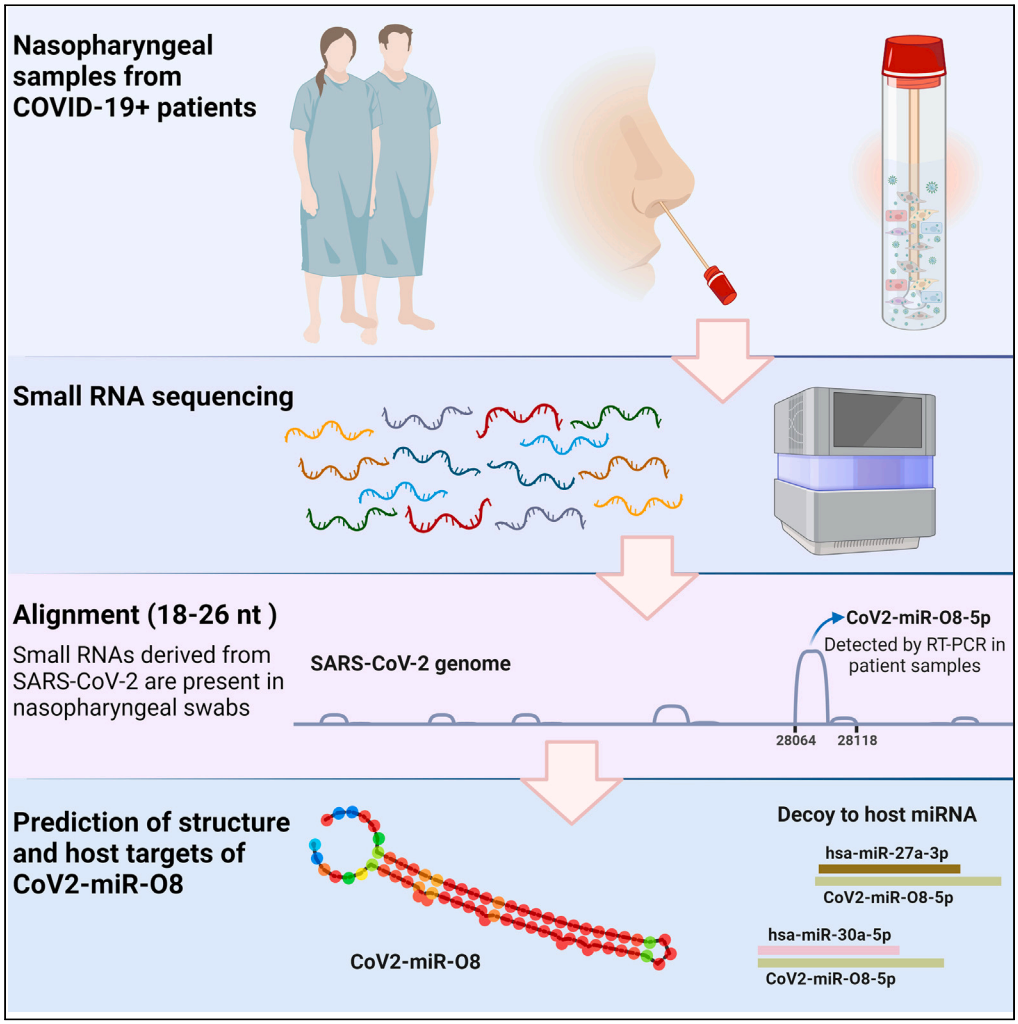
iScience

CellPress
OPEN ACCESS

Article

SARS-CoV-2 produces a microRNA CoV2-miR-O8 in patients with COVID-19 infection



Elise J. Tucker, Soon Wei Wong, Shashikanth Marri, ..., Jordan Y. Li, Chuan Kok Lim, Jonathan M. Gleadle

jonathan.gleadle@flinders.edu.au

Highlights

Small RNA CoV2-miR-O8 is present in nasopharyngeal swabs from patients with COVID-19

CoV2-miR-O8 has features of a microRNA forming a hairpin and binding Argonaute

CoV2-miR-O8 interacts with human microRNAs acting as a potential decoy

Tucker et al., iScience 27, 108719
January 19, 2024
https://doi.org/10.1016/j.isci.2023.108719

iScience

CellPress
OPEN ACCESS

Article

SARS-CoV-2 produces a microRNA CoV2-miR-O8 in patients with COVID-19 infection

Elise J. Tucker,[1,2,5] Soon Wei Wong,[1,2,5] Shashikanth Marri,[2] Saira Ali,[1,2] Anthony O. Fedele,[1] Michael Z. Michael,[2,3] Darling Rojas-Canales,[1,2] Jordan Y. Li,[1,2] Chuan Kok Lim,[4] and Jonathan M. Gleadle[1,2,6,*]

SUMMARY

Many viruses produce microRNAs (miRNAs), termed viral miRNAs (v-miRNAs), with the capacity to target host gene expression. Bioinformatic and cell culture studies suggest that SARS-CoV-2 can also generate v-miRNAs. This patient-based study defines the SARS-CoV-2 encoded small RNAs present in nasopharyngeal swabs of patients with COVID-19 infection using small RNA-seq. A specific conserved sequence (CoV2-miR-O8) is defined that is not expressed in other coronaviruses but is preserved in all SARS-CoV-2 variants. CoV2-miR-O8 is highly represented in nasopharyngeal samples from patients with COVID-19 infection, is detected by RT-PCR assays in patients, has features consistent with Dicer and Drosha generation as well as interaction with Argonaute and targets specific human microRNAs.

INTRODUCTION

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is an RNA virus that causes coronavirus disease 2019 (COVID-19). Coronaviruses primarily target the human respiratory system, and the family includes SARS-CoV-2, severe acute respiratory syndrome (SARS-CoV) and Middle East respiratory syndrome (MERS-CoV). SARS-CoV-2 causes a range of symptoms, and disease severity and fatality rates are largely dependent on viral strain, age, co-morbidities and vaccination status.

Several studies have demonstrated that DNA and RNA viruses, including SARS-CoV, HIV, and Ebola, can produce microRNAs (miRNAs), termed viral miRNAs (v-miRNAs).[1,2] MiRNAs are small non-coding RNAs with the ability to post-transcriptionally regulate target gene expression. v-miRNAs are usually generated by the host Drosha and Dicer machinery. The v-miRNA generates a primary miRNA (pri-miRNA) that usually forms a hairpin secondary structure and is then processed by Drosha/DiGeorge syndrome (DGCR8) protein to generate precursor miRNA (pre-miRNA). The pre-miRNA is additionally processed by Dicer to yield the mature miRNA duplex (21–25 nt). One strand of this duplex is loaded into the RNA-induced silencing complex (RISC) containing Argonaute (Ago), while the other passenger or star strand is degraded. The miRNA-RISC complex interacts with target transcripts to inhibit gene expression, though examples also exist of enhancement of target mRNA abundance.[3] The extent to which RNA viruses generate viral miRNAs and the mechanism and site of generation are, however, still incompletely understood and Dicer and Drosha independent mechanisms of generation have been described.

Do coronaviruses encode miRNAs? Morales et al.[2] identified three SARS-CoV virally encoded novel small RNAs (smRNA) in a small RNA-seq analysis of SARS-CoV infected mouse lungs. These smRNAs included svRNA-N, svRNA-nsp3.1 and svRNA-nsp3.2.[2] Importantly, they demonstrated that inhibition of svRNA-N significantly reduced *in vivo* lung pathology and cytokine expression suggesting that svRNAs may contribute to SARS-CoV pathogenesis. This raises the important question of whether SARS-CoV-2 encodes smRNAs which might have pathogenetic functions. Bioinformatic analyses have revealed that one of the SARS-CoV encoded miRNAs, svRNA-N, is highly conserved in SARS-CoV-2[4,5]. Additionally, bioinformatic studies of SARS-CoV-2 have also attempted to predict SARS-CoV-2 encoded miRNAs.[5–8] Merino et al. identified 12 candidate stem-loop structures in the SARS-CoV-2 viral genome and additionally confirmed the expression of six pre-miRNAs generating eight mature miRNA-like sequences in small RNA-seq data from SARS-CoV-2 infected cultures of the epithelial lung cancer cell line Calu-3.[7] Saini et al. suggested 26 putative mature SARS-CoV-2 encoded miRNAs[8] whilst Aydemir et al. predicted 40 SARS-CoV-2 v-miRNAs.[6] Importantly, Meng et al. defined three virally encoded short RNAs which could be detected by RT-PCR in nasopharyngeal swabs from patients with COVID-19 infection[5] indicating that the bioinformatically predicted sequences may be generated *in vivo*, however, their biological activity or relevance remains unknown.

