

UCSF

UC San Francisco Electronic Theses and Dissertations

Title

Systematic investigation of host-pathogen interactions identifies novel drug targets for HIV and SARS-CoV-2

Permalink

<https://escholarship.org/uc/item/84x302rk>

Author

Haas, Paige

Publication Date

2021

Supplemental Material

<https://escholarship.org/uc/item/84x302rk#supplemental>

Peer reviewed|Thesis/dissertation

Systematic investigation of host-pathogen interactions identifies novel drug targets for
HIV and SARS-CoV-2

by
Paige Haas

DISSERTATION

Submitted in partial satisfaction of the requirements for degree of
DOCTOR OF PHILOSOPHY

in

Biochemistry and Molecular Biology

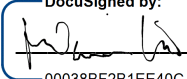
in the

GRADUATE DIVISION

of the

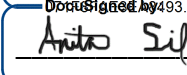
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

DocuSigned by:

00038BF2B1EE40C... Melanie Ott
Chair

DocuSigned by:

493... Nevan Krogan

DocuSigned by:

9BCB4DFEE8384D2... Anita Sil

Committee Members

ACKNOWLEDGEMENTS

Thank you to my advisor, Nevan Krogan, for the boundless opportunities I have enjoyed in your lab. You have supported my growth as a scientist and instilled a sense of adventure into my research. I have learned so much from your optimism, ambition, and sense of possibility, and I am grateful for your belief in me and support of my goals.

Thank you to my mentor, Robyn Kaake, for your steadfast encouragement and consistent support. Our weekly meetings were an essential place to organize my thoughts and get advice on both the scientific and personal challenges of graduate school, and the completion of my PhD would not have been possible without you. I have greatly enjoyed and benefited from your enthusiastic support, sincere love of science, and I feel fortunate to call you a friend.

Thank you to the members of my thesis committee, Anita Sil and Melanie Ott. Your thoughtful advice helped me grow scientifically and as a person. I want to thank you both for supporting my career goals without hesitation, tailoring my training, and celebrating my successes.

Thank you to my qualifying exam committee, Anita Sil, John Gross, Alexander Marson, and Danica Fujimori, for your sound input that shaped my project and refined my thinking as a scientist.

Thank you to the UCSF staff who have supported my journey through graduate school, especially Toni Hurley and Justin Gibson for your perspective and advice. Our conversations helped me persevere during graduate school and will continue to benefit me beyond my time at UCSF.

Thank you, Judd Hultquist, for your mentorship, especially during the conceptualization of my thesis project, my early technical training in the Krogan lab, and providing feedback that shaped my perspective as a scientist.

Thank you to all of the members of the Krogan lab for fostering an environment that was a joy to work in - dynamic, humorous, creative, with an abundance of mutual respect. I especially want to thank Michael McGregor for answering my many questions with calm expertise, and Margaret Soucheray who has saved my experiments more times than I can count. I want to thank Ernst Pulido for chats that brightened my day during weekends in lab, and Joe Hiatt for providing enthusiastic help and making long TC days fun. Thank you Beril Tutuncuoglu and Devin Cavero for fun conversations in and outside of the lab. Thank you Ujjwal Rathore and Vigneshwari Kumar for your kindness and skill that was a joy to work with, and Danielle Swaney and Erica Stevenson for your generosity with your time and expertise. Thank you, Lorena Zuliani-Alvarez, for providing key perspective, and Mehdi Bouhaddou for always being excited to help a student succeed.

Thank you to all of the members of the *Biologists Being Basic* podcast for creating an incredible opportunity to explore scientific communication and creativity. Our conversations were a highlight of my time in graduate school. A special thank you to Gina Nguyen and Alexa Rocourt for your media expertise that took our scientific communication to another level.

Thank you to my classmates in the Tetrad cohort of 2016, I feel honored to be part of such a talented group of scientists. I especially want to thank Adam Longhurst and Hannah Toutkoushian for your camaraderie and sense of humor that kept me laughing during graduate school. Your honesty and generosity helped me realize that I was not alone and gave me the courage to keep trying in hard times.

Thank you to my friends who I can always count on to cheer me on and cheer me up. Thank you Francesca Liburdy, Erin Deffenbaugh, and Laurel Callia for being a piece of home in San Francisco when you have come to visit. You keep me positive and help remind me of all the best things in life. Thank you, Lauren Dennis, for your excellent advice and endless encouragement that makes me feel seen and supported. Thank you, Andrea Chon, for our many phone calls that remind me of who I am when I have lost my way. Thank you, Rachel Lucero, for the great food, game nights, and keeping me inspired to be the best version of myself. And thank you Melanie Cernak for our life chats and adventures, I admire and benefit from your enduring optimism.

Thank you to my parents, Randy and Ann Haas, for your steady encouragement and support that allows me to pursue my dreams. It is powerful to know you are behind me during the different chapters of my life.

Thank you to my partner, Sean Minor, for believing in me before I could see it for myself, and for being proud of me before I felt I deserved it. Our partnership has been a place of comfort, happiness, and ever-expanding possibility, and sharing life with you is better than I could have imagined. I also want to thank our dog, Manchego, for the love and joy she's brought into our lives.

Thank you to my sister, Kelsey Haas - there aren't enough words. Thank you for being my role model, my best friend, and for helping me see how much I am capable of. Thank you for mentoring me as a scientist and for supporting me as my sister. We have weathered a lot together while in graduate school and shared plenty of good times too. I will always look back on the experience of earning my PhD and feel lucky that I got to do it with you.

CONTRIBUTIONS

The work presented in this dissertation was performed under the supervision and guidance of Nevan Krogan and Robyn Kaake. Additional guidance and insight were provided by thesis committee members Melanie Ott and Anita Sil.

Chapter 2 is a reprint of a *Journal of Proteome Research* perspective:

Reprinted with permission from Haas P, Muralidharan M, Krogan NJ, Kaake RM, Hüttenhain R. Proteomic Approaches to Study SARS-CoV-2 Biology and COVID-19 Pathology. *J Proteome Res.* 2021 Feb 5;20(2):1133-1152. doi: 10.1021/acs.jproteome.0c00764. Epub 2021 Jan 19. PMID: 33464917; PMCID: PMC7839417. Copyright 2021 American Chemical Society.

Chapter 3 is a manuscript in preparation:

Haas P, Rathore U, Kumar V, Hiatt J, Haas K, Bouhaddou M, Swaney DL, Stevenson E, Zuliani-Alvarez L, McGregor MJ, Braberg H, Soucheray M, Ott M, Hultquist JF, Marson A, Kaake RM, Krogan NJ. Genetic analysis of E3 ubiquitin ligases using CRISPR-Cas9 in primary human T cells identifies TRAF2 and UHRF1 as regulators of HIV infection and latency reversal. 2021. *Manuscript in preparation.*

Systematic investigation of host-pathogen interactions identifies novel drug targets for HIV and SARS-CoV-2

Paige Haas

ABSTRACT

HIV and SARS-CoV-2 are two major viral pathogens of the past 50 years, causing a global HIV/AIDS epidemic in the 1980s and a global COVID-19 pandemic in 2020, respectively. Both the HIV epidemic and COVID-19 pandemic are currently ongoing, and both viruses have infected millions of people with devastating consequences including long-term health effects and death. Like all viruses, HIV and SARS-CoV-2 rely upon host cell machinery to replicate, and characterizing these host-pathogen interactions informs our understanding of virology, pathology, and therapeutics. Host factors identified as critical for viral infection can serve as candidate drug targets for host-directed therapy (HDT), and novel drugs are desperately needed for both HIV and SARS-CoV-2. While anti-retroviral therapy (ART) halts the progression of HIV to AIDS, it is not curative, and ART treatment must be lifelong. While there are effective vaccines against SARS-CoV-2, therapeutics are needed for both unvaccinated and breakthrough infections. HDT, or drugging host factors instead of viral proteins, is an appealing strategy to treat both HIV and SARS-CoV-2 because it limits the potential for viral escape mutations. Systems biology approaches are promising tools to identify drug targets for HDT, as they capture a global and unbiased picture of infection. Here, we employ systems biology approaches to identify host factors important for SARS-CoV-2 and HIV infection that can serve as HDT candidates. In **Chapter 2**, we propose proteomic approaches to inform SARS-CoV-2 virology, COVID-19 pathology, and therapeutic strategies to combat the current pandemic. Specifically, we discuss how (1) mass spectrometry-based structural techniques can overcome limitations and complement traditional structural approaches to inform the dynamic structure of SARS-CoV-2 proteins, complexes, and virions; (2) virus–host protein–protein interaction

mapping can identify the cellular machinery required for SARS-CoV-2 replication; (3) global protein abundance and post-translational modification profiling can characterize signaling pathways that are rewired during infection; and (4) proteomic technologies can aid in biomarker identification, diagnostics, and drug development in order to monitor COVID-19 pathology and investigate treatment strategies. In **Chapter 3**, we combine proteomic and genetic systems biology approaches to identify E3 ubiquitin ligases critical for HIV infection and latency reversal. We use proteomics to identify 116 E3 ubiquitin ligases that are expressed in physiologically-relevant primary human CD4⁺ T cells, knock them out with CRISPR-Cas9, and identify 10 E3s that affect HIV infection: TRAF2, TRAF3, and PRPF19 have antiviral activity, and MARCH5, ZFP91, UHRF1, VPS18, NOSIP, PPIL2, RING1 have proviral activity. Importantly, we find that TRAF2 knockouts reverse latency in JLat and primary cell models, thus identifying TRAF2 as a candidate HDT drug target for “shock and kill” therapy towards an HIV cure. **Chapter 4** discusses the power of global, unbiased systems biology approaches to elucidate host-pathogen interactions that are critical for understanding the virology of and developing therapeutics for these two major pathogens of the past 50 years.

TABLE OF CONTENTS

Chapter 1: Introduction.....	1
Chapter 2: Proteomic approaches to study SARS-CoV-2 biology and COVID-19 pathology.....	7
Summary.....	8
Introduction.....	8
How do the dynamic structures of SARS-CoV-2 proteins and virions inform pathogenicity?.....	11
Cross-linking mass spectrometry (XL-MS).....	12
Hydrogen/deuterium exchange mass spectrometry (H/DX-MS).....	15
Intact protein mass spectrometry.....	16
Integrating complementary structural data.....	18
What cellular machinery does SARS-CoV-2 utilize during infection?.....	20
Affinity purification–mass spectrometry (AP-MS).....	20
Proximity-dependent labeling (PDL).....	22
Complex centric proteome profiling (CCPP).....	23
Toward an integrated virus–host PPI network.....	25
Which signaling pathways are rewired during SARS-CoV-2 infection?.....	26
Global phosphoproteomics.....	27
Global ubiquitylation profiling.....	28
Global protein abundance profiling and integrating signaling data.....	30
What are strategies to monitor COVID-19 pathology and investigate treatment strategies?.....	32
Targeted proteomics for monitoring disease progression and treatment response.....	32
Limited proteolysis-mass spectrometry (LiP-MS).....	34
Thermal proteome profiling (TPP).....	35

Activity-based protein profiling (ABPP).....	36
Prioritizing candidates from large-scale SARS-CoV-2 drug screens.....	38
Conclusion.....	38
Chapter 3: Genetic analysis of E3 ubiquitin ligases using CRISPR-Cas9 in primary human T cells identifies TRAF2 and UHRF1 as regulators of HIV infection and latency.....	48
Summary.....	49
Introduction.....	49
Results.....	53
Liquid chromatography tandem mass spectrometry (LC-MS/MS) analysis identifies 116 E3 ligases expressed in activated primary human CD4 ⁺ T cells...53	
CRISPR-Cas9 knockouts in primary CD4 ⁺ T cells identify novel E3s that regulate HIV infection.....	54
Network propagation analysis connects TRAF2, non-canonical NF-κB signaling, and HIV infection.....	58
Deletion of TRAF2 and UHRF1 reverses HIV latency in JLat and primary resting CD4 ⁺ T cell models.....	60
Discussion.....	64
Methods.....	68
Chapter 4: Discussion.....	97
References.....	100

LIST OF FIGURES

Figure 2.1 SARS-CoV-2 life cycle, open questions, and proteomic techniques.....	40
Figure 2.2 Overview of MS-based proteomics techniques proposed to study SARS-CoV-2....	41
Figure 3.1 Liquid chromatography tandem mass spectrometry (LC-MS/MS) analysis identifies 116 E3 ligases expressed in activated primary human CD4 ⁺ T cells.....	83
Figure 3.2 CRISPR-Cas9 knockouts in primary CD4 ⁺ T cells identify novel E3s that regulate HIV infection.....	84
Figure 3.3 Network propagation analysis connects TRAF2, non-canonical NF-κB signaling, and HIV infection.....	85
Figure 3.4 TRAF2 and UHRF1 knockouts reverse latency in JLat and resting primary human CD4 ⁺ T cell models.....	86
Figure 3.5 Proposed mechanism by which TRAF2 and UHRF1 transcriptionally reactivate the latent HIV-1 promoter.....	87
Figure S3.1 , related to Figure 3.1.....	88
Figure S3.2 , related to Figure 3.2.....	89
Figure S3.3 , related to Figure 3.3.....	90
Figure S3.4 , related to Figure 3.3.....	91
Figure S3.5 , related to Figure 3.3.....	92
Figure S3.6 , related to Figure 3.4.....	93
Figure S3.7 , related to Figure 3.4.....	94
Figure S3.8 Potential roles for 10 E3 hits in HIV infection.....	95

LIST OF TABLES

Table 2.1 Proteomic studies on SARS-CoV-2 highlighted in the perspective.....	42
Table S3.1 , related to Figure 3.1. Proteins identified by LC-MS/MS analysis of primary activated CD4 ⁺ T cells.....	96
Table S3.2 , related to Figure 3.2. CRISPR guides used in the CRISPR-Cas9 spreading HIV-1 infection assay.....	96
Table S3.3 , related to Figure 3.2. Raw and analyzed data from CRISPR-Cas9 knockout spreading HIV-1 infection assay.....	96
Table S3.4 , related to Figure 3.3. Network propagation of HIV pathways and E3 ubiquitin ligases that functionally affected HIV infection in our CRISPR-Cas9 knockout spreading HIV-1 infection assay.....	96
Table S3.5 , related to Figure 3.4. Raw and analyzed data from CRISPR-Cas9 knockouts in JLat models of HIV latency.....	96
Table S3.6 , related to Figure 3.4. Raw and analyzed data from CRISPR-Cas9 knockouts in a resting primary human CD4 ⁺ T cell model of HIV latency.....	96

LIST OF ABBREVIATIONS

A3	APOBEC3
AAV2	adeno-associated virus 2
ABPP	activity-based protein profiling
AIDS	acquired immunodeficiency syndrome
AP-MS	affinity purification–mass spectrometry
ART	antiretroviral therapy
CCPP	complex centric proteome profiling
CETSA	cellular thermal shift assay
CID	collision-induced dissociation
COVID-19	coronavirus disease 2019
crRNP	CRISPR ribonucleoprotein
cryo-EM	cryo-electron microscopy
cryo-ET	cryo-electron tomography
DIA	data-independent acquisition
DMF	dimethyl fumarate
DTSSP	3,3'-dithiobis(sulfo-succinimidylpropionate)
DUBs	deubiquitylating enzymes
E	envelope protein
ECD	electron-capture dissociation
ER	endoplasmic reticulum
ERGIC	ER–Golgi intermediate compartment
ETD	electron transfer dissociation
FDA	US Food and Drug Administration
GFR	growth factor receptor
GO	gene ontology

GPCRs	G-protein coupled receptors
HDT	host-directed therapy
H/DX-MS	hydrogen/deuterium exchange mass spectrometry
HECT	human thyroid receptor interacting protein
HIV	human immunodeficiency virus
IFN	interferon
IMMS	ion mobility mass spectrometry
JLat	Jurkat cell line model of HIV latency
L2FC	\log_2 fold change
LC-MS/MS	liquid chromatography tandem mass spectrometry
LiP-MS	limited proteolysis–coupled mass spectrometry
LiP-SMap	limited proteolysis–small molecule mapping
LiP	limited proteolysis
LRA	latency-reversing agent
LTR	long terminal repeat
M-PMV	Mason-Pfizer monkey virus
M	membrane protein
mAbs	monoclonal antibodies
MAVS	mitochondrial antiviral signaling
MMF	monomethyl fumarate
N	nucleocapsid protein
nMS	native mass spectrometry
NPC	nuclear pore complex
Nsps	nonstructural proteins
Orfs	open reading frames
PAM	protospacer adjacent motif

PDL-MS	proximity-dependent labeling mass spectrometry
PDL	proximity-dependent labeling
PPIs	protein–protein interactions
PRM	parallel reaction monitoring
PTMs	post-translational modifications
RBD	receptor-binding domain
RBR	RING-in-between-RING
RING	really interesting new gene
RTCs	replication and transcription complexes
S	spike protein
SARS-CoV-2	severe acute respiratory syndrome coronavirus 2
ssODN	single-stranded donor oligonucleotides
SUMO	small ubiquitin-like modifier
TAR	trans-activation response element
TDMS	top-down mass spectrometry
TIDE	tracking of indels by decomposition
TPP	thermal proteome profiling
TRESI-MS	time-resolved electrospray ionization mass spectrometry
UFM	ubiquitin fold modifier
UPS	ubiquitin-proteasome system
UVPD	ultraviolet photo-dissociation
VLPs	virus-like particles
XL-MS	cross-linking mass spectrometry

CHAPTER 1

Introduction

Human immunodeficiency virus (HIV) and severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) are two major viral pathogens of the past 50 years, causing a global HIV/acquired immunodeficiency syndrome (AIDS) epidemic in the 1980s and a global coronavirus disease 2019 (COVID-19) pandemic in 2020, respectively. Both the HIV epidemic and COVID-19 pandemic are currently ongoing, and both viruses have infected millions of people with devastating consequences of long-term health effects and death.

At the end of 2020, there were about 37.6 million people infected with HIV, of which 84% knew their HIV status, 73% were accessing anti-retroviral therapy (ART) with the goal of suppressing viral load, and 66% achieved viral load suppression¹. An unsuppressed viral load has devastating health consequences, including the progression to AIDS and death. It is estimated that 36.3 million people have died from HIV/AIDS since the beginning of the epidemic, with 680,000 people dying from AIDS-related illnesses in 2020². Individuals with an unsuppressed viral load not only risk the development of AIDS, but also risk transmission to others. 1.5 million people became newly infected with HIV during 2020, a number that is still too high but marks a 30% decrease in transmission from 2019¹. ART is effective at suppressing the viral load, preventing the progression to AIDS, and makes transmission to others less likely. However, ART is not curative, and requires constant lifelong treatment. ART drugs can cause side effects including nausea, diarrhea, dizziness, and fatigue, so while ART is life-saving it is not trivial³. There is currently no vaccine or cure for HIV. HIV disproportionately affects vulnerable and underrepresented populations, including gay and bisexual men, African Americans, and Latinos⁴. HIV infection was heavily stigmatized in the 1980s, and the media's use of the term "the gay plague" demonized the LGBT community and incited homophobia⁵. Stigmatization still endures today, as 80% of adults receiving HIV treatment report feeling internalized stigma⁶. HIV prevention and cure remain important research priorities.

As of August 2021, SARS-CoV-2 has infected over 200 million people worldwide and caused more than 4.3 million deaths. An estimated 10-30% of infected individuals have reported

lingering symptoms 2 months or longer after initial infection^{7,8}. These individuals, known as “COVID long-haulers”, experience diminished quality of life and debilitating health effects such as extreme fatigue, brain fog, chronic lung and cardiac issues, and negative impacts on mental health⁸. Remarkably, effective vaccines against COVID-19 became available as early as January 2021, representing the fastest vaccine development in human history, and to date nearly 4 million people have been vaccinated against COVID-19⁷. However, vaccine hesitancy has slowed progress fighting the pandemic, and infection continues to spread through communities via unvaccinated individuals and breakthrough infections in vaccinated individuals⁹. As with HIV, SARS-CoV-2 has exacerbated health inequities and disproportionately harmed vulnerable communities. One such example is those who do not have the ability to quarantine, including unhoused individuals and those who cannot work from home¹⁰. In parallel to the stigmatization of HIV, stigmatization of SARS-CoV-2 has incited a rise in anti-Asian racism. In the United States, President Donald Trump referred to the virus as “the Chinese virus”, and 2020 saw a drastic uptick in hate crimes against Asian Americans, rising nearly 150%^{11,12}. The many parallels between the HIV epidemic and SARS-CoV-2 pandemic emphasize the significant harm these viruses have caused to both society and human health, and demonstrate the urgent need for research.

An understanding of the virology of HIV and SARS-CoV-2 is an essential foundation for their study. HIV is a retrovirus with a genome composed of two strands of ~10kb positive-sense single-stranded RNA. Its genome encodes 3 structural proteins (1) the gag polyprotein that is processed into MA, CA, SP1, NC, SP2, and P6, (2) the pol polyprotein that is processed into RT, RNase H, IN, and PR, and (3) gp160 which is processed into gp120 and gp41), 2 essential regulatory elements (Tat and Rev), and 4 accessory regulatory proteins (Nef, Vif, Vpr, and Vpu). HIV gains entry into the cell by binding of the viral gp120 to the human CD4 receptor aided by CXCR4 or CCR5 co-receptors. Once inside the cell, the virus reverse transcribes its RNA genome into DNA, integrates into the host genome, transcribes viral mRNA, translates viral

proteins, and produces virions that are released via budding^{13,14}. If HIV infection is left untreated, it leads to the development of acquired immunodeficiency syndrome (AIDS). Currently, anti-retroviral therapy (ART) is a combination therapy of multiple drugs that inhibit key parts of the HIV life cycle. However, ART is not an HIV cure because it does not eliminate the reservoir of latently infected cells. The latent HIV reservoir is established early during infection. While the mechanisms governing the establishment of HIV latency are poorly understood, it occurs in cells in which the HIV proviral DNA is integrated into the host genome, but viral genes are not actively transcribed¹⁵⁻¹⁷. Because these cells are not producing virions, they are not killed by viral toxicity or immune clearance, and can persist for decades, with the ability to randomly reactivate at any time^{18,19}. For this reason, ART treatment must be lifelong.

SARS-CoV-2 is a single stranded, positive sense RNA virus. Its ~30 kb genome encodes 16 nonstructural proteins (Nsp 1–16), 4 structural proteins (Spike (S), Envelope (E), Membrane (M), and Nucleocapsid (N)), and 9 accessory proteins (Orf3a, Orf3b, Orf6, Orf7a, Orf7b, Orf8, Orf9b, Orf9c, and Orf10)^{20,21}. SARS-CoV-2 gains entry into the cell by binding of the viral S protein to the human angiotensin-converting enzyme 2 (ACE2) receptor and cleavage by the cellular serine protease TMPRSS2^{22,23}. After entry, the virus's positive sense RNA genome is immediately ready for polycistronic translation by the host ribosome. The viral genome is replicated by a viral RNA-dependent RNA polymerase holoenzyme. Full length genomic RNA is replicated via a negative-sense intermediate, and a nested set of subgenomic mRNAs encoding viral structural and accessory proteins are synthesized by continuous transcription and then translated either at the ER or in the cytoplasm. Virion assembly takes place in the ER-Golgi intermediate compartment (ERGIC), where N protein binds the RNA genome, virions bud from ER and Golgi membranes, and mature virions are released through a process similar to exocytosis^{24,25}.

Besides virology, there are many important areas of study for HIV and SARS-CoV-2. The development of accurate, widely distributable diagnostic tests is essential for reducing

transmission and treating infected individuals. Prevention of viral infection through vaccine development is also critical. For those who are infected, clinical research is important to study COVID-19 disease progression and testing the efficacy of drug treatments, and epidemiology studies how the virus spreads through a population. While this does not represent a comprehensive list of all areas of study, it provides a sense of the various size and scale of research questions pertinent to these viruses. This dissertation will focus on basic biology, meaning the molecular mechanisms by which SARS-CoV-2 and HIV infect their hosts.

Like all viruses, SARS-CoV-2 and HIV rely upon host cell machinery to replicate, and using basic biology to characterize these host-pathogen interactions informs our understanding of virology, pathology, and therapeutics. Host factors identified as critical for viral infection can serve as candidate drug targets, and novel drugs are desperately needed for both SARS-CoV-2 and HIV. Large-scale “omics” approaches, including genomics, transcriptomics, and proteomics are powerful technologies that can yield essential biological information about HIV and SARS-CoV-2 infection and identify druggable host factors. These systems biology approaches are particularly promising tools to study host-pathogen interactions because they capture a global and unbiased picture of infection.

Identification of host factors that HIV and SARS-CoV-2 commandeer and evade highlights vulnerabilities in their viral life-cycles that can be exploited for host-directed therapy (HDT). Since host mutation events are rare and not selected for viral survival, targeting factors encoded and expressed from the host genome limits the ability of the virus to develop escape mutations, an event that can occur when viral proteins targeted by a drug mutate under the drug’s selective pressure and over time develop drug resistance. While targeting multiple proteins and processes simultaneously can provide a high enough burden as to limit escape mutations, as is seen in HIV ART combination therapy, these drug regimens require consistent life-long adherence or HIV will rebound and re-establish active infection^{19,26}. Lapses in ART treatment provide opportunities for HIV to mutate and develop drug resistance²⁷⁻²⁹, a concern

avoided by HDT. HDT also provides advantages for the urgent treatment of COVID-19, as drugs targeting host proteins that are already FDA-approved for the treatment of other diseases can be repurposed for faster treatment of COVID-19.

Here, we use systems biology approaches to identify host factors important for SARS-CoV-2 and HIV infection that can serve as HDT candidate drug targets. In **Chapter 2**, we propose proteomic approaches to inform SARS-CoV-2 virology, COVID-19 pathology, and therapeutic strategies to combat the current pandemic. Specifically, we discuss how (1) mass spectrometry-based structural techniques can overcome limitations and complement traditional structural approaches to inform the dynamic structure of SARS-CoV-2 proteins, complexes, and virions; (2) virus–host protein–protein interaction mapping can identify the cellular machinery required for SARS-CoV-2 replication; (3) global protein abundance and post-translational modification profiling can characterize signaling pathways that are rewired during infection; and (4) proteomic technologies can aid in biomarker identification, diagnostics, and drug development in order to monitor COVID-19 pathology and investigate treatment strategies.

In **Chapter 3**, we combine proteomic and genetic systems biology approaches to identify E3 ubiquitin ligases critical for HIV infection and latency reversal. We use proteomics to identify 116 E3 ubiquitin ligases that are expressed in physiologically-relevant primary human CD4⁺ T cells, knock them out with CRISPR-Cas9, and identify 10 E3s that affect HIV infection: TRAF2, TRAF3, and PRPF19 have antiviral activity, and MARCH5, ZFP91, UHRF1, VPS18, NOSIP, PPIL2, RING1 have proviral activity. Importantly, we find that TRAF2 knockouts reverse latency in JLat and primary cell models, thus identifying TRAF2 as a candidate drug target for “shock and kill” therapy towards an HIV cure.

Chapter 4 discusses the significance of the work, including the power of global, unbiased systems biology approaches to elucidate host-pathogen interactions that are critical for combating two major pathogens of the past 50 years.

CHAPTER 2

Proteomic approaches to study SARS-CoV-2 biology and COVID-19 pathology

SUMMARY

The novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the causative agent of coronavirus disease 2019 (COVID-19), was declared a pandemic infection in March 2020. As of December 2020, two COVID-19 vaccines have been authorized for emergency use by the U.S. Food and Drug Administration, but there are no effective drugs to treat COVID-19, and pandemic mitigation efforts like physical distancing have had acute social and economic consequences. In this perspective, we discuss how the proteomic research community can leverage technologies and expertise to address the pandemic by investigating four key areas of study in SARS-CoV-2 biology. Specifically, we discuss how (1) mass spectrometry-based structural techniques can overcome limitations and complement traditional structural approaches to inform the dynamic structure of SARS-CoV-2 proteins, complexes, and virions; (2) virus–host protein–protein interaction mapping can identify the cellular machinery required for SARS-CoV-2 replication; (3) global protein abundance and post-translational modification profiling can characterize signaling pathways that are rewired during infection; and (4) proteomic technologies can aid in biomarker identification, diagnostics, and drug development in order to monitor COVID-19 pathology and investigate treatment strategies. Systems-level high-throughput capabilities of proteomic technologies can yield important insights into SARS-CoV-2 biology that are urgently needed during the pandemic, and more broadly, can inform coronavirus virology and host biology.

INTRODUCTION

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a novel coronavirus that causes coronavirus disease 2019 (COVID-19), and its outbreak in 2019 and the subsequent pandemic devastated global economies and human health. The *Coronaviridae* family of viruses, named for their crown-like appearance under an electron microscope, includes alpha, beta, gamma, and delta subgroups that infect a wide variety of mammals and birds, but

mutations facilitate cross-species infections and human-to-human transmission of the viruses³⁰. Seven coronaviruses are known to infect humans, two alpha coronaviruses (HCoV-229E and HCoV-NL63), and five beta coronaviruses (HCoV-OC43, HCoV-HKU1, MERS-CoV, SARS-CoV-1, and SARS-CoV-2)³¹. HCoV-229E, HCoV-NL63, HCoV-OC43, and HCoV-HKU1 regularly infect humans and cause common cold symptoms that are typically cleared but can progress to bronchiolitis or pneumonia^{32,33}. In contrast, MERS-CoV, SARS-CoV-1, and SARS-CoV-2 are more likely to cause severe respiratory disease^{20,34,35}. While the SARS-CoV-1 (2002/2003) and the MERS-CoV (2012) outbreaks had respective mortality rates of ~10% and 36%^{36,37}, the outbreaks were contained within specific geographic regions with only 8098 and 2494 verified infections^{38,39}. SARS-CoV-2 has a lower fatality rate but is more widespread than SARS-CoV-1 or MERS-CoV³⁷, with 75,704,857 confirmed cases worldwide as of December 21, 2020⁴⁰. There are currently over 50 COVID-19 vaccines in clinical trials⁴¹, and the US Food and Drug Administration (FDA) has authorized the Pfizer-BioNTech and Moderna COVID-19 mRNA vaccines for emergency use^{42,43}. This marks the fastest vaccine development in history, and widespread distribution of an effective vaccine would allow a safer end to stringent, long-term physical distancing that has had profound social and economic consequences⁴⁴. However, for people not protected by vaccination, there is an urgent need for effective COVID-19 treatment options. At this time, Remdesivir is the only drug approved by the FDA; however, the World Health Organization recommends against its use due to a lack of evidence for its efficacy^{45,46}.

During this unprecedented global crisis, the scientific community mobilized research efforts probing the mechanisms of SARS-CoV-2 infection and replication (Table 2.1)⁴⁷⁻⁴⁹. SARS-CoV-2 has a ~30 kb positive-sense RNA genome with as many as 14 open reading frames (Orfs) encoding 16 nonstructural proteins (Nsp 1–16), 4 structural proteins (Spike (S), Envelope (E), Membrane (M), and Nucleocapsid (N)), and 9 accessory proteins (Orf3a, Orf3b, Orf6, Orf7a, Orf7b, Orf8, Orf9b, Orf9c, and Orf10)^{20,21}. In the SARS-CoV-2 life cycle (Figure 2.1A),

viral entry is initiated by the binding of S protein to the human angiotensin-converting enzyme 2 (ACE2) receptor on the cell surface^{22,23}. This is followed by cleavage of S protein by the cellular serine protease TMPRSS2, which is required for fusion of viral and host cell membranes^{22,23}. After entry, the virus's positive sense RNA genome is immediately ready for polycistronic translation by the host ribosome, and ribosomal frameshifts allow for the expression of the Orf1a and Orf1b polyproteins²⁴. Autoproteolytic cleavage of Orf1a and Orf1b by viral proteases produces 16 Nsps²⁴. The viral genome is replicated by a viral RNA-dependent RNA polymerase holoenzyme, consisting of Nsp7, Nsp8, Nsp12, and Nsp14 for RNA transcription, capping, and proofreading^{24,25}. Replication takes place on double-membrane structures called replication and transcription complexes (RTCs) derived from and sometimes contiguous with the endoplasmic reticulum (ER)^{24,25}. RTCs are thought to both protect the virus from innate immune responses and concentrate the necessary viral components required for replication. Full length genomic RNA is replicated via a negative-sense intermediate, and a nested set of subgenomic mRNAs encoding viral structural and accessory proteins are synthesized by continuous transcription and then translated either at the ER or in the cytoplasm. Virion assembly takes place in the ER-Golgi intermediate compartment (ERGIC), where N protein binds the RNA genome, virions bud from ER and Golgi membranes, and mature virions are released through a process similar to exocytosis²². Understanding the underlying biology of SARS-CoV-2 infection, more specifically the host proteins and cellular processes that are essential for SARS-CoV-2 infection and replication, will identify targets for both drug repurposing and development of novel host-directed therapies.

Large-scale “omics” approaches, including genomics, transcriptomics, and proteomics are powerful technologies that could yield essential biological information about SARS-CoV-2 virology and COVID-19 pathology. Genomic approaches have been essential to investigate SARS-CoV-2 genome structure and similarities with related coronaviruses, among other foundational contributions to our understanding of this novel virus. With an understanding of the

SARS-CoV-2 genome, the field is now equipped to probe the actionable components of the virus: the proteins that the viral genome encodes. Proteomic approaches applied to SARS-CoV-2 allow the investigation of open questions of varying size and scale (Figure 2.1B). Proteomics can inform the structure of a single viral protein, the composition of a complete virion, and a global view of the host proteome during infection. Proteomic methods provide unique insight into the interaction between virus and host, including the host machinery co-opted for viral replication and signaling pathways that characterize the host response. Proteomic tools can also be used to probe interactions between compounds and proteins, and represent a powerful strategy for drug discovery. This perspective will discuss how proteomics can be leveraged to answer the following open and fundamental questions about SARS-CoV-2 biology. How do the dynamic structures of SARS-CoV-2 proteins and virions inform pathogenicity? What cellular machinery does SARS-CoV-2 utilize during infection? Which signaling pathways are rewired during SARS-CoV-2 infection? What are strategies to monitor COVID-19 pathology and investigate treatment strategies?

How do the dynamic structures of SARS-CoV-2 proteins and virions inform pathogenicity?

