Pacific Northwest
NATIONAL LABORATORY

PNNL-38328

Controlling Host Responses to Infection

September 2025

Hugh Mitchell
Sankarganesh Krishnamoorthy
Christopher R Anderton
Madelyn Berger, Arunima Bhattacharjee
Emma Carlson, Tony Chiang**
William Chrisler, Rosey Chu, Javier Flores
Damon Leach, Owen Leiser
Andy Lin, Irlanda Medrano, Nicholas Reichart
Adam Ryan, Gregory Vandergrift
Amy C Sims
** current contact information
chiang@math.washington.edu

U.S. DEPARTMENT of ENERGY

Prepared for the U.S. Department of Energy
under Contract DE-AC05-76RL01830

DISCLAIMER

PACIFIC NORTHWEST NATIONAL LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-AC05-76RL01830

Printed in the United States of America

Available to DOE and DOE contractors from
the Office of Scientific and Technical Information,
P.O. Box 62, Oak Ridge, TN 37831-0062
www.osti.gov
ph: (865) 576-8401
fox: (865) 576-5728
email: reports@osti.gov

Available to the public from the National Technical Information Service
5301 Shawnee Rd., Alexandria, VA 22312
ph: (800) 553-NTIS (6847)
or (703) 605-6000
email: info@ntis.gov
Online ordering: http://www.ntis.gov

PNNL-38328

Controlling Host Responses to Infection

September 2025

Hugh Mitchell
Sankarganesh Krishnamoorthy
Christopher R Anderton
Madelyn Berger, Arunima Bhattacharjee
Emma Carlson, Tony Chiang**
William Chrisler, Rosey Chu, Javier Flores
Damon Leach, Owen Leiser
Andy Lin, Irlanda Medrano, Nicholas Reichart
Adam Ryan, Gregory Vandergrift
Amy C Sims
** current contact information chiang@math.washington.edu

Prepared for
the U.S. Department of Energy
under Contract DE-AC05-76RL01830

Pacific Northwest National Laboratory
Richland, Washington 99354

Abstract

Pathogen invasion of host cells causes a myriad of functional changes including alterations of chromatin accessibility often limiting defense responses, shunting of cellular resources to centers of viral replication, and rearrangement of intracellular membranes to facilitate genome reproduction and progeny release. Systems biology approaches provide global snapshots of pathogen induced changes following infection and provide a variety of tools to begin to define how cellular homeostasis is disrupted, but improvements on these tools are required to determine how cellular functions are altered post infection. Chromatin accessibility techniques, biochemical assays to assess the activity of epigenetic enzymes, scalable sample collection platforms, and affinity-based probes were used to characterize how human respiratory viruses modify host responses in infected human lungs over time. These studies enhanced our knowledge of how pathogens usurp the host environment during infection and identify additional targets for future evaluations of medical countermeasures.

Summary

Systems biology approaches provide global snapshots of pathogen induced changes following infection and provide a variety of tools to begin to define how cellular homeostasis is disrupted, but improvements on these tools are required to determine how cellular functions are altered post infection. Chromatin accessibility techniques, biochemical assays to assess the activity of epigenetic enzymes, scalable sample collection platforms, and affinity-based probes were used to characterize how human respiratory viruses modify host responses in infected human lungs over time. These studies enhanced our knowledge of how pathogens usurp the host environment during infection and identify additional targets for future evaluations of medical countermeasures.

Acknowledgments

This research was supported by the Predictive Phenomics Initiative, under the Laboratory Directed Research and Development (LDRD) Program at Pacific Northwest National Laboratory (PNNL). PNNL is a multi-program national laboratory operated for the U.S. Department of Energy (DOE) by Battelle Memorial Institute under Contract No. DE-AC05-76RL01830.

Acronyms and Abbreviations

3D	3-Dimentional
ANOVA	Analysis of variance; statistical analysis
ATAC-Seq	Assay for transposase-accessible chromatin using sequencing
BCA	Bicinchonic acid
CARM1	Coactivator associated with arginine methyltransferase 1
CoV	Coronavirus
COVID19	Coronavirus infectious disease 2019
DNA	Deoxyribonucleic acid
EHMT1	Euchromatic histone methyltransferase 1 enzymes
ELISA	Enzyme-linked immunosorbent assay
EZH2	Enhancer of zeste homolog 2
FC	Fold change
FDR	False discovery rate
FRIP	Fraction of reads in peaks
GB	Gigabytes
GSK126	Small molecule inhibitor that works against EZH2
H1.4	Histone 1.4
H2B	Histone 2B
H3	Histone 3
H3K4	Histone 3 lysine 4
H3K27me3	Histone 2 lysine 27 with three methyl groups added to lysine 27
HCoV-229E	Human coronavirus 229E
HCoV	Human coronavirus
HPC	High performance computing
HuH7	Immortalized human liver cells
IC50	Infectious concentration 50
KEGG	Kyoto encyclopedia of genes and genomes
kDa	Kilodalton
K L	Lysine
MAD	Median absolute distance
MASIC MS/MS	Automated selected ion chromatogram generator
MB	Megabytes
MERS-CoV	Middle East respiratory syndrome coronavirus
METTL3	N6-adenosine-methyltransferase 70 kDa subunit
MOI	Multiplicity of infection
MRC5	Immortalized human lung fibroblasts

MSGF	Software that performs peptide identification
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
MTPP1	Methyltransferase photoaffinity probe derived from GSK126
MW	Molecular weight
nM	Nanomolar
nm	Nanometer
nanoPOTS	Nanodroplet processing in one pot for trace samples (proteomics)
OSX	Operating system
PBS	Phosphate buffered saline
PNNL	Pacific Northwest National Laboratory
PRC2	Polycomb repressive complex 2
RAM	Random access memory
RBBP4	RB binding protein 4, chromatin remodeling factor
SAM	S-adenosylmethionine
SARS-CoV 2	Severe acute respiratory syndrome coronavirus 2
SARS-CoV	Severe acute respiratory syndrome coronavirus
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SHMT2	Serine hydroxymethyltransferase 2
sd	Standard deviation
SPD	Serial presence detect
TSS	Transcription start sites
ug	Microgram
uL	Microliter
uM	Micromolar
UNC1999	Methyltransferase inhibitor
UV	Ultraviolet
vs	Versus
V	Version
ZLD1039	Methyltransferase inhibitor

Table of Contents

Figures

Tables

1.0 Introduction

1.1 Coronaviruses

Coronaviruses (CoV) are important emerging pathogens associated with mild and severe disease outcomes in humans and animals. Newly emerged CoV have caused significant global morbidity and mortality. In 2003, severe acute respiratory coronavirus (SARS-CoV) emerged from closely related group 2b bat CoV in Southeast Asia and spread to 33 nations, resulting in ~8,000 human cases and 800 deaths, worldwide. In 2012, Middle East respiratory syndrome CoV (MERS-CoV) emerged, causing ~2,200 cases with a ~36% mortality rate. MERS-CoV cases have been reported in 27 countries and the outbreak is still ongoing. In 2020, SARS-CoV 2 emerged and was identified as the causative agent of COVID19 and the global pandemic wreaking havoc with global economies and overwhelming healthcare systems. Like SARS-CoV and SARS-CoV 2, MERS-CoV disease severity is strongly influenced by aging and other co-morbidities (e.g., diabetes, obesity) and mortality rates exceed 50% in aged individuals (>60 years). These data not only underscore the highly pathogenic potential of an emerging CoV but reinforce the critical need for the development of new strategies to understand pathogenesis and reverse severe disease outcomes. In contrast, human coronavirus 229E (HCoV-229E), causes a mild upper respiratory tract disease in otherwise healthy individuals and circulates annually worldwide. Despite differences in disease outcomes the mild and severe disease HCoV share many features including genome configurations, genomic replication strategies, conserved viral proteins and in many cases altering similar host cellular networks to enhance and facilitate viral infection.

