Detection of SARS-CoV-2 infection by microRNA profiling of the upper respiratory tract

Ryan Farr

Commonwealth Scientific and Industrial Research Organisation

Christina Rootes

Commonwealth Scientific and Industrial Research Organisation

John Stenos

Barwon Health

Chwan Hong Foo

Exios Bio

Christopher Cowled

Commonwealth Scientific and Industrial Research Organisation

Cameron Stewart (✉ cameron.stewart@csiro.au)

Commonwealth Scientific and Industrial Research Organisation

Research Article

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

Posted Date: September 28th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-907655/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at PLOS ONE on April 5th, 2022. See the published version at <https://doi.org/10.1371/journal.pone.0265670>.

Abstract

Host biomarkers are increasingly being considered as tools for improved COVID-19 detection and prognosis. We recently profiled circulating host-encoded microRNA (miRNAs) during SARS-CoV-2 infection, revealing a signature that classified COVID-19 cases with 99.9% accuracy. Here we sought to develop a signature suited for clinical application by analyzing specimens collected using minimally invasive procedures. Eight miRNAs displayed altered expression in anterior nasal tissues from COVID-19 patients, with miR-142-3p, a negative regulator of interleukin-6 (IL-6) production, the most strongly upregulated. Supervised machine learning analysis revealed that a three-miRNA signature (miR-30c-2-3p, miR-628-3p and miR-93-5p) independently classifies COVID-19 cases with 100 % accuracy. This study further defines the host miRNA response to SARS-CoV-2 infection and identifies candidate biomarkers for improved COVID-19 detection.

Main Text

Host responses to SARS-CoV-2 infection are currently being examined as biomarkers for both improved detection of pre- or asymptomatic COVID-19 cases [1] and the prognosis of COVID-19 severity [2–4]. In contrast to viral RNA molecules, host biomarkers such as microRNAs (miRNAs) are relatively abundant in the patient during the early pre-symptomatic period. MicroRNAs are small non-coding RNAs that regulate biological processes, including the host antiviral immune response. During the earliest phase of viral infection, prior to symptom onset and detectable virions, the pathogen triggers signaling cascades in the innate effectors of the host immune system. These first line responders (e.g. myeloids) react rapidly, releasing expressed miRNAs in circulation.

We recently characterised changes in the circulating miRNA profile of human plasma observed during SARS-CoV-2 infection [1]. With many molecular COVID-19 tests employing nasal or nasopharyngeal swabs as analytes, here we analysed the miRNA profile in nasal swabs derived from COVID-19 patients and uninfected controls. Swabs of the anterior nares were collected by Barwon Health (Geelong, Australia) from members of the public undergoing COVID-19 testing between July and August 2020 (Table 1). Samples were also collected from persons defined as uninfected controls, who displayed no COVID-19 symptoms and returned negative SARS-CoV-2 PCR test results. Swabs were collected in universal transport medium and stored at -80°C until processed. Protocols for RNA purification, complementary DNA (cDNA) library preparation, next-generation sequencing, data processing, data analysis, and statical analysis are as described [1].

Small RNA seq resulted in 13–46 million (average 23 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 (18–26 nt) and quality, resulting in 2.4–10.6 million (average 5.4 million) reads per sample for further analysis. The majority of sequences were deemed high quality by FASTQC (data not shown). MiRDeep2 mapper was used to identify all known miRNA transcripts amongst the 20 samples (by mapping to the miRBase v22 human dataset) and read counts were determined for each mature miRNA transcript using

miRDeep2 quantifier. A total of 1,495 different 5p or 3p mature miRNA transcripts were detected, corresponding to 1,097 different precursors. The most abundant miRNA in the nasal swab dataset was hsa-miR-16-5p, followed by hsa-miR-29c-3p, hsa-miR-29a-3p and hsa-miR-223-3p (Fig. 1a). A total of 452 (28% of all detected miRNAs) had at least 100 reads (Fig. 1b). We did not observe a significant difference in the total number of miRNAs identified in infected versus uninfected patients (data not shown). Following data normalization, pairwise analyses was performed at the single-miRNA level using median normalised read counts from infected vs uninfected samples, revealing a high degree of similarity (Fig. 1c). Similar trends were also observed between individual samples (data not shown). By applying a stringent data filtering and normalisation strategy, miRNA expression between different individuals was demonstrated to exhibit a low level of biological and technical variation, confirming the suitability of this dataset to assess changes in miRNA expression between patient groups.

