3 whole genome (TKOv3) library. This library targets 18,053 protein-coding genes with 4 guides per gene and also includes 142 non-targeting control guides [85].

Our screening strategy is outlined in Figure 4A. The HIV-CRISPR vector is a lentiviral vector that contains the sequences for a guide RNA (from an sgRNA library) targeting a single gene, the Cas9 enzyme, a viral packaging signal (psi), and a repaired long terminal repeat region which allows transcription of a genomic RNA [78]. In addition to transcribing Cas9 and the encoded sgRNA, cells that are transduced with the HIV-CRISPR vector are also capable of making mRNA corresponding to full-length HIV-CRISPR genomes after infection with HIV-1. Thus, upon infection with an HIV strain of interest, newly-produced virions will also encapsidate the HIV-CRISPR genome *in trans*. Therefore, if a dependency factor has been knocked out, these cells will support less productive infection and fewer HIV-CRISPR genomes containing the sgRNA specific for the dependency factor will be packaged and released from infected cells. This relative depletion of sgRNAs in the viral supernatant, as compared to the genomic DNA [89] is the readout in which we infer knockout of a dependency factor. The advantage of this system is that CRISPR guides can be evaluated in bulk rather than arrayed in single wells, which makes a whole genome approach feasible.

A.


Viral supernatants of released virus containing HIVCRISPR-packaged guides
+ HIV-1 LAI
+ HIV-1 LAI-VSVG
+ HIV-1 Q23BG505
+ HIV-1 CH470TF
3 day infection period
Viral RNA extraction & RT-PCR Amplification/ NGS
Genomic DNA extraction & PCR Amplification/ NGS
Selected guide RNAs
Identify which guide RNAs are depleted in viral RNA
HIV-CRISPR cells: Jurkat T-lymphocyte cells transduced with HIVCRISPR lentivirus containing TKOv3 or HIVDEP guide libraries & Cas9
HIV-CRISPR cell pellets containing guide RNAs
Library representation of guide RNAs

B.


TKOv3
Previously described dependency factors
Genes
NTCs
-log10 MAGeCK Gene Score
Gene Targets
Top 365 genes
CD4
CXCR4
AMBRA1
UBE2M
CCNT1
NFKB1
KMT2A
MED26
LCK
UPF3B
RALA
KEAP1
MBNL1
UGP2
SPOP
DDA1
TRNAU1AP
CIC
XPR1
UBA3
OTUD5
POGLUT1
MEN1
TADA2B
RREB1
IWS1
GLRX3
LRP5L
RB1
AHDC1
FZR1

C.

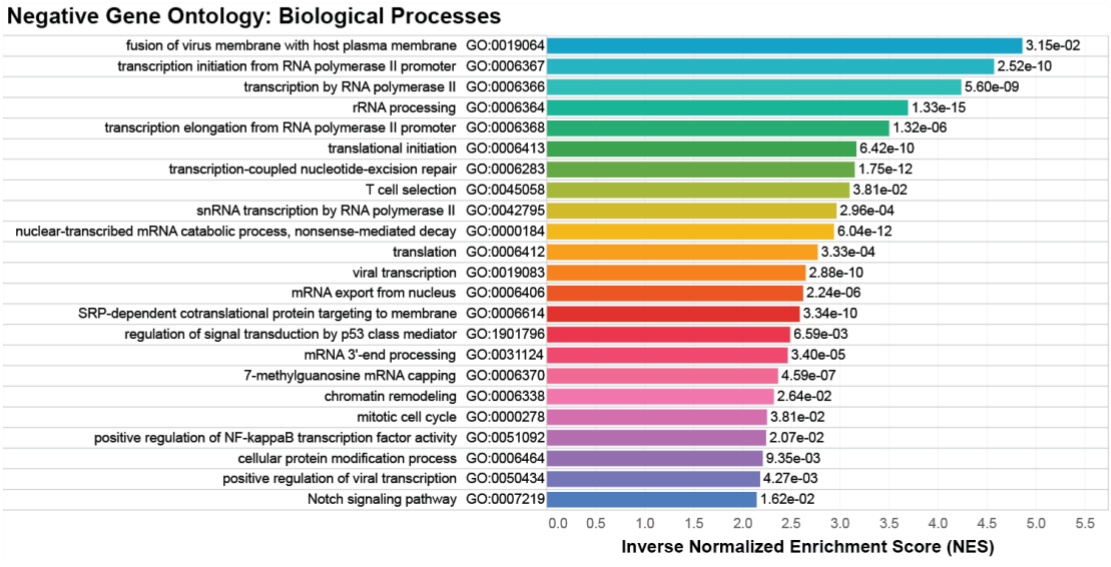

Negative Gene Ontology: Biological Processes
fusion of virus membrane with host plasma membrane GO:0019064 3.15e-02
transcription initiation from RNA polymerase II promoter GO:0006367 2.52e-10
transcription by RNA polymerase II GO:0006366 5.60e-09
rRNA processing GO:0006364 1.33e-15
transcription elongation from RNA polymerase II promoter GO:0006368 1.32e-06
translational initiation GO:0006413 6.42e-10
transcription-coupled nucleotide-excision repair GO:0006283 1.75e-12
T cell selection GO:0045058 3.81e-02
snRNA transcription by RNA polymerase II GO:0042795 2.96e-04
nuclear-transcribed mRNA catabolic process, nonsense-mediated decay GO:0000184 6.04e-12
translation GO:0006412 3.33e-04
viral transcription GO:0019083 2.88e-10
mRNA export from nucleus GO:0006406 2.24e-06
SRP-dependent cotranslational protein targeting to membrane GO:0006614 3.34e-10
regulation of signal transduction by p53 class mediator GO:1901796 6.59e-03
mRNA 3'-end processing GO:0031124 3.40e-05
7-methylguanosine mRNA capping GO:0006370 4.59e-07
chromatin remodeling GO:0006338 2.64e-02
mitotic cell cycle GO:0000278 3.81e-02
positive regulation of NF-kappaB transcription factor activity GO:0051092 2.07e-02
cellular protein modification process GO:0006464 9.35e-03
positive regulation of viral transcription GO:0050434 4.27e-03
Notch signaling pathway GO:0007219 1.62e-02
Inverse Normalized Enrichment Score (NES)

Figure 4. Genome-wide HIV-CRISPR screening to identify dependency factor candidates for HIV Dependency Factor guide library.

(A) HIV-CRISPR screening process. Jurkat-CCR5 cells containing a CRISPR knockout library are infected with HIV-1 in duplicate infections. Strains used in this study are listed about the arrow. Viral RNA and genomic DNA were collected 3 days post infection and sequences corresponding to sgRNAs present in virions (vRNA) and genomic DNA (gDNA) were quantified through deep sequencing. **(B)** MAGeCK Gene analysis of the Genome-wide screen showing the most depleted (dependency) genes in Jurkat T cells infected with HIV-1_{LAI} in duplicate infections. X-Axis: randomly arrayed target genes. Y-Axis: -Log10 MAGeCK score for each gene. Host factors previously reported as dependency factors are shown in Pink. Other genes in the library are shown in Purple. "Synthetic" Non-targeting control (NTC) genes generated in silico by iterate random binning of the 142 NTC sgRNA sequences to generate a negative control set which are shown in White. Gene names are shown for all hits with a -Log10 MAGeCK score greater than 6. The top 365 most depleted candidate genes (excluding the synthetic NTC genes) are indicated by the dashed black line. **(C)** Gene Set Enrichment Analysis of the genome-wide screen. The top 20 most enriched Negative Gene Ontologies (most enriched pathways of the depleted/ dependency factor candidate genes) are shown here in ranked order by Inverse Normalized Enrichment Score (NES). Adjusted p values are displayed next to each gene ontology.

We transduced Jurkat-CCR5 cells, a T cell line susceptible to HIV infection and engineered to express the CCR5 co-receptor, with the HIV-CRISPR vector containing the TKOv3 whole genome library. The cells were then infected with HIV-1 at an MOI of 0.5, and genomic DNA and viral RNA was harvested from cell pellets and viral supernatants, respectively three days after infection. These samples were then deep sequenced to quantify enrichment or depletion of guide RNAs in the viral supernatants compared to guide representation in the cellular genomic DNA [89]. Using the Model-based Analysis of Genome wide CRISPR/Cas9 Knockout (MAGeCK) screens algorithm and the log 2-fold change in gene depletion based on all four guide RNAs, statistical scores were assigned [90, 99] to delineate the most depleted genes, indicating factors important for infection (Figure 4B).

As expected given their essential role in viral entry, the receptor CD4 and the co-receptor CXCR4 (required for the CXCR4-tropic strain HIV-1_{LAI} used in this screen), were among the most depleted genes, ranking as the top scoring dependency factors (Figure 4B). Additionally, many other previously reported dependency factors were identified in the 300 top scoring dependency factors (shown as pink dots) that include all different parts of the HIV lifecycle including entry (CD4, CXCR4, LCK), integration and uncoating (PSIP1, TNPO1), transcription (NFKB1, SP1, CyclinT1 (CCNT1)) Mediator complex genes (MED7, MED10, MED16, MED18, MED23, MED26), POLR2A), and budding (TSG101, CHMP4B). Importantly, the non-targeting controls randomly binned to synthetic "genes" cluster near the bottom.

As a global approach to identify the most negatively enriched biological processes of the genome wide screen (candidate dependency factor pathways), we used the Gene Set Enrichment Analysis as part of the MAGeCK-Flute pipeline [90]. Of the top 20 negatively enriched Gene Ontology pathways (Figure 4C), three are classified as important for viral processes, including fusion of virus membrane with host plasma membrane (GO:0019064), viral transcription (GO:0019082), and positive regulation of viral transcription (GO:0050434). Notably, there were enriched pathways identified that may play roles for multiple stages of the viral life cycle, including entry/egress, transcription, mRNA capping and processing, translation, protein modification/trafficking, and T cell selection. Notwithstanding the caveat that the screen missed some expected dependency factors that did not cluster away from the NTCs (pink dots below the line in Figure 4B), this screening approach is capable of identifying dependency factors across the entire viral life-cycle and provides ample opportunity to investigate novel gene roles in the HIV life cycle.

An HIV-dependency sub-library screened with multiple HIV-1 strains

Smaller, targeted libraries are more powerful as it is easier to maintain suitable coverage, more guides can be utilized per gene, and the smaller screens can be done in much higher throughput in more replicates. Therefore, before doing any validation of the hits in the whole genome screen, we used data from our whole genome to inform the construction of a targeted CRISPR guide library called the HIVDEP (HIV Dependency Factor) library. We initiated this library by including guides targeting genes in the top scoring 500 from the TKOv3 genome-wide screen. After subtracting non-targeting controls from the rankings, this amounted to 368 genes. We also added back the known dependency factors that fell below this threshold in the screen including CCR5 and PAPSS1 which are essential for CCR5-dependent strains. Because transcription and chromatin modification/ regulation were major categories of negatively enriched genes in the whole genome screen (Figure 4C), we also performed an additional targeted screen using a custom-designed Human Epigenome (HuEpi) library consisting of 838 human epigenome and epigenetic regulator genes [100] and included any gene with <10% false discovery rate (FDR) from the HuEpi screen that were not already included in the top TKOv3 gene list (Figure 5A) as well as smaller number of negatively enriched genes in an ISG-related library [78] (Figure 5B), resulting in an additional 131 genes. Finally, 11 genes previously called "non-essential" [86] that were neither depleted nor enriched in our genome-wide screen were included as negative control genes. In total, the HIVDEP library consists of 525 genes targeted by 8 guides and 210 NTC guides for a total of 4,401 guides.

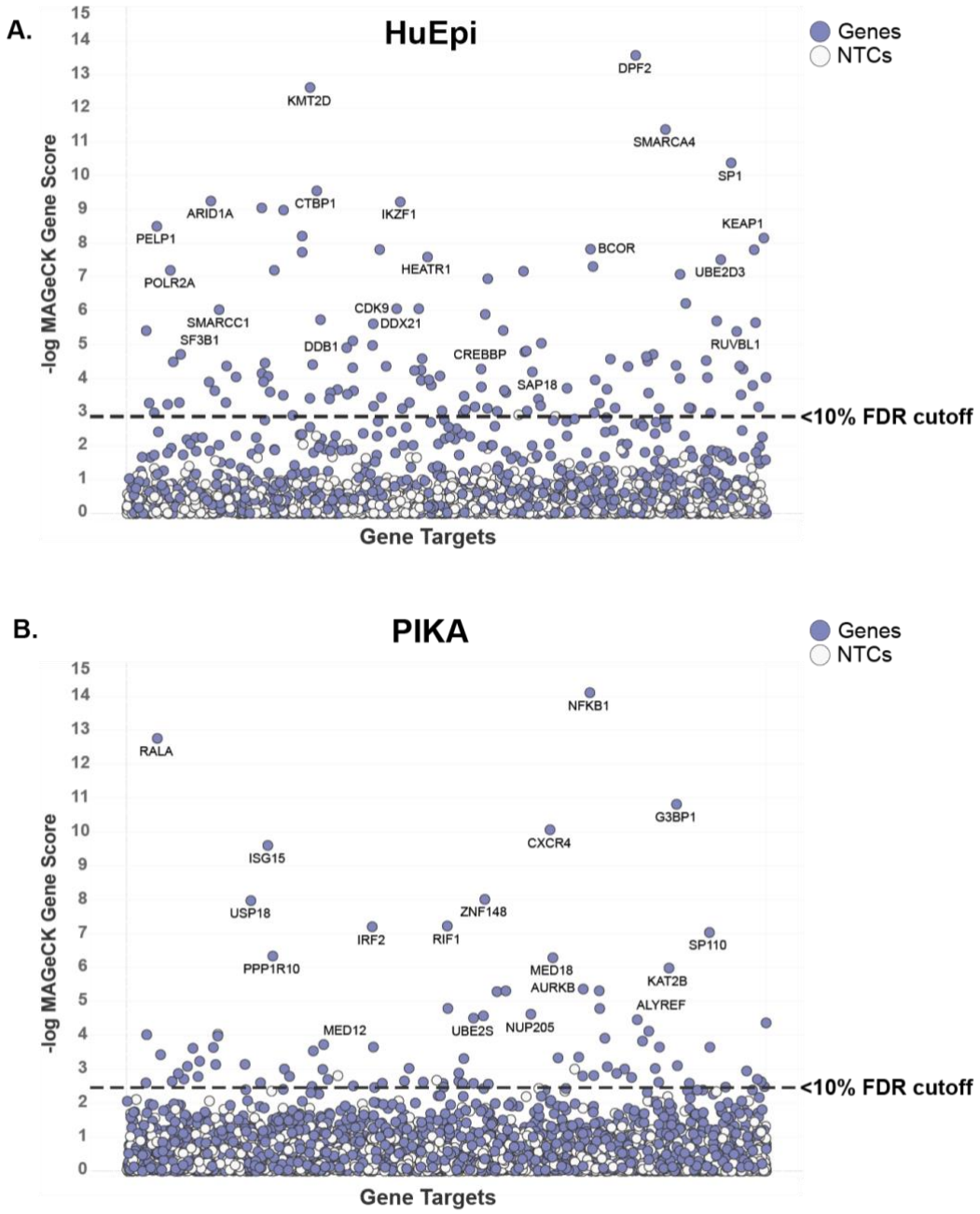


Figure 5. MAGeCK gene analysis of the HuEpi or PIKA CRISPR screens.

