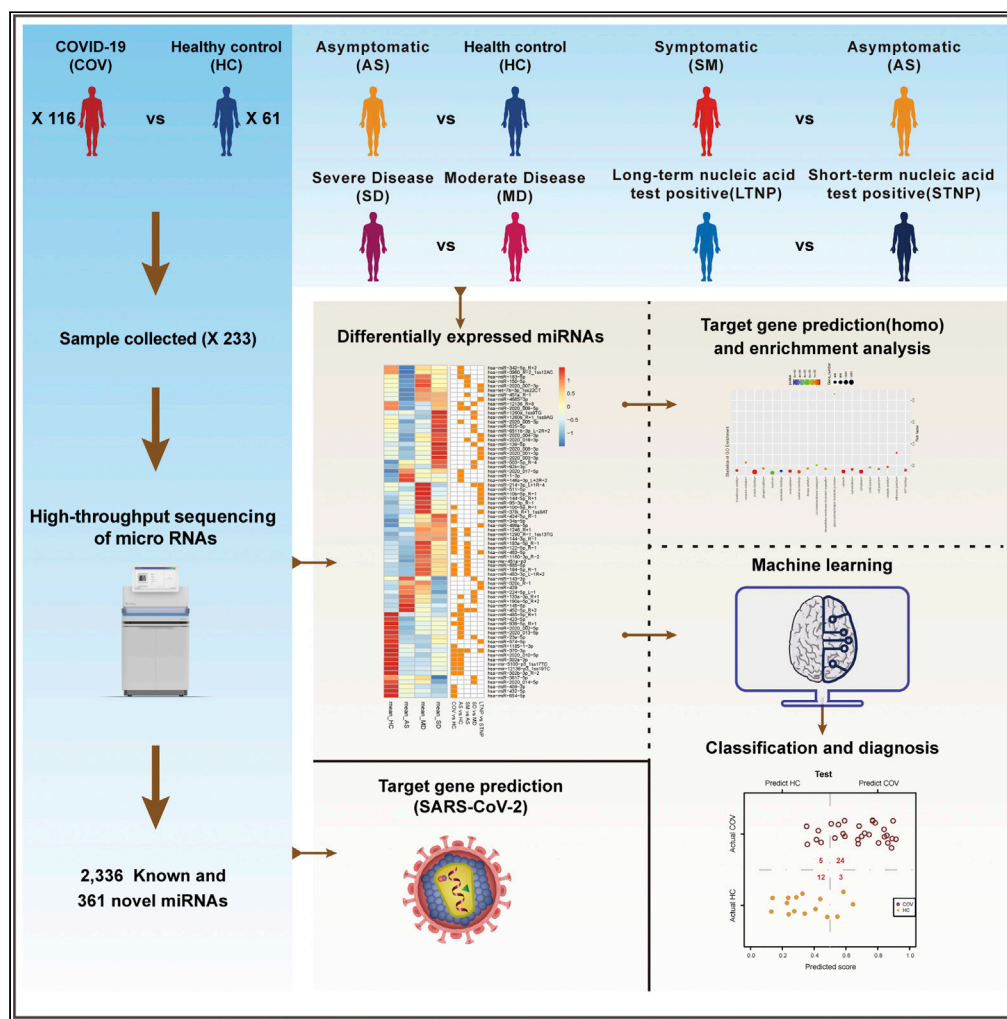


Article

Distinct miRNAs associated with various clinical presentations of SARS-CoV-2 infection



Qiqi Zeng, Xin Qi, Junpeng Ma, ..., Binghong Zhang, Bingyin Shi, Chengsheng Zhang

chenxi@spec-ally.com (X.C.)
huke-rmhospital@163.com (K.H.)
zhhong6288@163.com (B.Z.)
shibingy@126.com (B.S.)
cszhang99@126.com (C.Z.)

Highlights

2,336 known and 361 novel miRNAs identified in this study

85 miRNAs associated with COVID-19

A panel of miRNAs targeting the viral or cellular genes

Machine learning using miRNAs for classification of COVID-19



Article

Distinct miRNAs associated with various clinical presentations of SARS-CoV-2 infection

Qiqi Zeng,^{1,2,15} Xin Qi,^{1,2,15} Junpeng Ma,^{1,2,15} Fang Hu,^{1,2,15} Xiaorui Wang,^{1,2,15} Hongyu Qin,^{1,2,15} Mengyang Li,^{1,2,15} Shaoxin Huang,^{3,15} Yong Yang,^{3,15} Yixin Li,^{1,2} Han Bai,^{1,2} Meng Jiang,^{1,2} Doudou Ren,^{1,2} Ye Kang,^{1,2} Yang Zhao,⁴ Xiaobei Chen,⁵ Xi Ding,⁶ Di Ye,⁷ Yankui Wang,⁸ Jianguo Jiang,⁹ Dong Li,¹⁰ Xi Chen,^{3,11,*} Ke Hu,^{4,*} Binghong Zhang,^{12,*} Bingyin Shi,^{6,*} and Chengsheng Zhang^{1,2,13,14,16,*}

SUMMARY

MicroRNAs (miRNAs) have been shown to play important roles in viral infections, but their associations with SARS-CoV-2 infection remain poorly understood. Here, we detected 85 differentially expressed miRNAs (DE-miRNAs) from 2,336 known and 361 novel miRNAs that were identified in 233 plasma samples from 61 healthy controls and 116 patients with COVID-19 using the high-throughput sequencing and computational analysis. These DE-miRNAs were associated with SARS-CoV-2 infection, disease severity, and viral persistence in the patients with COVID-19, respectively. Gene ontology and KEGG pathway analyses of the DE-miRNAs revealed their connections to viral infections, immune responses, and lung diseases. Finally, we established a machine learning model using the DE-miRNAs between various groups for classification of COVID-19 cases with different clinical presentations. Our findings may help understand the contribution of miRNAs to the pathogenesis of COVID-19 and identify potential biomarkers and molecular targets for diagnosis and treatment of SARS-CoV-2 infection.

INTRODUCTION

The pandemic of the coronavirus disease 2019 (COVID-19), which was caused by SARS-CoV-2 infection, has posed unprecedented challenges to the international communities. As of December 29, 2021, 281,808,270 confirmed cases of COVID-19, including 5,411,759 deaths worldwide were reported to the World Health Organization ([WHO Coronavirus \(COVID-19\) Dashboard](#), 2021). The emerging variants of SARS-CoV-2 have resulted in the devastating surge of COVID-19 cases in several countries and have caused serious concerns ([Kupferschmidt and Wadman, 2021](#)). The clinical presentations and disease courses of SARS-CoV-2 infection are highly variable, ranging from asymptomatic infection (AS) to severe or critical conditions ([Rothe et al., 2020](#); [Buitrago-Garcia et al., 2020](#); [Guan et al., 2020](#); [Wu and McGoogan, 2020](#); [Price-Haywood et al., 2020](#)). While many studies have been conducted to understand the differences between the moderate disease (MD) and the severe disease (SD), the AS remains poorly characterized. In addition, some of the patients were tested positive for the nucleic acid of SARS-CoV-2 by RT-PCR but did not become negative for a longer period of time (≥ 60 days, herein designated as long-term nucleic acid test positive, LTNP) whereas others turned into negative for the viral nucleic acid in a shorter period of time (≤ 45 days, designated as short-term nucleic acid test positive, STNP). However, the underlying mechanisms responsible for the LTNP and STNP remain unknown.

MicroRNAs (miRNAs) are small single-stranded RNA molecules with a length of 18–28 nucleotides and have been suggested to play important roles in a variety of physiological and pathological processes in animals and humans, including development, immune responses, inflammation, and apoptosis ([Bartel, 2004](#)). In particular, miRNAs were involved in the replication process of a number of viruses, such as hepatitis C virus, HIV-1, and influenza viruses ([Lee et al., 2017](#); [Narla et al., 2018](#); [Zhao et al., 2019](#)). Recent studies have also suggested that miRNAs may be involved in the pathogenesis of COVID-19 ([Chauhan et al., 2021](#); [Arisan et al., 2020](#); [Hosseini Rad Sm and McLellan, 2020](#)). However, the miRNA alternations associated with SARS-CoV-2 infection and the underlying mechanisms are largely unknown. In this study, to investigate the potential contributions of miRNAs to the pathogenesis of COVID-19, we performed

¹Precision Medicine Center, The First Affiliated Hospital of Xi'an Jiaotong University, 277 Yanta West Road, Xi'an 710061, China

²The MED-X Institute, The First Affiliated Hospital of Xi'an Jiaotong University, Western China Science and Technology Innovation Harbor, Building 21, Xi'an 710000, China

³SpecAlly Life Technology Co, Ltd, Wuhan East Lake High-tech Development Zone, 666 Gaoxin Road, Wuhan 430075, China

⁴Department of Respiratory and Critical Care Medicine, Renmin Hospital of Wuhan University, 99 Zhangzhidong Road, Wuhan 430060, China

⁵Department of Infectious Diseases, The Renmin Hospital of Wuhan University, East Campus, East Lake New Technology Development Zone, Gaoxin 6th Road, Wuhan 430040, China

⁶Department of Endocrinology, The First Affiliated Hospital of Xi'an Jiaotong University, 277 Yanta West Road, Xi'an 710061, China

⁷Department of Rehabilitation, The Renmin Hospital of Wuhan University, 238 Jiefang Road, Wuchang District, Wuhan 430060, China

⁸Dialysis Center, The Renmin Hospital of Wuhan University, East Campus, East Lake New Technology Development Zone, Gaoxin 6th Road, Wuhan 430040, China

⁹LC-Bio Technologies (Hangzhou) Co., Ltd., Hangzhou 310000, China

¹⁰Department of Clinical Laboratory, The Renmin Hospital of Wuhan University, East Campus, East Lake New

Continued



high-throughput sequencing of the plasma miRNA library and identified 2,336 known and 361 novel miRNAs in 233 plasma samples from 61 healthy controls and 116 patients with COVID-19 with various clinical presentations. We revealed 85 differentially expressed miRNAs that were associated with the SARS-CoV-2 infections, disease severity, and viral persistence in the patients with COVID-19, such as hsa-miR-370-3p, hsa-miR-146a-3p, hsa-miR-885-5p, hsa-miR-214-3p, and hsa-miR-10b-5p. We also identified a panel of miRNAs that may directly target the viral genes of SARS-CoV-2. Moreover, we detected a number of miRNAs that may target the cellular genes involved in the life cycle of SARS-CoV-2, including ACE2, AXL, TMEM106B, and TOMM70. The gene ontology and KEGG pathway analyses of the significant miRNAs revealed their connections to immune responses, viral infections, and lung diseases. Finally, we established and tested a machine learning model using the differentially expressed miRNAs between various compared groups for classification of COVID-19 cases with different clinical presentations. Our findings may help understand the contribution of miRNAs to the pathogenesis of COVID-19 and identify potential biomarkers and molecular targets for the diagnosis and treatment of SARS-CoV-2 infection.

RESULTS

Plasma miRNA profiles in patients with COVID-19 with various clinical presentations

In this study, we first identified 2,336 known and 361 novel miRNAs in 233 plasma samples from 61 healthy controls (HC) and 116 SARS-CoV-2-infected subjects (COV) by high-throughput sequencing (Figure 1A; Data S2 and S3). Of note, we identified 2,336 of the 2,656 known miRNAs (approximately 88%) in the current database (<http://www.mirbase.org/ftp.shtml>), suggesting that our experimental procedures and the analytical pipelines were robust and well performed. We subsequently performed a series of computational and statistical analyses on these miRNAs to identify the target genes (both the viral and cellular genes), the differentially expressed miRNAs (hereinafter referred to as DE-miRNAs) between various groups (Figure 1A), Gene ontology (GO) and KEGG pathways, and built an XGBoost machine learning model to identify potential miRNA biomarkers for clinical classification and diagnosis of patients with COVID-19. Here, we compared the miRNA expression levels between eleven groups (COV vs HC, AS vs HC, SM vs HC, SM vs AS, MD vs AS, SD vs AS, SD vs MD, SD22 vs MD12, MD12 vs MD-R, SD22 vs SD-R, and LTNP vs STNP), respectively, and discovered 85 DE-miRNAs (Figure 1B; Tables 1, S2; Data S4 and S5). Further analysis revealed that 67 of the 85 DE-miRNAs were associated with immune responses, virus infections, and/or lung diseases (Figure 1B) (Pulati et al., 2019; Chen et al., 2019), suggesting that these DE-miRNAs may play important roles in the pathogenesis of COVID-19.