1.2 Epigenetics and respiratory virus infection

In human cells, transcription and translation of new mRNA and proteins relies on accessibility of histone wrapped genomic DNA or chromatin. The histone proteins (typically on lysine groups) can be post translationally modified by the addition of methyl or acyl groups and placement of these moieties determine whether the histone-wrapped DNA is covering (condensed, repressed) or has unwound the DNA (loose, active) at sites for transcription factor binding. Human respiratory virus infections induce a myriad of changes in infected host cells including multiple mechanisms to escape the innate immune response and allow the virus time to establish replication factories and begin assembly of progeny virions. Transcriptomic analysis of highly pathogenic human coronavirus infected cells revealed an 18 to 24 hours lag in the differential expression of genes post infection allowing the virus to “hide” from host defenses (Sims, Tilton et al. 2013, Menachery, Eisfeld et al. 2014). Following this lag, significant numbers of differentially expressed genes were detected, and a subset of these genes were categorized as interferon stimulated genes. Studies to determine how CoV and influenza virus infected host cells regulated differential gene synthesis demonstrated that availability of transcription factors could explain some, but not all the regulated gene expression. This was especially true for subsets of down regulated interferon stimulated genes. Subsequent studies that examined the accessibility of host chromatin suggested that the methylation state of histone 3 lysines tracked with whether interferon

stimulated genes were expressed post infection. Similarly, studies with HCoV-229E demonstrated significant changes to the accessibility of host chromatin post infection (Poppe, Wittig et al. 2017). Alterations were detected in both the methylation and acylation states of histones in the HCoV-229E studies suggesting that controlling host gene expression is a conserved mechanism used by respiratory viruses that cause both severe and less severe disease phenotypes. The mechanisms by which CoV are controlling host chromatin accessibility have not been defined.

2.0 Results and Discussion

2.1 Assays targeting human chromatin modifying enzymes

2.1.1 Photoaffinity-based probes

Chromatin accessibility is mediated by post translational modifications on histone proteins that control whether DNA is “open” for transcription or “closed”. Histone methylation and acetylation are two of seven histone modifications that can alter the chromatin landscape. Alterations in chromatin accessibility have been demonstrated for both low (HCoV-229E (Poppe, Wittig et al. 2017)) and high (MERS-CoV (Menachery, Eisfeld et al. 2014)) disease severity causing human coronaviruses however, little is known about the host enzymatic functions that alter chromatin accessibility.

GSK126 is a potent, selective, small-molecule inhibitor of Enhancer of Zeste Homolog 2 (EZH2), the catalytic subunit of the Polycomb Repressive Complex 2 (PRC2) responsible for tri-methylation of histone 3 (H3) at lysine (K) 27 (H3K27me3), a critical epigenetic mark associated with transcriptional repression (Margueron and Reinberg 2011, McCabe, Ott et al. 2012). By competitively binding to the S-adenosylmethionine (SAM) binding site of EZH2, GSK126 effectively inhibits its methyltransferase activity, leading to a global reduction of H3K27me3 levels. While GSK126 is a highly selective and potent inhibitor of EZH2, its broader molecular engagement within chromatin-associated complexes and potential context-dependent interactions remain insufficiently characterized. Developing a photoaffinity probe based on the GSK126 scaffold with a reporter group allowed covalent crosslinking and capture of its cellular and molecular targets (Smith and Collins 2015).

Probe design. Methyltransferase inhibitors (e.g., UNC1999, GSK126, ZLD1039 and related analogs) accommodate alkyl groups and maintain their capacity to block enzymatic functions at nanomolar concentrations (IC50 10 to 90 nM) (Yang, Li et al. 2016). The photoaffinity probe MTPP1 was obtained from one step alkylation of GSK126 (MTPP1). MTPP1 is complementary to the previously reported biotin probe UNC2399 (Konze, Ma et al. 2013), allowing for in cell treatment and photoactivation to covalently label probe targets. The labeled cells then can be used to install fluorescent dye or biotin using copper catalyzed click chemistry for whole cell fluorescent imaging and/or pull-down proteomics.

Human lung cell HCoV-229E infection and MTPP1 treatment. Human lung MRC5 (fibroblasts) cells were plated 24 hours prior to infection/treatment. Conditions included mock-infected cells +/- MTPP1 treatment, and HCoV-229E infected (multiplicity of infection (MOI) 1) cells +/- MTPP1 treatment (two concentrations 10 and 40 μM). HCoV-229E and MTPP1 were added simultaneously to the cells for 48 hours. Cells were treated with ultraviolet light for (360 nm) for seven minutes (inside the biological safety cabinet), media was removed, and cells were scraped into 1X phosphate buffered saline. Cells were lysed by a final volume of 0.5% sodium dodecyl sulfate. Samples were pelleted following lysis and then supernatants used in bicinchonic acid

(BCA) assays to determine protein concentrations. Samples were normalized to the lowest protein concentration. Click chemistry reactions were performed and samples reduced in 2X Sample (Laemmli buffer) prior to being resolved by electrophoresis on a 10 to 20% Tris Glycine protein gel. Labeled proteins were analyzed on a Typhoon laser scanner and then the gel was re-stained with Coomassie stain to look at total numbers of proteins (Figure 1).

Proteomic sample preparation and analysis. Proteomic samples were prepared by pulling down proteins via streptavidin-based enrichment. All treatments were done in triplicate. Proteome samples were clicked onto biotin by copper catalyzed click chemistry. Samples were incubated with pre-washed streptavidin beads on a rotating

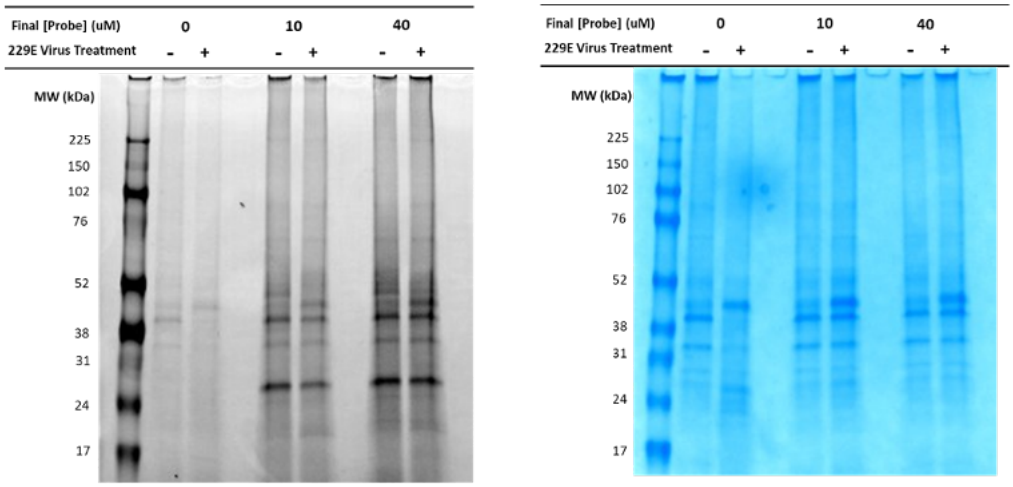

Figure 1: SDS-PAGE analysis of MTPP1 probe-labeling at fixed concentrations (10 μM and 40 μM) of mock-infected and HCoV-229E infected lung cells.

wheel. Beads were recovered and resuspended in digestion buffer with 0.25 μg trypsin. The resulting peptides were reconstituted in buffer. Samples were clarified by ultracentrifugation before being analyzed by mass spectrometry. LC-MS/MS measurements were made on peptide samples using a Exploris 480 mass spectrometer (Thermo Scientific, San Jose, CA) coupled to LC with C18 Evosep EV1137 Performance column for 15/30 SPD. Separations were performed at room temperature with a mobile phase consisted of (A) 0.1% formic acid in water and (B) 0.1% formic acid in acetonitrile over 1 h gradient. MASIC (MS/MS Automated Selected Ion Chromatogram generator)44 and MSGF+45 (v2023.01.12) were used for peptide abundance and identification, as previously described (Monroe, Shaw et al. 2008, Kim and Pevzner 2014).

