Altered microRNA expression in COVID-19 patients enables identification of SARS-CoV-2 infection

Ryan Farr
Commonwealth Scientific and Industrial Research Organisation https://orcid.org/0000-0003-1636-1815

Christina Rootes
Commonwealth Scientific and Industrial Research Organisation https://orcid.org/0000-0002-7448-8307

Louise Rowntree
Peter Doherty Institute for Infection and Immunity

Thi Nguyen
Peter Doherty Institute for Infection and Immunity

Luca Hensen
Peter Doherty Institute for Infection and Immunity

Lukasz Kedzierski
Peter Doherty Institute for Infection and Immunity

Allen Cheng
Monash University

Katherine Kedzierska
Department of Microbiology and Immunology, University of Melbourne at the Peter Doherty Institute for Infection and Immunity, Melbourne, Victoria, Australia https://orcid.org/0000-0001-6141-335X

Gough Au
Commonwealth Scientific and Industrial Research Organisation https://orcid.org/0000-0001-6923-8875

Glenn Marsh
CSIRO https://orcid.org/0000-0002-3469-1837

Seshadri Vasan
Commonwealth Scientific and Industrial Research Organisation https://orcid.org/0000-0002-7326-3210

Chwan Hong Foo
Exios Bio

Christopher Cowled
CSIRO

Cameron Stewart (Cameron.Stewart@csiro.au)
Commonwealth Scientific and Industrial Research Organisation

Article

Keywords: SARS-CoV-2, COVID-19, microRNA, biomarker, diagnostic

Posted Date: February 24th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-253459/v1

Abstract

The host response to SARS-CoV-2 infection provide insights into both viral pathogenesis and patient management. The host-encoded microRNA (miRNA) response to SARS-CoV-2 infection, however, remains poorly defined. Here we profiled circulating miRNAs from ten COVID-19 patients sampled longitudinally and ten age and gender matched healthy donors. We observed 55 miRNAs that were altered in COVID-19 patients during early-stage disease, with the inflammatory miR-31-5p the most strongly upregulated. Supervised machine learning analysis revealed that a three-miRNA signature (miR-423-5p, miR-23a-3p and miR-195-5p) independently classified COVID-19 cases with an accuracy of 99.9%. In a ferret COVID-19 model, the three-miRNA signature again detected SARS-CoV-2 infection with 99.7% accuracy, and distinguished SARS-CoV-2 infection from influenza A (H1N1) infection and healthy controls with 95% accuracy. Distinct miRNA profiles were also observed in COVID-19 patients requiring oxygenation. This study demonstrates that SARS-CoV-2 infection induces a robust host miRNA response that could improve COVID-19 detection and patient management

Main

As of February 2021, the COVID-19 pandemic, caused by infection with severe acute respiratory syndrome-associated coronavirus-2 (SARS-CoV-2) has resulted in over 107 million cases and 2.35 million deaths worldwide[1]. The outcome of SARS-CoV-2 infection varies widely from asymptomatic to severe disease associated with acute respiratory distress syndrome (ARDS) and death[2]. Several studies have established that host responses to infection play a critical role in determining disease outcomes in infected patients. For example, hyper-inflammatory responses including high levels of circulating cytokines (particularly interleukin (IL)-6, IL-8, and tumor necrosis factor (TNF)-α), lymphopenia and immune cell infiltration in infected organs are considered major determinants of COVID-19 severity[3-6]. As there are currently no approved curative treatments for COVID-19, the characterisation of host factors associated with SARS-CoV-2 pathogenesis is critically important for the design of novel therapies.

MicroRNAs (miRNAs) are a class of non-coding RNAs that regulate endogenous gene expression at the post-transcriptional level. In most instances, miRNAs function by interacting with the 3' untranslated region (3' UTR) of target mRNAs to induce degradation and translational repression[7]. There are currently over 2,600 human miRNAs listed in the miRNA registry (miRBase, version 22)[8] which are estimated to collectively regulate 60% of all human protein-coding genes[9]. The scientific rationale for investigating miRNAs during viral infections is two-fold. Firstly, miRNA profiles offer unique insight into cellular pathways associated with virus replication and pathogenesis. For instance, the human coronavirus OC43 potentiates NF-kB activation during infection by binding and sequestering miR-9, a negative regulator of NF-kB[10]. There is also evidence that coronaviruses co-opt the host miRNAs response to subvert antiviral immune responses. Infection by the *Alphacoronavirus* transmissible gastroenteritis virus (TGEV) downregulates miR-30a-5p expression, which disrupts the type I interferon response against TGEV[11]. Secondly, the characterisation of host miRNAs responses to virus infection informs the development of

biomarkers for improved disease detection and forecasting of disease outcome[12]. Several pathogenic viruses, including SARS-CoV-1, induce changes to the circulating host miRNA profile[13-17]. Interestingly, host miRNA responses to SARS-CoV-1 and influenza A virus differ based on virus type and pathogenicity[13], highlighting the potential for miRNAs to serve as diagnostic or prognostic biomarkers.

In this study we have investigated the circulating miRNA profiles in the plasma of ten COVID-19 patients and ten age and gender matched healthy donors. We observed that among patient samples collected during early-stage disease, COVID-19 induced differential expression of 55 host-encoded microRNAs, with miR-31, -4742 and -3125 strongly up-regulated and miR-1275, -3617 and -500b down-regulated. Logistic regression analysis revealed that measurement of three miRNAs (miR-423-5p, miR-23a-3p and miR-195-5p) could identify early-stage COVID-19 with 99.9% accuracy. As patients recovered from disease, the three-miRNA plasma signature returned to that of the healthy controls. The miRNA signature was shown to be robust in the ferret model of COVID-19 and could distinguish SARS-CoV-2 infection from seasonal influenza A infection. These findings suggest that miRNA profiling may be adopted to improve COVID-19 detection and patient management.

Results

Host miRNAs are altered in response to SARS-CoV-2 infection

Plasma samples were obtained from ten COVID-19 patients and ten age- and gender-matched healthy controls (Table 1). Longitudinal samples were available for some COVID-19 patients, categorized by visit (V), with V1 representing the plasma sample first taken following hospital admission. Plasma samples were first obtained from COVID-19 patients 2-15 days (average 8 days) post symptomatic disease onset. Small RNA deep sequencing resulted in 23-50 million (average 34 million) raw reads per sample, which have been submitted to the NCBI short read archive (SRA). Reads were trimmed of adaptors and filtered on length and quality, resulting in a loss of 29-74% (average 56%) of raw reads, leaving 8-35 million (average 15.2 million) reads per sample for further analysis (Supplementary Figure 1). The majority of sequences were deemed high quality by FASTQC (data not shown).

MiRDeep2 was used to identify all known miRNA transcripts amongst the 29 samples and read counts were determined for each mature miRNA transcript. Total counts included all reads that mapped to a locus (as opposed to reads matching the canonical/consensus sequence only). A total of 985 different mature miRNA transcripts were detected, corresponding to 756 different precursors (5p and 3p miRNAs were counted separately). We did not observe a significant difference in the total number of miRNAs identified in infected versus uninfected patients (data not shown). The most abundant miRNA in the plasma dataset was miR-16-5p, followed by miR-223-3p, let-7b-5p and miR-146a-5p.