Viral proteins dictate the virion's structure and shape, and carry out activities essential for viral replication. Studying the dynamic structures of SARS-CoV-2 proteins and intact virions is essential not only to understand their molecular functions, but also to facilitate the design of effective small molecule therapies that can disrupt virions, viral entry, and virus replication and egress. In addition to atomic-resolution structural approaches that are effective for smaller proteins and protein complexes (e.g., X-ray crystallography and NMR), advances in approaches like cryo-electron microscopy (cryo-EM) have made the analysis of larger, flexible, multistate proteins and protein complexes more feasible. However, these approaches still require reconstituted complexes often expressed and purified from non-native bacterial or other

expression systems, and can require non-native buffers and conditions to achieve structure determination. In contrast, diverse and complementary proteomic approaches including cross-linking mass spectrometry (XL-MS), hydrogen/deuterium exchange mass spectrometry (H/DX-MS), and intact protein mass spectrometry can illuminate structural features of proteins and protein complexes under near physiological conditions or even inside biologically relevant living cells or intact virions (Figure 2.2)⁵⁰. Unlike NMR or X-ray crystallography, these mass spectrometry (MS)-based proteomic structure techniques require relatively lower amounts of protein samples. Despite their benefits, XL-MS, H/DX-MS, and native MS (nMS) have been widely considered to be very niche with their applications requiring a high degree of specialized expertise and equipment not generally applied in standard MS experiments. Though there have been great advances in each field opening the technology to nonspecialists, this has limited their widespread adoption, particularly in rapid-response research. Still, each methodology provides a unique perspective, and when integrated with other structural techniques can provide a more comprehensive understanding of the dynamic complex structures essential for SARS-CoV-2 replication.

Cross-linking mass spectrometry (XL-MS)

XL-MS is a powerful approach that can probe protein–protein interaction (PPI) networks and interfaces while overcoming a number of structural biology limitations^{51,52}. Primarily this includes the ability to (1) capture structural information from transient and dynamic PPIs and PPI binding interfaces; (2) accommodate not only sample impurities but also structural heterogeneity, flexibility, and subunit composition; and (3) be applied to live cell and in vivo applications⁵². When combined with integrative modeling approaches (see below), XL-MS can provide distance restraints for dynamic structure determination. In recent years, a number of diverse XL-MS strategies have emerged though they all generally rely on chemical cross-linkers to covalently bridge adjacent proteins via reactive amino acid residues, followed by MS-based identification of cross-linked peptides^{52–54}. Cross-linked peptide searching and identification is

complicated by two main challenges: (1) the lower frequency of cross-linked peptides in a complex mixture of mainly unmodified peptides; and (2) the complexity of cross-linked peptide spectra which have two covalently linked contributing peptide sequences⁵⁵⁻⁵⁷. The majority of XL-MS strategies include some form of enrichment step to prioritize identification of cross-linked peptides either by their size, hydrophobicity, ion mobility, or charge. Improved software and the development of functionalized chemical cross-linkers that are isotope-coded, MS-cleavable, or contain a reporter ion have made the identification of cross-linked peptides more accurate and straightforward. While the scope of this perspective is not a comprehensive review of XL-MS approaches, recent in depth reviews of these methods can be found in refs⁵²⁻⁵⁴.

The practical applications for XL-MS vary from informing large macromolecular complexes^{58,59}, to conformationally flexible complexes^{60,61}, to more recent applications probing dynamic interfaces of challenging host-pathogen complexes⁶² and even virus-like particles (VLPs)⁶³. While cryo-electron tomography (cryo-ET) and cryo-EM have become very powerful techniques for visualizing viral protein interactions and intact viral particle structures, atomic resolution of dynamic interactions is still challenging. Cryo-ET shows ordered binding of part of the flexible receptor to the viral surface, with distal domains in multiple conformations. Using cryo-ET, Meyer et al. demonstrated the binding of adeno-associated virus 2 (AAV2) VLPs to the cell surface receptor AAVR⁶³ and by complementing with XL-MS data they were able to identify regions of AAVR in close proximity to AAV2 proteins. XL-MS data validated the localization of the PKD2 domain of AAVR, improved the placement of the PKD1 domain of AAVR, and explained the disordered EM density in the structure. By combining cryo-ET, cryo-EM, and XL-MS, collective limitations in size, scale, and resolution could be overcome to determine the structure of AAV2 bound to AAVR soluble domains and thus make biological predictions for the effects of neutralizing antibodies. Prchal et al. combined affinity purification, XL-MS, and NMR data to determine the structure and map interaction interfaces between the cytoplasmic tail of Mason-Pfizer monkey virus (M-PMV) envelope and matrix proteins⁶⁴. Similarly, XL-MS was

recently leveraged to study how the HIV protein Nef targets surface receptor CD4 for endocytosis to promote HIV infection⁶². XL-MS data captured flexible and unresolved components of the Nef-CD4-AP2 crystal structure, and confirmed the observation that Nef serves as a flexible connector between CD4 and clathrin AP2 to promote endocytosis and downregulation of CD4⁶². Finally, XL-MS allows unbiased structure characterization and identification of unknown structural features including additional protein components or post-translational modifications (PTMs). In Yu et al. the thiol-cleavable cross-linker 3,3'-dithiobis(sulfo-succinimidylpropionate) (DTSSP) was used to identify vimentin as a transient interactor of the SARS-CoV-1 Spike (S)-ACE2 virus–host protein complex⁶⁵. Vimentin expressed on the cell surface was found to be important for SARS-CoV-1 virus entry and Vero E6 cells pretreated with anti-vimentin antibodies showed >40% reduction in SARS-CoV-1 VLP uptake⁶⁵.

Though to date there have been no published studies using XL-MS to characterize SARS-CoV-2 virions or viral proteins, the studies above demonstrate some of the potential applications for XL-MS. With the advent of newer and faster instruments, improved XL-MS identification software, and optimized design of new cross-linkers for intracellular applications, the reality of large-scale unbiased cross-linked peptide identification of whole cell networks is approaching. This type of network and structure data collection is feasible only with XL-MS experiments. One particularly appealing application would be the identification of cross-linked peptides from SARS-CoV-2 infected cells. Comparison of XL-MS data of mock versus SARS-CoV-2 infected cells could illuminate not only the virus–host protein interactions, but also the viral–viral protein complexes and the changes in host–host interactions that occur during infection. By unifying this data with the more standard affinity purification mass spectrometry (AP-MS) approach, scientists could build not only a network model of SARS-CoV-2 infected cells, but also provide PPI interface data to inform protein complex structures. In addition to exploring SARS-CoV-2 infected cells, researchers could look at the virus itself. The same global

XL-MS application to purified SARS-CoV-2 virions could, in one experiment, characterize the virion proteome, identifying both viral and host proteins that make up the virion, as well as provide PPI interface data on virus–virus and virus–host interactions of the virion structure. Combined with sophisticated cryo-EM and cryo-ET images, including recent discoveries made by Yao et al.⁶⁶ and Liu et al.⁶⁷, these studies could enhance our understanding of and development of chemotherapies against SARS-CoV-2.

Hydrogen/deuterium exchange mass spectrometry (H/DX-MS)

H/DX-MS measures changes in mass associated with the exchange between protein backbone amide hydrogen and deuterium from the surrounding D₂O. The rate of exchange is dependent on the conformation, surface accessibility, inductive effect of the neighboring groups, stability of hydrogen bonding networks, and the intrinsic chemical properties of the underlying amino acid sequence^{68–71}. It is used to examine conformations of individual proteins or large protein complexes⁷², locate protein sites directly or indirectly involved in binding⁷³, probe for allosteric effects⁷⁴, monitor the folding dynamics of protein domains⁷⁵, examine intrinsic disorder⁷⁶, and provide insights into protein–membrane interactions⁷⁷. H/DX-MS is unique in probing conformational states with single-residue resolution. To perform residue-level measurements by H/DX-MS, suitable gas-phase fragmentation of the deuterated peptides by electron transfer dissociation (ETD) or electron-capture dissociation (ECD) is generally applied^{78,79}. Both ETD and ECD generate fragment ions with vibrationally cold energy, minimizing hydrogen scrambling, a phenomenon seen using other fragmentation strategies like collision-induced dissociation (CID)^{80–82}. In 2013, Resetca et al. developed a method called time-resolved electrospray ionization mass spectrometry (TRESI-MS), which uses a microfluidic chip in-line with all the steps involved in a “bottom-up” HDX workflow⁸³. This development provided faster sample preparation times and improved reproducibility to the point that it is now feasible to characterize rapid structural transitions that occur during protein folding⁸⁴, ligand

binding⁸⁵, or post-translational modification⁸⁶, applications mostly inaccessible to conventional techniques.

More recently, H/DX-MS has emerged as a potential tool in structural virology exploring the Hepatitis C E2 glycoprotein⁸⁷, influenza hemagglutinin⁸⁸, HIV envelope glycoprotein^{89–92}, and Ebola GP1/GP2⁹³. Using H/DX-MS, Ye et al. were able to explore the architecture and self-assembly of SARS CoV-2 N protein and showed that the addition of the C-terminal spacer B/N3 domain to the N2b domain mediates the formation of a homotetramer⁹⁴. These types of studies provide insights into the conformational dynamics of proteins in-solution that directly reflect the structural changes, mechanism of viral glycoprotein recognition, and virus neutralization caused by the binding of antibodies and small molecules. Further applications of H/DX-MS to SARS-CoV-2 could help to characterize (1) the effect of PTMs on SARS-CoV-2 protein structures such as glycosylation of S protein or phosphorylation of Orf9b and (2) the binding of small molecules or monoclonal antibodies (mAbs) targeting SARS-CoV-2 proteins, such as S protein. This could not only inform therapeutics, but also provide essential information on host immunity, and potentially the development of different types of neutralizing antibodies.

Intact protein mass spectrometry

While XL-MS and H/DX-MS rely on digesting protein complexes into peptides, intact protein MS enables the analysis of intact proteins and protein complexes. nMS is a unique intact protein MS methodology that maintains noncovalent interactions and stoichiometry of protein complexes in gas phase^{95,96} while denaturing top-down MS (TDMS) approaches identify proteoforms and PTMs. nMS combined with CID disrupts noncovalent interactions between protein subunits based on their strength, and can help decipher stoichiometry and topology of protein complexes. Furthermore, applications with ion mobility-MS (IMMS) have allowed the field to explore greater structural details of large macromolecular assemblies (for detailed review, see⁹⁷). Importantly, IMMS allows researchers to separate and characterize protein complexes and protein subunits in gas-phase based on their size and shape. It opens up

avenues for structural analysis of heterogeneous protein complexes allowing assessment of stoichiometry, topology, and cross-section of the assemblies and their subunits. IMMS can aid in generating hypotheses about complex structures, conformations, assembly, and disassembly and offers complementarity in structural biology⁹⁷.

nMS can capture transition events in protein complex assembly and disassembly, and more recently in proteolytic cleavage events like those that are essential to coronavirus replication and infection⁹⁷. As described above and shown in Figure 2.1, the translated coronavirus polyproteins are processed to produce Nsps. In order to investigate the processing of the SARS-CoV-1 polyprotein Nsp7–10 region by Mpro (the main protease), as well as the formation of intermediate products and complexes, Krichel et al. used nMS to demonstrate how the in vitro cleavage efficiencies resemble limited proteolysis of a folded protein and are influenced by tertiary polyprotein structure⁹⁸. In addition, they identified the heterotetrameric Nsp7 and Nsp8 complex as the predominant product, thus providing not only the order of proteolytic cleavage, but also the formation of postcleavage structures⁹⁸. While SARS-CoV-2 S protein and ACE2 are challenging to study by nMS given their extensive glycosylation, nMS with limited charge reduction⁹⁹ provided meaningful information about the complex between ACE2 and the receptor-binding domain (RBD) of the S protein as well as the role of heparin in destabilizing the ACE2/RBD association¹⁰⁰. Yang et al. showed that both short (pentasaccharide) and long (eicosasaccharide) heparin oligomers form a stoichiometric complex with the RBD, indicating a single binding site which alters the protein conformation and subsequently results in a decrease in its ability to associate with ACE2¹⁰⁰. This study suggests that nMS might be a powerful method for studying the interaction between drugs and their therapeutic targets. In addition to these studies, nMS has been applied to investigate viral particle assembly. While the multistep assembly of multiple identical proteins into an icosahedral virus capsid is poorly understood, using charge detection MS, Pierson et al. were able to detect trapped intermediates during the assembly of the hepatitis B virus capsid¹⁰¹. Subsequent cryo-

EM analysis indicated incomplete capsids rather than aberrant structures suggesting that the observed structures are on-path intermediates¹⁰¹. Application of these types of experiments to SARS-CoV-2, particularly if complemented with additional structural MS techniques like in vivo XL-MS, could characterize the assembly and structure of SARS-CoV-2 virions.

mAbs that target viral proteins are a promising class of therapeutics against infectious diseases like COVID-19¹⁰², Ebola¹⁰³, Hendra and Nipah virus infection¹⁰⁴. High-quality characterization of intact mAbs is essential and is commonly performed by TDMS. Similar to nMS, TDMS strategies avoid the use of proteolytic enzymes, providing intact masses of the molecules, which can determine the presence of multiple proteoforms^{105,106}. MS/MS fragmentation of the intact proteins and large protein subunits can provide amino acid resolution for interpretation of sequence heterogeneity and presence of PTMs^{106–108}. TDMS-associated technologies have significantly advanced over the past decades^{106,109}, mainly focusing on two areas of development: instrumentation and fragmentation approaches^{107,110}. Newer methods such as ultraviolet photodissociation (UVPD)¹¹¹, ECD¹¹², and ETD¹¹³ have radically increased the sequence coverage and PTM information for TDMS experiments. As a result, TDMS has emerged as a powerful means in basic, translational, and clinical research for protein identification and proteoform elucidation. Though no studies thus far have capitalized on TDMS for studying SARS-CoV-2 antibodies purified from convalescent sera, this application has the potential to aid in the identification of proteins and proteoforms that have higher specificity or neutralization potential, which could help in the design of mAb therapy, or in predicting reactions to new mutations.

Integrating complementary structural data

Complementary data collected from these diverse structural proteomic methodologies can be combined to maximize structural insight through integrative modeling. Integrative modeling synthesizes experimental data from structural, biochemical, proteomic and genetic studies to optimize a comprehensive and accurate model of protein complexes^{114–116}. In this

way, integrative modeling represents the most complete understanding of a structure that accounts for all data types. Integrative modeling is particularly powerful when defining macromolecular complexes for which atomic resolution might be feasible for individual or small, inflexible, subunits and subcomplexes, but not for the intact holocomplex. For example, integrative modeling was essential to synthesize cryo-EM, XL-MS, and crystal structure data to determine the dynamic structure of the yeast nuclear pore complex (NPC), a challenging macromolecular membrane bound structure consisting of 500+ proteins¹¹⁴. While cryo-EM and XL-MS heavily feature in integrative modeling, few studies have capitalized on the combined strengths of XL-MS, H/DX, and nMS. Particularly useful is the ability to define in-solution structures that are more representative of their native state, and in the case of XL-MS, can be performed in living cells.

To date, there are no integrative structures provided for SARS-CoV-2 viral proteins or complexes. Given the recent emergence of the virus, this can likely be attributed to the specialization and time required by each structure technology as well as for integrative modeling. However, applied structural proteomic technologies could be very powerful in studying SARS-CoV-2, as a number of SARS-CoV-2 proteins and virus–host complexes present structural challenges including size, cellular localization, and flexibility. Full-length SARS-CoV-2 Nsp3 is a large, multidomain, multifunctional, enzymatic, essential protein that has yet to be fully structurally characterized. Expression, in cell cross-linking, and purification of Nsp3 from human cells could provide structural information pertaining to adjacent Nsp3 domain residue contacts and could capture Nsp3-host interactors that may be transient in nature. Additional TDMS experiments of Nsp3 purified from human cells could identify essential proteoforms that exist in cells. By combining these data with high resolution X-ray diffraction data of individual domains and H/DX-MS studies probing individual domain interactions, integrative modeling could determine a comprehensive and dynamic structure for Nsp3 as it looks in human cells. Investigating SARS-CoV-2 viral entry with integrative modeling informed

by structural proteomics like XL-MS could also be particularly powerful to study S protein interactions and SARS-CoV-2 VLP interactions with the ACE2 receptor and the TMPRSS2 protease, as was demonstrated for AAV2 and AAVR in the example above⁶³. This type of XL-MS data would pair well with the nMS study by Yang et al. described above that focused on defining the ACE2-RBD S protein interaction. Combined with H/DX-MS studies that help identify solute exposed surfaces of free S protein, and protected interfaces of S bound to ACE2 or TMPRSS2, an integrative model could provide useful information about the structure in solution. Understanding the structure of these virus–host interaction interfaces will be critical for developing drugs that disrupt and prevent SARS-CoV-2 entry into host cells.

What cellular machinery does SARS-CoV-2 utilize during infection?

Like all viruses, SARS-CoV-2 does not encode all of the machinery required for its replication, and must co-opt host machinery for the production of progeny virions. Studying virus–host PPIs can identify essential host factors and provide mechanistic details into the viral life cycle (Figure 2.2). AP-MS is the most widely applied proteomics method to systematically characterize virus–host PPIs by expressing and purifying individual affinity tagged viral proteins in host cells. As a complementary method to classical AP-MS, proximity-dependent labeling (PDL) coupled to quantitative MS (PDL-MS) has emerged to study proximal PPIs, particularly those that are more transient or weaker interactions¹¹⁷. While AP-MS and PDL are powerful approaches to define host interactions of single viral proteins, recent developments in complex centric proteome profiling (CCPP) allow global mapping of cellular macromolecular complexes^{118,119}.

Affinity purification–mass spectrometry (AP-MS)

AP-MS is an established method that has been widely applied to systematically characterize virus–host PPIs by expressing individual affinity tagged viral proteins in host cells, purifying stably bound host protein interactors, and identifying those interactors by MS^{120–126}. In

a recent AP-MS study, Gordon et al. expressed and purified 26 of the 29 SARS-CoV-2 proteins from HEK293T cells, which led to the discovery of 332 virus–host PPIs¹²⁷. Subsequent chemoinformatic analysis identified 69 drugs and small molecules that target the SARS-CoV-2-human PPI network, which thus have the potential to disrupt host factor-dependent viral processes. Of 47 compounds tested, two classes of small molecules emerged that exhibited strong antiviral effects: those that modulate the activity of Sigma receptors, and those that inhibit mRNA translation. In a subsequent study, AP-MS efforts in HEK293T cells were expanded to map the full interactomes of SARS-CoV-1 and MERS-CoV¹²⁸. Virus–host interactions for each of the three highly pathogenic human coronavirus strains were compared to identify and understand pan-coronavirus molecular mechanisms, which revealed a high number of shared interactions between SARS-CoV-1 and SARS-CoV-2 and higher divergence comparing SARS-CoV-1 and 2 to MERS¹²⁸. One notable example is the mitochondrial outer membrane protein Tom70, which interacts with the mitochondria localized Orf9b for both SARS-CoV-1 and SARS-CoV-2¹²⁸. MERS does not have an Orf9b protein and thus did not show an interaction with Tom70. Interestingly, in the same study Tom70 was validated to be a host-dependency factor for SARS-CoV-2¹²⁸.

In another impressive large-scale study that integrated AP-MS and global proteomic data, Stukalov et al. systematically mapped the virus–host interactions for both SARS-CoV-2 and SARS-CoV-1 in A549 lung carcinoma cells¹²⁹. Performing global protein abundance, phosphorylation, and ubiquitylation profiling of cells overexpressing individual viral proteins, and integrating this data with SARS-CoV-2 PPI networks identified cellular pathways regulated by viral proteins¹²⁹. This integrated PPI network was then used to predict well-characterized selective drugs that could be targeted for host-directed therapies and identified drugs targeting AKT and matrix metalloproteases as having anti-SARS-CoV-2 activity¹²⁹. These studies demonstrate the potential of rapidly translating AP-MS data into druggable host factors with subsequent identification of repurposable drugs that have potent antiviral activity. Beyond the

translational potential of global host–pathogen interaction maps, these studies also show the impact and breadth of work that can be carried out through rapid response of international collaborations between interdisciplinary scientists.

Proximity-dependent labeling (PDL)

PDL relies on enzymes which catalyze covalent transfer of biotin or biotin derivatives to proteins in proximity of the enzyme, including promiscuous biotin protein ligases (BirA/BioID/TurboID)^{130,131} and engineered ascorbic acid peroxidase (APEX)¹³². To map PPIs using PDL, the enzyme is fused to a protein of interest, such as a viral protein, and upon addition of a substrate the proximal proteins will become biotinylated, followed by their enrichment using Streptavidin resin, and analysis using quantitative MS. The covalent labeling of proximal proteins in cells allows for the capture and identification of not only stable, but also transient and weaker interactions, allows purification under harsh lysis conditions, and provides additional spatial information about the subcellular location of the PPIs through proximal labeling. This was demonstrated for example by applying the APEX-based PDL approach to study agonist-induced changes of G-protein coupled receptors (GPCRs) interaction networks, which successfully characterized functionally relevant GPCR interactors with temporal and spatial resolution¹³³.

PDL has also been extended to study host–pathogen interactions. Coyaud et al. applied a combined AP-MS and BioID approach to systematically map the Zika–host interactome composed of over 1200 Zika–host interactions, which revealed extensive organellar targeting by Zika virus and a role of peroxisomes for Zika virus infection¹³⁴. A recent study introduced BirA into the viral genome of MHV (murine coronavirus) to define the microenvironment of the RTC¹³⁵. The study identified more than 500 proteins in proximity to the RTC and discovered a close association of viral RNA synthesis sites with the host translation machinery¹³⁵. Another PDL-MS study performed BioID for 27 SARS-CoV-2 proteins fused to the miniTurbo enzyme in a lung adenocarcinoma cell line¹³⁶, which revealed proximal interactions with 2242 host

proteins. The inclusion of several host subcellular markers as baits provided testable hypotheses regarding the functional consequences of virus–host interactions. For example the proximal interactome of the N protein revealed the depletion of proteins critical to the formation of stress granules, suggesting that the interaction between N protein and G3BP1 might prevent stress granule formation¹³⁶.

These studies demonstrate the power of PDL-MS as a complementary approach to AP-MS to map SARS-CoV-2-host interactions, particularly to characterize transient interactions and inform cellular location. SARS-CoV-2 relies on organellar targeting with replication taking place in membrane bound vesicles derived from the ER and Golgi. Application of PDL to key viral proteins such as members of the RTC (Nsp7, Nsp8, Nsp12, and Nsp14), or key host factors such as the Nsp6 interactor Sigma-1 receptor in the context of SARS-CoV-2 infection could track PPIs and their cellular location, essentially providing an intracellular GPS for virus–host PPIs throughout the viral replication cycle. As a recent extension of proximity labeling approaches, split enzymes have been engineered which enable contact-specific PDL. The two fragments of the split enzyme remain inactive apart, and become active upon reconstitution when they are driven together by PPIs^{137,138}. These sophisticated approaches could be particularly powerful in validating and functionally characterizing high-confidence virus–host PPIs identified by AP-MS and/or PDL-MS. A split enzyme PDL experiment could simultaneously validate a putative interaction, determine additional protein complex-specific subunits, and derive intracellular spatial information for the specific interaction. For example, splitting the labeling enzyme between a viral protein and an ER protein can inform protein interactions that take place on the surface of double-membrane vesicles of SARS-CoV-2 RTCs¹³⁹.

Complex centric proteome profiling (CCPP)

CCPP, samples are gently lysed to preserve native protein complexes which are then (1) separated into fractions using size exclusion chromatography; (2) digested into peptides and analyzed by highly sensitive data-independent acquisition (DIA) quantitative MS; and (3) protein

patterns across the fractions correlated to determine the composition of protein complexes. This approach has recently been utilized to quantitatively compare protein complexes in two distinct cell-cycle states, suggesting a model for disassembly of the nuclear pore complex during the transition from interphase to mitosis¹¹⁹. Host complexes are similarly disassembled and rearranged during viral infection, and CCPP would allow the capture of those changes during SARS-CoV-2 infection in a systematic fashion. Indeed, one of the limitations with AP-MS and PDL-MS strategies, which rely on individually tagged bait proteins, is that the captured PPIs are missing the context of the full suite of viral proteins in infected cells. Therefore, virus–host interactions dependent on multiple viral proteins, nucleic acids, or signaling will be absent in the final analysis. Characterizing SARS-CoV-2 infected cells using CCPP would avoid the necessity for tagged components, and can provide a global picture of the complexes viral proteins form and interact with, as well as how they manipulate cellular machinery through recruitment or dissociation of specific components.

The integration of CCPP with organelle fractionation could illuminate additional spatial information for host–pathogen interactions^{140–142}. Organelle fractionation alone was recently applied to determine targeting of viral proteins to distinct organelles and to define alterations in organellar proteome composition during Cytomegalovirus infection¹⁴⁰. Proteins localized to one compartment in uninfected cells and a different compartment during infection were identified as translocated proteins, with most translocations occurring between the plasma membrane, ER, Golgi, and lysosome. Notably, this approach identified MYO18A as a protein that translocates from the plasma membrane to the lysosome during Cytomegalovirus infection, and siRNA knockdown of MYO18A significantly decreased virion production. Because SARS-CoV-2 replication relies on trafficking through cellular compartments, especially the ER and Golgi, CCPP combined with organelle fractionation could represent a particularly powerful approach to identify how host proteins are globally hijacked and translocated throughout the SARS-CoV-2 replication cycle.

Toward an integrated virus–host PPI network

Complementary data from various virus–host PPI mapping approaches can be integrated to provide a model for how SARS-CoV-2 hijacks host machinery to replicate inside of cells. A model that accounts for the molecular players involved as well as the timing and localization of events throughout the viral replication cycle will aid in the rational design of novel drugs, the repurposing of existing drugs, and the development of combinatorial therapies (Figure 2.2). Given the initial successes of host–pathogen interaction mapping in standard, though less physiologically relevant cell lines like HEK293T cells, it is imperative to expand these efforts to clinically relevant cell models that mimic the infection in human lung epithelia. While expression of individual affinity-tagged viral proteins allows high-throughput discovery of stable virus–host interactions, this approach will miss highly transient or less stable interactions, as well as those interactions that rely on simultaneous expression of multiple viral proteins, such as the viral RNA-dependent RNA polymerase holoenzyme responsible for viral genome replication. Therefore, host–pathogen interaction studies should be performed in the context of viral infection. In addition, cross-linking coupled with AP-MS, or the application of global XL-MS experiments, can capture transient and weaker interactions from SARS-CoV-2 infected cells (see XL-MS section above). These experiments can provide PPI network data and inform virus–host protein complex structures. Furthermore, exciting work could be done to investigate not only SARS-CoV-2, but other human coronaviruses (i.e., HCoV-OC43, SARS-CoV-1, and MERS-CoV) in reservoir and cell lines from cross-species, such as bats. Through identification of conserved virus–host interactions across different coronavirus species, host factors that increase or mitigate pathogenicity, severity, infectivity, and cross-species barriers could be identified. These data would be a starting point for designing pan-coronavirus treatment strategies.

Which signaling pathways are rewired during SARS-CoV-2 infection?

During infection a complicated battle takes place wherein viruses must simultaneously rewire cellular pathways they need for replication while evading the host cell's innate immune defenses. Perturbations to PTMs allow viruses to quickly manipulate the host environment to control cell cycle, prioritize transcription and translation of viral products, and evade the immune response. For example, HIV-1 Vif degrades regulatory subunits of the key cellular phosphatase PP2A, which results in hyperphosphorylation of many cellular proteins, including substrates of the aurora kinases, and causes cell cycle arrest in G2¹⁴³⁻¹⁴⁵. Nuclear factor kB (NF-kB) is an innate immune pathway regulated by phosphorylation and ubiquitylation that has been co-opted by viruses such as HIV-1, which use NF-kB as a transcription factor to express viral mRNA¹⁴⁶. Phosphorylation is also key for viral trafficking throughout the viral replication cycle¹⁴⁷. For example, Ebola entry requires phosphoinositide-3 kinase signaling¹⁴⁸, Influenza replication relies on nuclear-cytoplasmic shuttling regulated by multiple phosphorylation sites on its nucleoprotein¹⁴⁹, and the phosphorylation of HIV protein p6 is required for virion budding and release from the cell¹⁵⁰. Given the important role of perturbing phosphorylation for viral infection, mapping phosphorylation events to kinases and phosphatases could allow for the identification of druggable targets and repurposing of FDA approved drugs¹⁵¹.

Ubiquitin signaling is important for viral replication, and inhibition of the host ubiquitin-proteasome system (UPS) can lead to restriction of viral entry and expression of viral proteins, as it has been shown for coronaviruses including mouse hepatitis virus and SARS-CoV-1¹⁵². Viruses often exploit the host UPS to target restriction factors (host factors that inhibit viral infection) for degradative ubiquitylation. One such example is HIV-1 Vif, which together with the transcriptional cofactor CBF β hijacks the Cul5-RING E3 ubiquitin ligase and acts as a non-native substrate adaptor to target cellular APOBEC3 restriction factors for polyubiquitylation and proteasomal degradation¹⁵³⁻¹⁶⁰. Unbiased proteomic approaches that selectively enrich and identify specific PTMs like phosphorylation and ubiquitylation can study global perturbations

during infection and highlight cellular signaling pathways that are critical for SARS-CoV-2 replication (Figure 2.2).

Global phosphoproteomics

Phosphorylation is a quick, reversible PTM that viruses use to (1) regulate viral proteins and (2) alter the stability, subcellular location, and enzymatic activity of host proteins to aid viral replication. Protein phosphorylation is regulated by kinases and phosphatases, which phosphorylate and dephosphorylate substrates, respectively. While serine, threonine and tyrosine are the most commonly phosphorylated and studied amino acids, other amino acids such as arginine, lysine, aspartic acid, glutamic acid, histidine and cysteine are also phosphorylated and though harder to study, have been implicated in various host biology^{161,162}.

MS-based proteomics has emerged as the method of choice to systematically study protein phosphorylation and its dynamics. To obtain higher sensitivity in identifying phosphorylation sites in complex protein samples such as cell lysates, several approaches have been developed to enrich phosphorylated peptides based on ion metal affinity chromatography, ion exchange chromatography, and antibody-based immunoprecipitation¹⁶³. Combined with developments in MS and computational approaches to identify, quantify and localize phosphorylation sites, tens of thousands of phosphosites can routinely be accurately quantified in a single experiment^{164–166}. The ability to subsequently map the quantified phosphorylation events to kinases might reveal druggable targets and repurposable kinase inhibitors that have been approved for the treatment of other diseases such as cancer^{167–169}. These inhibitors can then be readily tested for efficacy against viral infection and contribute to rapid development of effective treatment strategies for the SARS-CoV-2 pandemic.

The feasibility of this concept was recently demonstrated by applying global phosphorylation analysis to SARS-CoV-2 infected ACE2-expressing A549, Vero E6, and Caco-2 cells^{129,151,170,171}. The study in Vero E6 cells identified among others p38, CK2, CDKs, AXL, and PIKFYVE as dysregulated kinases and demonstrated that their pharmacological inhibition

restricted SARS-CoV-2 replication¹⁵¹. Bouhaddou et al. discovered that CK2 is activated by the N protein resulting in upregulation of CK2 cytoskeleton-related targets, which contributed to the formation of filopodial protrusions where virus particles seem to be budding from¹⁵¹. Klann et al. discovered not only extensive rearrangements of growth factor receptor (GFR) signaling, but also validated antiviral efficacy for multiple GFR inhibitors, thus demonstrating the central function of GFR signaling in SARS-CoV-2 infection¹⁷¹. Stukalov et al. followed a conceptually similar approach integrating global protein abundance, phosphorylation and ubiquitylation data measured during SARS-CoV-2 infection in ACE2-expressing A549 lung carcinoma cells to identify perturbation of cellular pathways by the virus¹²⁹. Targeting selected pathways with drugs, they identified several kinase and Matrix metalloproteinase inhibitors with significant antiviral effects against SARS-CoV-2¹²⁹.

Global ubiquitylation profiling

Ubiquitylation is another reversible PTM which is essential for the replication of many different virus families. As viruses do not typically encode their own ubiquitin machinery, members of most virus families exploit their host's UPS to ubiquitylate or deubiquitylate proteins to target them for proteasomal degradation, alter their trafficking, or to change their activity^{172,173}. Ubiquitylation requires E1 enzymes for activation, E2 enzymes for conjugation, and E3 ligases that recruit the protein substrate. Substrates can be mono- or polyubiquitylated, and polyubiquitin chains can be connected by any of ubiquitin's seven lysines (K6, K11, K27, K29, K33, K48, and K63) or N terminus. Different linkages allow for different signaling outcomes, for example, K48 linkage is associated with degradation by the proteasome, while K63 linkage is associated with nondegradative signaling including DNA damage repair, innate immunity, and intracellular trafficking. Cells encode a large number of deubiquitylating enzymes (DUBs) that catalyze the removal of ubiquitin from target substrates, are important for regulating a number of cellular processes, and have been implicated in a number of human diseases^{174–177}. Viruses

may hijack E3 ligases to establish novel ubiquitylation events or hijack DUBs to remove established ubiquitylation events in order to aid their replication.

Since enrichment increases the identification of ubiquitination sites, the most commonly used method for assessing global ubiquitylation relies on immunoprecipitation of peptides containing a lysine residue modified by diglycine, an adduct left at sites of ubiquitylation after trypsin digestion^{178,179}. To distinguish between ubiquitylation events that target proteins for degradation or change protein activity, global ubiquitylation enrichment can be combined with (1) proteasome inhibition which should stabilize proteins targeted for degradation and (2) global abundance proteomics which should identify proteins that are targeted for degradation as downregulated. In contrast to phosphorylation where many enzyme–substrate relationships are well characterized, ubiquitin ligase-substrate relationships are less defined. This complicates the mapping of ubiquitin ligases to ubiquitylation events, and hampers the identification of ubiquitin-related druggable targets. However, ubiquitylation involves a physical interaction between a virus and one or more host proteins, therefore global ubiquitin profiling can be combined with AP-MS and/or proximity labeling to identify relevant interactions with ligases, representing druggable target candidates.

Global ubiquitylation combined with global protein abundance analyses of ACE2-expressing A549 cells infected with SARS-CoV-2 identified virus-mediated ubiquitylation events and host proteins targeted for degradation¹²⁹. In this study, 1053 host and viral proteins, including the ACE2 receptor protein, were identified as having SARS-CoV-2-regulated ubiquitylation site. In addition to global ubiquitylation analysis, complementary approaches have suggested ubiquitin signaling as particularly important for SARS-CoV-2 infection. It has been shown by AP-MS that multiple E3 ubiquitin ligases physically interact with viral proteins, which have been subsequently validated to be functionally relevant for SARS-CoV-2 infection. ORF10 was found to interact with the CUL2ZYG11B E3 ligase complex¹²⁷ and knockdown of ZYG11B in Huh-7.5 hepatoma cells resulted in decreased SARS-CoV-2-induced cytopathic effect¹⁸⁰,

suggesting that ZYG11B is a dependency factor for SARS-CoV-2 infection. Another example is the E3 ligase MIB1, which was discovered as an interactor of Nsp9¹²⁷, and interestingly knockout of MIB1 increased the virus-induced cytopathic effect not only for SARS-CoV-2, but also for HCoV-229E, HCoV-NL63, HCoV-OC43¹⁸⁰. These results suggest that MIB1, which is known to function in TBK1 polyubiquitylation, which in turn is a signal integrator of multiple RIG-like receptors and positive regulator of IRF3, might establish an antiviral state that broadly controls coronavirus infection¹⁸⁰. Finally, the viral protein Nsp3 has DUB activity, and its inhibition by small molecules decreases SARS-CoV-2 replication¹⁸¹.

