scientific reports

OPEN

A potential role for SARS-CoV-2 small viral RNAs in targeting host microRNAs and modulating gene expression

Zachary T. Neeb[1,5], Alexander J. Ritter[2,5], Lokendra V. Chauhan[3], Sol Katzman[4], W. Ian Lipkin[3], Nischay Mishra[3] & Jeremy R. Sanford[1]

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) causes coronavirus disease (COVID-19) in humans, with symptoms ranging from mild to severe, including fatality. The molecular mechanisms surrounding the effects of viral infection on the host RNA machinery remain poorly characterized. We used a comparative transcriptomics approach to investigate the effects of SARS-CoV-2 infection on the host mRNA and sRNA expression machinery in a human lung epithelial cell line (Calu-3) and an African green monkey kidney cell line (Vero-E6). Upon infection, we observed global changes in host gene expression and differential expression of dozens of host miRNAs, many with known links to viral infection and immune response. Additionally, we discovered an expanded landscape of more than a hundred SARS-CoV-2-derived small viral RNAs (svRNAs) predicted to interact with differentially expressed host mRNAs and miRNAs. svRNAs are derived from distinct regions of the viral genome and sequence signatures suggest they are produced by a non-canonical biogenesis pathway. 52 of the 67 svRNAs identified in Calu-3 cells are predicted to interact with differentially expressed miRNAs, with many svRNAs having multiple targets. Accordingly, we speculate that these svRNAs may play a role in SARS-CoV-2 propagation by modulating post-transcriptional gene regulation, and that methods for antagonizing them may have therapeutic value.

Small non-coding RNAs (sRNAs) play diverse roles in gene regulation and genome integrity. Ranging from ~ 20–30 nt, this functionally diverse class of RNAs plays important roles in the regulation of many biological processes[1,2]. Short interfering RNAs and microRNAs (siRNA and miRNA, respectively) function in post-transcriptional control of gene expression by regulating messenger RNA (mRNA) translation and stability[3,4]. By contrast, Piwi-associated RNAs (piRNAs) control transcriptional silencing of transposable elements in eukaryotic germlines and the elimination of entire regions of ciliate genomes[5,6].

sRNAs are derived from larger precursor transcripts. Although biogenesis pathways for the different types of sRNAs vary widely depending on type and species, processing of precursors typically involves cleavage by the endonucleases Drosha and Dicer before they are loaded onto an Argonaute Family protein (either an AGO or a PIWI) to interact with downstream targets. However, there are a multitude of alternative mechanisms for sRNA biogenesis, some of which include Drosha/Dicer-independent mechanisms or even multiple Argonautes and Dicer-like proteins[7–10]. The level of complementarity of an sRNA to its target varies from as little as 8 nt in the 5′ seed region (miRNAs) up to 100% complementarity of the full sRNA sequence (siRNAs and piRNAs), and often is a determining factor in how an sRNA interacts with its targets. It has also been suggested for human miRNAs that their abundance directly affects whether they repress translation by active mRNA degradation, or by interaction with the 3′ UTR when miRNA levels are insufficient for widespread 3′ UTR binding[11]. Taken as a whole, small RNA biogenesis and function is complex and exceptions to rules regarding roles and mechanisms are quite common.

[1]Department of Molecular, Cell and Developmental Biology and Center for Molecular Biology of RNA, University of California Santa Cruz, Santa Cruz, CA, USA. [2]Department of Biomolecular Engineering, University of California Santa Cruz, Santa Cruz, CA, USA. [3]Center for Infection and Immunity, Mailman School of Public Health, Columbia University, New York, NY, USA. [4]UC Santa Cruz Genomics Institute, University of California Santa Cruz, Santa Cruz, CA, USA. [5]These authors contributed equally: Zachary T. Neeb and Alexander J. Ritter. ✉email: nm2641@cumc.columbia.edu; jsanfor2@ucsc.edu

Viruses such as SARS-CoV-2 remodel host gene expression programs. This occurs through a variety of mechanisms, including deregulation of post-transcriptional gene regulation. More recently, viral small non-coding RNAs have also been identified that may play important roles in adaptation and viral propagation[12–15] in infections with SARS-CoV-1[12], Influenza[13], Hepatitis A[14] and EV71[15]. During the COVID-19 pandemic, multiple groups also reported the existence of miRNA-like sRNAs derived from SARS-CoV-2, although the details of their biogenesis and specific function have yet to be elucidated[16–18]. Small viral RNAs (svRNAs) vary greatly in size and cellular abundance and have been implicated in a variety of different biological pathways related to infection and host immune evasion[19].

In this study, we characterize global changes in both host cell mRNA and miRNA expression using multiple sequencing-based methods taking a comparative transcriptomics approach in two very well-established primate cell lines[20–23]. This approach enables identification of conserved changes in host gene regulatory reponses during infection. We also report the discovery of a diverse landscape of svRNAs produced by SARS-CoV-2. Sequence alignments, along with expression dynamics suggest the intriguing hypothesis that SARS-CoV-2 small RNAs may directly regulate host transcripts, including microRNAs.

Results

SARS-CoV-2 infection induces global changes in host gene expression.   To investigate how SARS-CoV-2 infection influences host gene expression, we analyzed the transcriptomes of both a human lung epithelial cell line (Calu-3) and African green monkey kidney cell line (Vero-E6) during the course of SARS-CoV-2 infection. Both cell lines were infected with SARS-CoV-2 at a multiplicity of infection (MOI) of 0.01. RNA extracts were made from harvested cell pellets of infected Calu-3 cells at 0 h, 12 h, 24 h, 48 h and 72 h post-infection (hpi), and infected Vero-E6 cells at 4 h, 24 h, 48 h, 72 h, 120 h, 165 h and 216 hpi. RNA extracts of cell pellets from uninfected cells were also used as controls and processed from all time points. Similar to previous studies[24], the majority of reads from infected cells mapped to the host genomes (60.3–100.0% for Calu-3 cells, 88.2–95.6% for Vero-E6 cells), with the remainder mapping to the SARS-CoV-2 genome (Fig. 1A, Supplemental Table S1).

SARS-CoV-2 infection induces global changes in gene expression in both Calu-3 cells and Vero-E6 cells. DESeq2[25] analysis revealed 4,593 and 1,040 significantly differentially expressed genes, respectively (Fig. 1B, Supplemental Fig. S1). Differentially expressed genes in Calu-3 and Vero-E6 cells were used for downstream Gene Ontology Biological Process (GOBP) and KEGG Human Pathway enrichment analysis. Upregulated genes fell into enriched gene families related to innate immunity and inflammatory response pathways, while down-regulated genes tended to be involved in metabolic and other biosynthetic pathways (Fig. 1C–F, Supplemental Tables S2-S4). For both Calu-3 and Vero-E6 cells, we observed significant overlap between GOBP terms for upregulated and downregulated genes throughout the time course. By contrast, KEGG pathway overlap was only observed for the upregulated gene sets (Supplemental Fig. S2). Interestingly, for both cell types, we observed gene expression changes to be most significant from 24 to 48 hpi, most likely reflective of the time required to elicit cellular response to viral infection. Overall, we found SARS-CoV-2 infection to cause global gene expression changes in both Calu-3 and Vero-E6 cells.