MAGeCK Gene analysis of the **(A)** Human Epigenome/Epigenetic library (HuEpi) or **(B)** Packageable ISG Knockout Assembly library (PIKA) screen showing the most depleted (dependency) genes in Jurkat T cells. X-Axis: randomly arrayed target genes. Y-Axis: -Log10 MAGeCK score for each host factors previously reported as dependency factors are shown in Pink. Purple represents the candidate genes. Genes scoring >10% FDR cutoff are those above the dashed line were included in the creation of the

sublibrary HIVDEP. White = Non-targeting control (NTC) gene set generated in silico by randomly binning the 252 NTC sgRNA sequences into genes (six NTC guides per 841 genes) for HuEpi and 200 NTC sgRNA sequences into genes (eight NTC guides per 1,905 genes) for PIKA.

To identify specific host factors that are either important for all HIV-1 strains, or ones that are strain-specific, we used the HIVDEP library screened against genetically distinct HIV-1 strains (Table 1). Specifically, these include HIV-1 LAI, which is in Clade B and uses co-receptor CXCR4, HIV-1 Q23.BG505 which is a clade A CCR5-tropic strain [83], and a transmitted/ founder virus from clade B that is R5-tropic, HIV-1 CH470 T/F [101]. We also included an HIV-1 strain (LAI) that was deleted for its own envelope gene and pseudotyped with VSV-G to delineate entry-specific factors. Thus, we used HIV-1 strains that represent two clades (A and B), utilize different entry mechanisms, (-X4, -R5, or VSV-G), and are either lab-adapted vs. primary strains. Each screen was done in duplicate and the number of infected cells in each screen ranged from 6% to 23% (Table 1). In addition, we also performed screens with the HIV-1 LAI strain on two different occasions each with a separate transduction of the library followed by infection with different stocks of HIV-1 LAI (screens referred to as LAI and LAI redo in Figure 6).

HIV-1 strain	Clade	Co-receptor tropism	Traits	%p24 3dpi per biological replicate
LAI	B	CXCR4	Lab-adapted	LAI : R1 = 10.57; R2 = 10.33
				LAI redo : R1 = 7.05; R2 = 7.5
LAIΔenv-VSV-G	B	n/a	Lab-adapted/ alternate entry mechanism	R1 = 23.93 R2 = 23.9
Q23.BG505	A	CCR5	Primary isolate	R1 = 6.19 R2 = 6.32
CH470 TF	B	CCR5	Primary isolate/ Transmitted founder virus	R1 = 20.6 R2 = 22.9

Table 1. Viruses used in the HIVDEP screens reported in this study.
LAI was used twice to assess reproducibility of the screen (referred to as "LAI redo" in this paper). Percent infection of each biological replicate of each HIVDEP screen **(**as measured through flow cytometry quantification of cytoplasmic gag 3 days post infection) is shown in the last column.

We compared the MAGeCK scores of genes versus the synthetic non-targeting controls (NTC) in the genome-wide screen versus each of the screens done with the dependency factor-focused HIVDEP library using box and whisker plots (Figure 6A). While the mean score of the synthetic non-targeting controls in each screen are similar, the mean of the MAGeCK scores for the genes in each HIVDEP screen was significantly higher than the mean of the genes in the genome-wide TKOv3 screen (Figure 6A; Welch's t test). Moreover, the scores of the 95th percentile of genes (top horizontal line in each plot) were also increased in the HIVDEP library relative to the whole-genome screen for each of the HIV strains tested (Figure 6A). This indicates that the sub-library approach was successful in increasing the distance of noise to signal (NTCs to genes) relative to the whole genome screen. Moreover, waterfall plots of the top 20 hits from each screen (Figure 6B-6F) show that known dependency factors are identified such as

the receptor CD4, CCNT1 (CycT1), NFKB1, and members of the Mediator complex. Moreover, several novel genes score in the top 20 for multiple strains, including UBE2M, MBNL1, FBXW7, PCGF1, and PPP2CA, implicating those gene products in processes important for viral infection across HIV-1 strains.

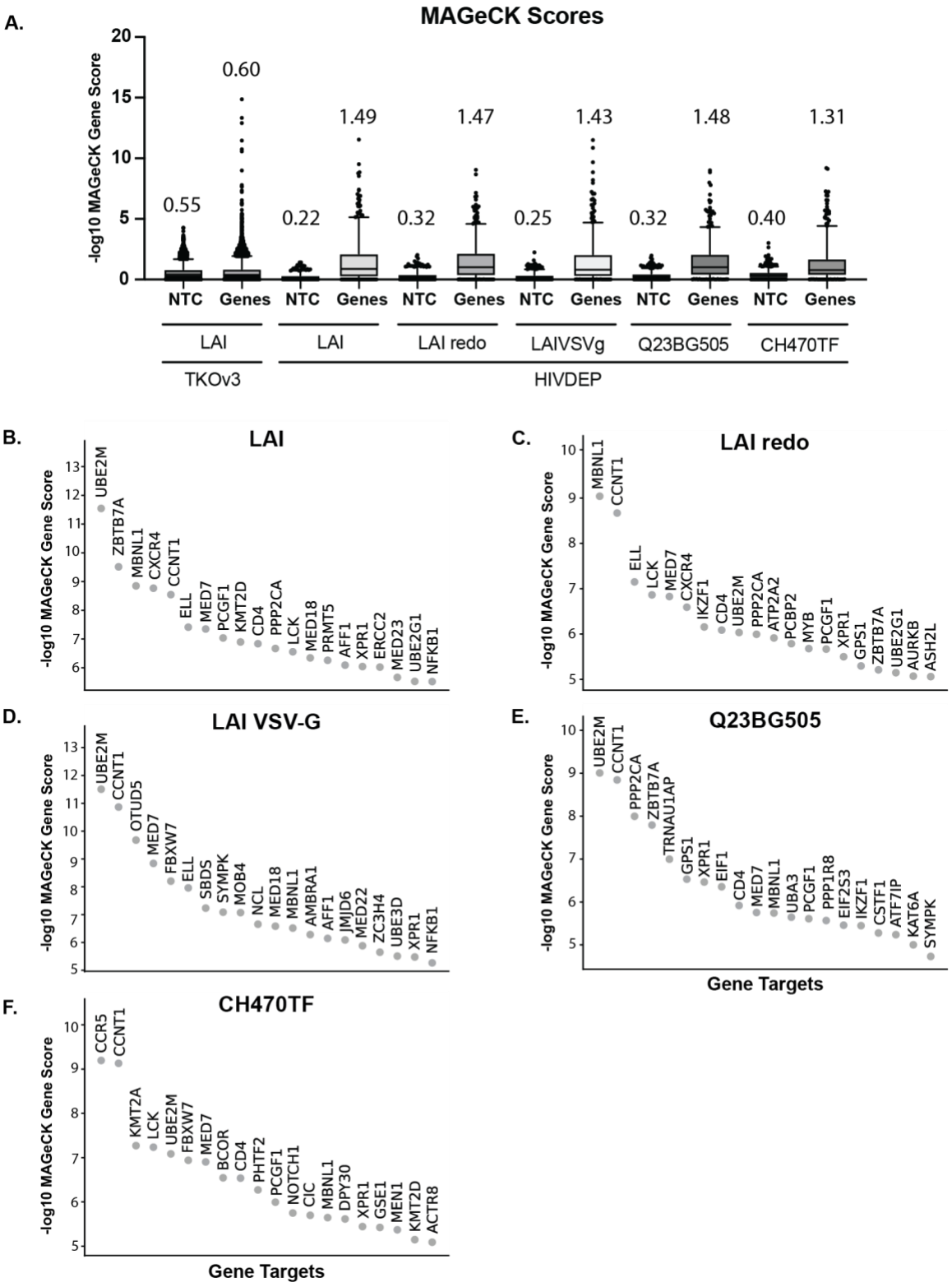

A.
MAGeCK Scores
-log10 MAGeCK Gene Score
0.55
0.60
0.22
1.49
0.32
1.47
0.25
1.43
0.32
1.48
0.40
1.31
NTC
Genes
LAI
LAI redo
LAIVSVg
Q23BG505
CH470TF
TKOv3
HIVDEP
B.
LAI
UBE2M
ZBTB7A
MBNL1
CXCR4
CCNT1
ELL
MED7
PCGF1
KMT2D
CD4
PPP2CA
LCK
MED18
PRMT5
AFF1
XPR1
ERCC2
MED23
UBE2G1
NFKB1
C.
LAI redo
MBNL1
CCNT1
ELL
LCK
MED7
CXCR4
IKZF1
CD4
UBE2M
PPP2CA
ATP2A2
PCBP2
MYB
PCGF1
XPR1
GPS1
ZBTB7A
UBE2G1
AURKB
ASH2L
D.
LAI VSV-G
UBE2M
CCNT1
OTUD5
MED7
FBXW7
ELL
SBDS
SYMPK
MOB4
NCL
MED18
MBNL1
AMBRA1
AFF1
JMJD6
MED22
ZC3H4
UBE3D
XPR1
NFKB1
E.
Q23BG505
UBE2M
CCNT1
PPP2CA
ZBTB7A
TRNAU1AP
GPS1
XPR1
EIF1
CD4
MED7
MBNL1
UBA3
PCGF1
PPP1R8
EIF2S3
IKZF1
CSTF1
ATF7IP
KAT6A
SYMPK
Gene Targets
F.
CH470TF
CCR5
CCNT1
KMT2A
LCK
UBE2M
FBXW7
MED7
BCOR
CD4
PHTF2
PCGF1
NOTCH1
CIC
MBNL1
DPY30
XPR1
GSE1
MEN1
KMT2D
ACTR8
Gene Targets

Figure 6. Iterative screening with HIVDEP sublibrary enriches for previously reported and candidate dependency factors.

MAGeCK score comparison of each CRISPR screen. Non-targeting control (NTC) sgRNA scores were randomly binned (four NTC guides per gene for TKOv3 or 8 NTC guides per gene for HIVDEP) to recapitulate the same number of genes in the respective libraries. Y-Axis: -Log10 MAGeCK scores. The mean MAGeCK scores for the synthetic NTC versus the genes are shown above each graph and are represented by the line within each box. The top 95th percentile of NTC/genes is represented as the top horizontal line. For statistical analyses, the MAGeCK scores of the synthetic non-targeting controls (shown here as NTCs) were compared to the MAGeCK scores of the Genes in each screen. The Gene MAGeCK scores per each screen were also compared across screens. Each comparison resulted in significant p values =<0.0001 = ****; Welch's t-test. Waterfall plots of the top 20 genes or all genes in each HIVDEP screen in descending order are shown for the following HIV-1 strains **(B)** LAI, **(C)** LAIredo, **(D)** LAI-VSV-G, **(E)** Q23BG505, and **(F)** CH470TF. The MAGeCK gene scores for each screen are listed in Supplemental Table 1.

Validation of hits from the combined HIV dependency factor screens across different cellular processes

While hits across two screens done with the same strain are similar, the absolute MAGeCK scores between the two are not identical (LAI and LAIredo in Figure 6B and 6C), Therefore, to compare hits across screens, we normalized our data with a z score analysis as previously described [93]. We thus created a ranked order list of both the average of all screen results by z score as well as the replicates from screens with each strain (Supplemental Table 3). We created pathway-focused heatmaps using the z scores for each replicate and the Gene Set Enrichment Analyses pathways established from the whole genome screen (Figure 7). As an initial control to determine if the screen could distinguish different strains, we looked at known entry factors that would be expected to be different between CXCR4-using, CCR5-using, and the VSV-G

pseudotyped viruses in the Binding and Viral Entry Gene Set (Figure 7A). The HIV-1 LAI screens correctly identify CXCR4, but not CCR5 nor PAPSS1, while the opposite is true of the CCR5-tropic viruses, Q23BG505 and CH470TF. Each of the strains that uses wild-type HIV-1 envelope require CD4, but the VSV-G pseudotyped HIV-1 does not. LCK (lymphoid specific Src Kinase) which is important for T cell signaling downstream of CD4 engagement and reported to be important for viral core transit from the plasma membrane to the nucleus [16] tracks with CD4 in our screens (Figure 7A).

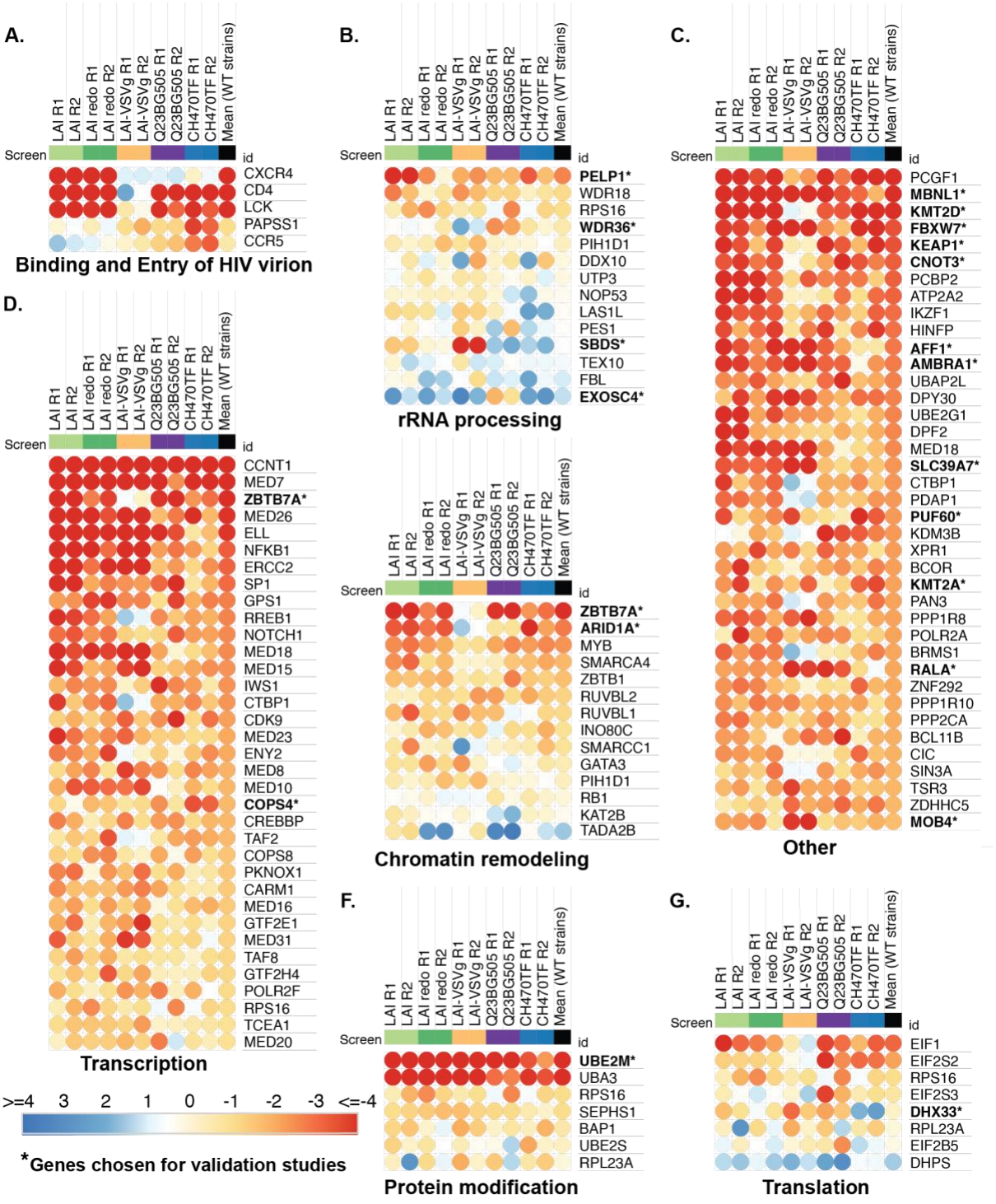


Figure 7. Common and differential use of host cellular pathways by HIV-1 strains.