Global protein abundance profiling and integrating signaling data

Viral perturbations in ubiquitin and phosphorylation signaling have downstream consequences at the protein level, and abundance proteomics is the method of choice to investigate how viral infection globally rewires the host proteome. Host proteins that are upregulated and downregulated during SARS-CoV-2 infection can be key host dependency and restriction factors that the virus utilizes or evades respectively to promote viral replication. Integrating protein abundance data with changes to protein ubiquitylation and phosphorylation can be an incredibly powerful way to identify not only individual proteins, but cellular pathways targeted by the virus, allowing scientists to build testable hypotheses of the molecular mechanisms of SARS-CoV-2 infection. Bojkova et al. analyzed the proteome of SARS-CoV-2 infected Caco-2 cells compared to mock infection, identified significantly perturbed pathways, and tested the antiviral effect of drug inhibition on those pathways¹⁸². In one notable example, 25 spliceosome proteins were found to be upregulated during infection, and the spliceosome inhibitor pladienolide B showed antiviral activity¹⁸². To better understand the mechanism of how SARS-CoV-2 rewires signaling pathways, PTM and abundance analysis can be performed after infection with gene deletion mutant viruses to identify which viral proteins are responsible for rewiring a given pathway. This approach has been applied to HIV to study phosphorylation and ubiquitin signaling associated with accessory proteins Vif, Vpr, and Vpu^{143,145,183}. Given its DUB

activity, Nsp3 represents a promising target to apply this type of study to SARS-CoV-2. PTM and abundance changes identified during infection with wild type SARS-CoV-2 but not with a Δ Nsp3 virus may represent proteins and PTMs regulated by Nsp3. Taken together, phosphorylation and ubiquitylation signaling integrated with protein abundance in SARS-CoV-2 infection is an important area of study to understand how the virus rewires host cells and identify druggable targets.

Despite the promise of data integration, its application for host–pathogen signaling has been difficult in part due to variability in and complexity of the individual data sets and in part due to the challenge of measuring both host and pathogen components (reviewed in ref¹⁸⁴). While identical time points or even split samples might be used to reduce reproducibility issues across data types, there will still be missing values that contribute to noise and inconsistency. In addition, a single protein might have multiple isoforms and multiple phosphorylation or ubiquitylation sites, with each potentially being unchanged, up- or down-regulated, in opposing ways. How to address this nonlinear integration is just the first step. Interpreting integrated data can also be challenging, as tracking meaningful changes across thousands of proteins, and forming testable hypotheses can lead to over- or under-interpretation. Finally, testing the importance of identified proteins or PTMs can be difficult. Genetic manipulation through knockout, knock-down, or mutation can be complicated by cellular and molecular redundancies. Additionally, to validate the specific modified site, site directed mutagenesis of endogenous genes needs to be performed, which is not established widely or in high-throughput fashion. Drug treatment can also complicate interpretation since there can be nonspecific interactions or cell-type related biology involved. Thus, a multipronged approach combining multiple functional validations can be key, though are often much lower in throughput. For instance, in Bouhaddou et al. parallel siRNA knockdowns and inhibitor treatments targeting the same p38 pathway were performed to demonstrate the p38 pathway as an essential signaling pathway for SARS-CoV-2¹⁵¹.

What are strategies to monitor COVID-19 pathology and investigate treatment strategies?

Managing the SARS-CoV-2 pandemic requires rapid and reliable diagnostics to trace disease spread, biomarkers to monitor disease severity, efficacious drug treatment to improve patient outcomes, and the development of vaccines to prevent the spread of disease. In addition to genomic-based approaches, proteomics might aid the high volume of diagnostic testing required for contact-tracing or be utilized to identify and monitor biomarkers of disease severity or treatment response (Figure 2.2). Improving patient outcomes also requires the identification of safe and effective drug treatments. As mentioned above, studying SARS-CoV-2 replication inside host cells can identify druggable targets. Proteomic technologies can also be employed subsequently to investigate binding of drugs to those targets and reveal potential off-target binding. These technologies include limited proteolysis-coupled mass spectrometry (LiP-MS), thermal proteome profiling (TPP), and activity-based protein profiling (ABPP) (Figure 2.2).

Targeted Proteomics for Monitoring Disease Progression and Treatment Response

Most biological samples taken from individuals suspected with COVID-19 are nasopharyngeal swabs, bronchoalveolar lavages, gargle samples, and blood samples, all of which represent complex proteomes with large dynamic range of protein concentrations. Thus, a highly sensitive method is required to allow for detection of SARS-CoV-2 proteins in complex diagnostic testing samples. Targeted proteomics approaches, in which a predefined set of proteins and their corresponding peptides are selectively and recursively isolated and then fragmented over their chromatographic elution time, offer highest sensitivity and quantitative accuracy¹⁸⁵. Several studies developed parallel reaction monitoring (PRM) assays for the detection of SARS-CoV-2 proteins^{186–189}, which were subsequently tested in dilute patient gargle samples¹⁸⁷, nasopharyngeal swab¹⁸⁹, and bronchoalveolar lavage samples¹⁸⁸. While these studies demonstrated that it is theoretically possible to detect viral proteins in patient samples using targeted proteomics, compared to PCR-based assays that can experimentally amplify signal and have high-throughput efficiency, the proteomic approach is limited by low sample

throughput and sensitivity. Given these limitations, proteomics approaches are to date not suitable for diagnostic purposes of COVID-19.

However, proteomics approaches might be powerful to study the biological factors contributing to COVID-19 severity and patient outcomes and identify potential predictive biomarkers. Patients with more severe COVID-19 infections accompanied by rapid deterioration of lung function are distinguished by significant immune dysregulation. The biological pathways that drive disease severity and lead to immune dysregulation remain poorly understood. The identification of differentially regulated proteins and pathways in plasma samples derived from uninfected and infected patient groups with varying disease severity (i.e., asymptomatic, moderate, and severe) would not only provide more information about the biological nature of the dysregulation, but might deliver biomarkers for predicting disease severity that can be measured noninvasively. The complexity and dynamic concentration range of the plasma proteome has posed challenges for reproducible and sensitive MS-based plasma proteome profiling¹⁸⁵. However, the performance of MS-based proteomics has matured, reaching a sensitivity and dynamic range which allows quantifying 500–700 proteins routinely and reproducibly across large patient cohorts, which makes it interesting for biomarker studies^{190,191}. These molecular measurements can then be correlated with clinical parameters to identify predictive biomarkers^{190,191}. Using unbiased proteomics to discover biomarkers for disease severity has the advantage that it might lead to more specific biomarkers for COVID-19 compared to established serological assays for cytokine and inflammatory proteins.

To identify potential biomarkers of COVID-19 severity, Messner et al. developed a low-cost and high-throughput platform that can handle 180 plasma samples within a single day¹⁹². Using sera and plasma from hospitalized patients they identified 27 putative proteins that are differentially expressed depending on the WHO severity grade of COVID-19¹⁹². These proteins include complement factors, proteins of the coagulation system, inflammatory modulators, and pro-inflammatory signaling molecules, thus capturing the host response to SARS-CoV-2

infection¹⁹². If validated in large independent patient cohorts, targeted proteomics assays, which in contrast to immunoassays do not rely on specific antibodies, could be developed rapidly for the proteomic signature to support clinical decision making and monitor treatment response¹⁸⁵.

Limited proteolysis-mass spectrometry (LiP-MS)

LiP-MS is a powerful approach for systematically characterizing protein conformational changes resulting from ligand or protein binding. In limited proteolysis (LiP), a broad specificity protease used at low enzyme to substrate ratio for a short time digests native protein extracts to generate large polypeptide fragments dependent on the protein's structural properties¹⁹³. The structure-specific protein fragments are then denatured and subjected to tryptic digestion to generate peptides that are amenable for MS analysis¹⁹³. An aliquot of the protein extract prior to LiP is fully digested with trypsin as a control and compared to the LiP sample¹⁹⁴. When analyzed by quantitative MS, the tryptic peptides containing the LiP sites, so-called conformotypic peptides, should have a lower abundance in the LiP digested sample compared to the trypsin-only control. Comparison of proteolytic signatures from samples that have been subjected to different conditions allows the identification of protein regions that underwent structural changes as a response to the perturbation¹⁹⁴.

Piazza et al. developed a chemoproteomics approach to systematically determine metabolite-protein interactions by combining LiP with quantitative MS on a cell extract in the presence and absence of metabolites to assess metabolite-induced structural alterations on a proteome-wide scale¹⁹⁵. The LiP-small molecule mapping (LiP-SMap) approach was applied to *Escherichia coli* to map the metabolite-protein interactions of 20 metabolites in an unbiased manner. This allowed not only the identification of 1678 metabolite-protein interactions with the majority being novel interactions, but also the determination of the structural regions that are affected by metabolite binding¹⁹⁵, which were found to be in close proximity to the binding site. Recently, the LiP-SMap approach was extended to enable systematic investigation of protein-small molecule interactions in complex eukaryotic proteomes¹⁹⁶. Due to the higher complexity of

the eukaryotic proteomes resulting in the initial discovery of multiple drug target candidates, Piazza et al. developed the machine-learning based LiP-Quant workflow, which performs LiP on lysate treated with a drug dilution series to score and prioritize target candidates¹⁹⁶. While the initial LiP-SMap identified 37 putative drug targets for cells treated with Rapamycin, the additional LiP-Quant scoring method could confirm FKBP1A as the highest-ranking candidate protein target¹⁹⁶. As of this publication, there have been no published studies using LiP-MS to study SARS-CoV-2 or any recently identified drugs that are in or being considered for clinical trials. However, in light of the ongoing large scale drug discovery efforts for COVID-19, the relatively simple experimental design of LiP-MS and its broad applicability make it an ideal technique to identify cellular targets of existing drugs in clinical trials or for prioritizing drugs or antibodies based on their off-target reactivity.

Thermal proteome profiling (TPP)

TPP combines cellular thermal shift assay (CETSA) with quantitative MS^{197,198}. The basic principle of TPP relies on denaturation and aggregation of proteins in cells at their intrinsic melting temperature, which results in solubility changes of proteins with increasing temperatures. By quantifying the abundance of the protein remaining in solution after subjecting cells to heat spanning a predetermined temperature gradient, a melting profile for a protein can be established. Protein conformational changes upon binding of small molecules or other ligands change the thermal stability of a protein and alter the melting profile. The combination of TPP with multiplexed quantitative MS-based proteomics can systematically determine melting profiles and melting temperature shifts upon drug treatment for thousands of expressed proteins simultaneously to discover drug targets and off-target binding.

As a proof of concept, TPP was applied to the promiscuous kinase inhibitors staurosporine and GSK3182571 with a known spectrum of targets, which induced shifts in the melting temperatures of many kinase targets¹⁹⁷. Interestingly, in addition to the proteins that are directly bound by the ligand, downstream pathway members like regulatory subunits of kinase

complexes displayed thermal stability shifts, likely as a result of altered PTMs, thus demonstrating that TPP could inform the drug's mechanism of action¹⁹⁷. The application of TPP was initially limited to soluble, mainly cytosolic proteins. However, since many ligand-binding receptors and drug targets represent transmembrane proteins, the method was extended to profile membrane-protein targets of small molecules by addition of a mild detergent during cell lysis which allows solubilization of membranes but does not solubilize protein aggregates¹⁹⁹. While initially only applied to cultured cells, most recently the method has been extended to tissue (tissue-TPP) and blood (blood-TPP) specimens²⁰⁰. Therefore, target and off-target binding could be characterized in primary cell studies, which is critical to predict adverse reactions of drugs. Following intravenous administration of the HDAC inhibitor panobinostat in mice, TPP identified Hdac1, Hdac2, and Ttc38 as known targets across all analyzed tissues, including spleen, liver, kidney, and lung. However, different off-target profiles were obtained across the different tissues which could be explained by heterogeneous levels of protein expression and drug exposure comparing the different tissues²⁰⁰. A recent study applied TPP to understand off-target effects of Remdesivir, a repurposed drug for treating COVID-19 patients, in uninfected HepG2 cells and identified the hexameric AAA+ ATPase Trip13 as a target of Remdesivir²⁰¹. Further experiments are necessary to investigate the relevance of the Remdesivir-Trip13 interaction for SARS-CoV-2, and in fact there is currently a lack of evidence that Remdesivir is effective against SARS-CoV-2²⁰².

Activity-based protein profiling (ABPP)

ABPP is a sophisticated chemical proteomic strategy that can be used to systematically interrogate cellular enzymes and discover in vivo inhibitors of enzymatic activity (reviewed by Niphakis et al.²⁰³ and Kahler et al.²⁰⁴). In ABPP, activity-based probes target a specific activity or structural feature of enzyme active sites in order to covalently modify the target with a reporter tag. The activity-based probe will have (1) a reactive group that forms the covalent bond to the target protein; (2) a binding group that directs the probe to active sites and typically mimics

natural substrates; and (3) a reporter/detection tag that allows measurement of the probe labeling. Depending on the assay and experimental design, labeled events can be visualized by fluorescence microscopy, flow cytometry, or SDS-PAGE, detected in vivo by radio-isotope detection, or can be enriched by affinity purification and analyzed by MS^{205,206}. Since probe labeling is dependent on active site accessibility, it can differentiate between inactive and active states of selective targets that would otherwise react to the probe, thus providing a unique opportunity to profile cells in various environmental, genetic, or conditional backgrounds. A number of enzymatic activities can be targeted, including cysteine proteases, cathepsins, kinases, metalloproteases, serine proteases, and oxidoreductases^{203,204,207}.

There are two main ABPP strategies that can be employed, those that are chemocentric and focused on the small molecule(s), and a target-centric approach which focuses on the enzyme. Chemocentric ABPP strategies can be used to discover cellular enzymes that are targets of specific drugs. This type of strategy is exemplified in an ABPP-based study that determined prodrug dimethyl fumarate (DMF), and not monomethyl fumarate (MMF) that DMF is converted to, is responsible for blocking activation of primary T cells in human and mice²⁰⁸. Activity-based probes have also been used to screen for drugs against specific targets. In an example of a target-based strategy, ABPP was used to identify targets of cellular ABHD2, a serine hydrolase involved in immune response, virus replication, and fertility²⁰⁹. In this study, novel inhibitors of ABHD2 that could be used to modulate sperm fertility were discovered and probed for off-target profiles. As of this publication, there have been no published studies using ABPP to study SARS-CoV-2 or any recently identified drugs that are in or being considered for clinical trials. One of the strengths of ABPP approaches is their applicability to detecting these events in cell models, primary cells, or in vivo animal models. Taken together, these approaches are particularly powerful for identifying protein/enzymatic cellular targets of drugs, characterizing and screening drugs against specific targets, and informing drug efficacy and safety as they are equipped to assess interaction and activity of compounds as well as identify off-target effects.

While ABPP methods are still very specialized in their broad application, they could provide unique insight into and help identify cellular targets of existing drugs in clinical trials or for prioritizing drugs based on their off-target reactivity for treatments of SARS-CoV-2.

Prioritizing candidates from large-scale SARS-CoV-2 drug screens

Large drug screening efforts are ongoing to identify drug candidates that could be repurposed as SARS-CoV-2 antivirals²¹⁰. To prioritize drugs demonstrating antiviral activity in vitro for further testing, it is imperative to assess target and off-target binding of promising small molecules from large-scale screens before moving them to animal models and testing them in humans. In light of the broad applicability of TPP, LiP-MS, and ABPP, these represent individually or in combination promising strategies for small molecule prioritization. To assess the effectiveness of novel or repurposed drugs for SARS-CoV-2 treatment in a noninvasive manner, drug discovery efforts could be accompanied by biomarker discovery and validation studies relying on global protein abundance and targeted proteomics approaches to identify biomarkers for monitoring treatment response in body fluids.

CONCLUSION

The SARS-CoV-2 pandemic necessitates the urgent study of viral protein structure, viral replication, virus–host interaction, and host response to gain a molecular understanding of pathogenicity and to investigate strategies for treatment options. The field of proteomics is well-positioned to inform open questions as discussed in this perspective, and indeed there is a precedent of important proteomic contributions in the study of other viral infections including HIV and influenza, among others. The pandemic has also impressed upon the need for cross-platform collaborations to quickly investigate, translate, and apply proteomic research into clinical outcomes. This type of rapid-response requires cross-discipline expertise and cooperation to not only generate, analyze, and integrate the data, but to interpret, functionally

validate, and translate findings into actionable hypotheses that can influence meaningful clinical studies¹⁸⁴.

Proteomic techniques are strikingly dynamic in the size and scale of the questions they can answer, including the study of viral infection from the structure of a single viral protein to tracking global changes in the human proteome during infection. Beyond size and scale, proteomics can also provide insight into timing and cellular location, which are critical to understanding how to combat SARS-CoV-2 infection and manage COVID-19 pathology. The application of existing proteomic tools to questions of SARS-CoV-2 biology is pressingly important, and this unprecedented pandemic also invokes an opportunity for creativity in developing new tools, and applying and integrating tools in new ways. As proteomic scientists, it is imperative that we contribute our unique proteomic perspective to open questions about SARS-CoV-2 biology, so we can provide an essential complement to studies in other fields as the scientific community works together to meet this monumental challenge. In fact, a number of studies have already demonstrated the impact and breadth of work that can be carried out by rapid response of international collaborations between interdisciplinary scientists.

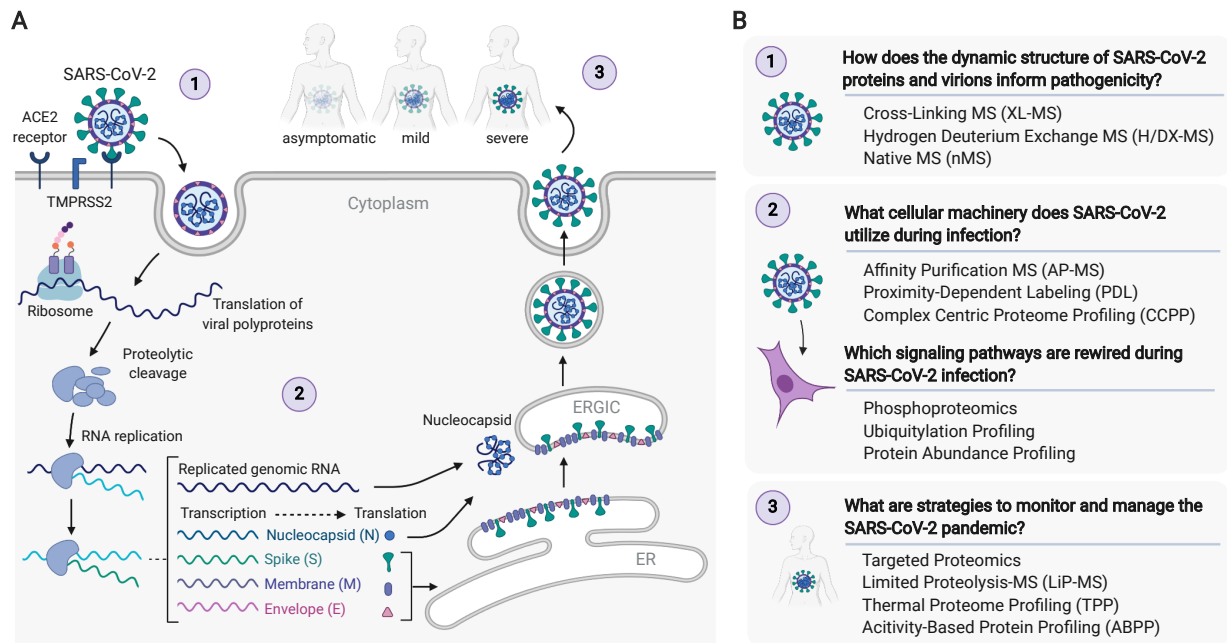


Figure 2.1 **A.** SARS-CoV-2 life cycle. **B.** Open questions to further our understanding of SARS-CoV-2 biology and proteomic techniques that can be leveraged to address these questions. Adapted from “Coronavirus Replication Cycle”, by BioRender.com (2020). Retrieved from <https://app.biorender.com/biorender-templates>.

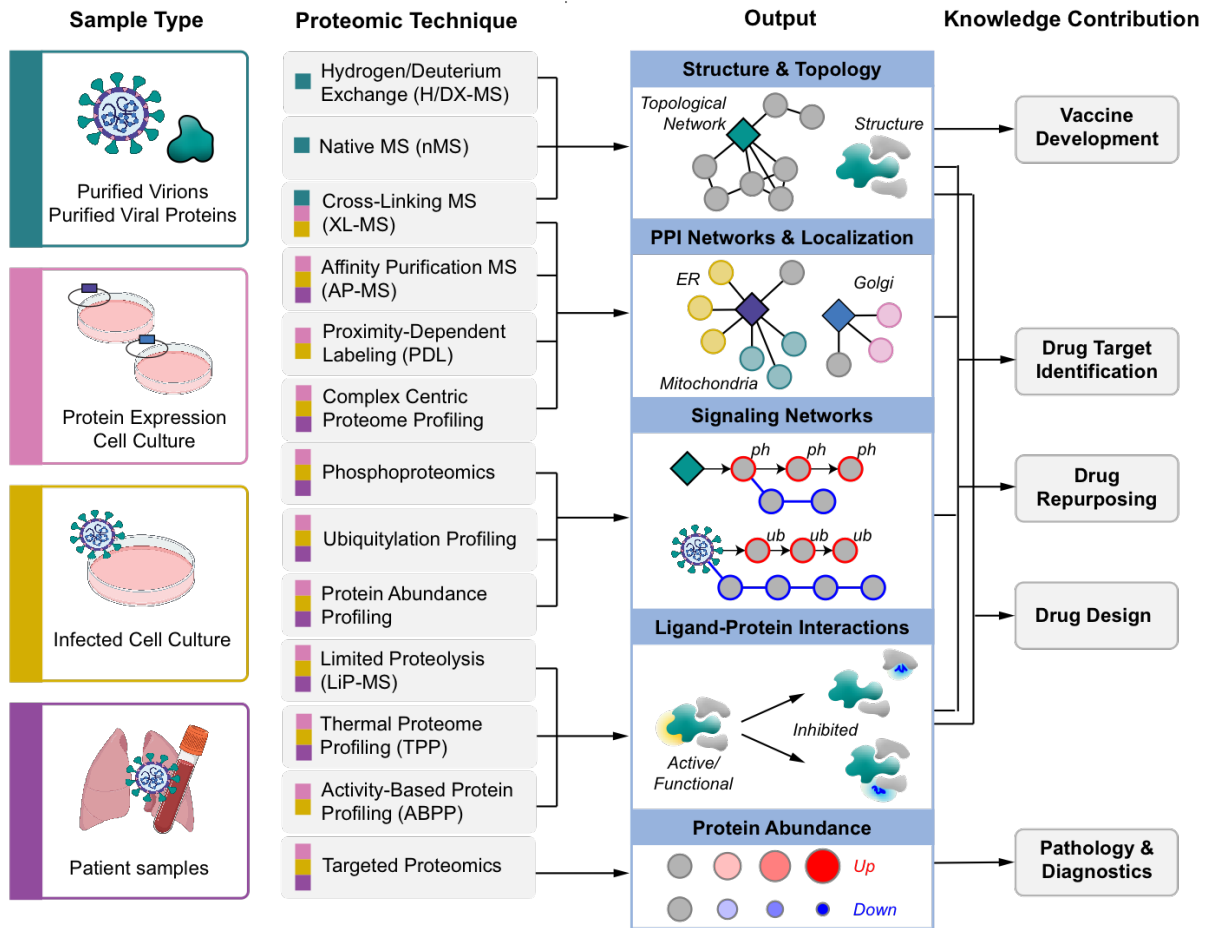


Figure 2.2 Overview of MS-based proteomics techniques proposed to study SARS-CoV-2, including sample types that can be used as input, molecular insights that can be obtained as output, and how the technologies can be integrated to inform SARS-CoV-2 biology and COVID-19 pathology.

Table 2.1 Proteomic studies on SARS-CoV-2 highlighted in the perspective.

Proteomic Technique	Sample Type	Study Objective	Reference
H/DX-MS	Purified protein	N protein structure	Ye <i>et al.</i> Architecture and Self-Assembly of the SARS-CoV-2 Nucleocapsid Protein. <i>bioRxiv</i> 2020. https://doi.org/10.1101/2020.05.17.100685 .
nMS	Purified protein	S-ACE2 virus-host protein complex structure; drug mechanism (heparin)	Yang <i>et al.</i> The Utility of Native MS for Understanding the Mechanism of Action of Repurposed Therapeutics in COVID-19: Heparin as a Disruptor of the SARS-CoV-2 Interaction with Its Host Cell Receptor. <i>Anal. Chem.</i> 2020. https://doi.org/10.1021/acs.analchem.0c02449 .
AP-MS	HEK293T cells expressing SARS-CoV-2 proteins	Virus-host protein-protein interactions; drug candidates	Gordon, D. E.; Jang, G. M.; Bouhaddou, M.; Xu, J.; Obernier, K.; White, K. M.; O'Meara, M. J.; Rezelj, V. V.; Guo, J. Z.; Swaney, D. L.; Tummino, T. A.; Huettnerlein, R.; Kaake, R. M.; Richards, A. L.; Tutuncuoglu, B.; Foussard, H.; Batra, J.; Haas, K.; Modak, M.; Kim, M.; Haas, P.; Polacco, B. J.; Braberg, H.; Fabius, J. M.; Eckhardt, M.; Soucheray, M.; Bennett, M. J.; Cakir, M.; McGregor, M. J.; Li, Q.; Meyer, B.; Roesch, F.; Vallet, T.; Mac Kain, A.; Miorin, L.; Moreno, E.; Naing, Z. Z. C.; Zhou, Y.; Peng, S.; Shi, Y.; Zhang, Z.; Shen, W.; Kirby, I. T.; Melnyk, J. E.; Chorba, J. S.; Lou, K.; Dai, S. A.; Barrio-Hernandez, I.; Memon, D.; Hernandez-Armenta, C.; Lyu, J.; Mathy, C. J. P.; Perica, T.; Pilla, K. B.; Ganesan, S. J.; Saltzberg, D. J.; Rakesh, R.; Liu, X.; Rosenthal, S. B.; Calviello, L.; Venkataramanan, S.; Liboy-Lugo, J.; Lin, Y.; Huang, X.-P.; Liu, Y.; Wankowicz, S. A.; Bohn, M.; Safari, M.; Ugur, F. S.; Koh, C.; Savar, N. S.; Tran, Q. D.; Shengjuler, D.; Fletcher, S. J.; O'Neal, M. C.; Cai, Y.; Chang, J. C. J.; Broadhurst, D. J.; Klippsten, S.; Sharp, P. P.; Wenzell, N. A.; Kuzuoglu, D.; Wang, H.-Y.; Trenker, R.; Young, J. M.; Cavero, D. A.; Hiatt, J.; Roth, T. L.; Rathore, U.; Subramanian, A.; Noack, J.; Hubert, M.; Stroud, R. M.; Frankel, A. D.; Rosenberg, O. S.; Verba, K. A.; Agard,

Proteomic Technique	Sample Type	Study Objective	Reference
			D. A.; Ott, M.; Emerman, M.; Jura, N.; von Zastrow, M.; Verdin, E.; Ashworth, A.; Schwartz, O.; d'Enfert, C.; Mukherjee, S.; Jacobson, M.; Malik, H. S.; Fujimori, D. G.; Ideker, T.; Craik, C. S.; Floor, S. N.; Fraser, J. S.; Gross, J. D.; Sali, A.; Roth, B. L.; Ruggero, D.; Taunton, J.; Kortemme, T.; Beltrao, P.; Vignuzzi, M.; García-Sastre, A.; Shokat, K. M.; Shoichet, B. K.; Krogan, N. J. A SARS-CoV-2 Protein Interaction Map Reveals Targets for Drug Repurposing. <i>Nature</i> 2020. https://doi.org/10.1038/s41586-020-2286-9 .
AP-MS	HEK293T cells expressing SARS-CoV-2, SARS-CoV-1, and MERS-CoV proteins	Virus-host protein-protein interactions; drug candidates	Gordon, D. E.; Hiatt, J.; Bouhaddou, M.; Rezelj, V. V.; Ulferts, S.; Braberg, H.; Jureka, A. S.; Obernier, K.; Guo, J. Z.; Batra, J.; Kaake, R. M.; Weckstein, A. R.; Owens, T. W.; Gupta, M.; Pourmal, S.; Titus, E. W.; Cakir, M.; Soucheray, M.; McGregor, M.; Cakir, Z.; Jang, G.; O'Meara, M. J.; Tummino, T. A.; Zhang, Z.; Foussard, H.; Rojc, A.; Zhou, Y.; Kuchenov, D.; Hüttenhain, R.; Xu, J.; Eckhardt, M.; Swaney, D. L.; Fabius, J. M.; Ummadi, M.; Tutuncuoglu, B.; Rathore, U.; Modak, M.; Haas, P.; Haas, K. M.; Naing, Z. Z. C.; Pulido, E. H.; Shi, Y.; Barrio-Hernandez, I.; Memon, D.; Petsalaki, E.; Dunham, A.; Marrero, M. C.; Burke, D.; Koh, C.; Vallet, T.; Silvas, J. A.; Azumaya, C. M.; Billesbølle, C.; Brilot, A. F.; Campbell, M. G.; Diallo, A.; Dickinson, M. S.; Diwanji, D.; Herrera, N.; Hoppe, N.; Kratochvil, H. T.; Liu, Y.; Merz, G. E.; Moritz, M.; Nguyen, H. C.; Nowotny, C.; Puchades, C.; Rizo, A. N.; Schulze-Gahmen, U.; Smith, A. M.; Sun, M.; Young, I. D.; Zhao, J.; Asarnow, D.; Biel, J.; Bowen, A.; Braxton, J. R.; Chen, J.; Chio, C. M.; Chio, U. S.; Deshpande, I.; Doan, L.; Faust, B.; Flores, S.; Jin, M.; Kim, K.; Lam, V. L.; Li, F.; Li, J.; Li, Y.-L.; Li, Y.; Liu, X.; Lo, M.; Lopez, K. E.; Melo, A. A.; Moss, F. R., 3rd; Nguyen, P.; Paulino, J.; Pawar, K. I.; Peters, J. K.; Pospiech, T. H., Jr; Safari, M.; Sangwan, S.; Schaefer, K.; Thomas, P. V.; Thwin,

Proteomic Technique	Sample Type	Study Objective	Reference
			<p>A. C.; Trenker, R.; Tse, E.; Tsui, T. K. M.; Wang, F.; Whitis, N.; Yu, Z.; Zhang, K.; Zhang, Y.; Zhou, F.; Saltzberg, D.; QCRG Structural Biology Consortium; Hodder, A. J.; Shun-Shion, A. S.; Williams, D. M.; White, K. M.; Rosales, R.; Kehrer, T.; Miorin, L.; Moreno, E.; Patel, A. H.; Rihn, S.; Khalid, M. M.; Vallejo-Gracia, A.; Fozouni, P.; Simoneau, C. R.; Roth, T. L.; Wu, D.; Karim, M. A.; Ghousaini, M.; Dunham, I.; Berardi, F.; Weigang, S.; Chazal, M.; Park, J.; Logue, J.; McGrath, M.; Weston, S.; Haupt, R.; Hastie, C. J.; Elliott, M.; Brown, F.; Burness, K. A.; Reid, E.; Dorward, M.; Johnson, C.; Wilkinson, S. G.; Geyer, A.; Giesel, D. M.; Baillie, C.; Raggett, S.; Leech, H.; Toth, R.; Goodman, N.; Keough, K. C.; Lind, A. L.; Zoonomia Consortium; Klesh, R. J.; Hemphill, K. R.; Carlson-Stevermer, J.; Oki, J.; Holden, K.; Maures, T.; Pollard, K. S.; Sali, A.; Agard, D. A.; Cheng, Y.; Fraser, J. S.; Frost, A.; Jura, N.; Kortemme, T.; Manglik, A.; Southworth, D. R.; Stroud, R. M.; Alessi, D. R.; Davies, P.; Frieman, M. B.; Ideker, T.; Abate, C.; Jouvenet, N.; Kochs, G.; Shoichet, B.; Ott, M.; Palmarini, M.; Shokat, K. M.; García-Sastre, A.; Rassen, J. A.; Grosse, R.; Rosenberg, O. S.; Verba, K. A.; Basler, C. F.; Vignuzzi, M.; Peden, A. A.; Beltrao, P.; Krogan, N. J. Comparative Host-Coronavirus Protein Interaction Networks Reveal Pan-Viral Disease Mechanisms. <i>Science</i> 2020, 370 (6521). https://doi.org/10.1126/science.abe9403</p>
AP-MS; phosphoproteomics; ubiquitylation profiling	A549 cells expressing SARS-CoV-2 proteins; ACE2-expressing A549 cells infected with SARS-CoV-2	Virus-host protein-protein interactions; transcriptome, proteome, phosphoproteome, and ubiquitome during infection; drug candidates	<p>Stukalov, A.; Girault, V.; Grass, V.; Bergant, V.; Karayel, O.; Urban, C.; Haas, D. A.; Huang, Y.; Oubraham, L.; Wang, A.; Hamad, S. M.; Piras, A.; Tanzer, M.; Hansen, F. M.; Engleitner, T.; Reinecke, M.; Lavacca, T. M.; Ehmann, R.; Wölfel, R.; Jores, J.; Kuster, B.; Protzer, U.; Rad, R.; Ziebuhr, J.; Thiel, V.; Scaturro, P.; Mann, M.; Pichlmair, A. Multi-Level Proteomics</p>

Proteomic Technique	Sample Type	Study Objective	Reference
			Reveals Host-Perturbation Strategies of SARS-CoV-2 and SARS-CoV, 2020, 2020.06.17.156455. https://doi.org/10.1101/2020.06.17.156455 .
PDL	A549 cells expressing SARS-CoV-2 proteins	Virus-host proximal protein interactions	Samavarchi-Tehrani, P.; Abdouni, H.; Knight, J. D. R.; Astori, A.; Samson, R.; Lin, Z.-Y.; Kim, D.-K.; Knapp, J. J.; St-Germain, J.; Go, C. D.; Larsen, B.; Wong, C. J.; Cassonnet, P.; Demeret, C.; Jacob, Y.; Roth, F. P.; Raught, B.; Gingras, A.-C. A SARS-CoV-2 – Host Proximity Interactome. https://doi.org/10.1101/2020.09.03.282103
Phosphoproteomics	Vero E6 cells infected with SARS-CoV-2	Phosphoproteome during infection; drug candidates	Bouhaddou, M.; Memon, D.; Meyer, B.; White, K. M.; Rezelj, V. V.; Correa Marrero, M.; Polacco, B. J.; Melnyk, J. E.; Ulferts, S.; Kaake, R. M.; Batra, J.; Richards, A. L.; Stevenson, E.; Gordon, D. E.; Rojc, A.; Obernier, K.; Fabius, J. M.; Soucheray, M.; Miorin, L.; Moreno, E.; Koh, C.; Tran, Q. D.; Hardy, A.; Robinot, R.; Vallet, T.; Nilsson-Payant, B. E.; Hernandez-Armenta, C.; Dunham, A.; Weigang, S.; Knerr, J.; Modak, M.; Quintero, D.; Zhou, Y.; Dugourd, A.; Valdeolivas, A.; Patil, T.; Li, Q.; Hüttenhain, R.; Cakir, M.; Muralidharan, M.; Kim, M.; Jang, G.; Tutuncuoglu, B.; Hiatt, J.; Guo, J. Z.; Xu, J.; Bouhaddou, S.; Mathy, C. J. P.; Gaulton, A.; Manners, E. J.; Félix, E.; Shi, Y.; Goff, M.; Lim, J. K.; McBride, T.; O’Neal, M. C.; Cai, Y.; Chang, J. C. J.; Broadhurst, D. J.; Klippsten, S.; De Wit, E.; Leach, A. R.; Kortemme, T.; Shoichet, B.; Ott, M.; Saez-Rodriguez, J.; tenOever, B. R.; Mullins, R. D.; Fischer, E. R.; Kochs, G.; Grosse, R.; García-Sastre, A.; Vignuzzi, M.; Johnson, J. R.; Shokat, K. M.; Swaney, D. L.; Beltrao, P.; Krogan, N. J. The Global Phosphorylation Landscape of SARS-CoV-2 Infection. <i>Cell</i> 2020, 182 (3), 685–712.e19.