We sought to identify miRNAs with significantly altered expression levels between healthy control (n=10) and COVID-19 V1 (n=7) samples. By using DESeq2 to perform count-based differential expression (DE) testing, a subset of miRNAs that were up- or down-regulated in infected patients relative to uninfected

controls were identified (Figure 1a, Supplementary Table 1). Using a False Discovery Rate (FDR) adjusted p-value <0.05, $\log_2$ fold change (FC) >1 and baseMean >5, this dataset consisted of 50 miRNAs, of which 20 were up-regulated (elevated in infected patients) and 30 were down-regulated. An additional 5 miRNAs were significantly DE in COVID-19 patients with $\log_2$FC values <1. The most highly up-regulated candidates in COVID-19 patients were miR-31-5p (associated with inflammatory disorders[18-20]) (Figure 1a), miR-3125 and miR-4742-3p, while the most down-regulated were miR-1275 (Figure 1a), miR-3617-5p and miR-500b-3p. The most statistically significant change was seen in miR-766-3p (Figure 1a), a known anti-inflammatory miRNA[21]. Unsupervised analysis of variance using principal components analysis (PCA) involving the 55 DE miRNAs showed tight clustering of patient groups (Figure 1b). qRT-PCR was employed to validate select DE miRNA expression (Figure 1c). Quantitation of circulating cytokines highlighted a significant increase in IL-6 (Figure 1d) during acute COVID-19 illness, supporting previous studies[4]. Other patient cytokine data is shown in Supplementary Table 2. One significantly reduced miRNA, miR-766-3p, has been shown to reduce IL-6 expression in a dose-dependent manner[21], while another, miR--4662a-5p (Figure 1c), is predicted to target IL-6 mRNA (TargetScan, release 7.2)[22]. The miRNAs may play a role in the hyperinflammatory state commonly seen in patients with COVID-19[23].

A three-miRNA signature accurately predicts COVID-19

Technologies most commonly utilized for COVID-19 diagnosis are virus-targeting molecular assays or serology, both of which can be associated with relatively high false-positive rates[24,25]. We therefore investigated whether, during the early stages of COVID-19, infected patients displayed a miRNA profile that could independently identify SARS-CoV-2 infection. A supervised machine learning method was implemented for the identification of the most predictive miRNAs and refined to identify the minimum targets necessary for accurate prediction and classification between healthy control and COVID-19 (V1) samples. A logistic regression model was implemented that randomly split the data into discovery and validation sets, trained and tested the model, which was repeated 1,000 times to determine reproducibility. The most predictive miRNAs were selected using recursive feature elimination (Figure 2a). Measuring a single miRNA (miR-195-5p) identified COVID-19 (V1) cases with ~90% accuracy, 95% precision, and 72% recall with a receiver operating characteristic area under the curve (ROC AUC) of 0.9. Measuring three miRNA targets (miR-423-5p, miR-23a-3p and miR-195-5p) in combination gave a model with 99.9% accuracy, 99.8% precision and 99.9% recall, with a ROC AUC of 1.0 (Figure 2b). Interestingly, the biomarker was comprised of two miRNAs DE in COVID-19 patients (miR-423-5p and miR-195-5p, both upregulated) and miR-23a-3p, which was not DE. Increasing candidates within the biomarker signature to more than three miRNAs did not improve test performance.

A decision boundary graph showed clear distinctions between healthy and infected patients based on these three miRNAs (Figure 2c). The decision boundary graph also clearly shows that each sample's grouping was predicted with a high degree of confidence (0% probability of healthy samples being identified as infected with SARS-CoV-2, and 100% probability of COVID-19 samples being detected as infected). The probability of a sample being infected with SARS-CoV-2 is determined by its distance from

the decision boundary. The absence of points close to the boundary supports the high predictive accuracy of this miRNA signature. Interestingly, samples taken at successive timepoints (V2, V3 and V4) cluster with the healthy controls, indicating a return to normal baseline and suggesting that the three-miRNA signature is associated with the early stages of COVID-19 (Figure 2d). The obvious exception to this is one sample from V2, V3, and V4 that clusters closer to the V1 COVID-19 samples (indicated by # in Figure 2d); these samples all came from the same participant who was treated in the intensive care unit (ICU) and had not recovered from COVID-19 at any of the timepoints sampled. To further support the hypothesis that this signature detects early symptomatic COVID-19, we tested the model on the later time points (V2,3 and 4) saw the accuracy reduced to 16.6% (Supplementary Figure 2).

Circulating miRNAs differ based on disease severity

Pro-inflammatory cytokines (IL-6, IL-8, TNF-α, IL-1β) are differentially expressed in COVID-19 patients according to severity[5]. We therefore investigated prospective differences between miRNA profiles in moderate and severe COVID-19 cases. In this analysis the need for infected patients to receive supplemental oxygen (O_2) or intubation was used as a proxy marker for severe disease, a metric previously used to categorise COVID-19 severity[4,5]. COVID-19 V1 patient samples were split into two groups (COVID-19 and COVID-19 + O_2) based on the need for supplemental oxygen or intubation and then compared to the healthy controls. Analysis revealed that COVID-19 patients requiring oxygenation had fewer DE miRNAs in circulation (15 vs 42) compared to patients not requiring oxygenation (Figure 3a, Supplementary Table 3). Interestingly, four miRNAs (let-7e-5p, miR-651-5p, miR-766-3p, and miR-4433b-5p) were differentially expressed in both groups, suggesting that these molecules might be potential candidates for stratifying patients based on severity. Indeed, the healthy, COVID-19 and COVID-19 + O_2 groups clustered based on the expression of these four miRNAs (Figure 3b).

The human COVID-19 signature also predicts ferret SARS-CoV-2 infection and differentiates it from influenza infection.

Finally, we investigated whether the biomarker of early-stage COVID-19 was robust in an animal model and could distinguish between different viral respiratory infections. To address this, infection studies were performed in domestic ferrets (*Mustla putorius furo*), a well-established model for human respiratory viruses, including SARS-CoV-2 and influenza virus[26]. Twenty adult ferrets were exposed to SARS-CoV-2 via the intranasal route and monitored for clinical signs, with four ferrets euthanized at 3, 5, 7, 9, and 14 days post-exposure (d.p.e.). The establishment of infection was confirmed by performing qRT-PCR for viral genomic RNA on tissues and swabs (Figure 4a). High viral load was detected in nasal wash samples from day 3, which declined over time and was negative in all ferrets by 14 d.p.e. Eleven ferrets were infected with influenza A(H1N1) virus via the intranasal route, with animals euthanized at days 1, 2, 3, 5, 6 and 7 d.p.e. Influenza virology data in tissue and swab samples is shown in Supplementary Figure 3. Viral load was detected in nasal wash samples from 1 to 7 d.p.e..

Small RNA from serum samples were profiled for miRNAs using the same methodology as the patient samples. Sera from 12 uninfected ferrets were included as controls. In the ferret model, the previously identified biomarker signature (miR-423-5p, miR-23a-3p and miR-195-5p) could independently distinguish uninfected ferrets from COVID-19 infected ferrets with 99.7 % accuracy, 99.5 % precision, 100 % recall, and a ROC AUC of 1.0 (Figure 4b). As with the human plasma samples, the decision boundary graph displayed high confidence in the predicted groupings (Figure 4c). Intriguingly, the miRNA biomarker still identified SARS-CoV-2 infection at 14 d.p.e., by which time ferrets were SARS-CoV-2 negative by nasal wash qRT-PCR, but with virus replication observed in the retroperitoneal lymph node tissue of 3 out of 4 ferrets (Figure 4a). In addition, the biomarker could distinguish SARS-CoV-2 infection from influenza infection and healthy control ferrets with 95 % accuracy, 95.5 % precision and 94.6 % recall (Figure 4d). The decision boundary graph comparing predicted grouping and true grouping is shown in Figure 4e.

Discussion

Here we present an unbiased profiling study of the circulating miRNAs in COVID-19 patients. In plasma samples obtained soon after the onset of disease symptoms (V1), a total of 55 miRNAs were DE, with several miRNAs more than 50-fold up-regulated (miR-31-5p, miR-3125, miR-4742-3p) or down-regulated (miR-1275, miR-3617-5p, miR-500b-3p) compared to basal miRNA expression levels in healthy donors. A response involving three miRNAs (miR-423-5p, miR-23a-3p and miR-195-5p) was consistently observed in COVID-19 patients at V1 and could independently classify SARS-CoV-2 infection with >99% accuracy. Ferret infection trials showed that this signature response was robust across species and was still valid during timepoints where SARS-CoV-2 replication was observed in internal organs but not in nasal wash samples. The biomarker was less predictive of SARS-CoV-2 infection in V2, V3 and V4 patient samples, suggesting this response is associated with early-stage COVID-19.

