Fixed single-cell RNA sequencing for understanding virus infection and host response

Hoang Van Phan,[1,2*] Michiel van Gent,[3,4*] Nir Drayman,[1,2*] Anindita Basu,[5] Michaela U. Gack,[3,4] Savaş Tay[1,2]

[1]Pritzker School of Molecular Engineering, The University of Chicago, Chicago, IL, USA

[2]Institute for Genomics and Systems Biology, The University of Chicago, Chicago, IL, USA

[3]Department of Microbiology, The University of Chicago, Chicago, IL, USA

[4]Florida Research and Innovation Center, Cleveland Clinic, Port Saint Lucie, FL 34987

[5]Department of Medicine, The University of Chicago, Chicago, IL, USA

*These authors contributed equally

Corresponding author: tays@uchicago.edu

ABSTRACT

Single-cell RNA sequencing studies requiring intracellular protein staining, rare-cell sorting, or pathogen inactivation are severely limited because current high-throughput methods are incompatible with paraformaldehyde treatment, a very common and simple tissue/cell fixation and preservation technique. Here we present FD-seq, a high-throughput method for droplet-based RNA sequencing of paraformaldehyde-fixed, stained and sorted single-cells. We used FD-seq to address two important questions in virology. First, by analyzing a rare population of cells supporting lytic reactivation of the human tumor virus KSHV, we identified *TMEM119* as a host factor that mediates reactivation. Second, we studied the transcriptome of lung cells infected with the

coronavirus OC43, which causes the common cold and also serves as a safer model pathogen for SARS-CoV-2. We found that pro-inflammatory pathways are primarily upregulated in abortively-infected or uninfected bystander cells, which are exposed to the virus but fail to express high level of viral genes. FD-seq is suitable for characterizing rare cell populations of interest, for studying high-containment biological samples after inactivation, and for integrating intracellular phenotypic with transcriptomic information.

Introduction

Single-cell RNA sequencing (scRNA-seq) has found many important biological applications, from discovery of new cell types[1] to mapping the transcriptional landscape of human embryonic stem cells[2]. Droplet-based scRNA-seq technologies, such as Drop-seq[3] and 10X Chromium[4], are particularly powerful due to their high throughput: thousands of single cells can be analyzed in a single experiment. However, even with these high-throughput techniques, analyzing rare cell populations remains a challenging task, often requiring protein-based enrichment for the cell population of interest before scRNA-seq[5]. Some cell types can be enriched using cell surface markers and fluorescent-activated cell sorting (FACS), while others require intracellular protein staining instead. For example, Foxp3 is an intracellular marker of regulatory T cells[6], and Oct4 and Nanog are intracellular reprogramming markers of induced pluripotent stem cells[7].

Intracellular protein staining requires cell fixation, which is usually achieved with methanol or paraformaldehyde (PFA) fixation. Drop-seq and 10X have been shown to be compatible with methanol-fixed cells[8,9]. In many intracellular staining applications, however, PFA fixation is necessary to increase the signal-to-background ratio[9–11]. PFA fixation is also a common method for cell and tissue preservation. PFA fixation presents more challenges for RNA sequencing than alcohol-based fixation because the nucleic acids are chemically cross-linked to

the intracellular proteins[5,9]. In order to retrieve high quality RNA from single PFA-fixed cells, an appropriate cross-link reversal protocol that maintains RNA integrity and minimizes RNA loss is crucial. Here we describe FD-seq (Fixed Droplet RNA sequencing), a method based on Drop-seq for RNA sequencing of PFA-fixed, stained and sorted single cells. We show that the relative RNA expression profile of fixed cells obtained by FD-seq is similar to that of live cells obtained by Drop-seq. FD-seq can also detect unspliced mRNA, allowing for advanced data analysis methods such as RNA velocity[12] in fixed single cells.

We used our method to study two important problems in virology. First, we investigated the host factors that influence Kaposi's sarcoma-associated herpesvirus (KSHV) reactivation in tumor cells. KSHV, also known as human herpesvirus type 8 (HHV-8), is a human gammaherpesvirus that causes a number of malignancies such as Kaposi's sarcoma, primary effusion lymphoma and multicentric Castleman's disease[13,14]. There is a considerable interest in unraveling the molecular details of the host factors that modulate KSHV latency and reactivation, because both latency and low-level reactivation are known to contribute to viral tumorigenesis[15], and therapeutic induction of reactivation could sensitize latently-infected cells to currently available anti-herpesvirus drugs[16]. Using FD-seq, we present a single-cell transcriptomic analysis of reactivated human primary effusion lymphoma (PEL) cells. We found that in reactivated cells, the expression levels of viral genes were extremely heterogeneous. Additionally, we found four host genes, *CDH1*, *CORO1C*, *ISCU* and *TMEM119*, whose levels correlated with the state of viral reactivation. Using live cell imaging and time-course study of viral gene expression, we found that ectopic expression of *TMEM119* gene induced KSHV reactivation, whereas its silencing led to lower reactivation. Bulk RNA-seq on *TMEM119* transfected cells showed that upregulated genes were enriched in the mitogen-activated protein kinase (MAPK) signaling pathway.

We then studied the immune response of human lung cells infected by OC43 coronavirus. The current coronavirus disease 2019 (COVID-19) pandemic has resulted in more than 500,000 deaths worldwide. Studying SARS-CoV-2, the etiological agent of COVID-19, requires Biosafety level 3 (BSL-3) facilities, which are often not readily available. In addition, time from sample collection (for example from a patient in a hospital) to analysis could pose a problem for maintaining cell viability and RNA quality. To make studying this virus more convenient, and to facilitate international collaboration, researchers can inactivate the patient cells with PFA so that the samples can be handled in BSL-2 facilities, and/or the samples can be preserved and transported internationally. In this study, we also show a possible application of FD-seq in the current pandemic by studying PFA-fixed cells that had been infected with OC43, a relative of SARS-CoV-2. OC43, like SARS-CoV-2, belongs to the family of betacoronaviruses, and is a major human pathogen that causes the common cold. OC43 has been successfully used to discover drugs that inhibit SARS-CoV-2 replication *in vitro*[17]. We found that human lung cells infected by OC43 were composed of two subpopulations with either low or high viral gene expression. Interestingly, pro-inflammatory immune signaling pathways, such as IL-17 and TNF signaling pathways, were upregulated in bystander cells that showed low or no viral gene expression. This finding suggests the possibility that the excessive cytokine release observed in severe COVID-19 patients[18] could be driven by bystander cells, and that these cells could serve as a potential therapeutic target.

RESULTS AND DISCUSSION

Design and optimization of FD-seq

In Drop-seq, each single cell is partitioned with a uniquely barcoded ceramic bead inside nanoliter droplets in oil, using a microfluidic device[3]. The cells are lysed inside the droplets, and their

mRNAs are captured by the oligonucleotides on the beads. After this, the droplets are broken to recover the beads, and the beads are extensively washed to remove uncaptured mRNAs. After pooling the beads, the barcoded mRNAs undergo reverse transcription, exonuclease I digestion of free bead oligonucleotides, and whole transcriptome amplification. Finally, the barcoded and amplified cDNAs are tagmented and sequenced.