SARS-CoV-2 expresses a diverse landscape of small RNAs.   We analyzed the small RNA transcriptome from mock and infected Calu-3 and Vero-E6 cells to uncover potential SARS-CoV-2-dependent gene regulatory mechanisms. Following infection, libraries from both cell lines contained reads mapping to the SARS-CoV-2 genome (Fig. 2A, Supplemental Fig. S3, Supplemental Table S1). We found that the distribution of svRNAs mapping to the SARS-CoV-2 genome was not uniform and that there were "pile-ups" in particular regions indicating the RNAs identified are not likely to be degradation products of the full-length viral genomic or subgenomic RNA, but may be derived from specific loci (Fig. 2A, Supplemental Fig. S3). Using Piranha[26], we identified 67 svRNAs from Calu-3 cells with a mean length of 22 nt and 97 svRNAs from Vero-E6 cells with a mean length of 25 nt (Fig. 2B, Supplemental Figs. S4, S5, Supplemental Table S3). 28 svRNA loci were shared between the two cell lines (Fig. 2B). svRNA expression is dynamic throughout the time course with peak expression of the svRNAs between 24 and 48 hpi for both species (Fig. 2C, Supplemental Fig. S6). Calu-3 svRNAs tend to cluster into three distinct types of expression profiles. Two svRNA groups showed transient peak expression at 24 and 48 hpi respectively and then returned to base-level, while another group showed sustained expression from 24 to 48 hpi (Fig. 2C). The dynamics of svRNA expression follows similar kinetics as changes in host mRNA and miRNA expression (Figs. 1B, 3A, Supplemental Fig. S1). To assess the conservation of svRNAs among variants of SARS-CoV-2 we intersected them with known mutations in variants of concern from the Uniprot Amino Acid mutations track on the UCSC Genome Browser. While 6 of the 67 Calu-3-specific svRNAs and 5 of the 66 conserved svRNAs overlapped with at least one mutation, only 3 and 4 of them respectively overlap in such a way that the seed sequence would be affected (Supplemental Table S8, Supplemental Fig. S14). We therefore speculate that these mutations would not significantly impact svRNA function. Further studies are needed to answer this question and we are highly interested in infecting Calu-3 cells with SARS-CoV-2 variants in the future to investigate the functional effects of these mutations.

To illuminate potential biogenesis pathways for SARS-CoV-2 svRNAs, we aligned the 5′ and 3′ ends of svRNAs searching for nucleotide biases. By contrast to miRNAs, the svRNAs did not possess sequence signatures with the typical 5′-U bias, suggesting a Dicer-independent processing pathway (Fig. 2D, Supplemental Fig. S7). Additionally, we found that only 10 of the 67 Calu-3 cell-derived svRNAs and 12 of the 97 Vero-E6 cell-derived svRNAs were predicted to have an upstream or downstream complementary sequence within 44 nt that may form a hairpin[27], often required for canonical miRNA processing. (Supplemental Table S6).
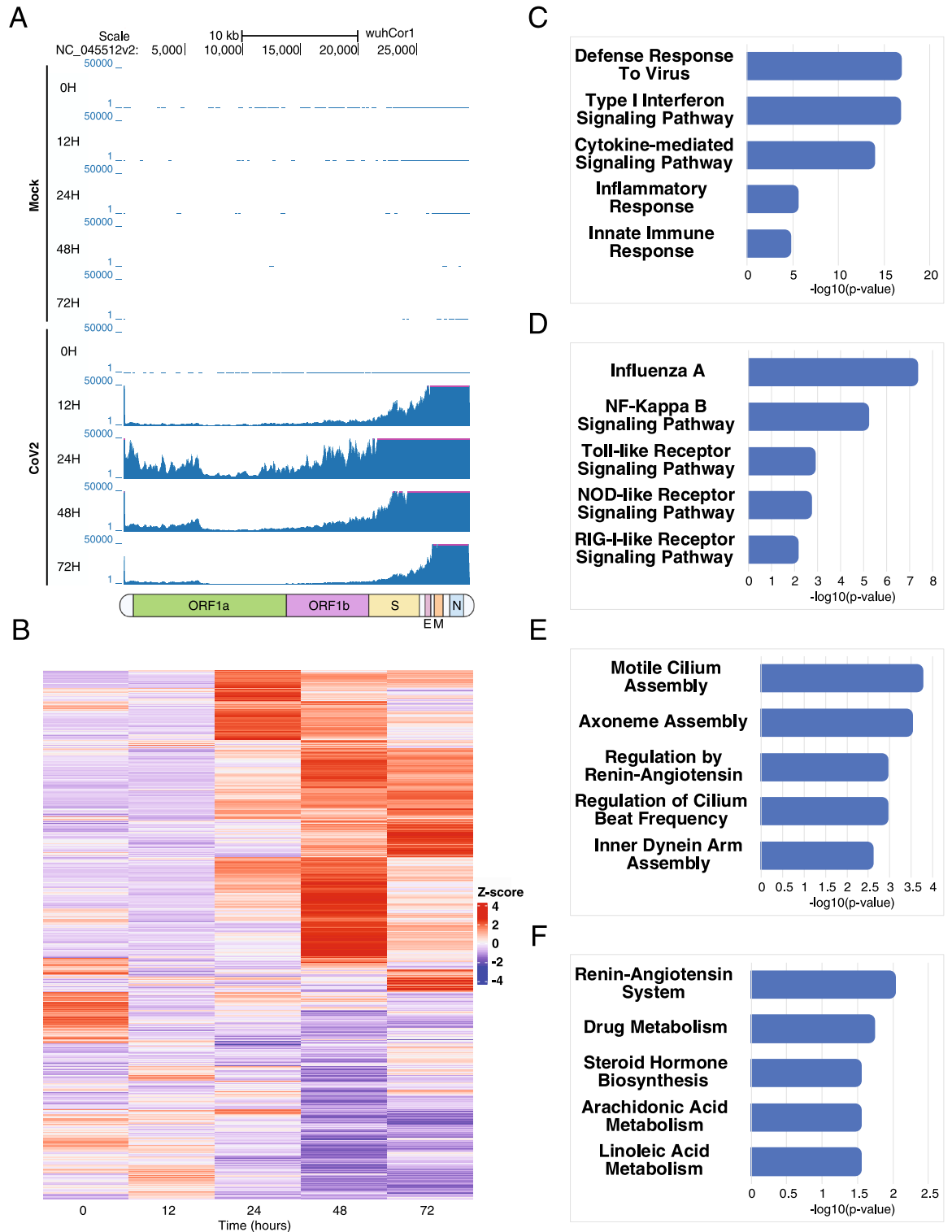
Figure 1. SARS-CoV-2 infection induces global changes in gene expression. (**A**) UCSC Genome Browser screenshots showing Calu-3 sample-derived mRNA sequencing reads mapped to the SARS-CoV-2 genome for infected and control cells. Underneath the screenshot is a schematic representation of the SARS-CoV-2 genome depicting the regions of ORF1a/1b, the spike protein (S), the envelope protein (E), the membrane protein (M) and the nucleocapsid protein (N). (**B**) Heatmap representation of differentially expressed Calu-3 genes over the time course of infection. 4593 differentially expressed genes are shown. Heatmaps were generated using hierarchical clustering. (**C**) Enriched Gene Ontology Biological Process (GOBP) terms are shown for upregulated Calu-3 genes. Only a subset of significantly enriched GOBP terms are shown. (**D**) Same as (**C**), but with enriched KEGG Human Pathways. (**E**) Enriched GOBP terms for downregulated Calu-3 genes. Only a subset of significantly enriched GOBP terms are shown. (**F**) Same as (**E**), but with enriched KEGG Human Pathways.
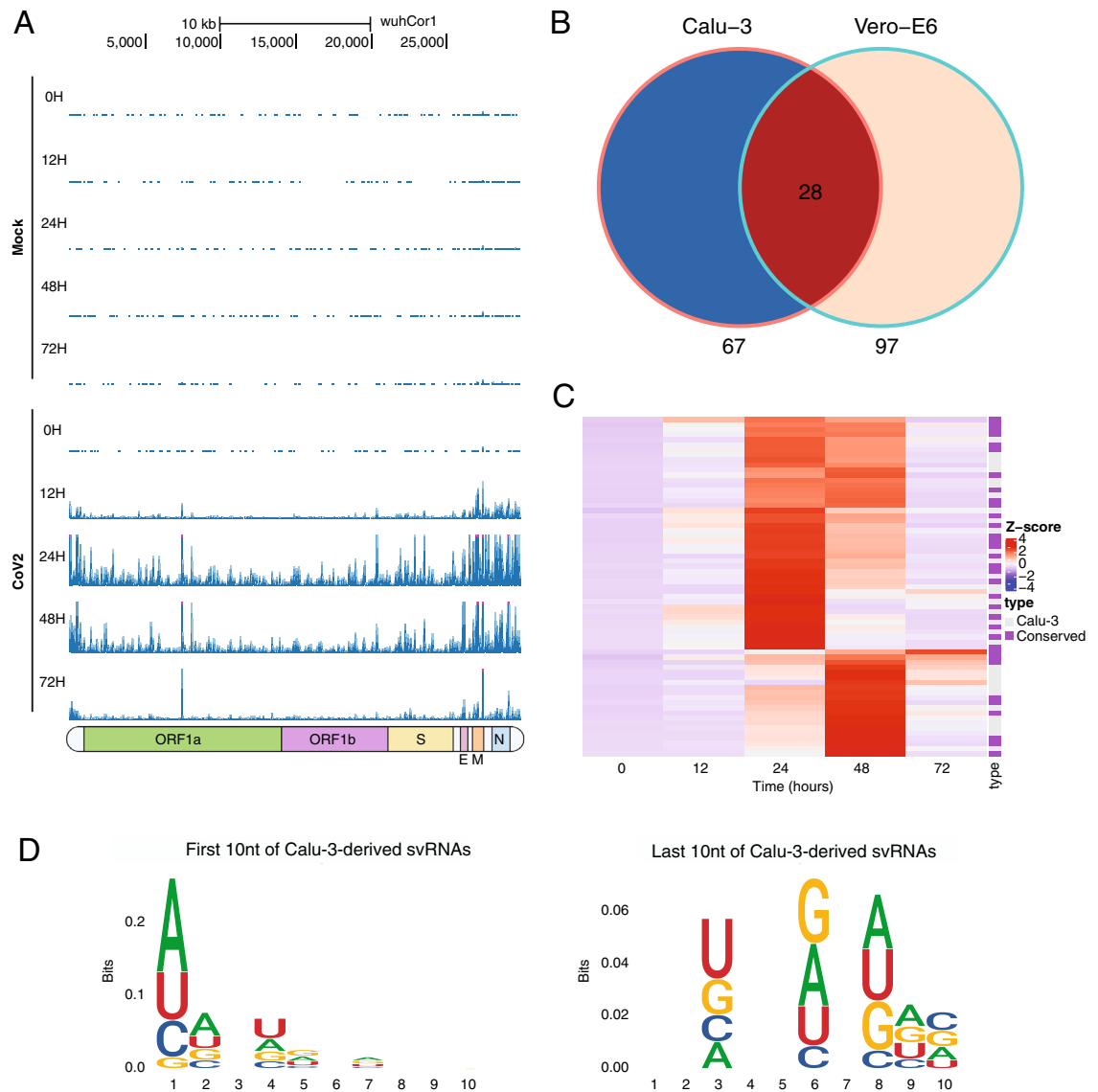
Figure 2. SARS-CoV-2 expresses a diverse landscape of small RNAs. (**A**) UCSC Genome Browser screenshots showing Calu-3 sample-derived sRNA sequencing reads mapped to the SARS-CoV-2 genome for both infected and control cells. Reads from infected cells do not map uniformly to the genome but instead form "pile ups" in particular regions. Underneath the screenshot is a schematic representation of the SARS-CoV-2 genome depicting the regions of ORF1a/1b, the spike protein (S), the envelope protein (E), the membrane protein (M) and the nucleocapsid protein (N). (**B**) Venn Diagram showing the number of svRNAs identified in Calu-3 and Vero-E6 cells, with conserved loci indicated. (**C**) Heatmap representation of Calu-3 cell-derived svRNA expression. Heatmaps were generated using hierarchical clustering. To the right of the heatmap is an annotation column indicating if the svRNA is species-specific or conserved between Calu-3 and Vero-E6 cells. (**D**) Sequence logos for the first ten and last ten nucleotides of the svRNAs identified in Calu-3 cells. svRNAs lack a 5'-U signature typical of many canonical miRNAs.