Proteomic Technique	Sample Type	Study Objective	Reference
Abundance proteomics; phosphoproteomics	Vero E6 cells infected with SARS-CoV-2	Transcriptome, proteome, and phosphoproteome during infection	Davidson <i>et al.</i> Characterisation of the Transcriptome and Proteome of SARS-CoV-2 Reveals a Cell Passage Induced in-Frame Deletion of the Furin-like Cleavage Site from the Spike Glycoprotein. <i>Genome Med.</i> 2020, 12 (1), 68.
Phosphoproteomics	Caco-2 cells infected with SARS-CoV-2	Phosphoproteome during infection; drug candidates	Klann <i>et al.</i> Growth Factor Receptor Signaling Inhibition Prevents SARS-CoV-2 Replication. <i>Mol. Cell</i> 2020. https://doi.org/10.1016/j.molcel.2020.08.006 .
Abundance proteomics	Caco-2 cells infected with SARS-CoV-2	Translatome and proteome during infection; drug candidates	Bojkova <i>et al.</i> Proteomics of SARS-CoV-2-Infected Host Cells Reveals Therapy Targets. <i>Nature</i> 2020, 583 (7816), 469–472.
Targeted proteomics	Vero E6 cells infected with SARS-CoV-2	Diagnostic methods	Bezstarosti <i>et al.</i> Targeted Proteomics for the Detection of SARS-CoV-2 Proteins. <i>bioRxiv</i> , 2020, 2020.04.23.057810. https://doi.org/10.1101/2020.04.23.057810 .
Targeted proteomics	Patient samples (gargle)	Diagnostic methods	Ihling <i>et al.</i> Mass Spectrometric Identification of SARS-CoV-2 Proteins from Gargle Solution Samples of COVID-19 Patients. <i>bioRxiv</i> , 2020, 2020.04.18.047878.
PRM	Vero E6 cells infected with SARS-CoV-2; patient samples (nasopharyngeal swabs, bronchoalveolar lavage)	Diagnostic methods	Zecha <i>et al.</i> Data, Reagents, Assays and Merits of Proteomics for SARS-CoV-2 Research and Testing. <i>Mol. Cell. Proteomics</i> 2020, 19 (9), 1503–1522.
Targeted proteomics	Patient samples (nasopharyngeal and oropharyngeal swabs)	Diagnostic methods	Cardozo, K. H. M.; Lebkuchen, A.; Okai, G. G.; Schuch, R. A.; Viana, L. G.; Olive, A. N.; Lazari, C. D. S.; Fraga, A. M.; Granato, C. F. H.; Pintão, M. C. T.; Carvalho, V. M. Establishing a Mass Spectrometry-Based System for Rapid Detection of SARS-CoV-2 in Large Clinical Sample Cohorts. <i>Nat. Commun.</i> 2020, 11 (1), 6201.

Proteomic Technique	Sample Type	Study Objective	Reference
Abundance proteomics	Patient samples (sera and plasma)	Biomarkers of COVID-19 disease severity	Messner <i>et al.</i> Ultra-High-Throughput Clinical Proteomics Reveals Classifiers of COVID-19 Infection. <i>Cell Syst</i> 2020. https://doi.org/10.1016/j.cels.2020.05.012
TPP	HepG2 cells treated with compounds	Off-target effects of COVID-19 drug candidates (remdesivir, hydroxychloroquine, and more)	Friman <i>et al.</i> CETSA MS Profiling for a Comparative Assessment of FDA Approved Antivirals Repurposed for COVID-19 Therapy Identifies Trip13 as a Remdesivir off-Target. <i>bioRxiv</i> 2020.

CHAPTER 3

Genetic analysis of E3 ubiquitin ligases using CRISPR-Cas9 in primary human T cells identifies TRAF2 and UHRF1 as regulators of HIV infection and latency

SUMMARY

During HIV infection host ubiquitination pathways are both activated as innate defenses against viral infection, and hijacked by viral factors to establish an environment optimal for virus replication. E3 ligases are the substrate specific component of a three-step enzymatic ubiquitylation machinery in cells, and provide an ability to fine tune response to different conditions. To better understand the role diverse ubiquitination pathways play in HIV infection, we first used mass spectrometry-based global proteomics to define the set of ubiquitin ligases that are expressed in primary human CD4⁺ T cells. Out of the approximately 380 total ligases, we identified 116 that were detected in at least one of three donors. Using an optimized CRISPR-based arrayed assay, we systematically deleted each of the 116 E3s from activated primary CD4⁺ T cells and infected them with HIV-1 reporter virus. Following infection with HIV, we determined that 10 E3s had significant effects on infection with 7 demonstrating proviral activity, including UHRF1, and 3 with antiviral effects, including TRAF2. We found deletion of either TRAF2 or UHRF1 in three JLat and one primary cell model of latency results in increased HIV transcription, suggesting they may be excellent targets for HIV latency reversal strategies.

INTRODUCTION

Human immunodeficiency virus (HIV) is dependent on a complex network of host cell machinery to establish infection, replicate, and transmit in humans. HIV primarily infects CD4⁺ T cells via an interaction with the CD4 receptor and either CCR5 or CXCR4 co-receptors. Once inside the cell, the virus reverse transcribes its RNA genome into DNA, integrates into the host genome, transcribes viral mRNA, translates viral proteins, and produces virions that are released via budding^{13,14}. Before the discovery of antiretroviral therapy (ART), most individuals infected with HIV developed acquired immunodeficiency syndrome (AIDS). AIDS is marked by CD4⁺ T cell depletion that weakens the immune system, and as a result opportunistic infections can be deadly. Currently, ART is a combination of multiple drugs that inhibit key parts of the HIV

life cycle to suppress the viral load, which prevents the progression to AIDS and decreases the risk of transmission to others. However, ART is not an HIV cure because it does not eliminate the reservoir of latently infected cells.

The latent HIV reservoir is established early during infection^{211,212}. While the mechanisms governing the establishment of HIV latency are poorly understood, it occurs in cells in which the HIV proviral DNA is integrated into the host genome, but viral genes are not actively transcribed¹⁵⁻¹⁷. Because these cells are not producing virions, they are not killed by viral toxicity or immune clearance, and can persist for decades, with the ability to randomly reactivate when treatment is stopped^{18,19}. For this reason, ART must be lifelong. A number of strategies have been proposed to combat latency and achieve a functional HIV cure, including “block and lock”, where the goal is to establish a deep state of latency such that reactivation never occurs, and the strategy that we will focus on in this paper, which is “shock and kill”. The goal of “shock and kill” is to “shock” latently-infected cells with latency reversing agents (LRAs) to produce virions in the presence of ART, and these virion-producing cells are then “killed” by viral toxicity or the host immune system²¹³. For “shock and kill” to be curative, latency reversal must be universally penetrant across the latent reservoir, likely requiring several rounds of the “shock and kill” treatment cycle to ensure complete success. While many LRAs have been identified²¹⁴, a universally penetrant combination therapy strategy has not been determined yet, and the discovery of novel complementary and highly potent LRAs is still needed. Towards this end, the elucidation of host genes, proteins, and pathways that govern latency maintenance and reversal is essential to the identification of synergistic combinatorial LRA therapies.

Discovering the proviral and antiviral host factors that HIV respectively commandeers and evades will highlight vulnerabilities in the HIV life-cycle that can be exploited for host-directed therapy (HDT). Since host mutation events are rare, targeting factors encoded from the host genome limits the ability of the virus to develop escape mutations which can result in resistance. While targeting multiple HIV proteins and processes simultaneously can provide a

high enough burden as to limit escape mutations, as is the case with standard ART, these drug regimens require consistent life-long adherence or HIV will rebound and re-establish active infection^{19,26}. Lapses in ART treatment provide opportunities for HIV to mutate and develop drug resistance²⁷⁻²⁹, a concern avoided by HDT.

The host ubiquitination signaling pathway is essential for a number of cellular processes including protein degradation and quality control, cellular localization, molecular interactions, and therefore many pathogens hijack this system to promote their own replication and survival^{215,216}. Ubiquitin as a signaling molecule is covalently conjugated to protein substrates in an ATP-dependent process catalyzed by three enzymes: E1 ubiquitin activating enzymes; E2 ubiquitin conjugating enzymes; and E3 ubiquitin ligases which catalyze the transfer of ubiquitin to the protein substrate. Because E3 ubiquitin ligases recognize specific substrates, they are the most selective, and therefore E3 specific drugs could have relatively limited off-target effects compared to drugs targeting E1 or E2s²¹⁷. The mechanism by which E3 ligases attach ubiquitin to the protein substrate is determined by their type: really interesting new gene (RING) E3s facilitate ubiquitin transfer directly from the E2 to the substrate, human thyroid receptor interacting protein (HECT) E3s bind ubiquitin themselves before transfer to the substrate, and RING-in-between-RING (RBR) E3s contain RING domains but bind ubiquitin themselves similar to HECT E3s²¹⁸.

E3 ubiquitin ligases can support and block HIV infection during the viral replication cycle. For example, after entry, the HIV protein Vpu targets the host receptor CD4 for degradation via β TRCP, a subunit of a multi-subunit E3 ubiquitin ligase²¹⁹. Degradation of CD4 prevents super-infection of a cell and promotes the synthesis of infectious virions²²⁰. The cullin 5 RING E3 ligase complex is hijacked by the HIV protein Vif to promote degradation of host APOBEC3 (A3) restriction factors (i.e. A3D, A3F, A3G, and A3H)^{153,157}. Unchecked, this family of cytosine deaminases binds to and hypermutates the HIV genome, effectively producing non-infectious virions^{155,158}. An additional E3 ligase, ARIH2, helps catalyze the mono-ubiquitylation of A3

proteins prior to subsequent poly-ubiquitylation and proteasomal degradation²²¹. In addition, ubiquitin pathways regulate a number of key processes important for HIV latency reversal, including epigenetic regulation and transcription²²²⁻²²⁴.

The importance of ubiquitin signaling has been reflected in genetic screens of host proteins important for HIV infection where the genes identified are enriched for ubiquitination machinery and ubiquitin-regulated pathways^{225,226}. In addition to genetic analysis, proteomic studies of HIV-host protein-protein interactions have found host interactors that are enriched for roles in ubiquitin signaling^{122,227}. However, these genetic and proteomic studies were performed in immortalized cell lines, as primary cells are difficult to work with and are less genetically-tractable. The study of ubiquitin signaling in immortalized cell lines has limitations, however, because ubiquitin regulates processes like apoptosis and cell cycle, which are perturbed by immortalization. Given the importance of ubiquitin pathways in HIV replication and host processes related to HIV pathogenesis, there is a need for a comprehensive study that clearly identifies E3s important to HIV infection in physiologically relevant primary CD4⁺ T cells.

Here, we study the functional role of E3 ubiquitin ligases in the context of HIV-1 infection in physiologically-relevant primary CD4⁺ T cells isolated from healthy human donors. In order to determine E3 ligases important for regulating HIV infection, we first identified which E3 proteins are expressed in primary CD4⁺ T cells. Mass spectrometry analysis identified 116 E3s which we systematically deleted in an arrayed format in primary CD4⁺ T cells and challenged with HIV infection. Through this assay, we identified 10 E3s that regulate HIV infection both positively and negatively and using network propagation of an integrated HIV network, we mapped host pathways connecting HIV-relevant genes, proteins, and pathways to our functional E3s. Our E3-HIV network identified enrichments in TNF and non-canonical NF- κ B pathways, and we found that knockout of TRAF2 activates non-canonical NF- κ B signaling. Excitingly, we found that TRAF2 knockout alone reverses latency in three Jurkat cell line models of latency (JLats), as does knockout of the epigenetic regulator UHRF1. To further investigate a role for TRAF2 and

UHRF1 in latency reversal, we developed a novel, physiologically-relevant model of latency reversal in primary resting CD4⁺ T cells isolated from healthy human donors. Similarly to our results in JLat models, we found that TRAF2 and UHRF1 knockouts stimulated p24/HIV production intracellularly, indicating increased viral transcription. Thus, we have identified TRAF2 and UHRF1 as exciting candidate drug targets for the development of novel latency reversing agents toward an HIV cure.

RESULTS

Liquid chromatography tandem mass spectrometry (LC-MS/MS) analysis identifies 116 E3 ligases expressed in activated primary human CD4⁺ T cells

In order to characterize the E3 ligases relevant to HIV infection, we first identified which E3s are expressed in our model of HIV infection: activated primary CD4⁺ T cells. To this end, CD4⁺ T cells were isolated by positive selection from blood donated by three healthy individuals, activated with anti-CD2, anti-CD3, and anti-CD28 antibodies and harvested after 6 days. Activated CD4⁺ T cells were lysed, proteins were proteolytically digested, and resulting peptides were de-salted and analyzed by liquid chromatography tandem mass spectrometry (LC-MS/MS analysis) (Figure 3.1A). To be as inclusive as possible for our initial screen, we considered as expressed in CD4⁺ T cells any protein that was identified in any of the three donors by at least one unique peptide. In total, we identified 5422 unique proteins across all three donors, with 5077 shared in at least two donors, and 4480 identified in all three (Figure S3.1, Table S3.1).

For this study we were primarily interested in E3 ligases that function as single subunits rather than multisubunit ligases (e.g. cullin-ring ligases) that would presumably be more challenging to deconvolute the genetic knockout phenotypes. We cross-referenced our list of 5429 expressed proteins against a database of 377 annotated human E3 ligases that perform ubiquitin ligase activity alone, excluding multi-subunit E3 ligase complexes²²⁸. We identified 116 E3 ligases that were expressed at the protein level in primary activated CD4⁺ T cells from three

healthy human donors. A similar number of E3 ligases were detected in each donor, with 106 detected in Donor A, 106 in Donor B, and 96 in Donor C. There was considerable donor-to-donor overlap, with 85 E3s identified in all three donors and 22 E3s identified in two donors, while 9 E3s were identified in only one donor (Figure 3.1B). We detected E3s measured across low and high abundance ranges relative to all proteins identified in the dataset, with the majority of E3s falling at mid-to-low abundance (Figure 3.1C). These proteins reflect different E3 types and span a variety of different mechanisms of ubiquitin ligation. We detected all but two types that are classified in the original E3 database: RING (including IBR), HECT, UBR, UBox, and zf-C2H2 E3s. We detected 86 RING E3s (including 2 IBR E3s) out of 331 in the database, 12 out of 30 HECT E3s, 4 out of 5 UBR E3s, 4 out of 5 UBox E3s, and 1 out of 2 zf-C2H2 E3s (Figure 3.1C). To include expressed E3s that ligate other ubiquitin-like molecules like small ubiquitin-like modifier (SUMO) family proteins and ubiquitin-fold modifier 1 (UFM1), we identified 8 SUMO E3s and 1 UFM E3 as expressed (Figure 3.1C).

CRISPR-Cas9 knockouts in primary CD4⁺ T cells identify novel E3s that regulate HIV infection

In order to identify which E3s have proviral or antiviral activity against HIV-1 in primary activated CD4⁺ T cells, we used an arrayed CRISPR-based knockout spreading HIV-1 infection assay. Briefly, we isolated and activated CD4⁺ T cells from healthy human donors, electroporated pre-formed CRISPR ribonucleoproteins (crRNPs) in a 96-well arrayed format, and infected them with replication-competent HIV-1 NL4-3 Nef:IRES:GFP (Figure 3.2A)^{229,230}. Every plate consisted of an outer-well border plated with media (to limit edge effects), and 3 non-targeting controls (NTCs) which do not align to any protospacer adjacent motif (PAM)-adjacent region of the human genome and thus should not edit the genome or affect HIV infection. Each plate also included four positive controls, which are known dependency factors for HIV-1 NL4-3: CXCR4, CDK9, LEDGF, and ELOB^{153,155,231–236}. The remaining wells targeted

E3 ligases, with each E3 gene being targeted by three individually designed guides to increase the likelihood that at least one would have high knockout efficiency (Table S3.2). The E3 genes were knocked out in activated CD4⁺ T cells from two independent donors and each guide was electroporated into separate wells and plated in technical triplicate. Samples were collected at 2, 4, and 6 days post-inoculation and the percent of cells that were GFP positive (% GFP+) was measured by flow cytometry (Figure 3.2A).

To avoid confounding phenotypes driven by low viability, cells were gated on lymphocytes by forward and side scatter to estimate viability, and samples with less than 21.8% lymphocytes and a count of less than 15,000 lymphocytes were removed from the analysis due to low viability (Figure S3.2). Using these viability estimates, we assessed the toxicity of each guide for each gene by flow cytometry, and most guides displayed limited toxicity (Figure 3.2B). In total, 140 individual replicates were removed from the dataset due to low viability (Figure S3.2, Table S3.3). The median percentage of HIV infection varied between donors and replicate plates, therefore to standardize across experiments we normalized by donor and plate. To this end, the % GFP+ was measured for each of the wells containing the 3 NTCs, 4 positive controls targeting HIV-1 dependency factors, and 65 E3 targeting guides. For every well the % GFP+ was first normalized to the median result of 3 NTCs. The normalized values were then averaged across the three technical replicates and the averaged value was log₂ transformed.

After normalizing the results, our NTCs and positive controls behaved as expected, with individual NTCs having minimal effect on HIV infection (less than a log₂ fold change (L2FC) or 1), and positive controls decreasing HIV infection by at least 2-fold (Figure 3.2C). CXCR4 is a co-receptor required for entry of the HIV-1 NL4-3 virus and on average decreased HIV infection by XX. LEDGF is an HIV Integrase cofactor that aids the integration of the viral genome into the host genome; CDK9 is an HIV Tat cofactor that aids viral transcription; and ELOB is an HIV Vif cofactor that aids in the degradation of the restriction factor APOBEC3G^{153,155,231–236}. Knocking out these genes decreased HIV infection at least 2-fold, while individual NTCs did not affect HIV

infection in either direction more than 2-fold (Figure 3.2C). Based on these results from the NTCs and positive controls, we set a threshold of L2FC >1 or <-1.

It is important to note that from these positive control genes, ELOB knockouts had the weakest phenotype, ranging from a L2FC of -2.23 to 0.89 with a median of -0.65 over 40 observations (Table S3.3). The median L2FC for ELOB does not make the threshold, which may be due to the mechanism by which ELOB acts during HIV infection. HIV co-opts ELOB to combat the antiviral activity of APOBEC3G and ensure the formation of infectious progeny virions¹⁵⁵. Therefore, knocking out ELOB will not affect the first but rather the subsequent rounds of infection. At the latest timepoint after infection (6 days), at which subsequent infections would have time to occur, ELOB has a median L2FC of -1.11, which does make the threshold. This observation highlights the importance of multiple timepoints to capture early and late-acting factors, and further reinforces the threshold of L2FC >1 or <-1.

Of the 116 expressed E3s, TRIM21 and PHRF1 did not have a pre-designed CRISPR guide RNA available for order, and were therefore removed from our study. We also included two E3s not detected by proteomic analysis: TRAF3 and BIRC3, as we identified other members of the NF- κ B-inducing kinase (NIK) regulatory complex in which they function (TRAF2 and BIRC2) in our proteomic analysis and wanted to probe this complex further.

We considered an E3 a hit if the knockout passed our L2FC >1 or <-1 threshold in the same direction at two timepoints in at least 2 guides in the same donor or the same guide in 2 donors. By defining biological replicates as either the same guide in two donors or two guides within the same donor, we maximized the potential to observe hits while maintaining rigor. For example, in the case where one donor reaches HIV infection saturation, it can be difficult to observe knockouts that increase HIV infection. In this case, having three biologically distinct replicates in the second donor provides greater confidence in those results. On the other hand, if only one of the three guides targeting a gene achieves high efficient editing, the other two guides may look as if there was no effect on HIV infection, however in this case having both

donors as biological replicates for the same guide can provide confidence in the results for a single guide. Using this criteria, we found 10 E3s that affect HIV infection: 7 with proviral activity (MARCH5, ZFP91, UHRF1, VPS18, NOSIP, PPIL2, RING1); and 3 with antiviral activity (TRAF3, TRAF2, PRPF19) (Figure 3.2D). Notably, 5 of these E3s make a more stringent cutoff of a L2FC >1.5 or <-1.5, and represent higher-confidence hits: TRAF2, TRAF3, MARCH5, UHRF1, and ZFP91 (Table S3.3).

These 10 E3s are enriched for roles regulating NF- κ B and transcription. Looking at the 5 high-confidence hits, TRAF2 and TRAF3 are TNF-associated factors that regulate NF- κ B and c-Jun N-terminal kinase (JNK) signaling, and are both members of the NIK regulatory complex that inhibits non-canonical NF- κ B signaling. Conversely, ZFP91 is an activator of non-canonical NF- κ B signaling by stabilizing NIK via K63 ubiquitination^{237,238}. ZFP91 physically interacts with HIV-1 Tat in HEK293T cells¹²² and was required for Tat-driven HIV-1 transcription in a HeLa cell model²³⁹. MARCH5 is a mitochondrial outer-membrane protein that regulates innate immunity, including activation of NF- κ B and toll-like receptor 7 (TLR7)^{240,241}. MARCH5 as an activator of TLR7 signaling may connect to HIV, as the TLR7 agonist GS-9620 was found to reverse HIV latency via Interferon (IFN)- α signaling²⁴². Given that TLR7 is not expressed in T cells (Table S3.1)²⁴³, the observed MARCH5 phenotype may be mediated by another member of the pathway or an alternative pathway altogether, such as mitochondrial antiviral signaling (MAVS) or NF- κ B^{241,244}. UHRF1 is an epigenetic regulator that recruits DNA methyltransferases and chromatin proteins. UHRF1 has recently been implicated in HIV latency maintenance, as it silences the HIV-1 promoter by DNA methylation and its downregulation in ex-vivo HIV+ peripheral blood mononuclear cells (PBMCs) increased latency reversal²⁴⁵.

Looking at the other 5 E3 hits, PRPF19 is a splicing factor that is also involved in transcription and DNA repair. While it does not have a known connection to HIV, it has been found to be a proviral factor for Influenza A H1N1 infection²⁴⁶. PPIL2 is part of the cyclophilin family and has been shown to bind to HIV-1 gag^{247,248}. RING1 is associated with the polycomb

group protein complex (PRC1) and acts as a transcriptional repressor, and PRC1/RING1 are involved in HIV latency^{249–252}. NOSIP is a regulator of nitric oxide production, is associated with homeostatic proliferation of T cells, and is differentially expressed in cells latently infected with HIV^{253,254}. VPS18, which is involved in vesicle-mediated protein trafficking, has previously been reported to have a role in the production and release of infectious virions²⁵⁵.

Notably, none of these ten genes were identified in four previous genome-wide siRNA screens performed in cell lines to identify host factors affecting HIV infection^{225,256–258}. This discrepancy may be due in part to differences in our methodologies, from the approach to genetic manipulation (CRISPR-Cas9 gene knockout vs. siRNA/shRNA knockdown), the delivery (electroporation vs. transfection/transduction), the cell types used (primary CD4⁺ T cells vs. cell lines), and the scope of study (E3s detected by proteomics vs. genome-wide). While other studies have implicated some of these E3s for roles in HIV infection, this is the first time that these 10 E3 ligases have been shown to affect HIV infection in physiologically-relevant primary activated CD4⁺ T cells.

Network propagation analysis connects TRAF2, non-canonical NF- κ B signaling, and HIV infection

To more systematically assess connections between our top 10 E3s and the pathways hijacked by HIV, we performed a network propagation-based analysis²⁵⁹. Network propagation is used to understand how genes of interest are interconnected within large biological networks that capture known functional and physical interaction information. Here, we independently propagated (1) human proteins associated with HIV extracted from pathway databases and (2) the top 10 E3s from this study across a large network comprised of ReactomeFI²⁶⁰, CORUM²⁶¹, and known HIV-host complexes¹²². We integrated the propagated networks gene-wise by multiplication and performed a permutation test to identify and extract the significant subnetwork, which we clustered into 30 smaller subnetworks. We then manually annotated

connections between any of the 30 subnetwork pieces and the 10 E3s identified in our screen (Figure 3.3A).

Of the 30 HIV pathway subnetworks we uncovered, 19 contained at least one of the 10 E3s from our knockout screen in HIV infected CD4⁺ T cells, including four of the five higher-confidence hits: TRAF2 (11 clusters), TRAF3 (9 clusters), ZFP91 (2 clusters), and MARCH5 (1 cluster). (Figure S3.3). Expectedly, TRAF2 and TRAF3 overlapped in 9 clusters, 5 of which had connections to NF- κ B signaling and function, including tumor necrosis factor mediated signaling which is an upstream activator of NF- κ B, activation of innate immune response (which represents multiple pathways including NF- κ B), and NIK NF- κ B signaling which is involved in activation of the non-canonical NF- κ B pathway. Consistent with these analyses, TRAF2 and TRAF3 are non-canonical NF- κ B inhibitors that function together in a complex with BIRC2 and BIRC3, a connection captured in cluster 8 (Figure 3.3B). This complex is also important for HIV latency reversal, as BIRC2 inhibition by AZD5582 was recently shown to increase transcription of HIV mRNA and reverse latency^{262,263}. Cluster 8 also contains NFKB1 and RELA, both transcription factors that bind the NF- κ B promoter in canonical NF- κ B signaling. Since our analysis and previous work^{264,265} connected TRAF2 and TRAF3 to NF- κ B signaling, we tested the effect of each gene knockout, alone or in combination with two latency reversing agents (LRAs), AZD5582 and GS-9620, for their ability to activate canonical and non-canonical NF- κ B. For comparison, the other high confidence E3 ligases MARCH5, ZFP91 and UHRF1 along with CXCR4 and NTCs were also tested. Western blot analysis was used to monitor non-canonical NF- κ B activation (processing of p100 to p52) and canonical NF- κ B activation (degradation of I κ B α). As expected, treatment with AZD5582 increased p52 levels in all cells tested whereas TLR7 agonist GS-9620 did not, a result that was anticipated as the latter drug is proposed to work only in dendritic cells (Figure 3.3C, Figure S3.5)²⁴². Deletion of TRAF2, but not the other genes tested, resulted in increased p52 levels, demonstrating that it is a regulator of non-canonical NF- κ B signaling (Figure 3.3C, Figure S3.5). Surprisingly, TRAF3 did not share this

phenotype, possibly owing to the fact that TRAF2 could be playing a role independent of TRAF3 and/or incomplete genetic ablation of TRAF3. More work will be required to understand the functional differences between TRAF2 and TRAF3 in these and other assays connected to HIV infection.

Interestingly, our analysis revealed several other connections between the E3 ligases uncovered in our screen and the clusters related to HIV function defined by network propagation. For example, the splicing factor PRPF19 was present in 10 clusters enriched for roles in TNF signaling, NF- κ B signaling, transcription and DNA repair (Figure S3.3, Figure S3.4). Also, PPIL2, a cyclophilin family member, was found in clusters C17 and C26, which are linked to transcription and mRNA processing (Figure S3.3, Figure S3.4). The nuclear factors RING1 and ZFP91 were enriched in clusters related to nuclear transport (C5) and transcription (C6) (Figure S3.3, Figure S3.4). Further studies will aid in understanding these and many of the other connections revealed by the network propagation analysis.

Deletion of TRAF2 and UHRF1 reverses HIV latency in JLat and primary resting CD4⁺ T cell models

While only TRAF2 was found to regulate non-canonical NF- κ B signaling, the other 4 high-confidence E3s were still connected to pathways connected to latency regulation and reversal, such as viral transcription, epigenetic regulation, and LRAs. To test whether our 5 high-confidence E3s regulate HIV latency reversal, we deleted each gene in Jurkat cell models of latency (JLats) using CRISPR. We used three JLat cell lines that have an integrated HIV promoter and a GFP reporter with low baseline expression, including 11.1s (long terminal repeat (LTR)-full length Δ env Δ Nef-GFP), A2s (LTR-Tat-GFP), and A72s (LTR-GFP)²⁶⁶. TRAF2, TRAF3, UHRF1, ZFP91, and MARCH5 were deleted in each JLat cell line by electroporation with pre-formed crRNPs using the guide that displayed the strongest phenotype in our previous screen. Cells recovered for 8 days after electroporation and were treated with media or TNF α , a

potent LRA, for 24 hours, and then were collected for analysis by flow cytometry (Figure 3.4A). TNF α , a canonical NF- κ B activator, was used because of the enrichment for TNF signaling and NF- κ B in the network propagation. The percentage of GFP+ cells was measured and compared to not only NTCs but also deletion of CXCR4, since its role in HIV entry is not relevant to JLats and should not have impact on latency reversal.

As expected, the percentage of untreated NTC cells expressing GFP is very low in all three cell lines (Figure 3.4B, Table S3.5). Compared to untreated cells, treatment with TNF α induces GFP expression in 11.1 cells by 19-fold, in A2s by 109-fold, and in A72s by 56-fold (Figure 3.4B, Table S3.5). In comparison to untreated NTC cells, deletion of TRAF2 increased GFP expression in 11.1, A2, and A72 cell lines 6-fold, 59-fold, and 27-fold, respectively (Figure 3.4B, Table S3.5). While knockout of TRAF2 knockout was not toxic, addition of TNF α to 11.1 cells resulted in toxicity that fell below our 30% viability threshold (Figure S3.6A, Table S3.5). However, TNF α -treated TRAF2 knockouts in A2s and A72s synergistically increased GFP expression 1.5-fold and 2.3-fold, respectively, compared to TNF α -treated NTCs (Figure S3.6B, Table S3.5).

While UHRF1 deletion in 11.1 cells resulted in toxicity based on our threshold (Figure S3.6A, Table S3.5), deletion in A2s and A72s increased GFP expression 27.4-fold and 22.6-fold, respectively (Figure 3.4B, Table S3.5), consistent with recent data connecting UHRF1 to HIV latency reversal²⁴⁵. Deletion of either CXCR4, TRAF3, ZFP91, or MARCH5 did not increase GFP expression in any of the latency cell systems used. (Figure 3.4B, Figure S3.6B).

To see if either TRAF2 and UHRF1 knockouts could impact LRA effectiveness in activating HIV transcription, we treated these deletions with a panel of LRAs, including JQ1, AZD5582, and GS-9620^{242,262,267,268}. JQ1 and AZD5582 induced GFP expression in JLats whereas TLR7 agonist GS-9620 did not, presumably because the TLR7 pathway is not present in these cells (Figure 3.4C)²⁴³. Unfortunately, deletion of UHRF1 combined with any of the LRAs resulted in toxicity in JLats (Figure S3.6C, Table S3.5), and treatment of TRAF2 knockouts with

TNF α , PHA, and PMA+Ionomycin also resulted in toxicity (Figure S3.6C, Figure S3.6D, Table S3.5). However, treatment of TRAF2 knockouts with JQ1 increased GFP expression compared to corresponding untreated cells, suggesting they may target different latency-reversing pathways (Figure 3.4C)²⁶⁹. Indeed, JQ1 is a p-TEFb activator that is known to have synergistic effects on reactivation of latent HIV in combination with NF- κ B activators²⁷⁰. In contrast, combination of AZD5582 with deletion of TRAF2 did not increase GFP expression (Figure 3.4C), a potential epistatic relationship that could imply that TRAF2 works in the same pathway that AZD5582 targets.

In order to test the latency reversal potential of TRAF2 and UHRF1 in a more physiologically relevant system, we developed a novel assay for testing latency reversal directly in primary resting CD4⁺ T cells (Figure 3.4D). To this end, we set up a system where resting CD4⁺ T cells were isolated from three healthy human donors by negative selection and 24 hours later, they were infected with a replication competent HIV-1 virus strain (HIV-LAI.2 WT, Subtype B, CXCR4 tropic) by spinoculation. They were immediately cultured in the presence of Saquinavir, a protease inhibitor that prevents the formation and release of virions, ultimately promoting latent infections²⁷¹. Three days following infection, the cells were electroporated with crRNPs in technical triplicate using either a non-targeting control guide (NTC-03, Table S3.2), a negative control HIV guide (a combination of three guides, two of which target HIV trans-activation response element (TAR) and one that targets the HIV LTR), or the TRAF2/UHRF1 guide resulting in the strongest phenotype in our previous screen (TRAF2-01 and UHRF1-01, Table S3.2). The HIV infection rate was 17.0-25.5% in all three donors on the day of electroporation (Table S3.6).

Cells were incubated with Saquinavir for 5 more days, which was then removed and replaced with fresh media containing Raltegravir, an integrase inhibitor. Removal of Saquinavir allows successful production of virions from latently infected cells, while Raltegravir inhibits integration of any newly produced viral particles, thereby preventing viral spread and allowing

an accurate estimation of the latency reversal^{271,272}. Cells were grown in the presence of Raltegravir for 2 days, at which point the supernatant was collected for p24 enzyme-linked immunosorbent assay (ELISA) and the cells were fixed and stained for intracellular HIV-1 p24. We achieved 50% knockout of TRAF2 and 74% knockout of UHRF1 in donor 1, 86% knockout of TRAF2 and 60% knockout of UHRF1 in donor 2, and 79% knockout of TRAF2 and 68% knockout of UHRF1 in donor 3, determined by tracking of Indels by DEcomposition (TIDE) analysis²⁷³. Intracellular p24 was measured by flow cytometry and supernatant p24 was measured by ELISA. Intracellular levels reflect latency reversal at the level of HIV transcription and translation, while p24 levels in the supernatant measure the assembly and release of virions.