Comparative pathway-focused heatmaps showing enriched or depleted sgRNAs across each HIVDEP screen. The pathways shown are derived from the top 20 most enriched Negative Gene Ontologies of the

genome-wide screen (Figure 4C). Any genes not included in the HIVDEP library were excluded. Z scores were calculated as described in [93]. Z scores on each heatmap are colored from red (lowest/ most depleted genes, i.e. dependency factors) to blue (highest/ most enriched genes, i.e. negative or restriction factors). The median NTC z score was 0.5 and marks the inflection in the color scale. **(A-G)**. Each biological replicate is represented as a separate column showing the mean scores across wild type (non-VSV-G-pseudotyped) strains. The "Transcription" and "Other" heatmaps were truncated to the top 40 hits each heatmap. Gene names that are bolded with an asterisk indicate they were chosen for validation studies in Figures 8, 9 and 12. "Other" are genes that were not assigned to any of the top Gene Ontology categories from Figure 4C.

The pathway-focused heatmaps also show enriched genes for each screen with a large number involved in transcription as well as rRNA processing, protein modification, chromatin remodeling, and translation (Figure 7). We then chose a subset of genes across different host cellular processes (Figure 7) to functionally validate using infectious virus in multi-round spreading infections with both LAI and with Q23BG505. We also picked some genes that scored lower in the screens in order to determine a cut-off from which we would be more confident of remaining hits that were not directly validated, as well as some genes that scored more highly for one strain versus another (see Figure 8A for the list ranked in descending order of average Z score as well as the Z score for each replicate of each gene). Our validation strategy used two guides per gene to knock out a candidate dependency factors in Jurkat-CCR5 cells. Positive controls included knocking out CD4, and negative controls included non-targeting control (NTC) guides, as well as guides targeting genes that are not involved in HIV or T cell biology, CD19 (B cell marker), and AAVS1 (the AAV integration site often used as a safe harbor locus [102-104]). The knockout cells were maintained as pools rather than clones to avoid clone-to-clone differences and were tested for gene editing as well as HIV infection within 2 weeks of generation.

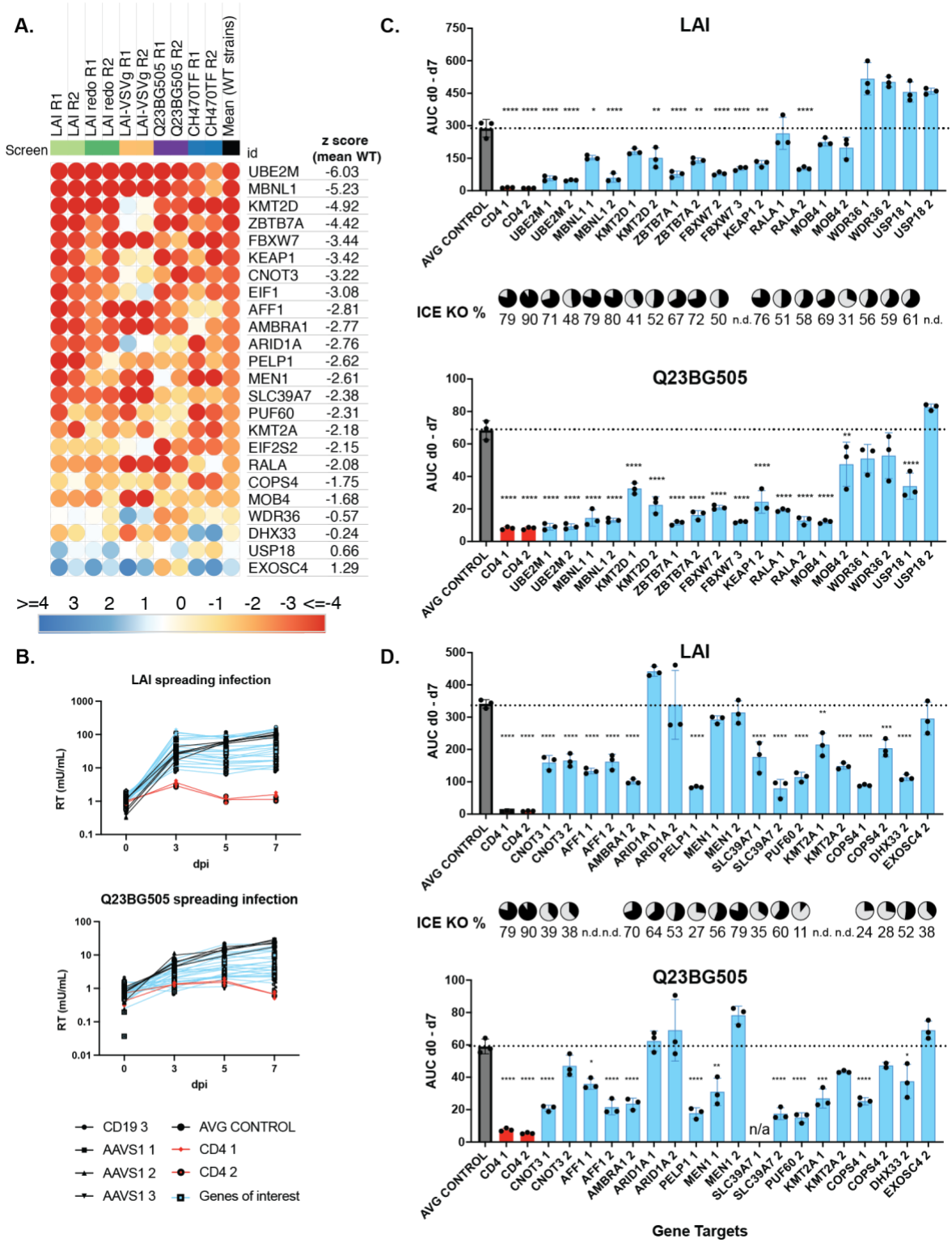

A.
LAI R1
LAI R2
LAI redo R1
LAI redo R2
LAI-VSVg R1
LAI-VSVg R2
Q23BG505 R1
Q23BG505 R2
CH470TF R1
CH470TF R2
Mean (WT strains)
Screen
id
z score (mean WT)
UBE2M -6.03
MBNL1 -5.23
KMT2D -4.92
ZBTB7A -4.42
FBXW7 -3.44
KEAP1 -3.42
CNOT3 -3.22
EIF1 -3.08
AFF1 -2.81
AMBRA1 -2.77
ARID1A -2.76
PELP1 -2.62
MEN1 -2.61
SLC39A7 -2.38
PUF60 -2.31
KMT2A -2.18
EIF2S2 -2.15
RALA -2.08
COPS4 -1.75
MOB4 -1.68
WDR36 -0.57
DHX33 -0.24
USP18 0.66
EXOSC4 1.29
>=4 3 2 1 0 -1 -2 -3 <=-4
B.
LAI spreading infection
RT (mU/mL)
dpi
Q23BG505 spreading infection
RT (mU/mL)
dpi
CD19 3
AAVS1 1
AAVS1 2
AAVS1 3
AVG CONTROL
CD4 1
CD4 2
Genes of interest
C.
LAI
AUC d0 - d7
ICE KO %
79 90 71 48 79 80 41 52 67 72 50 n.d. 76 51 58 69 31 56 59 61 n.d.
Q23BG505
AUC d0 - d7
D.
LAI
AUC d0 - d7
ICE KO %
79 90 39 38 n.d. n.d. 70 64 53 27 56 79 35 60 11 n.d. n.d. 24 28 52 38
Q23BG505
AUC d0 - d7
n/a
Gene Targets

Figure 8. Validation of curated top hits list

(**A)** Heatmap of the candidate genes used for validation studies, ordered by the mean z score of WT strains **(B)** Pooled Knockout Jurkat-CCR5) generated by transducing with lentiviral vectors encoding sgRNAs including positive control gene CD4, negative controls CD19 or AAVS1, or candidate dependency factor genes. Two knockout lines per gene were generated using the highest scoring sgRNAs selected from across each HIVDEP screen. Viral supernatants were collected at days 0, 3, 5, and 7 to assess overall effect on replication kinetics via reverse transcriptase activity at each timepoint. The spreading infections were performed over two batches. Y axis = Reverse Transcriptase milliUnits/ mL. Batch 1 is shown in panel B and Batch 2 is shown in panel Figure 9 **(C and D).** Area Under the Curve (AUC) was calculated for each cell line after 7 days of infection in Batch 1 (panel **C**) and Batch 2 (panel **D**) with either LAI or Q23BG505. Infection data of each guide is shown separately. For statistical analysis, all conditions are compared to the mean of the control cell lines (CD19 and AAVS1). One way Anova; Tukey's multiple corrections test, p-value = ns > 0.05, <0.05 = *, =<0.05 = **, =<0.001 = ***, =<0.0001 = ****. For each knockout line, Synthego ICE analysis was performed and knockout scores are displayed as pie charts in line with the corresponding gene target. n.d. = could not be determined

Each knockout pool was infected with HIV-1 strains at an MOI of 0.15. All infections were done in triplicate and virus growth was measured by assaying reverse transcriptase activity released into the supernatant 0, 3, 5, and 7 days after infection. Replication curves for the majority of the genes of interest (blue) lines were in-between the CD4-KO lines and negative control lines (black), exhibiting inhibition of infection or slower growth kinetics for both LAI infection and Q23BG505 infections (Figure 8B). Infections were done in two batches with similar controls in both and grouped as such (Figure 8C is batch 1 and Figure 8D is batch 2). We used all of the data points in the spreading infection to calculate an area under the curve (AUC) for each infection (Figure 8C, 8D). We find that knockout of UBE2M which was one of the top hits in each of the screens (Figure 8B-8F) had the strongest phenotype in the spreading infections for both viruses tested (Figure 8C, 8D). In order of their effects, the strongest hits in this set of genes (aside from the positive control of CD4) were UBE2M, PELP1, MBNL1, ZBTB7A, and

FBXW7. In addition, KMT2D, KMT2A, KEAP1, CNOT3, AFF1, AMBRA1, SLC39A7, PUF60, RALA, and COPS4 validated for both viruses. MOB4 (Figure 8C; Figure 9C, 9D) and WDR36 at day 7 (Figure 9C, 9D) validated for Q23BG505, but not for LAI. This was predicted for WDR36 since it's z score was lower for Q23BG505 (Figure 8A), although not expected for MOB4, which had similar z scores across the two viral strains in the HIVDEP screen (Figure 8A). We measured the RT activity in the supernatant at day 5 or at day 7 as the primary readout instead of AUC (Figure 9), and find that at both time points, knockout of UBE2M which was one of the top hits in each of the screens (Figure 6) had the strongest phenotype in the spreading infections.

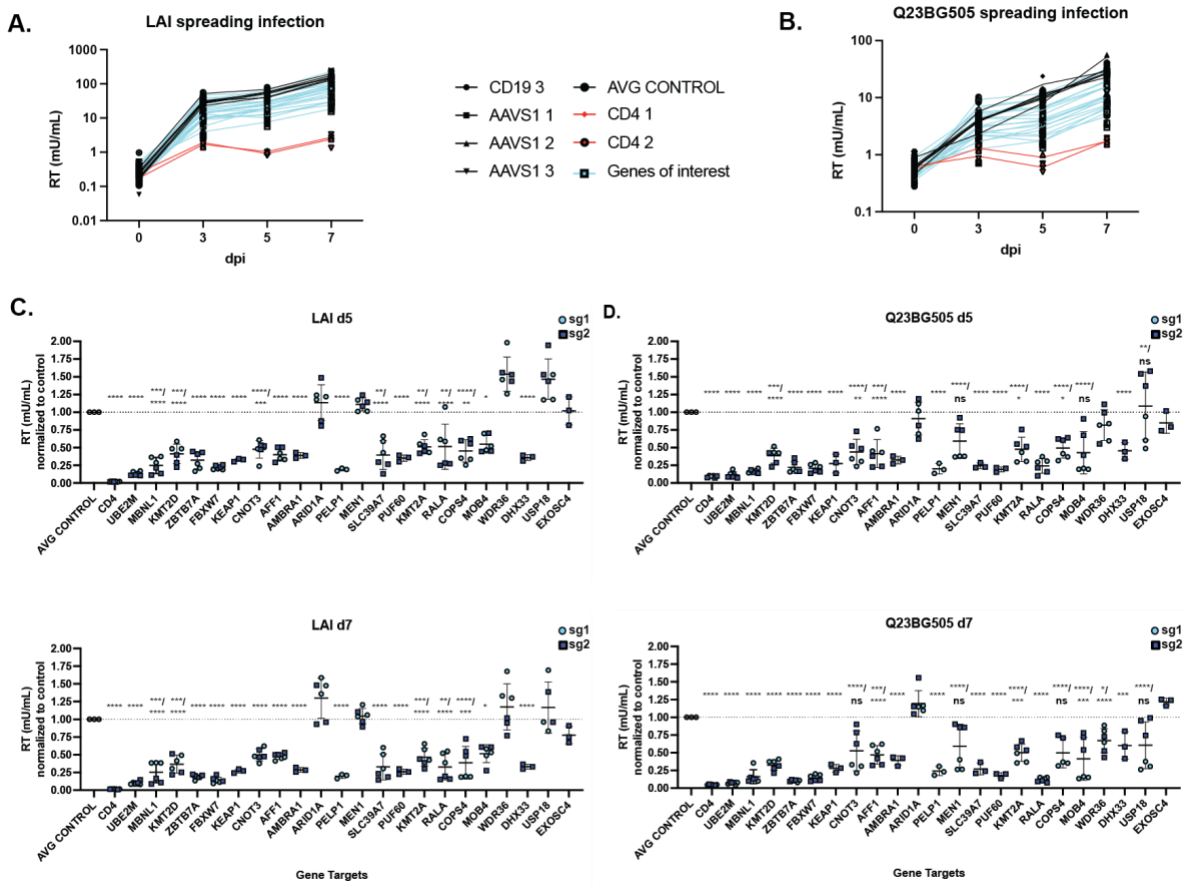


Figure 9. Validation of curated top hits at individual timepoints

(A-B) Pooled ZapKO-Jurkat-CCR5 cells were generated by transducing with sgRNA/lentiCRISPRv2 lentiviruses targeting either positive control gene CD4, negative controls CD19 or AAVS1, or candidate dependency factor genes. One or two knockout lines per gene were generated using the highest scoring sgRNAs across each HIVDEP screen, selected in puromycin for >10 days to allow for gene knockout, followed by knockout efficiencies determination by ICE analysis (I) or infection with either LAI or Q23BG505 (MOI = 0.15). Viral supernatants were collected at days 0, 3, 5, and 7 to assess overall effect on replication kinetics via reverse transcriptase activity at each timepoint. Representative graphs were included in the main text. A and B represent the genes in the other batch. Y axis = Reverse Transcriptase milliUnits/ mL. **(C, D)** Reverse Transcriptase activity at the d5 and d7 timepoints were normalized to the mean of the control cell lines (CD19 and AAVS1 knockout lines) to combine batch 1 and 2 for each virus. Both knockout lines per gene are displayed: sgRNA 1 is shown as cyan circles and sgRNA 2 is shown as dark blue squares. For statistical analysis, all conditions are compared to the mean of the control cell lines

(displayed as AVG CONTROLS). One way Anova; Tukey's multiple corrections test, p-value = ns > 0.05, <0.05 = *, =<0.05 = **, =<0.001 = ***, =<0.0001 = ****.