SARS-CoV-2 alters host microRNA expression. We hypothesized that global changes in gene expression could be mediated through changes in host sRNA expression, specifically miRNAs. We analyzed the host small RNA transcriptomes of infected and control Calu-3 cells (Fig. 3A, Supplemental Figs. S8–S10, Supplemental Table S1) and discovered that 28 miRNAs are differentially expressed throughout the infection time course (Fig. 3A, Supplemental Table S7). Consistent with our mRNA sequencing data, we observed differential expression of miRNAs at 24 hpi with the most significant changes occurring at 48 hpi (Fig. 1B).

SARS-CoV-2 svRNAs may hybridize with differentially expressed mRNAs and miRNAs. To investigate the potential function of svRNAs we used RNAHybrid[28,29] to identify targets within differentially expressed Calu-3 mRNAs and miRNAs. Of the svRNAs identified in Calu-3 cells, 59 of 67 have predicted mRNA targets within the 3'UTRs of differentially expressed (DE) genes (Tables 1 and 2, Supplemental Tables S2-3). RNAhybrid predictions identified svRNAs that can also form stable duplexes with miRNAs, 52 of which target
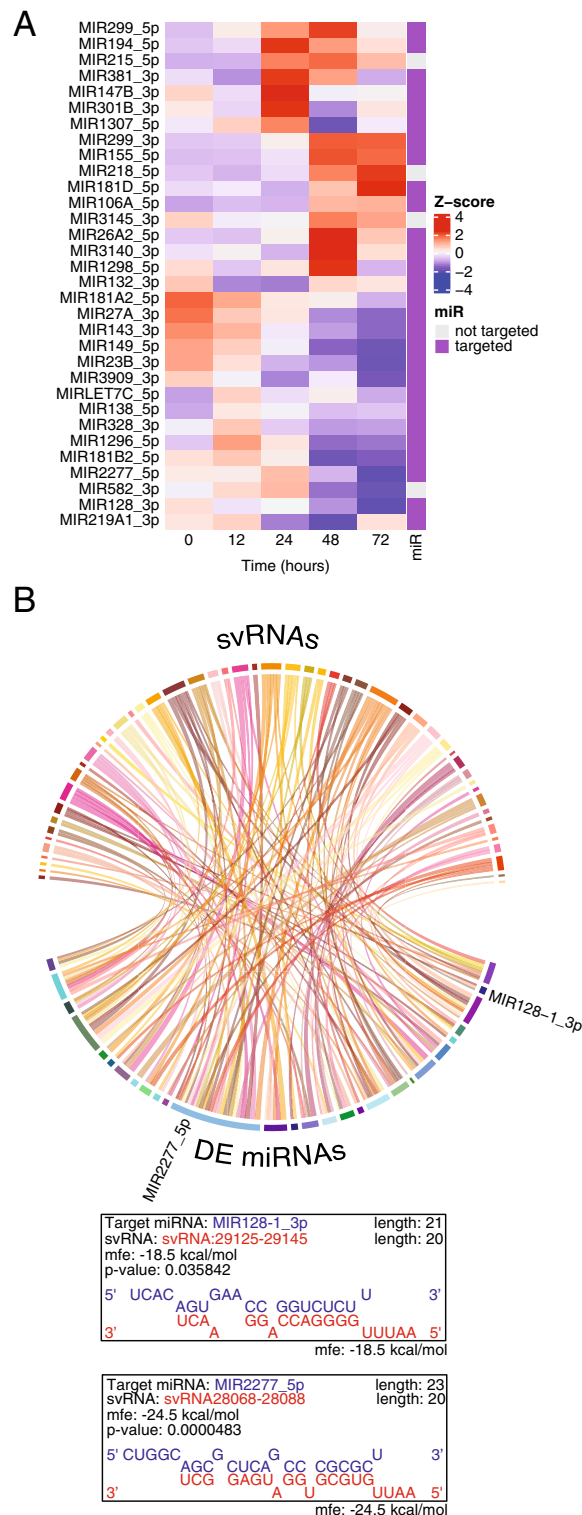
Figure 3. SARS-CoV-2 svRNAs may target host miRNAs. (**A**) Heatmap representation of the 28 differentially expressed Calu-3 miRNAs over the time course of infection. Heatmaps were generated using hierarchical clustering. To the right of the heatmap is an annotation column indicating if the miRNA is a predicted target of an svRNA. (**B**) Circos plot representation of predicted interactions between svRNAs and DE human miRNAs using RNAhybrid without seed-forcing. The ribbons are colored according to the svRNA that is predicted to interact with the miRNA to which it connects. There are many examples of svRNAs targeting more than one miRNA and of miRNAs being targeted by more than one svRNA. DE miRNAs mentioned in the Discussion section are labeled. Shown below are two examples of RNAhybrid prediction outputs for svRNAs and DE miRNAs pictured above.