DE-miRNAs associated with SARS-CoV-2 infection

We then compared the miRNA expression levels between the SARS-CoV-2-infected individuals (COV) and the healthy control (HC), and identified 112 upregulated (fold change > 1.5, $p \leq 0.05$) and 44 downregulated DE-miRNAs (fold change < 0.67, $p \leq 0.05$) (Figure 2A). To improve the accuracy and reliability of the DE-miRNA candidates, we removed the DE-miRNAs with the average expression level lower than 10 copies and retained 24 DE-miRNAs (Data S5). Further analysis suggested that 18, 12, and 17 of the DE-miRNAs were associated with virus infections, lung diseases, and/or immune responses, respectively (Figure 2B). Our data also indicate that the age and gender of the study subjects had no significant influence on the expression level and function of the DE-miRNAs (Figure 2B). A total of 3,108 target genes were mapped to these DE-miRNAs (e.g., CD40 for hsa-miR-370-3p and CXCL8 for hsa-miR-302a-3p) through the target gene prediction analysis (Data S6). Moreover, 936 significant GO terms and 59 KEGG pathways were derived from these targeted genes (Data S7 and S8). Here, we displayed the top 20 GO terms (Figure 2C) and KEGG pathways (Figure 2D), which were associated with immune responses (T cell homeostasis and interleukin-2 production), virus infection (Hepatitis B), lung diseases (non-small cell lung cancer), and other processes (HIF-1 signaling pathway), respectively. Furthermore, to test the hypothesis that the DE-miRNAs could be used as biomarkers for diagnosis of SARS-CoV-2 infection, we built an XGBoost machine learning model based on miRNA expression data from the 61 HC and 116 patients with COVID-19, leading to prioritization of eleven important miRNA variables (Figure 2E). This model reached an area under curve (AUC) of 0.9258 and correctly classified over 83% of COV and 87% of HC subjects in the training set (Figure 2F). In the testing set, 24 of 29 COV (83%) and 12 of 15 HC (80%) were correctly classified by this model (Figure 2G). Finally, we identified four DE-miRNAs (i.e., has-miR-34a-5p, has-miR-370-3p, has-miR-193a-5p_R-1, and has-miR-12136-P3_1ss19TC) by combining the differential expression data with the XGBoost machine learning, which could be used as potential biomarkers to distinguish the SARS-CoV-2-infected patients (COV) from the uninfected healthy individuals (HC) (Figure 2H).

Technology Development Zone, Gaoxin 6th Road, Wuhan 430040, China

¹¹Wuhan Institute of Biotechnology, Wuhan East Lake High-tech Development Zone, 666 Gaoxin Road, Wuhan 430040, China

¹²The Renmin Hospital of Wuhan University, East Campus, East Lake New Technology Development Zone, Gaoxin 6th Road, Wuhan 430040, China

¹³Cancer Center, The First Affiliated Hospital of Xi'an Jiaotong University, 277 Yanta West Road, Xi'an 710061, China

¹⁴The Jackson Laboratory for Genomic Medicine, Farmington, CT 06032, USA

¹⁵These authors contributed equally

¹⁶Lead contact

*Correspondence: chenxi@spec-ally.com (X.C.), huke-rmhospital@163.com (K.H.), zbhong6288@163.com (B.Z.), shibingy@126.com (B.S.), cszhang99@126.com (C.Z.)

<https://doi.org/10.1016/j.isci.2022.104309>

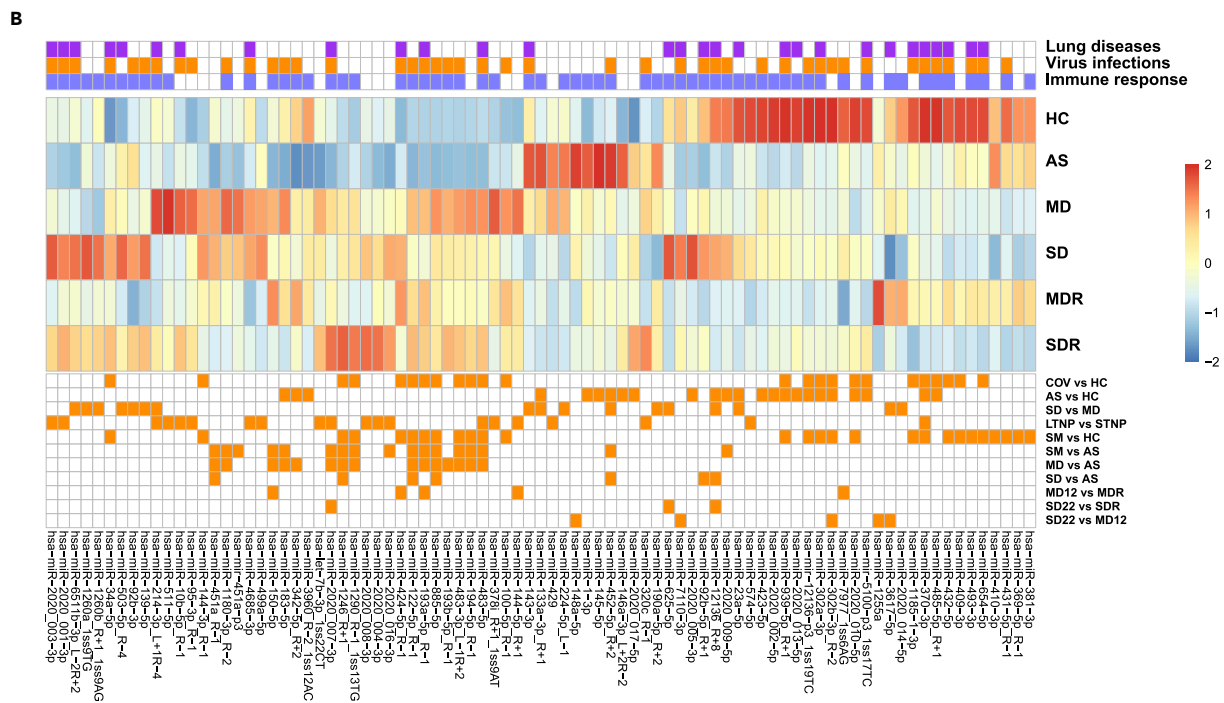
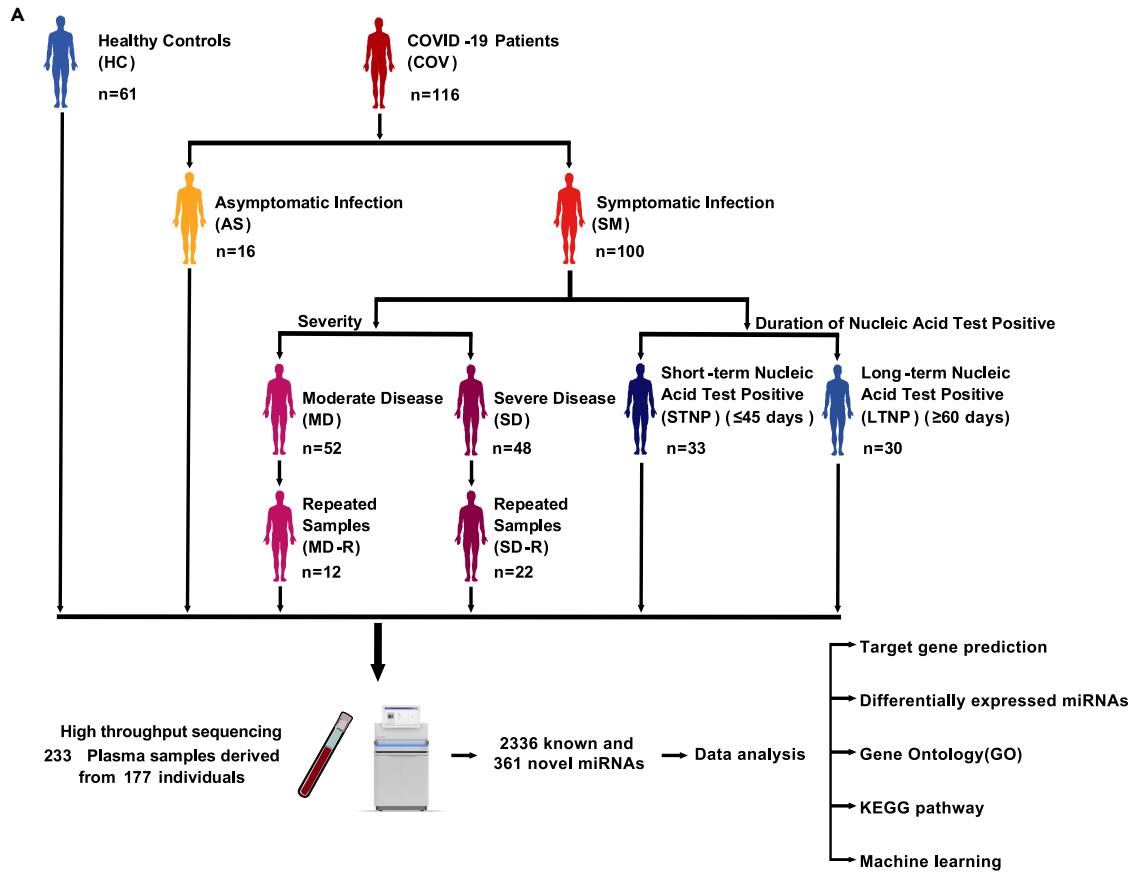


Figure 1. Study design and the differentially expressed miRNAs (DE-miRNAs)

(A), An overview of the study design. A total of 2,336 known and 361 novel miRNAs were identified in 233 plasma samples derived from 61 healthy controls (HC) and 116 SARS-CoV-2-infected patients (COV) by high throughput sequencing.
 (B), The heatmap of 85 DE-miRNAs from eleven compared groups: COV vs HC, AS vs HC, SM vs AS, SD vs MD, LTNP vs STNP, SM vs HC, MD vs AS, SD vs AS, SD22 vs MD12, MD12 vs MDR, and SD22 vs SDR. At the bottom of the figure, orange squares represent the DE-miRNAs between eleven compared groups. On the top, blue, orange, and purple squares represent the function of DE-miRNAs associated with immune responses, virus infections, or lung diseases, respectively.

DE-miRNAs associated with asymptomatic infection of SARS-CoV-2

Similar to the analysis described above, we compared the miRNA expression levels between the asymptomatic-infected individuals (AS) and the healthy control (HC), and identified 26 upregulated and 93 downregulated DE-miRNAs (Figure 3A). There are 24 DE-miRNAs left after removing the ones with the average expression level lower than 10 copies (Data S5). Further analysis suggested that 13, 8, and 20 of the DE-miRNAs were associated with virus infections, lung diseases, and/or immune responses, respectively (Figure 3B). Our data also indicate that the age and gender of the study subjects had no significant influence on the expression level and function of the DE-miRNAs (Figure 3B). A total of 3,184 target genes were mapped to these DE-miRNAs through the target gene prediction analysis (Data S6). Moreover, 952 significant GO terms and 71 KEGG pathways were derived from these targeted genes (Data S7 and S8). Here, we displayed the top 20 GO terms (Figure 3C) and KEGG pathways (Figure 3D), which were associated with immune responses, virus infection, lung diseases, and other processes, respectively. We also built an XGBoost machine learning model based on miRNA expression data from the 61 HC and 16 AS patients, leading to prioritization of nine important miRNA variables (Figure 3E). This model reached an area under curve (AUC) of 0.999 and correctly classified 100% of the AS and 96% of the HC subjects in the training set (Figure 3F). In the testing set, 3 of 4 COV (75%) and 13 of 1 H4C (93%) were correctly classified by this model (Figure 3G). Of note, the lower percentage of correct classification of the AS subjects in the testing set was most likely caused by the smaller sample size. Finally, we also identified four DE-miRNAs (has-miR-423-5p, has-miR-23a-5p, has-miR-146a-3p_2R-2, and has-miR-1-3p) by combining the differential expression data with the XGBoost machine learning, which could be used as potential biomarkers to distinguish the AS from the HC (Figure 3H).