This signature was not determined based on FC differences in miRNA expression between infected and control groups, as miR-423-5p and miR-195-5p were relatively mildly up-regulated COVID-19 patients (log_2FC 2.35 and 0.86, respectively), while miR-23-3p was non-significantly down-regulated (log_2FC -0.64, adjusted p-value=0.103). We hypothesize that a biomarker consisting of multiple miRNAs is more robust than one based on absolute or relative levels of a single miRNA. This multivariate approach, coupled with advanced machine learning analysis, can highlight a biomarker pattern that may not be identified via traditional DE analysis. In support of this, we note that the three-miRNA signature was robust in humans and ferrets, despite a relatively poor overlap in DE miRNAs observed in human and ferret COVID-19 samples (Supplementary Table 4). While miR-423-5p, miR-23a-3p and miR-195-5p measured in combination have not been defined previously as a biomarker for a specific disease, increased expression of circulating miR-423-5p is observed during heart failure[27] and pulmonary tuberculosis[28]. Increases in circulating miR-195-5p are associated with osteosarcoma[29], autism[30] and gestational diabetes mellitus[31]. Interestingly, increased plasma expression of miR-195-5p is also observed during HIV-1 infection, with miR-195-5p forming part of a four-miRNA signature that can identify HIV-1 infection with high confidence[32].

While host responses to infection are known to be critical in differential outcomes of SARS-CoV-2 infection, the role of miRNAs in COVID-19 pathogenesis is poorly understood. We observed that miR-31-5p was the most strongly up-regulated miRNA in COVID-19 patients, which may be related to its role in modulating inflammation. Transcription of miR-31-5p in endothelial cells is induced by TNF-α and triggers a negative feedback loop involving E-selectin to control inflammatory signalling[33]. MicroRNA-31-5p is also upregulated in inflamed ulcerative colitis mucosa[34], where it downregulates expression of the IL-13 receptor α-1 (*IL13RA1*) gene, which diminishes expression of signal transducer and activator of transcription 6 (*STAT6*), suppressor of cytokine signalling 1 (*SOCS1*) and eotaxin-3 (*CCL26)* expression. These findings raise the possibility that miR-31-5p is induced in response to acute stressors such as SARS-CoV-2 in order to curtail an excessive inflammatory response. Interestingly, miR-27a-5p (also up-regulated in V1 COVID-19 samples), is elevated in animal models of enterocolitis[35]. The up-regulation of miR-31-5p and miR-27a-5p in COVID-19 patients may reflect SARS-CoV-2 mediated gastrointestinal tract infection or inflammation[36]. Furthermore, the most statistically significant down-regulated miRNA was miR-766-3p, a previously identified anti-inflammatory miRNA. This miRNA was shown to reduce the expression of IL-6 in TNF-α stimulated MH7a cells [21] and so its reduction may be partially responsible for the characteristic IL-6 increase seen in COVID-19 patients[4]. In addition to miR-31-5p, miR-27a-5p, and miR-766-3p, we observed several miRNAs DE in COVID-19 patients that are poorly characterized from a functional perspective. Many miRNAs upregulated (miR-3125, miR-4742-3p, miR-2116-3p) or down-regulated (miR-3617-5p, miR-500b-3p, miR-3684) in COVID-19 patients have not been functionally characterized or previously observed in studies of miRNA responses to viral infection.

Current COVID-19 molecular tests target viral RNA for detection. Unfortunately, even the most advanced current molecular diagnostic tests (i.e. PCR or LAMP amplifying viral RNA) for SARS-CoV-2 require a relatively high viral load to accurately detect infection [37]. Thus, their sensitivity during the early pre-symptomatic phase of disease (incubation period), when the viral load is still low, is poor. Overall sensitivity of current PCR tests has been estimated to be as low as 30-70%[24,25], making it difficult to diagnose infections in many pre-symptomatic and some asymptomatic cases. Our study suggests that SARS-CoV-2 infection induces a miRNA response during the early stages of disease that involves three miRNAs (miR-423-5p, miR-23a-3p and miR-195-5p) that can independently identify COVID-19 cases and distinguish SARS-CoV-2 from influenza infections. Further studies involving larger patient groups, including pre-symptomatic, asymptomatic and mild (non-hospitalised) patients, are planned to assess whether this miRNA biomarker can improve COVID-19 detection rates.

Host molecules correlating with COVID-19 severity, such as the proinflammatory cytokine IL-6, are hypothesized to contribute to adverse COVID-19 outcomes and are the focus of ongoing clinical trials to assess treatments for severe COVID-19 (reviewed in [38]). Our study revealed differential miRNA responses in patients suffering moderate versus more severe COVID-19. Interestingly, four miRNAs (let-7e-5p, miR-651-5p, miR-766-3p, and miR-4433b-5p) were differentially expressed in both groups, suggesting that these molecules might be potential candidates for stratifying patients based on severity. All four miRNAs are predicted to target *HIF1AN* (hypoxia inducible factor 1, alpha subunit inhibitor) (TargetScan, release

7.2)[22] and so may play a role in the hypoxic response during COVID-19. Furthermore, previous studies have demonstrated altered let-7e-5p expression during hypoxic damage to the heart[39] and retina[40], supporting its role in the molecular response to oxygen deprivation. While the other three miRNAs have yet to be linked with hypoxia, miR-766-3p has an established anti-inflammatory role[21], and miR-4433b-5p is part of a biomarker signature of multi-drug resistant tuberculosis (an indicator of patient prognosis)[41].

In summary, this study exemplifies how analysis of miRNA responses to SARS-CoV-2 infection presents novel avenues in the characterization of cellular factors aiding in COVID-19 pathogenesis. It also presents novel opportunities for treatment and diagnosis of viral diseases. Targeting of pro-inflammatory miRNAs could present novel therapeutic opportunities against COVD-19, while miRNA profiling may aid in the disease detection and surveillance.

Method

Ethics statement

Human experimental work was conducted according to the Australian National Health and Medical Research Council Code of Practice. The study was approved by the Alfred Hospital (#280-14) and The University of Melbourne (#2056761, #1443389 and #1955465) Human Research Ethics Committees. The analysis of miRNAs from patient samples was approved by the CSIRO Human Research Ethics Committee (proposal # 2020_19). All animal studies were approved by the CSIRO Australian Centre for Disease Preparedness Animal Ethics Committee (document 1990 for SARS-CoV-2, document #1568 for influenza A(H1N1)) and conducted following the Australian National Health and Medical Research Council Code of Practice for the Care and Use of Animals for Scientific Purposes guidelines for housing and care of laboratory animals. All work with live animals was approved by the Institutional Animal Ethics Committee in accordance with guidelines from the Australian Code for the Care and Use of Animals for Scientific Purposes (8th Edition) and compliant with the Victoria State Prevention of Cruelty to Animals Act 1986 and Part 5 of the Prevention of Cruelty to Animals Regulations 2019. The facility has assurance from the US Office of Laboratory Animal Welfare Assurance (Legacy Assurance ID A5399-01). All animals were acclimatized for at least 7 days prior to entering the study, given food and water ad libitum, and monitored daily. Environmental enrichment was also provided in the cages during the study.