At the time of infection, we also quantified the gene editing efficiency for each pool which were generally between 50% and 80% (pie charts in Figure 8). Therefore, as these are less than complete knockouts, the degree of decreased virus replication in the absence of the candidate genes is likely an underestimate. However, a lack of gene editing would not explain the failure of AIRID1A nor MEN1 to validate (Figure 8D). There were not notable differences in the cell growth of the knockout pools, other than the pooled PELP1 KO cells early on in the selection process. PELP1 had lower editing efficiency, 27% as determined by ICE, which may indicate a larger outgrowth of the WT population with a smaller population of KO in the pool. Nonetheless, these results show that about 90% of these selected screen hits validated as HIV dependency factors to some degree.

In order to test the hypothesis that the strength of effect of the knockouts on HIV infectivity is correlated to its z score in the screen, we graphed the z score against the respective amount of virus replication represented by the AUC (Figure 8C and D) in spreading infections. Indeed, there was a negative correlation between the strain specific z scores and the amount of virus replication produced in each knockout pool (Figure 10A); LAI (R^2 = 0.39, p =0.001) and Q23BG505 (R^2 = 0.41, p =0.001). We also find that this correlation holds if we use the mean z score of all wt strains tested in the screen ((Figure 10B); LAI (R^2 =0.30, p = 0.007) and Q23BG505 (R^2 = 0.46, p = 0.0004)) which argues that despite strain differences, the list of hits can be used as a guide to identify HIV dependency factors. Thus, hits that scored higher in our screen are more likely to be authentic HIV dependency factors than those with lower z-scores.

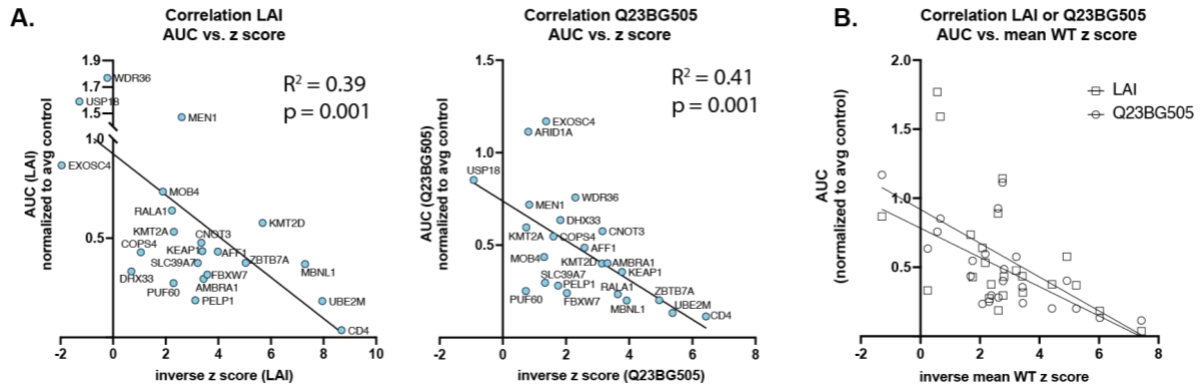


Figure 10. Correlation of gene z scores with AUC

(A) Area Under the Curve (AUC) was compared to the inverse z score of either LAI (left) or Q23BG505 (right). For statistical analysis, the mean biological replicate inverse z score per each gene for either LAI or Q23BG505, are compared to the mean AUC for infection of both guide knockouts per gene from Figure 8. Simple linear regression: LAI z score vs AUC (R^2 = 0.39, p =0.001); Q23BG505 z score vs AUC (R^2 = 0.41, p =0.001); **(B)** same as panel A, but the mean z score for all wild-type strains (Supplemental Table 3) was used rather than strain-specific z score: Mean WT z score vs LAI AUC (R^2 =0.30, p = 0.007); Mean WT inverse z score vs Q23BG505 AUC (R^2 = 0.46, p = 0.0004). Squares and solid line represent LAI; circles and dashed line represent Q23BG505.

At the lower end of hits in our validation experiments, MOB4 (mean wt z-score -1.683) validated for Q23BG505, though not LAI while each gene with z-scores greater than MOB4 did not validate except for DHX33. WDR36, the next highest z-score gene we chose for validation, did validate for Q23BG505 at day 7, but not for LAI, as predicted from the screen, though this phenotype was weak as the significance was lost when calculating AUC (Figure 8C). We therefore set two tiers of z score cutoffs with different levels of stringency to best identify hits for future validation studies. The most conservative tier, Tier 1, was set to the MOB4 z score cutoff (Supplemental Table 3). The Tier 2 cutoff was set to the WDR36 z score. This is especially stringent as there are well studied and previously described dependency factors that

score outside of Tier 1, but within Tier 2, including PSIP1/LEDGF (ranking #102 with mean wt z score of -1.37) which was also validated as a dependency factor in our system (Figure 12). This cutoff is likely still conservative as DHX33, which fell below Tier 2 (ranking #280), did validate for both viruses (Figure 8D), highlighting that there are likely other false negatives below this cutoff. We then conducted a literature review to identify how many of the 198 genes within Tiers 1 and 2 had already been identified in screens as HIV dependency factors as well as which ones have been experimentally validated. We find that 35 genes in this list had been previously identified in one or more previous screens [42-44, 70, 75]. We find that 57 of the 198 genes in our Tier 1 and 2 groups had been validated as HIV dependency factors through functional studies found in the literature (Supplemental table 3). Therefore, we report here an estimated over 140 unstudied novel HIV dependency factors, including the genes validated in Figures 8 and 12, with more confidence in the hits with higher z scores (Figure 10).

Combined HIV dependency screens highlight host entry factor preferences

As many experiments with HIV are done with VSV-G pseudotyped viruses to increase titers in single round infections, we were interested in further exploring hits with high differential z scores between the strains with wt HIV-1 envelope and VSV-G pseudotyped virus. Thus, we averaged the z scores for each gene across HIV-1 strains with wild-type envelopes and compared these scores with the z scores for the VSV-G pseudotyped virus (Figure 11).

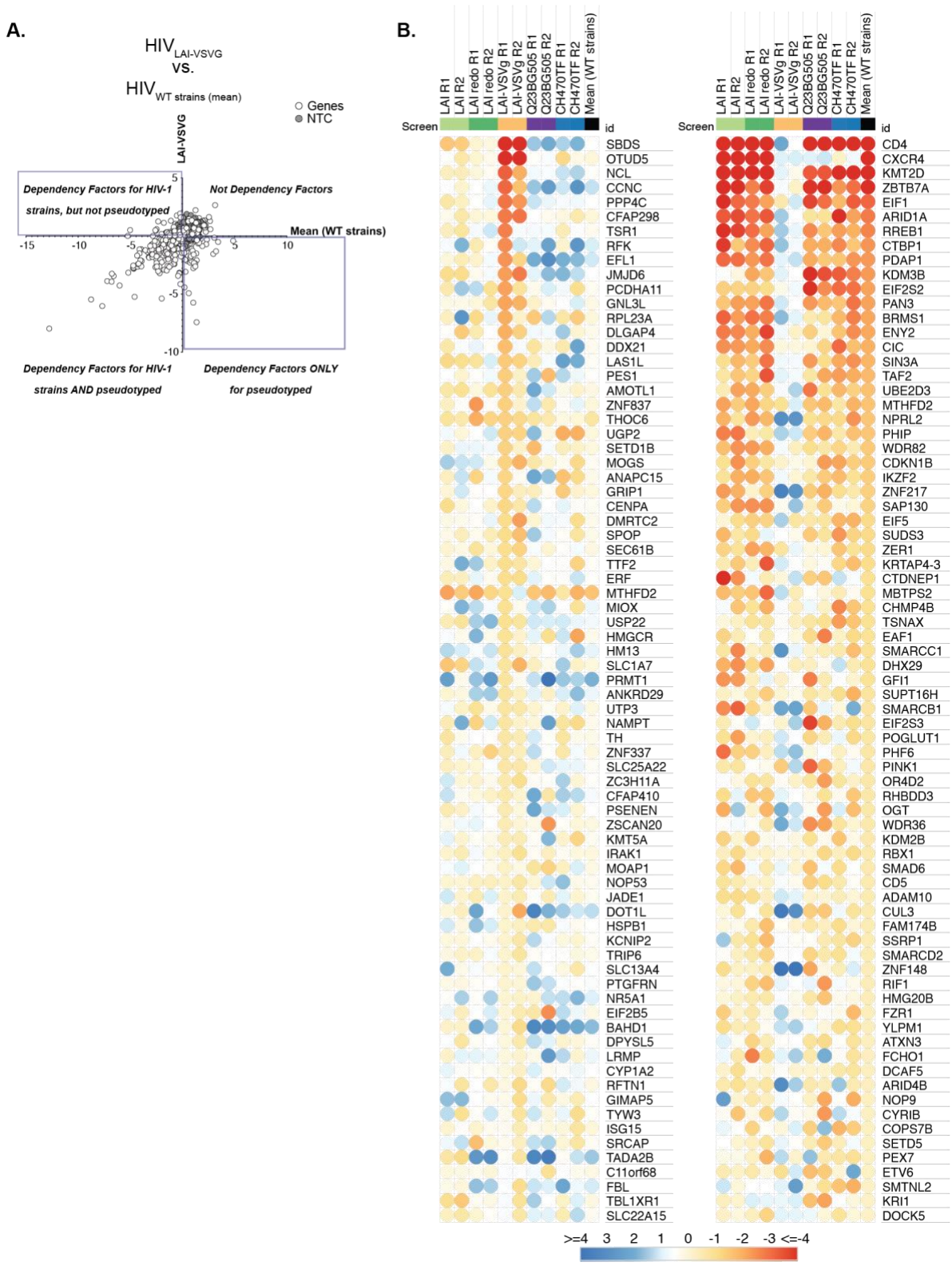

A.
HIV LAI-VSVG
vs.
HIV WT strains (mean)
Genes
NTC
LAI-VSVG
Mean (WT strains)
Dependency Factors for HIV-1 strains, but not pseudotyped
Not Dependency Factors
Dependency Factors for HIV-1 strains AND pseudotyped
Dependency Factors ONLY for pseudotyped
B.
LAI R1
LAI R2
LAI redo R1
LAI redo R2
LAI-VSVg R1
LAI-VSVg R2
Q23BG505 R1
Q23BG505 R2
CH470TF R1
CH470TF R2
Mean (WT strains)
Screen
id
SBDS
OTUD5
NCL
CCNC
PPP4C
CFAP298
TSR1
RFK
EFL1
JMJD6
PCDHA11
GNL3L
RPL23A
DLGAP4
DDX21
LAS1L
PES1
AMOTL1
ZNF837
THOC6
UGP2
SETD1B
MOGS
ANAPC15
GRIP1
CENPA
DMRTC2
SPOP
SEC61B
TTF2
ERF
MTHFD2
MIOX
USP22
HMGCR
HM13
SLC1A7
PRMT1
ANKRD29
UTP3
NAMPT
TH
ZNF337
SLC25A22
ZC3H11A
CFAP410
PSENEN
ZSCAN20
KMT5A
IRAK1
MOAP1
NOP53
JADE1
DOT1L
HSPB1
KCNIP2
TRIP6
SLC13A4
PTGFRN
NR5A1
EIF2B5
BAHD1
DPYSL5
LRMP
CYP1A2
RFTN1
GIMAP5
TYW3
ISG15
SRCAP
TADA2B
C11orf68
FBL
TBL1XR1
SLC22A15
CD4
CXCR4
KMT2D
ZBTB7A
EIF1
ARID1A
RREB1
CTBP1
PDAP1
KDM3B
EIF2S2
PAN3
BRMS1
ENY2
CIC
SIN3A
TAF2
UBE2D3
MTHFD2
NPRL2
PHIP
WDR82
CDKN1B
IKZF2
ZNF217
SAP130
EIF5
SUDS3
ZER1
KRTAP4-3
CTDNEP1
MBTPS2
CHMP4B
TSNAX
EAF1
SMARCC1
DHX29
GFI1
SUPT16H
SMARCB1
EIF2S3
POGLUT1
PHF6
PINK1
OR4D2
RHBDD3
OGT
WDR36
KDM2B
RBX1
SMAD6
CD5
ADAM10
CUL3
FAM174B
SSRP1
SMARCD2
ZNF148
RIF1
HMG20B
FZR1
YLPM1
ATXN3
FCHO1
DCAF5
ARID4B
NOP9
CYRIB
COPS7B
SETD5
PEX7
ETV6
SMTNL2
KRI1
DOCK5
>=4 3 2 1 0 -1 -2 -3 <=-4

Figure 11. Correlation matrix and extended heatmaps of entry specific host factors.
(A) Correlation matrix of LAI-VSV-G screen z scores vs. mean of the wild-type strains, LAI, LAI redo, Q23BG505, and CH470TF screen z scores per each gene. **(B)** Entire heatmap of z scores arranged by the mean of the wild-type strains, from most to least (top 75). **(C)** Entire heatmap of z scores arranged by the mean of the VSV-G-pseudotyped LAI, from most to least (top 75).

Using this correlation matrix, we ranked the 10 most highly ranked genes scoring as dependency factors only for the HIV-1 strains with wt envelopes, but not LAI-VSV-G (Figure 12A) and separately ranked the 10 most highly scoring genes for VSV-G pseudotyped HIV, but not for any of the wt HIV-1 stains (Figure 12B). We then functionally tested examples of each of these by knocking out the genes in Jurkat T cells and infecting them in a single round of infection with either an HIV-1 that either used VSV-G for entry or authentic HIV-1 envelope for entry. As positive controls, we used PSIP1 (also known as LEDGF), which should be needed for integration regardless of entry mechanism [105], and UBE2M which scored as one of the highest hits for all strains used in our screens regardless of entry mechanism or co-receptor (Figure 8). Negative controls included both Jurkat cells in which sgRNAs were used to target AAVS1, the safe harbor locus described above, as well as untransduced Jurkat cells.