DE-miRNAs associated with disease severity of COVID-19

When we compared the miRNA expression levels in the SD vs MD group using the same cutoff value described above (i.e. fold change >1.5 and $p \leq 0.05$ for the upregulated; fold change <0.67, and $p \leq 0.05$ for the downregulated DE-miRNAs), we only observed one significant DE-miRNA (has-miR-214-3p_L+1R_4). Therefore, we adjusted the fold change >1.2 for the upregulated whereas the fold change <0.83 for the downregulated miRNAs. As a result, we identified 47 upregulated and 37 downregulated DE-miRNAs (Figure 4A). There are 17 DE-miRNAs left after removing the ones with the average expression level lower than 10 copies (Data S5). Further analysis suggested that 7, 8, and 17 of the DE-miRNAs were associated with virus infections, lung diseases, and/or immune responses, respectively (Figure 4B). Our data also indicate that the age and gender of the study subjects had no significant influence on the expression level and function of the DE-miRNAs (Figure 4B). A total of 2,626 target genes were mapped to these DE-miRNAs through the target gene prediction analysis (Data S6). Moreover, 901 significant GO terms and 39 KEGG pathways were derived from these targeted genes (Data S7 and S8). Here, we displayed the top 20 GO terms (Figure 4C) and KEGG pathways (Figure 4D), which were associated with immune responses, virus infection, lung diseases, and other processes, respectively. We also built an XGBoost machine learning model based on miRNA expression data from the 48 SD and 52 MD patients, leading to prioritization of nine important miRNA variables (Figure 4E). This model reached an area under curve (AUC) of 0.996 and correctly classified 100% of the SD and 97% of the MD patients in the training set (Figure 4F). In the testing set, 10 of 12 SD (83%) and 9 of 13 MD (69%) were correctly classified by this model (Figure 4G). Of note, the lower percentage of correct classification of the SD and MD patients in the testing set was most likely caused by the smaller sample size. Finally, we identified five DE-miRNAs (i.e., has-miR-214-3p_L+1R_4, has-miR-143-3p, has-miR-224-5p_L-1, has-miR-452-3p_R+2, and hsa-miR-625-5p) by combining the differential expression data with the XGBoost machine learning, which could be used as potential biomarkers to distinguish the SD from the MD patients (Figure 4H). Furthermore, we examined the dynamic changes of the DE-miRNA expression levels in two of the patients with blood samples collected at multiple time points, including three samples collected from one fatal patient (COV012, COV031, COV042, and COV077) and three samples collected from a recovered patient (COV011, COV041, and COV078). The expression level of hsa-miR-122-5p_R-1 in SD was significantly

Table 1. DE-miRNAs associated with COVID-19 with various clinical presentations

miRNA	COV vs HC	AS vs HC	SM vs HC	SM vs AS	SD vs MD	SD22 vs MD12	MD12 vs MDR	SD22 vs SDR	LTNP vs STNP
hsa-miR-370-3p	↓ ^a	↓	↓						↓
hsa-miR-12136-p3_1ss19TC	↓	↓	↓						
hsa-miR-302b-3p_R-2	↓	↓	↓			↑			
hsa-miR-302a-3p	↓	↓	↓						
hsa-miR-939-5p_R+1	↓	↓	↓						
hsa-miR-2020_010-5p	↓	↓	↓						
hsa-miR-885-5p	↑ ^b		↑	↑					
hsa-miR-1246_R+1	↑		↑	↑					
hsa-miR-193a-5p_R-1	↑		↑	↑			↓		
hsa-miR-122-5p_R-1	↑		↑	↑					
hsa-miR-483-5p	↑		↑	↑					↑
hsa-miR-452-5p_R+2	↓ ^c	↑		↓					
hsa-miR-183-5p		↓		↑					
hsa-miR-2020_009-5p		↓		↑					
hsa-miR-342-5p_R+2		↓							
hsa-miR-423-5p		↓							
hsa-miR-2020_013-5p		↓							
hsa-miR-145-5p		↑							
hsa-miR-23a-5p		↓							
hsa-miR-12136_R+8		↓						↑	
hsa-miR-3960_R-2_1ss12AC		↓							
hsa-miR-146a-3p_L+2R-2		↑							
hsa-miR-1-3p		↑							
hsa-miR-2020_005-3p		↓							
hsa-miR-2020_002-5p		↓							
hsa-miR-2020_017-5p		↑							
hsa-miR-133a-3p_R+1		↑							
hsa-miR-92b-5p_R+1									
hsa-miR-625-5p								↑	
hsa-miR-7977_1ss6AG							↑		
hsa-miR-214-3p_L+1R-4					↓				↑
hsa-miR-452-5p_R+2					↓				
hsa-miR-23a-5p					↑				
hsa-miR-190a-5p_R+2					↓				
hsa-miR-95-3p_R-1									↑
hsa-miR-429									↑

COV, Coronavirus-infected cases; HC, Healthy Controls; AS, Asymptomatic Infection; SM, Symptomatic Infection; MD, moderate disease; SD, severe disease; MDR, moderate disease repeated samples; SDR, severe disease repeated samples; MD12, 12 moderate disease (single-person first blood samples); SD22, 22 severe disease (single-person first blood samples); LTNP, long-term nucleic acid test positive; STNP, short-term nucleic acid test positive.

^aThe expression of miRNA was upregulated or showed positive correlation.

^bThe expression of miRNA was downregulated or showed negative correlation.

^cNo significant difference.

higher than that in MD, suggesting that this miRNA could be positively correlated with the disease progression. Indeed, the level of hsa-miR-122-5p_R-1 in the fatal patient continuously increased as the disease worsened, whereas its expression restored to the normal level in the recovered patient (Figure 4I). On the other hand, the expression level of hsa-miR-224-5p_L-1 exhibited opposite changes in these

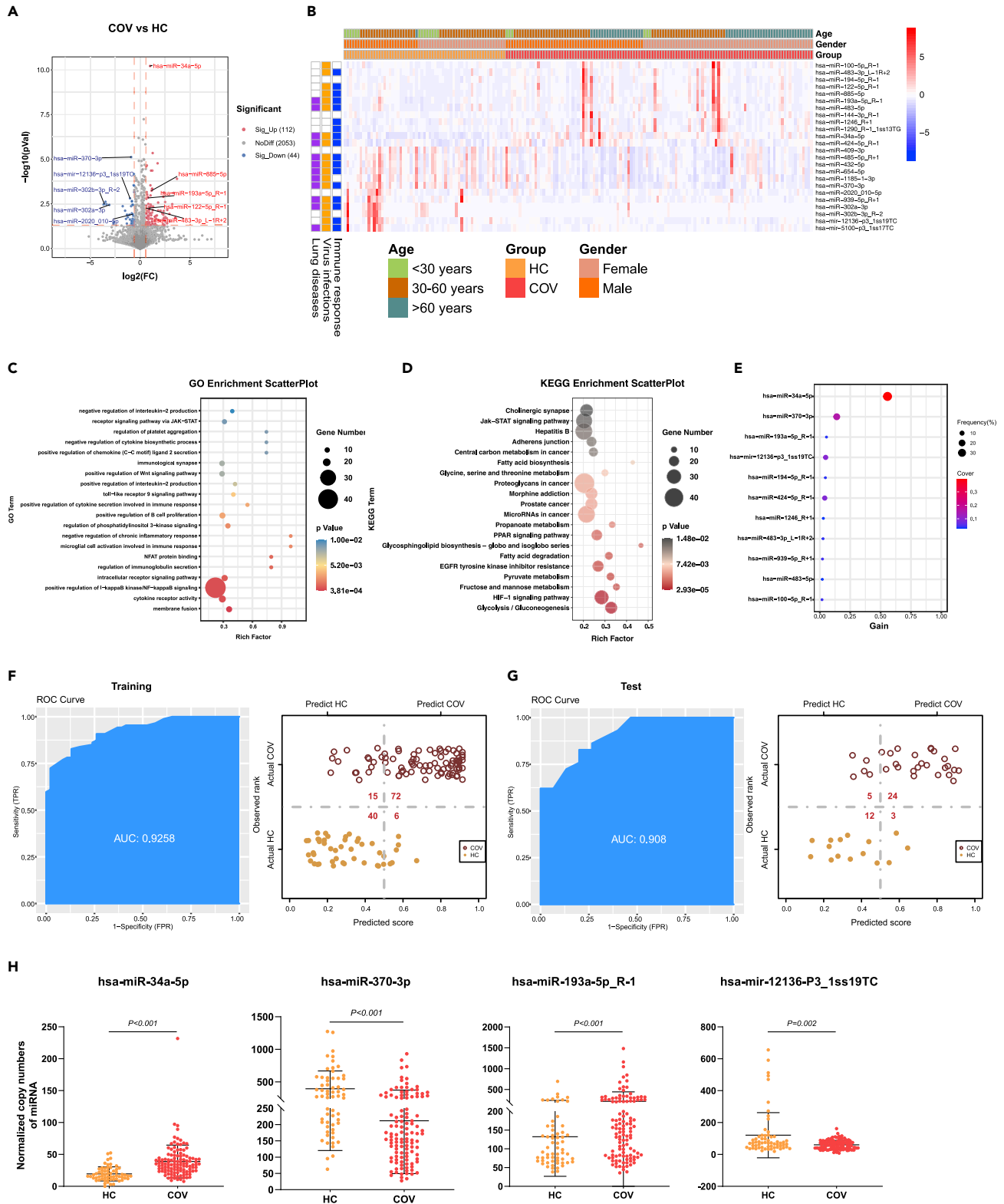


Figure 2. DE-miRNAs associated with SARS-COV-2 infections

(A), The volcano plot of COV vs HC. Red dots represent the upregulated DE-miRNAs (fold change > 1.5 , $p \leq 0.05$); blue dots represent the downregulated DE-miRNAs (fold change < 0.67 , $p \leq 0.05$); gray dots represent miRNAs without significant changes ($0.67 \leq \text{fold change} \leq 1.5$, $p > 0.05$).

Figure 2. Continued

- (B), The heatmap of 24 DE-miRNAs in COV vs HC. In the left panel, blue, orange, and purple squares represent the function of miRNAs associated with immune responses, virus infections, and lung diseases, respectively.
- (C), The GO analysis of the targeted genes in COV vs HC. Top 20 GO terms associated with immune responses, virus infections, and lung diseases were shown.
- (D), The KEGG analysis of the targeted genes in COV vs HC. Top 20 KEGG pathways associated with immune responses, virus infections, and lung diseases were shown.
- (E), The important DE-miRNAs prioritized by XGBoost analysis.
- (F), Receiver operating characteristic (ROC) and performance of the model in the training set.
- (G), ROC and performance of the model in the test set.
- (H), Normalized copy numbers of four representative DE-miRNAs in the COV vs HC group, the bars showing the mean value \pm standard deviation (sd) (COV n = 116, HC n = 61).

two patients (Figure 4J), suggesting that this miRNA could be negatively correlated with the disease progression. In this study, we also completed similar analysis on other compared groups (Figure 1B), including SM vs HC, SM vs AS, MD vs AS, SD vs AS, SD22 vs MD12, MD12 vs MD-R, and SD22 vs SD-R (Figures S1–S8 and Datas S5, S6, S7, and S8).

DE-miRNAs associated with the viral persistence of SARS-CoV-2 infection

We compared the miRNA alternations between the LTNP and STNP groups and identified 78 upregulated and 37 downregulated DE-miRNAs (Figure 5A). There are 20 DE-miRNAs left after removing the ones with the average expression level lower than 10 copies (Data S5). Further analysis suggested that 10, 8, and 11 of the DE-miRNAs were associated with virus infections, lung diseases, and/or immune responses, respectively (Figure 5B). Our data also indicate that the age and gender of the study subjects had no significant influence on the expression level and function of the DE-miRNAs (Figure 5B). A total of 1,848 target genes were mapped to these DE-miRNAs through the target gene prediction analysis (Data S6). Moreover, 594 significant GO terms and 44 KEGG pathways were derived from these targeted genes (Datas S7 and S8). Here, we displayed the top 20 GO terms (Figure 5C) and KEGG pathways (Figure 5D), which were associated with immune responses, virus infection, lung diseases, and other processes, respectively. We also built an XGBoost machine learning model based on miRNA expression data from the 30 LTNP and 33 STNP patients, leading to prioritization of 16 important miRNA variables (Figure 5E). This model reached an area under curve (AUC) of 0.977 and correctly classified 100% of the LTNP and 92% of the STNP patients in the training set (Figure 5F). In the testing set, 5 of 7 LTNP (71%) and 7 of 8 STNP (88%) were correctly classified by this model (Figure 5G). Of note, the lower percentage of correct classification of the LTNP and STNP patients in the testing set was most likely caused by the smaller sample size. Finally, we identified five DE-miRNAs (has-miR-429, has-miR-574-5p, has-miR-483-5p, has-miR-95-3p_R-1, and hsa-miR-378i_R+1_1ss9AT) by combining the differential expression data with the XGBoost machine learning, which could be used as potential biomarkers to distinguish the LTNP from the STNP patients (Figure 5H). Moreover, 9 of the 20 DE-miRNAs were positively or negatively correlated with the length (days) of the nucleic acid test positive for SARS-CoV-2 (Figure S8).