Ferret infection trials

Twenty outbred ferrets (approximately four months of age) were exposed to 4.64×10^4 $TCID_{50}$ of SARS-COV-2 (hCoV-19/Australia/VIC01/2020)[42] by the intranasal route. Prior to any manipulations, animals were immobilised with a mixture of ketamine HCl (5 mg/kg) and medetomidine (0.05 mg/kg); atimepazole was administered for reversal at a dose of 0.25 mg/kg. After virus exposure, animals were monitored for clinical signs of disease, and fever. They were randomly assigned to euthanasia on post-exposure days 3, 5, 7, 9 or 14, when clinical samples including nasal washes, serum and urine were collected together with multiple tissue specimens. Viral loads in tissues, swabs and nasal wash samples

were assessed by qRT-PCR [43]. Eleven ferrets (aged 4-6months) were exposed to 1 x 10^5 TCID50 of influenza A (H1N1) virus as described[44].

RNA isolation and heparinase treatment

Total RNA was isolated from 200 µL of human plasma and 150 µL of ferret serum using the miRNeasy micro kit (Qiagen) as per the manufacturer's instructions with one modification: glycogen (10 µg, Sigma Aldrich, G1767) was added as a carrier to each sample after lysis with Qiazol. As the human plasma samples were originally obtained using sodium heparin vacutainers, the eluted RNA was treated with 1U heparinase I (Sigma Aldrich, H2519) at 25°C for 30 min to remove any remaining heparin.

Next-generation sequencing

Complementary DNA libraries were generated using the QIAseq miRNA Library Kit and QIAseq miRNA NGS 48 Index IL (Qiagen) as per the manufacturer's protocol (handbook HB-2157-007 March 2020), with slight modifications: 5 µL of eluted RNA was used as the template and the libraries underwent 24 cycles of amplification. All libraries were analysed on the Bioanalyser 2100 using the High Sensitivity DNA Kit (Agilent) to ensure correct insert size and minimal adapter or primer carryover. Libraries were then sent to the Australian Genome Research Facility (AGRF) for 100 bp single end sequencing on the NovaSeq 6000 (Illumina). Due to technical issues, 1 V1 COVID samples could not be sequenced but was used in qRT-PCR validation.

Data pre-processing and differential expression

Adapters were trimmed using cutadapt[45] with a read length parameter (18 - 26 nucleotides). The remaining reads were examined using FastQC (www.bioinformatics.babraham.ac.uk/projects/fastqc/) to ensure high-quality data. miRDeep2 quantifier[46] was used to map and quantify reads against the latest miRBase human reference (version 22)[47]. Raw read counts were normalized and differential expression analysis was completed using the DESeq2[48] package in R. An adjusted False Discovery Rate (FDR) p-value of <0.05 was used to identify differentially expressed miRNAs.

Machine learning

All machine learning analysis was conducted using the scikit-learn[49] module in python. Normalized reads were first examined for highly correlated miRNAs; any pairs with a Pearson Rof >0.8 or <-0.8 had one member removed. Highly correlated features (miRNAs) can impact the performance of machine learning algorithms. Multicollinearity can cause skewed or misleading results, especially in models such as logistic regression. The remaining normalized miRNA counts were scaled using either a standard z-score transformation or a robust scaler (where the median is removed and the data is scaled according to the interquartile range). Feature selection was performed using recursive feature elimination (RFE) to identify the miRNAs that contributed the most to the classification model. For binary classification, a logistic regression model was used. For multiclass classification, a linear support vector classifier was used.

Once the optimal number of features (miRNAs) was selected, the data was PCA transformed. Each model underwent hyperparameter tuning using GridSearchCV. To assess the performance of the classification model, the data was randomly split into 70% labelled training data and 30% unlabelled test data, and the predicted classes of the test data samples were compared to the true classes. This process was repeated 1,000 times to ensure confidence in the classification performance. The machine learning models were assessed on their accuracy (how many of the predictions were correct), precision (how many of the predicted positives were true positives), and recall (how many of the true positives were found by the model). The logistic regression model was also assessed using the receiver operating characteristic area under the curve (ROC AUC), which is a succinct metric to describe a binary classification model[12].

qRT-PCR

MicroRNA cDNA was generating using the TaqMan Advanced miRNA cDNA Synthesis Kit (Applied Biosystems) with 2 µL of input RNA as per manufacturer's instructions. qPCR was conducted using 1X TaqMan Advanced miRNA Assay, 1X TaqMan Fast Universal PCR Master Mix, no AmpErase UNG (Applied Biosystems) and 5 µl of 1:10 diluted cDNA product, using standard PCR cycling conditions (95 °C for 20 sec, 40 cycles of 95 °C for 1 sec, 60 °C for 20 sec). Cycle threshold for all assays was set to 0.1. Data is presented as fold over detectable, as previously described[50], with a detectability cut off of C_T = 40.

Cytokine analysis

Plasma was diluted 1:2 and cytokine abundance measured using the LEGENDplex™ Human Inflammation Panel 1 kit, as per manufacturer's instructions (BioLegend).

Statistics

Statistical analyses were completed using the SciPy v1.6.0 package[51]. All measurements were taken from distinct samples. Differences in qRT-PCR results were examined using a one-sided Mann-Whitney U test due to the non-parametric nature of the fold-over-detectable transformation. Normality was tested using a combination skew and kurtosis test (scipy.stats.normaltest()). Differences in IL-6 expression was assessed using a one-sided t-test. A p-value <0.05 was considered significant. All statistics and p-values can be found in Supplementary Table 5.

Declarations

Acknowledgments

We are grateful for support from our colleagues at the Australian Centre for Disease Preparedness (https://www.grid.ac/institutes/grid.413322.5) for providing the facility used in the completion of this work. We acknowledge funding from the Coalition for Epidemic Preparedness Innovations (CEPI) for

supporting ferret COVID-19 studies. This project was supported by resources and expertise provided by CSIRO IMT Scientific Computing.

Author Contributions

R.J.F.: conception, design of the work, acquisition, analysis of data, interpretation of data, manuscript drafting and revision. C.L.R.: acquisition, analysis of data. L.C.R.: acquisition, analysis of data, manuscript drafting and revision. T.H.O.N.: acquisition, analysis of data. L.H.: acquisition, analysis of data. L.K.: acquisition, analysis of data. A.C.C.: acquisition, analysis of data. K.K.: acquisition, analysis of data, manuscript drafting and revision. G.G.A.: design of the work, acquisition, analysis of data, manuscript drafting and revision. G.A.M.: design of the work, acquisition, analysis of data, manuscript drafting and revision. S.S.V.: design of the work, acquisition, analysis of data, manuscript drafting and revision. C.H.F.: conception, interpretation of data, manuscript drafting and revision. C.C.: conception, acquisition, analysis of data, interpretation of data, manuscript drafting and revision. C.R.S.: conception, design of the work, interpretation of data, manuscript drafting and revision.

Declaration of Interests

CSIRO and Exios Bio have filed a provisional patent (docket no.380.35.0001U1) around the use of microRNAs for the early detection of COVID-19. R.J.F., C.C., C.H.F. and C.R.S. are inventors on the patent. The patent relates to microRNAs associated with detection of early-stage COVID-19 (data shown in figures 2 and 4 in the present study).

References

1 Chan, Y. P. *et al.* Biochemical, conformational, and immunogenic analysis of soluble trimeric forms of henipavirus fusion glycoproteins. *J Virol* **86**, 11457-11471, doi:JVI.01318-12 [pii] 10.1128/JVI.01318-12.

2 Wu, C. *et al.* Risk Factors Associated With Acute Respiratory Distress Syndrome and Death in Patients With Coronavirus Disease 2019 Pneumonia in Wuhan, China. *JAMA Intern Med* **180**, 934-943, doi:10.1001/jamainternmed.2020.0994 (2020).

3 Mehta, P. *et al.* COVID-19: consider cytokine storm syndromes and immunosuppression. *Lancet* **395**, 1033-1034, doi:10.1016/S0140-6736(20)30628-0 (2020).