To make Drop-seq suitable for PFA-fixed cells, it is critical that the reversal of cross-linked mRNAs should have a high RNA recovery rate and avoid cross-contamination, while being compatible with all downstream molecular biology steps. Thus, to modify Drop-seq in the least obtrusive way, FD-seq introduces a separate cross-link reversal step between droplet generation and the droplet breakage step. This allows all other steps of the standard Drop-seq protocol to remain unchanged, while restricting the uncross-linked mRNAs from a given cell inside the droplet during the reversal step.

First, we tested different conditions for cross-link reversal at bulk level by lysing PFA-fixed cells with Drop-seq lysis buffer, and extracting the total RNA using a commercial kit (**Fig. S1** and **S2a**, see Methods). We found that a 1-hour incubation at 56°C was sufficient to reverse PFA cross-linking, which agreed with another study[5]. The RNA yield could be improved further by the addition of proteinase K to the lysis buffer (**Fig. S2a**). However, there was an optimal concentration of 40 units/ml of proteinase K, above which the RNA yield actually decreased (**Fig. S1a**). We consistently obtained high quality total RNA for all tested concentrations of proteinase K, demonstrated by high RNA integrity numbers (above 8.0) (**Fig. S1b,c**). (The RNA integrity number is used to quantify the degree of total RNA degradation.)

To check for contamination from RNA across droplets, we performed species-mixing experiments on fresh live cells and fixed cells. For the live cell experiment, we used Drop-seq to

analyze a 1-to-1 mixture of BC3, a human PEL cell line, and mouse 3T3 cells. For the fixed cells experiment, we fixed the BC3 and 3T3 cells with PFA and permeabilized them with Triton-X separately, and analyzed a 1-to-1 mixture of the two fixed and permeabilized cell samples with FD-seq. For both experiments, the sequencing results were aligned to a combined human-mouse reference genome. The amount of human and/or mouse transcripts per single cell barcode was then calculated. Cells having both mouse and human transcripts, which indicate cross-droplet contamination and/or high multiplet (multiple cells captured in the same droplet) rate, were only minimally observed in both live and fixed experiments (~0.5% and ~1% mixed species rate for fresh live and fixed cells, respectively, **Fig. S2b**). This suggests that the modifications of FD-seq did not affect the single-cell capture feature of Drop-seq. Furthermore, the number of genes detected in fixed cells was comparable to that of live cells across species (median of 640 genes in fixed cells compared to 675 genes in live cells for human cells, and median of 734 genes compared to 724 genes for mouse cells) (**Fig. S2c**). Importantly, the relative expression levels of the detected genes correlated well between live and fixed cells (**Fig. S2d**), even though fewer transcripts were recovered from fixed cells than from live cells on average (**Fig. S2b, Fig. S3a**). FD-seq also did not affect the percentage of reads mapped to intron and exon regions (**Fig. S3b**). Taken together, these results demonstrate that FD-seq is a reliable method for whole transcriptome analysis of PFA-fixed single cells.

FD-seq reveals heterogeneity in KSHV reactivated single tumor cells

Next, we applied FD-seq to identify host factors that regulate KSHV reactivation. Upon primary infection, KSHV establishes latency that is characterized by the expression of a very limited set of viral gene products. Lytic reactivation involves expression of the complete viral genome and the

production of new viral particles[19]. Although many viral and host factors that regulate the reactivation process of KSHV have been identified[20], the exact physiological triggers have remained elusive. Furthermore, only a small proportion of cells enter the lytic replication, even when treated with known chemical inducing agents such as sodium butyrate (NaBut) and tetradecanoyl phorbol acetate (TPA)[13]. The reason for this low-level reactivation is currently not well understood, and we hypothesized that differences in the expression of specific host factors between individual cells contribute to the propensity of the cells to enter lytic reactivation. To test this hypothesis, we focused on the PEL cell line BC3 that is latently infected with KSHV.

To identify lytically reactivated cells, we performed flow cytometry analysis of the viral lytic cell-surface glycoprotein K8.1. As expected, we did not observe any appreciable spontaneous reactivation, as evident by undetectable K8.1 expression in untreated BC3 cells (**Fig. S4a**). However, treatment with NaBut and TPA resulted in 2-8% K8.1-positive, lytically reactivated cells after 36 hours (**Fig. S4b**). Enrichment of reactivated cells was thus required to obtain a sufficiently large population of reactivated cells to obtain sufficient statistical power. Therefore, TPA-treated BC3 cells were fixed and sorted by FACS based on the expression level of K8.1 protein (**Fig. S5**), and their transcriptomes were analyzed by FD-seq. We obtained 1,035 high-quality K8.1+ single cells, as well as 286 high-quality K8.1- cells in the FACS-sorted population using FD-seq. High dimensional clustering and visualization showed clear separation between reactivated and non-reactivated cells (**Fig. 1a**). We confirmed that FD-seq indeed enriched the cell population of interest: the K8.1+ cells expressed much higher *K8.1* mRNA levels compared to the K8.1- cells (**Fig. 1b**). The K8.1+ cluster also expressed more viral transcripts (on average, approximately 69% of the total detected transcripts in the K8.1+ cluster were of viral origin, compared to 4% in the K8.1- cluster) (**Fig. 1c**).

Herpesvirus reactivation involves the highly regulated, sequential expression of immediate-early, early, and late viral genes[21]. We found that the relative abundance of immediate-early viral genes increased early, then flattened after reaching 50% total viral transcripts (**Fig. S6a**). Early viral transcripts also increased early and monotonically with the abundance of total viral transcripts, without flattening like the immediate-early genes. On the other hand, late viral transcripts were only detected in cells with more than 25% total viral transcripts, and increased strongly with higher total viral transcript abundance.

The K8.1+ population was highly heterogeneous: the relative abundance of viral transcripts varied from below 50% to over 90% (**Fig. 1d**), and the correlation between viral genes was very poor (**Fig. S6b**). This agrees with two recent studies that showed the single-cell heterogeneity in herpes simplex virus type 1 (HSV-1) infection[22], and low correlation between viral gene expression in a murine gammaherpesvirus infection model[23]. One partial explanation is that the K8.1+ cells were at different stages of reactivation. Indeed, K8.1 protein expression level of the positive population varied within one order of magnitude, as quantified by flow cytometry (**Fig. S5**). However, this alone could not sufficiently explain the lack of correlation between viral genes, even among highly expressed viral genes (**Fig. S6c,d**). In summary, we showed that FD-seq enables scRNA-seq of a rare cell population of interest after fixation and FACS-enrichment, and found that KSHV reactivation is a highly heterogeneous process at the single-cell level.