Recently, Singh et al. analyzed SARS-CoV-2 infected cells in culture for the presence of putative viral encoded smRNAs.[9] The characteristics of the majority of such sequences (length and distribution throughout the viral genome) were interpreted as degradation fragments rather

[1]Department of Renal Medicine, Flinders Medical Centre, SA, Australia
[2]College of Medicine and Public Health, Flinders University, SA, Australia
[3]Department of Gastroenterology, Flinders Medical Centre, SA, Australia
[4]Infectious Diseases Laboratories, SA Pathology, SA, Australia
[5]These authors contributed equally
[6]Lead contact
*Correspondence: jonathan.gleadle@flinders.edu.au
https://doi.org/10.1016/j.isci.2023.108719

than Dicer generated miRNAs. However, two 22 nucleotide sequences (CoV2-miR-O7a.1 and CoV2-miR-O7a.2) present in virally infected cells were mapped to the beginning of ORF7a, required Dicer for generation in cell culture, could associate with Argonaute and were detected by RT-PCR in nasopharyngeal swabs of several patients with COVID-19 infection.[9] In a similar approach, Pawlica et al.[10] identified the same SARS-CoV-2 encoded smRNA (CoV2-miR-O7a) in RNA sequencing analysis of infected lung derived cell lines. This sequence associated with Argonaute, could be processed in cells, demonstrated Drosha-independent generation and was also detected by RT-PCR in nasopharyngeal samples from SARS.[10]

Since the initial identification of the SARS-CoV-2 virus an enormous number of clinical samples have been RNA sequenced, which has led to an understanding of the complexity of the viral transcriptome.[11] However, the method of RNA purification and sequencing can bias against the detection of short transcripts such as miRNAs.[12–14] In order to profile the SARS-CoV-2 encoded smRNAs in clinical specimens and to examine for the *in vivo* existence of these candidate v-miRNAs in clinical infection, we performed small RNA-seq analysis on RNA extracted from nasopharyngeal swabs obtained from patients with COVID-19 infection ascertained by a routine diagnostic RT-PCR assay.[15] In this work we detail the actual SARS-CoV-2 encoded smRNA transcriptome in nasopharyngeal swabs, suggest the existence of SARS-CoV-2 v-miRNAs in human infection, validate the presence of smRNA in infected patients and explore potential mRNA and microRNA targets.

RESULTS

Small RNAs derived from SARS-CoV-2 are present in nasopharyngeal swabs

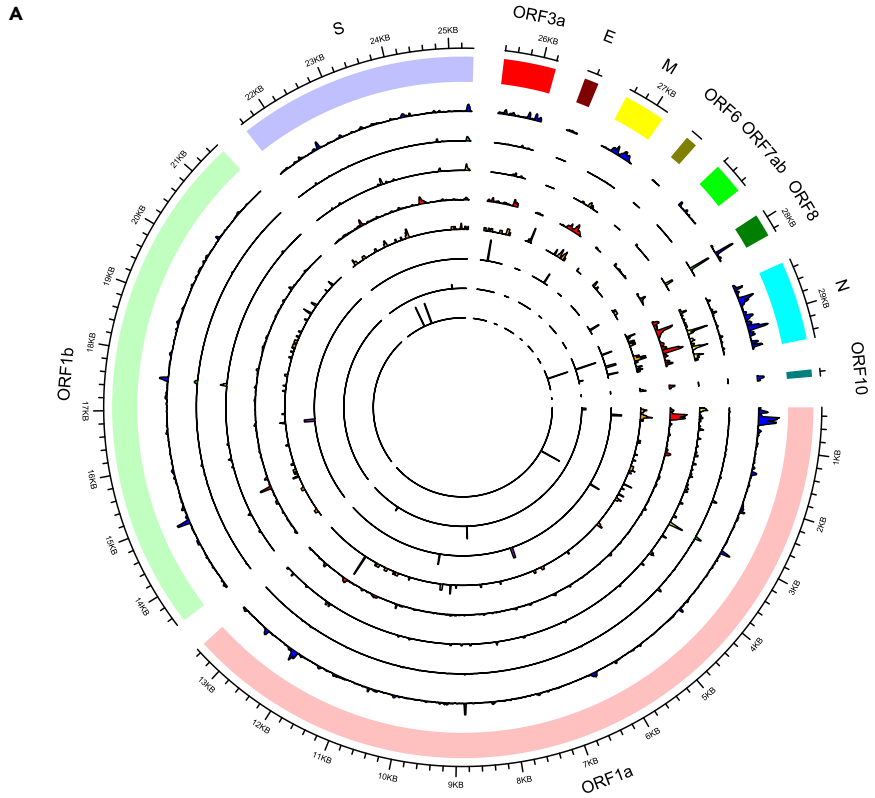
Initially, ten nasopharyngeal swab samples were obtained from patients with COVID-19 infection (determined by RT-PCR positivity using the diagnostic envelope (E) gene TaqMan qPCR assay[15] by SA Pathology). Nine of the samples yielded RNA but one (sample 4) did not and was therefore excluded from the subsequent small RNA-seq. The processed reads from each sample varied from 6 to 85 million, and after quality filtering a combined total of ~174 million reads remained (Table S1). The samples were analyzed for reads of human origin, SARS-CoV-2 origin (Tables S1 and S2) and reads corresponding to human miRNAs (miRBase V22 and genome release GRCh38) (Table S3). The percentage of reads mapping to the human genome varied from 25 to 96% across the different patient samples. The samples also varied substantially in the number of reads mapping to the SARS-CoV-2 genome from no reads (sample 6) to several thousand (sample 1, 1584 reads; sample 2, 1958 reads; and sample 5, 1802 reads) with on average 0.002% of smRNAeq reads mapping to the SARS-CoV-2 genome across the 9 patient samples (Table S1). Comparison of normalized (CPM) read counts highlighted that the reads were not uniformly distributed across the SARS-CoV-2 genome but were clustered to specific regions, notably to the 5′ and 3′ regions of the genome (Figure 1A). The pattern of sequencing reads was closely shared between the different samples (e.g., compare samples 1, 2, 3, 5 and 7 in Figure 1A) indicating a non-random detection or abundance of sequences. The number of reads mapping to each region of the SARS-CoV-2 genome varied significantly with the greatest numbers of reads mapping to ORF1ab and N (Table S1), followed by S and ORF8, while other regions showed no or minimal reads (E, ORF6, ORF7b, ORF10). The size distribution of the SARS-CoV-2 encoded smRNA reads is shown in Figure 1B with a peak at 22–23 nucleotides in keeping with the usual length of miRNAs.

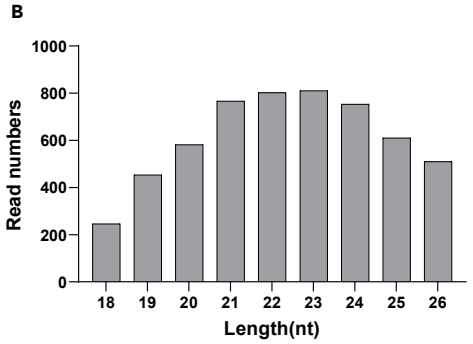
CoV2-miR-O8 (SARS-CoV-2 miRNA-like ORF8-derived) small RNA is a highly expressed and conserved small RNA sequence

The SARS-CoV-2 encoded smRNA sequences with the greatest number of reads are shown in Figure 1C. Interestingly, 7 of the 10 of the most highly represented sequences corresponded to genome position 28064–28088 within ORF8, albeit with variation in lengths within that sequence from 18 to 24 nucleotides. This highly represented region was therefore analyzed in more detail and is subsequently referred to as CoV2-miR-O8 (SARS-CoV-2 miRNA-like ORF8-derived smRNA). To examine for the presence and conservation of this sequence in related viruses and in different strains of SARS-CoV-2, the presence of the CoV2-miR-O8 sequence was determined in the reference sequence for SARS-CoV-2 and in other coronaviruses (Figures 1D and S1). The sequence was perfectly conserved in the reference sequence (NC_045512.2), and in the Alpha (B.1.1.7), Beta (B.1.351), Gamma (P.1), Delta (B.617.2), and Omicron (B.1.1.529, BA.4, BA.5) variants (https://genome.ucsc.edu/covid19.html) and no mutations in this sequence were found (Figure 1D). However, there appears to be no orthologous region in other strains of human coronavirus including SARS-CoV-1 and MERS (Figure 1).