svRNA	Counts	# mRNA targets	Top enriched GO term for targets
svRNA2875-2897	165	39	Negative regulation of CD4-positive, alpha–beta T cell proliferation (GO:2000562)
svRNA26927-26947	134	18	Negative regulation of CD4-positive, alpha–beta T Cell proliferation (GO:2000562)
svRNA28260-28289	89	1	Gamma-aminobutyric acid metabolic process (GO:0009448)
svRNA27059-27089	88	16	Regulation of protein import into nucleus (GO:0042306)
svRNA27107-27128	80	11	Heart field specification (GO:0003128)
svRNA26820-26840	77	2	Mesoderm morphogenesis (GO:0048332)
svRNA5572-5597	64	4	Regulation of oxidoreductase activity (GO:0051341)
svRNA37-63	57	2	Ribosomal Large Subunit Assembly (GO:0000027)
svRNA18943-18965	47	1	N/A
svRNA29039-29064	45	5	Extracellular structure organization (GO:0043062)

Table 1. List of the ten most abundant Calu-3-derived svRNAs with DE mRNA targets. Table listing the ten most abundant svRNAs with DE mRNA targets, normalized counts, number of differentially expressed mRNA targets and the top enriched GO term for targets. Counts were normalized to the mean read depth of all sample libraries.

Top enriched GO terms for downregulated mRNA targets of svRNAs	Top enriched GO terms for upregulated mRNA targets of svRNAs
Quinone catabolic process (GO:1901662)	Cellular response to cytokine stimulus (GO:0071345)
Leukotriene B4 metabolic process (GO:0036102)	Cytokine-mediated signaling pathway (GO:0019221)
Menaquinone metabolic process (GO:0009233)	Regulation of interferon-gamma production (GO:0032649)
Vitamin K metabolic process (GO:0042373)	Cellular response to interferon-gamma (GO:0071346)
Fat-soluble vitamin metabolic process (GO:0006775)	Negative regulation of natural killer cell mediated cytotoxicity (GO:0045953)
Organic hydroxy compound catabolic process (GO:1901616)	Regulation of apoptotic cell clearance (GO:2000425)
Fat-soluble vitamin catabolic process (GO:0042363)	Positive regulation of apoptotic cell clearance (GO:2000427)
Axoneme assembly (GO:0035082)	interferon-gamma-mediated Signaling Pathway (GO:0060333)
Diterpenoid metabolic process (GO:0016101)	Regulation of immune response (GO:0050776)
Phosphate ion transport (GO:0006817)	Negative regulation of natural killer cell mediated immunity (GO:0002716)

Table 2. List of the top enriched GO terms for DE mRNA targets of Calu-3-derived svRNAs. Table listing the top ten significantly enriched GO terms for all upregulated and all downregulated mRNA targets of Calu-3-derived svRNAs in our data set.

svRNA	Counts	# miRNA targets	miRNA
svRNA26741-26761	165	4	**MIR128_3p**, MIR1298_5p, **MIR1307_5p**, **MIR2277_5p**
svRNA28573-28595	134	1	**MIR2277_5p**
svRNA2868-2897	88	3	**MIR106A_5p**, **MIR2277_5p**, MIR381_3p
svRNA28704-28725	80	2	**MIR138_5p**, **MIR2277_5p**
svRNA28553-28573	77	4	**MIR132_3p**, **MIR2277_5p**, **MIR23B_3p**, MIR299_5p
svRNA18943-18965	57	1	**MIR149_5p**
svRNA27578-27602	47	2	MIR26A2_5p, MIR381_3p
svRNA524-546	45	1	**MIR2277_5p**
svRNA5423-5445	45	3	MIR1298_5p, **MIR181B2_5p**, **MIR3909_3p**
svRNA28068-28088	41	1	**MIR132_3p**

Table 3. List of the ten most abundant Calu-3-derived svRNAs with DE miRNA targets. Table listing the ten most abundant svRNAs with DE miRNA targets, normalized counts, number of differentially expressed miRNA targets and miRNA target names. miRNAs in bold are downregulated in our data set. Counts were normalized to the mean read depth of all sample libraries.
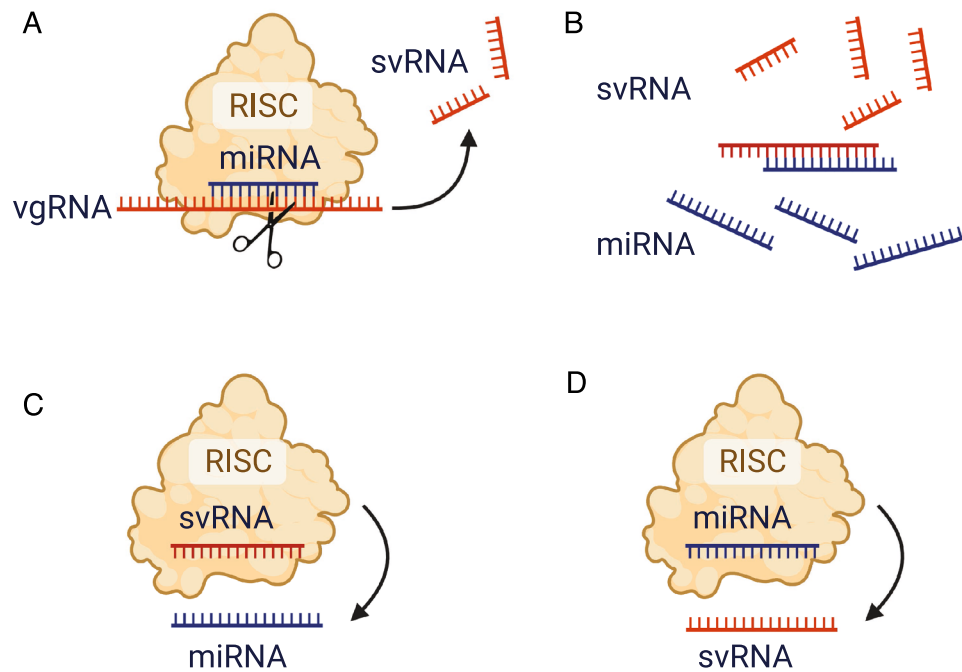
Figure 4. Proposed svRNA-miRNA interactions upon SARS-CoV-2 infection. The illustrated model is briefly described in the Discussion section. svRNAs may interact with miRNAs in multiple ways, none of which are mutually exclusive. (**A**) A RISC-miRNA complex targets complementary sequences in the viral genomic RNA. svRNAs are produced by cleavage and go on to perform downstream functions, such as interacting with host mRNAs and miRNAs. (**B**) There is direct hybridization of svRNAs and miRNAs without association with RISC, leading to sequestration of miRNAs, preventing them from interacting with endogenous targets. (**C**) A RISC-svRNA complex targets miRNAs and prevents them from interacting with endogenous targets. (**D**) A RISC-miRNA complex targets svRNAs, preventing the complex from interacting with endogenous targets. Created with BioRender.com.