TMEM119 **facilitates KSHV reactivation**

Next, we sought to identify host genes that modulate KSHV reactivation. However, due to the large variation in the proportion of viral transcripts between cells (~4% in the K8.1- population and up to 90% in highly reactivated cells, **Fig. 1d**), reduced expression of a host gene can merely result from the lower coverage of host transcripts overall, and does not necessarily mean that it is

downregulated in the reactivated cells. Indeed, the majority of differentially expressed genes were negatively correlated with the relative abundance of viral transcripts (**Fig. S7a**). While it is known that overall transcriptional host shut-off is induced by KSHV[24], we instead primarily focused on host genes that were upregulated in the reactivated population, since their higher expression level would not be confounded by the high amount of viral transcripts.

We found four host genes whose expression strongly correlated with the degree of KSHV reactivation: *ISCU*, *CDH1*, *CORO1C* and *TMEM119* (q-value < 10^{-100} for all four genes). We observed that the expression of *ISCU* was relatively low in non-reactivated cells, and increased significantly with the abundance of viral transcripts (**Fig. 1e**). On the other hand, *CDH1*, *CORO1C* and *TMEM119* were mostly undetectable in non-reactivated cells, and increased by 1-2 orders of magnitude as reactivation progressed. We also validated the upregulation of *ISCU*, *CORO1C* and *TMEM119* in bulk cell samples by RT-qPCR (**Fig. S7b,c**).

Because the higher abundance of these transcripts in highly reactivated cells may suggest that these genes modulate KSHV reactivation efficiency, we tested the effect of their overexpression on KSHV reactivation. To this end, we exogenously expressed these genes in HEK293T.rKSHV219 cells, which are latently infected with the recombinant KSHV.219 virus strain that constitutively expresses GFP and further encodes RFP under control of the viral lytic PAN promoter[25]. Live-cell imaging showed a clear increase in RFP expression, which is indicative of viral reactivation, upon ectopic expression of *TMEM119* relative to a GFP-encoding control vector, whereas overexpression of the other three genes had no significant effect on KSHV reactivation efficiency (**Fig. 1f, Fig. S8a-c**). Conversely, silencing of endogenous *TMEM119* reduced the level of KSHV reactivation in TPA-treated cells (**Fig. 1g, Fig S8b,c**). Corroborating these results, RT-qPCR analysis of KSHV lytic gene expression showed that *TMEM119*

overexpression enhanced viral gene expression compared to the GFP control (**Fig. 1h**), while *TMEM119* silencing had the opposite effect (**Fig. 1i**).

To determine the mechanism by which *TMEM119* enhances KSHV reactivation, we performed bulk transcriptome analysis of HEK293T.rKSHV and naïve HEK293T cells following ectopic expression of *TMEM119*. We found that in HEK293T cells, 202 genes were significantly upregulated upon *TMEM119* overexpression (**Fig. 1j** and **Supplementary File 1**). KEGG pathway analysis[26] showed that these genes were highly enriched for the MAPK signaling pathway (**Fig. 1l**). In *TMEM119* transfected HEK293T.rKSHV cells, 994 genes were upregulated and showed enrichment in the HSV-1 infection pathway, especially the many zinc finger protein-encoding genes (**Fig. 1k,l** and **Supplementary File 2**). This is probably because both HSV-1 and KSHV belong to the Herpesvirus family. While further study is needed to fully characterize the molecular mechanism by which *TMEM119* modulates KSHV reactivation, our data suggests that *TMEM119* promotes viral reactivation via the MAPK pathway, which has previously been implicated in KSHV reactivation[27,28].

FD-seq reveals pro-inflammatory bystander cells after coronavirus OC43 infection

We next applied FD-seq to study infection of the betacoronavirus OC43 in single human lung cells. OC43 is a human pathogen that causes the common cold, and also a close relative of SARS-CoV-2. OC43 has been successfully used as a safer (BSL-2) model pathogen in COVID-19 research. Drug screening studies showed that many drugs that inhibit replication of OC43 also inhibit SARS-CoV-2 *in vitro*[17]. First, we infected human lung epithelial A549 cells with coronavirus OC43 at a multiplicity of infection (MOI) of 1 and fixed the cells with PFA, along

with a fixed mock infected sample, and performed FD-seq. This experiment emulated a scenario where researchers fix SARS-CoV-2-infected sample with PFA and store them for later analysis or international transportation to their collaborators. We obtained 1,167 and 1,924 high-quality single cells from the mock infected and OC43-infected cells, respectively. After regressing out the effects of cell cycle variations (**Fig. S9a, b**), high dimensional clustering yielded three main clusters of cells (**Fig. 2a**). Cluster 0 mainly contained the mock infected cell, while clusters 1 and 2 mainly contained OC43-infected cells (**Fig. 2b**). At MOI 1, more than 70% of the cells were infected (defined as cells showing non-zero expression of viral genes) (**Fig. 2c**). Interestingly, this infected population consists of two subpopulations based on the number of viral transcripts detected: a larger subpopulation containing a low number of viral transcripts (below 100 UMIs, median 2 UMIs) and a smaller subpopulation containing a much high number of viral transcripts (above 100 UMIs, median 428 UMIs) (**Fig 2d**). This small subpopulation of highly infected cells corresponds to cluster 2 (**Fig. 2e, Fig. S9c, d**).

Next, we sought to investigate the expression profile of OC43 viral genes as a function of total viral gene level. Within cluster 2, gene *N* (which encodes the viral nucleocapsid protein[29]) was uniformly highly expressed (**Fig. 2f**). The expression of the remaining viral genes was more heterogeneous, with *ORF1ab*, *M* (membrane gene) and *S* (spike gene) being more highly expressed overall compared to *E* (envelope gene), *HE* (hemagglutinin-esterase gene) and *ns12.9* (which encodes a 12.9 kDa non-structural protein). *ns2* (which encodes the non-structural protein 2) was the least abundant viral gene. These patterns of viral gene expression level were also true when looking at all infected cells (**Fig. S10**).

Cluster 1 consisted of cells that were exposed to the virus but failed to express high levels of viral genes (**Fig. 2d and Fig. S9c, d**), and were thus either abortively-infected or uninfected

bystander cells. We found that these bystander cells were enriched in pro-inflammatory genes, such as *CXCL1*, *CXCL5*, *CCL2* and *NFKBIA* (**Fig. 2g, Fig. S11,** and **Supplementary File 3**). *CXCL1* and *CXCL5* encode the protein ligands of the CXC chemokine receptor 2 (CXCR2), and are crucial to neutrophil recruitment[30]. CCL2 encodes a different chemokine that is responsible for monocyte recruitment from the bone marrow[31]. *NFKBIA* encodes the inhibitor IκBα to the transcription factor NF-κB, which is an important regulator of the immune response[32]. KEGG pathway analysis showed that cluster 1 was enriched for three main pro-inflammatory pathways: TNF, IL-17 and NF-κB signaling pathways (**Fig. 2g**). On the other hand, the ribosome biogenesis pathway is enriched in the upregulated genes in this cluster 2 (**Fig. 2g**), most probably because the virus inside the cells in cluster 2 hijacks the host's machinery for their own RNA replication. These findings are in agreement with our previous characterization of HSV-1 infection at the single-cell level[22], which showed that most HSV-1 infected cells did not express high levels of viral genes, and that these abortively-infected cells were the main producers of antiviral gene transcripts and the initiators of the type I interferon signaling.