DE-miRNAs that may target the viral sequences of SARS-CoV-2 or the cellular genes associated with the viral life cycle

First, among the 2,336 known and 361 novel miRNAs, we detected 58 known and 7 novel miRNAs that may target at the viral sequences of SARS-CoV-2 genome (Data S9), including 4, 20, 16, and 61 miRNAs that target the membrane protein (M), nucleocapsid protein (N), spike protein (S), and open reading frame (ORF1ab), respectively (Figure 6A). In addition, by reviewing the published literature and the relevant databases including The Human Gene Database GeneCards (<https://www.genecards.org>), we have identified a panel of miRNAs that may target the cellular genes involved in the life cycle of SARS-CoV-2 replication including the entry and post-entry events, such as miRNA let-7b targeting ACE2 receptor, hsa-miR-1-3p targeting AXL receptor, hsa-miR-122-5p targeting ADAM17, hsa-miR-485-5p targeting DC-SIGN, hsa-miR-424-5p targeting furin, hsa-miR-1185-1-3p targeting HAT(HAT1), hsa-miR-183-5p targeting integrin (ITGB1), hsa-miR-1-3p targeting Neuropilin-1 (NRP1), hsa-miR-485-5p targeting TMPRSS4, hsa-miR-34a-5p targeting AR, hsa-miR-423-5p targeting Cyclophilin A, hsa-miR-377-5p targeting Cathepsin B, hsa-miR-148a-5p targeting TMEM106B, hsa-miR-193a-5p_R-1 targeting TOMM70 (TOM70), and hsa-miR-1-3p and hsa-miR-7110-3p targeting VMP1 (Figure 6B). Our findings indicate that the cellular miRNAs may directly and/or indirectly act on the life cycle of SARS-CoV-2 and involve in the pathogenesis of COVID-19.

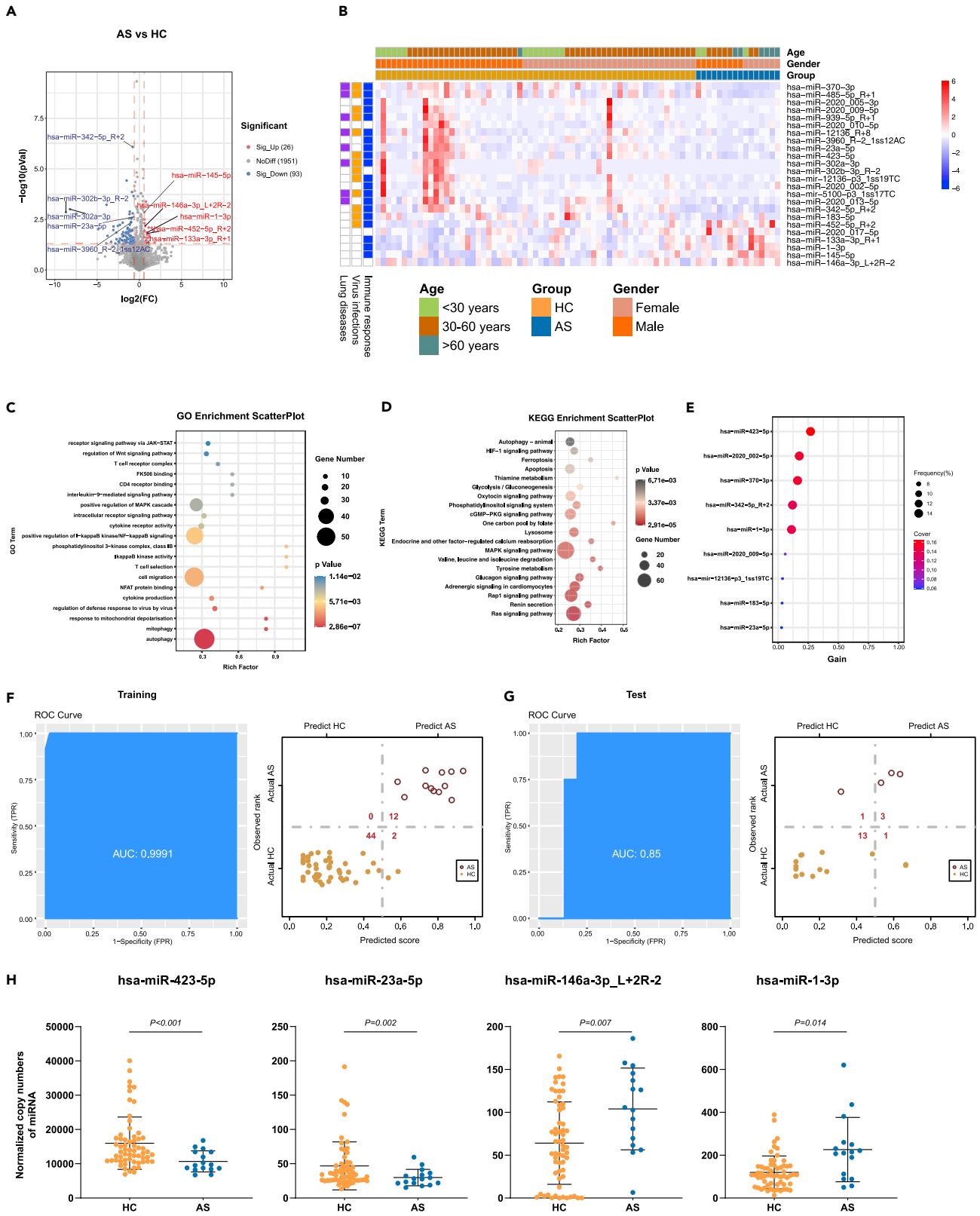


Figure 3. DE-miRNAs associated with the asymptomatic infection of SARS-CoV-2

(A), The volcano plot of AS vs HC. Red dots represent the upregulated DE-miRNAs (fold change > 1.5, $p \leq 0.05$); blue dots represent the downregulated DE-miRNAs (fold change < 0.67, $p \leq 0.05$); gray dots represent miRNAs without significant changes ($0.67 \leq \text{fold change} \leq 1.5$, $p > 0.05$).
(B), The heatmap of 24 DE-miRNAs in AS vs HC. In the left panel, blue, orange, and purple squares represent the function of miRNAs associated with immune responses, virus infections, and lung diseases, respectively.
(C), The GO analysis of the targeted genes in AS vs HC. Top 20 GO terms associated with immune responses, virus infections, and lung diseases were shown.
(D), The KEGG analysis of the targeted genes in AS vs HC. Top 20 KEGG pathways associated with immune responses, virus infections, and lung diseases were shown.
(E), The important DE-miRNAs prioritized by XGBoost analysis.
(F), Receiver operating characteristic (ROC) and performance of the XGBoost model in the training set.
(G), ROC and performance of the XGBoost model in the test set.
(H), Normalized copy numbers of four representative DE-miRNAs in the AS vs HC group, the bars showing the mean \pm sd (AS n = 16, HC n = 61).

DISCUSSION

In this study, we identified a large number of DE-miRNAs that were associated with the virus infection, clinical symptoms, disease severity, and the viral persistence of the patients with COVID-19 as well as a panel of miRNAs that may target the viral sequences of SARS-CoV-2 or the cellular genes associated with the viral life cycle (Figures 1–7; Table 1). To our knowledge, this is the first large-scale study of miRNA profiles by high-throughput sequencing on the plasma samples from the AS subjects in parallel with the HC individuals and the SM patients with various clinical presentations. We identified six DE-miRNAs that could be used as biomarkers to distinguish the SARS-CoV-2-infected individuals from the healthy controls, and four of the six biomarkers were suggested to be associated with virus infection, inflammation, immune responses, and lung diseases (Pulati et al., 2019; Ramanathan et al., 2019; Peng et al., 2018). In particular, hsa-miR-302b-3p_R-2 and hsa-miR-302a-3p were differentially expressed in COV vs HC, AS vs HC, and SM vs HC and significantly downregulated in the SARS-CoV-2-infected patients, suggesting that hsa-miR-302 may be used as a potential biomarker to distinguish patients with COVID-19 from normal controls. The hsa-miR-302 family members have been demonstrated to play vital roles in cell proliferation and the regulation of tumor growth (Maadi et al., 2016; Subramanyam et al., 2011; Lin et al., 2011). Previous studies have shown that influenza A virus (IAV) infection may cause downregulation of miRNA-302a and miR-302c, which may suppress IAV-induced cytokine storm and prevent the translation of nuclear factor-kappa B (NF- κ B) from the cytosol to the nucleus, respectively (Gui et al., 2015). Both hsa-miR-302a and hsa-miR-302b may target on the C-X-C motif chemokine ligand 8 (CXCL8) gene, which was associated with the patients with acute respiratory distress syndrome (ARDS) by promoting neutrophil migration to lung interstitium and alveolar space (Williams et al., 2017). CXCL8 was also suggested to play a vital role in the inflammatory diseases of the lungs, such as asthma and chronic obstructive pulmonary disease (COPD) (Huang et al., 2017). These findings suggested that miR-302 may have a protective effect in SARS-CoV-2-infected patients through targeting at the CXCL8-linked pathways.

The hsa-miR-146a-3p was significantly upregulated in AS subjects compared with HC. Previous studies suggested that hsa-miR-146a-3p was a key factor in the regulation of innate immunity, viral infection, inflammation, and tumor development (Testa et al., 2017; Fei et al., 2020). Fei et al. found that miR-146a suppresses the NF- κ B pathway by downregulating TLR3 and TRAF6 to alleviate the inflammatory response during coxsackievirus B infection (Fei et al., 2020). The host's innate immune system provides the first line of defense against viral invasions so that the body can effectively fight the virus infection. It is conceivable that through the regulatory loop, the inflammation can be controlled to a level adequate to eliminate the invading viruses without causing significant damage to the host. Our results indicated that hsa-miR-146a-3p may play a protective role in the patients with asymptomatic infection and serve as a biomarker for diagnosis and therapeutic approach for AS.