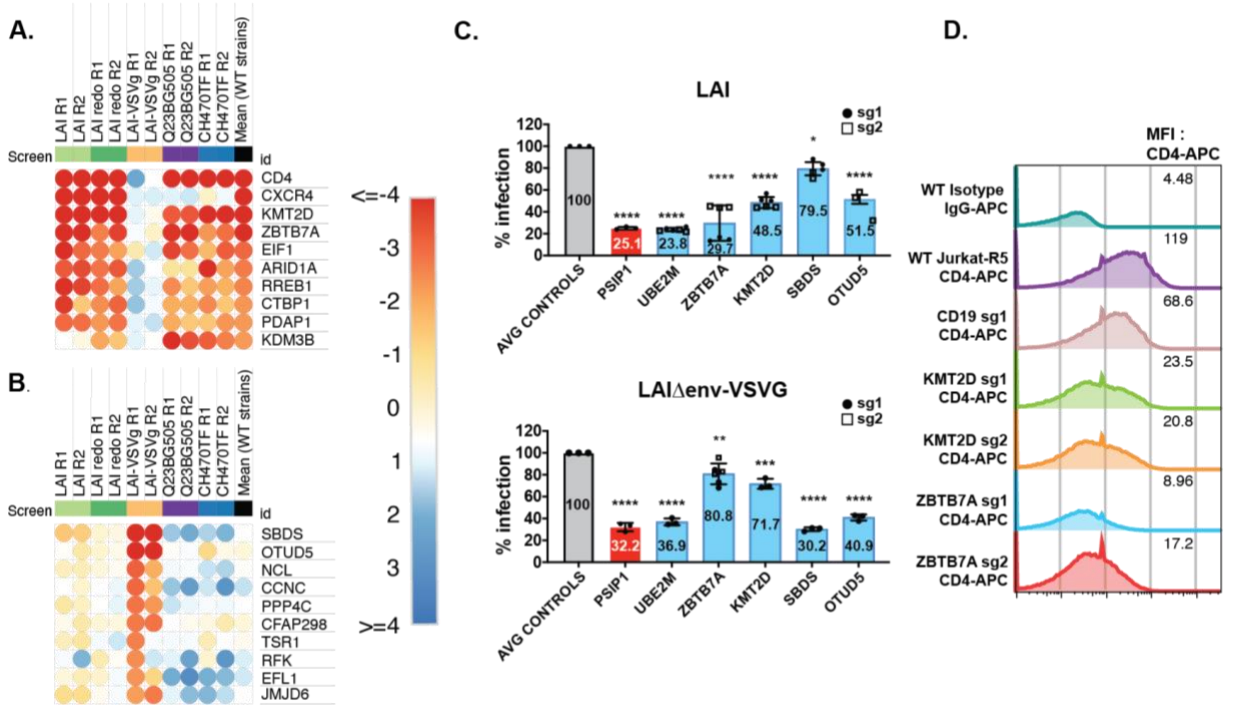


Figure 12. Entry-specific host factors.

(A) Heatmap of the top 10 most depleted sgRNAs based on the mean of the wild-type strains ,but not for VSV-G pseudotyped HIV-1. **(B)** Heatmap of the top 10 most depleted sgRNAs for VSV-G-pseudotyped HIV-1, but not for HIV-1 with wt HIV. Matrix for determination of panels A and B is in Figure 11 with the heatmap z score values in Supplemental Table 3. **(C)** Jurkat -CCR5 cell pools edited for gene targets of interest were created by transducing wild-type Jurkat-CCR5 cells with lentiCRISPRv2 sgRNA constructs using sgRNAs for CD19 (B cell marker used as a negative control), PSIP1 encoding for p75/LEDGF (positive control), and genes of interest UBE2M, ZBTB7A, KMT2D, SBDS, and OTUD5), puromycin selected for at least 10 days. Knockout pools were infected with HIV 1-LAI or VSV-G-pseudotyped HIV-1 which both encode luciferase in place of the nef gene. Luciferase expression in infected wt or knockout cells was quantified 2 days post infection. All infections were done in triplicate using two pools of knockouts with one or two different sgRNA guides per gene. Data for the two different guides are shown as black circles (sg1) and open squares (sg2) for each pooled knockout cell line. The mean percent of luciferase activity of all replicates relative to the control cells is displayed on each bar. (**D)** Surface CD4 expression of WT or KO Jurkat-ZsGreen/CCR5 cells was quantified using flow cytometry for CD4-APC. MFI for each cell line is shown in each respective row. Two different knockout pools corresponding to two different guides per gene (with the exception of CD19) are shown.

We find that the decrease in infectivity of the cells knocked out for both PSIP1 and UBE2M was similar regardless of whether entry was through HIV-1 env or VSV-G (Figure 12C). Moreover, as predicted from the screen data, ZBTB7A and KMT2D were important for infection with HIV-1 envelope, but had a minimal effect on infection with the HIV-1 pseudotyped with VSV-G (Figure 12C). In contrast, knockout of SBDS affected the VSV-G pseudotyped virus, but not HIV-1 envelope-mediated entry while OTUD5 knockout had an effect on both viruses but decreased the VSV-G pseudotype to a greater extent. Examination of CD4 levels on the pools of knockout cells show that cells with ZBTB7A and KMT2D have reduced levels of cell surface CD4 relative to the control knockout cells (Fig. 12D). These results indicate some HIV dependency factors act indirectly on the virus by affecting receptor availability on the cell surface, and other factors that affect pseudotyped viruses may not be relevant for wt HIV. Thus, our screens with whole genome and more focused libraries identify both shared and strain-specific HIV dependency factors across a broad range of cellular processes.

Discussion

We used HIV-CRISPR screening first at the genome-wide scale, and then at a smaller, targeted scale, to uncover novel genes that act as HIV-1 dependency factors. By using the HIV-CRISPR screening technique which assesses enrichment or depletion of guide-encoding genomes based on HIV-1 release from infected cells, we are able to establish a powerful functional assay to identify factors that promote HIV infection across the entire viral life cycle in a T lymphocyte cell line. Additionally, we were able to uncover dependency factors across multiple strains, from different clades and with different co-receptor tropisms. In addition to the

genes above that were validated, we also conducted GSEA pathway analyses to investigate which host pathways to which HIV-1 is most dependent in these screens, with transcription-related pathways being the most represented.

Transcription and chromatin remodeling factors as a major axis of HIV host dependency

Of the enriched pathways we identified, there were six transcription-related pathways (Figure 4C) which comprised a list of 58 genes that were included in the HIVDEP library as well an enriched pathway that includes chromatin remodeling genes. These are the largest pathway focused heatmaps, demonstrating an emphasis on transcription as a major focal point for host factors that promote HIV replication (Figure 7A). Some of the most enriched genes included previously reported dependency factors (positive controls) such as pTEFb components, CCNTI and CDK9, which are important for transcriptional elongation of HIV transcripts, (mean wt z scores: -7.97 and -2.18), transcription factors NFkB and SP1 (mean wt z scores: -3.24 and -3.02), and several Mediator complex genes (MED7, MED26, MED18, MED15, MED23, MED8, MED10, MED16, MED20) (Supplemental Table 3). In addition to Tat-recruited pTEFb members, the Super Elongation Complex (SEC) has also been reported to be required for efficient Tat transactivation and elongation [106, 107]. ELL is a member of the SEC and scored highly for all strains, except for CH470TF. Similarly, AFF1, which encodes for a scaffold protein of the SEC, also scored highly for each virus except CH470TF. AFF1 was also identified as a dependency factor in primary T cells (Hiatt et al. 2022). This is consistent with our data as we see a reduction of infection for both LAI and Q23BG505 in AFF1 KO Jurkat cells (Figure 8D). CREBBP, which encodes for CREB-binding protein (CBP), is recruited by Tat to the viral

LTR [108], and acetylation of Tat by CBP/p300 has been shown to be important for transcriptional activation at the LTR [109, 110]. Here, we have shown that CREBBP is important for four different viruses as it scored highly in our screens for LAI, LAI-VSV-G, Q23BG505, and CH470TF (mean zscore: -1.767).

PELP1, one of the top hits of this study, has previously been identified in an siRNA screen in HeLa-CD4 cells as a restriction factor [111]. However, both our screens and validation experiments suggest that it is acts as a dependency factor in Jurkat cells (mean replicate wt z score: -2.617) (Figures 7B, 8D). PELP1 has been shown to interact with SETDB1, a methyltransferase, oncogene, and restriction factor that effectively inhibits Tat activity by methylation [112]. It is possible that PELP1 promotion of SETDB1-Akt activation [113] may draw SETDB1 away from Tat, thus allowing transactivation. Therefore, PELP1 co-regulation of HIV-1 transcription factors and SETDB1 activity could explain the dramatic reduction we see upon gene knockout in Jurkat T cells.

ZBTB7A (also known as FBI-1 and LRF) is a zinc-finger protein involved in a diverse array of activities involving transcriptional co-repressors [114]. It has also previously been identified as a protein that binds the HIV-1 LTR and associates with HIV-1 Tat [115, 116]. Notably, ZBTB7A was recently identified to affect human coronavirus 229E through modulation of oxidative stress [117]. However, we find that ZBTB7A was a hit for each virus except for the VSV-G pseudotyped virus (mean replicate z score of 0.2) and resulted in lower cell surface CD4 levels in the knockout pools implying ZBTB7A transcriptionally regulates pathways important for CD4 levels rather than HIV transcription itself. Similarly, KMT2D, also known as MLL2, is a lysine methyltransferase, but known for post-translational modification of histones reviewed in [118] appears to act indirectly on HIV through effects on CD4 levels. Thus, several of the

implicated factors in this study, and others, likely act indirectly on viral replication. For example, SEC62, a component of protein translocation machinery in the ER membrane, was identified in an HIV-CRISPR screen performed in THP-1 cells [78]. Knockdown of SEC62 in THP-1 cells also reduced surface CD4 expression and wild type HIV-1 infection, but not VSV-G psuedotyped HIV-1 infection. Therefore, SEC62 also acts as a dependency factor indirectly, through affecting receptor availability [78].

Cellular Protein Modification as HIV dependency factors

Our screens identified multiple genes involving cullin-mediated ubiquitin ligase complexes. Cullin ring ligases are multi-subunit E3 ubiquitin ligase complexes responsible for ubiquitylating ~20% of cellular proteins targeted for degradation [119]. Cullin ring ligases (CRLs) must undergo neddylation to be activated. Neddylation involves the transfer of ubiquitin-like molecule NEDD8 to a lysine residue of a substrate, subsequently affecting activity, conformation, and/or subcellular localization. Notably, UBE2M, which neddylates CRL-1, CRL-2, CRL-3, and CRL-4 via the substrate specific E3 ubiquitin ligase Rbx1 (mean wt zscore: -0.54), was one of the top scoring hits across all screens, including the genome-wide LAI screen and the HIVDEP LAI VSV-G screen, following essential dependency factors CD4 and CyclinT1 (mean wt zscore CCNT1: -7.97; mean wt zscore CD4: -7.42; mean wt zscore UBE2M: -6.03). Knockout of UBE2M in Jurkat cells exhibited the most potent inhibition of infection of all the validated genes except for the positive control, CD4 (Figure 8). Neddylation and specifically UBE2M has been shown previously to be important for activation of CRL-4 and CRL-5 for efficient degradation of restriction factors SAMHD1 by HIV-2 Vpx and APOBEC3G by Vif,

respectively [120]. However, as neither of these factors are important in Jurkat T cells infected with wild type HIV, we hypothesize we have identified other effects of UBE2M neddylation for HIV-1 replication. UBA3 is one of two components of the NEDD8 E1 Activating Enzyme (NAE) heterodimer, just upstream of UBE2M in the neddylation process [121]. UBA3 scored extremely well across all screens, including the LAI-VSV-G pseudotyped virus (Figure 7F, mean wt zscore: -4.258; mean zscore of all: -4.61). Notably, inhibitors targeting the NAE complex have been effective in inhibiting HIV-1 infection [120], further validating these degradation pathways as important for the virus to replicate.

FBXW7, a substrate receptor for degradation of proteins dependent on phosphorylation status has a large array of transcriptional effects in cells [122] and was also validated in our screens (Figure 7C; Figure 8) as an HIV dependency factor. FBXW7 has many substrates through which there are downstream effects. Interestingly, FBXW7 substrate CCNE2 (encodes for protein Cyclin E1), and Cdk2 and Cyclin E have previously been shown to induce HIV Tat activity by phosphorylation of the Ser16 residue, thus acting as a dependency factor [123-126]. It is possible that HIV-1 hijacks FBXW7 to sequester it, thus preventing or co-opting degradation of CCNE2 or one of its other substrates that have no previous implication for HIV-1 infection.

Additional HIV Dependency Factors

Of the genes chosen for validation studies, we validated 88% of these as dependency factors for at least one virus (Figure 8). Thus, as there are 72 genes in Tier 1 and 198 genes in Tier 2, if we use this percentage of validation rate as a rough estimate across the genes in Tier 1 and Tier 2, we estimate our CRISPR screen has, at minimum, identified at least more 156 HIV

dependency factors in addition to the 18 validated here (Supplemental Table 3). However, this is likely a minimum number as there are known HIV dependency factors that are below #198 (Supplemental Table 3) on the list. For example, known dependency factors for budding TSG101 (ranking at #299 with mean wt z score = -0.12) and CHMP4B (ranking at #147 with mean wt z score = -0.91) did not meet this cutoff. Thus, the absence of a gene in our top list is not evidence of its lack of effect. We find a negative correlation between z score and the magnitude of effect in spreading infections (Figure 10), in general, the higher the z score, the less likely lack of that gene product will affect HIV replication.

There are certain caveats to our screen results. First, the screen is done in Jurkat T cells, therefore, host genes that are more important in primary cells could be missed in our system. There was only a small overlap between our HIVDEP guide library and the library of interactors that were screened and validated in primary cells [75], however, we did identify some of the same genes , (including AFF1, which was validated in both studies) suggesting that our screens have identified factors relevant to primary CD4+ T cell infection (Supplemental Table 3). Of the previously mentioned RNAi screens and CRISPR screens, only one used the same cell type used in this study [96]. Yeung et al. 2009 used an shRNA whole genome screening approach, however, there were only 3 overlapping genes, implicating that there are likely differences in identification based on viruses used for each screen and/or significance criteria. HIVDEP was also designed based on a genome-wide screen using LAI, a CXCR4-tropic strain, therefore some genes that may only be important for another strain, could also have been excluded from the selection for genes included in the sublibrary HIVDEP, resulting in missing some relevant host factors. Future screening with HIVDEP in other HIV-permissive cell types, most importantly primary CD4+ T cells, may identify other host factors important for HIV-1 infection.

Alternatively, a new HIVDEP library can be generated based on other HIV-1 strains to potentially include more candidates based on strain-specific dependencies. Despite these limitations, we have identified and validated many new dependency factors for HIV-1, including factors that are important across clades, as well as many that are strain-specific. This list of host dependency factors can be used to further explore which pathways HIV-1 must hijack and has substantial implications for anti-HIV-1 therapeutic design and HIV cure strategies.

Chapter 4. Perspectives and Future Directions

Although very effective antiviral therapy against HIV has been developed, as the target of these drugs are viral proteins, there will always be the problem of evolution of the virus towards resistance. An alternative approach would be to target host proteins that are necessary for virus replication. My thesis work has identified a large number of possible candidates for this approach, as well as many interesting host proteins that can be studied for understanding the basic biology of HIV replication and its dependence on host factors.