Conclusion

Here we present FD-seq, a novel method for high-throughput droplet-based sequencing of PFA-fixed single cells. FD-seq is particularly useful for sequencing rare subpopulations of cells that require intracellular staining and FACS-enrichment, and for rendering BSL-3 samples safe for BSL-2 condition. We used FD-seq to study the process of KSHV lytic reactivation at the single-cell level, and found that reactivation is very heterogeneous, and that the expression levels of viral genes correlated poorly with one another. We found four host factors to be strongly correlated with viral reactivation, namely *CDH1*, *CORO1C*, *ISCU* and *TMEM119*. *ISCU* is involved in the production of the iron-sulfur cluster assembly enzyme, and a mutation in *ISCU* is linked to

myopathy[33]. *CDH1* (E-cadherin) is an adhesion protein, and different mutations in *CDH1* are associated with different cancer conditions[34]. *CORO1C* encodes a protein that belongs to the WD repeat protein family. Lastly, *TMEM119* is a transmembrane protein, and is a marker for human microglial cells[35]. Using live-cell imaging and time-course study, we found *TMEM119* to have the most pronounced effect among the four genes in enhancing the degree of viral reactivation. Bulk RNA-seq suggested that this effect of *TMEM119* could be mediated through the MAPK signaling pathway.

We showed a potential application of FD-seq for investigating the host response in a coronavirus infection *in vitro* model, which also suggests its possible use in COVID-19 pandemic research. We found that most of the cells after exposure to the OC43 coronavirus were unable to support high level of viral gene expression, and instead upregulated pro-inflammatory signaling pathways. Therefore, these bystander cells are the main producers of pro-inflammatory gene products such as cytokines during OC43 infection *in vitro*. Because cytokine storm has been shown to be associated many viral infections[36] and with COVID-19 disease severity[18], and because of the close relationship between OC43 and SARS-CoV-2, it would be interesting to investigate whether cytokine storm is indeed driven by these bystander cells in COVID-19 patients. If this is true, the bystander cells can be a therapeutic target for these patients.

FD-seq is a valuable tool for integrating protein activity with transcriptome information. Transcription factor expression levels and phosphorylation can be integrated with whole transcriptome analysis by combining intracellular protein staining with FD-seq. Furthermore, FD-seq is compatible with RNA velocity analysis, which relies on detection of unspliced mRNA molecules[12] (**Fig. S12**). In future studies, FD-seq could potentially be improved for sequencing

formalin-fixed paraffin-embedded (FFPE) tissues, the achievement of which would enable high-throughput single-cell sequencing of readily available samples in tissue banks.

Data availability

The raw data, metadata and count data are deposited in NBCI's Gene Expression Omnibus (accession number GSE156988).

Author contribution

H.V.P., M.v.G. and N.D. planned the experiments. H.V.P. optimized FD-seq, and performed FD-seq, Drop-seq and bulk RNA-seq experiments. M.v.G. performed KSHV reactivation experiments, transfection and silencing experiments. N.D. performed live-imaging experiments and OC43 infection experiment. H.V.P., M.v.G. and N.D. analyzed the data. A.B. provided guidance on optimizing FD-seq and analyzing scRNA-seq data. M.U.G. and S.T. provided guidance on the experiments. H.V.P., M.v.G., N.D. and S.T. wrote the manuscript. All authors commented on the manuscript.

Acknowledgement

We thank Heather Eckart for providing advice on Drop-seq. This study was supported in part by U.S. National Institutes of Health (NIH) grants (R21 AI118509 and R01 AI087846 to M.U.G.).

Competing interests

The authors declare no competing interests.

Methods

Microfluidic device fabrication

The microfluidic device was fabricated based on the AutoCAD file provided in Macosko et al.[3] The mold was made from SU8-3050 photoresist of 100 μm height. Then, polydimethylsiloxane (PDMS, Momentive RTV615) was cast on the mold to form the microfluidic devices. Plasma bonding was used to bond the PDMS device to a glass slide, and the microfluidic channels were coated with Aquapel.

Cell culture

HEK293T (ATCC) cells were maintained in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% (v/v) fetal bovine serum (FBS), 2 mM GlutaMAX (Gibco), and 1% (v/v) penicillin-streptomycin (P/S, Gibco). HEK293T.rKSHV.219 cells were generated by infecting HEK293T cells with rKSHV.219[25], and selecting with 1 μg/mL puromycin. The KSHV-positive human PEL cell line BC3 was cultured in Roswell Park Memorial Institute (RPMI) supplemented with 20% (v/v) FBS, 2 mM GlutaMAX, and 1% (v/v) P/S. A549 cells (ATCC) were maintained in DMEM supplemented with 10% FBS, without antibiotics. The 3T3 cells ($p65^{-/-}$ 3T3 mouse embryonic fibroblast cells expressing p65-DsRed and H2B-GFP nucleus marker[32]) were cultured in DMEM supplemented with 10% (v/v) fetal bovine calf serum (HyClone), 1% (v/v) GlutaMAX and 1x P/S.

Optimization of RNA extraction condition for fixed and permeabilized cells

BC3 cells were harvested, centrifuged at 300g for 3 min to remove the cell media, and washed once with PBS and 1% BSA (molecular biology grade, Gemini Bio-Products). To fix the cells, 4% PFA in PBS (Santa Cruz Biotechnology) was added to the cells and incubated at room temperature for 15 minutes. Next, the paraformaldehyde was discarded, and the cells was washed once with wash buffer (PBS with 1% BSA and 40 units/ml of RNase inhibitor, murine (NEB)). To

permeabilize the cells, 0.1% Triton X-100 (molecular biology grade, Acros Organics) diluted in the wash buffer was added to the cells, and incubated at room temperature for 15 minutes. Then, after adding the wash buffer, cells were pelleted, the supernatant was discarded, and the cells were wash once more with the wash buffer. To mimic the condition of antibody staining, cells were resuspended in wash buffer and incubated on ice for 1 hour, washed twice with the wash buffer and finally stored in wash buffer.

To serve as a positive control, RNA from bulk live cells was extracted using the RNeasy Plus Mini Kit (Qiagen). Cells were first suspended in a combination of the standard Drop-seq lysis buffer with PBS at equal volume ratio (to make the lysis buffer's final concentration the same as that in droplets), incubated at room temperature for 10 minutes, and placed on ice for at least 5 minutes. Next, the cell lysate was combined with 350 µl of RLT plus buffer from the RNeasy kit, and processed according to the manufacturer's protocol.

To extract RNA from bulk fixed cells, fixed and permeabilized cells (as described above) were suspended in a combination of Drop-seq lysis buffer spiked in with various concentrations of proteinase K (NEB), and PBS at equal volume ratio. Next, the sample was at 56ºC for 1 hour, left at room temperature for 10 minutes, and placed on ice for at least 5 minutes. Finally, the cell lysate was combined with 350 µl of RLT plus buffer, and processed according to the RNeasy Plus Mini Kit's protocol.