CoV2-miR-O8 has characteristics of a microRNA

To explore whether this and other smRNA sequences had other characteristics in keeping with miRNAs, the highly represented sequences and their adjacent sequences were examined for predicted pre-miR duplex structures. The detailed structurome analysis of the SARS-CoV-2 genome by Andrews et al. was utilised.[16] The 524 uniquely stable predicted secondary structures were inspected for small RNA sequence representation in the RNA-seq analysis. Six highly represented smRNA sequences coincided with predicted hairpin structures (Figures 2A and S2). The presence of hairpin structures corresponding to the highly expressed smRNA sequences might indicate relative stability (for instance, by protection from physicochemical or enzymatic-mediated degradation), or the presence of duplex structures capable of Dicer and Drosha processing. However, in examining the representation of these sequences in the smRNA-seq data, there was greater abundance of either the 5p or the 3p arms of the predicted hairpins. This was particularly marked for CoV2-miR-O8 and apparent in all the patient samples (Figure 2B). This is in keeping with the marked differences in 5p and 3p mature miRNA expression seen in miRNAs with asymmetric degradation of the passenger or * strand and highly suggestive of Argonaute mediated generation.





C

Sequence	Genome position	Length (nt)	No. reads	Gene
AATTGAATTGTGCGTGGATGAGGC	28064-28087	24	56	ORF8
AATTGTGCGTGGATGAGGC	28069-28087	19	34	ORF8
ATTGTGCGTGGATGAGGC	28070-28087	18	28	ORF8
ATTGTGCGTGGATGAGGCT	28070-28088	19	26	ORF8
AATTGAATTGTGCGTGGATGAGGCT	28064-28088	25	25	ORF8
TGAATTGTGCGTGGATGAGGC	28067-28087	21	21	ORF8
GTGAAATTGTCGGTGGACA	2235-2253	19	20	ORF1a
ACTTTCGATCTCTTGTAGATCTGTT	38-63	26	19	5′UTR
AGGAACTAATCAGACAAGGAACT	29139-29161	23	18	N
ATTGAATTGTGCGTGGATGAGGCT	28065-28088	24	17	ORF8

D

Reference	Wuhan (NC_045512.2)	AATTGAATTGTGCGTGGATGAGGCT
Variants	Alpha/Beta/Delta/Gamma/Omicron	AATTGAATTGTGCGTGGATGAGGCT
	AATTGAATTGTGCGTGGATGAGGC	
	AATTGTGCGTGGATGAGGC	
	ATTGTGCGTGGATGAGGC	
	ATTGTGCGTGGATGAGGCT	
CoV2-miR-O8	AATTGAATTGTGCGTGGATGAGGCT	
	TGAATTGTGCGTGGATGAGGC	
	ATTGAATTGTGCGTGGATGAGGCT	

Figure 1. SARS-CoV-2 patient nasopharyngeal samples express diverse smRNAs

(A) SARS-CoV-2 patient nasopharyngeal smRNA-seq reads (not size selected) are unevenly distributed across the SARS-CoV-2 genome. The outer ring represents the SARS-CoV-2 viral gene structure (ORF1a, ORF1b, S, ORF3a, E, M, ORF6, ORF7ab, ORF8, N, ORF10) and genome positions (nt), the eight inner rings represent read distribution from patients 1, 2, 3, 5, 7, 8, 9 and 10, respectively.

(B) Frequency distribution of 18–26 nt size selected reads based on length.

(C) Combined total read counts and genome position for the ten most abundant 18–26 nt smRNA sequences from all patient SARS-CoV-2 nasopharyngeal swabs.

(D) The CoV2-miR-O8 sequence and other over-represented sequences in the same region do not contain mutations compared to the reference sequence (NC_045512.2), or Alpha (B.1.1.7), Beta (B.1.351), Gamma (P.1), Delta (B.617.2), and Omicron (B.1.1.529, BA.4, BA.5) variants.

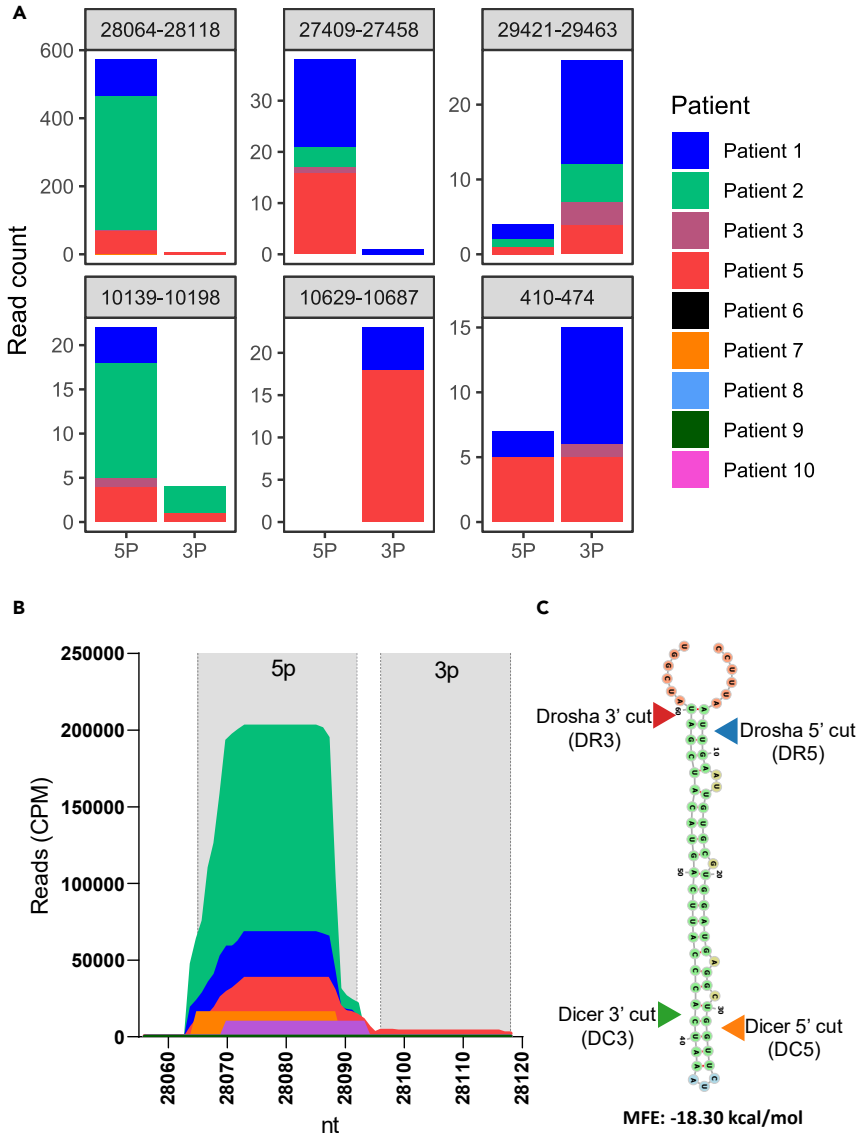


Figure 2. SARS-CoV-2 smRNA sequences coincided with predicted hairpin structures

(A) Highly abundant nasopharyngeal smRNA sequences corresponded to regions of predicted hairpins.

(B) Frequency distribution of reads normalized by counts per million, relative to all 18–26 nt SARS-CoV-2 mapped reads. This distribution pertains to the CoV2-miR-O8 predicted hairpin region.

(C) The RNA secondary structure of CoV2-miR-O8, as determined by RNAfold, is displayed alongside its minimal free energy (MFE) and annotated with predicted Dicer and Drosha cleavage sites.

To examine for the potential role of Argonaute in the generation of this transcript, the Argonaute immunoprecipitation of SARS-CoV-2 infected cells experiment undertaken by Pawlica et al. was examined.[10] In their pan-Argonaute immunoprecipitation experiments, CoV2-miR-O8 sequences were associated with Argonaute in infected A549 cells, though, as with CoV2-miR-O7a and other predicted hairpin structures, to a lesser extent than most host miRNAs (Figure S3). This provides support for the role for this sequence as a miRNA and for Argonaute in its action.