differentially expressed miRNAs in our data set (Fig. 3B, Table 3, Supplemental Fig. S11, Supplemental Table S7). A larger portion of svRNA-target interactions were predicted to occur with downregulated miRNAs than with downregulated mRNA targets throughout the time course (Supplemental Fig. S10). Interestingly, we found numerous examples of host miRNAs that can pair with multiple svRNAs (Figs. 3B, 4, Supplemental Fig. S11). In addition, we also observed examples of svRNAs that can potentially hybridize to a broad array of miRNAs (Figs. 3B, 4, Table 3, Supplemental Fig. S11). Because the Vero-E6 genome has not been extensively annotated to include miRNA sequences, we were unable to perform RNAhybrid predictions against Vero-E6 miRNAs, but found that of the 61 Calu-3-derived svRNAs capable of targeting host miRNAs, 27 had conserved loci with Vero-E6-derived svRNAs.

Discussion

We analyzed the transcriptomes of two primate cell lines during SARS-CoV-2 infection and observed global changes in gene expression. Consistent with previous work, we found enrichment for upregulated genes involved in the RIG-I pathway, Toll-like Receptor (TLR) signaling pathways, NF-kappa Beta pathway, and type 1 interferon (IFN) signaling[16,30–34] (Supplemental Fig. S12). Downregulated genes tended to fall into enriched gene families involved with metabolic and other biosynthetic pathways, consistent with SARS-CoV-2-mediated metabolic dysregulation observed in human patients[35–37]. Metabolic dysregulation may lead to an increase in the inflammatory and auto-immune responses as observed in severe Covid-19 patients. We also identified more than a hundred svRNAs capable of interacting with host transcripts, including miRNAs. svRNAs are expressed from specific loci, within ORF7a, and the spike (S) and nucleocapsid (N) genes. In addition to having predicted target sites in differentially expressed host transcript 3'-UTRs like canonical miRNAs, 52 of the 67 Calu-3 sample-derived svRNAs are also predicted to form stable duplexes with differentially expressed miRNAs, many of which have been implicated in other types of viral respiratory infections (VRIs)[38]. For example, miR-128, miR-2277 and miR-155 are known to be associated with rhinovirus infection, Middle Eastern Respiratory Syndrome (MERS) and SARS-CoV-2, respectively. We see many examples of svRNAs with predicted target sites in more than one differentially expressed miRNA, with 38 of the 67 having two or more predicted targets. svRNAs are most highly expressed at 24 h and 48 hpi, which corresponds to the time points at which we see the most significant changes in host gene and miRNA expression. This expression dynamic suggests an interplay between svRNAs and host miRNAs, in which svRNAs have the capability to interact with host miRNAs, therefore affecting downstream target mRNA expression.

Here we describe an expanded landscape of SARS-CoV-2-derived small RNAs including all four svRNAs previously reported by the Cheng, Meng and Pawlica groups[16–18] (Supplemental Fig. S13). In agreement with what the Meng and Cheng groups observed, we identified dozens of svRNAs derived from the nucleocapsid (N) gene, many that overlap the sequences reported in those studies, albeit with varying starting coordinates and lengths[16,17] (Supplemental Fig. S13). In addition to recapitulating the existence of svRNAs derived from the N gene, we identified an svRNA (svRNA27406-27428) derived from ORF7a that almost perfectly overlaps with the sequence reported by Pawlica and colleagues[18] (Supplemental Fig. S13). We reason that our experiment has increased sensitivity because we use a lower MOI (0.01) than previous studies (0.05–5), and performed an extended time course experiment[16,18]. In contrast to previous studies, we have shown SARS-CoV-2 infection to cause significant changes in global host miRNA expression and have additionally predicted the svRNAs identified here to interact with differentially expressed host miRNAs.

miRNAs are typically processed via a Dicer-mediated biogenesis pathway. Our data suggest that many of the svRNAs are generated instead via a Dicer-independent processing mechanism. Lack of a 5′-U bias in the sequence analyses for svRNAs indicates that biogenesis may occur in a non-canonical Dicer-independent manner. Other groups have also reported svRNAs derived from SARS viruses lacking a 5′-U sequence signature and it has been suggested that these small RNAs may utilize the slicer activity of Ago2 for biogenesis rather than the traditional miRNA/siRNA processing machinery, similar to what has been observed for the well-studied human miR-451[10,12,17]. Morales and colleagues observed svRNA expression following knockdown of the RNase III nucleases Drosha and Dicer[12], suggesting these canonical enzymes are not necessary for svRNA biogenesis. We also identified svRNAs derived from ORF7a and the N gene that overlap with svRNAs reported by the Meng, Cheng and Pawlica groups, but with varied starting coordinates and/or sizes. This may suggest that isomiRs exist for svRNAs, potentially caused by alternative processing events such as 5′/3′ trimming variants due to imprecise cleavage of precursors[39]. Further studies on svRNAs derived from SARS-CoV-2 are necessary to elucidate key players in their biogenesis and to validate their potential to interact with host transcripts to affect host gene expression.

Based on our observations, we propose a functional role for svRNAs in which deregulation of the host mRNA expression program may be a direct effect of svRNAs sequestering or "sponging" host miRNAs (Fig. 4). There are many examples of known miRNA sponges in the literature, such as *HSUR 1* and *HSUR 2* from *Herpesvirus saimiri*[40], and it is entirely possible that the svRNAs presented here may function similarly[41–43]. An efficient way to post-transcriptionally alter a gene network or biological system would be to sequester/act on regulators of the network itself, rather than individual genes separately. As canonical human miRNAs often have hundreds of predicted target genes each, production of svRNAs that interfere with this regulation may confer an evolutionary advantage for SARS-CoV-2 and other viruses by modulating host gene expression on a global scale, increasing the likelihood of viral propagation. It has been previously suggested that viral RNA derived from SARS-CoV-2 may act as a host miRNA sponge to aid in evading the host immune response and our findings add more credence to this hypothesis[44]. Based on RNAhybrid seed-target predictions, 22 of 33 differentially expressed host miRNAs are capable of targeting svRNAs, 11 of which are upregulated during the time course. It is possible that in addition to svRNAs sponging host miRNAs, they also directly target miRNAs with or without RISC, preventing the host miRNAs from interacting with endogenous targets (Fig. 4). These potential svRNA-miRNA interactions are not mutually exclusive and may occur simultaneously, ultimately leading to the same outcome: global changes in host gene expression. Direct interactions between svRNAs and their predicted targets must still be validated experimentally to assess the full extent to which these interactions occur during viral infection and also to tease apart the actual targeting mechanism itself. In the future, these interactions could be easily antagonized or blocked using locked nucleic acids (LNAs) or other methods and potentially used as targets for the therapeutic treatment of patients infected with the virus.