4 Zhang, X. *et al.* Viral and host factors related to the clinical outcome of COVID-19. *Nature* **583**, 437-440, doi:10.1038/s41586-020-2355-0 (2020).

5 Del Valle, D. M. *et al.* An inflammatory cytokine signature predicts COVID-19 severity and survival. *Nat Med*, doi:10.1038/s41591-020-1051-9 (2020).

6 Xu, Z. *et al.* Pathological findings of COVID-19 associated with acute respiratory distress syndrome. *Lancet Respir Med* **8**, 420-422, doi:10.1016/S2213-2600(20)30076-X (2020).

7 Bartel, D. P. Metazoan MicroRNAs. *Cell* **173**, 20-51, doi:10.1016/j.cell.2018.03.006 (2018).

8 Griffiths-Jones, S., Grocock, R. J., van Dongen, S., Bateman, A. & Enright, A. J. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res* **34**, D140-144, doi:10.1093/nar/gkj112 (2006).

9 Friedman, R. C., Farh, K. K., Burge, C. B. & Bartel, D. P. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res* **19**, 92-105, doi:10.1101/gr.082701.108 (2009).

10 Lai, F. W., Stephenson, K. B., Mahony, J. & Lichty, B. D. Human coronavirus OC43 nucleocapsid protein binds microRNA 9 and potentiates NF-kappaB activation. *J Virol* **88**, 54-65, doi:10.1128/JVI.02678-13 (2014).

11 Ma, Y. *et al.* 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**, doi:10.1128/JVI.00728-18 (2018).

12 Tribolet, L. *et al.* MicroRNA Biomarkers for Infectious Diseases: From Basic Research to Biosensing. *Front Microbiol* **11**, 1197, doi:10.3389/fmicb.2020.01197 (2020).

13 Peng, X. *et al.* Integrative deep sequencing of the mouse lung transcriptome reveals differential expression of diverse classes of small RNAs in response to respiratory virus infection. *mBio* **2**, doi:10.1128/mBio.00198-11 (2011).

14 Han, M. G. *et al.* Serum MicroRNA Expression Profiling in Mice Infected with Rabies Virus. *Osong Public Health Res Perspect* **2**, 186-191, doi:10.1016/j.phrp.2011.11.043 (2011).

15 Duy, J. *et al.* Circulating microRNA profiles of Ebola virus infection. *Sci Rep* **6**, 24496, doi:10.1038/srep24496 (2016).

16 Cowled, C. *et al.* Circulating microRNA profiles of Hendra virus infection in horses. *Sci Rep* **7**, 7431, doi:10.1038/s41598-017-06939-w (2017).

17 Tambyah, P. A. *et al.* microRNAs in circulation are altered in response to influenza A virus infection in humans. *PLoS One* **8**, e76811, doi:10.1371/journal.pone.0076811 (2013).

18 Xu, N. *et al.* MicroRNA-31 is overexpressed in psoriasis and modulates inflammatory cytokine and chemokine production in keratinocytes via targeting serine/threonine kinase 40. *J Immunol* **190**, 678-688, doi:10.4049/jimmunol.1202695 (2013).

19 Shi, J. *et al.* MiR-31 Mediates Inflammatory Signaling to Promote Re-Epithelialization during Skin Wound Healing. *J Invest Dermatol* **138**, 2253-2263, doi:10.1016/j.jid.2018.03.1521 (2018).

20 Tian, Y. *et al.* MicroRNA-31 Reduces Inflammatory Signaling and Promotes Regeneration in Colon Epithelium, and Delivery of Mimics in Microspheres Reduces Colitis in Mice. *Gastroenterology* **156**, 2281-2296 e2286, doi:10.1053/j.gastro.2019.02.023 (2019).

21 Hayakawa, K. *et al.* MicroRNA-766-3p Contributes to Anti-Inflammatory Responses through the Indirect Inhibition of NF-kappaB Signaling. *Int J Mol Sci* **20**, doi:10.3390/ijms20040809 (2019).

22 Agarwal, V., Bell, G. W., Nam, J. W. & Bartel, D. P. Predicting effective microRNA target sites in mammalian mRNAs. *Elife* **4**, doi:10.7554/eLife.05005 (2015).

23 Chen, G. *et al.* Clinical and immunological features of severe and moderate coronavirus disease 2019. *J Clin Invest* **130**, 2620-2629, doi:10.1172/JCI137244 (2020).

24 Kanne, J. P., Little, B. P., Chung, J. H., Elicker, B. M. & Ketai, L. H. Essentials for Radiologists on COVID-19: An Update-Radiology Scientific Expert Panel. *Radiology*, 200527, doi:10.1148/radiol.2020200527 (2020).

25 Ai, T. *et al.* Correlation of Chest CT and RT-PCR Testing in Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases. *Radiology*, 200642, doi:10.1148/radiol.2020200642 (2020).

26 Belser, J. A., Katz, J. M. & Tumpey, T. M. The ferret as a model organism to study influenza A virus infection. *Dis Model Mech* **4**, 575-579, doi:10.1242/dmm.007823 (2011).

27 Tijsen, A. J. *et al.* MiR423-5p as a circulating biomarker for heart failure. *Circ Res* **106**, 1035-1039, doi:10.1161/CIRCRESAHA.110.218297 (2010).

28 Tu, H. *et al.* Elevated pulmonary tuberculosis biomarker miR-423-5p plays critical role in the occurrence of active TB by inhibiting autophagosome-lysosome fusion. *Emerg Microbes Infect* **8**, 448-460, doi:10.1080/22221751.2019.1590129 (2019).

29 Lian, F., Cui, Y., Zhou, C., Gao, K. & Wu, L. Identification of a plasma four-microRNA panel as potential noninvasive biomarker for osteosarcoma. *PLoS One* **10**, e0121499, doi:10.1371/journal.pone.0121499 (2015).

30 Mundalil Vasu, M. *et al.* Serum microRNA profiles in children with autism. *Mol Autism* **5**, 40, doi:10.1186/2040-2392-5-40 (2014).

31 Wang, J. *et al.* Serum miR-195-5p is upregulated in gestational diabetes mellitus. *J Clin Lab Anal* **34**, e23325, doi:10.1002/jcla.23325 (2020).

32 Biswas, S., Haleyurgirisetty, M., Lee, S., Hewlett, I. & Devadas, K. Development and validation of plasma miRNA biomarker signature panel for the detection of early HIV-1 infection. *EBioMedicine* **43**, 307-316, doi:10.1016/j.ebiom.2019.04.023 (2019).

33 Suarez, Y., Wang, C., Manes, T. D. & Pober, J. S. Cutting edge: TNF-induced microRNAs regulate TNF-induced expression of E-selectin and intercellular adhesion molecule-1 on human endothelial cells: feedback control of inflammation. *J Immunol* **184**, 21-25, doi:10.4049/jimmunol.0902369 (2010).

34 Gwiggner, M. *et al.* MicroRNA-31 and MicroRNA-155 Are Overexpressed in Ulcerative Colitis and Regulate IL-13 Signaling by Targeting Interleukin 13 Receptor alpha-1. *Genes (Basel)* **9**, doi:10.3390/genes9020085 (2018).

35 Yu, R. Q. *et al.* Small RNA Sequencing Reveals Differentially Expressed miRNAs in Necrotizing Enterocolitis in Rats. *Biomed Res Int* **2020**, 5150869, doi:10.1155/2020/5150869 (2020).

36 Beattie, R. M., Ashton, J. J. & Penman, I. D. COVID-19 and the gastrointestinal tract: recent data. *Frontline Gastroenterol* **11**, 371-374, doi:10.1136/flgastro-2020-101602 (2020).

37 Kucirka, L. M., Lauer, S. A., Laeyendecker, O., Boon, D. & Lessler, J. Variation in False-Negative Rate of Reverse Transcriptase Polymerase Chain Reaction-Based SARS-CoV-2 Tests by Time Since Exposure. *Ann Intern Med* **173**, 262-267, doi:10.7326/M20-1495 (2020).