MASIC and MSGF+ results were aggregated using Mage software and a 5% false discovery rate (FDR) based on Q-value was used for initial filtering. Subsequent analyses were conducted using chemoprotR (https://code.emsl.pnl.gov/multiomics-analyses/chemoproteomicsR/-/tree/main). Peptide identifications by organism were filtered using a MSGF spectral probability value of 1 x 10 -9 corresponding to a 1.9% FDR at the peptide level, calculated using a target-decoy approach (Elias and Gygi 2007). Once filtered, peptide redundancies were summed. Potential outliers were calculated

using a robust Mahalanobis distance, calculated by using multiple metrics such as the correlation of samples to other samples in the same protocol, skewness, and MAD (median absolute distance) of the molecule abundance profile, and the proportion of missing values. Data were then normalized using mean centering with a group specific back transformation. Peptide-level data were rolled up to protein-level using a summation method. Reverse hits and contaminants were then removed from the dataset (Matzke, Waters et al. 2011).

For quantitative analysis, ANOVA was employed to calculate fold changes and corresponding p-values between the groups of interest. For qualitative presence/absence data, a G-test was used to assess significant differences in protein detection. Proteins were considered statistically significant if they exhibited a p-value < 0.05 and a fold change (FC) > 1.5.

MTPP1 probe treated mock-infected and infected cells were successfully labeled with probe and over 350 proteins were enriched via streptavadin beads (Table 1). 394 proteins enriched specifically in HCoV-229E infected samples while 382 proteins were detected in MTPP1 probe treated mock-infected cells but only 187 of these proteins were detected under both conditions. These results suggest that HCoV-229E infection alters possible MTPP1 binding sites.

Table 1: Proteins enriched in mock-infected vs HCoV-229E infected affinity-based probe treated cells.

Condition	*Number of proteins enriched*
Probe enriched in mock-infected cells	382
Probe enriched in HCoV-229E infected cells	394
Probe enriched and detected in both	187

We performed functional enrichment on the proteins identified from HCoV-229E infected (Figure 2) experimental groups to identify the top cellular pathways affected by the MTPP1 probe. Enriched pathways of note in the proteins from HCoV-229E infected cells include: ‘infectious disease pathway’, ‘pathway pertinent to protein folding sorting modification translocation and degradation’, and ‘pathway pertinent to DNA replication and repair cell cycle maintenance of genomic integrity RNA and protein biosynthesis’. Future studies will explore and verify the roles of individual proteins within these pathways to determine their contributions to HCoV-229E replication.

2.1.2 Biochemical characterization of coronavirus infection induced methyltransferase and deacetylase activity.

Previous studies have demonstrated that human coronaviruses alter the epigenetic landscape of infected cells causing the opening and condensing of chromatin to optimize host transcription for viral replication (Menachery, Eisfeld et al. 2014, Poppe, Wittig et al. 2017). To further characterize which epigenetic post translational

modifications on histone 3 or DNA are most active during HCoV-229E infection, a variety of biochemical activity kits were tested with lysate from mock-infected and HCoV-229E infected cells. Control samples (positive and negative controls) to verify kit functionality were tested side by side with 10 μg protein per well from human lung cell nuclear extracts (Abcam extraction kit ab113474) that were treated with saline or infected with HCoV-229E for 24 hours. Biochemical kits measuring histone 3 lysine 27 (H3K27) methyltransferase activity (Abcam ab113454- closed chromatin), H3K4 methyltransferase activity (ab11352- open chromatin), and EpiQuik DNA methyltransferase Activity ELISA (P-3139) all according to manufacturer's recommendations were evaluated. For all kits that were assessed the control reactions worked as expected but no methyltransferase or deacetylase activity was detected in mock-infected or HCoV-229E infected nuclear lysates. Assessment was performed on frozen and freshly harvested nuclear lysates with similar results.

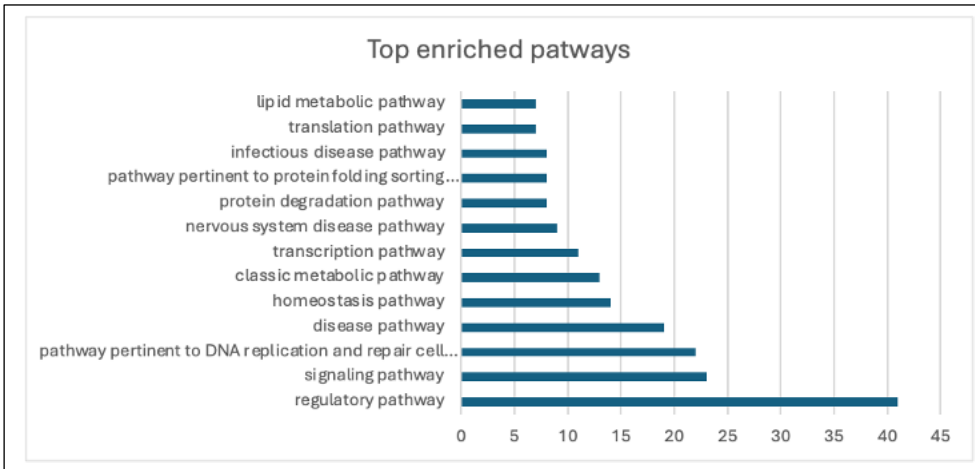

Figure 2: Proteins enriched in HCoV-229E infected cells. Pathways are significantly enriched in proteins from infected cells using Fisher's exact test for over-representation analysis.

2.2 Modifying high throughput sample collection platform for secretome/excreteome capture from mammalian cells

SubTap is a novel 3D printed culturing platform (Birer-Williams, Chu et al. 2021) developed and engineered to resemble a 96 well plate to provide replicable and quick detection of excreted/secreted metabolites and proteins via direct infusion mass spectrometry analysis. Initial development and subsequent studies have used this technology to explore the secretome from strains of bacteria (Birer-Williams, Chu et al. 2021). Our studies sought to determine the metabolites and proteins excreted and secreted from healthy and virally infected human lung cells. The SubTap plates were printed in several different configurations, where mono- or co-cultures could be in direct contact (1 compartment) or be physically separated from each other (e.g., 2, 3, or 4 compartments). The multiple compartment configurations were designed to study how the secretome/exometabolome changes if healthy cells and infected cells can exchange chemical signals without being in direct contact. If successful, the SubTap platform would

allow for high throughput and scalable exometabolome and secretome screening of healthy and infected human cells in a platform that allows investigation of extra cellular signaling as well as repeated sampling of the same wells for extended time courses.

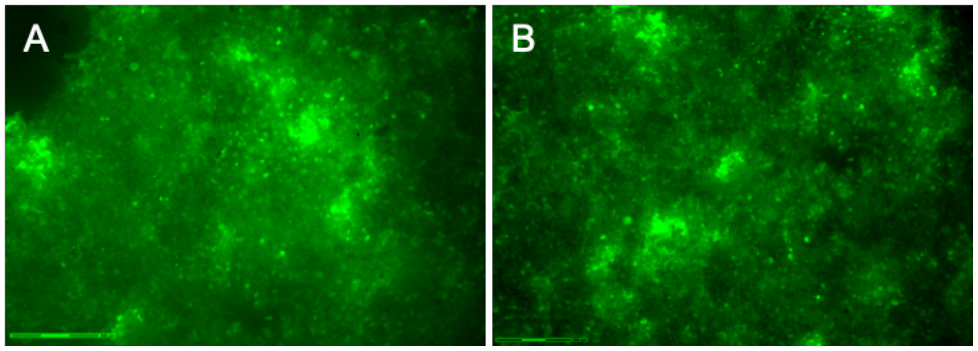

Figure 3: HuH7 cells plated in modified SubTap on lysine coated membranes. A-B. HuH7 cells were stained with a living cell dye and images taken at 5 to 7 days post plating. The scale bar at the bottom lefthand corner of each image represents 300 μm.