STAR methods

Cell culture and viruses.
African green monkey kidney (Vero-E6) cells and human lung epithelial (Calu-3) cells were obtained from the ATCC (Manassas, VA). The cells were cultured in Dulbecco's modified Eagle's medium (ThermoFisher Scientific, Waltham, MA, USA) containing 1% heat inactivated fetal bovine serum (ThermoFisher Scientific), 100 U/mL penicillin, and 100 μg/mL streptomycin. The cells were incubated in 95% air and 5% CO_2 at 37 °C.

Virus infection and total RNA extraction.
Vero-E6 and Calu-3 cells (2×10^5 cells/well in 6-well plates) were cultured in DMEM containing 1% FBS at 37 °C in a CO_2 incubator overnight. The cells were washed once with phosphate-buffered saline (PBS), prototypic SARS-CoV-2 Washington strain (USA/WA1/2020) in growth media at MOI 0.01 was added into each desired well, and the plates were incubated for 1.5 h at 37 °C in a CO_2 incubator. After incubation, 3 mL of DMEM containing 1% FBS was added to each well and the plates were incubated at 37 °C in a CO_2 incubator. Infected Calu-3 cells were harvested at 0 h, 12 h, 24 h, 48 h and 72 hpi (in triplicate) and Vero-E6 cells were harvested at 4 h, 24 h, 48 h, 72 h, 120 h, 168 h and 216 hpi (in duplicate).

For harvesting, cell culture supernatants were collected and mixed with Trizol (Fisher Scientific) at a ratio of 1:1. Cells were detached using Trypsin–EDTA 0.25% (ThermoFIsher Scientific), followed by centrifugation at 1000 rpm for 5 min at 4 °C to form a cell pellet which in turn was dissolved in 250 μl of Trizol. Both cell supernatant in Trizol and cell pellet in Trizol were stored at – 80 °C until further processing. SARS-CoV-2 amplification and cell culture procedures were performed according to biosafety level 3 (BSL-3) conditions.

Total RNA was extracted using the TRI Reagent protocol for isolation of RNA with the following modification: one additional RNA ethanol wash step was included. After the total RNA was solubilized in ddH_2O, one overnight ethanol precipitation step was included for further purification of the total RNA.

Illumina sequencing of sRNA libraries. Total RNA was isolated from cell culture pellets as described above. Total RNA from infected Calu-3 cells, infected Vero-E6 cells and uninfected controls were used for sRNA library preparation. 1 µg of total RNA for each sample was used for sRNA library preparation using the NEXT-FLEX small RNA-Seq Kit v3 following the manufacturer's protocol (Perkin Elmer Applied Genomics). Adapter dilution was not performed and 17 cycles of PCR amplification were used before using the gel-free size selection and cleanup protocol. Pooled sRNA sequencing libraries were sequenced on an Illumina HiSeq 4000 at the UC Davis Sequencing Core Facility, generating 100 bp single-end reads.

Illumina sequencing of mRNA libraries. Total RNA was isolated for cell culture pellets as described above. Total RNA from infected Calu-3 cells, infected Vero-E6 cells and uninfected controls were used for mRNA library preparation. 2 µg of total RNA for each sample was used for mRNA library preparation using the NEXT-FLEX Rapid Directional RNA-Seq Kit 2.0 following the manufacturer's protocol (Perkin Elmer Applied Genomics). Before library preparation, total RNA samples were subjected to Poly(A) selection and purification using the NEXTFLEX Poly(A) Beads Kit 2.0 following the manufacturer's protocol (Perkin Elmer Applied Genomics). Based on sample RNA integrity numbers (RINS) and total RNA input, 9 min fragmentation time was used, adapter concentration was diluted by half and 9 cycles of PCR amplification were used before bead clean up and elution. For Calu-3 cell samples, pooled mRNA sequencing libraries were sequenced on an Illumina NovaSeq S4 at the UC Davis Sequencing Core Facility, generating 150 bp paired-end reads. For Vero-E6 samples, pooled mRNA sequencing libraries were sequenced on an Illumina HiSeq 4000, generating 100 bp single-end reads.

Analyses of Calu-3 mRNA sequencing data. The full-length paired-end reads were mapped with Bowtie 2[45] against the set of human repeat-masker elements in the hg38 assembly, as well as a constant poly-A sequence. Reads that were mapped to these elements were removed from further processing. The repeat-filtered full-length reads were mapped with STAR 2.5.3a[46], using the "–alignEndsType EndToEnd" option, to a single joint target consisting of the hg38 genome assembly plus the wuhCor1 genome. In cases of multiply-mapped reads, only the best mapping was retained. The STAR mapped bam files were divided into separate hg38 and wuhCor1 files for further analysis.

Gene-by-gene coverage was extracted from the hg38 mappings that overlapped any potential exon for each gene in a gene model of hg38. The total coverage for each gene was divided by the total paired-end read length of each mapping to extract read counts as input to DESeq2[25]. DESeq2 was run to compare all replicates of conditions as follows:

1. At each time point, the control vs the infected samples
2. For the controls, comparisons between the time points
3. For the infected samples, comparisons between the time points.

Analyses of Vero-E6 mRNA sequencing data. The single-end reads were trimmed at the 3′ end to leave 50 bp for mapping. The trimmed reads were mapped with Bowtie 2[45] against a set of repeat elements in the Vero-E6 genome. The trimmed, repeat-filtered and wuhCor1-filtered, single-end reads were mapped with STAR 2.5.3a, using the "–alignEndsType EndToEnd" option, to chlSab2 genome assembly. In cases of multiply-mapped reads, only the best mapping was retained. Potential PCR duplicates (single-end reads mapped to the identical coordinates) were removed from the chlSab2 mappings.

Gene-by-gene coverage was extracted from the chlSab2 mappings that overlapped any potential exon for each gene in a gene model of chlSab2. The total coverage for each gene was divided by the single-end read length of each mapping to extract read counts as input to DESeq2.

DESeq2 was run to compare all replicates of conditions:

1. At each time point, the control vs the infected samples
2. For the controls, comparisons between the time points
3. For the infected samples, comparisons between the time points.

Gene ontology (GO) and KEGG human pathway analyses. GO process annotations and KEGG human pathway annotations were retrieved using the web tool Enrichr[47–50]. The p-values of enrichment of GO process and KEGG pathways were determined using Enrichr on differentially expressed upregulated and down-regulated genes from both Calu-3 and Vero-E6 data sets.

svRNA discovery from sRNA-seq data. Potential svRNAs were identified from sRNA-seq reads using a stepwise bioinformatic approach which incorporates established tools and bespoke analysis methods. Prerequisite packages include: STAR v2.7.8a[46], BEDOPS v2.4.40[51], bedtools v2.28.0[52], Piranha v1.2.1[26], Bowtie 2 v2.4.4[45] and RNAhybrid v2.1.2[29]. The svRNA-discovery pipeline entails the following steps:

1. sRNA-seq libraries were first mapped against their respective host genomes (GRCh38 for human, chlSab2 for African Green Monkey) and against the SARS-CoV-2 genome (wuhCor1) using STAR[53].
2. Mapping files from the infected samples from Calu-3 and Vero-E6 were subsetted for sequences between 20 and 30 nt in length that mapped to the SARS-CoV-2 genome. Piranha peaks were called for pooled reads from infected samples for each time point in each of the two cell types using bins of size 50 and a background threshold of 0.95.