We also identified a number of DE-miRNAs that may help differentiate the SM from AS. In particular, hsa-miR-122-5p, hsa-miR-1246, and hsa-miR-885-5p were suggested to be involved in the virus infection, immune responses, and inflammation (Suo et al., 2018; Lu et al., 2020; Janssen et al., 2013; Hussein et al., 2017). For instance, the stability and propagation of hepatitis C virus (HCV) was dependent on a functional interaction between the HCV genome and liver-expressed miR-122 (Hussein et al., 2017). The hsa-miR-122 could promote HCV RNA accumulation, and serum miR-122-5p was significantly higher in patients in different stages of HBV infection than in controls (Lak et al., 2020). HCV infection may lead to dyslipidaemia, and miR-122 can regulate hepatic lipid metabolism and inflammation (Lee et al., 2015). Targeting miR-122 may be an effective treatment strategy for HCV infection (van der Ree et al., 2014). In addition, miR-122-5p may protect acute lung injury via regulation of DUSP4/ERK signaling in pulmonary microvascular

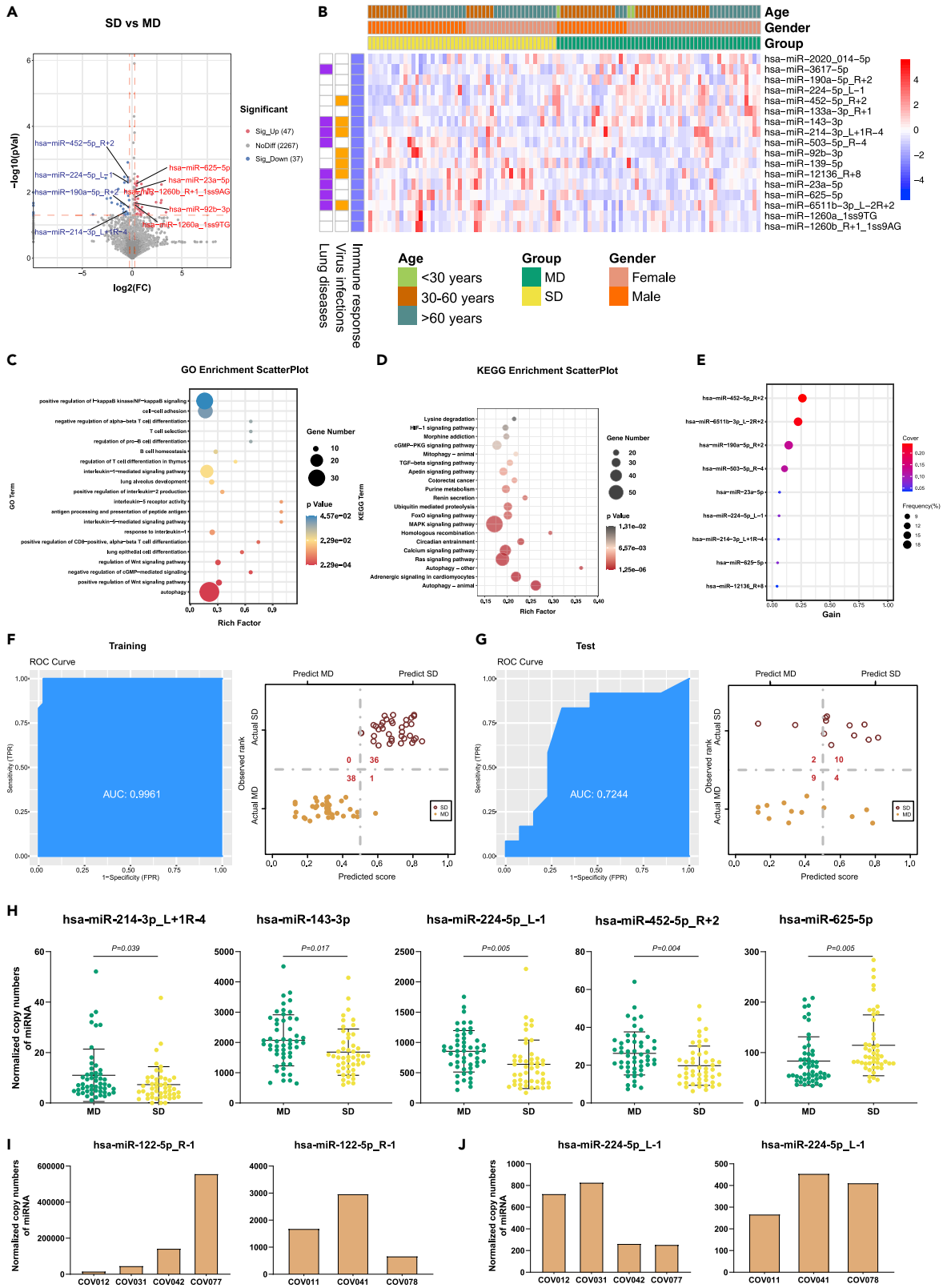


Figure 4. DE-miRNAs associated with the disease severity of COVID-19, see also Figures S1–S8

- (A), The volcano plot of SD vs MD. Red dots represent the upregulated DE-miRNAs (fold change > 1.2, $p \leq 0.05$); blue dots represent the downregulated DE-miRNAs (fold change < 0.83, $p \leq 0.05$); gray dots represent miRNAs without significant changes ($0.83 \leq \text{fold change} \leq 1.2$, $p > 0.05$).
- (B), The heatmap of 17 DE-miRNAs in SD vs MD. In the left panel, blue, orange, and purple squares represent the function of miRNAs associated with immune responses, virus infections, and lung diseases, respectively.
- (C), The GO analysis of the targeted genes in SD vs MD. Top 20 GO terms associated with immune responses, virus infections, and lung diseases were shown.
- (D), The KEGG analysis results of targeted genes in SD vs MD. Top 20 KEGG pathways associated with immune responses, virus infections, and lung diseases were shown.
- (E), The important miRNAs prioritized by XGBoost analysis.
- (F), Receiver operating characteristic (ROC) and performance of the XGBoost model in the training set.
- (G), ROC and performance of the XGBoost model in the test set.
- (H), Normalized copy numbers of five representative DE-miRNAs in the SD vs MD group, the bars showing the mean \pm SD (SD n = 48, MD n = 52).
- (I), The dynamic change of the copy numbers of hsa-miR-122-5p_R-1 in a fatal patient (COV012, COV031, COV042, and COV077) and a recovered patient (COV011, COV041, and COV078).
- (J), The dynamic change of the copy numbers of hsa-miR-224-5p_L-1 in a fatal patient (COV012, COV031, COV042, and COV077) and a recovered patient (COV011, COV041, and COV078).

endothelial cells and may help differentiate moderate from severe patients with COVID-19 (Lu et al., 2020; Fujita et al., 2021). Nakao et al. reported that miR-122-5p may target ADAM metalloproteinase domain 17 (ADAM17) and become a promising strategy for hepatocellular carcinoma treatment (Nakao et al., 2014). GO analysis showed that ADAM17 enriched in T cell differentiation in thymus. There are robust T cell responses against SARS-CoV-2 in patients with COVID-19 (Neidleman et al., 2020), suggesting that miR-122-5p may be involved in the immune responses in COVID-19. We observed that hsa-miR-214-3p_L+1R-4 was significantly downregulated in SD compared with the MD patients, and was associated with the disease severity. The hsa-miR-214-3p has been suggested to be associated with various cancers, inflammatory diseases, and the related signaling pathways (Liu et al., 2019; Yang et al., 2019; Yan et al., 2020; Li and Wang, 2017).

Another important aspect of this study is the identification of 20 DE-miRNAs in the LTNP vs STNP groups, nine of which were significantly correlated with the length of the viral persistence in patients with COVID-19, including hsa-miR-483-5p and hsa-miR-429. The expression level of hsa-miR-483-5p was significantly higher in the LTNP subjects, indicating that hsa-miR-483-5p was positively correlated with the disease course (Figure 4E). The hsa-miR-483-5p was suggested to be associated with influenza virus infection, acute lung injury, chronic pulmonary obstruction, and lung cancer (Hassan et al., 2019; Shen et al., 2017; Wang et al., 2018). The expression of hsa-miR-483-5p was up-regulated in lung injury tissues and the increase of hsa-miR-483-5p may protect acute lung injury (Leng et al., 2020). In addition, hsa-miR-429 was previously reported to be antiviral in respiratory diseases (Sardar et al., 2020). Kozak et al. have demonstrated that miR-429 may directly target SESN3 (Sestrin 3) (Kozak et al., 2019). Functional analysis showed that SESN3 enriched in the p53 signaling pathway and could be regulated by p53 (Budanov et al., 2010). SARS-CoV-2 could induce p53 signaling pathway in lymphocytes (Xiong et al., 2020). Therefore, hsa-miR-429 may target SESN3 through p53 signaling pathway and correlated with the viral persistence. Moreover, hsa-let-7b-3p_1ss22CT was up-regulated in LTNP compared with STNP. ACE2 is a major cellular receptor for SARS-CoV-2 and Let-7b may downregulate ACE2 expression (Zhang et al., 2019, 2021), suggesting that the upregulation of let-7b in the LTNP patients may lead to the downregulation of ACE2 and subsequently prevent the hosts from virus infection.

In this study, we have identified 63 miRNAs that may directly target the viral sequences of SARS-CoV-2 genome and a panel of miRNAs that may target the cellular genes involved in the life cycle of SARS-CoV-2 replication. The exact roles of these miRNA in viral infection, persistence, and pathogenesis of COVID-19 remain unclear at this time, and they warrant additional studies. In addition, we have successfully established and tested a machine learning model for differentiation of SARS-CoV-2-infected patients from uninfected individuals and classification of patients with COVID-19 based on the differentially expressed miRNAs. Currently, detection of viral nucleic acid by RT-PCR is the gold standard method for clinical diagnosis of SARS-CoV-2 infection. Because multiple factors may affect the sensitivity and specificity of RT-PCR assays, our machine learning model may become a novel complementary method to RT-PCR for clinical diagnosis of COVID-19.

In conclusion, we identified a large number of DE-miRNAs that were associated with the virus infection, clinical symptoms, disease severity, and the viral persistence of the patients with COVID-19, as well as a

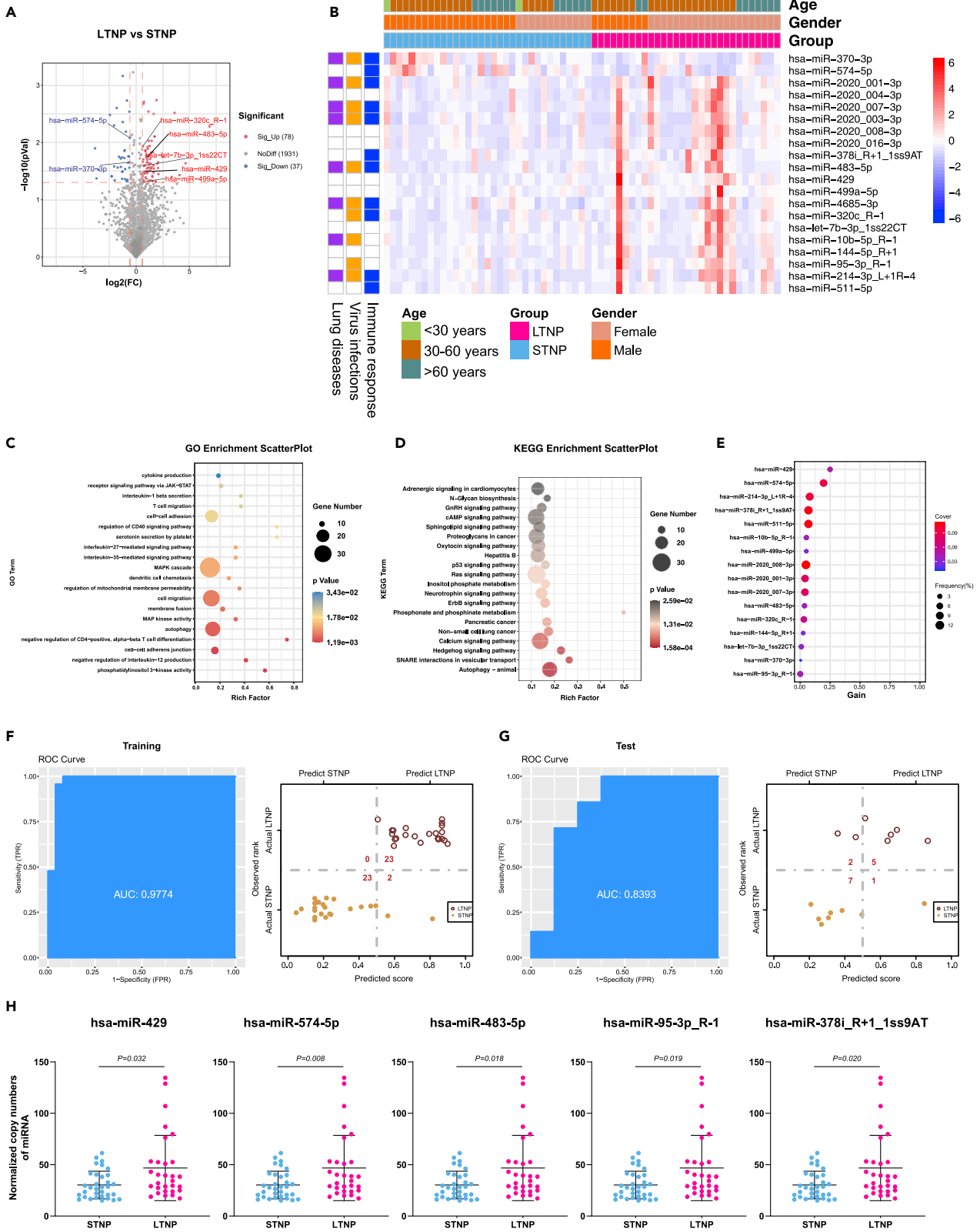


Figure 5. DE-miRNAs associated with the viral persistence of SARS-CoV-2 infection, see also Figure S8

(A), The volcano plot of LTNP vs STNP. Red dots represent the upregulated DE-miRNAs (fold change > 1.5, $p \leq 0.05$); blue dots represent the downregulated DE-miRNAs (fold change < 0.67, $p \leq 0.05$); gray dots represent miRNAs without significant changes ($0.67 \leq \text{fold change} \leq 1.5$, $p > 0.05$).