To examine for features consistent with a role for Dicer and Drosha in the generation of the mature CoV2-miR-O8 transcript, the sequences were examined for predicted hairpin generation and Dicer and Drosha cleavage sites. Figure 2C shows the predicted duplex structure and sites of Drosha and Dicer cleavage which in concert would generate a 5p mature miRNA of 23 nucleotides. The predictions of the most probable 5′ and 3′ cleavage sites by Drosha and Dicer are indicated in Figure 2C. However, cleavage can be imprecise and adjacent, less probable cleavage sites can also be predicted. This variability in cleavage site, combined with exonuclease action, is likely to explain the variation in lengths of smRNA sequences derived from this locus that are demonstrated in the smRNA sequencing.[17]

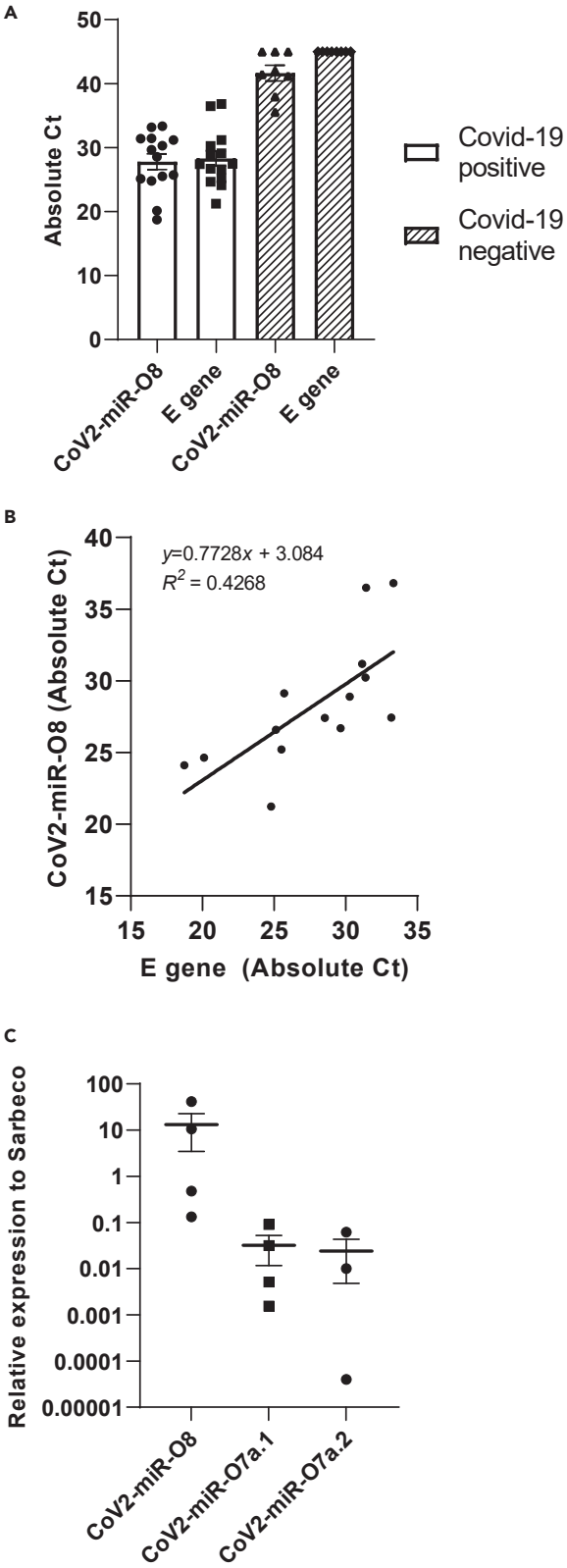

A
Absolute Ct
50
40
30
20
10
0
CoV2-miR-O8
E gene
CoV2-miR-O8
E gene
Covid-19 positive
Covid-19 negative
B
CoV2-miR-O8 (Absolute Ct)
E gene (Absolute Ct)
y=0.7728x + 3.084
R2 = 0.4268
40
35
30
25
20
15
15
20
25
30
35
C
Relative expression to Sarbeco
100
10
1
0.1
0.01
0.001
0.0001
0.00001
CoV2-miR-O8
CoV2-miR-O7a.1
CoV2-miR-O7a.2

Figure 3. CoV2-miR-O8 expression is associated with COVID-19 infection

(A) CoV2-miR-O8 expression was detectable in COVID-19 positive patient nasopharyngeal swab samples and undetectable in COVID-19 negative patient nasopharyngeal swab samples. Results are expressed as means ± standard error of the mean.

(B) CoV2-miR-O8 expression level positively correlates with SARS-CoV-2 viral load as detected by RT-qPCR. Simple linear regression used to plot the line of best fit and R^2 the coefficient of determination and equation of the line indicated.

(C) CoV2-miR-O8 is more abundant than CoV2-miR-O7a.1 and CoV2-miR-O7a.2 in SARS-CoV-2 positive patients. Results are expressed as means ± standard error of the mean.

CoV2-miR-O8 can be detected by RT-PCR in nasopharyngeal swabs of patients with COVID-19 infection

To validate the presence of CoV2-miR-O8 identified from the COVID-19 patient nasopharyngeal small RNA-seq experiment, an additional cohort of patients with and without COVID-19 were examined as well as the original 9 samples from the small RNA-seq experiment. Stem-loop RT-qPCR assays were designed and used to determine the presence of CoV2-miR-O8 in RNA from nasopharyngeal swabs of COVID-19 positive or negative patients as determined by the envelope (E) gene diagnostic RT-PCR assay (Corman et al., 2020). It was possible to detect CoV2-miR-O8 expression in all 15 positive patients (Figure 3A) but in none of the negative patients. There was a significant correlation between the level of expression of CoV2-miR-O8 and the level of expression of the envelope (E) gene diagnostic RT-PCR test. Furthermore, the Ct values of the two RT-PCR assays were at a comparable level (Figure 3B). In a subset of patients, the level of CoV2-miR-O8 in infected patients appeared greater (100-fold) than both putative v-miRNA miR-O7a.1 and miR-O7a.2 (Figure 3C).

To examine for the existence of CoV2-miR-O8 in other settings, the small RNA sequencing in SARS-CoV-2 infected A549 cells undertaken by Pawlica et al.[10] was interrogated for the presence of miR-O8-5p sequences. Multiple transcripts corresponding to miR-O8-5p were observed in keeping with the existence of this v-miRNA in cell culture models of infection (Figure S4).

Interactions of CoV2-miR-O8 with host RNA

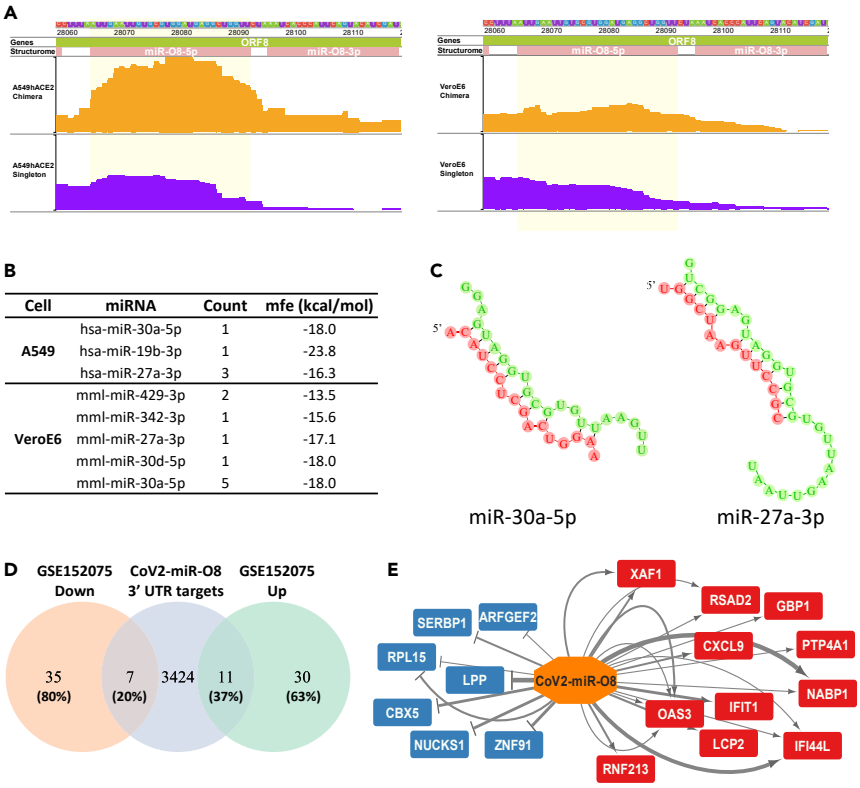
In a recent study, Fossat et al.[18] utilized covalent ligation of endogenous Argonaute-bound RNAs (CLEAR)-cross-linking immunoprecipitation (CLIP) to directly map viral and cellular microRNA interactions during SARS-CoV-2 infection in cultured cells. We utilized their publicly deposited data to further analyze the interactions of CoV2-miR-O8 with human RNA sequences. Analysis of the immunoprecipitated SARS-CoV-2 encoded RNA revealed that multiple sequences corresponding to CoV2-miR-O8 were purified, particularly of the 5′ strand (Figure 4A), providing further support for the presence of this transcript and in keeping with Argonaute binding and microRNA generation. Their data was also examined for the presence of chimeric transcripts containing both CoV-miR-O8 sequence together with human mRNA or microRNA sequences that would indicate direct binding targets of CoV-miR-O8. In keeping with function as a v-miRNA, a limited number of chimeric transcripts were detected consistent with human mRNA targets but were too short in length to unambiguously assign identity. Interestingly, chimeric sequences were identified that contained both CoV-miR-O8 sequence and human microRNA sequences (e.g., miR-30a-5p, and miR-27a-3p) in both A549 and VeroE6 cells, suggesting interaction between these sequences (Figure 4B). To examine the potential nature of this mechanism RNAhybrid analyses were undertaken and predicted highly stable hybrid interactions between these human microRNAs and CoV-miR-O8 (Figure 4C). A functional enrichment analysis of the targets of the human microRNAs shown to interact with CoV2-miR-O8-5p in the Fossat CLEAR-CLIP data[18] was performed (Figure S5) and interestingly the second most significant enrichment was with ‘viral process’.