38 Melo Silva Junior, M. L., Souza, L. M. A., Dutra, R., Valente, R. G. M. & Melo, T. S. Review on therapeutic targets for COVID-19: insights from cytokine storm. *Postgrad Med J*, doi:10.1136/postgradmedj-2020-138791 (2020).

39 Chouvarine, P., Legchenko, E., Geldner, J., Riehle, C. & Hansmann, G. Hypoxia drives cardiac miRNAs and inflammation in the right and left ventricle. *J Mol Med (Berl)* **97**, 1427-1438, doi:10.1007/s00109-019-01817-6 (2019).

40 Desjarlais, M. *et al.* MicroRNA expression profile in retina and choroid in oxygen-induced retinopathy model. *PLoS One* **14**, e0218282, doi:10.1371/journal.pone.0218282 (2019).

41 Wang, C. *et al.* A Group of Novel Serum Diagnostic Biomarkers for Multidrug-Resistant Tuberculosis by iTRAQ-2D LC-MS/MS and Solexa Sequencing. *Int J Biol Sci* **12**, 246-256, doi:10.7150/ijbs.13805 (2016).

42 Caly, L. *et al.* Isolation and rapid sharing of the 2019 novel coronavirus (SARS-CoV-2) from the first patient diagnosed with COVID-19 in Australia. *Med J Aust* **212**, 459-462, doi:10.5694/mja2.50569 (2020).

43 Corman, V. M. *et al.* Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Euro Surveill* **25**, doi:10.2807/1560-7917.ES.2020.25.3.2000045 (2020).

44 Rockman, S. *et al.* Control of pandemic (H1N1) 2009 influenza virus infection of ferret lungs by non-adjuvant-containing pandemic and seasonal vaccines. *Vaccine* **30**, 3618-3623, doi:10.1016/j.vaccine.2012.03.043 (2012).

45 Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *2011* **17**, 3, doi:10.14806/ej.17.1.200 (2011).

46 Friedlander, M. R., Mackowiak, S. D., Li, N., Chen, W. & Rajewsky, N. miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res* **40**, 37-52, doi:10.1093/nar/gkr688 (2012).

47 Kozomara, A., Birgaoanu, M. & Griffiths-Jones, S. miRBase: from microRNA sequences to function. *Nucleic Acids Res* **47**, D155-D162, doi:10.1093/nar/gky1141 (2019).

48 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).

49 Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python *Journal of Machine Learning Research* **12**, 2825 - 2830 (2011).

50 Hardikar, A. A., Farr, R. J. & Joglekar, M. V. Circulating microRNAs: understanding the limits for quantitative measurement by real-time PCR. *J Am Heart Assoc* **3**, e000792, doi:10.1161/JAHA.113.000792 (2014).

51 Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* **17**, 261-272, doi:10.1038/s41592-019-0686-2 (2020).

Table

Table 1. Participant and sample visit details

	HEALTHY	COVID-19	P-value	
Participants, N	10	10	-	
Female, %	60	60	ns[#]	
Age, years (mean ± SD)	53 ± 17.6	53.5 ± 17.2	ns[##]	
Required oxygen therapy, N (%)	-	4 (40%)	-	
COVID-19 Sample Visits	V1	V2	V3	V4
Days post symptom onset (mean ± SD)	8 ± 4	13 ± 6	38 ± 1	51 ± 24
Participants, N	8[]	5	4	3

ns = non-significant [#] Chi-Square test [##] Normality test and two-sided t-test

[]Due to technical issues, only 7 V1 samples were sequenced, however all samples were used for qRT-PCR validation.

Figures

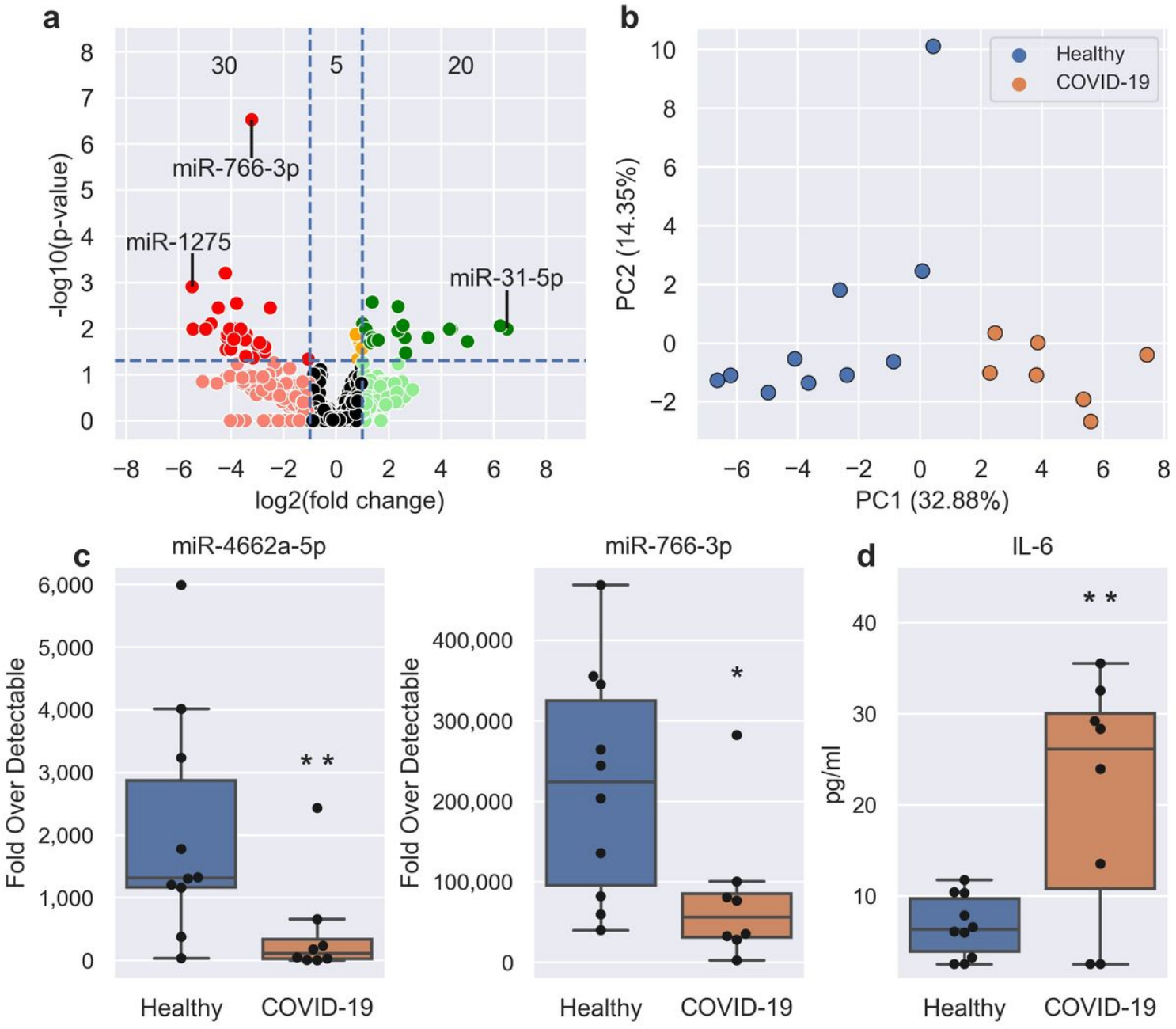

Figure 1

SARS-CoV-2 induces circulating miRNA and cytokine changes. a, Volcano plot showing the increased (green) and decreased (red) DE miRNAs in V1 COVID-19 patients when compared to healthy controls. Horizontal dotted line is the p-value cut-off (False Discovery Rate, FDR<0.05) and the vertical lines are the fold change cut-off (>2 FC). Orange miRNAs are statistically significant but are not >2 FC. The number of statistically significant miRNAs in each section are shown. The most up-regulated, down-regulated, and statistically significant miRNAs have been labelled. b, PCA plot showing the separation of healthy (blue, n=10) and COVID-19 V1 (orange, n=7) samples using the 55 DE miRNAs. c-d, Boxplots of (C) select qRT-PCR validated miRNAs and (D) IL-6 expression in healthy (blue, n=10) and COVID-19 V1 (orange, n=8)

samples. Boxes are the 25th - 75th percentile, line is the median, and whiskers are 1.5x IQR. * p-value < 0.05, ** p-value < 0.01.