(B), The heatmap of 20 DE-miRNAs in LTNP vs STNP. In the left panel, blue, orange, and purple squares represent the function of miRNAs associated with immune responses, virus infections, and lung diseases, respectively.

(C), The GO analysis of the targeted genes in LTNP vs STNP. Top 20 GO terms associated with immune responses, virus infections, and lung diseases were shown.

(D), The KEGG analysis of the targeted genes in LTNP vs STNP. Top 20 KEGG pathways associated with immune responses, virus infections, and lung diseases were shown.

(E), The important DE-miRNAs prioritized by XGBoost analysis.

(F), Receiver operating characteristic (ROC) and performance of the XGBoost model in the training set.

(G), ROC and performance of the XGBoost model in the test set.

(H), Normalized copy numbers of five representative DE-miRNAs in the LTNP vs STNP group, the bars showing the mean \pm SD (LTNP $n = 30$, STNP $n = 33$).

panel of miRNAs that may target the host genes involved in a variety of pathways or the viral genomic regions of SARS-CoV-2. Our findings may help us understand the potential contributions of microRNAs to the pathogenesis of COVID-19 and identify novel biomarkers for the diagnosis and treatment of SARS-CoV-2 infection.

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 SUBJECT DETAILS**
 - Ethics statement
 - Case definition and study cohort
- **METHOD DETAILS**
 - Blood sample collection and plasma preparation
 - RNA extraction, library construction and high throughput sequencing
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
 - The miRNA-seq data analysis
 - Identification of differentially expressed miRNAs
 - Target genes prediction
 - The comparison of sequences between the SARS-CoV-2 genome and miRNAs
 - Gene ontology (GO) and KEGG pathway enrichment analysis
 - Machine learning
 - Statistical analysis

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2022.104309>.

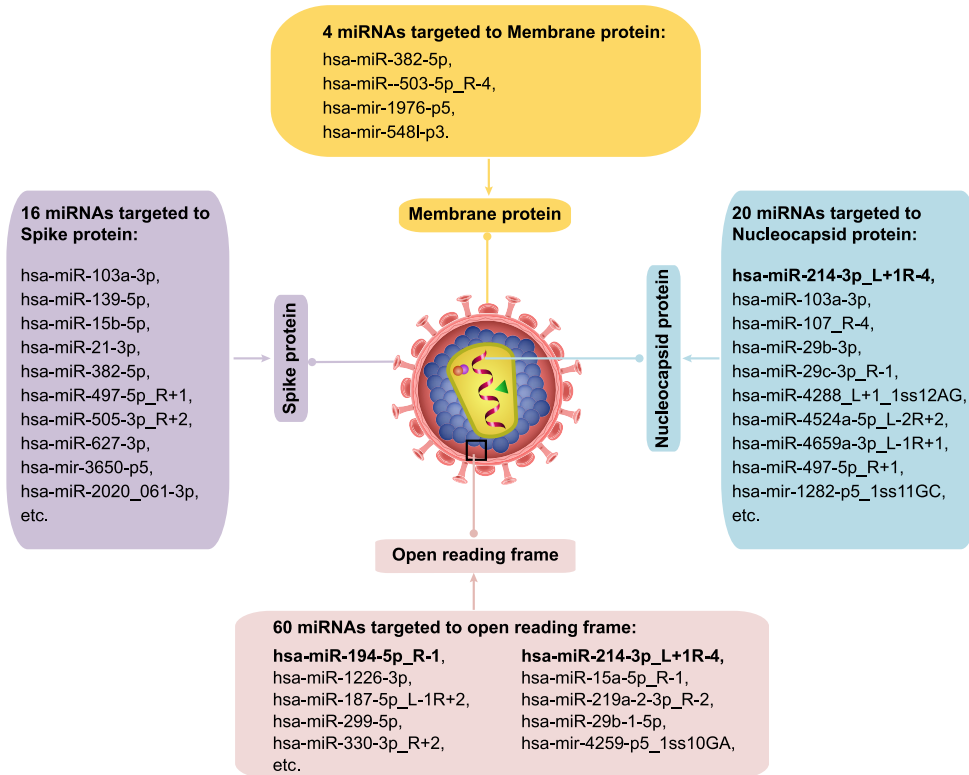
ACKNOWLEDGMENTS

We are extremely grateful to all the patients and their families for participation of this study and providing the valuable information. We also thank our colleagues at The First Affiliated Hospital of Xi'an Jiaotong University and The Renmin Hospital of Wuhan University for their kind help and strong support throughout the course of this study. This work was supported by the Department of Science and Technology of Shaanxi Province [Grant No. 2020ZDXM2-SF-02] and the operational funds from The First Affiliated Hospital of Xi'an Jiaotong University.

AUTHOR CONTRIBUTIONS

Conceptualization: CZ and BS; Methodology: CZ, QZ, XQ, JM, FH, ML, XW, HQ, SH, YY, YL, HB, MJ, DR, YT, LF, YZ, XC, QS, ZS, JM, JW, DY, JH, YW, JH, XD, and JJ; Resources: CZ, BS, BZ, KH, and DL; Writing: CZ, QZ, XQ, JM, YL, XC, BZ, and BS; Review and editing: CZ, BS; Supervision: CZ, BS, BZ, KH, and XC; Funding: CZ and BS; Project administration: CZ and BS.

A



B

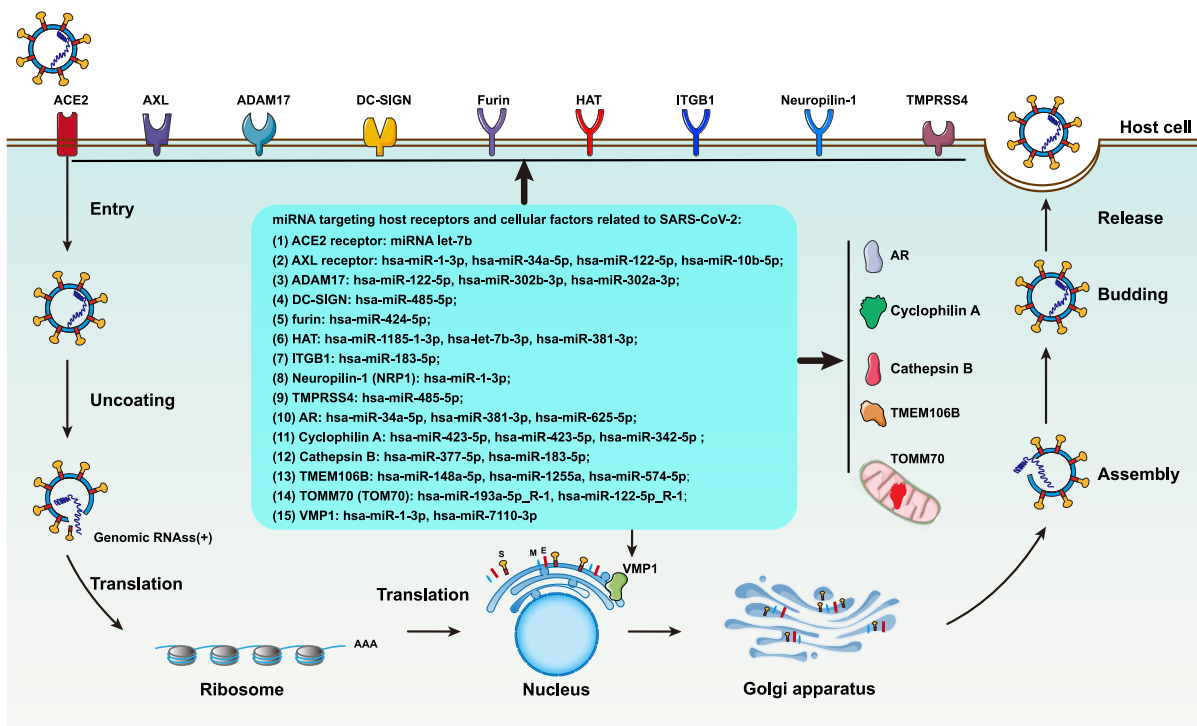


Figure 6. The miRNAs targeting the SARS-CoV-2 viral genome and the cellular genes associated with the viral life cycle

(A), The miRNAs targeting the SARS-CoV-2 viral genome. There are 65 miRNAs that may target the SARS-CoV-2 viral genome, including membrane protein, spike protein, nucleocapsid protein, and open reading frame.

(B), The miRNAs targeting the cellular genes associated with the life cycle of SARS-CoV-2.

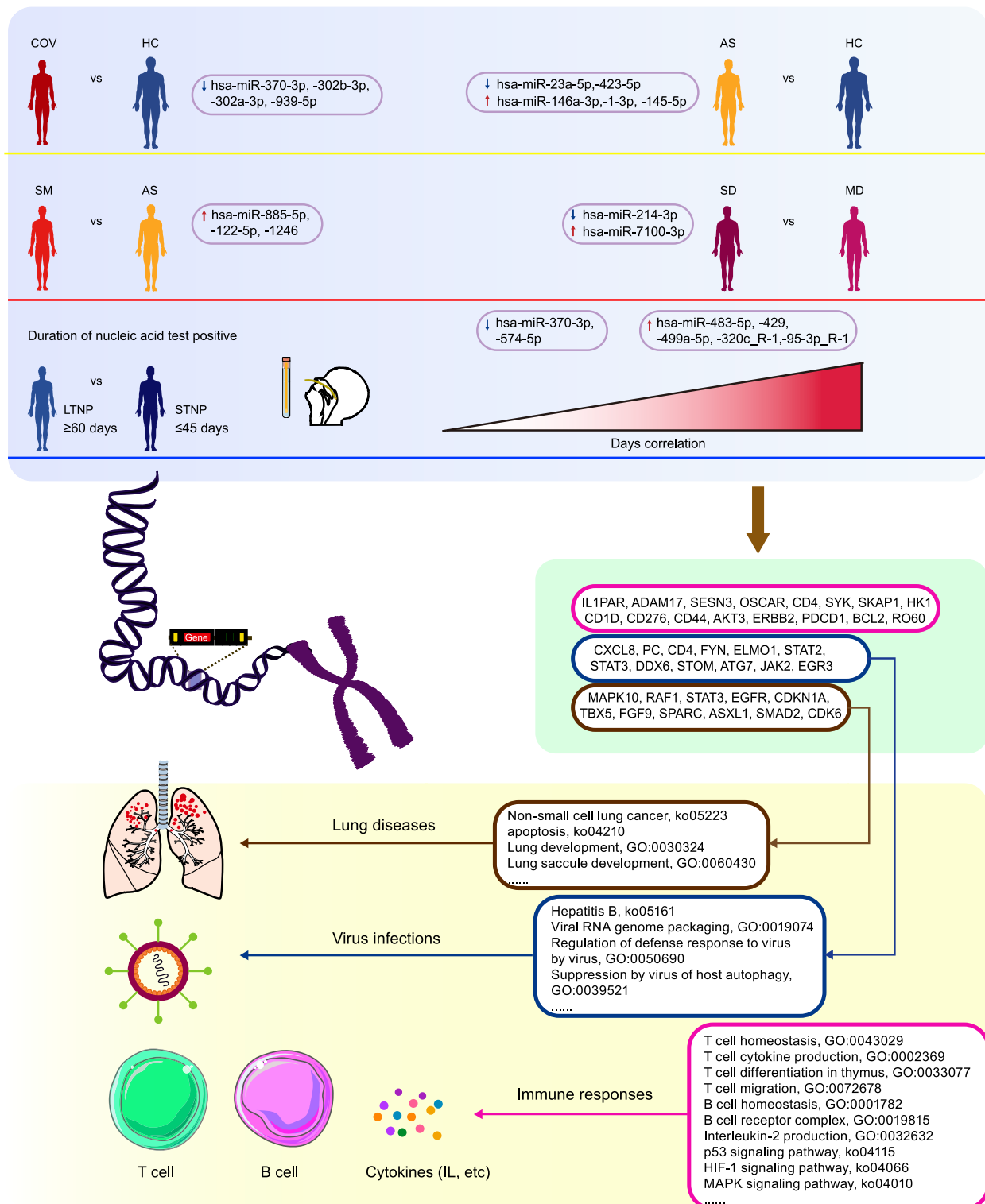


Figure 7. Potential contributions of the representative DE-miRNAs to the pathogenesis of COVID-19

Representative miRNAs, which were associated with immune responses, virus infections, and lung diseases, were selected from the five compared groups (COV vs HC, AS vs HC, SM vs AS, SD vs MD, and LTNP vs STNP) to help understand the pathogenesis of COVID-19. COV, SARS-CoV-2-infected individuals; HC, Healthy controls; AS, Asymptomatic infection; SM, Symptomatic infection; MD, Moderate disease; SD, Severe disease; LTNP, Long-term nucleic acid test positive; STNP, Short-term nucleic acid test positive.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

We worked to ensure gender balance in the recruitment of human subjects. We worked to ensure ethnic or other types of diversity in the recruitment of human subjects. The author list of this paper includes contributors from the location where the research was conducted who participated in the data collection, design, analysis, and/or interpretation of the work.