Using DESeq2 to perform count-based differential expression (DE) testing, a subset of miRNAs that were up- or down-regulated in COVID-19 cases relative to uninfected controls were identified (Fig. 2a, 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 6 miRNAs, of which four were up-regulated (elevated in infected patients) and two were down-regulated. An additional two 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-142-3p (Fig. 2a), miR-486-5p, and miR-451a, while the most down-regulated were miR-3065-3p (Fig. 2a) and miR-3065-5p. The most statistically significant change was seen in miR-142-3p. Unsupervised analysis of variance using principal components analysis (PCA) involving the eight DE miRNAs showed tight clustering of patient groups (Fig. 2b). Differences in miRNA expression for miR-142-3p, miR-3065-3p and miR-93-5p are shown in Fig. 2c. Upon comparing miRNAs differentially expressed in COVID-19 patients in nasal swabs and plasma [1], two miRNAs (miR-142-3p and miR-3065-3p) were DE in both datasets (Fig. 3a), while miRNAs DE in nasal swabs for the most part showed agreement in terms of upregulation or downregulation without being statistically significant in plasma (Fig. 3b).

We next investigated if, similar to miRNA profiles in plasma, changes in the nasal swab profile could independently classify 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 number needed for accurate prediction. The most predictive miRNAs were selected using recursive feature elimination (Fig. 4a). Measuring three miRNA targets (miR-30c-2-3p, miR-628-3p and miR-93-5p) in combination gave a model with 100% accuracy, 100% precision and 100% recall, with a ROC AUC of 1.0 (Fig. 4b). This composite biomarker was comprised of two miRNAs DE in COVID-19 patients (miR-628-3p (downregulated) and miR-93-5p (upregulated)) and miR-30c-2-3p, which was not DE (Fig. 4d). A decision boundary graph showed clear distinctions between healthy and infected patients based on these three miRNAs (Fig. 4c).

Several miRNAs DE in COVID-19 patient nasal swabs are associated with inflammation. Elevated expression of miR-142-3p has been reported in Crohn's disease and ulcerative colitis, where elevated

levels of miR-142-3p are observed in colon, blood and saliva [5]. Separate studies demonstrated a correlation between elevated miR-142-3p and circulating IL-6 levels in inflammatory bowel disease [6] and miR-142-3p/IL-6 production in dendritic cells stimulated with lipopolysaccharide (LPS) [7]. Studies using luciferase reporters carrying wild-type and altered IL-6 3'UTR confirm *IL6* as a miR-142-3p target gene [7], while miR-142-3p acts in a dose-dependent manner to inhibit *IL6* transcription in polymorphonuclear leukocytes stimulated with LPS [8]. miR-93-5p, also up-regulated in COVID-19 anterior nasal tissues, inhibits the production of IL-6, TNF and IL-1 β in osteoarthritis and diabetic nephropathy models through regulation of high mobility group proteins HMGB1 and HMGA2, respectively [9, 10]. IL-6 is one of the key mediators of viral cytokine storm and inflammation in patients with severe COVID-19 [11]. It is intriguing to speculate that miR-142-3p and miR-93-5p are induced to counteract potentially deleterious effects of elevated IL-6 in COVID-19 patients, a response associated with respiratory failure and death [12]. Intriguingly, the most downregulated miRNA in nasal swabs from COVID-19 patients, the relatively poorly-characterised miR-3065-3p, is also down-regulated in inflamed placental tissue and significantly reduced by LPS stimulation [13]. Additionally, other miRNAs responsive to SARS-CoV-2 infection has no known links to inflammation but have been observed in infection [14, 15].