Intracellular p24 staining measured by flow cytometry showed increased p24+ cells in TRAF2 knockouts compared to NTC in all three donors, increasing from an average of 5.91% p24+ cells to 10.2% in donor 1, 4.18% to 7.12% in donor 2, and 4.06% to 11.67% in donor 3 (Figure 3.4E), representing statistically significant fold changes of 1.73, 1.70, and 2.88 respectively. UHRF1 knockout also resulted in a significant increase in intracellular p24 staining in two out of three donors, increasing from an average of 5.91% p24+ cells to 11.33% in donor 1, and 4.06% to 10.26% in donor 3 (Figure 3.4E), corresponding to fold changes of 1.92 and 2.53. These results suggest that both TRAF2 and UHRF1 play a role in HIV latency and are involved in regulating HIV transcription. We also measured p24 levels in cell supernatants by ELISA and found an increase in TRAF2 knockouts compared to NTC in one of three donors (6.03-fold in donor 3, fold change > 1.5 and p value > 0.05) (Figure S3.7B, Table S3.6). Similarly, p24 levels in cell supernatant were higher in UHRF1 knockouts compared to NTC in two of three donors (1.89-fold in donor 1 and 7.39-fold in donor 3, both fold change > 1.5 and p value > 0.05), correlating with the two donors that showed an increase in intracellular p24. (Figure S3.7B, Table S3.6).

DISCUSSION

In this study, we have investigated the role of E3 ubiquitin ligases in HIV infection and latency directly in primary human CD4⁺ T cells. Ubiquitin signaling regulates many processes that exist in primary cells, including apoptosis and cell cycle, that are perturbed in immortalized cell lines. Therefore, we decided to probe the role of E3 ligases directly in primary T cells and identified more therapeutically valuable targets for HIV infection as well as latency reversal. To accomplish this, we first identified proteins expressed in activated primary CD4⁺ T cells from three healthy donors by mass spectrometry and found 113 E3s were expressed in at least one donor. Using our recently developed arrayed CRISPR-Cas9 knockout screening pipeline for studying HIV-host interactions in primary activated CD4⁺ T cells²³⁰, we systematically and individually deleted 116 expressed E3s (and related family members). We found that 10 had a significant impact on HIV infection both positively (3 antiviral: TRAF2, TRAF3, PRPF19) and negatively (7 proviral: MARCH5, ZFP91, UHRF1, VPS18, NOSIP, PPIL2, RING1). Previous systematic genetic screens of HIV host factors in cell lines^{225,256–258} did not overlap with any of our 10 E3 hits, emphasizing the importance of using physiologically-relevant, primary CD4⁺ T cells. While we have hypotheses about the role of these 10 E3s in the HIV lifecycle (Figure S3.8), further work will be required to elucidate the mechanism by which they are functionally affecting infection.

Through a computational process called network propagation²⁵⁹, we used a number of HIV datasets and databases to find connections between several of these E3 ligases with a number of biological pathways and processes connected to HIV infection, including NF-κB signaling. Guided by this analysis, we found that TRAF2 is a key regulator of non-canonical NF-κB signaling in primary T cells. Consistent with this, previous work has shown other proteins that function in the same pathway as TRAF2, notably BIRC2 and BIRC3, are also involved in regulating NF-κB signaling^{262,274,275}. Interestingly, deletion of TRAF3, which is known to function with TRAF2, did not have an effect on NF-κB signaling. These results suggest that TRAF2 is

functioning independently of TRAF3 and more work will be needed to understand the functional differences between these two related ubiquitin ligases.

To further study the E3 ligases found to be functionally-relevant to HIV infection, especially the ones that provided more significant effects (TRAF2, TRAF3, MARCH5, ZFP91, UHRF1), we used HIV latency models to test their potential for use in HIV cure strategies. We found that deletion of either TRAF2 or UHRF1 resulted in reversal of HIV latency in multiple JLat models. Deletion of TRAF2 and UHRF1 resulted in spontaneous reactivation of latent HIV transcription, and did not require treatment with LRAs to observe latency reversal caused by gene deletion. Treatment of gene knockouts with LRAs informed pathways of action, as TRAF2 deletion treated with TNF α (canonical NF- κ B activator) and JQ1 (p-TEFb activator) was synergistic, suggesting that TRAF2 deletion functions in different pathways, while treatment with AZD5582 (non-canonical NF- κ B activator) was epistatic, suggesting that TRAF2 functions in this pathway. Treatment of UHRF1 knockouts with LRAs was toxic and therefore less informative.

To probe latency reversal in a more physiologically-relevant system, we developed a novel latency reversal assay in resting primary human CD4⁺ T cells. This model has overcome previous limitations of modeling latency in resting T cells, including achieving enough HIV infection to observe latency reversal phenotypes and achieving efficient gene editing in primary resting CD4⁺ T cells. While our JLat models of latency are Env-deleted and therefore only report HIV transcription and translation, our primary cell model allows us to probe both HIV transcription and translation as well as virion production. The primary cell model also recapitulates multiple integration sites of latently-infected cells, while clonal Jurkat models have only a single integration site. Further, the signaling pathways and molecules regulating latency reversal, especially ubiquitin pathways that are of special interest in this study, are closer to physiological conditions in the primary cell model than in JLat cell lines.

In our novel primary cell model of HIV latency, TRAF2 deletion increased intracellular p24 in all three donors, and UHRF1 deletion increased intracellular p24 in two of three donors.

These results are consistent with the phenotypes observed in JLat cells and suggest transcriptional and translational reactivation of latent HIV-1. TRAF2 deletion increased supernatant p24 in one of three donors, and further work will be required to confirm this increase in virion production observed in the single donor is a significant, donor-dependent phenotype, or if TRAF2 deletion increases HIV-1 transcription but not virion production. UHRF1 deletion increased supernatant p24 in two of three donors, and was consistent with the two of three donors that showed increased intracellular p24, indicating that UHRF1 deletion does increase virion production. For both intracellular and supernatant p24 data, it is notable that TRAF2 and UHRF1 deletions alone reversed latency without treatment with LRAs.

The mechanism by which TRAF2 knockouts increase HIV infection and reverse latency requires further investigation, however we have proposed a model based on our literature review and observation of non-canonical NF- κ B activation by Western blot (Figure 3.5). The NIK regulatory complex is comprised of TRAF2, TRAF3, BIRC2, and BIRC3 and functions to suppress non-canonical NF- κ B signaling. TRAF2 binds BIRC2 and BIRC3, which K48 ubiquitinate NIK to lead to its degradation. TRAF2 also binds TRAF3, which binds NIK to bring it into proximity for ubiquitination by BIRC2 and BIRC3. When NIK is degraded, the non-canonical NF- κ B pathway is inactive. However, upon TRAF2 knockout, BIRC2 and BIRC3 are no longer in proximity to NIK for K48 ubiquitination, and NIK is allowed to accumulate. NIK then phosphorylates I κ ka, which phosphorylates p100. p100 is normally sequestered in the cytoplasm, however, upon phosphorylation it is recognized by SCF ^{β TRCP} ubiquitin ligase complex and processed into p52. p52 in complex with RELB can then translocate to the nucleus and bind the HIV LTR to promote transcription of viral mRNA. Increased viral transcription leads to increased active HIV infection (seen in our screen) and latency reversal (as seen in JLat and primary cell models).

However, our evidence suggests that TRAF2 is likely affecting more pathways than just non-canonical NF- κ B. For example, the combination of TRAF2 knockout and LRA treatment

with BIRC2 inhibitor AZD5582 did not reverse latency in JLats nearly to the same extent as with TRAF2 knockout alone. BIRC2 and TRAF2 are members of the same NIK regulatory complex, yet there is no observed synergistic effect of combined TRAF2 knockout and BIRC2 inhibition. In addition, TRAF3 knockout also did not reverse latency in JLats to the same extent observed in the TRAF2 knockout. Elucidating additional pathways by which TRAF2 reverses HIV latency will be important for development of novel LRAs that take advantage of TRAF2-mediated signaling pathways and regulation of HIV latency reversal.

In addition to TRAF2, we found that UHRF1 is another key regulator of HIV infection and HIV latency. UHRF1 deletion decreased active infection in primary activated CD4⁺ T cells, and increased latency reversal in JLat and primary resting CD4⁺ T cell models. The mechanism of action of UHRF1 during HIV infection requires further investigation, but we have proposed a model based on a recent pre-print publication that suggests a role for UHRF1 in epigenetic regulation of HIV-1 (Figure 3.5)²⁴⁵. UHRF1 may recruit DNA methyltransferases to silence the HIV-1 promoter by DNA methylation, resulting in repressed HIV transcription to maintain HIV latency. In this recent study, knockdown of UHRF1 in JLat models and drug inhibition of UHRF1 in *ex vivo* HIV+ patient samples reactivated HIV-1 transcription and virion production. Consistent with this, we have observed in our study that UHRF1 knockout increased HIV-1 transcription in JLat models and increased both transcription and virion production in a resting primary human CD4⁺ T cell model of latency.

Interestingly, in our study, UHRF1 deletion decreased active infection but increased latency reversal. This suggests two independent roles for UHRF1 at different stages of HIV infection, where UHRF1 is required for the establishment of initial infection, but then hinders HIV-1 production and contributes to the maintenance of latency. Both roles can be considered proviral, as UHRF1 is needed for effective initial infection, and its maintenance of latency aids long-term infection in the host. While a potential mechanism for UHRF1 in latency maintenance

has been recently described²⁴⁵, future work is needed to identify a mechanistic role for UHRF1 in initial HIV infection.

All together, we have comprehensively studied E3 ligases expressed in primary human CD4⁺ T cells for a functional role in HIV infection and identified 10 E3s enriched for roles in NF- κ B regulation, two of which also have a functional role in HIV latency reversal. We have developed a novel assay for studying the impact of gene deletions on latency reversal in resting primary human CD4⁺ T cells that is of use to the broader HIV research community. Deletion of TRAF2 and UHRF1 reverses latency in JLat and primary cell models, making them exciting candidates for development of novel LRAs, as well as further mechanistic study into their role in HIV infection and latency.

METHODS

Cell lines

Jurkat (JLat) cell lines 11.1, A2, and A72²⁶⁶ were cultured in complete Roswell Park Memorial Institute (RPMI) media, consisting of RPMI-1640 media (Corning) with 10% FBS (Gibco), 50ug/ml penicillin-streptomycin (Fisher), and 2mM L-Glutamine (Corning) at 37°C/5% CO₂/humid/dark. JLat cell lines were a generous gift from Melanie Ott's lab.

HEK293T cell line was cultured in complete Dulbecco's Modified Eagle Medium (DMEM) media consisting of DMEM media (Corning) with 10% FBS (Gibco) and 50ug/ml penicillin-streptomycin (Corning) at 37°C/5% CO₂/humid/dark.

Isolation and activation of primary CD4⁺ T cells

CD4⁺ T cells were isolated from healthy human donors using residuals from leukoreduction chambers after Trima Apheresis (Blood Centers of the Pacific, Vitalant). By positive selection, CD4⁺ T cells were isolated by the FABian cell isolation system (Cellclopedia) and IBA

Lifesciences CD4 isolation kit for FABian (IBA Lifesciences, Fisher Scientific). By negative selection, peripheral blood mononuclear cells (PBMCs) were isolated by Ficoll centrifugation using SepMate tubes (STEMCELL, per manufacturer's instructions). PBMCs were used fresh or frozen at 250 million cells in 1ml of 90% FBS 10% DMSO and stored in liquid nitrogen. CD4⁺ T cells were isolated from PBMCs by magnetic negative selection using an EasySep Human CD4⁺ T Cell Isolation Kit (STEMCELL, per manufacturer's instructions). Isolated CD4⁺ T cells were activated for 72 hours by plate-bound CD3 (Tonbo Biosciences clone UCHT1) (150ul of 10ug/ml CD3 in PBS in non-TC treated 48-well plates, incubated for at least 12 hours at 4°C, then aspirated before adding cells) and 5ug/ml CD28 (Tonbo Biosciences clone CD28.2) in complete RPMI media, consisting of RPMI-1640 media (Corning) with 10% FBS (Gibco), 50ug/ml penicillin-streptomycin (Corning), 5mM sodium pyruvate (Corning), 5mM (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) (Hyclone), and 20IU/ml IL-2 (Miltenyi Biotec) at 37°C/5% CO₂/humid/dark.

Sample preparation for global abundance proteomic analysis of primary CD4⁺ T cells

Primary CD4⁺ T cells from three healthy human donors were activated by anti-CD2/anti-CD3/anti-CD28 beads (Miltenyi Biotec T cell activation/expansion kit, human) at a 1:1 bead:cell ratio and cultured in complete media (RPMI-1640 media (Corning) with 10% FBS (Gibco), 50ug/ml penicillin-streptomycin (Corning), 5mM sodium pyruvate (Corning), 5mM (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) (Hyclone), and 20IU/ml IL-2 (Miltenyi Biotec)) at 37°C/5% CO₂/humid/dark for 6 days. Cells were pelleted by centrifugation at 400g for 5 min and washed 3 times with PBS. After the last wash, cell pellets were resuspended in 100ul of lysis buffer (8M urea, 0.1M ABC pH 8, 150mM NaCl, 1 mini-cOmplete protease inhibitor, 1 PhosSTOP phosphatase inhibitor brought to 10ml with H₂O) and frozen at -80°C. Samples were thawed on ice and lysed by probe sonication in three pulses of 20% amplitude for 15 seconds. Lysates were clarified by centrifugation at 16,100x g at 4°C for 30 min. Protein concentration

was measured by Bradford assay (Thermo) and 0.16 mg of total protein was reduced with 4 mM tris(2-carboxyethyl)phosphine (TCEP) for 30 min at room temperature and alkylated with 10 mM iodoacetamide for 30 min at room temperature in the dark. Excess iodoacetamide was quenched with 10 mM 1,4-dithiothreitol (DTT) for 30 min at room temperature in the dark. The samples were diluted four-fold in 100 mM ammonium bicarbonate, pH 8.0, to final urea concentration of 2 M. Samples were incubated with 4 µg of sequencing grade modified trypsin (Promega) and incubated at room temperature with rotation for 18 hr. The digests were acidified by addition of 10% trifluoroacetic acid (TFA) to a final concentration of 0.3% trifluoroacetic acid (pH ~2). Insoluble material was removed by centrifugation at 16,000 g for 10 min. Peptides were desalted using UltraMicroSpin C18 columns (The Nest Group). The columns were activated with 200 µL of 80% acetonitrile (ACN), 0.1% TFA, and equilibrated 3 times with 200 µL of 0.1% TFA. Peptide samples were applied to the columns, and the columns were washed 3 times with 200 µL of 0.1% TFA. Peptides were eluted with 140 µL of 50% ACN, 0.25% formic acid and lyophilized. The samples were fractionated using the Pierce High pH Reversed-Phase Peptide Fractionation Kit (Thermo cat. no. 84828) according to the manufacturer's instructions. Samples were dissolved in 0.1% TFA, bound to the column, and washed with water. Any peptides not binding to the column (flow-through) or eluted during the water wash (column wash) were also analyzed by MS. Eight fractions per sample were then eluted by a stepwise gradient of acetonitrile (5, 7.5, 10, 12.5, 15, 17.5, 20, and 50% acetonitrile) in 0.1% triethylamine. All samples were lyophilized prior to MS analysis.

Mass spectrometry analysis for global abundance proteomics of primary CD4⁺ T cells

Fractionated samples were resuspended in 4% formic acid, 4% acetonitrile solution, separated by reversed-phase HPLC using a Thermo Easy n1200 LC (Thermo Scientific) using an in house packed integrafrit column (360 µm O.D. x 75 µm I.D.) packed with 25 cm of 1.8 µm Reprosil C18 particles (Dr. Maisch-GMBL). Mobile phase A consisted of 0.1% FA in water and mobile

phase B consisted of 80% acetonitrile (ACN)/0.1% FA. Peptides were separated at flow rate 300nL/minute by the following 2hr gradient: 8% to 18% B over 52 minutes; 18-38% B over 56 minutes; and 10 min at 88% B. Eluting peptides were analyzed by an Orbitrap Fusion Lumos Tribrid Mass Spectrometer (Thermo Scientific). Data was collected in positive ion mode with MS1 detection in profile mode in the orbitrap using 120,000 resolution, 350-1350 m/z scan range, 25 ms maximum injection, and an AGC target of 5e5. MS2 fragmentation was performed on charge states from 2-5, MIPS mode = peptide, with a 40s dynamic exclusion after a single selection, and 10ppm +/- mass tolerance. MS2 data was collected in centroid mode at a turbo scan rate in the ion trap with HCD (32% normalized collision energy), 15ms maximum injection time, 2e4 AGC, 0.7mz quadrupole isolation window, and 120 m/z first mass.

All raw MS data were searched with MaxQuant (v 1.6.2.6) against the human proteome (Uniprot canonical protein sequences downloaded March 21, 2018). Peptides and proteins were filtered to 1% false discovery rate. MaxQuant default parameters were used with the exception that label-free quantification was turned on, with match between runs set to 0.7 min. Mass spectrometry data files (raw and search results) have been deposited to the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>) via the PRIDE partner repository with dataset identifier PXD028127 (user name: reviewer_pxd028127@ebi.ac.uk, password OAiNusVR).

CRISPR-Cas9 RNP generation and electroporation in activated primary CD4⁺ T cells

Lyophilized tracrRNA (Horizon) and crRNA (all guide sequences designed by Horizon, Table S3.2) were resuspended at 160uM in 10 mM Tris-HCL (7.4 pH) with 150 mM KCl. crRNPs were made by incubating 5ul of 160uM tracrRNA with 5ul of 160uM guide RNA and incubating for 30 minutes at 37°C, then 10ul of 40uM Cas9 (Macrolab) were added and incubated for 15 minutes at 37°C. Stocks of crRNPs were plated in 3.5ul aliquots in a 96-well plate and frozen for up to 6 months before use. Primary CD4⁺ T cells from 2-3 healthy human donors were ready for

electroporation after 3 days of activation by plate-bound CD3 and in-solution CD28. 3.5ul of crRNPs were pipette mixed with 0.3×10^6 primary activated CD4⁺ T cells in 20ul of Lonza electroporation buffer P3 and electroporated in Lonza 96-well nucleocuvette plates with code EH-115 using the Lonza 4D nucleofector core unit and Lonza 96-well shuttle device (Lonza). 100ul complete media (RPMI-1640 media (Corning) with 10% FBS (Gibco), 50ug/ml penicillin-streptomycin (Corning), 5mM sodium pyruvate (Corning), 5mM (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) (Hyclone), and 20IU/ml IL-2 (Miltenyi Biotec)) was added immediately after electroporation and cells were rested for 30 minutes at 37°C/5% CO₂/humid/dark. Cells were then transferred to 96 well flat-bottom plates, brought to 200ul with complete media, and cultured with anti-CD2/anti-CD3/anti-CD28 beads (Miltenyi Biotec T cell activation/expansion kit, human) at a 1:1 bead:cell ratio at 37°C/5% CO₂/humid/dark. Cells were fed with complete media every 2-3 days until they were plated into replicates.

Replica plating edited activated primary CD4⁺ T cells before HIV infection

Six days after electroporation, activated primary CD4⁺ T cells were plated into 3 technical replicates in 96-well U-bottom plates and brought up to 150ul with complete media RPMI-1640 media (Corning) with 10% FBS (Gibco), 50ug/ml penicillin-streptomycin (Corning), 5mM sodium pyruvate (Corning), 5mM (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) (Hyclone), and 20IU/ml IL-2 (Miltenyi Biotec). Plates were “edgeless”, meaning that 200ul of media was plated instead of sample in rows A and H and columns 1 and 12, to prevent edge effects attributed to greater levels of evaporation. Experimental replica plated cells were infected with HIV one day after cells were plated in technical triplicate. In addition to the experimental replica plated cells, 2x60ul of cells were collected and stored to analyze editing efficiency for each sample at the DNA and protein levels. Cellular DNA was collected from one sample by resuspending cells in QuickExtract DNA extraction solution (Lucigen) and incubating at 65°C for 20 min then 95°C for 20 min. Protein samples were collected by resuspending cells in 2.5x

Laemmli Sample Buffer (25mM Tris pH 6.8, 8% glycerol, 0.8% SDS, 2% 2-mercaptoethanol, 0.02% bromophenol blue) and incubation at 95°C for 20 min.

HIV virion production

15cm plates were seeded with 5e6 HEK293T cells in 25ml complete media (DMEM media (Corning) with 10% FBS (Gibco) and 50ug/ml penicillin-streptomycin (Corning)) 24 hours before transfection and cultured at 37°C/5% CO₂/humid/dark. Per 15cm plate, 10ug of HIV-1 NL4-3 Nef:IRES:GFP or HIV-1 LAI.2 plasmid DNA was mixed with 250ul serum-free DMEM media, and separately, 30ul PolyJet DNA In Vitro Transfection Reagent (Signa Gen Laboratories) was mixed with 250ul serum-free DMEM media (Corning). Plasmid DNA and PolyJet solutions were combined, pipette mixed twice, and incubated for 15 min at room temperature. The mixture was then added dropwise to seeded HEK293T cells in complete media and incubated in a BSL-3 setting at 37°C/5% CO₂/humid/dark for 48 hours. Supernatants were collected and kept at 4°C, and cells were fed with 25ml complete media and incubated at 37°C/5% CO₂/humid/dark for 24 hours. After 24hrs, the second set of supernatants were collected and added to the previously collected supernatants. Combined supernatant was filtered through a 0.22um PVDF filter (Steriflip). For virus precipitation, 5.5ml 50% PEG-6000 was mixed with 2.5ml 4M NaCl, to which 24ml of supernatant was added. Samples were inverted 5 times to mix and incubated 4-8 hours at 4°C. Precipitated virus was pelleted at 3500rpm for 20 min at 4°C and resuspended in 500ul PBS. All prepared virus was combined and made into ≤1ml aliquots in 2ml micro tubes, flash frozen on dry ice, and stored at -80°C until use.

HIV spreading infection in activated primary CD4⁺ T Cells

2.5ul of HIV-1 NL4-3 Nef:IRES:GFP was mixed with 47.5ul complete media (RPMI-1640 media (Corning) with 10% FBS (Gibco), 50ug/ml penicillin-streptomycin (Corning), 5mM sodium pyruvate (Corning), 5mM (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES)

(Hyclone), and 20IU/ml IL-2 (Miltenyi Biotec)) and added to replica plated cells (150ul of cells in complete media in 96 well U-bottom plates) to achieve ~2% infection after 48 hours. Plates were centrifuged at 1200 xg for 2 hours at room temperature before incubating at 37°C/5% CO₂/humid/dark. 2 days after infection, 75ul of sample was added to 75ul 2% paraformaldehyde in PBS and remaining samples were fed with 75ul complete media. 4 days after infection, 75ul of sample was added to 75ul 2% paraformaldehyde in PBS and remaining samples were fed with 75ul complete media. 6 days after infection, 150ul of sample was added to 50ul 4% paraformaldehyde in PBS and any remaining sample was discarded.

Flow cytometry and computational analysis of GFP-positive (HIV-infected) edited primary activated CD4⁺ T cells

Fixed cells were analyzed by an Attune NxT Acoustic Focusing Cytometer (ThermoFisher), recording all events in a 100uL sample volume after one 150uL mixing cycle. Data were exported as FCS3.0 files and analyzed with a consistent template in FlowJo. Cells were gated on live lymphocytes, singlets, excluding autofluorescence, and then the percentage of GFP+ cells was quantified. The data was exported to CSV, including the count and percentage of lymphocytes, the count and percentage of GFP+ cells, and the count and percentage of autofluorescent cells. The data was imported into R using RStudio and annotated with the sample information including donor, target gene, guide, days after infection, plate, and well. Wells with less than 21.8% lymphocytes or less than a count of 15,000 lymphocytes were excluded from analysis. Each well was normalized to the median of 3 non-targeting controls on its same plate for lymphocyte and GFP+ counts and percentages. After normalization, technical triplicates were averaged for the count and percentage of lymphocytes and the count and percentage of GFP+ cells and reported as a Log₂ fold change. The non-normalized averages for lymphocyte count and percentage and GFP+ count and percentage were also reported, and the data was exported as a CSV. For hit-calling, hits were defined as genes whose knockout

yielded a log2 fold change in infection ≥ 1 or ≤ -1 in the same direction in at least two timepoints for (1) two guides within the same donor, or (2) the same guide in two donors.

Network propagation integrative analysis

We performed a network propagation-based analysis to identify pathways and protein complexes that converged between genes with known association with HIV pathogenesis and the top E3s identified in this study. Specifically, we used a heat-diffusion kernel analogous to random walk with restart (RWR, also known as insulated diffusion and personalized PageRank) which better captures the local topology of the interaction network compared to a general heat diffusion process. The process is captured by the steady-state solution as follows:

$$P_{SS} = \alpha(I - (1 - \alpha)W)^{-1}P_0 \quad (\text{eq. 1})$$

where P_{SS} represents the vector of propagated values at steady-state, P_0 is the initial labeling (genes of interest from molecular studies), W is the normalized version of the adjacency matrix of the underlying network (in this implementation $W = AD^{-1}$, where A is the unnormalized adjacency matrix, and D is the diagonal degree matrix of the network), I is the identity matrix, and α denotes the restart probability (here, $\alpha=0.2$), which is the probability of returning to the previously visited node, thus controlling the spread through the network.

To create our base network, we merged ReactomeFI²⁶⁰, CORUM²⁶¹, and HIV-human complexes¹²². We then performed two independent propagations: one for human genes associated with HIV and the other for the top 10 E3s in this study. For the former, we extracted all genes that were members of pathway terms containing the word “HIV” in the MSigDB pathway database (c2.cp.v7.4.symbols.gmt) and were also members of the base network ($n=282$). To seed each propagation, genes of interest were labeled in a binary on or off fashion. After propagation, the two propagated networks were integrated by multiplying across them, gene-wise. Such an operation is used to create a gene list ranked to prioritize genes with high scores from both propagated datasets. To control for nodes with high degree (i.e. many

connections), which due to their heightened connectivity are biased to receive higher propagation scores, we conducted a permutation test. Specifically, we simulated random propagations by shuffling the positive scores to random genes, repeating this 20,000 times per propagation run. Next, we derived an empirical p-value by calculating the fraction of random propagation runs greater than or equal to the true propagation run for each gene.

The significant subnetwork was created by extracting genes with p-value ≤ 0.05 from the base network, requiring they form a singular connected component (i.e. possessed at least one connection to another gene within this set; $n=447$). The resulting significant subnetwork was clustered into smaller subnetwork clusters using the `cluster_walktrap` function (iGraph; `steps=10`) in conjunction with the `cophenetic` function in R (stats package) to calculate a distance matrix²⁷⁶. The `cluster_walktrap` function was weighted using weights derived from similarity of gene co-membership within Gene Ontology and pathway (c2.cp.v7.4.symbols.gmt and c5.all.v7.1.symbols.gmt from MSigDB as well as CORUM) terms. The final dendrogram was cut (i.e. clustered) using the `dynamicTreeCut` package in R (`minClusterSize = 7`, `deepSplit = 4`, `method = 'hybrid'`). This resulted in 30 subnetwork clusters (#0-29; Figure S3.3). Lastly, Gene Ontology (GO) enrichment analysis (biological process) was performed for each of the 30 resulting subnetwork clusters to identify biological processes associated with each cluster (Figure S3.4).

AZD5582 and GS-9620 treatment and collection of edited activated primary CD4⁺ T cells to probe for NF- κ B activation by Western blot

Primary activated CD4⁺ T cells were isolated, activated, and edited as described above with guides NTC-02, MARCH5-01, TRAF2-01, TRAF3-03, UHRF1-01, ZFP91-02 (designed by Horizon, Table S3.2). 8 days after editing, activated primary CD4⁺ T cells were treated with 100nM of AZD5582 (a generous gift from Sumit Chanda's lab), 100nM GS-9620 (Cayman Chemical), or untreated with complete media (RPMI-1640 media (Corning) with 10% FBS

(Gibco), 50ug/ml penicillin-streptomycin (Corning), 5mM sodium pyruvate (Corning), 5mM (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) (Hyclone), and 20IU/ml IL-2 (Miltenyi Biotec)). Cells were collected 24 hours after treatment by resuspension in 2.5x Laemmli Sample Buffer (25mM Tris pH 6.8, 8% glycerol, 0.8% SDS, 2% 2-mercaptoethanol, 0.02% bromophenol blue) and incubation for 20 minutes at 95°C. 10ul of sample were separated by SDS-PAGE on precast Criterion 26-well 4-20% TGX gels (BioRad) at 90 Volts for 100 minutes. Proteins were transferred to PVDF membranes at 4°C for 90 minutes at 25 milliAmps in 20% methanol transfer buffer. PVDF membranes were blocked with 5% milk in TBST for 1 hour. Blocked membranes were incubated with primary antibodies against p100/p52 (1:1000) (Cell Signaling Technology), IκBα (1:1000) (Cell Signaling Technology), or GAPDH (1:1000) (Sigma) at 4°C rocking overnight, washed 6 times for 5 minutes each with TBST, and incubated with secondary antibody (goat anti-rabbit IgG-HRP (1:5000-1:10,000) (Cell Signaling Technology) or goat anti-mouse IgG-HRP (1:5000) (Cell Signaling Technology)) for 1 hour. Membranes were washed with TBST 6 times for 5 minutes and incubated with 2ml of Pierce ECL western blotting substrate (for anti-IκBα and anti-GAPDH) (Thermo Fisher Scientific) or SuperSignal West Pico PLUS chemiluminescent substrate (for anti-p100/p52) (Thermo Fisher Scientific) for 3-5 minutes. Blots were exposed with autoradiography film (Thomas Scientific), and developed with a medical film processor (Konica Minolta Medical & Graphic).

CRISPR-Cas9 editing and LRA treatment of 11.1, A2, and A72 JLat cells

Editing was performed as described above in activated primary CD4⁺ T cells with the following exceptions: crRNA guides used were NTC-02, MARCH5-01, TRAF2-01, TRAF3-03, UHRF1-01, ZFP91-02 (designed by Horizon, Table S3.2), no stimulation beads were used, and complete media was RPMI-1640 media (Corning) with 10% FBS (Gibco), 50ug/ml penicillin-streptomycin (Fisher), and 2mM L-Glutamine (Corning). 300,000 JLat cells were resuspended in 20ul Lonza electroporation buffer SE (Lonza) and electroporated with code CM-137. Cells were fed with

complete media every 2-3 days until drug treatment. 8 days after electroporation with crRNPs, JLat cells were treated with 10ng/ml TNFalpha (PeproTech), 100 nM AZD5582 (a generous gift from Sumit Chanda's lab), 100 nM GS-9620 (Cayman Chemical), 10ug/ml PHA (a generous gift from Melanie Ott's lab), 16mM PMA (Ott lab) + 0.5uM Ionomycin (Ott lab), 625 nM JQ1 (Ott lab), or untreated with complete media. Cells were collected 24 hours after treatment by resuspension in 1% paraformaldehyde in PBS.

Flow cytometry and computational analysis of GFP-positive edited 11.1, A2, and A72 JLat cells

Fixed cells were analyzed by an Attune NxT Acoustic Focusing Cytometer (ThermoFisher), recording all events in a 100uL sample volume after one 150uL mixing cycle. Data were exported as FCS3.0 files and analyzed with a consistent template in FlowJo. Cells were gated on live lymphocytes, singlets, and then the percentage of GFP+ cells was quantified. The data was exported to CSV, including the count and percentage of lymphocytes and the count and percentage of GFP+ cells. 9 replicates (3 biological replicates with 3 technical replicates each) were averaged for the percentage of lymphocytes and percentage of GFP+ cells and standard deviation was calculated. Data was removed if the average percentage of lymphocytes was below the viability cutoff of 30%. For the 7 guides/2 treatments experiment (Figure 3.4B, Figure S3.6B), fold change and p value were calculated by comparison with the NTC of the same treatment within the same cell line. Significance was defined as a fold change ≥ 1.5 and a p value < 0.05 . For the 3 guides/7 treatments experiment (Figure 3.4C, Figure S3.6D), fold change and p value were calculated by comparison with the untreated condition of the same gene knockout within the same cell line. Significance was defined as a fold change ≥ 1.5 and a p value < 0.05 .

CRISPR-Cas9 RNP generation for electroporation in resting primary CD4⁺ T cells

crRNA guides targeted TRAF2 (TRAF2-01, Table S3.2) and UHRF1 (UHRF1-01, Table S3.2), a non-targeting control guide (NTC-03, Table S3.2) and a mixture of three guides targeting HIV; two targeting HIV-1 TAR and one targeting HIV-1 LTR region (TAR1:

GTTAGACCAGATTTGAGCCT, TAR2: GTTAGACCAGATTTGAGCCT, LTR:

AGAGCTCCCAGGCTCAAATC). Lyophilized tracrRNA (Horizon) and crRNA were resuspended at 160uM in 10 mM Tris-HCL (7.4 pH) with 150 mM KCl. Pre-formed crRNPs were made by incubating 1.5ul of 160uM tracrRNA with 1.5ul of 160uM guide RNA and incubating for 30 minutes at 37°C. Next, 1.2ul of 100nmol/ul Single-stranded donor oligonucleotides (ssODN; sequence:

TTAGCTCTGTTTACGTCCCAGCGGGCATGAGAGTAACAAGAGGGTGTGGTAATATTACGGT ACCGAGCACTATCGATAACAATATGTGTCATACGGACACG, IDT) was added and incubated at 37°C for 5 minutes. Next, 3ul of 40uM Cas9 (Macrolab) was added slowly taking care to avoid precipitation, pipetting up and down several times to ensure complete resuspension of the RNP complex, and incubated for 15 minutes at 37°C. crRNPs were aliquoted at 7.2ul and frozen at -80°C for up to 6 months before use as described in the next section. Along with guides targeting genes of interest, each experiment also included a non-targeting control guide (Horizon NTC#3, U-007503-01) and a mixture of three guides, two targeting HIV-1 TAR and one targeting HIV-1 LTR region (TAR1: GTTAGACCAGATTTGAGCCT, TAR2: GTTAGACCAGATTTGAGCCT, LTR: AGAGCTCCCAGGCTCAAATC).

Modeling effect of gene deletions on HIV latency reversal in resting primary CD4⁺ T cells

Primary human CD4⁺ T cells were isolated from PBMC enriched leukapheresis products (Leukopaks) from three healthy donors, following informed written consent (StemCell Technologies), by magnetic negative selection (StemCell, 17952) to avoid activation. These cells were cultured in resting-cell complete RPMI media (rcRPMI; RPMI-1640 media (Corning)

with 10% FBS (Gibco), 50ug/ml penicillin-streptomycin (Corning), 5mM sodium pyruvate (Corning), 5mM (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) (Hyclone), 10IU/ml IL-2 (Miltenyi Biotec), and 5ng/ml IL-7 (R&D Systems)). The next day, unstimulated primary CD4⁺ T cells were spinoculated with a replication-competent HIV strain (LAI.2, subtype B, CXCR4 tropic) as described above. The infected cells were immediately cultured at 5e6 cells/ml density in rcRPMI media in the presence of 10uM Saquinavir for five days to promote viral integration and latency as described in the Greene model of latency²⁷¹. The media was replenished after every two days. After five days of culturing in Saquinavir, the cells were electroporated with crRNPs by mixing 7.2 ul of crRNP-ssODN solution with 1.5e6 unstimulated HIV-infected CD4⁺ T cells in 18ul of Lonza electroporation buffer P2. The electroporation was done in Lonza 96-well nucleocuvette plates with code EH-110 on a Lonza 4D nucleofactor core unit and Lonza 96-well shuttle device (Lonza). After the nuke, the cells were again cultured in the presence of saquinavir for five more days to allow degradation of the proteins encoded by targeted genes. After which, they were moved to crRPMI media containing 60uM Raltegravir, a potent HIV integrase inhibitor to allow reactivation of the latent virus while preventing spreading infection. After two days of culture in the presence of Raltegravir, approximately 10,000 cells were collected from each sample for TIDE analysis and the remaining cells were pelleted down for intracellular P24 staining. The supernatant was collected, frozen and stored at -80°C until use for p24 ELISA.