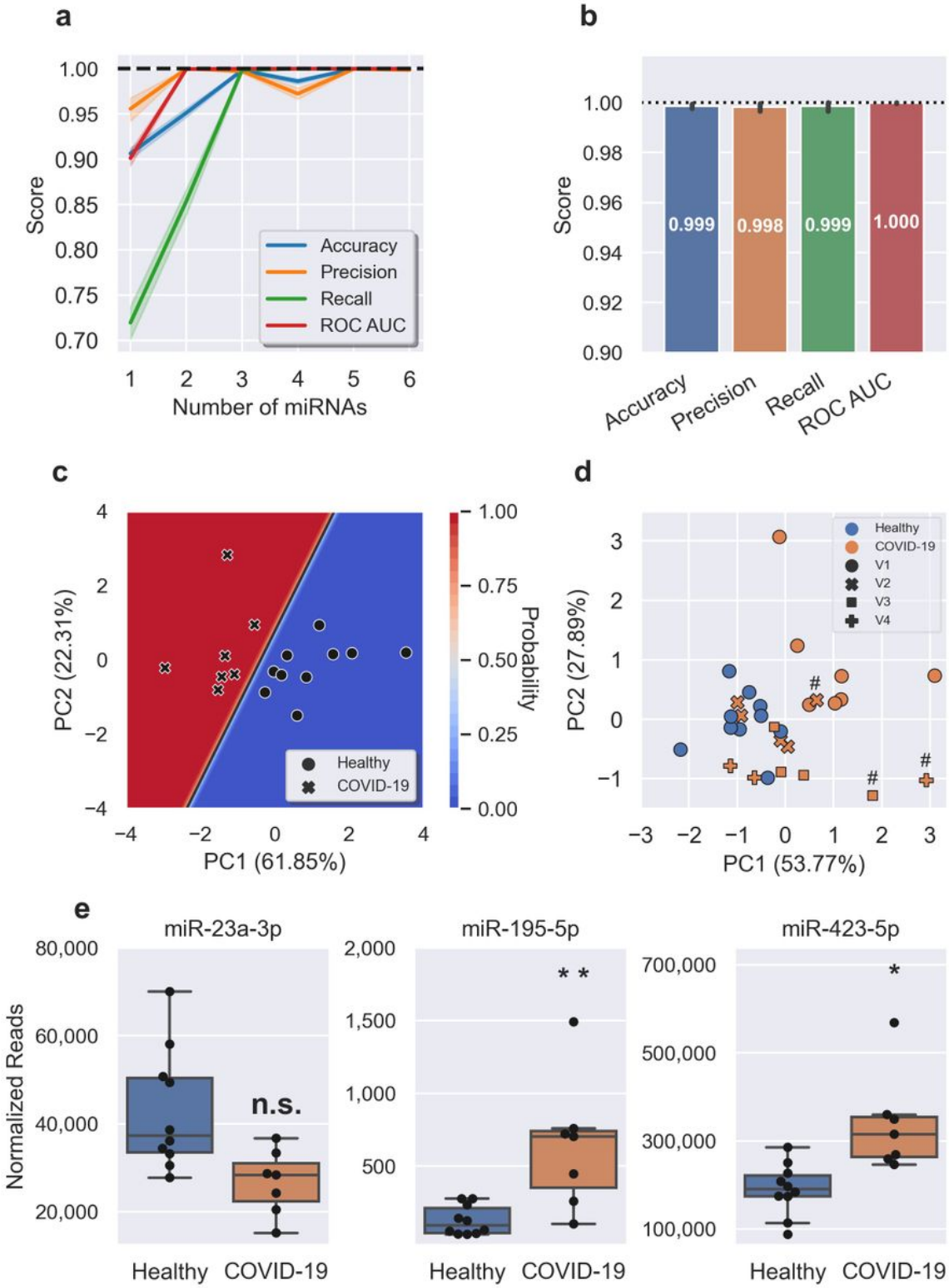

Figure 2

A three miRNA signature classifies COVID with 99.9% accuracy. a, Feature (miRNA) selection lineplot showing the impact of increasing numbers of miRNAs on the performance of a logistic regression model. MicroRNAs were selected using recursive feature elimination to identify the most important miRNAs.

Each combination of miRNAs was randomly assessed 1,000 times. Shaded areas are the 95% CI, and the dotted line is a perfect (100%) score. b, Barplot showing the average score of the three-miRNA signature in predicting healthy controls and COVID-19 patients. Error bars are the 95% CI after 1,000 random iterative assessments. c, Decision boundary graph showing the logistic regression decision point (solid black line) and the probability a person is infected with SARS-CoV-2 (blue to red shading). Datapoints are healthy (circles, n=10) and COVID-19 V1 (crosses, n=7) samples. d, PCA plot based on the three miRNA signature showing all healthy (blue, n=10) and COVID-19 (orange, n=19) samples. Subsequent V2 (crosses, n=5), V3 (squares, n=4), and V4 (plus signs, n=3) samples cluster with the healthy controls, apart from those denoted with a hash (#) – these all came from one participant that was treated in ICU and had not recovered at any visit. e, Boxplots of each of the signature miRNAs in healthy (blue, n=10) and COVID-19 V1 (orange, n=7) samples. Boxes are the 25th - 75th percentile, line is the median, and whiskers are 1.5x IQR. * FDR adjusted p-value < 0.05, ** FDR adjusted p-value < 0.01. n.s. non-significant.