One of the more dangerous features of COVID-19 is its ability for sustaining human-to-human transmission pre- and asymptotically [16]. U.S. CDC estimates that 40% of transmission occurs prior to symptom onset [16, 17]. Furthermore, approximately 35% of COVID-19 infections remain asymptomatic throughout the entire course of the disease [16]. These traits of COVID-19 have facilitated its rapid spread leading to the current deadly global pandemic, and highlights that innovations are required to fill gaps in the SARS-CoV-2 diagnostic landscape. Here we have shown that positive COVID-19 PCR test results correspond to a change in the nasal swab miRNA profile that can independently classify disease cases. Further studies involving larger patient groups, including pre-symptomatic, asymptomatic and different (e.g. severity, variants) infections are planned to assess whether this pattern is observed during the COVID-19 incubation period (median 6.5 days) and would thus have real-world application for improved disease detection or prognosis. As miRNA responses are reflective of the host response to infection, miRNA biomarkers could also provide clinical utility in the provision of infection evidence to reduce false-negative rates with PCR testing [17].

List Of Abbreviations

CDC
Centre for Disease Control
cDNA
complementary DNA
COVID-19
coronavirus disease of 2019
DE
differential expression
FDR

false discovery rate

IL

interleukin

LPS

lipopolysaccharide

miRNA

microRNA

PCA

principle component analysis

PCR

polymerase chain reaction

Declarations

Ethics approval and consent to participate:

The analysis of miRNAs from clinical samples was approved by the CSIRO Human Research Ethics Committee (proposal # 2020_19_LR) with all patient information de-identified and samples randomised prior to RNA isolation.

Consent for publication:

Not applicable

Availability of data and materials

The datasets generated and/or analysed during the current study are available in the NCBI short read archive (SRA) (<https://www.ncbi.nlm.nih.gov/sra>)

Competing 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.

Funding:

This work was supported by the Commonwealth Scientific and Industrial Research Organisation (CSIRO).

Author contributions:

- Ryan J. Farr: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing
- Christina L. Rootes: Data curation, Investigation, Project administration
- John Stenos: Data curation, Formal analysis, Writing – review & editing

- Chwan Hong Foo: Conceptualization, Formal analysis, Writing – original draft, Writing – review & editing
- Christopher Cowled: Conceptualization, Data curation, Formal analysis, Methodology, Writing – original draft, Writing – review & editing
- Cameron R. Stewart: Conceptualization, Data curation, Formal analysis, Funding acquisition, Methodology, Project administration, Supervision, Writing – original draft, Writing – review & editing

Acknowledgements:

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.

References

1. Farr, R.J., et al., *Altered microRNA expression in COVID-19 patients enables identification of SARS-CoV-2 infection*. PLoS Pathog, 2021. **17**(7): p. e1009759.
2. Del Valle, D.M., et al., *An inflammatory cytokine signature predicts COVID-19 severity and survival*. Nat Med, 2020. **26**(10): p. 1636-1643.
3. Zhang, X., et al., *Viral and host factors related to the clinical outcome of COVID-19*. Nature, 2020. **583**(7816): p. 437-440.
4. Kreutmair, S., et al., *Distinct immunological signatures discriminate severe COVID-19 from non-SARS-CoV-2-driven critical pneumonia*. Immunity, 2021. **54**(7): p. 1578-1593 e5.
5. Schaefer, J.S., et al., *MicroRNA signatures differentiate Crohn's disease from ulcerative colitis*. BMC Immunol, 2015. **16**: p. 5.
6. Nijakowski, K. and A. Surdacka, *Salivary Biomarkers for Diagnosis of Inflammatory Bowel Diseases: A Systematic Review*. Int J Mol Sci, 2020. **21**(20).
7. Sun, Y., et al., *Targeting of microRNA-142-3p in dendritic cells regulates endotoxin-induced mortality*. Blood, 2011. **117**(23): p. 6172-83.
8. Huang, H.C., et al., *MicroRNA-142-3p and let-7g Negatively Regulates Augmented IL-6 Production in Neonatal Polymorphonuclear Leukocytes*. Int J Biol Sci, 2017. **13**(6): p. 690-700.
9. Meng, Y., et al., *Knockdown of exosomemediated IncPVT1 alleviates lipopolysaccharideinduced osteoarthritis progression by mediating the HMGB1/TLR4/NFkappaB pathway via miR935p*. Mol Med Rep, 2020. **22**(6): p. 5313-5325.