Treatment of BC3 cells, flow cytometry and FACS-sorting of reactivated cells

BC3 cells were mock-treated or treated with a final concentration of 20 ng/mL tetradecanoyl phorbol acetate (TPA, Millipore Sigma) for 36 hours to induce KSHV reactivation. Cells were pelleted by centrifugation at 300g for 3 min and fixed with 4% PFA in PBS for 15 min at room

temperature followed by permeabilization with 0.1% (v/v) Triton X-100 in PBS with 1% (w/v) BSA for 15 min at room temperature. After washing and blocking the cells in PBS supplemented with 4% BSA, staining was performed with an anti-KSHV K8.1 antibody (sc-65446, Santa Cruz, 1:50 dilution) and an Alexa Fluor 488-conjugated goat-anti-mouse secondary antibody (A10667, Life Technologies, 1:4000 dilution) in PBS/1% BSA supplemented with RNase inhibitor, murine (NEB). K8.1-positive and negative cell populations were analyzed on an LSRFortessa flow cytometer (BD Biosciences) and/or sorted on a FACSAriaIIIu cell sorter (BD Biosciences).

Infection of A549 cells

OC43 (ATCC) were grown and titrated on A549 cells. For FD-seq experiments, A549 cells were seeded in 6-well plates and infected with OC43 at an MOI of 1 the following day. Cells were incubated at 33C for 4 days and were subsequently harvested for FD-seq.

FD-seq protocol

The protocol for FD-seq is based on the McCaroll lab's online Drop-seq protocol[3]. In summary, there are only two significant differences: 40 units/ml proteinase K was added to the Drop-seq lysis, and a heating step was added between the droplet generation and droplet breakage steps. First, the barcoded beads were suspended at 120,000 beads/ml in a modified Drop-seq lysis buffer: 200 mM Tris pH 7.5, 6% Ficoll type 400, 0.2% Sarkosyl, 20 mM EDTA, 50 mM DTT and 40 units/ml proteinase K (NEB). Cells were suspended in PBS with 0.01% BSA at 100,000 cells/ml. For droplet generation, the cells, beads and oil were injected at flow rates of 3, 3 and 12 ml/h, respectively, using syringe pumps. After droplet generation, the droplets were heated on a heat block at 56°C for 1 hour to reverse the PFA cross-links, then incubated at room temperature for 10 minutes and kept on ice for at least 5 minutes. After this, the droplets were broken, the beads were

collected, washed, and subjected to reverse transcription and exonuclease I digestion as per standard Drop-seq protocol.

For whole transcriptome amplification (WTA), the beads were distributed into a 96-well plate such that each well contained 5,000 beads. A 50 ul PCR reaction was set up in each well (1X KAPA HiFi Hotstart Readymix, 0.8 uM TSO_PCR primer) using the following thermal cycle program: 95C 3 min; 4 cycles of 98C 20s, 65C 45s, 72C 3 min; 10 cycles of 98C 20s, 67C 20s, 72C 3 min; 72C 5 min; 4C hold. After PCR finishes, the WTA products were cleaned up with 0.6X Ampure XP beads (Beckman Coulter), pooled and quantified with TapeStation (Agilent).

To prepare the sequencing library, 450pg of WTA products were used for each tagmentation reaction (Nextera XT DNA Library Prep Kit, Illumina). First, the WTA products were incubated with the tagment enzyme at 55C for 5 min. Next, the neutralization buffer was added, and the sample was incubated at room temperature for 5 min, in order to neutralize the enzyme. After this, the Nextera PCR Master Mix and custom primers P5-TSO_Hybrid and Nextera_N70X were added, and the sample was thermocycled as following: 95C 30s; 12 cycles of 95C 10s, 55C 30s, 72C 30s; 72C 5 min; 4C hold. Finally, the tagmentation products were cleaned up with 0.6X Ampure XP beads and quantified with TapeStation. The sequences of the primers are given in Table S1.

Plasmids and transfections

Expression plasmids encoding human *CDH1*, *TMEM119*, *ISCU* and *CORO1C*, as well as KSHV *ORF50* and nls-eGFP under control of the human EF1α promoter in the pLV backbone were ordered from VectorBuilder. All sequences were verified by Sanger sequencing. HEK293T.rKSHV219 cells were reverse transfected with 0.8 μg plasmid DNA using

Lipofectamine2000 (Life Technologies) following the manufacturer's instructions and seeded in 12 or 24-well plates for analysis by live-cell imaging or qRT-PCR as indicated. For silencing experiments, HEK293T.rKSHV219 cells were reverse transfected with 40 nM siRNA targeting *TMEM119* (Dharmacon siGenome SMARTpool M-018636-01-0005), *CDH1* (M-003877-02-0005), *CORO1C* (M-017331-00-0005), ISCU (M-012837-03-0005), or non-targeting control (D-001206-14-05) in 12 or 24-well plates using Lipofectamine RNAiMAX reagent (Life Technologies) following the manufacturer's protocol. After 48 hours, the cells were treated with 2 ng/mL TPA as indicated and KSHV reactivation efficiency was assessed by live cell imaging or qRT-PCR.

Reverse transcription and real-time PCR (RT-qPCR)

Total RNA was extracted from cells using the HP Total RNA Kit (OMEGA Bio-Tek) following the manufacturer's instructions. RT-qPCR was performed using either a one-step or a two-step protocol. For the one-step protocol, RT-qPCR were performed with equal amounts (25-500 ng) of RNA using the SuperScript III Platinum One-Step qRT-PCR kit with ROX (Thermo Fisher Scientific) on a 7500 Fast Real-Time PCR Machine (Applied Biosystems). Premixed master mixes containing TaqMan primers and probes for each individual gene were purchased from IDT (*GAPDH*, *TMEM119*, *CORO1C*, *ISCU*, and *CDH1*) or Applied Biosystems (18S).

For the two-step protocol, reverse transcription was performed using the ProtoScript II First Strand cDNA Synthesis Kit (NEB) according to the manufacturer's protocol. The kit's oligo-dT primer was used as the primer in this step. Next, the RT products were diluted by 20-fold in water, and used as the input for qPCR (Luna Universal qPCR Master Mix, NEB). The primers' sequences are given in Table S2. Relative expression level of each target gene was calculated by

normalizing for GAPDH or 18S levels using the Comparative Ct Method (ΔΔCt Method) and presented relative to the control sample.

Live-cell imaging experiments

HEK293T.rKSHV were seeded in 24-well plates and transfected with plasmids for over-expression or siRNA for knock-down experiments. Cells were then imaged on a Nikon Ti-Eclipse containing an environmental chamber (37C, 100% humidity, 5% CO2). Images were acquired every 4 hours for 4 days.