Prediction of CoV2-miR-O8 interactions with host mRNA

We utilized the miRanda algorithm to predict potential interactions between CoV2-miR-O8 and the 3′ UTR, 5′ UTR and CDS of human host mRNAs. We identified 3442 mRNA targets that could potentially be regulated by CoV2-miR-O8. Whilst the CLEAR-CLIP data had not clearly defined mRNA targets in cell culture models of infection, we wished to explore potential targets in clinical infection. To investigate changes in mRNA targets during COVID-19 infection, we analyzed publicly available nasopharyngeal swab RNA-seq data (GSE186651) for upregulated and downregulated genes during COVID-19 infection. We then overlapped these gene lists with the predicted targets of CoV2-miR-O8 (Table S4). Our findings showed that 20% of the downregulated genes during COVID-19 were predicted targets of CoV2-miR-O8, while 37% of the upregulated genes during COVID-19 were also targets of CoV2-miR-O8 (Figure 4D). We also undertook analysis of predicted targeting of the 5′-UTR and coding sequences but only 3 additional targets were found predicted in the 5′UTRs (RPL6, PDE4B, GBP5), and only one in coding sequence (ZNF91). Interestingly, the genes targeted by CoV2-miR-O8 and upregulated during COVID-19 were associated with the type I interferon signaling pathway, including RSAD2, OAS3, IFIT1, and XAF1 (Figures 4E; Table S5). These results suggest a potential role of CoV2-miR-O8 in regulating antiviral immune responses during COVID-19 infection. The suggestion of upregulation of genes associated with the type I interferon signaling pathway directly by CoV2-miR-O8 is surprising given the more commonly accepted downregulation of target genes and the tendency of viruses to suppress interferon signaling. It requires further experimental validation as does defining targets of other SARS-CoV-2 encoded v-miRNAs.

DISCUSSION

This patient-based study defines the SARS-CoV-2 encoded smRNAs present in nasopharyngeal swabs of patients with COVID-19 infection using smRNA-seq. A specific conserved sequence (CoV2-miR-O8) has been identified that is not expressed by other coronaviruses but is preserved in all SARS-CoV-2 variants. It is highly represented in nasopharyngeal samples from patients with COVID-19 infection and indeed appears more abundant than other smRNA sequences initially defined in cell culture studies. This work is in keeping with the other studies



Cell	miRNA	Count	mfe (kcal/mol)
A549	hsa-miR-30a-5p	1	-18.0
	hsa-miR-19b-3p	1	-23.8
	hsa-miR-27a-3p	3	-16.3
VeroE6	mml-miR-429-3p	2	-13.5
	mml-miR-342-3p	1	-15.6
	mml-miR-27a-3p	1	-17.1
	mml-miR-30d-5p	1	-18.0
	mml-miR-30a-5p	5	-18.0

Figure 4. CoV2-miR-O8 targets host microRNA and mRNAs in COVID-19 infected patients

Analysis of miR-O8-5p in CLEAR-CLIP data from COVID-19-infected cells.

(A) The chimera and singleton reads of miR-O8-5p and miR-O8-3P region in A549-hACE2 and VeroE6 cells infected with COVID-19 (WashU genome browser visualization).

(B) Chimera sequences were mapped to mature miRNAs from miRBase22, and read counts were determined. RNAhybrid analysis shows the minimum free energy hybridization of miR-O8-5p and mature miRNAs.

(C) Predicted target sites for miR-O8-5p (green) interaction with miR-30a-5p (left) and miR-27a-3p (right) (red).

(D) Venn diagram illustrating the predicted targets of CoV2-miR-O8 (blue) overlapping with the upregulated (green) and downregulated (orange) genes in COVID-19 positive patients' nasopharyngeal swab RNA-seq data (GSE152075). The percentage of overlap indicates the number of genes that were differentially expressed in their RNA-seq analysis and are also potential targets of CoV2-miR-O8.

(E) Network plot showing the genes targeted by CoV2-miR-O8, with upregulated genes represented in red and downregulated genes in blue. The thickness of the edges represents the total score from miRanda output of target prediction. The network was generated using Cytoscape.

suggesting v-miRNA generation by SARS-CoV-2[5,9,10] and emphasizes the importance of studies in infected people to accurately profile the existence and abundance of small viral encoded RNAs which are predicted from bioinformatic or cell culture studies.

This sequence, CoV2-miR-O8, can be detected by RT-PCR assays in nasopharyngeal swabs from patients, though further studies will be required to understand the extent to which its detection correlates with active infection, severity and recovery. It is present in these samples at substantially higher levels than CoV2-miR-O7a.1 and CoV2-miR-O7a.2 and indicates the importance of studying the level and potential RNA targets of v-miRNAs in patient samples and not only in cell culture studies. The methods of RNA purification and detection might influence the relative detection of different microRNAs. It also remains an open question whether detection of CoV2-miR-O8 or other v-miRNAs might be more sensitive or specific to infection stage than the existing diagnostic assays. This study provides evidence for the existence of SARS-CoV-2 encoded small RNAs in patient-based samples but has not addressed if there are any correlations of expression of such sequences with clinical severity of infection. There is now a need for clinical trials to explore the association of expression of CoV2-miR-O8 and other virally encoded microRNAs with clinical features of acute COVID-19 infection and post-acute sequelae of SARS-CoV-2 infection (PASC).

The CoV2-miR-O8 sequence has several features consistent with those of a v-miRNA, notably Dicer and Drosha generation, interaction with Argonaute and prediction of targeting specific human RNAs. The greater abundance of either the 5p or the 3p arms of the predicted hairpins for CoV2-miR-O8 and the other hairpins, apparent in all the patient samples (Figure 2B) is in keeping with the marked differences in 5p and 3p mature miRNA expression seen in miRNAs with asymmetric degradation of the passenger or * strand and highly suggestive of Argonaute mediated generation, though it is not possible to completely exclude the explanation that other sequence features or biophysical properties have led to the presence and asymmetric stability of the arms of these hairpin sequences. Furthermore, the binding of

Argonaute to virally encoded hairpin structures might mediate an anti-viral defense mechanism. The differing abundance of these v-miRNAs between cell culture and patient samples is significant and requires consideration in any interpretation of their pathogenetic significance.

The existence of chimeric sequences containing both CoV2-miR-O8 and human microRNA sequence in the CLEAR-CLIP experiments undertaken by Fossat et al.[18] in two different types of SARS-CoV-2 infected cells provides more direct evidence for association between these RNAs. However, it is not possible to discern whether this is due to interaction of these mammalian miRNA sequences and the SARS-CoV-2 RNA encoding CoV2-miR-O8, or whether CoV2-miR-O8 could be acting as a decoy for these microRNAs. It is noteworthy that suppression of these interacting human microRNAs (miR-30a and miR-27a) has been seen in other viral infections[19–22] suggesting a plausible decoy function for CoV2-miR-O8 but it is also possible that host anti-viral responses might explain these interactions or that CoV2-miR-O8 interacts with lncRNAs.

Limitations of the study

An important limitation of this work is lack of direct demonstration of mRNA suppression by CoV2-miR-O8. The lack of hybrids between any bioinformatically predicted targeted mRNAs of CoV2-miR-O8 in the analysis of Fossat et al. (2023) CLEAR-CLIP data in cell culture studies casts doubt on the security of those target predictions. Direct evidence of targets is challenging to achieve directly in patient-based studies but manipulations of expression of miR-O8 in cell culture models of SARS-CoV-2 infection could enhance identification of mRNA targets. The conservation of this v-miRNA sequence, its occurrence in both cell and patient-based samples and the potential regulation of human microRNA targets is of clear interest. It remains to be determined if sequence specific suppression of this or other SARS-CoV-2 encoded v-miRNAs could have therapeutic utility.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Human nasopharyngeal samples
- METHOD DETAILS
  - RNA purification
  - Small RNAseq
  - Small RNAseq bioinformatic analysis
  - CoV-2-miR-O8 sequence analysis
  - Discovery of candidate viral miRNAs using SARS-CoV-2 RNA structurome
  - Dicer and Drosha cleavage site prediction
  - Identification of CoV-2-miR-O8 complementary sites on human 3′UTR
  - RT-qPCR – SARS-CoV-2 diagnostic assay
  - RT-qPCR – Small RNA assays
  - CLEAR-CLIP analysis of the miR-O8-5p and miRNA interactome
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2023.108719.