Implications for druggable targets and polymorphisms in the human population

The approach for using host factors as druggable targets for HIV infection has been proposed previously for the entry receptors. One current example of this is the targeting of the receptor CD4 and co-receptor CCR5 for autologous or allogenic stem cell transplantation [127, 128] . Through robust mechanistic studies and the natural presence of a polymorphism within the human population that confers resistance to HIV, CCR5 has been an attractive host target. Individuals in the population that possess a 32 base-pair deletion in CCR5, resulting in truncation of the encoded protein that leads to decreased detection of the co-receptor at the cell surface and resistance to HIV infection [129-131]. The delta-32 deletion in CCR5 is present at about 10% frequency in populations originating from Northern latitudes with 1% frequency of homozygous delta-32. The alteration within a single allele (heterozygotes) confers resistance and slows the progression to AIDS, whereas those that have this deletion in both alleles (homozygotes) are resistant to HIV infection by CCR5-tropic strains, which are the predominantly circulating strains worldwide. Famously, Timothy Ray Brown, initially known as “The Berlin Patient,” was

the first person to be cured of HIV/AIDS. He received an allogenic stem cell transplantation as treatment for acute myeloid leukemia from a donor who was homozygous for the Δ32 mutation in CCR5 [127]. He continued his fight with cancer, and following another stem cell transplantation from the same donor remained in remission, but excitingly lived the rest of his life free of HIV/AIDS. Well-tolerated and safe CCR5 agonists, such as Maraviroc, have been effective against R5-tropic strains[132]. However, there are limitations to the efficacy as individuals co-infected with X4-tropic and R5-tropic strains can have an outrgrowth of X4-tropic strains and there is evidence of acquisition of mutations in the V3 loop of gp120, leading to drug resistance[132]. While CCR5 deletions are the best example of human polymorphisms that affect HIV infection, other rare polymorphisms have also been described that affect HIV infection in vitro, for example, a natural human polymorphism in TNP03 confers protection from HIV infection [133], while natural human polymorphisms in SAMHD1 confer increased sensitivity to HIV infection [134]. As human genetics outside of CCR5 and the HLA locus play a role in HIV progression (reviewed in [135]and [136]), there is the potential that polymorphisms in some of the genes identified here will impact HIV infections *in vivo.*

The more interesting genes identified here to follow-up as therapeutic targets will be either those identified to be important for all strains, or those that act on the CCR5 using strains such as Q23BG505. To follow up on these genes, one could use drugs or inhibitors to block activity of these genes or pathways. Small molecule inhibitors for some of them are already available, including some of the genes I identified and validated in my screen: UBE2M [137] and SLC39A7 [138]. Notably, MLN4924 (aka Pevonedistat) is a drug that is in Phase 3 clinical trials for patients with Higher-risk Myelodysplastic Syndromes (HR MDS), Chronic Myelomonocytic Leukemia (CMML), or Low-Blast Acute Myelogenous Leukemia (AML). This drug is a

regulator of the NAE (Nedd-activating enzyme), of which UBA3, another highly scoring host factor just upstream of validated factor UBE2M in the Neddylation cascade and could be a promising clinical drug for HIV-1 infection treatment. Next steps would be testing in animal models such as in the SHIV model of non-human primates [139] or the BLT mice model of HIV infection [140].

There are certain limitations that determine whether or not these can be used as druggable targets. First, these factors cannot be essential to the cell since those factors would also inhibit cellular mRNA transcription and be lethal to the cell. However, by nature of my knockout screen, cells with a gene whose knockout is acutely toxic would have already been screened out since they would not survive long enough to be infected in my protocol. Further mechanistic studies must be done to determine whether or not these criteria will be met for both the host factor and for the developed drugs.

Pathway-focused heatmaps and implications for HIV biology

In Chapter 3, I used pathway-focused heatmaps as a method to hone in on factors within a common pathway. If multiple factors scored well for a virus, this implicated a larger importance for a pathway or potentially further evidence of importance for multi-subunit complexes that HIV may hijack. These pathway-focused heatmaps were used to choose genes of interest for further validation studies in this thesis work.

However, there are additional pathways identified in the HIV dependency screen that were not discussed in Chapter 3. Of these pathways were the combined mRNA-related pathways, which included mRNA 3' processing, mRNA capping, and mRNA export (Figure 7). Of these factors, there are several that interact with the TREX/THOC mRNA export complex, which may

be of importance for mRNA export of transcripts for other host dependency factors, or recruitment of dependency factors for 3'end processing, such as CDK11 [141]. Other genes have been implicated before, including members of the TFIIH complex, which is important for RNA pol II transcription and is part of the pre-initiation complex at sites of transcription. Two of the ten subunits are shown in this heatmap: ERCC2 (aka XPD) an RNA helicase and the highest scoring gene in this heatmap and CDK7, a kinase important for regulating other kinases and the RNA pol II by phosphorylating the CTD of the cytoplasmic tail. Although CDK7 did not score well for any of the screens (mean z score 0.44), we know of it's biological importance during HIV infection[142]. XPD on the other hand has previously been implicated to act as a restriction factor for HIV by acting with XPB to degrade HIV cDNA transcripts [143], however, this gene was the top 20th scoring gene across wild type strains (mean wt z score: -3.15) and 43rd hit across all strains (mean overall z score: -3.56). It has been shown that XPD acts as a flexible tether for CDK7 to TFIIH, allowing for access to nearby substrates which includes CDK9 and RNA pol II [144]. Therefore, our data and these associations suggest that ERCC2 may need further studies to determine the impact on HIV-1 infection by this host factor.

Another pathway focused heatmap of the HIV dependency factors shows "regulation of signal transduction by p53 class mediator", GO: 1901796 (data not shown). AURKB (encoding for aurora kinase B), is the highest scoring gene in this heatmap (mean wt z score -2.72; mean overall z score: -3.11). Aurora kinases are difficult to study due to their roles in mitosis and meiosis, therefore knockout or knockdown may have cytotoxic effects on the cell. AURKB is important for condensation and separation of chromatids during cytokinesis (insert citation). Targeting of Aurora kinase B and other CPC complex proteins to the centrosomes and midsomes for regulation of chromosome separation is triggered by the CRL-3 ubiquitin ligase complex,

which contains the Cullin 3 protein of the cullin family discussed in chapter 3. Prior to 2018, there had been no role previously described for HIV biology. Nunes et al. reported that upon HIV infection of primary CD4+ T cells, Aurora kinase B expression was upregulated, however the mechanism by which this happens and why expression increases is still under investigation [145]. Through mass spectrometry-based proteomics in Jurkat cells, Johnson et al. showed that aurora kinase A and aurora kinase B protein abundance was decreased in the absence of Vpr. However, upon aurora kinase B inhibitor treatment of HIV-1 infected primary T cells, there was no impact of infection, though there was a reduction in cell viability [146]. Although knockout of this gene could cause growth defects in Jurkat T cells, HIVCRISPR screening purifies out lethal gene knockout by comparing enrichment of the guide targeting the gene of interest to what was packaged (through successful viral infection), and is in the viral supernatant. It is possible that the half-life of aurora kinase b is longer than the screening period, therefore further experiments would be required, possibly with siRNA knockdown, rather than CRISPR knockout, to determine the role of this gene for HIV infection.

The pathway focused heatmaps used and discussed here were created based on the Gene Set Enrichment Analysis (GSEA) of the genome-wide screen due to the unbiased set of genes in the genome-wide library, TKOv3. Although HIVDEP is skewed based on the most enriched genes of the genome wide screen and two other smaller screens with targeted libraries for epigenetic/epigenomic factors or interferon-stimulated genes, we can still use these enrichment analyses to rank the most enriched pathways per virus and visualize pathway-focused heatmaps based on enriched pathways within each screen. This can give a new perspective on which genes may or may not be strain-specific and can also be used to guide us in future validation studies.

Pros and Cons of HIV-CRISPR screening with focused CRISPR libraries

In my thesis work, I was able to identify many factors that both at the genome-wide screening scale and at the HIVDEP smaller scale, scored consistently high across multiple HIV-1 strains, demonstrating a common dependency across HIV-1 clades. HIV-CRISPR screening has many advantages that address and overcome challenges of other formatted CRISPR screen and siRNA screen designs. Firstly, using lentivirus as a delivery format for our guide libraries and the Cas9, we are able to screen in permissive cell types that are relevant to HIV infection. Another large advantage HIV-CRISPR screening has is the ability to survey hundreds or thousands of genes in a pooled fashion. The use of the guide RNAs as barcodes bypasses the need to separate each knockout (or in the case of siRNA screening, knock down lines). Therefore, we did not need robots to singly transfect clones within a 96 well plate and the downstream steps can also be streamlined to harvest, extract, amplify, and sequence the genetic material of the pelleted cells (containing the genomic DNA) or the viral supernatant (containing the packaged barcodes) altogether. For this reason, I could successfully screen at the large scale, genome-wide, within 2 T-75 flasks.

Using larger screen data (TKOv3), and other screens using focused libraries for human epigenome/ epigenetic factors (HuEpi) or interferon stimulated genes (PIKA), I designed a smaller targeted library for dependency factors. The strengths of the HIVCRISPR technology combined with the small size of HIVDEP (525 genes / 8 guides per gene), allowed me to screen with multiple viruses which use different co-receptors, are lab-adapted or primary strains, or are from different clades. Additionally, the with fewer guides to maintain proper coverage, each of these HIVDEP screens were performed in 2 wells of a 6-well plate, 1 per replicate.

Though this screening technique has many pros and has led to promising prospective dependency factors for multiple strains, there are also caveats to the screen design. For instance, there are some genes we expected to identify, like some of the nucleoporin genes identified in the siRNA screens, that did not score as dependency factors in our screen or scored below our cutoff. This could be because knocking out essential genes is lethal to the cell. The exception to this, and the reason we hypothesize some essential genes have still been identified in our screen, i.e. POL2RA (which encodes for subunit A of RNA polymerase II), is that some proteins may have a longer half-life than the puromycin selection period (10 days) and screening period (3 days of infection). In these cases, we expect these gene "knockouts" are closer to "knockdowns" because the functional mRNA production will have been abrogated, though there may be lingering protein expression from before the transduction.

Another reason some of the genes we expected to identify did not come out of our screens could be due to cell-type specificity. Many of the studies that have done validation experiments used 293Ts, TZM.bl, or other cell types. We used Jurkat cells in these screens, which are T lymphocytic cells, though may confer differences to other HIV-1 permissive cells, including 293T, TZM.Bl, primary T cells, monocytic lines (THP-1), primary macrophages, or other T cell lines including MT4 or SupT1 cells. We see this cell type specificity when we compare the dependency factors identified in OhAinle et al 2018, as compared to the screens in this dissertation. OhAinle et al. used THP-1 cells and the IFN stimulated gene focused library (PIKA), however used one of the same strains in this work, LAI. Although there were probably not many within the library, some positive controls CXCR4 and NFKB1 (also identified here), were identified, and novel factors SEC62, CD169, and TLR2 were also identified and validated. One of the screens used to create HIVDEP used PIKA, LAI, and Jurkat-R5 cells and these

factors did not fall within 10% FDR, therefore they were not included in the final gene list for HIVDEP. They also did not rank within the top 500 of the genome-wide screen. It is still possible that these are relevant in T cells, however, the effect may be below the level of detection we have for HIVCRISPR screening in Jurkats. Encouragingly, we did see some overlap of the validated dependency factors from Hiatt et al. 2022 found by CRISPR screening in primary T cells, suggesting that our screens have identified factors relevant to primary CD4+ T cell infection. Their library was limited to host factors known to interact with HIV proteins, however there was still some overlap between our screens and theirs, including AFF1, which was validated in both studies.

Two other CRISPR screening approaches could be applied for identifying HIV dependency factors, both of which have their own benefits and caveats. CRISPRi, also known as CRISPR inhibition, and CRISPRa, also known as CRISPR activation, both use catalytically dead Cas9 proteins that lack the cutting ability (dCas9), but are able to probe loss or gain of function in a pooled fashion, respectively. CRISPRi uses a fused dCas9 with a transcription repressor domain (for example, the Krüppel-associated box KRAB domain) and is recruited to the transcriptional start site (TSS) to repress transcription [147]. CRISPRa, in contrast, recruits multiple transcriptional activators to the TSS that are either fused to the dCas9, recruited to a protein scaffold fused to dCas9 and a single-chain variable fragment (scFv) targeting a GCN4 epitope, or to an RNA scaffold fused to the sgRNA, all with the end result of overexpressing the target endogenous genes [147]. As reviewed in [147], there are many advantages to CRISPRi/a individually and combinatorially. With CRISPRi, we can intentionally create a “knockdown” state with a higher efficiency than siRNA targeting [148] and in a pooled approach, rather than in a 96-well format, to assess loss of function in a reversible manner. The CRISPRa approach, on

the other hand, could tell us the opposing side, i.e. upon overexpression of the target genes in question, does HIV-1 infection increase (implicating this gene as a dependency factor), decrease (restriction or negative factor), or stay the same (no role in HIV biology). Importantly, we can use both CRISPRi and CRISPRa to interrogate complementary phenotypes for dependency factors for HIV— for example, overexpression of a gene encoding for one subunit of a large protein complex may not show an appreciable impact on HIV infection, however knockdown may inhibit the function of the complex. There are technically challenging hurdles to both approaches, including the fact that targeting a bidirectional promoter may affect neighboring gene expression [149]. To further support the idea that coupling CRISPRa/i screens can be informative, Schmidt et al. recently reported genome-wide CRISPRa and CRISPRi screens in primary T cells, revealing gene networks responsible for interleukin-2 and interferon- γ production [150]. Similar to HIV-CRISPR screening, the library and Cas9 was delivered via lentivirus. In this case, the lentivirus contained dCas9-VP64 (for CRISPRa) or dCas9-mCherry-KRAB (for CRISPRi). Additionally, after rounds of stimulation and passaging, the populations were restimulated and stained for FACS sorting prior to gDNA extraction. In the future, it may be possible to adapt and integrate the CRISPRi or CRISPRa approaches into the HIVCRISPR platform. This would bypass the FACS sorting limitations, which may include cell fragility and loss, and allow for maximal recovery for extractions. The field of CRISPR has rapidly evolved in the past decade and a half and the future of CRISPR screening for novel viral factors is promising.

Follow up experiments on validated genes in this thesis

Aside from following up on any of the other genes proposed in this study as potential dependency factors, there are genes specifically validated in this thesis in Jurkat cells. We challenged CRISPR knockout cells of 24 genes with either single-round infections with LAI or LAI-VSVg, or multi-round infections with LAI or Q23BG505. Nearly all of these genes validated, though some only validated with a strain-specific pattern, i.e. SBDS only validating with LAI psuedotyped with VSVg. Deeper understanding of how this factor regulates VSVg-mediated entry has the potential to improve lentiviral delivery platforms. We would also want to complement the gene expression back into these knockout lines using lentiviruses encoding the gene with nonsense mutations in the sgRNA target region or using CRISPR activation to measure upregulation of gene expression (discussed later in this section).

I would also determine where in the HIV lifecycle these dependency factors act. As an example, I will describe the approach we would use to uncover the mechanism of dependency for HIV-1 and UBE2M, one of our most potent HIV dependency factors in this study and one of two E2 ligases that conjugate NEDD8, a ubiquitin-like protein, also known as neddylation. Our screens identified genes encoding for proteins involved in both the neddylation process and substrate receptor specificity of three CRL complexes. The neddylation process begins when NEDD8 is adenylated and activated by the E1 activating heterodimer NAE (consisting of NAE1 and UBA3) [121]. Although NAE1 did not score high enough in the genome wide library to be included in the HIVDEP library, UBA3 scored extremely well across screens, including LAI-VSVg psuedotyped (mean wt z score: -4.258; mean z score of all: -4.61). Next, NEDD8, via NAE, is transferred to the E2 NEDD8-conjugating enzymes, UBE2M or UBE2F. UBE2F is known to neddylate CRL5, which is responsible for degrading potent restriction factor

APOBEC3G using HIV-1 Vif as a substrate receptor (Yu et al. 2003). UBE2F was included the genome-wide screen, however since this screen assesses dependency factors and knockout effects within the producer cell, rather than the target cell (which is where APOBEC3G would initiate cytidine deaminase activity for RT inhibition), we expect this is why UBE2F was not a hit. Notably, UBE2M, the other neddylating E2, was one of the top scoring hits across all screens, including the genome-wide screen and LAI-VSV-G, following essential dependency factors CD4 and CyclinT1 (mean wt z score CCNT1: -7.97; mean wt z score CD4: -7.42; mean wt z score UBE2M: -6.03).