10. Wang, X., et al., *LncRNA HCP5 knockdown inhibits high glucose-induced excessive proliferation, fibrosis and inflammation of human glomerular mesangial cells by regulating the miR-93-5p/HMGA2 axis*. BMC Endocr Disord, 2021. **21**(1): p. 134.
11. Lee, D.W., et al., *Current concepts in the diagnosis and management of cytokine release syndrome*. Blood, 2014. **124**(2): p. 188-95.
12. Chen, L.Y.C., et al., *Confronting the controversy: interleukin-6 and the COVID-19 cytokine storm syndrome*. Eur Respir J, 2020. **56**(4).
13. Jang, H.Y., et al., *Identification of microRNAs and their target genes in the placenta as biomarkers of inflammation*. Clin Exp Reprod Med, 2020. **47**(1): p. 42-53.
14. McGowan, K., K.J. Simpson, and J. Petrik, *Expression Profiles of Exosomal MicroRNAs from HEV- and HCV-Infected Blood Donors and Patients: A Pilot Study*. Viruses, 2020. **12**(8).
15. Cui, L., et al., *Serum microRNA expression profile distinguishes enterovirus 71 and coxsackievirus 16 infections in patients with hand-foot-and-mouth disease*. PLoS One, 2011. **6**(11): p. e27071.
16. .
17. Kucirka, L.M., et al., *Variation in False-Negative Rate of Reverse Transcriptase Polymerase Chain Reaction-Based SARS-CoV-2 Tests by Time Since Exposure*. Ann Intern Med, 2020. **173**(4): p. 262-267.

Tables

Please see the supplementary files section to view the tables.

Figures

A.

B.

C.

Figure 1

Overview of host-encoded miRNAs in nasal swab samples. A, Treemap plot displaying the relative abundance of the most highly-expressed miRNAs in nasal swab samples. The most prevalent miRNA was miR-16-5p, followed by miR-29a-3p, miR-29c-3p and miR-223-3p. B, Line plot showing the relative abundance of all host miRNAs identified in nasal swab samples. A total of 1495 miRNAs were detected across all 20 samples, of which 452 were detected at greater than 100 reads. C, Scatter plot illustrating inter-sample variance at the single-miRNA level. Each individual point represents a single mature miRNA, shown as the median of DESeq2-normalized read counts in each group and drawn in log10 scale on both axes. MiRNAs that lie exactly on the diagonal midline have equal expression in both groups, while miRNAs located further away from the midline are potential candidates for differential expression.

Figure 2

SARS-CoV-2 induces significant changes in the miRNA profile from patient nasal swabs. A, Volcano plot showing the increased (green) and decreased (red) DE miRNAs in 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 (adjusted P-value < 0.05) in each section are shown: < -1 Log2 FC (2 miRNAs), between -1 and 1 Log2 FC (2 miRNAs), and > 1 Log2 FC (4 miRNAs). The most up-regulated, down-regulated, and statistically significant miRNAs have been labelled. B, PCA plot showing the separation of healthy (blue) and COVID-19 (orange) samples using the 8 DE miRNAs. C, Boxplots of select miRNAs in healthy (blue) and COVID-19 (orange) samples. Boxes are the 25th - 75th percentile, line is the median, and whiskers are 1.5x IQR. ** p-value < 0.01, *** p-value < 0.001.

A **B** **Figure 3**

Comparisons of DE miRNAs induced by SARS-CoV-2 infection in nasal swabs and plasma. A, Venn diagram identifying DE miRNAs in plasma and nasal swabs miRNA datasets. B, Results from DE miRNA analysis from nasal swab and plasma datasets for the 8 miRNAs listed.

Figure 4

A miRNA signature in nasal swabs classifies COVID with 100% 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) and COVID-19 (crosses) samples. D, Boxplots of each of the signature miRNAs in healthy (blue) and COVID-19 (orange) 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.

Supplementary Files

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

- [Table1.xlsx](#)
- [SupplementaryTable1.xlsx](#)