ACKNOWLEDGMENTS

This work was supported by Flinders University College of Medicine and Public Health COVID-19 grant scheme and by the Flinders Medical Centre Renal Unit Research Fund. We would like to thank David Lynn and the South Australian Genomics Centre for their assistance with obtaining funding for this project and generating the small RNA sequencing data, SA Pathology for their support in obtaining clinical samples, Paul Monis, Alex Keegan, Melody Lau from SA Water for useful discussions and the patients for their involvement and access to clinical samples.

AUTHOR CONTRIBUTIONS

E.J.T., A.O.F., and M.Z.M. undertook the sample preparation. E.J.T. designed and undertook the subsequent RNA assays. E.J.T., S.M., S.A., A.O.F., and S.W.W. undertook the initial RNA sequencing and bioinformatic analyses. S.W.W. performed the analysis of Dicer and Drosha cleavage, structurome analysis and target prediction. J.Y.L. obtained the ethics approval, assisted in grant application and sample collection. C.K.L. coordinated and facilitated the provision of patient samples. E.J.T. and S.W.W. generated the final figures. E.J.T., S.W.W., D.R.C., and M.Z.M. contributed to the conception and design of the study and the interpretation of the results. J.M.G., E.J.T., S.W.W., and D.R.C. wrote the paper which was reviewed and approved by all authors. J.M.G. conceived and designed the study.

DECLARATION OF INTERESTS

The authors have no competing interests.

Received: June 23, 2023
Revised: September 28, 2023
Accepted: December 11, 2023
Published: December 13, 2023

REFERENCES

1. Mishra, R., Kumar, A., Ingle, H., and Kumar, H. (2019). The Interplay Between Viral-Derived miRNAs and Host Immunity During Infection. Front. Immunol. *10*, 3079.
2. Morales, L., Oliveros, J.C., Fernandez-Delgado, R., tenOever, B.R., Enjuanes, L., and Sola, I. (2017). SARS-CoV-Encoded Small RNAs Contribute to Infection-Associated Lung Pathology. Cell Host Microbe *21*, 344–355.
3. Place, R.F., Li, L.C., Pookot, D., Noonan, E.J., and Dahiya, R. (2008). MicroRNA-373 induces expression of genes with complementary promoter sequences. Proc. Natl. Acad. Sci. USA *105*, 1608–1613.
4. Boroojeny, A., and Chitsaz, H. (2020). SARS-CoV-2 orthologs of pathogenesis-involved small viral RNAs of SARS-CoV. Preprint at arXiv. https://doi.org/10.48550/arXiv.2007.05859.
5. Meng, F., Siu, G.K., Mok, B.W., Sun, J., Fung, K.S.C., Lam, J.Y., Wong, N.K., Gedefaw, L., Luo, S., Lee, T.M.H., et al. (2021). Viral MicroRNAs Encoded by Nucleocapsid Gene of SARS-CoV-2 Are Detected during Infection, and Targeting Metabolic Pathways in Host Cells. Cells *10*.
6. Aydemir, M.N., Aydemir, H.B., Korkmaz, E.M., Budak, M., Cekin, N., and Pinarbasi, E. (2021). Computationally predicted SARS-COV-2 encoded microRNAs target NFKB, JAK/STAT and TGFB signaling pathways. Gene Rep. *22*, 101012.
7. Merino, G.A., Raad, J., Bugnon, L.A., Yones, C., Kamenetzky, L., Claus, J., Ariel, F., Milone, D.H., and Stegmayer, G. (2021). Novel SARS-CoV-2 encoded small RNAs in the passage to humans. Bioinformatics *36*, 5571–5581.
8. Saini, S., Saini, A., Thakur, C.J., Kumar, V., Gupta, R.D., and Sharma, J.K. (2020). Genome-wide computational prediction of miRNAs in severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) revealed target genes involved in pulmonary vasculature and antiviral innate immunity. Mol. Biol. Res. Commun. *9*, 83–91.
9. Singh, M., Chazal, M., Quarato, P., Bourdon, L., Malabat, C., Vallet, T., Vignuzzi, M., van der Werf, S., Behillil, S., Donati, F., et al. (2022). A virus-derived microRNA targets immune response genes during SARS-CoV-2 infection. EMBO Rep. *23*, e54341.
10. Pawlica, P., Yario, T.A., White, S., Wang, J., Moss, W.N., Hui, P., Vinetz, J.M., and Steitz, J.A. (2021). SARS-CoV-2 expresses a microRNA-like small RNA able to selectively repress host genes. Proc. Natl. Acad. Sci. USA *118*, e2116668118.
11. Kim, D., Lee, J.Y., Yang, J.S., Kim, J.W., Kim, V.N., and Chang, H. (2020). The Architecture of SARS-CoV-2 Transcriptome. Cell *181*, 914–921.e10.
12. Brown, R.A.M., Epis, M.R., Horsham, J.L., Kabir, T.D., Richardson, K.L., and Leedman, P.J. (2018). Total RNA extraction from tissues for microRNA and target gene expression analysis: not all kits are created equal. BMC Biotechnol. *18*, 16.
13. Leshkowitz, D., Horn-Saban, S., Parmet, Y., and Feldmesser, E. (2013). Differences in microRNA detection levels are technology and sequence dependent. RNA *19*, 527–538.
14. Wong, R.K.Y., MacMahon, M., Woodside, J.V., and Simpson, D.A. (2019). A comparison of RNA extraction and sequencing protocols for detection of small RNAs in plasma. BMC Genom. *20*, 446.
15. Corman, V.M., Landt, O., Kaiser, M., Molenkamp, R., Meijer, A., Chu, D.K., Bleicker, T., Brünink, S., Schneider, J., Schmidt, M.L., et al. (2020). Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. Euro Surveill. *25*.
16. Andrews, R.J., O'Leary, C.A., Tompkins, V.S., Peterson, J.M., Haniff, H.S., Williams, C., Disney, M.D., and Moss, W.N. (2021). A map of the SARS-CoV-2 RNA structurome. NAR Genom. Bioinform. *3*, lqab043.
17. Starega-Roslan, J., Krol, J., Koscianska, E., Kozlowski, P., Szlachcic, W.J., Sobczak, K., and Krzyzosiak, W.J. (2011). Structural basis of microRNA length variety. Nucleic Acids Res. *39*, 257–268.
18. Fossat, N., Lundsgaard, E.A., Costa, R., Rivera-Rangel, L.R., Nielsen, L., Mikkelsen, L.S., Ramirez, S., Bukh, J., and Scheel, T.K.H. (2023). Identification of the viral and cellular microRNA interactomes during SARS-CoV-2 infection. Cell Rep. *42*, 112282.
19. Buck, A.H., Perot, J., Chisholm, M.A., Kumar, D.S., Tuddenham, L., Cognat, V., Marcinowski, L., Dölken, L., and Pfeffer, S. (2010). Post-transcriptional regulation of miR-27 in murine cytomegalovirus infection. RNA *16*, 307–315.
20. Lin, X., Yu, S., Ren, P., Sun, X., and Jin, M. (2020). Human microRNA-30 inhibits influenza virus infection by suppressing the expression of SOCS1, SOCS3, and NEDD4. Cell Microbiol. *22*, e13150.
21. Ma, Y., Wang, C., Xue, M., Fu, F., Zhang, X., Li, L., Yin, L., Xu, W., Feng, L., and Liu, P. (2018). The Coronavirus Transmissible Gastroenteritis Virus Evades the Type I Interferon Response through IRE1alpha-Mediated Manipulation of the MicroRNA miR-30a-5p/SOCS1/3 Axis. J. Virol. *92*, e00728-18.
22. Marcinowski, L., Tanguy, M., Krmpotic, A., Rädle, B., Lisnić, V.J., Tuddenham, L., Chane-Woon-Ming, B., Ruzsics, Z., Erhard, F., Benkartek, C., et al. (2012). Degradation of cellular mir-27 by a novel, highly abundant viral transcript is important for efficient virus replication in vivo. PLoS Pathog. *8*, e1002510.
23. Andrews, S. (2010). FastQC: a quality control tool for high throughputsequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc.
24. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. *17*, 3.
25. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.
26. Gu, Z., Gu, L., Eils, R., Schlesner, M., and Brors, B. (2014). circlize Implements and enhances circular visualization in R. Bioinformatics *30*, 2811–2812.
27. Thorvaldsdóttir, H., Robinson, J.T., and Mesirov, J.P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Briefings Bioinf. *14*, 178–192.
28. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.
29. Bell, J., and Hendrix, D.A. (2021). Predicting Drosha and Dicer Cleavage Sites with DeepMirCut. Front. Mol. Biosci. *8*, 799056.
30. Enright, A.J., John, B., Gaul, U., Tuschl, T., Sander, C., and Marks, D.S. (2003). MicroRNA targets in Drosophila. Genome Biol. *5*, R1.
31. Hannon, G.J. (2010). FASTX-toolkit.
32. Bushnell, B. (2014). BBMap: A Fast, Accurate, Splice-Aware Aligner.
33. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.
34. Krüger, J., and Rehmsmeier, M. (2006). RNAhybrid: microRNA target prediction easy, fast and flexible. Nucleic Acids Res. *34*, W451–W454.
35. Li, D., Purushotham, D., Harrison, J.K., Hsu, S., Zhuo, X., Fan, C., Liu, S., Xu, V., Chen, S., Xu, J., et al. (2022). WashU Epigenome Browser update 2022. Nucleic Acids Res. *50*, W774–W781.
36. Chen, S., He, C., Li, Y., Li, Z., and Melançon, C.E. (2021). A computational toolset for rapid identification of SARS-CoV-2, other viruses and microorganisms from sequencing data. Briefings Bioinf. *22*, 924–935.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Human nasopharyngeal swabs	South Australia Pathology	Southern Adelaide Clinical Human Research Ethics Committee EC00188
Chemicals, peptides, and recombinant proteins		
TRIzol LS	Thermo Fisher Scientific	Cat# 10296010
Glycogen, RNA grade	Thermo Fisher Scientific	Cat# R0551
EDTA (0.1 mM), pH 8.0, RNase-free	Thermo Fisher Scientific	Cat# AM9912
Critical commercial assays		
Quant-iT™ RNA Assay Kits	Thermo Fisher Scientific	Cat# Q32884
LabChip GX Touch 24 RNA reagent	Perkin Elmer	Cat# CLS 960010
NEBNext® Multiplex Small RNA Library Prep Set for Illumina®	New England Biolabs	Cat# E7300S
AMPure XP Beads	Beckman Coulter	Cat# A63880
SuperScript IV VILO	Thermo Fisher Scientific	Cat# 11756050
TaqMan MicroRNA Reverse Transcription kit	Thermo Fisher Scientific	Cat# 4366597
TaqMan™ Fast Advanced Master Mix	Thermo Fisher Scientific	Cat# 4444557
PowerUp™ SYBR™ Green Master Mix	Thermo Fisher Scientific	Cat# A25742
Deposited data		
smRNA-seq human SARS-CoV-2 nasopharyngeal swab RNA	This paper	GSE230080
Oligonucleotides		
E_gene_F1_Forward: ACAGGTACGTTAATAGTTAATAGCGT	Corman et al.[15]	E gene
E_gene_R2_Reverse Primer ATATTGCAGCAGTACGCACACA	Corman et al.[15]	E gene
E gene_TaqMan Probe: (ABY) ACACTAGCCATCCTTACTGCGCTTCG(QSY7)	Corman et al.[15]	E gene
Reverse transcription primer: CoV2-miR-O8: CTCAACTGGTGTCGTGGAGTCGGCA ATTCAGTTGAGAGCCTCAT	This paper	CoV2-miR-O8 28064-28088
Forward primer: CoV2-miR-O8: ACACTCCAGCTGGGAATTGA ATTGTGCGTGGATG	This paper	CoV2-miR-O8 28064-28088
Reverse Primer: Universal for smRNAs: CTCAAGTGTCGTGGAGTCGGCAA	Singh et al.[9]	Universal tag
Reverse transcription primer: CoV2-miR-7a.1: CTCAACTGGTGTCGTGGAGTCGGCAA TTCAGTTGAGAGTGTTAT	Singh et al.[9]	CoV2-miR-7a.1
Forward primer: CoV2-miR-7a.1: ACACTCCAGCTGGGTTTTCTTGGCACTGAT	Singh et al.[9]	CoV2-miR-7a.1
Software and algorithms		
GraphPad Prism	GraphPad Software, Inc	www.graphpad.com