After printing, the SubTap plates were sterilized prior to use for mammalian tissue culture. All components of the plate and membrane were plasma cleaned and exposed to ultraviolet light for 3 hours and stored at room temperature prior to use. SubTap capture (bottom) and top plates plus membrane were again treated with UV light immediately prior to use. All subsequent manipulation steps took place inside a biological safety cabinet. For the top plate (prior to assembly with the capture plate), each well/membrane was treated with 100 μL 0.01% sterile poly-L-lysine solution (Millipore Sigma) for 5 minutes at room temperature, washed once with sterile tissue culture grade water (Corning tissue culture grade sterile water), and the membrane allowed to dry at room temperature for at least 2 hours inside the biological safety cabinet. To capture secreted metabolites from mammalian cells, the wells in the capture plate (lower plate half) were filled with 20 μL 0.5% low melting point agarose solution (Corning tissue culture grade sterile water; SeaKem low melt agarose Lonza) and the solution allowed to solidify at room temperature before assembling with the top plate. For each experiment, HuH7 cells were harvested, counted, and 15,000 cells per well added 24 hours prior to infection. The day after plating, media was removed and 100 μL HCoV-229E stock diluted in 1X sterile PBS inoculum (multiplicity of infection MOI 2, or 2 infectious viral particles per cell) was added to the appropriate wells. Control wells without cells were treated with media alone and the remaining wells were mock-infected, and the plate was incubated at 34°C and 5% CO_2 for 24 hours. At 24 hours post infection, media was removed and the capture and top plates were gently pried apart. The capture plate agarose was allowed to dry overnight in the biological safety cabinet and then was stored at -80 °C prior to analysis for metabolites species.

Several rounds of plates were evaluated by plating cells only, allowing them to grow to confluence, staining the cells with a fluorescent dye and examining the cells by microscopy (Figure 3). Clumps of living cells (brighter green) were detected on the membrane for several days following plating suggesting that it was possible to culture mammalian cells using SubTap. Two sets of plates were infected with HCoV-229E and

samples collected in the capture plate (as described above). However, there were issues with media leaking between the wells and attempts at performing metabolomics from the capture plate did not result in consistent data preventing downstream analysis.

2.3 ATAC-Seq analysis of HCoV-229E induced changes in chromatin accessibility

ATAC-Seq Basics. To optimize the intracellular environment for replication, upon entry viruses need to commandeer host transcription and protein synthesis to block host innate immune responses and ensure that the proteins, lipids and metabolites required for successful infection are present. The expression of transcripts from genes is regulated by a variety of mechanisms but one of the most important is the accessibility of a gene's promoter for transcription factor binding. Accessibility is regulated by chemical modifications of genomic DNA and/or the histone proteins that protect genomic DNA. Post translational modifications of the histone proteins can lead to an "open" histone configuration that allow transcription to take place (euchromatin) or a "closed" confirmation (heterochromatin) where the genes in that section of chromatin are not accessible for transcription. Viruses have been shown to employ a variety of mechanisms to alter chromatin accessibility including encoding histone mimics and methyltransferases (Tarakhovsky and Prinjha 2018, Sutto-Ortiz, Tcherniuk et al. 2021, Kee, Thudium et al. 2022, Ramdhan and Li 2022). As technologies improve, we are learning more about how viruses manipulate gene expression in host cells. To explore chromatin accessibility for human coronavirus infection, we employed a new technology called ATAC-Seq. This technology takes advantage of an engineered hyperactive version of the enzyme Tn5 transposases that has been "preloaded" with sequencing adaptors to identify regions of the chromosome that are open or accessible for transcription (Grandi, Modi et al. 2022). This technology allows the user to identify all of the accessible chromatin regions following their perturbation of choice and requires no specific information about gene specific promoter or enhancer sites. The details provided here focus on the workflow our group established for analyzing bulk ATAC-Seq data.

Construction and testing of bulk ATAC-Seq analysis pipeline. We developed an analysis pipeline for transposase-accessible chromatin sequencing (ATAC-Seq) data derived from bulk samples, which brought together publicly available R packages in addition to command-line tools designed for analysis of bulk ATAC-Seq data and can be run on any computer running a Linux-like operating system such as Ubuntu or Apple OSX (Figure 4). Of note, more powerful computing resources were necessary for larger input sequence read files (> 1 GB), as use of the pipeline required both increased memory/processing speed and parallel processing abilities. We tested this pipeline (Figure 4) using selected read files from two publicly available datasets (Corces, Trevino et al. 2017, Xu, Wen et al. 2021)and generated various output streams suitable for downstream biological analysis, including: 1) quality metrics with mapping statistics across chromosomes and distribution of sequence library insert sizes across nucleosome states, 2) annotation of gene regions comprising open chromatin regions based on known transcription start sites, 3) profiles of open chromatin regions by

sample across the entire chromosome, 4) gene ontology enrichment of identified open chromatin regions for cellular components, biological processes, and molecular function, and 5) differential analysis of annotated gene regions between pairs of sample groups.

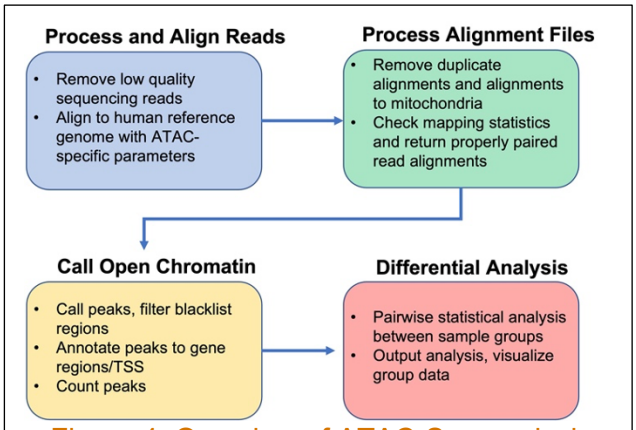

Figure 4: Overview of ATAC-Seq analysis pipeline

Analysis of test data and experimental design. The total number of reads mapping to each identified open chromatin region ("peaks") for all test samples were used as input for a power analysis to inform our experimental design. Power analyses were based on a two-sample t-test applied to log2-transformed data. A type 1 error of 5% was assumed, and the variance levels used to generate each power/sample size curve in Figure 5 were based on median (standard deviation (sd) = 0.751), third quartile (sd = 0.977), and maximum (sd = 2.846) values observed across chromatin regions in the test data. Use of these higher quartiles for estimates of data variability provided a conservative estimate for the number of samples required to achieve a desired power level. Power analysis indicated that these data allow the identification of low-level (two-fold) changes in peak abundance between treatments using a reasonable number of experimental replicates, in the case of our test data, as few as 3-4 replicates (Figure 5).

To determine how viral infection alters host cell chromatin accessibility, immortalized human cells were treated with media alone, ultra-violet light (UV) inactivated viral particles or replication competent human coronavirus 229E (HCoV-229E) for 24 hours prior to ATAC-Seq analysis. For each condition, ten replicates were harvested to ensure robustness against sample loss during sample processing and data analysis. Our experimental setup comprised three sample categories: 1) mock-infected HuH7 cells, 2) HuH7 cells treated with UV-inactivated HCoV-229E (to determine whether viral replication is required to elicit host response phenotypes in the presence of inactivated virions), and 3) HuH7 cells infected with HCoV-229E.

Potential batch effects were mitigated in our experimental design by randomizing the distribution of replicate samples for each treatment group prior to sample barcoding (four sense and four antisense primers were used) and assignment to one of four Illumina HiSeq lanes, such that every combination of primers and sequencing lane contained similar numbers of each sample type.