Received: August 12, 2021

Revised: January 4, 2022

Accepted: April 22, 2022

Published: May 20, 2022

REFERENCES

- Agarwal, V., Bell, G.W., Nam, J.W., and Bartel, D.P. (2015). Predicting effective microRNA target sites in mammalian mRNAs. *eLife* 4, e05005. <https://doi.org/10.7554/eLife.05005>.
- Arisan, E.D., Dart, A., Grant, G.H., Arisan, S., Cuhadaroglu, S., Lange, S., and Uysal-Onganer, P. (2020). The prediction of miRNAs in SARS-CoV-2 genomes: hsa-miR databases identify 7 key miRNAs linked to host responses and virus pathogenicity-related KEGG pathways significant for comorbidities. *Viruses* 12, 614. <https://doi.org/10.3390/v12060614>.
- Bartel, D.P. (2004). MicroRNAs. *Cell* 116, 281–297. [https://doi.org/10.1016/s0092-8674\(04\)00045-5](https://doi.org/10.1016/s0092-8674(04)00045-5).
- Budanov, A.V., Lee, J.H., and Karin, M. (2010). Stressin' Sestrins take an aging fight. *EMBO Mol. Med.* 2, 388–400. <https://doi.org/10.1002/emmm.201000097>.
- Buitrago-Garcia, D., Egli-Gany, D., Counotte, M.J., Hossmann, S., Imeri, H., Ipekci, A.M., Salanti, G., and Low, N. (2020). Occurrence and transmission potential of asymptomatic and presymptomatic SARS-CoV-2 infections: a living systematic review and meta-analysis. *PLoS Med.* 17, e1003346. <https://doi.org/10.1371/journal.pmed.1003346>.
- Chauhan, N., Jaggi, M., Chauhan, S.C., and Yallapu, M.M. (2021). COVID-19: fighting the invisible enemy with microRNAs. *Expert Rev. Anti Infect Ther.* 19, 137–145. <https://doi.org/10.1080/14787210.2020.1812385>.
- Chen, T.Q., and Guestrin, C. (2016). XGBoost: a scalable tree boosting system. In *KDD '16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794. <https://doi.org/10.1145/2939672.2939785>.
- Chen, X.M., Zhao, Y., Wu, X.D., Wang, M.J., Yu, H., Lu, J.J., Hu, Y.J., Huang, Q.C., Huang, R.Y., and Lu, C.J. (2019). Novel findings from determination of common expressed plasma exosomal microRNAs in patients with psoriatic arthritis, psoriasis vulgaris, rheumatoid arthritis, and gouty arthritis. *Discov. Med.* 28, 47–68.
- Enright, A.J., John, B., Gaul, U., Tuschl, T., Sander, C., and Marks, D.S. (2003). MicroRNA targets in *Drosophila*. *Genome Biol.* 5, R1. <https://doi.org/10.1186/gb-2003-5-1-r1>.
- Fei, Y., Chaulagain, A., Wang, T., Chen, Y., Liu, J., Yi, M., Wang, Y., Huang, Y., Lin, L., Chen, S., et al. (2020). MiR-146a down-regulates inflammatory response by targeting TLR3 and TRAF6 in Cocksackievirus B infection. *RNA* 26, 91–100. <https://doi.org/10.1261/rna.071985.119>.
- Fujita, Y., Hoshina, T., Matsuzaki, J., Yoshioka, Y., Kadota, T., Hosaka, Y., Fujimoto, S., Kawamoto, H., Watanabe, N., Sawaki, K., et al. (2021). Early prediction of COVID-19 severity using extracellular vesicle COPB2. *J. Extracell. Vesicles* 10, e12092. <https://doi.org/10.1002/jev2.12092>.
- Government of China. (2021). New Coronavirus Pneumonia Prevention and Control Program, 8nd ed.. <http://www.nhc.gov.cn/jkj/s3577/202105/6f1e8ec6c4a540d99fafef52fc86d0f8/files/4a860a7e85d14d55a22fbab0bbe77cd9.pdf>.
- Guan, W.J., Ni, Z.Y., Hu, Y., Liang, W.H., Ou, C.Q., He, J.X., Liu, L., Shan, H., Lei, C.L., Hui, D.S.C., et al. (2020). Clinical characteristics of coronavirus disease 2019 in China. *N. Engl. J. Med.* 382, 1708–1720. <https://doi.org/10.1056/NEJMoa2002032>.
- Gui, S., Chen, X., Zhang, M., Zhao, F., Wan, Y., Wang, L., Xu, G., Zhou, L., Yue, X., Zhu, Y., and Liu, S. (2015). Mir-302c mediates influenza A virus-induced IFN β expression by targeting NF- κ B inducing kinase. *FEBS. Lett.* 589, 4112–4118. <https://doi.org/10.1016/j.febslet.2015.11.011>.
- Hassan, A.S., Elgendy, N.A., Tawfik, N.A., and Elnasser, A.M. (2019). Serum miR-483-5p and miR-133a as biomarkers for diagnosis of hepatocellular carcinoma post-hepatitis C infection in Egyptian patients. *Egypt. J. Immunol.* 26, 31–40.
- Hosseini Rad Sm, A., and McLellan, A.D. (2020). Implications of SARS-CoV-2 mutations for genomic RNA structure and host microRNA targeting. *Int. J. Mol. Sci.* 21, 4807. <https://doi.org/10.3390/ijms21134807>.
- Huang, Z.W., Lien, G.S., Lin, C.H., Jiang, C.P., and Chen, B.C. (2017). p300 and C/EBP β -regulated IKK β expression are involved in thrombin-induced IL-8/CXCL8 expression in human lung epithelial cells. *Pharm. Res.* 121, 33–41. <https://doi.org/10.1016/j.phrs.2017.04.020>.
- Hussein, N.A.E.M., Kholi, Z.A.E., Anwar, M.M., Ahmad, M.A., and Ahmad, S.M. (2017). Plasma miR-22-3p, miR-642b-3p and miR-885-5p as diagnostic biomarkers for pancreatic cancer. *J. Cancer Res. Clin. Oncol.* 143, 83–93. <https://doi.org/10.1007/s00432-016-2248-7>.
- Janssen, H.L., Reesink, H.W., Lawitz, E.J., Zeuzem, S., Rodriguez-Torres, M., Patel, K., van der Meer, A.J., Patock, A.K., Chen, A., Zhou, Y., et al. (2013). Treatment of HCV infection by targeting microRNA. *N. Engl. J. Med.* 368, 1685–1694. <https://doi.org/10.1056/NEJMoa1209026>.
- Kozak, J., Wdowiak, P., Maciejewski, R., and Torres, A. (2019). Interactions between microRNA-200 family and Sestrin proteins in endometrial cancer cell lines and their significance to anoikis. *Mol. Cell. Biochem.* 459, 21–34. <https://doi.org/10.1007/s11010-019-03547-2>.
- Kozomara, A., Birgaoanu, M., and Griffiths-Jones, S. (2019). miRBase: from microRNA sequences to function. *Nucleic Acids Res.* 47, D155–D162. <https://doi.org/10.1093/nar/gky1141>.
- Kupferschmidt, K., and Wadman, M. (2021). Delta variant triggers new phase in the pandemic. *Science* 372, 1375–1376. <https://doi.org/10.1126/science.372.6549.1375>.
- Lak, R., Yaghobi, R., and Garshasbi, M. (2020). Importance of miR-125a-5p and miR-122-5p expression in patients with HBV infection. *Cell Mol. Biol.* 66, 1–8. <https://doi.org/10.14715/cmb/2020.66.5.1>.
- Lee, C.H., Kim, J.H., and Lee, S.W. (2017). The role of MicroRNA in pathogenesis and as markers of HCV chronic infection. *Curr. Drug Targets* 18, 756–765. <https://doi.org/10.2174/1389450117666160401125213>.
- Lee, T.H., Matta, B., King, B.D., Hodges, M.R., Tillmann, H.L., and Patel, K. (2015). MicroRNA-122 associates with serum apolipoprotein B but not liver fibrosis markers in CHC genotype 1 infection. *J. Med. Virol.* 87, 1722–1726. <https://doi.org/10.1002/jmv.24230>.
- Leng, C., Sun, J., Xin, K., Ge, J., Liu, P., and Feng, X. (2020). High expression of miR-483-5p aggravates sepsis-induced acute lung injury. *J. Toxicol. Sci.* 45, 77–86. <https://doi.org/10.2131/jts.45.77>.
- Li, J.A., and Wang, Y.D. (2017). Downregulation of miR-214-3p May Contribute to Pathogenesis of Ulcerative Colitis via Targeting STAT6. *Biomed.*