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
FastQC	Andrews[23]	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
Cutadapt v2.3	Martin[24]	https://cutadapt.readthedocs.io/en/stable/
Bowtie2	Langmead and Salzberg[25]	https://bowtie-bio.sourceforge.net/bowtie2/manual.shtml
Circlize package in R	Gu et al.[26]	https://github.com/jokergoo/circlize
Integrative Genomics Viewer (IGV)	Thorvaldsdottir et al.[27]	https://www.igv.org/
bedtools	Quinlan et al.[28]	https://bedtools.readthedocs.io/en/latest/
DeepMirCut	Bell et al.[29]	https://github.com/JimBell/deepMirCut
miRanda	Enright et al.[30]	https://cbio.mskcc.org/miRNA2003/miranda.html
FASTX-Toolkit	Hannon et al.[31]	http://hannonlab.cshl.edu/fastx_toolkit/
BBMap	Bushnell et al.[32]	https://jgi.doe.gov/data-and-tools/software-tools/bbtools/bb-tools-user-guide/bbmap-guide/
samtools idxstats	Li et al.[33]	http://www.htslib.org/doc/samtools-idxstats.html
RNAhybrid	Krüger et al.[34]	https://bibiserv.cebitec.uni-bielefeld.de/rnahybrid
WashU Epigenome Browser	Li et al.[35]	http://epigenomegateway.wustl.edu/browser/
Other		
SARS-CoV-2 genome	https://www.ncbi.nlm.nih.gov/	NC_045512.2
Homo sapiens genome	https://www.gencodegenes.org/	GRCh38
Stable predicted RNA structures in SARS COV-2 genome	https://structurome.bb.iastate.edu/download/sars-cov-2-extractedstructures	SARS-CoV-2-ExtractedStructures
SARS-CoV-2 RNA-seq	https://www.ncbi.nlm.nih.gov/	GSE152075
SARS-CoV-2 Genome Browser	https://genome.ucsc.edu/covid19.html	Covid19
SARS-CoV-2 anti-pan-Ago RNA immunoprecipitation (RIP) datasets	https://www.ncbi.nlm.nih.gov/geo/	GSM5554778 - GSM5554789
Human miRNA database	https://www.mirbase.org	GRCh38
SARS-CoV-2 CLEAR-CLIP datasets	https://www.ncbi.nlm.nih.gov/geo	GSE201894

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Jonathan Gleadle (Jonathan.gleadle@flinders.edu.au).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Small RNA-seq data have been deposited at GEO and are publicly available as of the date of publication. Accession number is GSE230080 as listed in the key resources table.
- This paper does not report original code.
- Any additional information required to re-analyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Human nasopharyngeal samples

Nasopharyngeal swabs were obtained from SA Pathology (Adelaide, Australia) in accordance with ethics approval from the Southern Adelaide Clinical Human Research Ethics Committee (SAC HREC) (EC00188), informed consent was obtained from all subjects. The nasopharyngeal swab samples were categorised as positive or negative for COVID-19 as previously determined by SA Pathology's routine diagnostic testing of the envelope (E) gene.[15]

METHOD DETAILS

RNA purification

Upon collection, the specimens were sampled into 1 mL of viral transport fluid following SA Pathology's standard combined throat and deep nasal swab collection protocol. RNA was isolated from 0.5 mL of viral transport fluid using 1.5 mL of TRIzol LS (Thermo Fisher Scientific) and chloroform as per the manufacturer's recommendations. RNA was precipitated with isopropanol supplemented with 20 μg of glycogen (Thermo Fisher Scientific) as a carrier. RNA pellets were washed with 75% ethanol and resuspended in 30 μL 0.1 mM nuclease-free EDTA (Thermo Fisher Scientific). RNA concentration was determined using a NanoDrop 8000 Spectrophotometer (Thermo Fisher Scientific) and Qubit Fluorometer (Thermo Fisher Scientific), and LabChip GX Touch 24 (PerkinElmer) for quality assessment. RNA from ten nasopharyngeal COVID-19 positive samples were processed for small RNAseq analysis. An additional 10 nasopharyngeal COVID-19 positive samples and 8 nasopharyngeal COVID-19 negative samples were used as a validation panel.