3. Piranha peaks were pooled for each cell type, and unique reads mapping to them were aligned to their respective host genomes using Bowtie 2 to remove any potential multi-mapping or host-mapped reads from downstream analysis.
4. Of the remaining sequences, those that met our count-based cutoffs were filtered for sequences whose counts across infected replicates and time points were the highest among overlapping sequences. Some overlapping sequences whose counts were especially relatively high were retained. The count-based cutoffs for a given sequence were as follows:

    a. It must be present in at least 5 infected samples for Calu-3 and at least 4 for Vero-E6 (due to the difference in replicates).
    b. Its total raw counts from infected samples must be at least 20.
    c. If present in any control samples, it must not be present in all control samples.
    d. If present in any control samples, its total raw counts from control samples must be less than or equal to 10% of the total raw counts from infected samples.

5. Duplexes between host transcripts (all expressed mRNAs and miRNAs) and SARS-CoV-2-derived sequences from Step 4 were then predicted using RNAhybrid[28,29], which is a tool that is typically used to predict the hybridization, with the smallest minimum free energy, of a short RNA molecule with a longer one. SARS-CoV-2-derived sequences with at least 1 statistically significant (p-value < 0.05) predicted duplex with a host transcript were then considered potential SARS-CoV-2-derived svRNAs. The resultant target predictions were then subsetted for those involving host targets that were significantly differentially expressed (p-value < 0.05, $\log_2$ fold change ≥ 1.0).
6. Potential svRNAs from both cell types with overlaps to each other (across cell types) greater than or equal to 60% were categorized as svRNAs derived from "conserved loci". We define conserved loci as loci in the SARS-CoV-2 genome from which potential svRNAs were detected in both the Calu-3 and Vero-E6 experiments. Additionally, overlapping svRNAs within cell type (i.e. shifted 1-2nt down the SARS-CoV-2 genome sequence) that were retained due to their high relative counts were denoted as potential viral isomiRs—in this case being svRNAs from the same approximate locus but with slightly different ends owed to imprecise precursor cleavage.
7. Lastly, to investigate the potential of identified svRNAs to be products of hairpin precursors, we used RNAfold[27] to predict the structure of each svRNA with the addition of 30–45 nt upstream and downstream separately. In order to call potential hairpin precursors for svRNAs, the structure predictions needed to have: only a single loop—with no shared nucleotides with the mature svRNA sequence, at least 14 base pairings in the stem, and a maximum minimum free energy (MFE) of − 4.3 kcal/mol[54]. Hairpins that passed these filters were visualized using Forna[55].

Data availability
All RNA-sequencing data can be accessed at the National Center for Biotechnology Institute (NCBI) Gene Expression Omnibus (GEO) database: GSE197521.

Code availability
No new code was developed for this study.

References
1. Michelini, F., Jalihal, A. P., Francia, S., Meers, C., Neeb, Z. T., Rossiello, F., Gioia, U., Aguado, J., Jones-Weinert, C., Luke, B., Biamonti, G., Nowacki, M., Storici, F., Carninci, P., Walter, N. G. & d'Adda di Fagagna, F. From 'Cellular' RNA to 'Smart' RNA: Multiple Roles of RNA in genome stability and beyond. *Chem. Rev.* **118,** 4365–4403 (2018).
2. Moazed, D. Small RNAs in transcriptional gene silencing and genome defence. *Nature* **457,** 413–420 (2009).
3. Alshaer, W. *et al.* siRNA: Mechanism of action, challenges, and therapeutic approaches. *Eur. J. Pharmacol.* **905,** 174178 (2021).
4. Ergin, K. & Çetinkaya, R. in 1–32 (2022).
5. Iwasaki, Y. W., Siomi, M. C. & Siomi, H. PIWI-interacting RNA: Its biogenesis and functions. *Annu. Rev. Biochem.* **84,** 405–433 (2015).
6. Allen, S. E. & Nowacki, M. Roles of noncoding RNAs in ciliate genome architecture. *J. Mol. Biol.* **432,** 4186–4198 (2020).
7. Furrer, D. I., Swart, E. C., Kraft, M. F., Sandoval, P. Y. & Nowacki, M. Two sets of piwi proteins are involved in distinct sRNA pathways leading to elimination of germline-specific DNA. *Cell Rep.* **20,** 505–520 (2017).
8. Hoehener, C., Hug, I. & Nowacki, M. Dicer-like enzymes with sequence cleavage preferences. *Cell* **173,** 234-247.e7 (2018).
9. Sandoval, P. Y., Swart, E. C., Arambasic, M. & Nowacki, M. Functional diversification of dicer-like proteins and small RNAs required for genome sculpting. *Dev. Cell* **28,** 174–188 (2014).
10. Cheloufi, S., Dos Santos, C. O., Chong, M. M. W. & Hannon, G. J. A dicer-independent miRNA biogenesis pathway that requires Ago catalysis. *Nature* **465,** 584–589 (2010).
11. Valencia-Sanchez, M. A., Liu, J., Hannon, G. J. & Parker, R. Control of translation and mRNA degradation by miRNAs and siRNAs: Table 1. *Genes Dev.* **20,** 515–524 (2006).
12. Morales, L. *et al.* SARS-CoV-encoded small RNAs contribute to infection-associated lung pathology. *Cell Host Microbe* **21,** 344–355 (2017).
13. Perez, J. T., Varble, A., Sachidanandam, R., Zlatev, I., Manoharan, M., Garcia-Sastre, A. & tenOever, B. R. Influenza A virus-generated small RNAs regulate the switch from transcription to replication. *Proc. Natl. Acad. Sci.* **107,** 11525–11530 (2010).
14. Shi, J. *et al.* Novel microRNA-like viral small regulatory RNAs arising during human hepatitis A virus infection. *FASEB J.* **28,** 4381–4393 (2014).