Sample preparation and library preparation. Samples harvested from mock-infected, UV-inactivated virus treated, and HCoV-229E infected cells were harvested using Active Motif's Bulk ATAC-Seq kit to generate bar-coded libraries for RNA Seq (according to manufacturer's recommendations). Initial quality control of the sequencing libraries

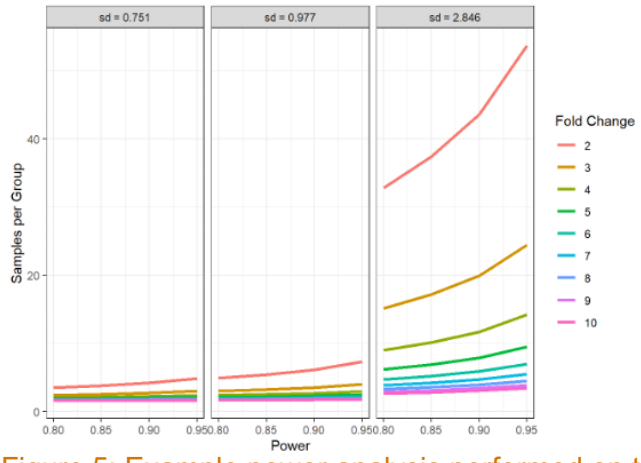

Figure 5: Example power analysis performed on test data, modeling three different levels of data variance.

revealed that our libraries had an average fragment size greater than the expected size distribution. Ideally, the fragment size distribution for an ATAC-Seq sequencing library peaks at around 150-300 bases, with a rapidly decreasing tail to around 800 bases. However, our libraries had size distributions peaking between 650-800 bases (Figure 6). We hypothesized that the most likely cause of the abnormal distribution was due to under-tagmentation during sample preparation. Tagmentation is the process by which hyperactive Tn5 transposase fragments bind sequencing adapters to regions of open chromatin, and under-tagmentation can result from several factors including insufficient transposase-chromatin incubation time or excess cell death during sample preparation. To overcome this unanticipated challenge, a size selection step was performed to enrich for appropriately sized libraries prior to sequencing. The end goal was to obtain 50 million paired-end reads per sample.

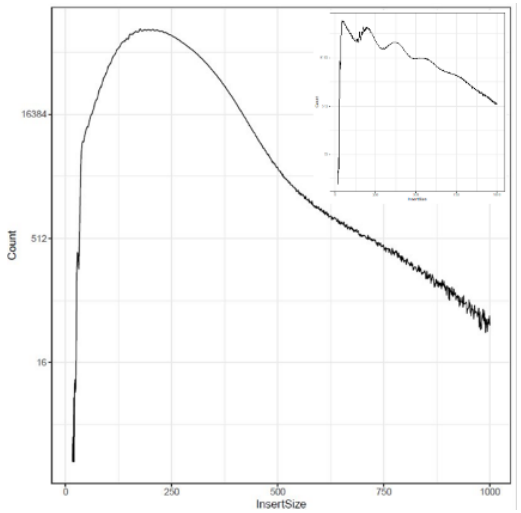

Figure 6: Visualization of fragment size distribution (one sample from our dataset). Inset, ideal size distribution from test showing peaks indicative of open chromatin, followed by mononucleosome-bound chromatin, dinucleosome, etc.

After sequencing, several samples with either abnormally high or low numbers of sequencing reads were identified and removed from further analysis. They included one mock-infected sample, one UV-treated virus sample, and four samples that were infected with HCoV-229E. Despite removal of these samples, our initial power analyses supported continuing with the analysis as the remaining number of replicates per group were sufficient to ensure adequately powered (>80%) analyses. To process the substantially higher number of reads in our experimental data, the pipeline was ported to use HPC resources, to allow sufficient memory, parallel processing, and storage capabilities, reducing the time required for sample analysis.

Initial analysis of experimental data quality. Several informative metrics of data quality were generated during the initial steps of our ATAC-Seq data processing pipeline. Prior to mapping reads to reference sequence, the distribution of fragment sizes is plotted. Ideally, fragment distribution follows the pattern show in the inset of Figure 6 (derived from test data (Corces, Trevino et al. 2017)), with pronounced peaks corresponding to nucleosome-free regions, followed by mononucleosome, dinucleosome, etc. The size

distribution of our experimental library fragments was not ideal, with most fragments corresponding to mononucleosome-bound chromatin. Nevertheless, we observed enough fragments corresponding to open chromatin (<150 bases) to proceed (Figure 6).

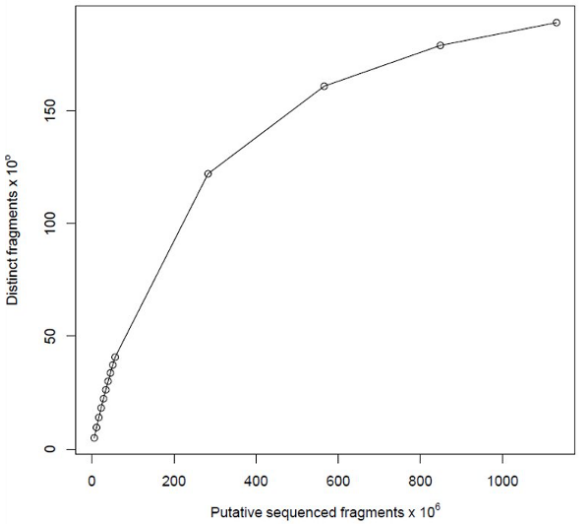

Figure 7: Plot of example library complexity

After mapping to the human reference chromosome GRCh38 using Bowtie2(Langmead and Salzberg 2012), four additional quality metrics were generated: library complexity, mapping distribution, fraction of reads in peaks, and distribution of peak annotation. Library complexity is determined as a function of distinct fragments relative to the total number of sequenced fragments. Our experimental data show an ideal complexity curve, with an early steep slope tapering off as additional fragments are added (Figure 7).

Distribution of read mapping across 24 chromosomes was similar to test data (Figure 8). Importantly, reads mapping to the mitochondrial chromosome were excluded from downstream analysis, as the nature of this chromosome makes it hyper-available to the Tn5 transposase during tagmentation and therefore not informative to ATAC-Seq. The final quality metric produced by our analysis pipeline is the Fraction of Reads in Peaks (FRIP), which quantified the number of reads that map to regions of open chromatin after ATAC peaks were identified. An ideal FRIP is above 0.3, with usable but suboptimal FRIP ranges from 0.1-0.29. Our data averaged a FRIP of 0.16, which was lower than ideal and likely related to the fragment size distribution discussed above. This metric is anticipated to improve as sample collection procedures are optimized.

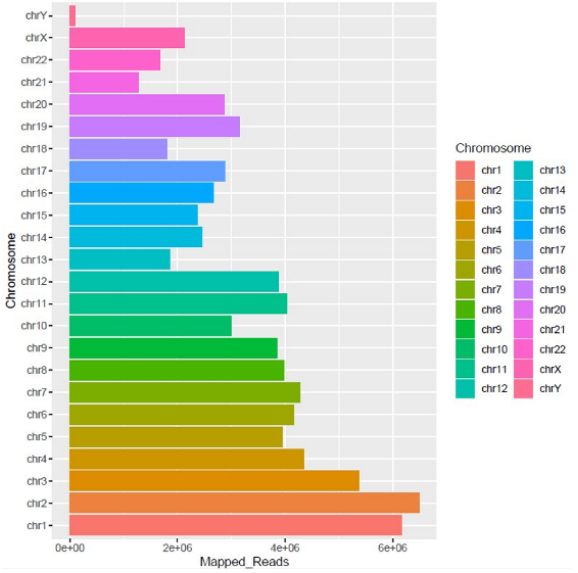

Figure 8. Read mapping distribution across all human chromosomes. Reads mapping to the mitochondrial chromosome were excluded as they were not informative to ATAC-Seq analysis.

ATAC-Seq peaks identified by our pipeline (Zhang, Liu et al. 2008) were annotated based on their proximity to known transcription start sites (TSS) within the human genome, the distribution of which is shown in Figure 9. In typical ATAC-Seq data, around 25% of peaks map to promoter regions less than 1000 nucleotides from TSS and around 50% map to intronic and distal intergenic regions. Our study, peaks mapping to promoter regions accounted for 31.5%, 30.5%, and 35.5% of total peaks in mock infected, UV-treated, and infected samples, respectively, while peaks mapping to intronic and distal intergenic regions

accounted for 55.6%, 56.5%, and 52.2% in the same samples. The annotation distributions are approximately equal across sample treatments, increasing confidence in the comparability of data between these sample types.