Small RNAseq

Libraries were prepared from nasopharyngeal RNA from SARS-CoV-2 positive patients using the NEBNext Multiplex Small RNA Library Prep Set for Illumina (New England Biolabs) as per the manufacturer's protocol. PCR amplification was performed using 15 cycles and size selection was performed using AMPureXP beads (Beckman Coulter) according to the manufacturer's protocol. Libraries were QC checked for quality and quantitation of RNA on the Lab Chip (PerkinElmer). Pooled libraries were sequenced on the Illumina NextSeq 500 1 x 75 cycle High Output kit.

Small RNAseq bioinformatic analysis

FastQC[23] was used to quality check the smRNAseq sequencing reads. Cutadapt version 2.3[24] was used to trim reads with adaptors (AAGATCGGAAGAGCACACGTCT), remove low quality reads and remove reads with a length less than 18 bp. The cleaned reads were first aligned to the Homo sapiens genome (GRCh38, Gencode) using Bowtie2[25] with the following parameters: -q -D 20 -R 3 -N 0 -L 18 -i S,1,0.50. The unaligned reads were then aligned to SARS-CoV-2 genome (NC_045512.2, NCBI). The whole read coverage was visualised as a circos plot using the Circlize package in R[26] with mapped reads divided into different bins and distribution calculated at each position to produce smooth and uniform coverage. The alignment process was repeated with 18–26 nucleotides size selected reads using awk command. The size distribution of SARS-CoV-2 mapped reads and the overrepresented smRNAseq patient reads were determined. Integrative Genomics Viewer (IGV)[27] was used to visually inspect the bam file showing smRNAseq read distribution across the SARS-CoV-2 genome and select sequences for subsequent validation. The reads of human origin were aligned to the miRBase GRCh38 human miRNA database.

CoV-2-miR-O8 sequence analysis

The CoV-2-miR-O8 containing region of genome position 28064-28088 bp was compared for sequence composition between SARS-CoV-2 (B.1.1.7), Beta (B.1.351), Gamma (P.1), Delta (B.617.2), and Omicron (B.1.1.529, BA.4, BA.5) strains using the UCSC Genome Browser view of SARS-CoV-2 genomic datasets (https://genome.ucsc.edu/covid19.html). SARS-CoV-2 anti-pan-Ago RNA immunoprecipitation (RIP) datasets GSM5554778 – GSM5554789 were searched for the presence of the CoV2-miR-O8 sequence using the Sequence Read Archive interface.

Discovery of candidate viral miRNAs using SARS-CoV-2 RNA structurome

524 uniquely stable predicted RNA structures in SARS-CoV-2 genome were obtained from Andrews et al.[16]; https://structurome.bb.iastate.edu/download/sars-cov-2-extractedstructures). Based on dot-bracket notation in the ScanFold-Fold Pairs and MFE refold column, the hairpin RNA structures were then divided into 5p and 3p sequences. 165 hairpin structures were considered to be candidate pre-miRs if the resulting 5p or 3p sequence was 16–28 nucleotides in length. SARS-CoV-2 mapped reads (18–26 nt) were counted if 80% of the reads overlapped with the 5p or 3p candidate pre-miRs reference using bedtools coverage.[28]

Dicer and Drosha cleavage site prediction

A neural network-based algorithm DeepMirCut[29] was applied to predict cleavage sites for Drosha and Dicer on the 165 hairpin structures (30 nt extended on each end) using sequence, dot-bracket structure array, and RNA secondary structures context, thereby confirming the region of the precursor resulting in a potential mature miRNA.

Identification of CoV-2-miR-O8 complementary sites on human 3′UTR

Target prediction software miRanda[30] was used to identify potential interaction between CoV2-miR-O8 and human mRNA transcripts. Briefly, 3′ UTR sequences were extracted from Homo sapiens genome (GRCh38, Gencode) fasta file based on annotation gff3 (version 42, Gencode). Local alignments of miRNA:UTR was then carried out by miRanda to assess the thermodynamic folding energy of the duplex. The output targets with strong interaction (total energy < −30 kcal/mol) were retained for further analysis. An additional filtering was incorporated using differentially expressed genes RNA-seq retrieved from nasopharyngeal swabs (GSE152075) to remove any putative interactions where the target transcript is not expressed in the tissue or cell line of interest.

RT-qPCR – SARS-CoV-2 diagnostic assay

The envelope (E) gene SARS-CoV-2 diagnostic assay[15] was used to verify the COVID-19 status of the patient nasopharyngeal RNA samples. The envelope (E) gene TaqMan assay following primers and probe were synthesised by Thermo Fisher Scientific E_Sarbeco_F1 ACAGGTA CGTTAATAGTTAATAGCGT; E_Sarbeco_R2 ATATTGCAGCAGTACGCACACA; and E_Probe (ABY)ACACTAGCCATCCTTACTGCGCTTCG (QSY7). RNA template from SARS-CoV-2 positive and negative patient samples were used for reverse transcription reactions and subsequent qPCR. Reverse transcription reactions were prepared from each RNA sample in a 10 μL volume using the SuperScript IV VILO (Thermo Fisher) according to the manufacturer's instructions. qPCR assays were setup in triplicate using 2 μL cDNA (1:2 dilution), 5 μL 2x TaqMan Fast Advanced Mastermix (Thermo Fisher); 0.5 μL E_Sarbeco TaqMan Assay primer mix (Thermo Fisher); and 2.5 μL of water. Reactions were cycled in a Rotorgene Q (Qiagen): 95°C for 20 s, followed by 50 cycles of 95°C for 5 s, 60°C for 30 s, acquiring fluorescence in the yellow channel at the end of each extension step. Cycling threshold scores were determined for each reaction.

RT-qPCR – Small RNA assays

Stem-loop RT-qPCR assays (see key resources table) were designed to the novel small RNA sequence CoV-2-miR-O8 based on Chen et al.[36] with the universal tag and reverse primer as per Singh et al. Stem-loop RT-qPCR assays for CoV2-miR-7a.1 from Singh et al.[9] were synthesised to assess the expression of these selected small RNA sequences. RNA template from SARS-CoV-2 positive and negative patient nasopharyngeal swabs were used for reverse transcription reactions and subsequent qPCR. Reverse transcription reactions were prepared for each RNA sample in a 7.5 μL volume using assay specific primers and the TaqMan MicroRNA Reverse Transcription kit (Thermo Fisher) according to the manufacturer's instructions. Each microRNA reverse transcription reaction was diluted 1:3 prior to qPCR. qPCR assays were setup in triplicate using 2 μL cDNA (1:3 dilution), 5 μL 2x PowerUP SYBR Green mastermix (Thermo Fisher); 0.5 μL forward primer; 0.5 μL reverse primer; and 2 μL of water. Reactions were cycled in a Rotorgene Q (Qiagen): 95°C for 5 s, followed by 40 cycles of 95°C for 5 s, 60°C for 30 s, acquiring fluorescence in the green channel at the end of each extension step. Cycling threshold scores were determined for each reaction.

CLEAR-CLIP analysis of the miR-O8-5p and miRNA interactome

To analyze the CLEAR-CLIP data for miR-O8-5p from the study by Fossat et al.,[18] raw sequencing files from A549hACE2 and VeroE6 cells infected with SARS-CoV-2 were obtained from GSE201894 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE201894). Adapters were trimmed using the FASTX-Toolkit (fastx_clipper), and unique molecular identifiers (UMIs) were collapsed using BBMap (clumpify). Cleaned reads were initially aligned to the SARS-CoV-2 genome using Bowtie 2.[25]with the following parameters: –local -D 20 -R 2 -N 0 -L 15 -i S,1,0.50 –ma 3. Reads without soft-clipped regions were considered SARS-CoV-2-mapped singleton reads. For chimeric reads, soft-clipped regions of the mapped reads were separated, leaving the SARS-CoV-2-mapped reads. SARS-CoV-2-mapped reads from singleton and chimera were then filtered for 18–34 nt. BAM files of the aligned reads were converted to BigWig files for visualization on the WashU Epigenome Browser. Adjacent chimera sequences were mapped directly to mature miRNAs from either Homo sapiens (hsa) or Macaca mulatta (mml) obtained from miRBase 22 release, and the read counts were determined using samtools idxstats. RNAhybrid was performed to identify the minimum free energy hybridization of miR-O8-5p with mature miRNAs.

QUANTIFICATION AND STATISTICAL ANALYSIS

Data were analyzed using the GraphPad Prism 9 software package (GraphPad Software, San Diego, CA, USA) and R version 4.3.1 (R Core Team). The graphs represent the mean ± standard error of the mean (SEM) of 6–17 biological samples. Simple linear regression was carried out to find correlation between the expression of CoV2-miR-08 and the E gene.