Measurement of latency reversal in edited resting primary CD4⁺ T cells by intracellular p24 staining and flow cytometry

Pelleted cells were first stained extracellularly with Ghost Dye Violet 510 (Tonbo Biosciences) for live-dead discrimination. Briefly, cells were pelleted at 300g for 5 minutes in a v-bottom 96-well plate and the media was removed. Cells were suspended in a 1:500 dilution of the antibody in MACS buffer (0.5% bovine serum albumin (BSA) and 2 mM EDTA in PBS) at a concentration

of roughly 10^3 cells/ μ L and incubated for 20 minutes at RT. Cells were pelleted again, washed with MACS buffer, and suspended in 1% paraformaldehyde-PBS for fixation prior to intracellular staining. Intracellular staining was performed on infected, fixed resting primary CD4⁺ T cell populations with 1:100 dilution of p24-FITC antibody (KC57, Beckman Coulter) using FOXP3 Fix/Perm buffer set (Biolegend) as per manufacturer's instructions. After intracellular p24 staining, the cells were resuspended in 1% paraformaldehyde PBS for flow cytometric analysis on an Attune NxT Acoustic Focusing Cytometer (ThermoFisher), recording all events in a 180uL sample volume after one 200uL mixing cycle. Data were exported as FCS3.0 files and analyzed in FlowJo. Cells were gated on lymphocytes, side-scatter singlets, forward-scatter singlets, live cells and then the percentage of p24+ cells was quantified. The data was exported to CSV, including the count and percentage of live cells and the count and percentage of p24+ cells. Three technical replicates for each of the three biological replicates were averaged and standard deviation was calculated. The fold change and p value were calculated by comparison with the NTC of the same donor.

Measurement of the effects of gene deletions on virus production from latently infected resting CD4⁺ T cells by supernatant p24 ELISA

Latency reversal and HIV-1 production was measured by performing P24-ELISA with the culture supernatants using Lenti-X p24 Rapid Titer kit (Takara). Briefly, the culture supernatant from the edited and control samples were diluted 1:40 fold in the complete DMEM medium used for ELISA as per manufacturer's instructions. The OD450 value for the blank control was subtracted from the value of each sample. Three technical replicates for each of the three biological replicates were averaged and standard deviation was calculated. The fold change in supernatant p24 (proxy for HIV production) and p value were calculated by comparison with the NTC of the same donor.

TIDE analysis of edited primary CD4⁺ T cells and JLat cells

The efficiency of the gene knockout in each case was assessed by performing TIDE analysis with the cell lysates²³⁰. Briefly, 10ul of cell-suspension was added to 20ul of QuickExtract buffer (Lucigen QE09050) in a PCR plate and heated at 65°C for 6 minutes and 98°C for 2 minutes to prepare cell lysates containing genomic DNA. PCR amplification over the cut site was performed using this genomic DNA to generate a template for sanger sequencing, which was then analyzed for indel percentage using the TIDE webtool. Primers were designed with the Primer3 tool in Benchling. PCR amplification was performed using NEBNext Ultra II Q5 2X Master Mix (New England Biolabs M0544L), 10µM primer pair, and approximately 100ng template DNA. PCR amplicons were subsequently sent for cleanup and Sanger sequencing. Mutational efficiency was determined by comparison of non-targeting and gene-targeting sample chromatograms using the TIDE Web Tool²⁷³. The TIDE output calculates the percentage of insertions and deletions (indels) from these chromatograms. The total efficiency of indel generation provides a reasonable estimate of knockout percentage in a cell population following crRNP treatment.

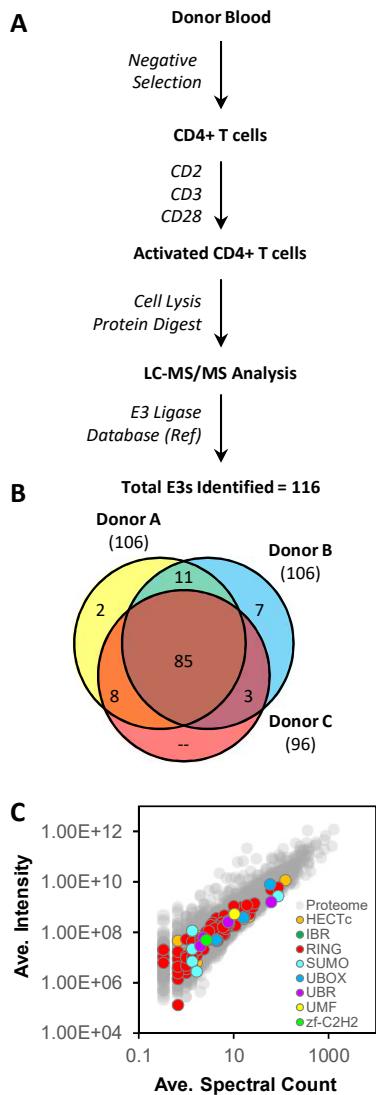


Figure 3.1. Liquid chromatography tandem mass spectrometry (LC-MS/MS) analysis identifies 116 E3 ligases expressed in activated primary human CD4⁺ T cells. **A.** Workflow describing the proteomic identification of expressed E3s by LC-MS/MS in primary activated CD4⁺ T cells from three healthy human donors. **B.** Venn Diagram showing the overlap of the 116 E3s identified as expressed by LC-MS/MS in primary activated C4⁺ T cells from three healthy human donors. **C.** Scatter plot showing protein E3 type and their abundance (by average (ave.) spectral count and average intensity for each protein identified in all three donors (for proteins not identified in all three donors, the missing value was ignored in the average calculation). E3 ligases are represented by colored markers according to their type (see legend inset), and non-E3 ligase proteins are represented by grey markers) relative to all other genes detected by LC-MS/MS analysis in primary activated CD4⁺ T cells from three healthy human donors.

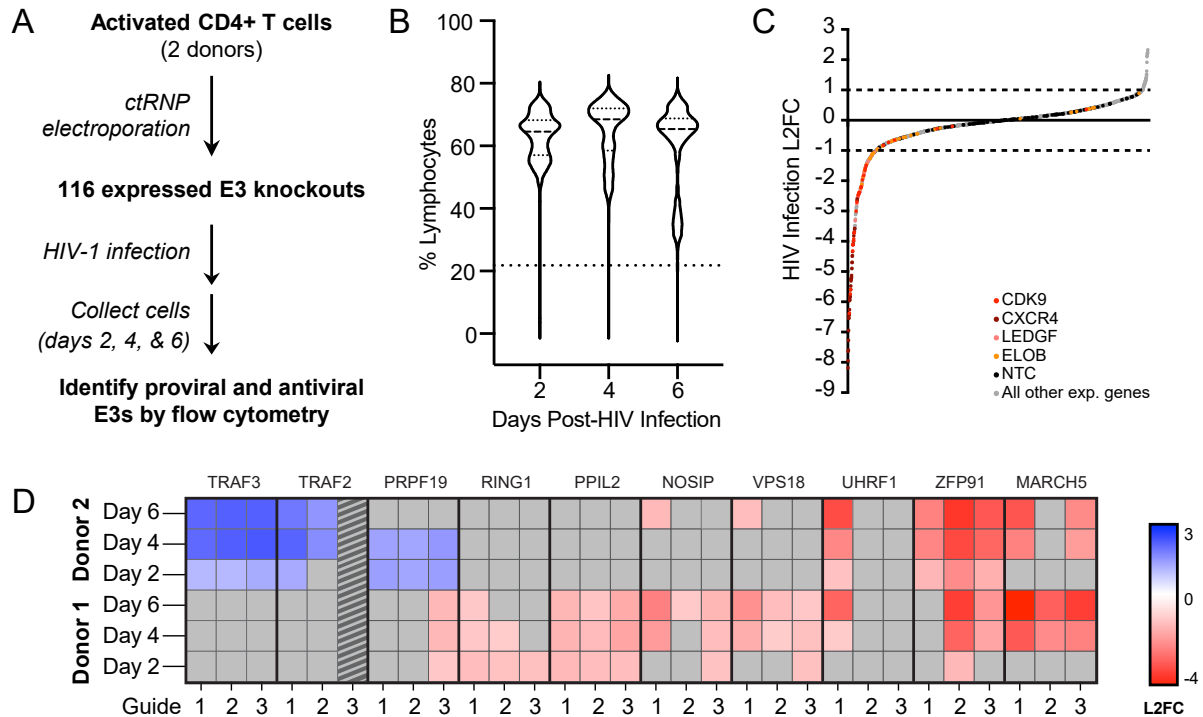


Figure 3.2. CRISPR-Cas9 knockouts in primary CD4⁺ T cells identify novel E3s that regulate HIV infection. **A.** Workflow showing CRISPR knockout and screening of 116 E3 knockouts for a functional role in HIV-1 infection in primary activated CD4⁺ T cells from healthy human donors. **B.** Flow cytometry quantification of the % lymphocytes across all donors, guides, and technical replicates of the E3 knockouts at day 2, 4 and 6 post HIV infection. A viability estimate filter of 21.8% lymphocytes (Figure S3.2) is marked by a dashed line. **C.** Flow cytometry quantification of HIV infection (% GFP positive cells normalized to NTCs and log₂ transformed) for all gene knockouts, in all donors, at 2, 4, and 6 days after HIV-1 infection in primary activated CD4⁺ T cells from all healthy human donors. Known dependency factor controls are marked in color, NTCs are marked in black, all E3s are grey. **D.** Heatmap showing HIV infection for 10 E3 hits. Hits were defined as a gene knockout that yielded a HIV infection log₂ fold change (L2FC) compared to NTC of ≥ 1 or ≤ -1 in the same direction in at least two timepoints for two guides within the same donor, or the same guide in two donors. Data points with a L2FC < 1 or > -1 are colored grey. Data from TRAF2 guide 3 was removed as it fell below our viability estimate filter. “Donor 1” and “donor 2” do not correspond to the same human donors throughout the screen but are arbitrarily listed in random order, as different donors were used to complete the screen in multiple rounds.

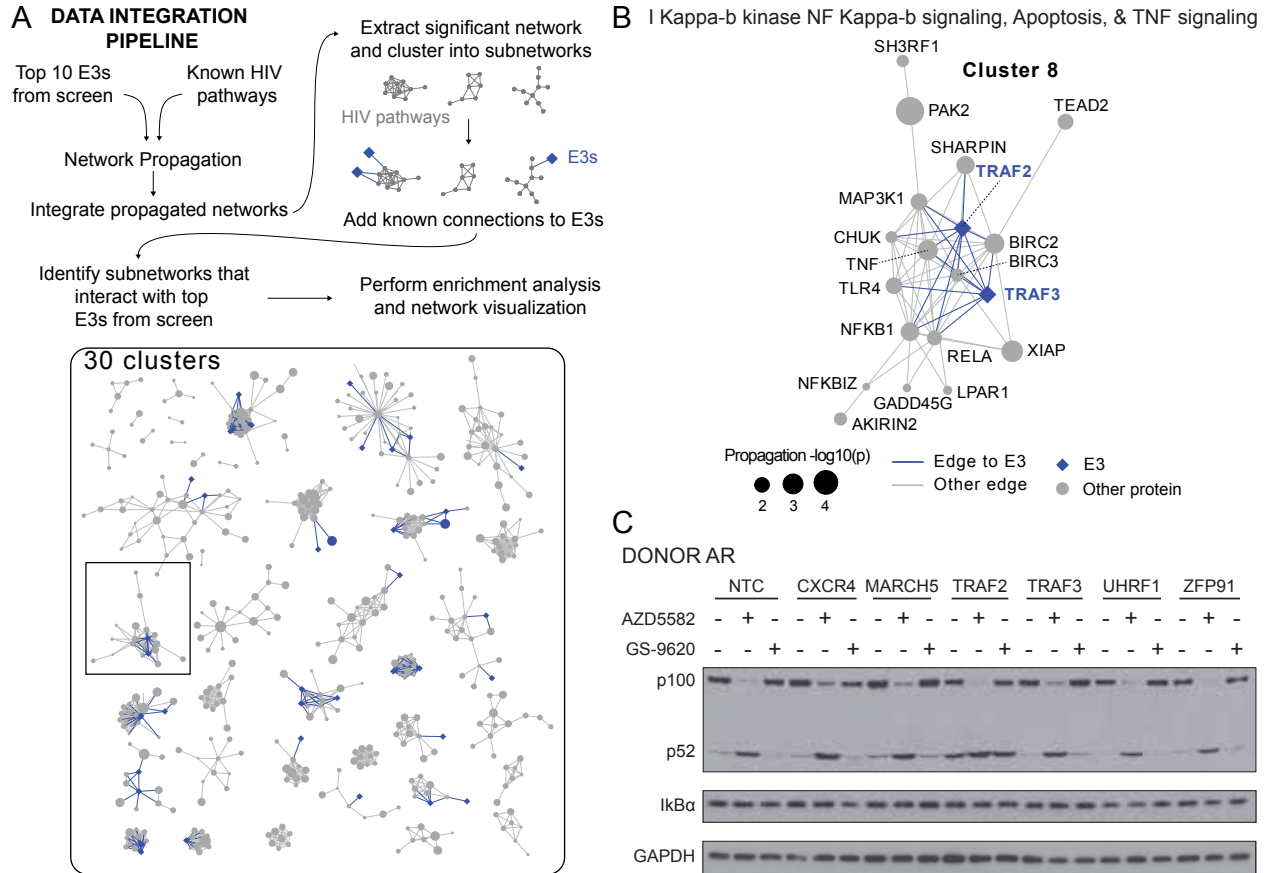


Figure 3.3. Network propagation analysis connects TRAF2, non-canonical NF- κ B signaling, and HIV infection. **A.** Data integration pipeline overview. The top E3 hits from this study and human proteins with prior association with HIV were integrated using network propagation. The significant subnetwork from the propagation was further clustered into 30 subnetwork pieces (or “clusters”) and known functional and physical connections between each cluster and the top E3s from this study were added. This analysis serves to highlight the pathways and complexes that are associated with both HIV pathogenesis and the E3s from this study. **B.** One of the clusters extracted from the network propagation results (C8) that possess significant enrichment in NF- κ B related pathway terms. E3s from this study are annotated in blue diamonds. Grey circles denote human proteins within the base network. Size of the circle denotes the $-\log_{10}(p)$ -value of the propagation result. **C.** Western blot analysis probing for non-canonical NF- κ B activation (marked by processing of p100 to p52), and non-canonical NF- κ B activation (marked by I κ B α degradation), with a GAPDH loading control in gene knockouts in primary activated CD4⁺ T cells from healthy human donor AR, which is one representative donor of three (Figure S3.5). NTC is a non-targeting control.

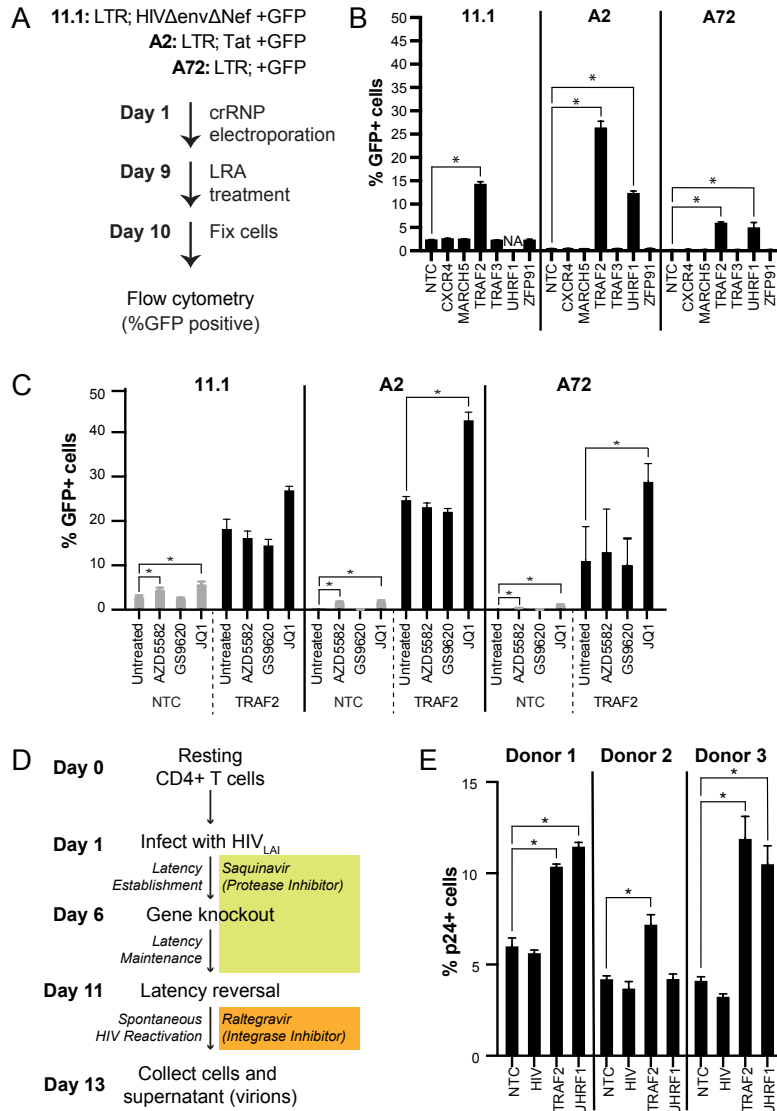


Figure 3.4. TRAF2 and UHRF1 knockouts reverse latency in JLat and resting primary human CD4⁺ T cell models. **A.** Workflow describing gene knockout and treatment with latency-reversing agents (LRAs) in JLat cell lines 11.1, A2, and A72. **B.** Flow cytometry quantification of latency reversal (% GFP+ cells) for gene knockouts that were untreated (no LRAs) in JLat cell lines 11.1, A2, and A72. One sample that did not pass the viability cutoff of 30% lymphocytes was removed and denoted “NA”. Significance was defined as a fold change ≥ 1.5 and a p value < 0.05 compared to the untreated non-targeting control (NTC) within the same cell line. **C.** Flow cytometry quantification of latency reversal (% GFP+ cells) for NTCs and TRAF2 knockouts that were untreated or treated with the LRAs AZD5582, GS-9620, or JQ1 in JLat cell lines 11.1, A2, and A72. Significance was defined as a fold change ≥ 1.5 and a p value < 0.05 compared to the untreated condition of the same gene knockout within the same cell line. **D.** Workflow describing gene knockout and spontaneous reactivation in a novel primary resting CD4⁺ T cell latency model. **E.** Flow cytometry quantification of latency reversal (% p24+ cells) for gene knockouts in primary resting CD4⁺ T cells. HIV knockout is a negative control that consists of three guides, two targeting the HIV trans-activation response element (TAR) and one targeting the HIV long-terminal repeat (LTR). Significance was defined as a fold change ≥ 1.5 and a p value < 0.05 compared to NTC within the same donor.

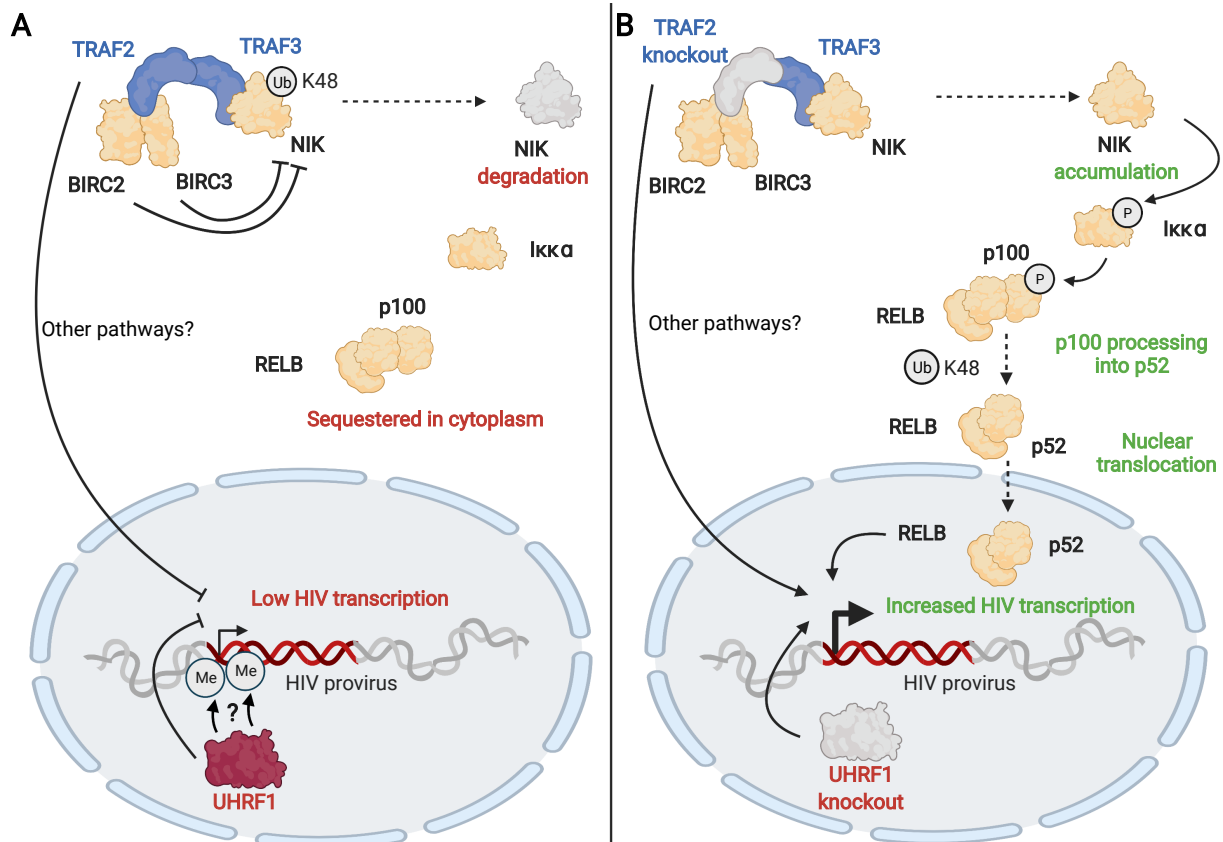


Figure 3.5. Proposed mechanism by which TRAF2 and UHRF1 transcriptionally reactivate the latent HIV-1 promoter. A. Low HIV-1 transcription in the presence of TRAF2 and UHRF1. **B.** Increased HIV-1 transcription upon TRAF2 knockout and UHRF1 knockout. Image created with BioRender.

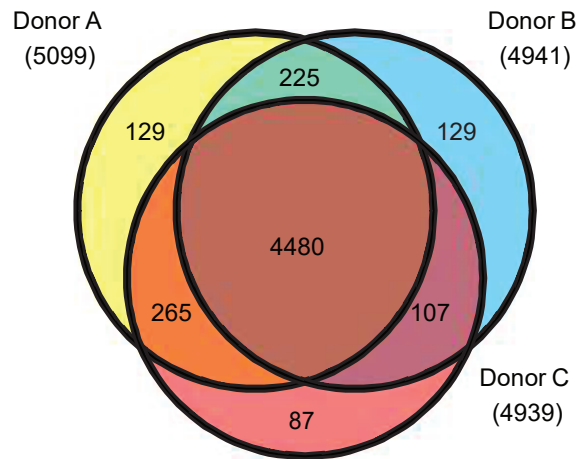


Figure S3.1, related to Figure 3.1. Overlap of 5429 proteins identified as expressed in primary activated CD4⁺ T cells from three healthy human donors.

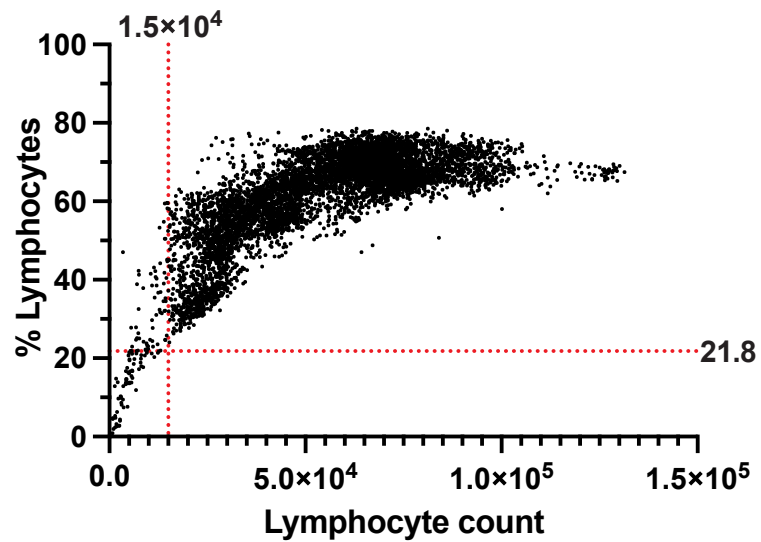


Figure S3.2, related to Figure 3.2. Scatter plot showing the flow cytometry analysis of the % lymphocytes by the lymphocyte count of all knockout events from all donors, guides, technical replicates, and all three days of cell collection (2, 4, and 6 days after HIV-1 infection). Using this plot to estimate viability, we filtered out the cells that fell below 15000 lymphocyte count and that were less than 21.8% lymphocytes. In total, this filtered out 1% of the total events.

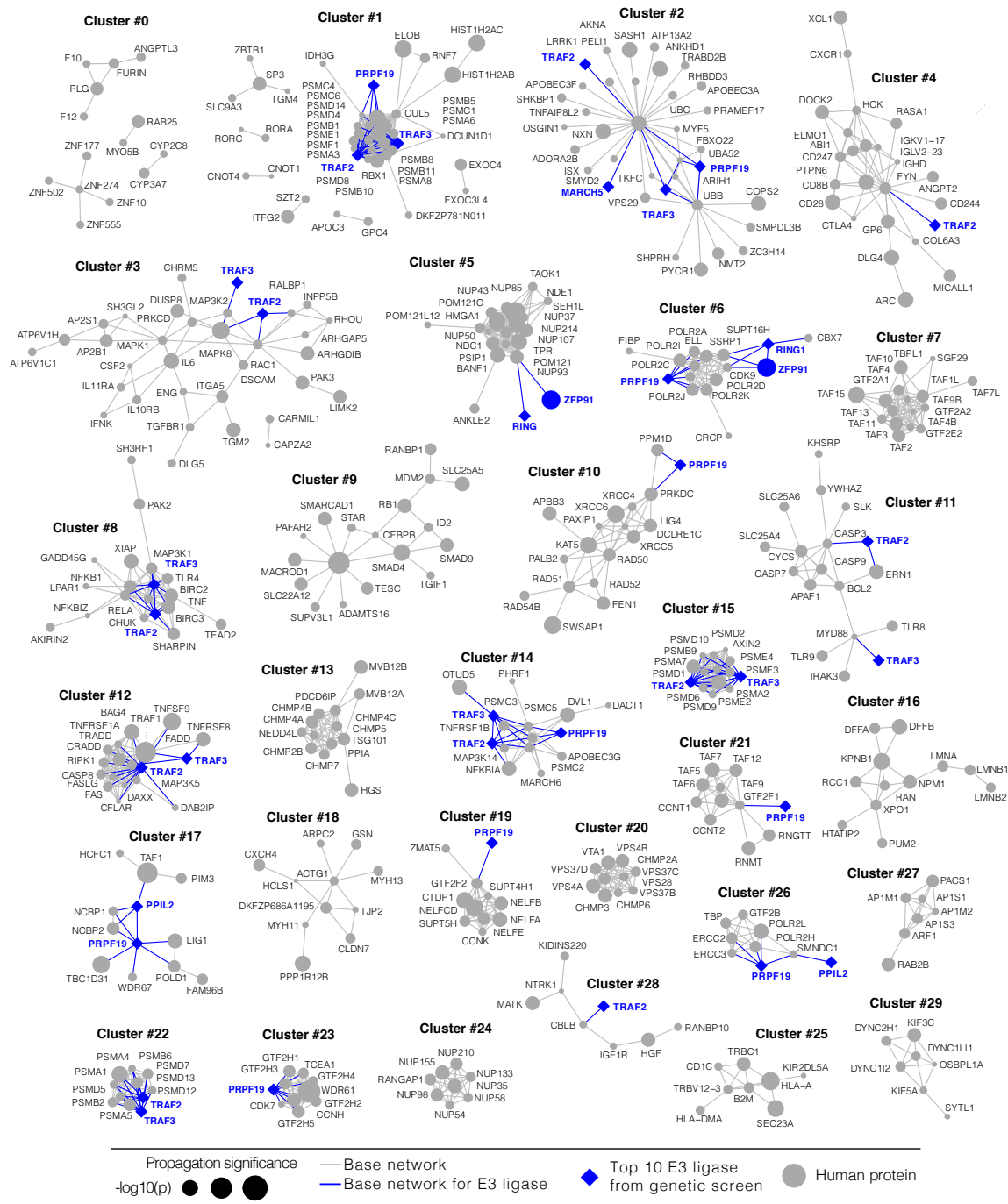


Figure S3.3, related to Figure 3.3. Final clustered subnetworks from network propagation analysis. Integrative network propagation analysis was performed by merging networks propagated with either genes associated with HIV pathogenesis or the top E3s identified in this study (see Methods). The significant subnetwork was created by extracting genes (grey circles) with p -value ≤ 0.05 from the base network, requiring they form a singular connected component (i.e. possessed at least one connection to another gene within this set; $n=447$). Circle size corresponds to the $-\log_{10}(p)$ -value from the propagation analysis. The resulting significant subnetwork was clustered into 30 smaller subnetwork clusters and E3s with known connections to each cluster were overlaid onto each cluster (blue diamonds and lines).

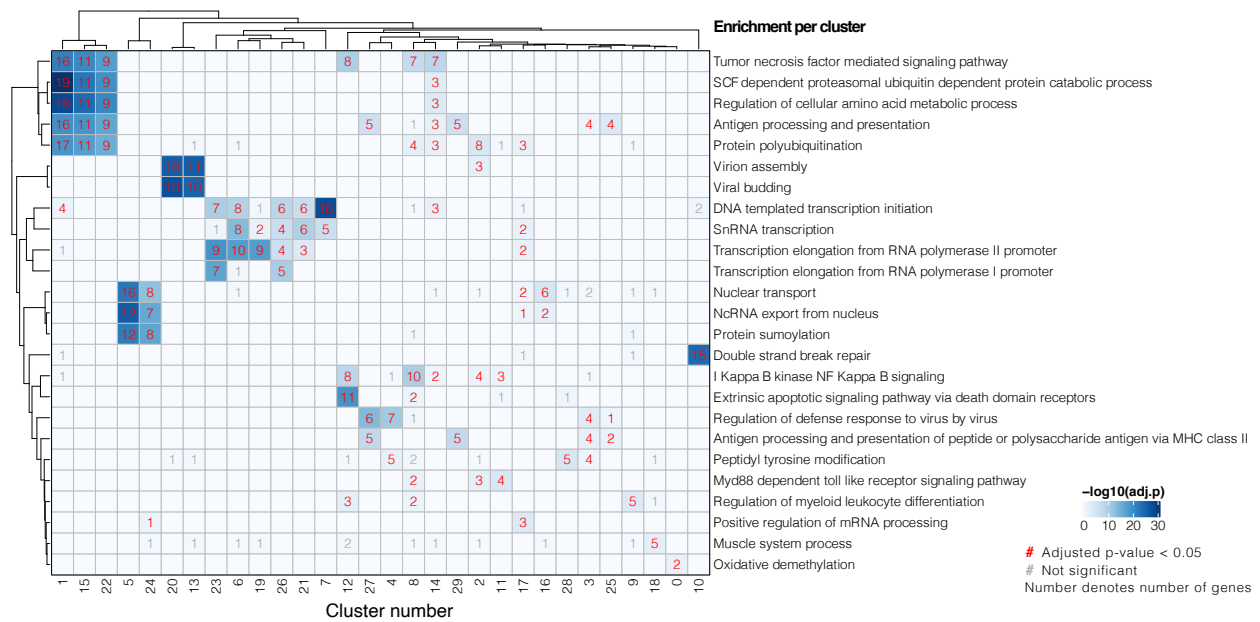
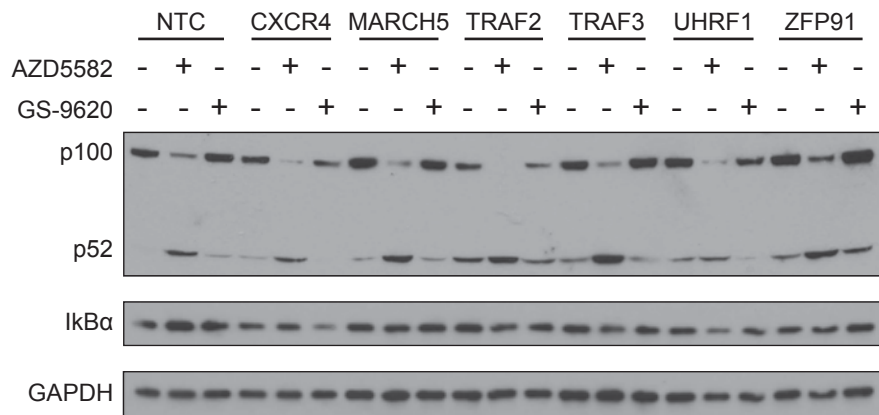


Figure S3.4, related to Figure 3.3. Enrichment for subnetwork clusters (annotated along the columns) identified in Figure S3.3. Gene Ontology (GO) enrichment analysis (biological process) was performed for each of the 30 resulting subnetwork clusters to identify biological processes associated with each cluster. Color corresponds to the $-\log_{10}(\text{adj.p})$ from the GO Biological Process enrichment analysis. Numbers denote the number of genes enriched in each cluster for each term; if significant (adjusted p-value < 0.05) they are colored red, otherwise grey.