Knockout of UBE2M dramatically decreased release of virus into the supernatant after multiple rounds of infection and decreased HIV-1 luciferase production in a single round infection assay. The fact that UBE2M knockout has an effect on single round infections where the readout is expression of a gene from the viral promoter rules out an effect of UBE2M on late events in the viral lifecycle such as virion assembly or budding. To investigate whether or not viral entry is dependent on UBE2M expression, we can use the β-lactamase Vpr (BlaM-Vpr) assay which assesses each step in the viral life cycle through the fusion step [151]. This assay measures fusion through the incorporation and fluorescence of the chimeric protein BlaM-Vpr (which fluoresces in the target cell post fusion and cleavage by host proteases), therefore, if binding and entry is inhibited, there won't be any emission. If entry is unaffected, we would use different assays to measure early RT products, measure nuclear entry of viral DNA, and linker-mediated PCR followed by next generation sequencing to monitor proviral integration into the genome. Another possible role for UBE2M in the viral lifecycle is at the level of transcription of the viral genome (or viral mRNAs). In this case, I would look at the effect of UBE2M knockouts on cells that already have an integrated provirus (to eliminate early events) and

measure viral mRNA levels both in knockout cells and in knockout cells where UBE2M has been re-introduced on inducible vectors.

UBE2M is a protein that can have a large impact on many cellular processes through activation the CRL1-4, therefore it is quite possible that we can see subtle effects at multiple stages of the life cycle due to effects of different substrates that get regulated through CRL-mediated degradation. Currently there are multiple small molecule inhibitors (including DI-591 [137]) that interrupt the interaction between UBE2M and DCN1, the E3 ligase used for CRL-3 complexes [152]. We can use one of these inhibitors, and others that block UBE2M-CRL interactions as they are developed and tested, to block individual arms of UBE2M-NEDD8-CRL activation and systematically measure the effects caused by different CRL and substrates UBE2M function contributes to. One would also like to know to what degree this is dependency is necessary for successful infection in primary T cells. With the publication of the large number of likely HIV dependency factors, my hope is that other labs will also follow up on additional experiments we can conduct to pinpoint where in the life cycle this factor acts and how HIV hijacks these proteins to gain a greater understanding of HIV biology.

Broad importance for understanding more about HIV biology

At the start of my thesis work, it appeared to me that we, as a field, had already arrived at a vast understanding of HIV, the HIV life cycle, and ways in which we can keep the virus at bay through pre-exposure prophylaxis (PrEP) and anti-retroviral therapies. I now understand that there are intricate, interconnected networks, pathways, and elegant modes of which HIV has adapted to use our own cellular proteins, beyond what we know as the HIV life cycle. We must continue to evolve our understanding of all the ways HIV is able to infect our cells, overcome

blocks (immune or even limited resource blocks intracellularly), and egress to infect a new cell if we are able to one day find a cure against this virus that continues to adapt, infect new hosts, and maintain itself in latent reservoirs ready to rebound. It is my hope that my thesis work can contribute to our basic understanding of HIV biology and the advancement towards a cure against HIV/AIDS.

References

1. Woolhouse, M.E.J. and L. Brierley, *Epidemiological characteristics of human-infective RNA viruses.* Sci Data, 2018. **5**: p. 180017.
2. Barre-Sinoussi, F., et al., *Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS).* Science, 1983. **220**(4599): p. 868-71.
3. Gallo, R.C., et al., *Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS).* Science, 1983. **220**(4599): p. 865-7.
4. Centers for Disease, C., *Kaposi's sarcoma and Pneumocystis pneumonia among homosexual men--New York City and California.* MMWR Morb Mortal Wkly Rep, 1981. **30**(25): p. 305-8.
5. Zhu, T., et al., *An African HIV-1 sequence from 1959 and implications for the origin of the epidemic.* Nature, 1998. **391**(6667): p. 594-7.
6. Gryseels, S., et al., *A near full-length HIV-1 genome from 1966 recovered from formalin-fixed paraffin-embedded tissue.* Proc Natl Acad Sci U S A, 2020. **117**(22): p. 12222-12229.
7. Worobey, M., et al., *Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960.* Nature, 2008. **455**(7213): p. 661-4.
8. van Heuvel, Y., et al., *Infectious RNA: Human Immunodeficiency Virus (HIV) Biology, Therapeutic Intervention, and the Quest for a Vaccine.* Toxins (Basel), 2022. **14**(2).
9. Maddon, P.J., et al., *The T4 gene encodes the AIDS virus receptor and is expressed in the immune system and the brain.* Cell, 1986. **47**(3): p. 333-48.
10. McDougal, J.S., et al., *Binding of HTLV-III/LAV to T4+ T cells by a complex of the 110K viral protein and the T4 molecule.* Science, 1986. **231**(4736): p. 382-5.

11. Arthos, J., et al., *HIV-1 envelope protein binds to and signals through integrin alpha4beta7, the gut mucosal homing receptor for peripheral T cells.* Nat Immunol, 2008. **9**(3): p. 301-9.

12. Cicala, C., et al., *The integrin alpha4beta7 forms a complex with cell-surface CD4 and defines a T-cell subset that is highly susceptible to infection by HIV-1.* Proc Natl Acad Sci U S A, 2009. **106**(49): p. 20877-82.

13. Saphire, A.C., et al., *Syndecans serve as attachment receptors for human immunodeficiency virus type 1 on macrophages.* J Virol, 2001. **75**(19): p. 9187-200.

14. Geijtenbeek, T.B., et al., *DC-SIGN, a dendritic cell-specific HIV-1-binding protein that enhances trans-infection of T cells.* Cell, 2000. **100**(5): p. 587-97.

15. Engelman, A. and P. Cherepanov, *The structural biology of HIV-1: mechanistic and therapeutic insights.* Nat Rev Microbiol, 2012. **10**(4): p. 279-90.

16. Vasiliver-Shamis, G., et al., *Human immunodeficiency virus type 1 envelope gp120-induced partial T-cell receptor signaling creates an F-actin-depleted zone in the virological synapse.* J Virol, 2009. **83**(21): p. 11341-55.

17. McDonald, D., et al., *Visualization of the intracellular behavior of HIV in living cells.* J Cell Biol, 2002. **159**(3): p. 441-52.

18. Dharan, A. and E.M. Campbell, *Role of Microtubules and Microtubule-Associated Proteins in HIV-1 Infection.* J Virol, 2018. **92**(16).

19. Carnes, S.K., J. Zhou, and C. Aiken, *HIV-1 Engages a Dynein-Dynactin-BICD2 Complex for Infection and Transport to the Nucleus.* J Virol, 2018. **92**(20).

20. Monroe, K.M., et al., *IFI16 DNA sensor is required for death of lymphoid CD4 T cells abortively infected with HIV.* Science, 2014. **343**(6169): p. 428-32.

21. Dharan, A., et al., *Nuclear pore blockade reveals that HIV-1 completes reverse transcription and uncoating in the nucleus.* Nat Microbiol, 2020. **5**(9): p. 1088-1095.

22. Burdick, R.C., et al., *HIV-1 uncoats in the nucleus near sites of integration.* Proc Natl Acad Sci U S A, 2020. **117**(10): p. 5486-5493.

23. Khan, A.U., et al., *Role of Nucleoporins and Transport Receptors in Cell Differentiation.* Front Physiol, 2020. **11**: p. 239.

24. Dharan, A., et al., *KIF5B and Nup358 Cooperatively Mediate the Nuclear Import of HIV-1 during Infection.* PLoS Pathog, 2016. **12**(6): p. e1005700.

25. Ocwieja, K.E., et al., *HIV integration targeting: a pathway involving Transportin-3 and the nuclear pore protein RanBP2.* PLoS Pathog, 2011. **7**(3): p. e1001313.

26. Zila, V., et al., *Cone-shaped HIV-1 capsids are transported through intact nuclear pores.* Cell, 2021. **184**(4): p. 1032-1046 e18.

27. De Iaco, A., et al., *TNPO3 protects HIV-1 replication from CPSF6-mediated capsid stabilization in the host cell cytoplasm.* Retrovirology, 2013. **10**: p. 20.

28. Maertens, G.N., et al., *Structural basis for nuclear import of splicing factors by human Transportin 3.* Proc Natl Acad Sci U S A, 2014. **111**(7): p. 2728-33.

29. Jang, S., et al., *Differential role for phosphorylation in alternative polyadenylation function versus nuclear import of SR-like protein CPSF6.* Nucleic Acids Res, 2019. **47**(9): p. 4663-4683.

30. Jacques, D.A., et al., *HIV-1 uses dynamic capsid pores to import nucleotides and fuel encapsidated DNA synthesis.* Nature, 2016. **536**(7616): p. 349-53.

31. Muller, T.G., et al., *HIV-1 uncoating by release of viral cDNA from capsid-like structures in the nucleus of infected cells.* Elife, 2021. **10**.

32. Muller, T.G., et al., *Nuclear Capsid Uncoating and Reverse Transcription of HIV-1*. Annu Rev Virol, 2022.

33. Hu, W.S. and S.H. Hughes, *HIV-1 reverse transcription.* Cold Spring Harb Perspect Med, 2012. **2**(10).

34. Nishitsuji, H., et al., *Augmentation of reverse transcription by integrase through an interaction with host factor, SIP1/Gemin2 Is critical for HIV-1 infection.* PLoS One, 2009. **4**(11): p. e7825.

35. Leng, J., et al., *A cell-intrinsic inhibitor of HIV-1 reverse transcription in CD4(+) T cells from elite controllers.* Cell Host Microbe, 2014. **15**(6): p. 717-728.

36. Achuthan, V., et al., *Capsid-CPSF6 Interaction Licenses Nuclear HIV-1 Trafficking to Sites of Viral DNA Integration.* Cell Host Microbe, 2018. **24**(3): p. 392-404 e8.

37. Lewinski, M.K., et al., *Retroviral DNA integration: viral and cellular determinants of target-site selection.* PLoS Pathog, 2006. **2**(6): p. e60.

38. Sowd, G.A., et al., *A critical role for alternative polyadenylation factor CPSF6 in targeting HIV-1 integration to transcriptionally active chromatin.* Proc Natl Acad Sci U S A, 2016. **113**(8): p. E1054-63.

39. Shun, M.C., et al., *LEDGF/p75 functions downstream from preintegration complex formation to effect gene-specific HIV-1 integration.* Genes Dev, 2007. **21**(14): p. 1767-78.

40. Singh, P.K., et al., *LEDGF/p75 interacts with mRNA splicing factors and targets HIV-1 integration to highly spliced genes.* Genes Dev, 2015. **29**(21): p. 2287-97.

41. Ruiz, A., et al., *Characterization of the influence of mediator complex in HIV-1 transcription.* J Biol Chem, 2014. **289**(40): p. 27665-76.

42. Konig, R., et al., *Global analysis of host-pathogen interactions that regulate early-stage HIV-1 replication.* Cell, 2008. **135**(1): p. 49-60.

43. Zhou, H., et al., *Genome-scale RNAi screen for host factors required for HIV replication.* Cell Host Microbe, 2008. **4**(5): p. 495-504.

44. Brass, A.L., et al., *Identification of host proteins required for HIV infection through a functional genomic screen.* Science, 2008. **319**(5865): p. 921-6.

45. Kao, S.Y., et al., *Anti-termination of transcription within the long terminal repeat of HIV-1 by tat gene product.* Nature, 1987. **330**(6147): p. 489-93.

46. Fujinaga, K., et al., *Dynamics of human immunodeficiency virus transcription: P-TEFb phosphorylates RD and dissociates negative effectors from the transactivation response element.* Mol Cell Biol, 2004. **24**(2): p. 787-95.

47. Wei, P., et al., *A novel CDK9-associated C-type cyclin interacts directly with HIV-1 Tat and mediates its high-affinity, loop-specific binding to TAR RNA.* Cell, 1998. **92**(4): p. 451-62.

48. Isel, C. and J. Karn, *Direct evidence that HIV-1 Tat stimulates RNA polymerase II carboxyl-terminal domain hyperphosphorylation during transcriptional elongation.* J Mol Biol, 1999. **290**(5): p. 929-41.

49. Cullen, B.R., *Nuclear mRNA export: insights from virology.* Trends Biochem Sci, 2003. **28**(8): p. 419-24.

50. Guerrero, S., et al., *HIV-1 replication and the cellular eukaryotic translation apparatus.* Viruses, 2015. **7**(1): p. 199-218.

51. Soto-Rifo, R., et al., *DEAD-box protein DDX3 associates with eIF4F to promote translation of selected mRNAs.* EMBO J, 2012. **31**(18): p. 3745-56.

52. Amorim, R., et al., *HIV-1 transcripts use IRES-initiation under conditions where Cap-dependent translation is restricted by poliovirus 2A protease.* PLoS One, 2014. **9**(2): p. e88619.

53. Ohlmann, T., C. Mengardi, and M. Lopez-Lastra, *Translation initiation of the HIV-1 mRNA.* Translation (Austin), 2014. **2**(2): p. e960242.

54. Kuzembayeva, M., et al., *Life of psi: how full-length HIV-1 RNAs become packaged genomes in the viral particles.* Virology, 2014. **454-455**: p. 362-70.

55. Onafuwa-Nuga, A. and A. Telesnitsky, *The remarkable frequency of human immunodeficiency virus type 1 genetic recombination.* Microbiol Mol Biol Rev, 2009. **73**(3): p. 451-80, Table of Contents.

56. Tang, Y., et al., *Cellular motor protein KIF-4 associates with retroviral Gag.* J Virol, 1999. **73**(12): p. 10508-13.

57. Abrahamyan, L.G., et al., *Novel Staufen1 ribonucleoproteins prevent formation of stress granules but favour encapsidation of HIV-1 genomic RNA.* J Cell Sci, 2010. **123**(Pt 3): p. 369-83.

58. Mouland, A.J., et al., *The double-stranded RNA-binding protein Staufen is incorporated in human immunodeficiency virus type 1: evidence for a role in genomic RNA encapsidation.* J Virol, 2000. **74**(12): p. 5441-51.

59. Chatel-Chaix, L., et al., *Identification of Staufen in the human immunodeficiency virus type 1 Gag ribonucleoprotein complex and a role in generating infectious viral particles.* Mol Cell Biol, 2004. **24**(7): p. 2637-48.

60. Mallery, D.L., et al., *Cellular IP6 Levels Limit HIV Production while Viruses that Cannot Efficiently Package IP6 Are Attenuated for Infection and Replication.* Cell Rep, 2019. **29**(12): p. 3983-3996 e4.

61. Dick, R.A., et al., *Inositol phosphates are assembly co-factors for HIV-1.* Nature, 2018. **560**(7719): p. 509-512.

62. Hurley, J.H. and A.K. Cada, *Inside job: how the ESCRTs release HIV-1 from infected cells.* Biochem Soc Trans, 2018. **46**(5): p. 1029-1036.

63. Kleinpeter, A.B. and E.O. Freed, *HIV-1 Maturation: Lessons Learned from Inhibitors.* Viruses, 2020. **12**(9).

64. Pornillos, O. and B.K. Ganser-Pornillos, *Maturation of retroviruses.* Curr Opin Virol, 2019. **36**: p. 47-55.

65. Napoli, C., C. Lemieux, and R. Jorgensen, *Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in trans.* Plant Cell, 1990. **2**(4): p. 279-289.

66. Fire, A., et al., *Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans.* Nature, 1998. **391**(6669): p. 806-11.

67. Kimpton, J. and M. Emerman, *Detection of replication-competent and pseudotyped human immunodeficiency virus with a sensitive cell line on the basis of activation of an integrated beta-galactosidase gene.* J Virol, 1992. **66**(4): p. 2232-9.