15. Weng, K.-F. *et al.* A cytoplasmic RNA virus generates functional viral small RNAs and regulates viral IRES activity in mammalian cells. *Nucleic Acids Res.* **42**, 12789–12805 (2014).
16. Meng, F. *et al.* Viral MicroRNAs encoded by nucleocapsid gene of SARS-CoV-2 are detected during infection, and targeting metabolic pathways in host cells. *Cells* **10**, 1762 (2021).
17. Cheng, Z., Cheng, L., Lin, J., Lunbiao, C., Chunyu, L., Guoxin, S., Rui, X., Xiangnan, G., Changxing, L., Yan, C., Baoli, Z. & Wei, Z. Verification of SARS-CoV-2-Encoded small RNAs and contribution to Infection-Associated lung inflammation. *bioRxiv* 2021.05.16.444324 (2021).
18. Pawlica, P., Yario, T. A., White, S., Wang, J., Moss, W. N., Hui, P., Vinetz, J. M. & Steitz, J. A. SARS-CoV-2 expresses a microRNA-like small RNA able to selectively repress host genes. *Proc. Natl. Acad. Sci. U. S. A.* **118,** (2021).
19. Withers, J. B. *et al.* Idiosyncrasies of viral noncoding RNAs provide insights into host cell biology. *Annu. Rev. Virol.* **6**, 297–317 (2019).
20. Park, B. K. *et al.* differential signaling and virus production in Calu-3 cells and vero cells upon SARS-CoV-2 infection. *Biomol. Ther.* **29**, 273–281 (2021).
21. Chu, H., Chan, J. F.-W., Yuen, T. T.-T., Shuai, H., Yuan, S., Wang, Y., Hu, B., Yip, C. C.-Y., Tsang, J. O.-L., Huang, X., Chai, Y., Yang, D., Hou, Y., Chik, K. K.-H., Zhang, X., Fung, A. Y.-F., Tsoi, H.-W., Cai, J.-P., Chan, W.-M., Ip, J. D., Chu, A. W.-H., Zhou, J., Lung, D. C., Kok, K.-H., To, K. K.-W., Tsang, O. T.-Y., Chan, K.-H. & Yuen, K.-Y. Comparative tropism, replication kinetics, and cell damage profiling of SARS-CoV-2 and SARS-CoV with implications for clinical manifestations, transmissibility, and laboratory studies of COVID-19: an observational study. *The Lancet Microb.* **1**, e14–e23. https://doi.org/10.1016/s2666-5247(20)30004-5 (2020)
22. Zaliani, A. *et al.* Cytopathic SARS-CoV-2 screening on VERO-E6 cells in a large-scale repurposing effort. *Sci. Data* **9**, 405 (2022).
23. Ogando, N. S. *et al.* SARS-coronavirus-2 replication in Vero E6 cells: Replication kinetics, rapid adaptation and cytopathology. *J. Gen. Virol.* **101**, 925–940 (2020).
24. Kim, D. *et al.* The architecture of SARS-CoV-2 transcriptome. *Cell* **181**, 914-921.e10 (2020).
25. 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).
26. Uren, P. J. *et al.* Site identification in high-throughput RNA–protein interaction data. *Bioinformatics* **28**, 3013–3020 (2012).
27. Mathews, D. H. *et al.* Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 7287–7292 (2004).
28. Kruger, J. & Rehmsmeier, M. RNAhybrid: microRNA target prediction easy, fast and flexible. *Nucleic Acids Res.* **34**, W451–W454 (2006).
29. Rehmsmeier, M., Steffen, P., Höchsmann, M. & Giegerich, R. Fast and effective prediction of microRNA/target duplexes. *RNA* **10**, 1507–1517 (2004).
30. DeDiego, M. L. *et al.* Inhibition of NF- B-mediated inflammation in severe acute respiratory syndrome coronavirus-infected mice increases survival. *J. Virol.* **88**, 913–924 (2014).
31. Nelemans, T. & Kikkert, M. viral innate immune evasion and the pathogenesis of emerging RNA virus infections. *Viruses* **11**, 961 (2019).
32. Xiong, Y. *et al.* Transcriptomic characteristics of bronchoalveolar lavage fluid and peripheral blood mononuclear cells in COVID-19 patients. *Emerg. Microbes Infect.* **9**, 761–770 (2020).
33. Chen, Y., Liu, Q. & Guo, D. Emerging coronaviruses: Genome structure, replication, and pathogenesis. *J. Med. Virol.* **92**, 418–423 (2020).
34. Li, C.-X. *et al.* Whole-transcriptome rna sequencing reveals significant differentially expressed mRNAs, miRNAs, and lncRNAs and related regulating biological pathways in the peripheral blood of COVID-19 patients. *Mediators Inflamm.* **2021**, 1–22 (2021).
35. Ajaz, S. *et al.* Mitochondrial metabolic manipulation by SARS-CoV-2 in peripheral blood mononuclear cells of patients with COVID-19. *Am. J. Physiol. Cell Physiol.* **320**, C57–C65 (2021).
36. Gassen, N. C. *et al.* SARS-CoV-2-mediated dysregulation of metabolism and autophagy uncovers host-targeting antivirals. *Nat. Commun.* **12**, 3818 (2021).
37. Bruzzone, C., Bizkarguenaga, M., Gil-Redondo, R., Diercks, T., Arana, E., García de Vicuña, A., Seco, M., Bosch, A., Palazón, A., San Juan, I., Laín, A., Gil-Martínez, J., Bernardo-Seisdedos, G., Fernández-Ramos, D., Lopitz-Otsoa, F., Embade, N., Lu, S., Mato, J. M. & Millet, O. SARS-CoV-2 infection dysregulates the metabolomic and lipidomic profiles of serum. *iScience* **23**, 101645 (2020).
38. Mirzaei, R. *et al.* The emerging role of microRNAs in the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection. *Int. Immunopharmacol.* **90**, 107204 (2021).
39. Morin, R. D. *et al.* Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. *Genome Res.* **18**, 610–621 (2008).
40. Cazalla, D., Yario, T. & Steitz, J. A. Down-regulation of a host microRNA by a Herpesvirus saimiri noncoding RNA. *Science* **328**, 1563–1566 (2010).
41. Ebert, M. S. & Sharp, P. A. MicroRNA sponges: Progress and possibilities. *RNA* **16**, 2043–2050 (2010).
42. Zhou, Q. *et al.* Novel insights into MALAT1 function as a MicroRNA sponge in NSCLC. *Front. Oncol.* **11**, 758653 (2021).
43. Xiao, J. *et al.* Long noncoding RNA TRPM2-AS acts as a microRNA sponge of miR-612 to promote gastric cancer progression and radioresistance. *Oncogenesis* **9**, 29 (2020).
44. Bartoszewski, R. *et al.* SARS-CoV-2 may regulate cellular responses through depletion of specific host miRNAs. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **319**, L444–L455 (2020).
45. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
46. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
47. Chen, E. Y., Tan, C. M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G. V., Clark, N. R. & Ma'ayan, A. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128 (2013).
48. Kuleshov, M. V., Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S. L., Jagodnik, K. M., Lachmann, A., McDermott, M. G., Monteiro, C. D., Gundersen, G. W. & Ma'ayan, A. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44,** W90–W97 (2016).
49. Xie, Z., Bailey, A., Kuleshov, M. V., Clarke, D. J. B., Evangelista, J. E., Jenkins, S. L., Lachmann, A., Wojciechowicz, M. L., Kropiwnicki, E., Jagodnik, K. M., Jeon, M. & Ma'ayan, A. Gene set knowledge discovery with Enrichr. *Curr. Protocols* **1** (2021).
50. Chen, E. Y. Enrichr. at https://maayanlab.cloud/Enrichr.
51. Neph, S. *et al.* BEDOPS: High-performance genomic feature operations. *Bioinformatics* **28**, 1919–1920 (2012).
52. Quinlan, A. R. & Hall, I. M. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
53. Kent, W. J. *et al.* The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).
54. Stepanowsky, P., Levy, E., Kim, J., Jiang, X. & Ohno-Machado, L. Prediction of MicroRNA precursors using parsimonious feature sets. *Cancer Inform.* **13**, 95–102 (2014).
55. Kerpedjiev, P., Hammer, S. & Hofacker, I. L. Forna (force-directed RNA): Simple and effective online RNA secondary structure diagrams. *Bioinformatics* **31**, 3377–3379 (2015).

Author contributions

Conceptualization, Z.T.N., N.M., and J.R.S.; methodology, Z.T.N., N.M., J.R.S.; software, A.G. and S.K.; formal analysis, A.G. and S.K.; investigation, Z.T.N. and L.V.C.; resources, N.M. and J.R.S.; data curation, A.G. and S.K.; writing—original draft, Z.T.N.; writing—review and editing, Z.T.N., A.G., L.V.C., S.K., W.I.L., N.M., and J.R.S.; visualization, Z.T.N. and A.G.; supervision, W.I.L., N.M. and J.R.S.; funding acquisition, N.M. and J.R.S.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-022-26135-9.

Correspondence and requests for materials should be addressed to N.M. or J.R.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.