A

DONOR AQ

**B**

DONOR AS

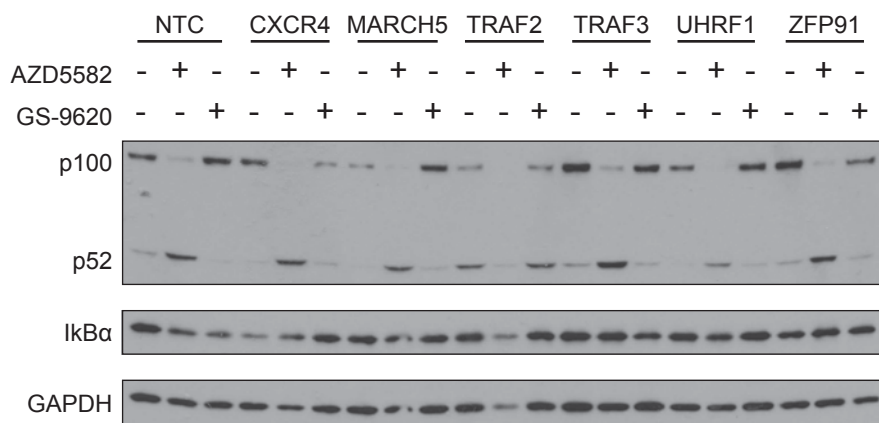


Figure S3.5, related to Figure 3.3. A. Western blot analysis probing for non-canonical NF- κ B activation (marked by processing of p100 to p52), and non-canonical NF- κ B activation (marked by I κ B α degradation), with a GAPDH loading control in gene knockouts in primary activated CD4⁺ T cells from healthy human donor AQ. **B.** Same as in (A) but from healthy human donor AS.

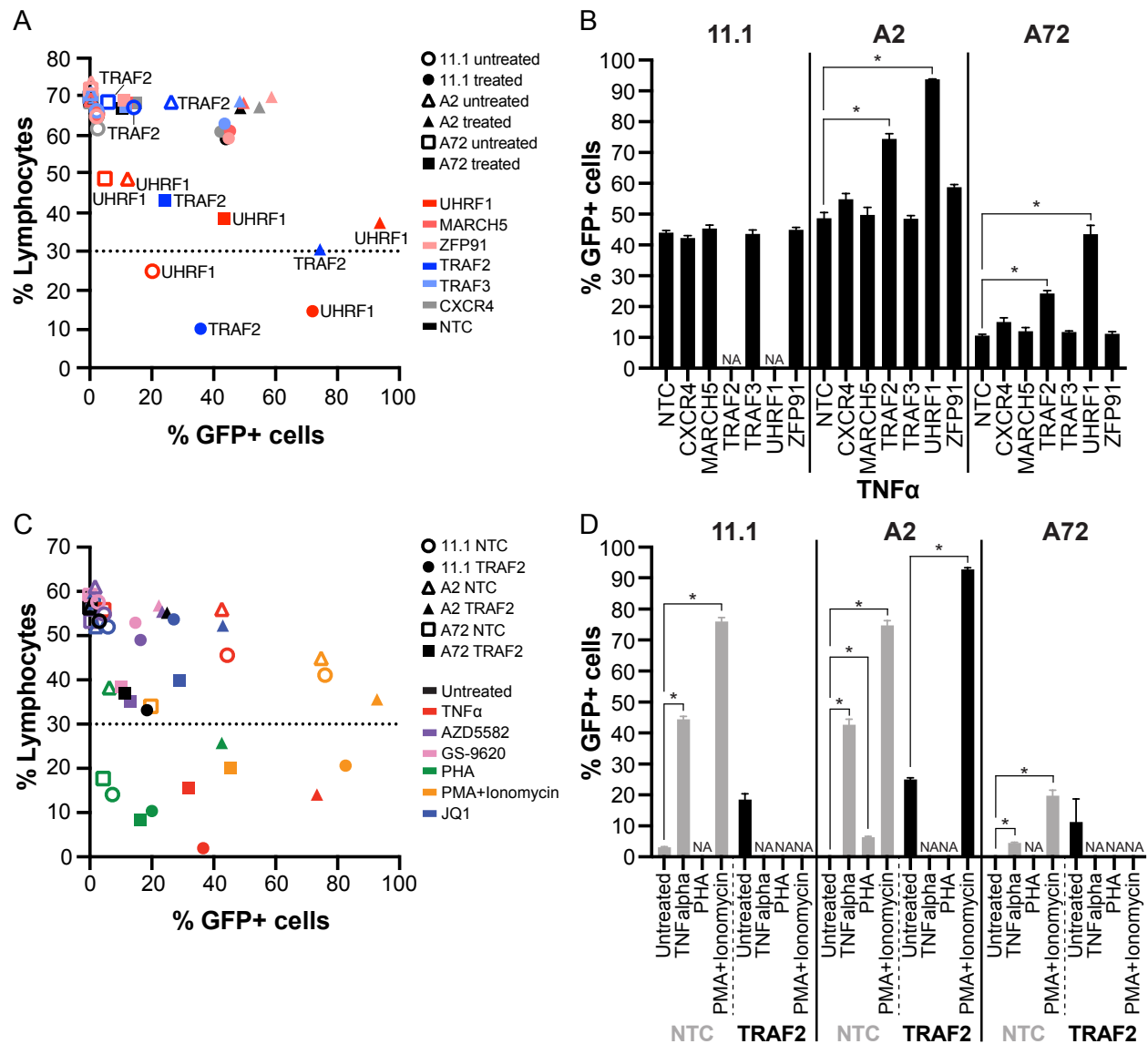


Figure S3.6, related to Figure 3.4. A. Flow cytometry quantification of viability (% lymphocytes) relative to HIV latency reversal (%GFP+ cells) for gene knockouts treated with TNF α or untreated in JLat cell lines 11.1, A2, and A72. NTC is a non-targeting control. A viability cutoff was set at 30% lymphocytes. **B.** Flow cytometry quantification of latency reversal (%GFP+ cells) for gene knockouts treated with TNF α in JLat cell lines 11.1, A2, and A72. Samples that did not pass the viability cutoff and were removed and denoted “NA”. Significance was defined as a fold change ≥ 1.5 and a p value < 0.05 compared to the TNF α -treated NTC within the same cell line. **C.** Flow cytometry quantification of viability (% lymphocytes) relative to HIV latency reversal (% GFP+ cells) for NTC or TRAF2 knockouts treated with a panel of latency-reversing agents (LRAs) in JLat cell lines 11.1, A2, and A72. A viability cutoff was set at 30% lymphocytes. **D.** Flow cytometry quantification of latency reversal (% GFP+ cells) for NTCs and TRAF2 knockouts that were untreated or treated with the LRAs TNF α , PHA, or PMA+Ionomycin in JLat cell lines 11.1, A2, and A72. Samples that did not pass the viability cutoff and were removed and denoted “NA”. Significance was defined as a fold change ≥ 1.5 and a p value < 0.05 compared to the untreated condition of the same gene knockout within the same cell line.

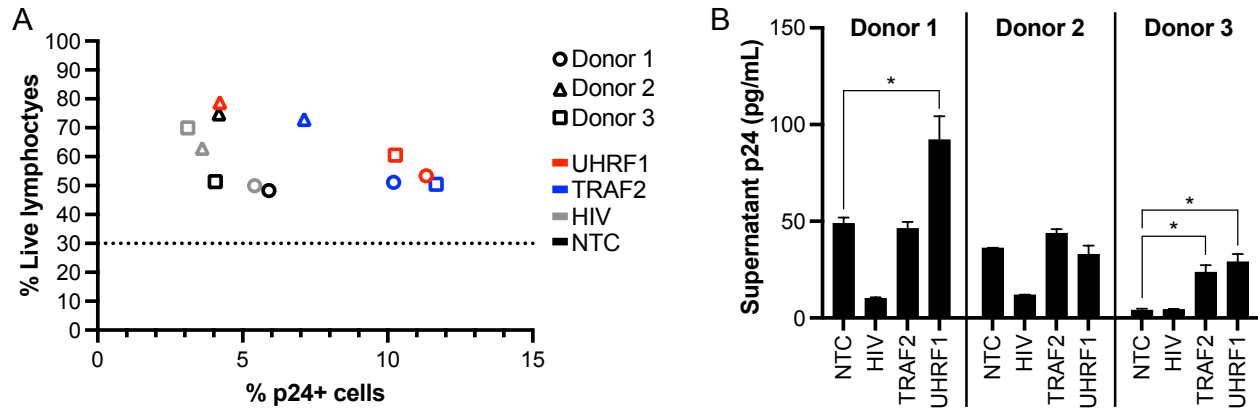


Figure S3.7, related to Figure 3.4. A. Flow cytometry quantification of viability (% live lymphocytes) relative to HIV latency reversal (%p24+ cells) for gene knockouts in primary resting CD4⁺ T cells from three healthy human donors. NTC is a non-targeting control and HIV is a negative control that consists of three guides, two targeting the HIV trans-activation response element (TAR) and one targeting the HIV long-terminal repeat (LTR). A viability cutoff was set at 30% live lymphocytes. **B.** ELISA quantification of latency reversal (supernatant p24 (pg/ml)) for gene knockouts in primary resting CD4⁺ T cells from three healthy human donors. Significance was defined as a fold change ≥ 1.5 and a p value < 0.05 compared to NTC within the same donor.

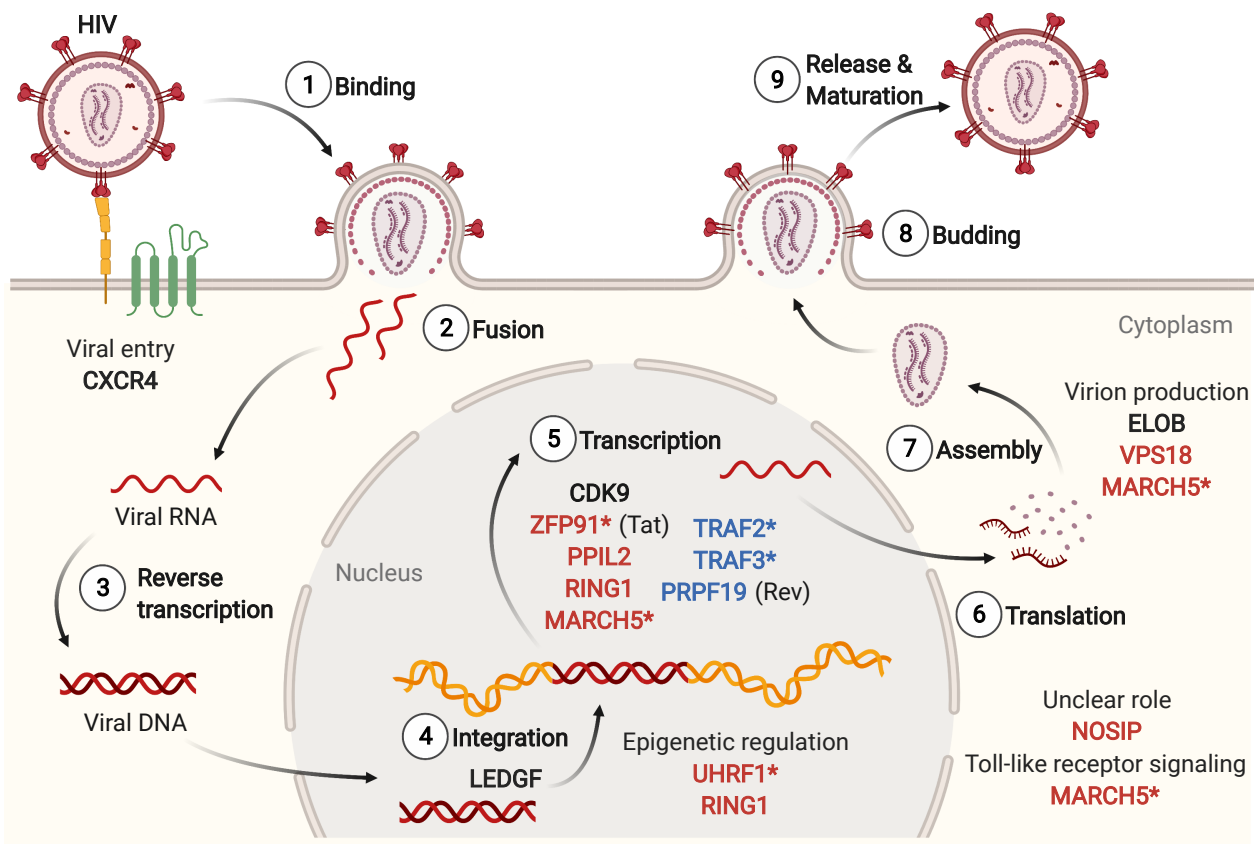


Figure S3.8, related to Discussion. Potential roles for 10 E3 hits in HIV infection. Image created with BioRender.

SUPPLEMENTARY TABLES

Table S3.1, related to Figure 3.1. Proteins identified by LC-MS/MS analysis of primary activated CD4⁺ T cells.

Table S3.2, related to Figure 3.2. CRISPR guides used in the CRISPR-Cas9 spreading HIV-1 infection assay.

Table S3.3, related to Figure 3.2. Raw and analyzed data from CRISPR-Cas9 knockout spreading HIV-1 infection assay.

Table S3.4, related to Figure 3.3. Network propagation of HIV pathways and E3 ubiquitin ligases that functionally affected HIV infection in our CRISPR-Cas9 knockout spreading HIV-1 infection assay.

Table S3.5, related to Figure 3.4. Raw and analyzed data from CRISPR-Cas9 knockouts in JLat models of HIV latency.

Table S3.6, related to Figure 3.4. Raw and analyzed data from CRISPR-Cas9 knockouts in a resting primary human CD4⁺ T cell model of HIV latency.

CHAPTER 4

Discussion

In this dissertation, I have leveraged the power of global, unbiased systems biology approaches to study host-pathogen interactions critical for combating two major pathogens of the past 50 years: SARS-CoV-2 and HIV. While my perspective on SARS-CoV-2 includes proposed studies (and not completed experiments), my coauthors and I outlined key questions and proteomic technologies to address them. Proteins are the actionable components of the cell, and proteomics provides a global picture of how proteins interact to achieve health and disease. In our perspective, we not only proposed proteomic technologies that could be leveraged to study each part of the viral life cycle, we described how different datasets could be integrated to provide a more complete picture of infection. We wrote and published the perspective within the first year of the COVID-19 pandemic, which required us to keep on top of rapidly evolving information about the novel SARS-CoV-2 virus, synthesize complementary information, and prioritize the most impactful areas of study. Finally, we called upon the broader proteomic research community to do our part to contribute to the study of the novel SARS-CoV-2 virus and propose urgently-needed treatments for COVID-19. The research strategies proposed in the perspective, especially the integrative synthesis of different types of data, provide a blueprint for the proteomic study of any virus.

In my manuscript, I leveraged proteomics and genetic screening in primary CD4⁺ T cells to study HIV infection and latency reversal. I identified E3 ubiquitin ligases as important signaling molecules in HIV infection, and was motivated to study them further using primary CD4⁺ T cells. While other studies of HIV host factors had been performed in cell lines (Zhou et al. 2008; Yeung et al. 2009; König et al. 2008; Brass et al. 2008), the use of primary cells in this study is critical because ubiquitin regulates cell cycle and apoptosis, processes that are perturbed in immortalized cell lines. Because work in primary cells is cumbersome, it was important to prioritize which E3s to test for a functional impact on HIV infection. Instead of relying on previously published datasets from cell lines or observing expression at the RNA level, we created a novel proteomic dataset identifying E3s expressed at the protein level in

primary CD4⁺ T cells. When screening expressed E3s for a functional role in HIV infection, we chose an arrayed screen in primary CD4⁺ T cells, the physiologically-relevant cell type. The arrayed screen format allowed us to control for viability, as E3s often regulate essential processes.

We continued the use of primary cells when testing for HIV latency reversal with the development of a novel latency reversal assay in primary resting CD4⁺ T cells. Because resting CD4⁺ T cells do not divide, we had to be even more stringent about the factors tested. To narrow our candidates, we leveraged Jurkat (JLat) cell line models of latency, showing the utility of cell lines to allow for higher-throughput screening, followed by validation of top candidates in primary cells. The identification of TRAF2 and UHRF1 as novel factors in HIV latency reversal provides novel candidate drug targets for shock and kill therapy towards an HIV cure.

Together, the proposed systems biology approaches to study SARS-CoV-2 in my perspective and the systems-level genetic and proteomic work studying HIV in my manuscript demonstrate the power of global, unbiased approaches in the identification of novel therapeutic candidates for two of the most devastating viruses of our current time.

REFERENCES

1. Global Statistics. *HIV.gov* <https://www.hiv.gov/hiv-basics/overview/data-and-trends/global-statistics> (2021).
2. Global HIV & AIDS statistics - Fact sheet. *UNAIDS* <https://www.unaids.org/en/resources/fact-sheet>.
3. HIV Treatment. *Centers for Disease Control and Prevention* <https://www.cdc.gov/hiv/basics/livingwithhiv/treatment.html> (2021).
4. HIV Populations at Greatest Risk. *Centers for Disease Control and Prevention* <https://www.cdc.gov/hiv/policies/hip/risk.html> (2020).
5. Homophobia and HIV. *Avert* <https://www.avert.org/professionals/hiv-social-issues/homophobia> (2019).
6. Standing up to stigma. <https://www.hiv.gov/hiv-basics/overview/making-a-difference/standing-up-to-stigma> (2020).
7. Coronavirus COVID-19 Dashboard. *World Health Organization* <https://covid19.who.int/>.
8. Schmidt, C. COVID-19 long haulers. *Nat. Biotechnol.* **39**, 908–913 (2021).
9. COVID-19 State of Vaccine Confidence Insights Report. *Centers for Disease Control and Prevention* <https://www.cdc.gov/vaccines/covid-19/downloads/SoVC-report-12.pdf> (2021).
10. COVID-19 Health Equity Considerations and Racial and Ethnic Minority Groups. *Centers for Disease Control and Prevention* <https://www.cdc.gov/coronavirus/2019-ncov/community/health-equity/race-ethnicity.html> (2021).
11. Reja, M. Trump's 'Chinese Virus' tweet helped lead to rise in racist anti-Asian Twitter content: Study. *ABC News* <https://abcnews.go.com/Health/trumps-chinese-virus-tweet-helped-lead-rise-racist/story?id=76530148> (2021).

12. Yam, K. Anti-Asian hate crimes increased by nearly 150% in 2020, mostly in N.Y. and L.A., new report says. *NBC News* <https://www.nbcnews.com/news/asian-america/anti-asian-hate-crimes-increased-nearly-150-2020-mostly-n-n1260264> (2021).
13. Ramdas, P., Sahu, A. K., Mishra, T., Bhardwaj, V. & Chande, A. From Entry to Egress: Strategic Exploitation of the Cellular Processes by HIV-1. *Front. Microbiol.* **11**, 559792 (2020).
14. Suzuki, Y. & Craigie, R. The road to chromatin - nuclear entry of retroviruses. *Nat. Rev. Microbiol.* **5**, 187–196 (2007).
15. Lusic, M. & Siliciano, R. F. Nuclear landscape of HIV-1 infection and integration. *Nat. Rev. Microbiol.* **15**, 69–82 (2017).
16. Chun, T. W. *et al.* Quantification of latent tissue reservoirs and total body viral load in HIV-1 infection. *Nature* **387**, 183–188 (1997).
17. Finzi, D. *et al.* Latent infection of CD4⁺ T cells provides a mechanism for lifelong persistence of HIV-1, even in patients on effective combination therapy. *Nat. Med.* **5**, 512–517 (1999).
18. Chun, T. W. *et al.* Presence of an inducible HIV-1 latent reservoir during highly active antiretroviral therapy. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 13193–13197 (1997).
19. Ho, Y.-C. *et al.* Replication-competent noninduced proviruses in the latent reservoir increase barrier to HIV-1 cure. *Cell* **155**, 540–551 (2013).
20. Wu, F. *et al.* A new coronavirus associated with human respiratory disease in China. *Nature* **579**, 265–269 (2020).
21. Chan, J. F.-W. *et al.* Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerg. Microbes Infect.* **9**, 221–236 (2020).
22. Fung, T. S. & Liu, D. X. Human Coronavirus: Host-Pathogen Interaction. *Annu. Rev. Microbiol.* **73**, 529–557 (2019).

23. Hoffmann, M. *et al.* SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* **181**, 271–280.e8 (2020).
24. V'kovski, P., Kratzel, A., Steiner, S., Stalder, H. & Thiel, V. Coronavirus biology and replication: implications for SARS-CoV-2. *Nat. Rev. Microbiol.* (2020) doi:10.1038/s41579-020-00468-6.
25. Sola, I., Almazán, F., Zúñiga, S. & Enjuanes, L. Continuous and Discontinuous RNA Synthesis in Coronaviruses. *Annu Rev Virol* **2**, 265–288 (2015).
26. Cohn, L. B. *et al.* HIV-1 integration landscape during latent and active infection. *Cell* **160**, 420–432 (2015).
27. Court, R. *et al.* Random lopinavir concentrations predict resistance on lopinavir-based antiretroviral therapy. *Int. J. Antimicrob. Agents* **48**, 158–162 (2016).
28. Cohen, K. *et al.* A Clinical Prediction Rule for Protease Inhibitor Resistance in Patients Failing Second-Line Antiretroviral Therapy. *J. Acquir. Immune Defic. Syndr.* **80**, 325–329 (2019).
29. van Zyl, G. U. *et al.* Low lopinavir plasma or hair concentrations explain second-line protease inhibitor failures in a resource-limited setting. *J. Acquir. Immune Defic. Syndr.* **56**, 333–339 (2011).
30. Su, S. *et al.* Epidemiology, Genetic Recombination, and Pathogenesis of Coronaviruses. *Trends Microbiol.* **24**, 490–502 (2016).
31. Human Coronavirus Types. *Centers for Disease Control and Prevention* <https://www.cdc.gov/coronavirus/types.html> (2020).
32. Common Human Coronaviruses. *Centers for Disease Control and Prevention* <https://www.cdc.gov/coronavirus/general-information.html>.
33. van der Hoek, L. Human coronaviruses: what do they cause? *Antivir. Ther.* **12**, 651–658 (2007).

34. Drosten, C. *et al.* Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N. Engl. J. Med.* **348**, 1967–1976 (2003).
35. Zaki, A. M., van Boheemen, S., Bestebroer, T. M., Osterhaus, A. D. M. E. & Fouchier, R. A. M. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N. Engl. J. Med.* **367**, 1814–1820 (2012).
36. de Wit, E., van Doremalen, N., Falzarano, D. & Munster, V. J. SARS and MERS: recent insights into emerging coronaviruses. *Nat. Rev. Microbiol.* **14**, 523–534 (2016).
37. Petersen, E. *et al.* Comparing SARS-CoV-2 with SARS-CoV and influenza pandemics. *Lancet Infect. Dis.* (2020) doi:10.1016/S1473-3099(20)30484-9.
38. SARS Basics Fact Sheet. *Centers for Disease Control and Prevention*
<https://www.cdc.gov/sars/about/fs-sars.html>.
39. Middle East Respiratory Syndrome Coronavirus. *World Health Organization*
<https://www.who.int/emergencies/mers-cov/en/>.
40. Coronavirus Disease (COVID-19) Dashboard. *World Health Organization*
<https://covid19.who.int/>.
41. COVID-19 Vaccines. *World Health Organization*
<https://www.who.int/emergencies/diseases/novel-coronavirus-2019/covid-19-vaccines>.
42. FDA Takes Key Action in Fight Against COVID-19 By Issuing Emergency Use Authorization for First COVID-19 Vaccine. *U.S. Food and Drug Administration*
<https://www.fda.gov/news-events/press-announcements/fda-takes-key-action-fight-against-covid-19-issuing-emergency-use-authorization-first-covid-19> (2020).
43. Moderna COVID-19 Vaccine. *US Food & Drug Administration*
<https://www.fda.gov/emergency-preparedness-and-response/coronavirus-disease-2019-covid-19/moderna-covid-19-vaccine>.

44. Ashraf, B. N. Economic impact of government interventions during the COVID-19 pandemic: International evidence from financial markets. *J Behav Exp Finance* **27**, 100371 (2020).
45. Grein, J. *et al.* Compassionate Use of Remdesivir for Patients with Severe Covid-19. *N. Engl. J. Med.* **382**, 2327–2336 (2020).
46. WHO recommends against the use of remdesivir in COVID-19 patients. *World Health Organization* <https://www.who.int/news-room/feature-stories/detail/who-recommends-against-the-use-of-remdesivir-in-covid-19-patients> (2020).
47. Kumar, S., Nyodu, R., Maurya, V. K. & Saxena, S. K. Morphology, Genome Organization, Replication, and Pathogenesis of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). in *Coronavirus Disease 2019 (COVID-19): Epidemiology, Pathogenesis, Diagnosis, and Therapeutics* (ed. Saxena, S. K.) 23–31 (Springer Singapore, 2020).
48. Lu, R. *et al.* Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* **395**, 565–574 (2020).
49. Wan, Y., Shang, J., Graham, R., Baric, R. S. & Li, F. Receptor Recognition by the Novel Coronavirus from Wuhan: an Analysis Based on Decade-Long Structural Studies of SARS Coronavirus. *J. Virol.* **94**, (2020).
50. Dülfer, J., Kadek, A., Kopicki, J.-D., Krichel, B. & Uetrecht, C. Structural mass spectrometry goes viral. *Adv. Virus Res.* **105**, 189–238 (2019).
51. Iacobucci, C., Götze, M. & Sinz, A. Cross-linking/mass spectrometry to get a closer view on protein interaction networks. *Curr. Opin. Biotechnol.* **63**, 48–53 (2020).
52. Piotrowski, C. & Sinz, A. Structural Investigation of Proteins and Protein Complexes by Chemical Cross-Linking/Mass Spectrometry. *Adv. Exp. Med. Biol.* **1105**, 101–121 (2018).
53. Chavez, J. D. & Bruce, J. E. Chemical cross-linking with mass spectrometry: a tool for systems structural biology. *Curr. Opin. Chem. Biol.* **48**, 8–18 (2019).

54. O'Reilly, F. J. & Rappsilber, J. Cross-linking mass spectrometry: methods and applications in structural, molecular and systems biology. *Nat. Struct. Mol. Biol.* **25**, 1000–1008 (2018).
55. Rappsilber, J. The beginning of a beautiful friendship: cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes. *J. Struct. Biol.* **173**, 530–540 (2011).
56. Yilmaz, Ş. *et al.* Cross-linked peptide identification: A computational forest of algorithms. *Mass Spectrom. Rev.* **37**, 738–749 (2018).
57. Liu, F. & Heck, A. J. R. Interrogating the architecture of protein assemblies and protein interaction networks by cross-linking mass spectrometry. *Curr. Opin. Struct. Biol.* **35**, 100–108 (2015).
58. Lin, D. H. & Hoelz, A. The Structure of the Nuclear Pore Complex (An Update). *Annu. Rev. Biochem.* **88**, 725–783 (2019).
59. Fernandez-Martinez, J. *et al.* Structure and Function of the Nuclear Pore Complex Cytoplasmic mRNA Export Platform. *Cell* **167**, 1215–1228.e25 (2016).
60. Tüting, C., Iacobucci, C., Ihling, C. H., Kastritis, P. L. & Sinz, A. Structural analysis of 70S ribosomes by cross-linking/mass spectrometry reveals conformational plasticity. *Sci. Rep.* **10**, 12618 (2020).
61. Klykov, O., van der Zwaan, C., Heck, A. J. R., Meijer, A. B. & Scheltema, R. A. Missing regions within the molecular architecture of human fibrin clots structurally resolved by XL-MS and integrative structural modeling. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 1976–1987 (2020).
62. Kwon, Y. *et al.* Structural basis of CD4 downregulation by HIV-1 Nef. *Nat. Struct. Mol. Biol.* (2020) doi:10.1038/s41594-020-0463-z.
63. Meyer, N. L. *et al.* Structure of the gene therapy vector, adeno-associated virus with its cell receptor, AAVR. *Elife* **8**, (2019).

64. Prchal, J., Sýs, J., Junková, P., Lipov, J. & Ruml, T. The interaction interface of Mason-Pfizer monkey virus matrix and envelope proteins. *J. Virol.* (2020) doi:10.1128/JVI.01146-20.
65. Yu, Y. T.-C. *et al.* Surface vimentin is critical for the cell entry of SARS-CoV. *J. Biomed. Sci.* **23**, 14 (2016).
66. Yao, H. *et al.* Molecular Architecture of the SARS-CoV-2 Virus. *Cell* **183**, 730–738.e13 (2020).
67. Liu, C. *et al.* The Architecture of Inactivated SARS-CoV-2 with Postfusion Spikes Revealed by Cryo-EM and Cryo-ET. *Structure* **28**, 1218–1224.e4 (2020).
68. Konermann, L., Pan, J. & Liu, Y.-H. Hydrogen exchange mass spectrometry for studying protein structure and dynamics. *Chem. Soc. Rev.* **40**, 1224–1234 (2011).
69. Hvidt, A. & Nielsen, S. O. Hydrogen exchange in proteins. *Adv. Protein Chem.* **21**, 287–386 (1966).
70. Englander, S. W. & Kallenbach, N. R. Hydrogen exchange and structural dynamics of proteins and nucleic acids. *Q. Rev. Biophys.* **16**, 521–655 (1983).
71. Masson, G. R. *et al.* Recommendations for performing, interpreting and reporting hydrogen deuterium exchange mass spectrometry (HDX-MS) experiments. *Nat. Methods* **16**, 595–602 (2019).
72. Harrison, R. A. & Engen, J. R. Conformational insight into multi-protein signaling assemblies by hydrogen-deuterium exchange mass spectrometry. *Curr. Opin. Struct. Biol.* **41**, 187–193 (2016).
73. Chalmers, M. J., Busby, S. A., Pascal, B. D., West, G. M. & Griffin, P. R. Differential hydrogen/deuterium exchange mass spectrometry analysis of protein-ligand interactions. *Expert Rev. Proteomics* **8**, 43–59 (2011).

74. Englander, J. J. *et al.* Protein structure change studied by hydrogen-deuterium exchange, functional labeling, and mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 7057–7062 (2003).
75. Englander, S. W. & Mayne, L. The nature of protein folding pathways. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 15873–15880 (2014).
76. Balasubramaniam, D. & Komives, E. A. Hydrogen-exchange mass spectrometry for the study of intrinsic disorder in proteins. *Biochim. Biophys. Acta* **1834**, 1202–1209 (2013).
77. Vadas, O., Jenkins, M. L., Dornan, G. L. & Burke, J. E. Using Hydrogen-Deuterium Exchange Mass Spectrometry to Examine Protein-Membrane Interactions. *Methods Enzymol.* **583**, 143–172 (2017).
78. Rand, K. D., Adams, C. M., Zubarev, R. A. & Jørgensen, T. J. D. Electron capture dissociation proceeds with a low degree of intramolecular migration of peptide amide hydrogens. *J. Am. Chem. Soc.* **130**, 1341–1349 (2008).
79. Zehl, M., Rand, K. D., Jensen, O. N. & Jørgensen, T. J. D. Electron transfer dissociation facilitates the measurement of deuterium incorporation into selectively labeled peptides with single residue resolution. *J. Am. Chem. Soc.* **130**, 17453–17459 (2008).
80. Jørgensen, T. J. D., Gårdsvoll, H., Ploug, M. & Roepstorff, P. Intramolecular migration of amide hydrogens in protonated peptides upon collisional activation. *J. Am. Chem. Soc.* **127**, 2785–2793 (2005).
81. Ferguson, P. L. *et al.* Hydrogen/deuterium scrambling during quadrupole time-of-flight MS/MS analysis of a zinc-binding protein domain. *Anal. Chem.* **79**, 153–160 (2007).
82. Sterling, H. J. & Williams, E. R. Real-time hydrogen/deuterium exchange kinetics via supercharged electrospray ionization tandem mass spectrometry. *Anal. Chem.* **82**, 9050–9057 (2010).

83. Resetca, D. & Wilson, D. J. Characterizing rapid, activity-linked conformational transitions in proteins via sub-second hydrogen deuterium exchange mass spectrometry. *FEBS J.* **280**, 5616–5625 (2013).
84. Pan, J., Wilson, D. J. & Konermann, L. Pulsed hydrogen exchange and electrospray charge-state distribution as complementary probes of protein structure in kinetic experiments: implications for ubiquitin folding. *Biochemistry* **44**, 8627–8633 (2005).
85. Resetca, D., Haftchenary, S., Gunning, P. T. & Wilson, D. J. Changes in signal transducer and activator of transcription 3 (STAT3) dynamics induced by complexation with pharmacological inhibitors of Src homology 2 (SH2) domain dimerization. *J. Biol. Chem.* **289**, 32538–32547 (2014).
86. Zhu, S. *et al.* Hyperphosphorylation of intrinsically disordered tau protein induces an amyloidogenic shift in its conformational ensemble. *PLoS One* **10**, e0120416 (2015).
87. Khan, A. G. *et al.* Structure of the core ectodomain of the hepatitis C virus envelope glycoprotein 2. *Nature* **509**, 381–384 (2014).
88. Garcia, N. K., Guttman, M., Ebner, J. L. & Lee, K. K. Dynamic changes during acid-induced activation of influenza hemagglutinin. *Structure* **23**, 665–676 (2015).
89. Davenport, T. M. *et al.* Isolate-specific differences in the conformational dynamics and antigenicity of HIV-1 gp120. *J. Virol.* **87**, 10855–10873 (2013).
90. Guttman, M. *et al.* CD4-induced activation in a soluble HIV-1 Env trimer. *Structure* **22**, 974–984 (2014).
91. Guttman, M., Váradi, C., Lee, K. K. & Guttman, A. Comparative glycoprofiling of HIV gp120 immunogens by capillary electrophoresis and MALDI mass spectrometry. *Electrophoresis* **36**, 1305–1313 (2015).
92. Kim, M. *et al.* Antibody mechanics on a membrane-bound HIV segment essential for GP41-targeted viral neutralization. *Nat. Struct. Mol. Biol.* **18**, 1235–1243 (2011).

93. Bale, S. *et al.* Ebola virus glycoprotein needs an additional trigger, beyond proteolytic priming for membrane fusion. *PLoS Negl. Trop. Dis.* **5**, e1395 (2011).
94. Ye, Q., West, A. M. V., Silletti, S. & Corbett, K. D. Architecture and self-assembly of the SARS-CoV-2 nucleocapsid protein. *bioRxiv* (2020) doi:10.1101/2020.05.17.100685.
95. Chandler, S. A. & Benesch, J. L. Mass spectrometry beyond the native state. *Curr. Opin. Chem. Biol.* **42**, 130–137 (2018).
96. Brogan, W. C., 3rd *et al.* Collection and handling of clinical blood samples to assure the accurate measurement of cocaine concentration. *J. Anal. Toxicol.* **16**, 152–154 (1992).
97. Uetrecht, C., Rose, R. J., van Duijn, E., Lorenzen, K. & Heck, A. J. R. Ion mobility mass spectrometry of proteins and protein assemblies. *Chem. Soc. Rev.* vol. 39 1633–1655 (2010).
98. Krichel, B., Falke, S., Hilgenfeld, R., Redecke, L. & Uetrecht, C. Processing of the SARS-CoV pp1a/ab nsp7–10 region. *Biochem. J* **477**, 1009–1019 (2020).
99. Abzalimov, R. R. & Kaltashov, I. A. Electrospray ionization mass spectrometry of highly heterogeneous protein systems: protein ion charge state assignment via incomplete charge reduction. *Anal. Chem.* **82**, 7523–7526 (2010).
100. Yang, Y., Du, Y. & Kaltashov, I. A. The utility of native MS for understanding the mechanism of action of repurposed therapeutics in COVID-19: heparin as a disruptor of the SARS-CoV-2 interaction with its host cell receptor. *Anal. Chem.* (2020) doi:10.1021/acs.analchem.0c02449.
101. Pierson, E. E. *et al.* Detection of late intermediates in virus capsid assembly by charge detection mass spectrometry. *J. Am. Chem. Soc.* **136**, 3536–3541 (2014).
102. Wang, C. *et al.* A human monoclonal antibody blocking SARS-CoV-2 infection. *Nat. Commun.* **11**, 2251 (2020).
103. Saphire, E. O., Schendel, S. L., Gunn, B. M., Milligan, J. C. & Alter, G. Antibody-mediated protection against Ebola virus. *Nat. Immunol.* **19**, 1169–1178 (2018).