68. Bushman, F.D., et al., *Host cell factors in HIV replication: meta-analysis of genome-wide studies.* PLoS Pathog, 2009. **5**(5): p. e1000437.

69. Barrangou, R., et al., *CRISPR provides acquired resistance against viruses in prokaryotes.* Science, 2007. **315**(5819): p. 1709-12.

70. Park, R.J., et al., *A genome-wide CRISPR screen identifies a restricted set of HIV host dependency factors.* Nat Genet, 2017. **49**(2): p. 193-203.

71. Jager, S., et al., *Global landscape of HIV-human protein complexes.* Nature, 2011. **481**(7381): p. 365-70.

72. Jager, S., et al., *Vif hijacks CBF-beta to degrade APOBEC3G and promote HIV-1 infection.* Nature, 2011. **481**(7381): p. 371-5.

73. Faust, T.B., et al., *The HIV-1 Tat protein recruits a ubiquitin ligase to reorganize the 7SK snRNP for transcriptional activation.* Elife, 2018. **7**.

74. Chen, S.H., et al., *CRL4(AMBRA1) targets Elongin C for ubiquitination and degradation to modulate CRL5 signaling.* EMBO J, 2018. **37**(18).

75. Hiatt, J., et al., *A functional map of HIV-host interactions in primary human T cells.* Nat Commun, 2022. **13**(1): p. 1752.

76. Fu, C., et al., *Human Gene Functional Network-Informed Prediction of HIV-1 Host Dependency Factors.* mSystems, 2020. **5**(6).

77. Molly Ohainle, L.P., Jolien Vermiere, Ferdinand Roesch, Daryl Humes, Ryan Basom, Jeffrey J. Delrow, Julie Overbaugh, Michael Emerman, *A Virus-Packageable CRISPR Screen Identifies Host Factors Mediating Interferon Inhibition of HIV.* elife, 2018.

78. OhAinle, M., et al., *A virus-packageable CRISPR screen identifies host factors mediating interferon inhibition of HIV.* Elife, 2018. **7**.

79. Puryear, W.B., et al., *Interferon-inducible mechanism of dendritic cell-mediated HIV-1 dissemination is dependent on Siglec-1/CD169.* PLoS Pathog, 2013. **9**(4): p. e1003291.

80. Gummuluru, S., C.M. Kinsey, and M. Emerman, *An in vitro rapid-turnover assay for human immunodeficiency virus type 1 replication selects for cell-to-cell spread of virus.* J Virol, 2000. **74**(23): p. 10882-91.

81. Bartz, S.R. and M.A. Vodicka, *Production of high-titer human immunodeficiency virus type 1 pseudotyped with vesicular stomatitis virus glycoprotein.* Methods, 1997. **12**(4): p. 337-42.

82. Haddox, H.K., et al., *Mapping mutational effects along the evolutionary landscape of HIV envelope.* Elife, 2018. **7**.

83. Poss, M. and J. Overbaugh, *Variants from the diverse virus population identified at seroconversion of a clade A human immunodeficiency virus type 1-infected woman have distinct biological properties.* J Virol, 1999. **73**(7): p. 5255-64.

84. Fenton-May, A.E., et al., *Relative resistance of HIV-1 founder viruses to control by interferon-alpha.* Retrovirology, 2013. **10**: p. 146.

85. Hart, T., et al., *Evaluation and Design of Genome-Wide CRISPR/SpCas9 Knockout Screens.* G3 (Bethesda), 2017. **7**(8): p. 2719-2727.

86. Wang, T., et al., *Identification and characterization of essential genes in the human genome.* Science, 2015. **350**(6264): p. 1096-101.

87. Meier, J.A., F. Zhang, and N.E. Sanjana, *GUIDES: sgRNA design for loss-of-function screens.* Nat Methods, 2017. **14**(9): p. 831-832.

88. Labun, K., et al., *CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing.* Nucleic Acids Res, 2019. **47**(W1): p. W171-W174.

89. Roesch, F. and M. OhAinle, *HIV-CRISPR: A CRISPR/Cas9 Screening Method to Identify Genes Affecting HIV Replication.* Bio Protoc, 2020. **10**(9): p. e3614.

90. Wang, B., et al., *Integrative analysis of pooled CRISPR genetic screens using MAGeCKFlute.* Nat Protoc, 2019. **14**(3): p. 756-780.

91. Langmead, B., et al., *Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.* Genome Biol, 2009. **10**(3): p. R25.

92. Subramanian, A., et al., *Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.* Proc Natl Acad Sci U S A, 2005. **102**(43): p. 15545-50.

93. Schneider, W.M., et al., *Genome-Scale Identification of SARS-CoV-2 and Pan-coronavirus Host Factor Networks.* Cell, 2021. **184**(1): p. 120-132 e14.

94. Vermeire, J., et al., *Quantification of reverse transcriptase activity by real-time PCR as a fast and accurate method for titration of HIV, lenti- and retroviral vectors.* PLoS One, 2012. **7**(12): p. e50859.

95. Conant, D., et al., *Inference of CRISPR Edits from Sanger Trace Data.* CRISPR J, 2022. **5**(1): p. 123-130.

96. Yeung, M.L., et al., *A genome-wide short hairpin RNA screening of jurkat T-cells for human proteins contributing to productive HIV-1 replication.* J Biol Chem, 2009. **284**(29): p. 19463-73.

97. Gordon, D.E., et al., *A Quantitative Genetic Interaction Map of HIV Infection.* Mol Cell, 2020. **78**(2): p. 197-209 e7.

98. Ohainle, M., et al., *TRIM34 restricts HIV-1 and SIV capsids in a TRIM5alpha-dependent manner.* PLoS Pathog, 2020. **16**(4): p. e1008507.

99. Li, W., et al., *MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens.* Genome Biol, 2014. **15**(12): p. 554.

100. Hsieh, E., et al., *A modular CRISPR screen identifies individual and combination pathways contributing to HIV-1 latency.* 2022.

101. Parrish, N.F., et al., *Phenotypic properties of transmitted founder HIV-1.* Proc Natl Acad Sci U S A, 2013. **110**(17): p. 6626-33.

102. Smith, J.R., et al., *Robust, persistent transgene expression in human embryonic stem cells is achieved with AAVS1-targeted integration.* Stem Cells, 2008. **26**(2): p. 496-504.

103. Hockemeyer, D., et al., *Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases.* Nat Biotechnol, 2009. **27**(9): p. 851-7.

104. Chu, V.T., et al., *Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells.* Nat Biotechnol, 2015. **33**(5): p. 543-8.

105. Llano, M., et al., *An essential role for LEDGF/p75 in HIV integration.* Science, 2006. **314**(5798): p. 461-4.

106. Liu, R., et al., *ELL-associated factors EAF1/2 negatively regulate HIV-1 transcription through inhibition of Super Elongation Complex formation.* Biochim Biophys Acta Gene Regul Mech, 2020. **1863**(5): p. 194508.

107. He, N., et al., *HIV-1 Tat and host AFF4 recruit two transcription elongation factors into a bifunctional complex for coordinated activation of HIV-1 transcription.* Mol Cell, 2010. **38**(3): p. 428-38.

108. Marzio, G., et al., *HIV-1 tat transactivator recruits p300 and CREB-binding protein histone acetyltransferases to the viral promoter.* Proc Natl Acad Sci U S A, 1998. **95**(23): p. 13519-24.

109. Deng, L., et al., *Acetylation of HIV-1 Tat by CBP/P300 increases transcription of integrated HIV-1 genome and enhances binding to core histones.* Virology, 2000. **277**(2): p. 278-95.

110. Ott, M., et al., *Acetylation of the HIV-1 Tat protein by p300 is important for its transcriptional activity.* Curr Biol, 1999. **9**(24): p. 1489-92.

111. Liu, L., et al., *A whole genome screen for HIV restriction factors.* Retrovirology, 2011. **8**: p. 94.

112. Van Duyne, R., et al., *Lysine methylation of HIV-1 Tat regulates transcriptional activity of the viral LTR.* Retrovirology, 2008. **5**: p. 40-40.

113. Liu, Z., et al., *SETDB1 interactions with PELP1 contributes to breast cancer endocrine therapy resistance.* Breast Cancer Res, 2022. **24**(1): p. 26.

114. Gupta, S., et al., *Emerging role of ZBTB7A as an oncogenic driver and transcriptional repressor.* Cancer Lett, 2020. **483**: p. 22-34.

115. Pendergrast, P.S., et al., *FBI-1 can stimulate HIV-1 Tat activity and is targeted to a novel subnuclear domain that includes the Tat-P-TEFb-containing nuclear speckles.* Mol Biol Cell, 2002. **13**(3): p. 915-29.

116. Pessler, F. and N. Hernandez, *Flexible DNA binding of the BTB/POZ-domain protein FBI-1.* J Biol Chem, 2003. **278**(31): p. 29327-35.

117. Zhu, X., et al., *ZBTB7A promotes virus-host homeostasis during human coronavirus 229E infection.* Cell Rep, 2022. **41**(4): p. 111540.

118. Wang, L.H., et al., *The MLL3/4 H3K4 methyltransferase complex in establishing an active enhancer landscape.* Biochem Soc Trans, 2021. **49**(3): p. 1041-1054.

119. Soucy, T.A., et al., *An inhibitor of NEDD8-activating enzyme as a new approach to treat cancer*. Nature, 2009. **458**(7239): p. 732-6.

120. Nekorchuk, M.D., et al., *HIV relies on neddylation for ubiquitin ligase-mediated functions.* Retrovirology, 2013. **10**: p. 138.

121. Walden, H., et al., *The structure of the APPBP1-UBA3-NEDD8-ATP complex reveals the basis for selective ubiquitin-like protein activation by an E1.* Mol Cell, 2003. **12**(6): p. 1427-37.

122. Thirimanne, H.N., et al., *Global and context-specific transcriptional consequences of oncogenic Fbw7 mutations.* Elife, 2022. **11**.

123. Strohmaier, H., et al., *Human F-box protein hCdc4 targets cyclin E for proteolysis and is mutated in a breast cancer cell line.* Nature, 2001. **413**(6853): p. 316-22.

124. Ivanov, A., et al., *HIV-1 Tat phosphorylation on Ser-16 residue modulates HIV-1 transcription.* Retrovirology, 2018. **15**(1): p. 39.

125. Deng, L., et al., *HIV-1 Tat interaction with RNA polymerase II C-terminal domain (CTD) and a dynamic association with CDK2 induce CTD phosphorylation and transcription from HIV-1 promoter.* J Biol Chem, 2002. **277**(37): p. 33922-9.

126. Nekhai, S., et al., *HIV-1 Tat-associated RNA polymerase C-terminal domain kinase, CDK2, phosphorylates CDK7 and stimulates Tat-mediated transcription.* Biochem J, 2002. **364**(Pt 3): p. 649-57.

127. Hutter, G., et al., *Long-term control of HIV by CCR5 Delta32/Delta32 stem-cell transplantation.* N Engl J Med, 2009. **360**(7): p. 692-8.

128. Khalid, K., et al., *Stem Cell Therapy and Its Significance in HIV Infection.* Cureus, 2021. **13**(8): p. e17507.

129. Liu, R., et al., *Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection.* Cell, 1996. **86**(3): p. 367-77.

130. Dean, M., et al., *Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CKR5 structural gene. Hemophilia Growth and Development Study, Multicenter AIDS Cohort Study, Multicenter Hemophilia Cohort Study, San Francisco City Cohort, ALIVE Study.* Science, 1996. **273**(5283): p. 1856-62.

131. Samson, M., et al., *Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene.* Nature, 1996. **382**(6593): p. 722-5.

132. Woollard, S.M. and G.D. Kanmogne, *Maraviroc: a review of its use in HIV infection and beyond.* Drug Des Devel Ther, 2015. **9**: p. 5447-68.

133. Rodriguez-Mora, S., et al., *The mutation of Transportin 3 gene that causes limb girdle muscular dystrophy 1F induces protection against HIV-1 infection.* PLoS Pathog, 2019. **15**(8): p. e1007958.

134. Berger, A., et al., *SAMHD1-deficient CD14+ cells from individuals with Aicardi-Goutieres syndrome are highly susceptible to HIV-1 infection.* PLoS Pathog, 2011. **7**(12): p. e1002425.

135. McLaren, P.J., et al., *Polymorphisms of large effect explain the majority of the host genetic contribution to variation of HIV-1 virus load.* Proc Natl Acad Sci U S A, 2015. **112**(47): p. 14658-63.

136. McLaren, P.J. and M. Carrington, *The impact of host genetic variation on infection with HIV-1.* Nat Immunol, 2015. **16**(6): p. 577-83.

137. Zhou, H., et al., *A potent small-molecule inhibitor of the DCN1-UBC12 interaction that selectively blocks cullin 3 neddylation.* Nat Commun, 2017. **8**(1): p. 1150.

138. Nolin, E., et al., *Discovery of a ZIP7 inhibitor from a Notch pathway screen.* Nat Chem Biol, 2019. **15**(2): p. 179-188.

139. Del Prete, G.Q., J.D. Lifson, and B.F. Keele, *Nonhuman primate models for the evaluation of HIV-1 preventive vaccine strategies: model parameter considerations and consequences.* Curr Opin HIV AIDS, 2016. **11**(6): p. 546-554.

140. Karpel, M.E., C.L. Boutwell, and T.M. Allen, *BLT humanized mice as a small animal model of HIV infection.* Curr Opin Virol, 2015. **13**: p. 75-80.

141. Pak, V., et al., *CDK11 in TREX/THOC Regulates HIV mRNA 3' End Processing.* Cell Host Microbe, 2015. **18**(5): p. 560-70.

142. Rice, A.P., *Roles of CDKs in RNA polymerase II transcription of the HIV-1 genome.* Transcription, 2019. **10**(2): p. 111-117.

143. Yoder, K., et al., *The DNA repair genes XPB and XPD defend cells from retroviral infection.* Proc Natl Acad Sci U S A, 2006. **103**(12): p. 4622-7.

144. Rimel, J.K., et al., *Selective inhibition of CDK7 reveals high-confidence targets and new models for TFIIH function in transcription.* Genes Dev, 2020. **34**(21-22): p. 1452-1473.

145. Nunes, J.M., et al., *Modulation of epigenetic factors during the early stages of HIV-1 infection in CD4(+) T cells in vitro.* Virology, 2018. **523**: p. 41-51.

146. Johnson, J.R., et al., *Global post-translational modification profiling of HIV-1-infected cells reveals mechanisms of host cellular pathway remodeling.* Cell Rep, 2022. **39**(2): p. 110690.

147. Kampmann, M., *CRISPRi and CRISPRa Screens in Mammalian Cells for Precision Biology and Medicine.* ACS Chem Biol, 2018. **13**(2): p. 406-416.

148. Gilbert, L.A., et al., *CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes.* Cell, 2013. **154**(2): p. 442-51.

149. Rosenbluh, J., et al., *Complementary information derived from CRISPR Cas9 mediated gene deletion and suppression.* Nat Commun, 2017. **8**: p. 15403.

150. Schmidt, R., et al., *CRISPR activation and interference screens decode stimulation responses in primary human T cells.* Science, 2022. **375**(6580): p. eabj4008.

151. Cavrois, M., J. Neidleman, and W.C. Greene, *HIV-1 Fusion Assay.* Bio Protoc, 2014. **4**(16).

152. Fang, Y., et al., *Targeting the DCN1-UBC12 protein-protein interaction: novel approaches and future directions.* Future Med Chem, 2019. **11**(22): p. 2869-2871.