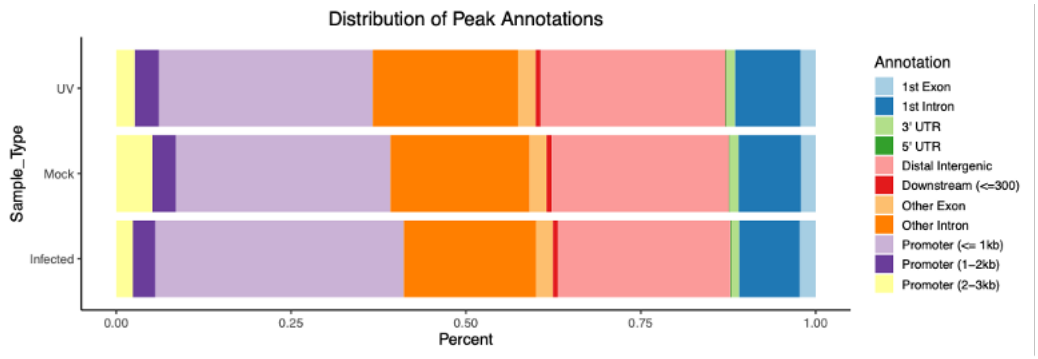

Figure 9. Distribution of ATAC-Seq peak annotations relative to transcription start sites (TSS)

Differential analysis of ATAC peaks. Three comparisons of peak abundance were made between sample treatments using DESeq2, which normalizes peak data to total read count 1) mock-infected vs. HCoV-229E infected, 2) UV-inactivated virus treatment vs. HCoV-229E infected, and 3) mock-infected vs. UV-inactivated virus treatment.

Comparisons were initially made at the peak level, with potentially multiple peaks corresponding to a given coding region/gene, because ATAC-Seq ultimately gives a measure of differentially accessible chromatin, not necessarily differential gene expression. Principal component analysis was conducted using data from these comparisons. This analysis demonstrated a clear global separation between HCoV-229E infected cells and both UV-inactivated virus treated and mock-infected cells (Figure 10). Additionally, it showed a lack of separation between mock-infected and UV-inactivated virus treated cells indicating a requirement for viral replication for eliciting the phenotypes observed in infected cells.

Chromatin accessibility differential detection analysis was reduced in scope to comparing HCoV-229E infected samples to either mock-infected or UV-inactivated virus treated samples. Results from these comparisons are summarized in Table 2 and visualized as volcano plots in Figure 11.

Table 2: Summary of ATAC-Seq peak differential analysis

Comparison	*Total number of quantified peaks*	*Total number of significant peaks*)	*Significant peaks +- 2-fold change*	*Peaks with increasing abundance*	*Peaks with decreased abundance*
Mock vs. UV	114,207	0	0	0	0
Mock vs. infected	114,207	15,186	6,089	3,439	2,650
UV vs. infected	114,207	12,503	4,997	1,982	3,015

When analyzing differential detection, we chose to impose an additional cutoff of ±2-fold change in peak abundance to simplify exploration of the data and to adhere to the power analysis performed on test data. For our sample set, between-sample variance was lower than in the test data, suggesting we can detect smaller effect sizes at a reasonable power, however in our analysis the more conservative cutoff was used. We compared the differential peak calls from the two conditions (mock-infected vs. infection, UV-inactivated virus treatment vs. HCoV-229E infection) to determine overlap between the two (Figure 12) sets of conditions. We observed a significant overlap in peak abundance between the two sets of treatments, with 67% of differential peak calls being shared. Encouragingly, differentially abundant peaks identified in HCoV-229E infected cells common to both UV-inactivated virus treated and mock-infected comparisons exhibited the same change, e.g., they were both increased or both decreased. There were no instances of peak abundance changes in the opposite direction between the two conditions.

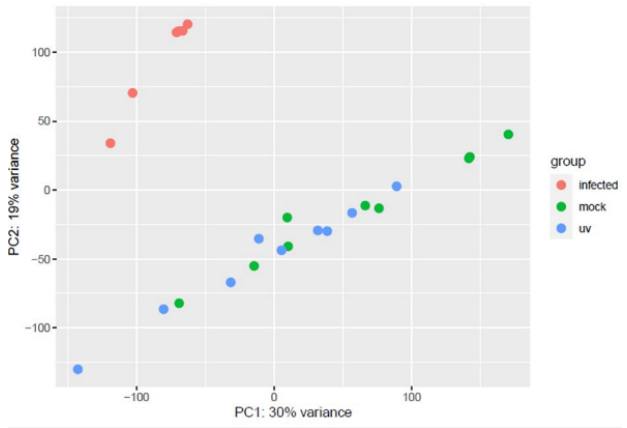

Figure 10. Principal coordinate analysis plot of differentially abundant ATAC-Seq peaks.

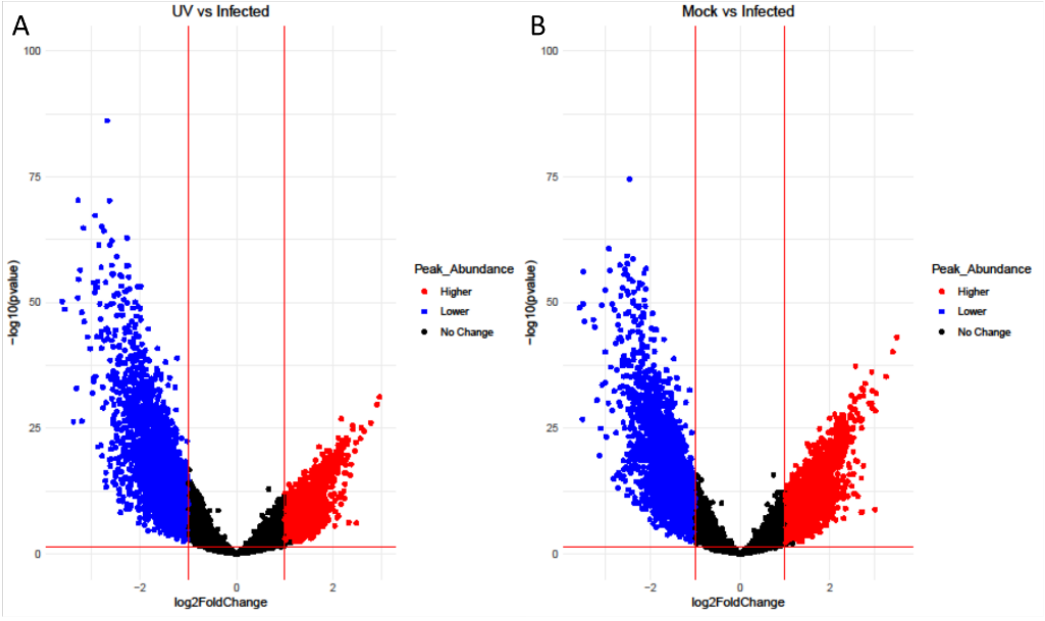

Figure 11. Volcano plots of differentially abundant ATAC peaks for UV vs. HCoV-229E infected (A) and mock-infected vs HCoV-229E infected (B). Vertical lines represent 2-fold higher or lower abundance. Horizontal lines indicate p-value cutoff of 0.05.

To assess the biological relevance of differential chromatin accessibility, we subjected significantly different peaks to pathway analysis using the R package PathfindR(Ulgen, Ozisik et al. 2019), which uses a clustering algorithm to detect pathway enrichment. Prior to pathway analysis, peaks were annotated to the nearest gene using known TSS locations. PathfindR filters redundant gene calls from analysis, leaving only data from the entry with the lowest p-value for a given gene. This analysis identified a total of 206 enriched pathway terms, with 179 terms shared between UV-inactivated virus treatment and mock-infected comparisons (Figure 13). The top 20 enriched pathways for each condition are visualized in Figure 14. Genes comprising enriched pathways were visualized within PathfindR by mapping them to the Kyoto Encyclopedia of Genes and Genomes (KEGG; (Kanehisa and Goto 2000, Kanehisa 2019, Kanehisa, Furumichi et al. 2021)) database. The top hit for both sets of comparisons were for adherens, Adherens junctions are protein complexes present at the interfaces between cells, and are linked to the actin cytoskeleton (Meng and Takeichi 2009). In HCoV-229E infected cells the chromatin regulating adherens junction gene transcription is not accessible (or closed) suggesting that infection reduces cellular adhesion.