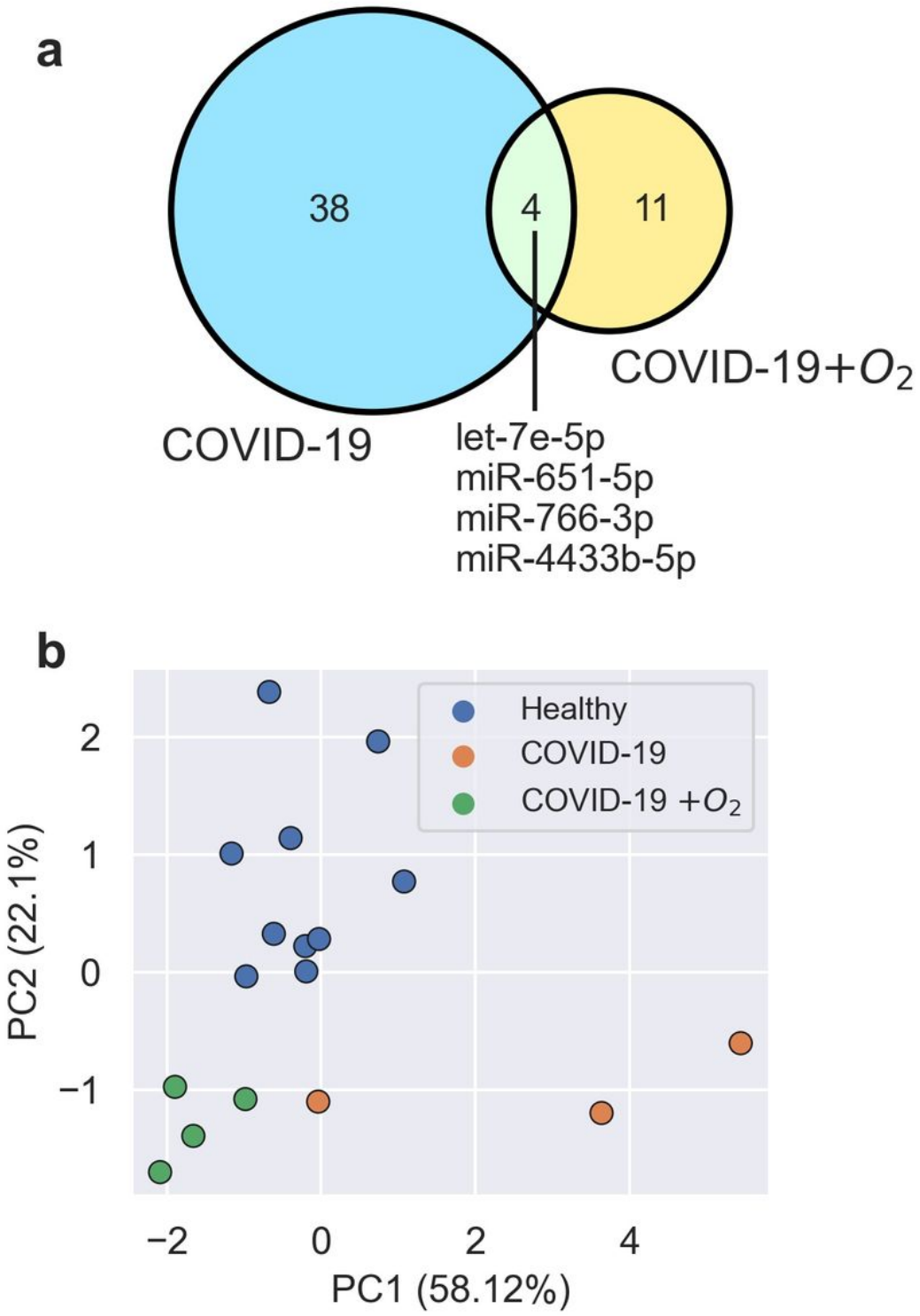

Figure 3

Differential miRNA profiles based on COVID-19 severity. a, Venn diagram of COVID-19 (light blue) and COVID-19 + O2 (yellow) V1 DE miRNAs when compared to healthy controls. b, PCA plot based on the four common DE miRNAs. Healthy (blue, n=10), COVID-19 (orange, n=3) and COVID-19 + O2 (green, n=4) V1 samples.

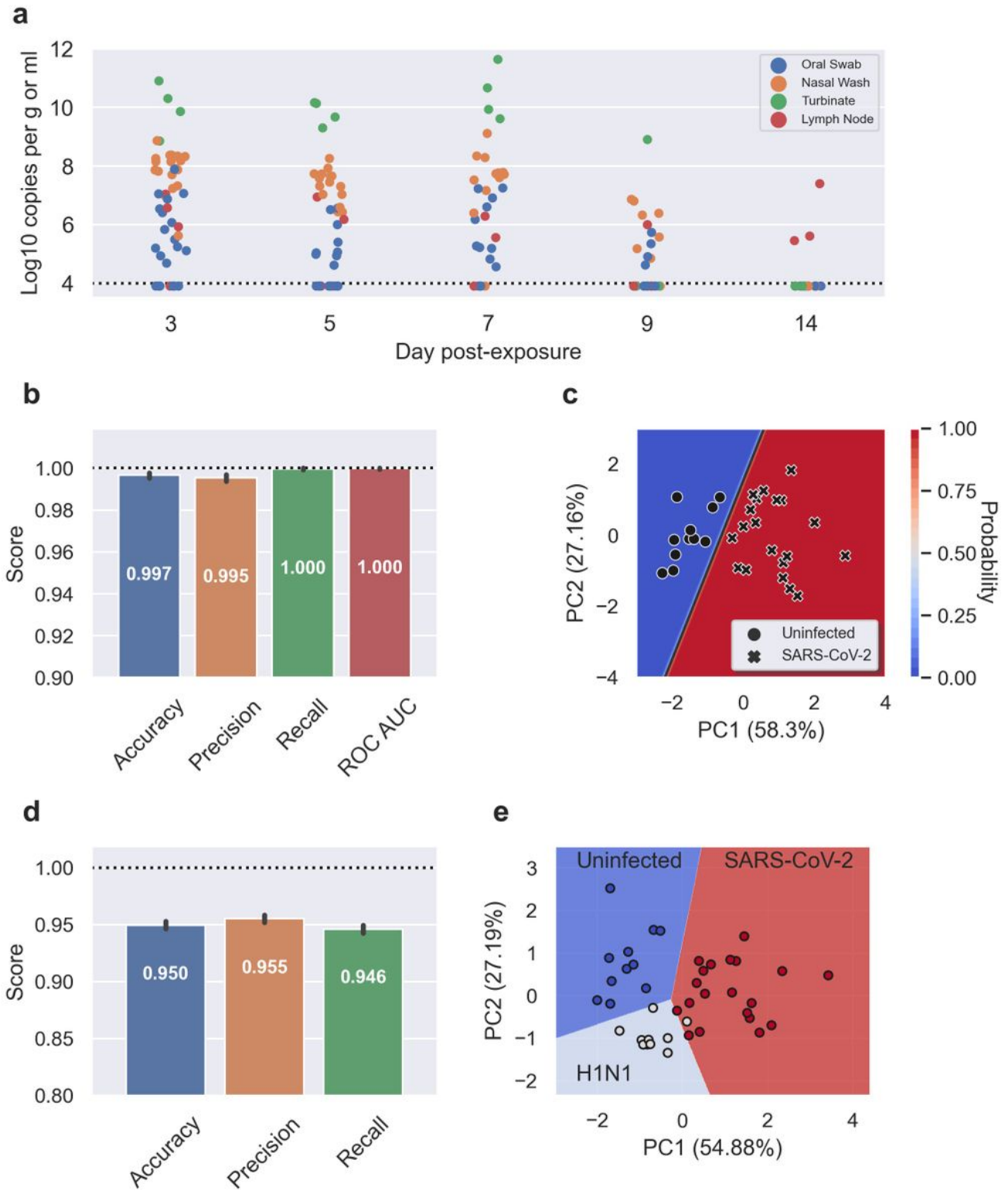

Figure 4

Human miRNA signature accurately identifies influenza and SARS-CoV-2 infection in a ferret model. a, Detection of SARS-CoV-2 viral genomic RNA in the retroperitoneal lymph node (blue), nasal wash (orange), oral swab (green), and turbinate tissue (red) of infected ferrets (n=20, swab and wash samples taken from every ferret at each time point, tissue samples were analysed from the 4 euthanized ferrets at each time point). Data is presented as log10 copies per g of tissue or ml of sample. b, Final metrics of the

trained logistic regression model to identify uninfected or SARS-CoV-2 infected ferrets. Dotted line is a perfect score (or 100%). Error bars are 95% CI for 1,000 random assessments. c, Decision boundary graph showing the logistic regression decision point (solid black line) and the probability a sample is infected with SARS-CoV-2 (blue to red shading). Datapoints are uninfected (circles, n=11) and SARS-CoV-2 infected (crosses, n=20) ferrets. d, Final metrics of the trained linear support vector classifier model to identify uninfected, influenza A (H1N1) virus, or SARS-CoV-2 infected ferrets. Dotted line is a perfect score (or 100%). Error bars are 95% CI for 1,000 random assessments. As ROC AUC is a measure of binary classification (two groups) it is omitted here. e, Decision boundary graph showing the linear support vector classifier decision points and predicted groups: uninfected (blue, n=11), influenza A (H1N1) virus infected (light blue, n=11) or SARS-CoV-2 infected (red, n=20) ferrets.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryInformation.docx