Bulk RNA-seq

Total RNA was extracted from cells with the HP Total RNA Kit as above, and quantified using NanoDrop. mRNA was isolated from 100 ng of total RNA per sample using oligo dT magnetic beads (NEB) according to the manufacturer's protocol. The isolated mRNA was then used to prepare sequencing libraries using NEBNext Ultra II Non-directional RNA Library Prep Kit for Illumina and NEBNext Multiplex Oligos for Illumina (Index sets 1 and 2) (NEB). Next, the sequencing libraries were quantified using High Sensitivity Fragment Analyzer (Agilent), normalized to 30 nM, pooled at equimolar, and sequenced with NextSeq High-output kit (single-read 75 bp).

Sequencing and alignment

Drop-seq and FD-seq libraries were sequenced on a NextSeq 550 machine with the following read distribution: 20 bp for read 1, 60 bp for read 2, and 8 bp for index 1 read. The custom read 1 sequencing primer Read1CustomSeqB was used in place of Illumina's read 1 primer (Table S1). Alignment was performed using Drop-seq tools version 1.13.

Bulk RNA-seq libraries were sequenced on a NextSeq 550 machine with the following read distribution: 75 bp for read 1, and 6 bp for index 1 read. Alignment was performed using STAR aligner (version 2.7.3a) and counted using featureCounts. Alignemnt was performed to a concatenated version of the human genome GRCh38 and OC43 genome NC_006213.1[29].

scRNA-seq data analysis

Data analysis was done using R packages Monocle 2[37–39]q and Seurat 3[40,41]. Data visualization was done using R and Python. For species mixing data, a cell is classified as human or mouse if more than 90% of the detected transcripts are mapped to human or mouse, respectively. dropEst pipeline[42] was used with argument –L eiEIBA to annotate reads mapping to exon, intron or exon/intron spanning regions. For the BC3 data, we first discarded cells with more than 3,000 genes detected, and cells with fewer than 1,000 UMIs or more than 10,000 UMIs. Then we discarded genes detected in fewer than 5 cells. To find the relative expression of each gene, we first normalized the cells using Monocle 2. To cluster the cells, we first performed dimension reduction with t-SNE using the first 4 principal components, then clustered the cells by setting rho_threshold = 17 and delta_threshold = 11. We then set the percentage of viral transcripts as the pseudotime parameter, and performed a "pseudotime analysis" in order to find genes correlated with the abundance of viral transcripts. For the A549 data, we discarded cells with fewer than 1,000 UMIs or more than 20,000 UMIs. We then discarded genes detected in fewer than 5 cells. We then used Seurat 3 to remove the cell cycle effects, normalize and scale the data. To cluster the cells, we used the first 10 principal components for the FindNeighbors and RunTSNE functions, and resolution=0.2 for the FindClusters function. To find the cluster markers, we used min.pct=0.25 and logfc.threshold=0.25 for the FindAllMarkers function.

Bulk RNA-seq data analysis

Bulk RNA-seq data was analyzed using the R package DESeq2[43]. To find the enriched KEGG pathways, we uploaded the list of genes that were upregulated in TMEM119 overexpression versus GFP overexpression (log2 fold change > 1, FDR < 0.05) to g:Profiler[26].

RNA velocity analysis

The output files of Drop-seq tools DetectBeadSynthesisErrors function were processed with the dropEst pipeline[42] to tag spliced and unspliced transcripts, and the results were analyzed with Python velocyto package[12].

References

1. Villani, A.-C. *et al.* Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science* **356**, eaah4573 (2017).
2. Yan, L. *et al.* Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. *Nat. Struct. Mol. Biol.* **20**, 1131 (2013).
3. Macosko, E. Z. *et al.* Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).
4. Zheng, G. X. Y. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).
5. Thomsen, E. R. *et al.* Fixed single-cell transcriptomic characterization of human radial glial diversity. *Nat. Methods* **13**, 87–93 (2016).

6. Hori, S., Nomura, T. & Sakaguchi, S. Control of Regulatory T Cell Development by the Transcription Factor Foxp3. *Science* **299**, 1057–1061 (2003).
7. Brambrink, T. *et al.* Sequential Expression of Pluripotency Markers during Direct Reprogramming of Mouse Somatic Cells. *Cell Stem Cell* **2**, 151–159 (2008).
8. Alles, J. *et al.* Cell fixation and preservation for droplet-based single-cell transcriptomics. *BMC Biol.* **15**, 44 (2017).
9. Katzenelenbogen, Y. *et al.* Coupled scRNA-Seq and Intracellular Protein Activity Reveal an Immunosuppressive Role of TREM2 in Cancer. *Cell* (2020) doi:10.1016/j.cell.2020.06.032.
10. Pollice, A. A. *et al.* Sequential paraformaldehyde and methanol fixation for simultaneous flow cytometric analysis of DNA, cell surface proteins, and intracellular proteins. *Cytometry* **13**, 432–444 (1992).
11. Krutzik, P. O. & Nolan, G. P. Intracellular phospho-protein staining techniques for flow cytometry: Monitoring single cell signaling events. *Cytometry A* **55A**, 61–70 (2003).
12. La Manno, G. *et al.* RNA velocity of single cells. *Nature* **560**, 494–498 (2018).
13. Wen, K. W. & Damania, B. Kaposi sarcoma-associated herpesvirus (KSHV): Molecular biology and oncogenesis. *Cancer Lett.* **289**, 140–150 (2010).
14. Cesarman, E. Gammaherpesviruses and Lymphoproliferative Disorders. *Annu. Rev. Pathol. Mech. Dis.* **9**, 349–372 (2014).
15. Manners, O., Murphy, J. C., Coleman, A., Hughes, D. J. & Whitehouse, A. Contribution of the KSHV and EBV lytic cycles to tumourigenesis. *Curr. Opin. Virol.* **32**, 60–70 (2018).
16. Zhou, F. *et al.* Oncolytic Reactivation of KSHV as a Therapeutic Approach for Primary Effusion Lymphoma. *Mol. Cancer Ther.* **16**, 2627–2638 (2017).

17. Drayman, N. *et al.* Drug repurposing screen identifies masitinib as a 3CLpro inhibitor that blocks replication of SARS-CoV-2 in vitro. *bioRxiv* 2020.08.31.274639 (2020) doi:10.1101/2020.08.31.274639.
18. Huang, C. *et al.* Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet* **395**, 497–506 (2020).
19. Ganem, D. KSHV and the pathogenesis of Kaposi sarcoma: listening to human biology and medicine. *J. Clin. Invest.* **120**, 939–949 (2010).
20. Aneja, K. K. & Yuan, Y. Reactivation and Lytic Replication of Kaposi's Sarcoma-Associated Herpesvirus: An Update. *Front. Microbiol.* **8**, (2017).
21. Arias, C. *et al.* KSHV 2.0: A Comprehensive Annotation of the Kaposi's Sarcoma-Associated Herpesvirus Genome Using Next-Generation Sequencing Reveals Novel Genomic and Functional Features. *PLoS Pathog.* **10**, e1003847 (2014).
22. Drayman, N., Patel, P., Vistain, L. & Tay, S. HSV-1 single-cell analysis reveals the activation of anti-viral and developmental programs in distinct sub-populations. *eLife* **8**, e46339 (2019).
23. Berger, J. N. *et al.* Redefining De Novo Gammaherpesvirus Infection Through High-Dimensional, Single-Cell Analysis of Virus and Host. *bioRxiv* 2020.08.11.203117 (2020) doi:10.1101/2020.08.11.203117.
24. Glaunsinger, B. & Ganem, D. Lytic KSHV Infection Inhibits Host Gene Expression by Accelerating Global mRNA Turnover. *Mol. Cell* **13**, 713–723 (2004).
25. Vieira, J. & O'Hearn, P. M. Use of the red fluorescent protein as a marker of Kaposi's sarcoma-associated herpesvirus lytic gene expression. *Virology* **325**, 225–240 (2004).