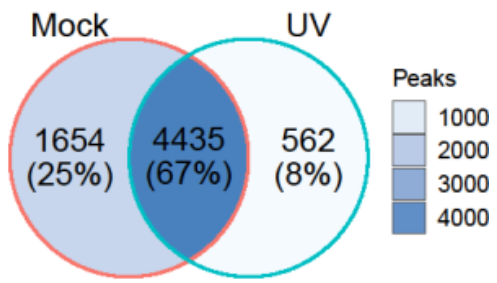

Figure 12. Overlap of differentially abundant ATAC-Seq peak calls in HCoV-229E infected samples relative to mock-infected and UV-treated samples.

Next steps. Evaluating single-cell ATAC-Seq analysis software: In addition to analysis of bulk ATAC-Seq data, we began exploring analysis software for single-cell ATAC-Seq experimental data. Analysis of single-cell data is nontrivial and requires significant computational resources, such as the HPC infrastructure at PNNL. We have identified software purpose-built for processing and alignment of single-cell ATAC sequence data (Cell Ranger 10x Genomics) and have successfully processed ~10 publicly available test datasets of varying sizes using HPC resources. We also tested two software packages designed for analysis of single cell ATAC-Seq data. Loupe (10x Genomics) was designed to work seamlessly with Cell Ranger output as a graphical user interface, while ArchR, is an R package that allows users a high degree of customization and granularity for data analysis but requires some knowledge of the R language. We will use Cell Ranger,

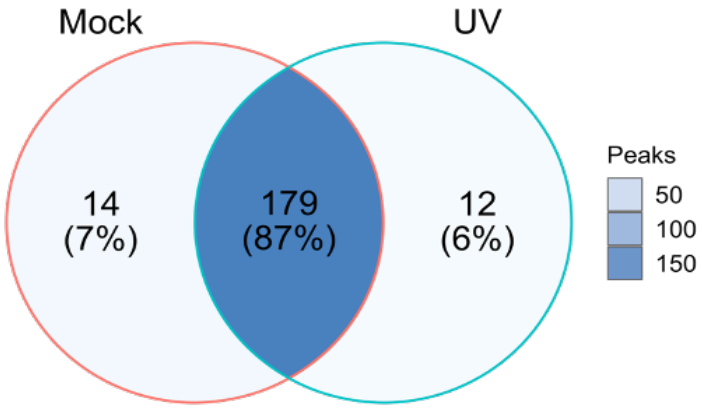

Figure 13. Overlap of enriched pathway terms between mock-infected samples and UV-treated samples.

Loupe, and/or ArchR for future studies with single-cell ATAC-Seq sequencing datasets. These software packages are also applicable to other experimental techniques which measure relative abundance of nucleic acids, such as single cell RNA Seq or 3' expression libraries.

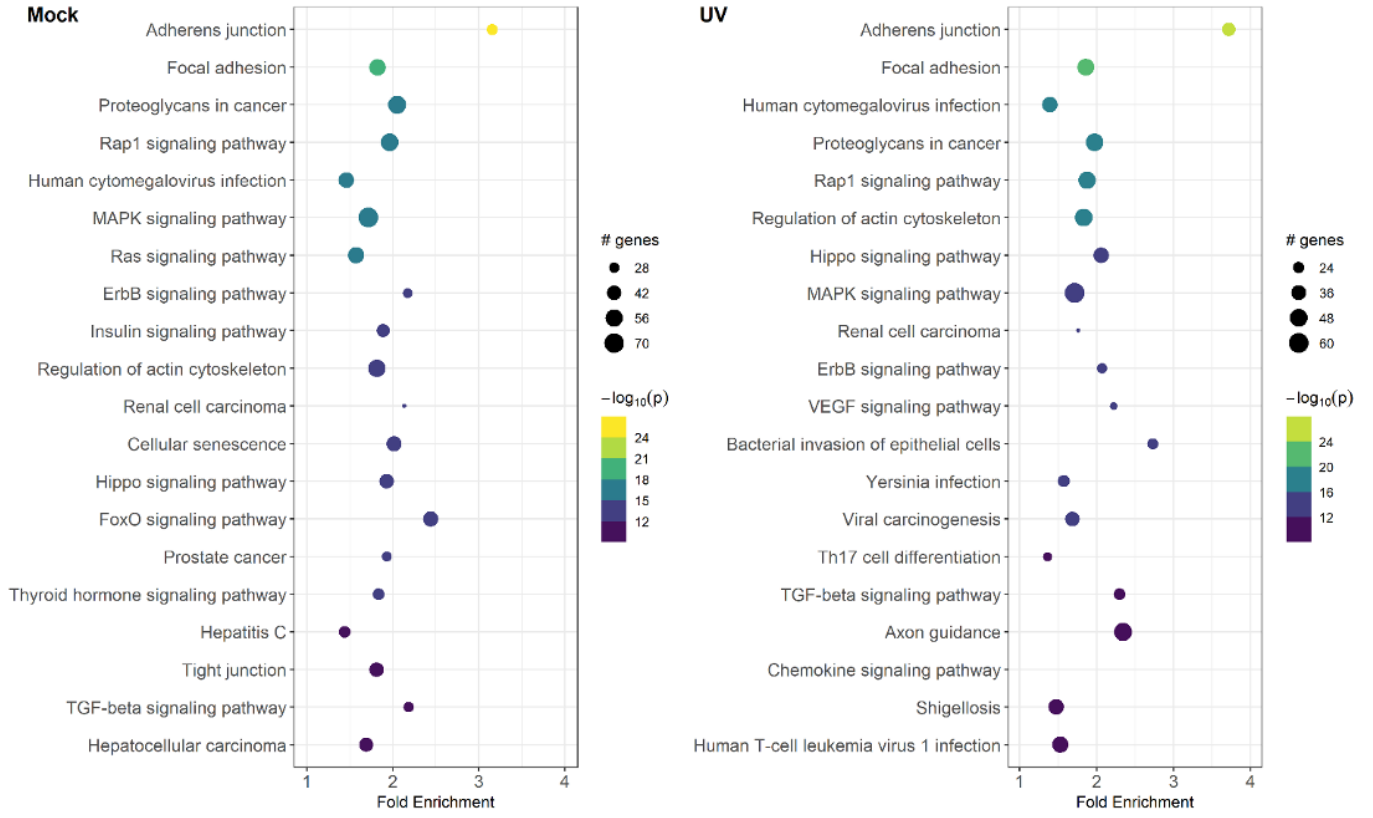

Figure 14. Top 20 enriched pathway terms for HCoV-229E ATAC-Seq data compared to mock-infected samples (A) and UV-treated samples (B).

3.0 References

Birer-Williams, C. M. C., R. K. Chu, C. R. Anderton and E. S. Wright (2021). "SubTap, a Versatile 3D Printed Platform for Eavesdropping on Extracellular Interactions." mSystems **6**(4): e0090221.

Corces, M. R., A. E. Trevino, E. G. Hamilton, P. G. Greenside, N. A. Sinnott-Armstrong, S. Vesuna, A. T. Satpathy, A. J. Rubin, K. S. Montine, B. Wu, A. Kathiria, S. W. Cho, M. R. Mumbach, A. C. Carter, M. Kasowski, L. A. Orloff, V. I. Risca, A. Kundaje, P. A. Khavari, T. J. Montine, W. J. Greenleaf and H. Y. Chang (2017). "An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues." Nat Methods **14**(10): 959-962.