104. Prabakaran, P. *et al.* Potent human monoclonal antibodies against SARS CoV, Nipah and Hendra viruses. *Expert Opin. Biol. Ther.* **9**, 355–368 (2009).
105. Srzentić, K. *et al.* Interlaboratory Study for Characterizing Monoclonal Antibodies by Top-Down and Middle-Down Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **31**, 1783–1802 (2020).
106. Lermyte, F., Tsybin, Y. O., O'Connor, P. B. & Loo, J. A. Top or Middle? Up or Down? Toward a Standard Lexicon for Protein Top-Down and Allied Mass Spectrometry Approaches. *J. Am. Soc. Mass Spectrom.* **30**, 1149–1157 (2019).
107. Fornelli, L. *et al.* Top-down proteomics: Where we are, where we are going? *J. Proteomics* **175**, 3–4 (2018).
108. McLafferty, F. W. *et al.* Top-down MS, a powerful complement to the high capabilities of proteolysis proteomics. *FEBS J.* **274**, 6256–6268 (2007).
109. Resemann, A. *et al.* Full validation of therapeutic antibody sequences by middle-up mass measurements and middle-down protein sequencing. *MAbs* **8**, 318–330 (2016).
110. Donnelly, D. P. *et al.* Best practices and benchmarks for intact protein analysis for top-down mass spectrometry. *Nat. Methods* **16**, 587–594 (2019).
111. Brodbelt, J. S. Photodissociation mass spectrometry: new tools for characterization of biological molecules. *Chem. Soc. Rev.* **43**, 2757–2783 (2014).
112. McLafferty, F. W. *et al.* Electron capture dissociation of gaseous multiply charged ions by Fourier-transform ion cyclotron resonance. *J. Am. Soc. Mass Spectrom.* **12**, 245–249 (2001).
113. Syka, J. E. P., Coon, J. J., Schroeder, M. J., Shabanowitz, J. & Hunt, D. F. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 9528–9533 (2004).
114. Kim, S. J. *et al.* Integrative structure and functional anatomy of a nuclear pore complex. *Nature* **555**, 475–482 (2018).

115. Rout, M. P. & Sali, A. Principles for Integrative Structural Biology Studies. *Cell* **177**, 1384–1403 (2019).
116. Gutierrez, C. *et al.* Structural dynamics of the human COP9 signalosome revealed by cross-linking mass spectrometry and integrative modeling. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 4088–4098 (2020).
117. Samavarchi-Tehrani, P., Samson, R. & Gingras, A.-C. Proximity Dependent Biotinylation: Key Enzymes and Adaptation to Proteomics Approaches. *Mol. Cell. Proteomics* **19**, 757–773 (2020).
118. Heusel, M. *et al.* Complex-centric proteome profiling by SEC-SWATH-MS. *Mol. Syst. Biol.* **15**, e8438 (2019).
119. Heusel, M. *et al.* A Global Screen for Assembly State Changes of the Mitotic Proteome by SEC-SWATH-MS. *Cell Syst* **10**, 133–155.e6 (2020).
120. Batra, J. *et al.* Protein Interaction Mapping Identifies RBBP6 as a Negative Regulator of Ebola Virus Replication. *Cell* **175**, 1917–1930.e13 (2018).
121. Eckhardt, M. *et al.* Multiple Routes to Oncogenesis Are Promoted by the Human Papillomavirus-Host Protein Network. *Cancer Discov.* **8**, 1474–1489 (2018).
122. Jäger, S. *et al.* Global landscape of HIV-human protein complexes. *Nature* **481**, 365–370 (2011).
123. Jean Beltran, P. M., Federspiel, J. D., Sheng, X. & Cristea, I. M. Proteomics and integrative omic approaches for understanding host-pathogen interactions and infectious diseases. *Mol. Syst. Biol.* **13**, 922 (2017).
124. Luo, Y. *et al.* HIV-host interactome revealed directly from infected cells. *Nat Microbiol* **1**, 16068 (2016).
125. Shah, P. S. *et al.* Comparative Flavivirus-Host Protein Interaction Mapping Reveals Mechanisms of Dengue and Zika Virus Pathogenesis. *Cell* **175**, 1931–1945.e18 (2018).

126. Watanabe, T. *et al.* Influenza virus-host interactome screen as a platform for antiviral drug development. *Cell Host Microbe* **16**, 795–805 (2014).
127. Gordon, D. E. *et al.* A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* (2020) doi:10.1038/s41586-020-2286-9.
128. Gordon, D. E. *et al.* Comparative host-coronavirus protein interaction networks reveal pan-viral disease mechanisms. *Science* **370**, (2020).
129. Stukalov, A. *et al.* Multi-level proteomics reveals host-perturbation strategies of SARS-CoV-2 and SARS-CoV. (2020) doi:10.1101/2020.06.17.156455.
130. Roux, K. J., Kim, D. I., Raida, M. & Burke, B. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J. Cell Biol.* **196**, 801–810 (2012).
131. Branon, T. C. *et al.* Efficient proximity labeling in living cells and organisms with TurboID. *Nat. Biotechnol.* **36**, 880–887 (2018).
132. Hung, V. *et al.* Proteomic mapping of the human mitochondrial intermembrane space in live cells via ratiometric APEX tagging. *Mol. Cell* **55**, 332–341 (2014).
133. Lobingier, B. T. *et al.* An Approach to Spatiotemporally Resolve Protein Interaction Networks in Living Cells. *Cell* **169**, 350–360.e12 (2017).
134. Coyaud, E. *et al.* Global Interactomics Uncovers Extensive Organellar Targeting by Zika Virus. *Mol. Cell. Proteomics* **17**, 2242–2255 (2018).
135. V'kovski, P. *et al.* Determination of host proteins composing the microenvironment of coronavirus replicase complexes by proximity-labeling. *Elife* **8**, (2019).
136. Samavarchi-Tehrani, P. *et al.* A SARS-CoV-2 – host proximity interactome. (2020) doi:10.1101/2020.09.03.282103.
137. Schopp, I. M. *et al.* Split-BioID a conditional proteomics approach to monitor the composition of spatiotemporally defined protein complexes. *Nat. Commun.* **8**, 15690 (2017).

- 138.Han, Y. *et al.* Directed Evolution of Split APEX2 Peroxidase. *ACS Chem. Biol.* **14**, 619–635 (2019).
- 139.Cho, K. F. *et al.* Split-TurboID enables contact-dependent proximity labeling in cells. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 12143–12154 (2020).
- 140.Jean Beltran, P. M., Mathias, R. A. & Cristea, I. M. A Portrait of the Human Organelle Proteome In Space and Time during Cytomegalovirus Infection. *Cell Syst* **3**, 361–373.e6 (2016).
- 141.Itzhak, D. N., Tyanova, S., Cox, J. & Borner, G. H. Global, quantitative and dynamic mapping of protein subcellular localization. *Elife* **5**, (2016).
- 142.Geladaki, A. *et al.* Combining LOPIT with differential ultracentrifugation for high-resolution spatial proteomics. *Nat. Commun.* **10**, 331 (2019).
- 143.Greenwood, E. J. *et al.* Temporal proteomic analysis of HIV infection reveals remodelling of the host phosphoproteome by lentiviral Vif variants. *Elife* **5**, (2016).
- 144.Marelli, S. *et al.* Antagonism of PP2A is an independent and conserved function of HIV-1 Vif and causes cell cycle arrest. *Elife* **9**, (2020).
- 145.Naamati, A. *et al.* Functional proteomic atlas of HIV infection in primary human CD4⁺ T cells. *Elife* **8**, (2019).
- 146.Roulston, A., Lin, R., Beauparlant, P., Wainberg, M. A. & Hiscott, J. Regulation of human immunodeficiency virus type 1 and cytokine gene expression in myeloid cells by NF-kappa B/Rel transcription factors. *Microbiol. Rev.* **59**, 481–505 (1995).
- 147.Keating, J. A. & Striker, R. Phosphorylation events during viral infections provide potential therapeutic targets. *Rev. Med. Virol.* **22**, 166–181 (2012).
- 148.Saeed, M. F., Kolokoltsov, A. A., Freiberg, A. N., Holbrook, M. R. & Davey, R. A. Phosphoinositide-3 kinase-Akt pathway controls cellular entry of Ebola virus. *PLoS Pathog.* **4**, e1000141 (2008).

149. Zheng, W. *et al.* Phosphorylation controls the nuclear-cytoplasmic shuttling of influenza A virus nucleoprotein. *J. Virol.* **89**, 5822–5834 (2015).
150. Müller, B., Patschinsky, T. & Kräusslich, H.-G. The late-domain-containing protein p6 is the predominant phosphoprotein of human immunodeficiency virus type 1 particles. *J. Virol.* **76**, 1015–1024 (2002).
151. Bouhaddou, M. *et al.* The Global Phosphorylation Landscape of SARS-CoV-2 Infection. *Cell* **182**, 685–712.e19 (2020).
152. Raaben, M. *et al.* The ubiquitin-proteasome system plays an important role during various stages of the coronavirus infection cycle. *J. Virol.* **84**, 7869–7879 (2010).
153. Yu, X. *et al.* Induction of APOBEC3G ubiquitination and degradation by an HIV-1 Vif-Cul5-SCF complex. *Science* **302**, 1056–1060 (2003).
154. Mehle, A., Goncalves, J., Santa-Marta, M., McPike, M. & Gabuzda, D. Phosphorylation of a novel SOCS-box regulates assembly of the HIV-1 Vif-Cul5 complex that promotes APOBEC3G degradation. *Genes Dev.* **18**, 2861–2866 (2004).
155. Mehle, A. *et al.* Vif overcomes the innate antiviral activity of APOBEC3G by promoting its degradation in the ubiquitin-proteasome pathway. *J. Biol. Chem.* **279**, 7792–7798 (2004).
156. Sheehy, A. M., Gaddis, N. C., Choi, J. D. & Malim, M. H. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature* **418**, 646–650 (2002).
157. Sheehy, A. M., Gaddis, N. C. & Malim, M. H. The antiretroviral enzyme APOBEC3G is degraded by the proteasome in response to HIV-1 Vif. *Nat. Med.* **9**, 1404–1407 (2003).
158. Conticello, S. G., Harris, R. S. & Neuberger, M. S. The Vif protein of HIV triggers degradation of the human antiretroviral DNA deaminase APOBEC3G. *Curr. Biol.* **13**, 2009–2013 (2003).

159. Stopak, K., de Noronha, C., Yonemoto, W. & Greene, W. C. HIV-1 Vif blocks the antiviral activity of APOBEC3G by impairing both its translation and intracellular stability. *Mol. Cell* **12**, 591–601 (2003).
160. Jäger, S. *et al.* Vif hijacks CBF- β to degrade APOBEC3G and promote HIV-1 infection. *Nature* **481**, 371–375 (2011).
161. Hardman, G. *et al.* Strong anion exchange-mediated phosphoproteomics reveals extensive human non-canonical phosphorylation. *EMBO J.* **38**, e100847 (2019).
162. Fuhs, S. R. *et al.* Monoclonal 1- and 3-Phosphohistidine Antibodies: New Tools to Study Histidine Phosphorylation. *Cell* **162**, 198–210 (2015).
163. Li, X. S., Yuan, B. F. & Feng, Y. Q. Recent advances in phosphopeptide enrichment: strategies and techniques. *Trends Analyt. Chem.* (2016).
164. Bekker-Jensen, D. B. *et al.* Rapid and site-specific deep phosphoproteome profiling by data-independent acquisition without the need for spectral libraries. *Nat. Commun.* **11**, 787 (2020).
165. Searle, B. C., Lawrence, R. T., MacCoss, M. J. & Villén, J. Thesaurus: quantifying phosphopeptide positional isomers. *Nat. Methods* **16**, 703–706 (2019).
166. Humphrey, S. J., Azimifar, S. B. & Mann, M. High-throughput phosphoproteomics reveals in vivo insulin signaling dynamics. *Nat. Biotechnol.* **33**, 990–995 (2015).
167. Hernandez-Armenta, C., Ochoa, D., Gonçalves, E., Saez-Rodriguez, J. & Beltrao, P. Benchmarking substrate-based kinase activity inference using phosphoproteomic data. *Bioinformatics* **33**, 1845–1851 (2017).
168. Ochoa, D. *et al.* An atlas of human kinase regulation. *Mol. Syst. Biol.* **12**, 888 (2016).
169. Krug, K. *et al.* A Curated Resource for Phosphosite-specific Signature Analysis. *Mol. Cell. Proteomics* **18**, 576–593 (2019).

170. Davidson, A. D. *et al.* Characterisation of the transcriptome and proteome of SARS-CoV-2 reveals a cell passage induced in-frame deletion of the furin-like cleavage site from the spike glycoprotein. *Genome Med.* **12**, 68 (2020).
171. Klann, K. *et al.* Growth factor receptor signaling inhibition prevents SARS-CoV-2 replication. *Mol. Cell* (2020) doi:10.1016/j.molcel.2020.08.006.
172. Mahon, C., Krogan, N. J., Craik, C. S. & Pick, E. Cullin E3 ligases and their rewiring by viral factors. *Biomolecules* **4**, 897–930 (2014).
173. Viswanathan, K., Früh, K. & DeFilippis, V. Viral hijacking of the host ubiquitin system to evade interferon responses. *Curr. Opin. Microbiol.* **13**, 517–523 (2010).
174. Shin, J. Y. *et al.* Deubiquitination Reactions on the Proteasome for Proteasome Versatility. *Int. J. Mol. Sci.* **21**, (2020).
175. Li, T. & Zou, C. The Role of Deubiquitinating Enzymes in Acute Lung Injury and Acute Respiratory Distress Syndrome. *Int. J. Mol. Sci.* **21**, (2020).
176. Chakraborty, J. & Ziviani, E. Deubiquitinating Enzymes in Parkinson's Disease. *Front. Physiol.* **11**, 535 (2020).
177. Celebi, G., Kesim, H., Ozer, E. & Kutlu, O. The Effect of Dysfunctional Ubiquitin Enzymes in the Pathogenesis of Most Common Diseases. *Int. J. Mol. Sci.* **21**, (2020).
178. Xu, G., Paige, J. S. & Jaffrey, S. R. Global analysis of lysine ubiquitination by ubiquitin remnant immunoaffinity profiling. *Nat. Biotechnol.* **28**, 868–873 (2010).
179. Kim, W. *et al.* Systematic and quantitative assessment of the ubiquitin-modified proteome. *Mol. Cell* **44**, 325–340 (2011).
180. Hoffmann, H.-H. *et al.* Functional interrogation of a SARS-CoV-2 host protein interactome identifies unique and shared coronavirus host factors. *bioRxiv* (2020) doi:10.1101/2020.09.11.291716.
181. Klemm, T. *et al.* Mechanism and inhibition of the papain-like protease, PLpro, of SARS-CoV-2. *EMBO J.* e106275 (2020).

- 182.Bojkova, D. *et al.* Proteomics of SARS-CoV-2-infected host cells reveals therapy targets. *Nature* **583**, 469–472 (2020).
- 183.Johnson, J. R. *et al.* Global post-translational modification profiling of HIV-1-infected cells reveals mechanisms of host cellular pathway remodeling. (2020)
doi:10.1101/2020.01.06.896365.
- 184.Eckhardt, M., Hultquist, J. F., Kaake, R. M., Hüttenhain, R. & Krogan, N. J. A systems approach to infectious disease. *Nat. Rev. Genet.* **21**, 339–354 (2020).
- 185.Hüttenhain, R. *et al.* Reproducible quantification of cancer-associated proteins in body fluids using targeted proteomics. *Sci. Transl. Med.* **4**, 142ra94 (2012).
- 186.Bezstarosti, K., Lamers, M. M., Haagmans, B. L. & Demmers, J. A. A. Targeted Proteomics for the Detection of SARS-CoV-2 Proteins. *bioRxiv* (2020) doi:10.1101/2020.04.23.057810.
- 187.Ihling, C. *et al.* Mass Spectrometric Identification of SARS-CoV-2 Proteins from Gargle Solution Samples of COVID-19 Patients. *bioRxiv* (2020) doi:10.1101/2020.04.18.047878.
- 188.Zecha, J. *et al.* Data, Reagents, Assays and Merits of Proteomics for SARS-CoV-2 Research and Testing. *Mol. Cell. Proteomics* **19**, 1503–1522 (2020).
- 189.Cardozo, K. H. M. *et al.* Establishing a mass spectrometry-based system for rapid detection of SARS-CoV-2 in large clinical sample cohorts. *Nat. Commun.* **11**, 6201 (2020).
- 190.Geyer, P. E. *et al.* Plasma Proteome Profiling to detect and avoid sample-related biases in biomarker studies. *EMBO Mol. Med.* **11**, e10427 (2019).
- 191.Geyer, P. E., Holdt, L. M., Teupser, D. & Mann, M. Revisiting biomarker discovery by plasma proteomics. *Mol. Syst. Biol.* **13**, 942 (2017).
- 192.Messner, C. B. *et al.* Ultra-High-Throughput Clinical Proteomics Reveals Classifiers of COVID-19 Infection. *Cell Syst* (2020) doi:10.1016/j.cels.2020.05.012.
- 193.Fontana, A. *et al.* Probing protein structure by limited proteolysis. *Acta Biochim. Pol.* **51**, 299–321 (2004).

194. Feng, Y. *et al.* Global analysis of protein structural changes in complex proteomes. *Nat. Biotechnol.* **32**, 1036–1044 (2014).
195. Piazza, I. *et al.* A Map of Protein-Metabolite Interactions Reveals Principles of Chemical Communication. *Cell* **172**, 358–372.e23 (2018).
196. Piazza, I. *et al.* A machine learning-based chemoproteomic approach to identify drug targets and binding sites in complex proteomes. *Nat. Commun.* **11**, 4200 (2020).
197. Savitski, M. M. *et al.* Tracking cancer drugs in living cells by thermal profiling of the proteome. *Science* **346**, 1255784 (2014).
198. Martinez Molina, D. *et al.* Monitoring drug target engagement in cells and tissues using the cellular thermal shift assay. *Science* **341**, 84–87 (2013).
199. Reinhard, F. B. M. *et al.* Thermal proteome profiling monitors ligand interactions with cellular membrane proteins. *Nat. Methods* **12**, 1129–1131 (2015).
200. Perrin, J. *et al.* Identifying drug targets in tissues and whole blood with thermal-shift profiling. *Nat. Biotechnol.* **38**, 303–308 (2020).
201. Friman, T., Chernobrovkin, A., Molina, D. M. & Arnold, L. H. CETSA MS profiling for a comparative assessment of FDA approved antivirals repurposed for COVID-19 therapy identifies Trip13 as a Remdesivir off-target. *bioRxiv* (2020).
202. WHO recommends against the use of remdesivir in COVID-19 patients. (2020).
203. Niphakis, M. J. & Cravatt, B. F. Enzyme inhibitor discovery by activity-based protein profiling. *Annu. Rev. Biochem.* **83**, 341–377 (2014).
204. Kahler, J. P., Vanhoutte, R. & Verhelst, S. H. L. Activity-Based Protein Profiling of Serine Proteases in Immune Cells. *Arch. Immunol. Ther. Exp.* **68**, 23 (2020).
205. Bar-Peled, L. *et al.* Chemical Proteomics Identifies Druggable Vulnerabilities in a Genetically Defined Cancer. *Cell* **171**, 696–709.e23 (2017).
206. Kok, B. P. *et al.* Discovery of small-molecule enzyme activators by activity-based protein profiling. *Nat. Chem. Biol.* (2020) doi:10.1038/s41589-020-0555-4.

207. Fuerst, R. & Breinbauer, R. Activity-Based Protein Profiling (ABPP) of Oxidoreductases. *Chembiochem* (2020) doi:10.1002/cbic.202000542.
208. Blewett, M. M. *et al.* Chemical proteomic map of dimethyl fumarate–sensitive cysteines in primary human T cells. *Science Signaling* vol. 9 rs10–rs10 (2016).
209. Baggelaar, M. P. *et al.* ABHD2 Inhibitor Identified by Activity-Based Protein Profiling Reduces Acrosome Reaction. *ACS Chem. Biol.* **14**, 2943 (2019).
210. Riva, L. *et al.* Discovery of SARS-CoV-2 antiviral drugs through large-scale compound repurposing. *Nature* (2020) doi:10.1038/s41586-020-2577-1.
211. Chun, T. W. *et al.* Early establishment of a pool of latently infected, resting CD4(+) T cells during primary HIV-1 infection. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 8869–8873 (1998).
212. Whitney, J. B. *et al.* Rapid seeding of the viral reservoir prior to SIV viraemia in rhesus monkeys. *Nature* **512**, 74–77 (2014).
213. Deeks, S. G. HIV: Shock and kill. *Nature* vol. 487 439–440 (2012).
214. Rasmussen, T. A. & Lewin, S. R. Shocking HIV out of hiding: where are we with clinical trials of latency reversing agents? *Curr. Opin. HIV AIDS* **11**, 394–401 (2016).
215. Valerdi, K. M., Hage, A., van Tol, S., Rajsbaum, R. & Giraldo, M. I. The Role of the Host Ubiquitin System in Promoting Replication of Emergent Viruses. *Viruses* **13**, (2021).
216. Wimmer, P. & Schreiner, S. Viral Mimicry to Usurp Ubiquitin and SUMO Host Pathways. *Viruses* **7**, 4854–4872 (2015).
217. Bielskienė, K., Bagdonienė, L., Mozūraitienė, J., Kazbarienė, B. & Janulionis, E. E3 ubiquitin ligases as drug targets and prognostic biomarkers in melanoma. *Medicina* **51**, 1–9 (2015).
218. Morreale, F. E. & Walden, H. Types of Ubiquitin Ligases. *Cell* **165**, 248–248.e1 (2016).
219. Binette, J. *et al.* Requirements for the selective degradation of CD4 receptor molecules by the human immunodeficiency virus type 1 Vpu protein in the endoplasmic reticulum. *Retrovirology* **4**, 75 (2007).

220. Wildum, S., Schindler, M., Münch, J. & Kirchhoff, F. Contribution of Vpu, Env, and Nef to CD4 down-modulation and resistance of human immunodeficiency virus type 1-infected T cells to superinfection. *J. Virol.* **80**, 8047–8059 (2006).
221. Hüttenhain, R. *et al.* ARIH2 Is a Vif-Dependent Regulator of CUL5-Mediated APOBEC3G Degradation in HIV Infection. *Cell Host Microbe* **26**, 86–99.e7 (2019).
222. Rathore, A. *et al.* CRISPR-based gene knockout screens reveal deubiquitinases involved in HIV-1 latency in two Jurkat cell models. *Sci. Rep.* **10**, 5350 (2020).
223. Ma, X. *et al.* TRIM28 promotes HIV-1 latency by SUMOylating CDK9 and inhibiting P-TEFb. *Elife* **8**, (2019).
224. Bosque, A. *et al.* Benzotriazoles Reactivate Latent HIV-1 through Inactivation of STAT5 SUMOylation. *Cell Rep.* **18**, 1324–1334 (2017).
225. König, R. *et al.* Global analysis of host-pathogen interactions that regulate early-stage HIV-1 replication. *Cell* **135**, 49–60 (2008).
226. Bushman, F. D. *et al.* Host cell factors in HIV replication: meta-analysis of genome-wide studies. *PLoS Pathog.* **5**, e1000437 (2009).
227. Liu, Y. *et al.* Proteomic profiling of HIV-1 infection of human CD4⁺ T cells identifies PSGL-1 as an HIV restriction factor. *Nat Microbiol* **4**, 813–825 (2019).
228. Medvar, B., Raghuram, V., Pisitkun, T., Sarkar, A. & Knepper, M. A. Comprehensive database of human E3 ubiquitin ligases: application to aquaporin-2 regulation. *Physiol. Genomics* **48**, 502–512 (2016).
229. Münch, J. *et al.* Nef-mediated enhancement of virion infectivity and stimulation of viral replication are fundamental properties of primate lentiviruses. *J. Virol.* **81**, 13852–13864 (2007).
230. Hultquist, J. F. *et al.* CRISPR-Cas9 genome engineering of primary CD4⁺ T cells for the interrogation of HIV-host factor interactions. *Nat. Protoc.* **14**, 1–27 (2019).

231. Berson, J. F. *et al.* A seven-transmembrane domain receptor involved in fusion and entry of T-cell-tropic human immunodeficiency virus type 1 strains. *J. Virol.* **70**, 6288–6295 (1996).
232. Deng, H. *et al.* Identification of a major co-receptor for primary isolates of HIV-1. *Nature* **381**, 661–666 (1996).
233. Feng, Y., Broder, C. C., Kennedy, P. E. & Berger, E. A. HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. *Science* **272**, 872–877 (1996).
234. Zhu, Y. *et al.* Transcription elongation factor P-TEFb is required for HIV-1 tat transactivation in vitro. *Genes Dev.* **11**, 2622–2632 (1997).
235. Mancebo, H. S. *et al.* P-TEFb kinase is required for HIV Tat transcriptional activation in vivo and in vitro. *Genes Dev.* **11**, 2633–2644 (1997).
236. Llano, M. *et al.* An essential role for LEDGF/p75 in HIV integration. *Science* **314**, 461–464 (2006).
237. Jin, X. *et al.* An atypical E3 ligase zinc finger protein 91 stabilizes and activates NF-kappaB-inducing kinase via Lys63-linked ubiquitination. *J. Biol. Chem.* **285**, 30539–30547 (2010).
238. Jin, H. R., Jin, X. & Lee, J. J. Zinc-finger protein 91 plays a key role in LIGHT-induced activation of non-canonical NF-κB pathway. *Biochem. Biophys. Res. Commun.* **400**, 581–586 (2010).
239. Faust, T. B. *et al.* PJA2 ubiquitinates the HIV-1 Tat protein with atypical chain linkages to activate viral transcription. *Sci. Rep.* **7**, 45394 (2017).
240. Banoth, B. & Cassel, S. L. Mitochondria in innate immune signaling. *Transl. Res.* **202**, 52–68 (2018).
241. Shi, H.-X. *et al.* Mitochondrial ubiquitin ligase MARCH5 promotes TLR7 signaling by attenuating TANK action. *PLoS Pathog.* **7**, e1002057 (2011).

242. Tsai, A. *et al.* Toll-Like Receptor 7 Agonist GS-9620 Induces HIV Expression and HIV-Specific Immunity in Cells from HIV-Infected Individuals on Suppressive Antiretroviral Therapy. *J. Virol.* **91**, (2017).
243. Hornung, V. *et al.* Quantitative expression of toll-like receptor 1-10 mRNA in cellular subsets of human peripheral blood mononuclear cells and sensitivity to CpG oligodeoxynucleotides. *J. Immunol.* **168**, 4531–4537 (2002).
244. Yoo, Y.-S. *et al.* The mitochondrial ubiquitin ligase MARCH5 resolves MAVS aggregates during antiviral signalling. *Nat. Commun.* **6**, 7910 (2015).
245. Verdikt, R. *et al.* Novel Role of UHRF1 in DNA methylation-mediated repression of latent HIV-1. *bioRxiv* 2021.06.15.448539 (2021) doi:10.1101/2021.06.15.448539.
246. Kuo, R.-L. *et al.* Interactome Analysis of the NS1 Protein Encoded by Influenza A H1N1 Virus Reveals a Positive Regulatory Role of Host Protein PRP19 in Viral Replication. *J. Proteome Res.* **15**, 1639–1648 (2016).
247. Luban, J., Bossolt, K. L., Franke, E. K., Kalpana, G. V. & Goff, S. P. Human immunodeficiency virus type 1 Gag protein binds to cyclophilins A and B. *Cell* **73**, 1067–1078 (1993).
248. Braaten, D. & Luban, J. Cyclophilin A regulates HIV-1 infectivity, as demonstrated by gene targeting in human T cells. *EMBO J.* **20**, 1300–1309 (2001).
249. Satijn, D. P. *et al.* RING1 is associated with the polycomb group protein complex and acts as a transcriptional repressor. *Mol. Cell. Biol.* **17**, 4105–4113 (1997).
250. Sharma, A. L. *et al.* CBF-1 Promotes the Establishment and Maintenance of HIV Latency by Recruiting Polycomb Repressive Complexes, PRC1 and PRC2, at HIV LTR. *Viruses* **12**, (2020).
251. Khan, S., Iqbal, M., Tariq, M., Baig, S. M. & Abbas, W. Epigenetic regulation of HIV-1 latency: focus on polycomb group (PcG) proteins. *Clin. Epigenetics* **10**, 14 (2018).

252. Marian, C. A. *et al.* Small Molecule Targeting of Specific BAF (mSWI/SNF) Complexes for HIV Latency Reversal. *Cell Chem Biol* **25**, 1443–1455.e14 (2018).
253. Bieberich, F. *et al.* A Single-Cell Atlas of Lymphocyte Adaptive Immune Repertoires and Transcriptomes Reveals Age-Related Differences in Convalescent COVID-19 Patients. *Front. Immunol.* **12**, 701085 (2021).
254. Bradley, T., Ferrari, G., Haynes, B. F., Margolis, D. M. & Browne, E. P. Single-Cell Analysis of Quiescent HIV Infection Reveals Host Transcriptional Profiles that Regulate Proviral Latency. *Cell Rep.* **25**, 107–117.e3 (2018).
255. Tomita, Y. *et al.* The cellular factors Vps18 and Mon2 are required for efficient production of infectious HIV-1 particles. *J. Virol.* **85**, 5618–5627 (2011).
256. Zhou, H. *et al.* Genome-scale RNAi screen for host factors required for HIV replication. *Cell Host Microbe* **4**, 495–504 (2008).
257. Brass, A. L. *et al.* Identification of host proteins required for HIV infection through a functional genomic screen. *Science* **319**, 921–926 (2008).
258. Yeung, M. L., Houzet, L., Yedavalli, V. S. R. K. & Jeang, K.-T. A genome-wide short hairpin RNA screening of jurkat T-cells for human proteins contributing to productive HIV-1 replication. *J. Biol. Chem.* **284**, 19463–19473 (2009).
259. Cowen, L., Ideker, T., Raphael, B. J. & Sharan, R. Network propagation: a universal amplifier of genetic associations. *Nat. Rev. Genet.* **18**, 551–562 (2017).
260. Wu, G., Feng, X. & Stein, L. A human functional protein interaction network and its application to cancer data analysis. *Genome Biol.* **11**, R53 (2010).
261. Giurgiu, M. *et al.* CORUM: the comprehensive resource of mammalian protein complexes-2019. *Nucleic Acids Res.* **47**, D559–D563 (2019).
262. Pache, L. *et al.* BIRC2/cIAP1 Is a Negative Regulator of HIV-1 Transcription and Can Be Targeted by Smac Mimetics to Promote Reversal of Viral Latency. *Cell Host Microbe* **18**, 345–353 (2015).

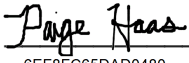
263. Nixon, C. C. *et al.* Systemic HIV and SIV latency reversal via non-canonical NF- κ B signalling in vivo. *Nature* **578**, 160–165 (2020).
264. Vallabhapurapu, S. *et al.* Nonredundant and complementary functions of TRAF2 and TRAF3 in a ubiquitination cascade that activates NIK-dependent alternative NF- κ B signaling. *Nat. Immunol.* **9**, 1364–1370 (2008).
265. Zarnegar, B. J. *et al.* Noncanonical NF- κ B activation requires coordinated assembly of a regulatory complex of the adaptors cIAP1, cIAP2, TRAF2 and TRAF3 and the kinase NIK. *Nat. Immunol.* **9**, 1371–1378 (2008).
266. Jordan, A., Bisgrove, D. & Verdin, E. HIV reproducibly establishes a latent infection after acute infection of T cells in vitro. *EMBO J.* **22**, 1868–1877 (2003).
267. Huang, H. *et al.* A Novel Bromodomain Inhibitor Reverses HIV-1 Latency through Specific Binding with BRD4 to Promote Tat and P-TEFb Association. *Front. Microbiol.* **8**, 1035 (2017).
268. Dashti, A. *et al.* SMAC Mimetic Plus Triple-Combination Bispecific HIVxCD3 Retargeting Molecules in SHIV.C.CH505-Infected, Antiretroviral Therapy-Suppressed Rhesus Macaques. *J. Virol.* **94**, (2020).
269. Darcis, G. *et al.* An In-Depth Comparison of Latency-Reversing Agent Combinations in Various In Vitro and Ex Vivo HIV-1 Latency Models Identified Bryostatin-1+JQ1 and Ingenol-B+JQ1 to Potently Reactivate Viral Gene Expression. *PLoS Pathog.* **11**, e1005063 (2015).
270. Jiang, G. *et al.* Synergistic Reactivation of Latent HIV Expression by Ingenol-3-Angelate, PEP005, Targeted NF- κ B Signaling in Combination with JQ1 Induced p-TEFb Activation. *PLoS Pathog.* **11**, e1005066 (2015).
271. Lassen, K. G., Hebbeler, A. M., Bhattacharyya, D., Lobritz, M. A. & Greene, W. C. A flexible model of HIV-1 latency permitting evaluation of many primary CD4 T-cell reservoirs. *PLoS One* **7**, e30176 (2012).

272. Spina, C. A. *et al.* An in-depth comparison of latent HIV-1 reactivation in multiple cell model systems and resting CD4⁺ T cells from aviremic patients. *PLoS Pathog.* **9**, e1003834 (2013).
273. Brinkman, E. K., Chen, T., Amendola, M. & van Steensel, B. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res.* **42**, e168 (2014).
274. Sun, S.-C. The noncanonical NF- κ B pathway. *Immunol. Rev.* **246**, 125–140 (2012).
275. Sun, S.-C. Non-canonical NF- κ B signaling pathway. *Cell Res.* **21**, 71–85 (2011).
276. Csardi, G., Nepusz, T. & Others. The igraph software package for complex network research. *InterJournal, complex systems* **1695**, 1–9 (2006).

Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

DocuSigned by:

6EF8FC65DAD0480... Author Signature

9/1/2021
Date