26. Raudvere, U. *et al.* g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* **47**, W191–W198 (2019).

27. Qin, D. *et al.* Activation of PI3K/AKT and ERK MAPK signal pathways is required for the induction of lytic cycle replication of Kaposi's Sarcoma-associated herpesvirus by herpes simplex virus type 1. *BMC Microbiol.* **11**, 240 (2011).

28. Guito, J. & Lukac, D. M. KSHV Reactivation and Novel Implications of Protein Isomerization on Lytic Switch Control. *Viruses* **7**, 72–109 (2015).

29. St-Jean, J. R. *et al.* Human respiratory coronavirus OC43: genetic stability and neuroinvasion. *J. Virol.* **78**, 8824–8834 (2004).

30. De Filippo, K. *et al.* Mast cell and macrophage chemokines CXCL1/CXCL2 control the early stage of neutrophil recruitment during tissue inflammation. *Blood* **121**, 4930–4937 (2013).

31. Shi, C. & Pamer, E. G. Monocyte recruitment during infection and inflammation. *Nat. Rev. Immunol.* **11**, 762–774 (2011).

32. Tay, S. *et al.* Single-cell NF-κB dynamics reveal digital activation and analogue information processing. *Nature* **466**, 267–271 (2010).

33. Mochel, F. *et al.* Splice Mutation in the Iron-Sulfur Cluster Scaffold Protein ISCU Causes Myopathy with Exercise Intolerance. *Am. J. Hum. Genet.* **82**, 652–660 (2008).

34. Berx, G., Becker, K.-F., Höfler, H. & Van Roy, F. Mutations of the human E-cadherin (CDH1) gene. *Hum. Mutat.* **12**, 226–237 (1998).

35. Satoh, J. *et al.* TMEM119 marks a subset of microglia in the human brain: Human microglial marker TMEM119. *Neuropathology* **36**, 39–49 (2016).

36. Tisoncik, J. R. *et al.* Into the Eye of the Cytokine Storm. *Microbiol. Mol. Biol. Rev.* **76**, 16–32 (2012).

37. Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381 (2014).

38. Qiu, X. *et al.* Single-cell mRNA quantification and differential analysis with Census. *Nat. Methods* **14**, 309 (2017).

39. Qiu, X. *et al.* Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods* **14**, 979 (2017).

40. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).

41. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e21 (2019).

42. Petukhov, V. *et al.* dropEst: pipeline for accurate estimation of molecular counts in droplet-based single-cell RNA-seq experiments. *Genome Biol.* **19**, 78 (2018).

43. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

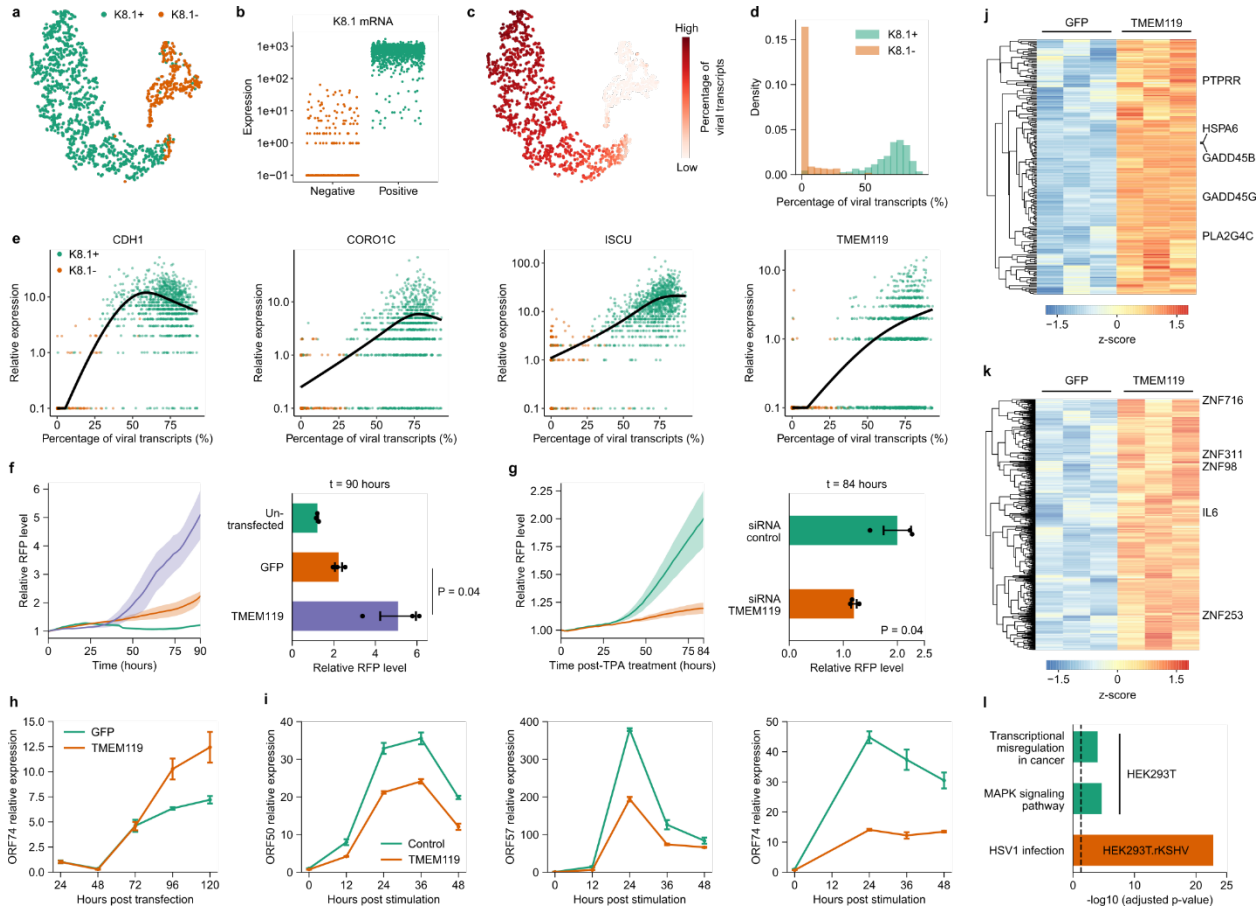

Figure 1. FD-seq identifies *TMEM119* as a host factor that modulates KSHV reactivation. (**a**) t-SNE plot of K8.1+ (green, lytically reactivated) and K8.1- (red, latently infected) BC3 cells as analyzed by FD-seq. (**b**) KSHV K8.1 mRNA levels in the K8.1- and K8.1+ sorted populations. (**c**) t-SNE plot of reactivated and non-reactivated cells as in (a), colored by the percentage of total viral transcripts. (**d**) Histogram showing the percentages of detected transcripts that are from KSHV in K8.1+ and K8.1- BC3 cells. (**e**) Correlation between the percentage of KSHV transcripts in the cell (x-axis) and relative expression of the host factors *CDH1*, *CORO1C*, *ISCU* and *TMEM119*. Dots indicate individual cells in the K8.1+ or K8.1- populations. (**f**) Live-cell imaging analysis of RFP expression, which is indicative of KSHV reactivation, in untransfected HEK293T.rKSHV219 cells or cells transfected with *TMEM119* or GFP control (left), and end-point quantification of RFP expression at 90 h (right). (**g**) Live-cell imaging analysis of RFP expression in HEK293T.rKSHV219 cells transfected with control or TMEM119-targeting siRNAs for 48 h, followed by treatment with 2 ng/mL TPA (left), and end-point quantification of RFP expression at 84 h (right). (f,g) The ribbons indicate the s.e.m. of the data. P-values: one-sided

Welch's t-test. (**h**) Time-course qRT-PCR analysis of KSHV *ORF74* abundance in HEK293T.rKSHV219 cells transfected with *TMEM119* or GFP control. (**i**) Time-course qRT-PCR analysis of KSHV *ORF50*, *ORF57*, and *ORF74* levels in HEK293T.rKSHV219 cells transfected with control or *TMEM119*-targeting siRNAs as in (g). (f-i) Data are represented as mean ± s.e.m. (n=3 biological replicates). (**j,k**) Heatmap showing the genes upregulated in (j) HEK293T or (k) HEK293T.rKSHV219 cells transfected with *TMEM119* compared to the GFP control (n=3 biological replicates). (j,k) Shown are 5 most upregulated genes related to the (j) MAPK signaling and (k) HSV1 infection pathways. (**l**) Bar plot showing the p-values of overrepresented KEGG pathways of the cells in (j,k).

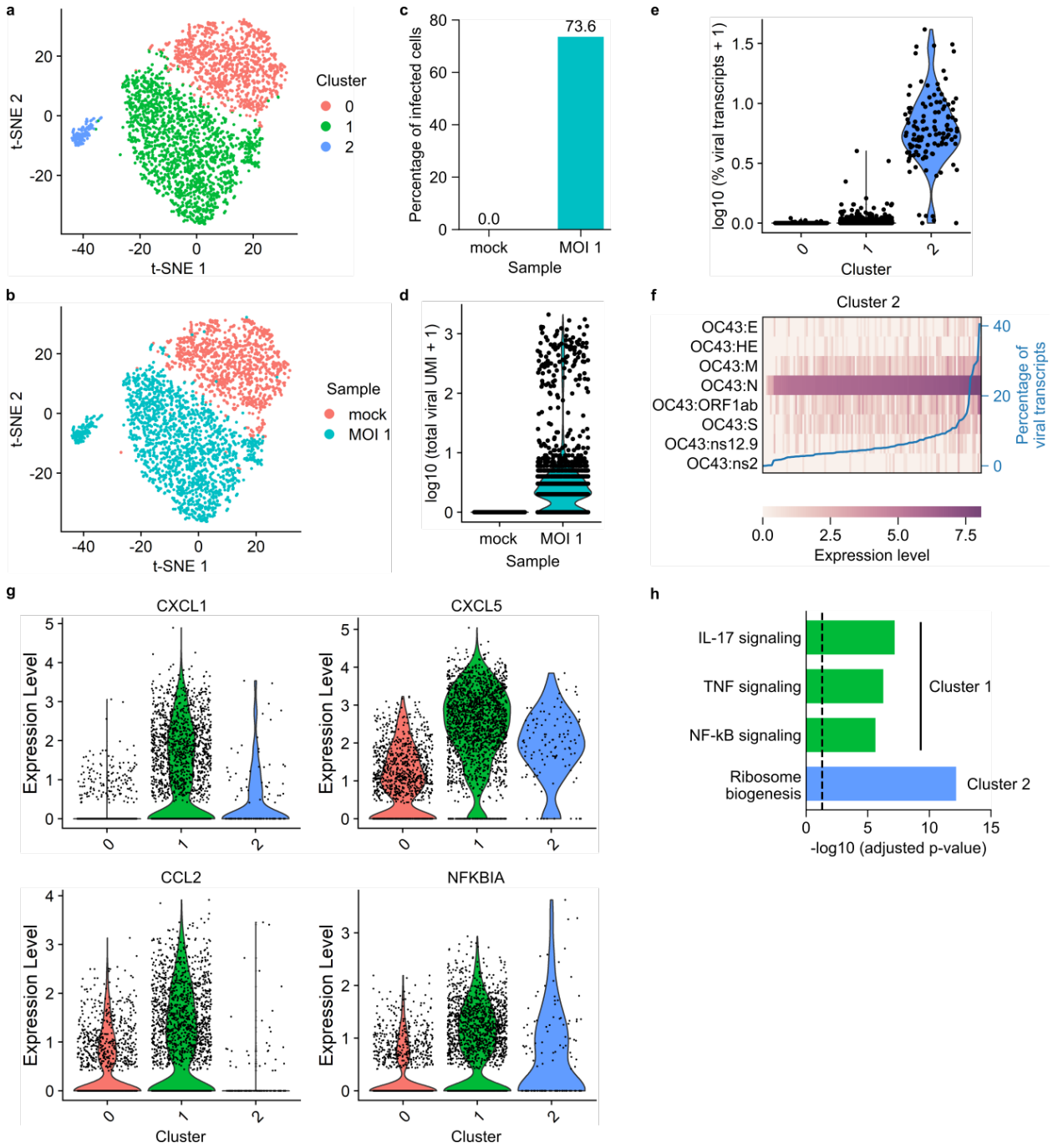

Figure 2. FD-seq reveals pro-inflammatory bystander cells after coronavirus OC43 infection. (**a-b)** t-SNE plots with cells colored by (a) cluster identity or (b) sample type. (**c)** Bar plot showing the percentage of infected cells (defined as cells that expressed at least 1 viral transcript) of mock infected and MOI 1 sample. (**d)** Violin plot showing distribution of total viral transcript counts. (**e)** Violin plot showing the distribution of the percentage of total viral transcript by cluster identity.

(**f**) Heatmap showing the relative expression level of each viral gene of each single cell in cluster 2. The blue line shows each cell's percentage of total viral transcripts. (**g**) Violin plots showing the expression of four representative immune-related genes that are upregulated in cluster 0. (**h**) Bar plot showing the adjusted P-values of upregulated KEGG pathways in clusters 1 and 2. The vertical dashed line indicates P-value = 0.05.