Elias, J. E. and S. P. Gygi (2007). "Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry." Nat Methods **4**(3): 207-214.

Grandi, F. C., H. Modi, L. Kampman and M. R. Corces (2022). "Chromatin accessibility profiling by ATAC-seq." Nat Protoc **17**(6): 1518-1552.

Kanehisa, M. (2019). "Toward understanding the origin and evolution of cellular organisms." Protein Sci **28**(11): 1947-1951.

Kanehisa, M., M. Furumichi, Y. Sato, M. Ishiguro-Watanabe and M. Tanabe (2021). "KEGG: integrating viruses and cellular organisms." Nucleic Acids Res **49**(D1): D545-D551.

Kanehisa, M. and S. Goto (2000). "KEGG: kyoto encyclopedia of genes and genomes." Nucleic Acids Res **28**(1): 27-30.

Kee, J., S. Thudium, D. M. Renner, K. Glastad, K. Palozola, Z. Zhang, Y. Li, Y. Lan, J. Cesare, A. Poleshko, A. A. Kiseleva, R. Truitt, F. L. Cardenas-Diaz, X. Zhang, X. Xie, D. N. Kotton, K. D. Alysandratos, J. A. Epstein, P. Y. Shi, W. Yang, E. Morrisey, B. A. Garcia, S. L. Berger, S. R. Weiss and E. Korb (2022). "SARS-CoV-2 disrupts host epigenetic regulation via histone mimicry." Nature **610**(7931): 381-388.

Kim, S. and P. A. Pevzner (2014). "MS-GF+ makes progress towards a universal database search tool for proteomics." Nat Commun **5**: 5277.

Konze, K. D., A. Ma, F. Li, D. Barsyte-Lovejoy, T. Parton, C. J. Macnevin, F. Liu, C. Gao, X. P. Huang, E. Kuznetsova, M. Rougie, A. Jiang, S. G. Pattenden, J. L. Norris, L. I. James, B. L. Roth, P. J. Brown, S. V. Frye, C. H. Arrowsmith, K. M. Hahn, G. G. Wang, M. Vedadi and J. Jin (2013). "An orally bioavailable chemical probe of the Lysine Methyltransferases EZH2 and EZH1." ACS Chem Biol **8**(6): 1324-1334.

Langmead, B. and S. L. Salzberg (2012). "Fast gapped-read alignment with Bowtie 2." Nat Methods **9**(4): 357-359.

Margueron, R. and D. Reinberg (2011). "The Polycomb complex PRC2 and its mark in life." Nature **469**(7330): 343-349.

Matzke, M. M., K. M. Waters, T. O. Metz, J. M. Jacobs, A. C. Sims, R. S. Baric, J. G. Pounds and B. J. Webb-Robertson (2011). "Improved quality control processing of peptide-centric LC-MS proteomics data." Bioinformatics **27**(20): 2866-2872.

McCabe, M. T., H. M. Ott, G. Ganji, S. Korenchuk, C. Thompson, G. S. Van Aller, Y. Liu, A. P. Graves, A. Della Pietra, 3rd, E. Diaz, L. V. LaFrance, M. Mellinger, C. Duquenne, X. Tian, R. G. Kruger, C. F. McHugh, M. Brandt, W. H. Miller, D. Dhanak, S. K. Verma, P. J. Tummino and C. L. Creasy (2012). "EZH2 inhibition as a therapeutic strategy for lymphoma with EZH2-activating mutations." Nature **492**(7427): 108-112.

Menachery, V. D., A. J. Eisfeld, A. Schafer, L. Josset, A. C. Sims, S. Proll, S. Fan, C. Li, G. Neumann, S. C. Tilton, J. Chang, L. E. Gralinski, C. Long, R. Green, C. M. Williams, J. Weiss, M. M. Matzke, B. J. Webb-Robertson, A. A. Schepmoes, A. K. Shukla, T. O. Metz, R. D. Smith, K. M. Waters, M. G. Katze, Y. Kawaoka and R. S. Baric (2014). "Pathogenic influenza viruses and coronaviruses utilize similar and contrasting

approaches to control interferon-stimulated gene responses." mBio **5**(3): e01174-01114.
Meng, W. and M. Takeichi (2009). "Adherens junction: molecular architecture and regulation." Cold Spring Harb Perspect Biol **1**(6): a002899.
Monroe, M. E., J. L. Shaw, D. S. Daly, J. N. Adkins and R. D. Smith (2008). "MASIC: a software program for fast quantitation and flexible visualization of chromatographic profiles from detected LC-MS(/MS) features." Comput Biol Chem **32**(3): 215-217.
Poppe, M., S. Wittig, L. Jurida, M. Bartkuhn, J. Wilhelm, H. Muller, K. Beuerlein, N. Karl, S. Bhuju, J. Ziebuhr, M. L. Schmitz and M. Kracht (2017). "The NF-kappaB-dependent and -independent transcriptome and chromatin landscapes of human coronavirus 229E-infected cells." PLoS Pathog **13**(3): e1006286.
Ramdhan, P. and C. Li (2022). "Targeting Viral Methyltransferases: An Approach to Antiviral Treatment for ssRNA Viruses." Viruses **14**(2).
Sims, A. C., S. C. Tilton, V. D. Menachery, L. E. Gralinski, A. Schafer, M. M. Matzke, B. J. Webb-Robertson, J. Chang, M. L. Luna, C. E. Long, A. K. Shukla, A. R. Bankhead, 3rd, S. E. Burkett, G. Zornetzer, C. T. Tseng, T. O. Metz, R. Pickles, S. McWeeney, R. D. Smith, M. G. Katze, K. M. Waters and R. S. Baric (2013). "Release of severe acute respiratory syndrome coronavirus nuclear import block enhances host transcription in human lung cells." J Virol **87**(7): 3885-3902.
Smith, E. and I. Collins (2015). "Photoaffinity labeling in target- and binding-site identification." Future Med Chem **7**(2): 159-183.
Sutto-Ortiz, P., S. Tcherniuk, N. Ysebaert, P. Abeywickrema, M. Noel, A. Decombe, F. Debart, J. J. Vasseur, B. Canard, D. Roymans, P. Rigaux, J. F. Eleouet and E. Decroly (2021). "The methyltransferase domain of the Respiratory Syncytial Virus L protein catalyzes cap N7 and 2'-O-methylation." PLoS Pathog **17**(5): e1009562.
Tarakhovsky, A. and R. K. Prinjha (2018). "Drawing on disorder: How viruses use histone mimicry to their advantage." J Exp Med **215**(7): 1777-1787.
Ulgen, E., O. Ozisik and O. U. Sezerman (2019). "pathfindR: An R Package for Comprehensive Identification of Enriched Pathways in Omics Data Through Active Subnetworks." Front Genet **10**: 858.
Xu, W., Y. Wen, Y. Liang, Q. Xu, X. Wang, W. Jin and X. Chen (2021). "A plate-based single-cell ATAC-seq workflow for fast and robust profiling of chromatin accessibility." Nat Protoc **16**(8): 4084-4107.
Yang, X., F. Li, K. D. Konze, J. Meslamani, A. Ma, P. J. Brown, M. M. Zhou, C. H. Arrowsmith, H. U. Kaniskan, M. Vedadi and J. Jin (2016). "Structure-Activity Relationship Studies for Enhancer of Zeste Homologue 2 (EZH2) and Enhancer of Zeste Homologue 1 (EZH1) Inhibitors." J Med Chem **59**(16): 7617-7633.
Zhang, Y., T. Liu, C. A. Meyer, J. Eeckhoute, D. S. Johnson, B. E. Bernstein, C. Nusbaum, R. M. Myers, M. Brown, W. Li and X. S. Liu (2008). "Model-based analysis of ChIP-Seq (MACS)." Genome Biol **9**(9): R137.

Pacific Northwest National Laboratory

902 Battelle Boulevard
P.O. Box 999
Richland, WA 99354

1-888-375-PNNL (7665)

www.pnnl.gov