- Res. Int. 2017, 8524972. <https://doi.org/10.1155/2017/8524972>.
- Lin, S.L., Chang, D.C., Lin, C.H., Ying, S.Y., Leu, D., and Wu, D.T.S. (2011). Regulation of somatic cell reprogramming through inducible mir-302 expression. *Nucleic Acids Res.* 39, 1054–1065. <https://doi.org/10.1093/nar/gkq850>.
- Liu, C., Shang, Z., Ma, Y., Ma, J., and Song, J. (2019). HOTAIR/miR-214-3p/FLOT1 axis plays an essential role in the proliferation, migration, and invasion of hepatocellular carcinoma. *Int. J. Clin. Exp. Pathol.* 12, 50–63.
- Lu, Z., Feng, H., Shen, X., He, R., Meng, H., Lin, W., and Geng, Q. (2020). MiR-122-5p protects against acute lung injury via regulation of DUSP4/ERK signaling in pulmonary microvascular endothelial cells. *Life Sci.* 256, 117851. <https://doi.org/10.1016/j.lfs.2020.117851>.
- Maadi, H., Moshtaghan, A., Taha, M.F., Mowla, S.J., Kazerooni, A., Haass, N.K., and Javeri, A. (2016). Multimodal tumor suppression by miR-302 cluster in melanoma and colon cancer. *Int. J. Biochem. Cel. Biol.* 81, 121–132. <https://doi.org/10.1016/j.biocel.2016.11.004>.
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *J. Comput. Biol.* 17, 3. <https://doi.org/10.14806/ej.17.1.200>.
- Nakao, K., Miyaaki, H., and Ichikawa, T. (2014). Antitumor function of microRNA-122 against hepatocellular carcinoma. *J. Gastroenterol.* 49, 589–593. <https://doi.org/10.1007/s00535-014-0932-4>.
- Narla, V., Bhakta, N., Freedman, J.E., Tanriverdi, K., Maka, K., Deeks, S.G., Ganz, P., and Hsue, P. (2018). Unique circulating MicroRNA profiles in HIV infection. *J. Acquir. Immune Defic. Syndr.* 79, 644–650. <https://doi.org/10.1097/qai.0000000000001851>.
- Neidleman, J., Luo, X., Frouard, J., Xie, G., Gill, G., Stein, E.S., McGregor, M., Ma, T., George, A.F., Kusters, A., et al. (2020). SARS-CoV-2-Specific T cells exhibit phenotypic features of helper function, lack of terminal differentiation, and high proliferation potential. *Medicine* 1, 100081. <https://doi.org/10.1016/j.xcrm.2020.100081>.
- Peng, N., Yang, X., Zhu, C., Zhou, L., Yu, H., Li, M., Lin, Y., Wang, X., Li, Q., She, Y., et al. (2018). MicroRNA-302 cluster downregulates enterovirus 71-induced innate immune response by targeting KPNA2. *J. Immunol.* 201, 145–156. <https://doi.org/10.4049/jimmunol.1701692>.
- Price-Haywood, E.G., Burton, J., Fort, D., and Seoane, L. (2020). Hospitalization and mortality among black patients and white patients with covid-19. *N. Engl. J. Med.* 382, 2534–2543. <https://doi.org/10.1056/NEJMsa2011686>.
- Pulati, N., Zhang, Z., Gulimilamu, A., Qi, X., and Yang, J. (2019). HPV16(+)-miRNAs in cervical cancer and the anti-tumor role played by miR-5701. *J. Gene. Med.* 21, e3126. <https://doi.org/10.1002/jgm.3126>.
- Ramanathan, S., Shenoda, B.B., Lin, Z., Alexander, G.M., Huppert, A., Sacan, A., and Ajit, S.K. (2019). Inflammation potentiates miR-939 expression and packaging into small extracellular vesicles. *J. Extracell. Vesicles* 8, 1650595. <https://doi.org/10.1080/20013078.2019.1650595>.
- Rothe, C., Schunk, M., Sothmann, P., Bretzel, G., Froeschl, G., Wallrauch, C., Zimmer, T., Thiel, V., Janke, C., Guggemos, W., et al. (2020). Transmission of 2019-nCoV infection from an asymptomatic contact in Germany. *N. Engl. J. Med.* 382, 970–971. <https://doi.org/10.1056/NEJMc2001468>.
- Sardar, R., Satish, D., and Gupta, D. (2020). Identification of novel SARS-CoV-2 drug targets by host MicroRNAs and transcription factors Co-regulatory interaction network analysis. *Front. Genet.* 11, 571274. <https://doi.org/10.3389/fgene.2020.571274>.
- Shen, Z., Tang, W., Guo, J., and Sun, S. (2017). miR-483-5p plays a protective role in chronic obstructive pulmonary disease. *Int. J. Mol. Med.* 40, 193–200. <https://doi.org/10.3892/ijmm.2017.2996>.
- Subramanyam, D., Lamouille, S., Judson, R.L., Liu, J.Y., Bucay, N., Derynck, R., and Blesch, R. (2011). Multiple targets of miR-302 and miR-372 promote reprogramming of human fibroblasts to induced pluripotent stem cells. *Nat. Biotechnol.* 29, 443–448. <https://doi.org/10.1038/nbt.1862>.
- Suo, T., Chen, G.Z., Huang, Y., Zhao, K.C., Wang, T., and Hu, K. (2018). miRNA-1246 suppresses acute lung injury-induced inflammation and apoptosis via the NF- κ B and Wnt/ β -catenin signal pathways. *Biomed. Pharmacother.* 108, 783–791. <https://doi.org/10.1016/j.biopha.2018.09.046>.
- Testa, U., Pelosi, E., Castelli, G., and Labbaye, C. (2017). miR-146 and miR-155: two key modulators of immune response and tumor development. *Noncoding RNA* 3, 22. <https://doi.org/10.3390/ncrna3030022>.
- van der Ree, M.H., van der Meer, A.J., de Bruijne, J., Maan, R., van Vliet, A., Welzel, T.M., Zeuzem, S., Lawitz, E.J., Rodriguez-Torres, M., Kupcova, V., et al. (2014). Long-term safety and efficacy of microRNA-targeted therapy in chronic hepatitis C patients. *Antiviral. Res.* 111, 53–59. <https://doi.org/10.1016/j.antiviral.2014.08.015>.
- Wang, F., Zhang, X., Zhong, X., Zhang, M., Guo, M., Yang, L., Li, Y., Zhao, J., and Yu, S. (2018). Effect of miR-483-5p on apoptosis of lung cancer cells through targeting of RBM5. *Int. J. Clin. Exp. Pathol.* 11, 3147–3156.
- WHO Coronavirus (COVID-19) Dashboard. <https://www.who.int/>.
- Williams, A.E., José, R.J., Mercer, P.F., Brealey, D., Parekh, D., Thickett, D.R., O’Kane, C., McAuley, D.F., and Chambers, R.C. (2017). Evidence for chemokine synergy during neutrophil migration in ARDS. *Thorax* 72, 66–73. <https://doi.org/10.1136/thoraxjnl-2016-208597>.
- World Health Organization. (2020a). COVID-19: surveillance, case investigation and epidemiological protocols. <https://www.who.int/internal-publications-detail/considerations-in-the-investigation-of-cases-and-clusters-of-covid-19>.
- World Health Organization. (2020b). Laboratory testing for 2019 novel coronavirus (2019-nCoV) in suspected human cases. <https://www.who.int/health-topics/coronavirus/laboratory-diagnostics-for-novel-coronavirus>.
- Wu, Z., and McGoogan, J.M. (2020). Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China. *J. Am. Med. Assoc.* 323, 1239–1242. <https://doi.org/10.1001/jama.2020.2648>.
- Xiong, Y., Liu, Y., Cao, L., Wang, D., Guo, M., Jiang, A., Guo, D., Hu, W., Yang, J., Tang, Z., et al. (2020). Transcriptomic characteristics of bronchoalveolar lavage fluid and peripheral blood mononuclear cells in COVID-19 patients. *Emerg. Microbes Infect.* 9, 761–770. <https://doi.org/10.1080/22221751.2020.1747363>.
- Yan, Z., Zang, B., Gong, X., Ren, J., and Wang, R. (2020). MiR-214-3p exacerbates kidney damages and inflammation induced by hyperlipidemic pancreatitis complicated with acute renal injury. *Life. Sci.* 241, 117118. <https://doi.org/10.1016/j.lfs.2019.117118>.
- Yang, Y., Li, Z., Yuan, H., Ji, W., Wang, K., Lu, T., Yu, Y., Zeng, Q., Li, F., Xia, W., and Lu, S. (2019). Reciprocal regulatory mechanism between miR-214-3p and FGFR1 in FGFR1-amplified lung cancer. *Oncogene* 8, 50. <https://doi.org/10.1038/s41389-019-0151-1>.
- Zhang, R., Su, H., Ma, X., Xu, X., Liang, L., Ma, G., and Shi, L. (2019). MiRNA let-7b promotes the development of hypoxic pulmonary hypertension by targeting ACE2. *Am. J. Physiol. Lung. Cell. Mol. Physiol.* 316, L547–L557. <https://doi.org/10.1152/ajplung.00387.2018>.
- Zhang, S., Amahong, K., Sun, X., Lian, X., Liu, J., Sun, H., Lou, Y., Zhu, F., and Qiu, Y. (2021). The miRNA: a small but powerful RNA for COVID-19. *Brief. Bioinformatics* 22, 1137–1149. <https://doi.org/10.1093/bib/bbab062>.
- Zhao, L., Zhang, X., Wu, Z., Huang, K., Sun, X., Chen, H., and Jin, M. (2019). The downregulation of MicroRNA hsa-miR-340-5p in IAV-infected A549 cells suppresses viral replication by targeting RIG-I and OAS2. *Nucleic Acids* 14, 509–519. <https://doi.org/10.1016/j.omtn.2018.12.014>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Human blood samples (COVID-19)	The Renmin Hospital of Wuhan University-China	Data S1
Human blood samples (Healthy donors)	The First Affiliated Hospital of Xi'an Jiaotong University-China	Data S1
Chemicals, peptides, and recombinant proteins		
PBS	ThermoFisher	Cat# 20,012,068
Ethanol	Sigma	Cat# E7023
SUPERase•In™ RNase Inhibitor	ThermoFisher	Cat# AM2694
Critical commercial assays		
Plasma/Serum Circulating and Exosomal RNA Purification Kit	Norgen	Cat#51000
Next Multiplex Small RNA Library Prep Set	New England Biolabs	Cat#E7300
Qubit dsDNA HS Assay Kit	Invitrogen	Cat#Q32854
Deposited data		
The raw sequence data http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE166160	This paper	GEO: GSE166160
Software and algorithms		
R	https://www.cran.r-project.org	V4.0.2
Prism (software)	https://www.graphpad.com	v8.0.2
Human Genomes	ftp://ftp.ensembl.org/pub/release-96/fasta/homo_sapiens/dna/	
Hairpin RNA Structures Sequences Predication	http://ma.tbi.univie.ac.at/cgi-bin/RNAfold.cgi	
ComputationalTarget Prediction Algorithms TargetScan	http://www.targetscan.org/vert_71/docs/help.html	
ComputationalTarget Prediction Algorithms miRanda	http://cbio.mskcc.org/miRNA2003/miranda.html	
Bowtie	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml	V1.0.0
OmicStudio	http://www.omicstudio.cn/tool	
Xgboost	Chen et al., 2019 https://dl.acm.org/doi/10.1145/2939672.2939785#pill-authors__content	V1.4.1.1

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Chengsheng Zhang (cszhang99@126.com).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- The raw sequence data have been deposited at GEO and are publicly available as of the data of publication. Accession number are listed in the [key resources table](#). Source data are provided with this paper.
- This manuscript doesn't report original code. R is a free software environment for statistical computing and graphics (<https://www.r-project.org/>). Machine learning using XGBoost method with the R package "xgboost" (https://dl.acm.org/doi/10.1145/2939672.2939785#pill-authors__content).

- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Ethics statement

This study was approved by the Ethics Committee of The First affiliated Hospital of Xi'an Jiaotong University (XJTU1AF2020LSK-015) and The Renmin Hospital of Wuhan University (WDRY2020-K130). All participants enrolled in this study offered the written informed consent by themselves or their surrogates. The high throughput sequencing of plasma samples was performed on existing samples collected during standard diagnostic tests, posing no extra burden to patients.

Case definition and study cohort

The definition and classification of all COVID-19 patients in this study follow the Guidelines of the World Health Organization and the "Guidelines on the Diagnosis and Treatment of the Novel Coronavirus Infected Pneumonia" developed by the National Health Commission of People's Republic of China ([World Health Organization, 2020a, 2020b](#); [World Health Organization, 2020a, 2020b](#); [Government of China, 2021](#)). The study cohort included 16 asymptomatic subjects (AS, $n = 16$) and 100 symptomatic patients (SM, $n = 100$) consisting of 52 moderate disease (MD, $n = 52$), 12 repeated samples from the MD patients (MD-R, $n = 12$), 48 severe disease (SD, $n = 48$), and 22 repeated samples from the SD patients (SD-R, $n = 22$). The SM group was further divided into the long-term nucleic acid test positive (LTNP, $n = 30$) and the short-term nucleic acid test positive (STNP, $n = 33$) sub-groups. In this study, based on the clinical observation that most of the COVID-19 patients hospitalized in the Renmin Hospital in Wuhan became negative for the nucleic acid test within 45 days, we therefore defined the STNP was ≤ 45 days whereas the LTNP was >60 days ([Figure 1A](#); [Table 1](#) and [S1](#); [Data S1](#)). The demographic features, clinical laboratory testing results and other relevant information were provided in [Data S1](#) and [Table S1](#).

METHOD DETAILS

Blood sample collection and plasma preparation

The peripheral blood was collected into the standard EDTA-K2 Vacuette Blood Collection Tubes (Jiangsu Yuli Medical Equipment Co., Ltd, China; Cat.Y09012282) and stored at room temperature or 4°C until processed. The plasma was prepared after centrifugation of the whole blood sample at 2500 rpm for 20 min and stored in the -80°C freezer until used for the studies. All the experimental procedures were completed inside a biosafety level 2 (BSL-2) laboratory at the Department of Clinical Diagnostic Laboratories, Renmin Hospital of Wuhan University. A total of 233 plasma samples were collected from 61 healthy controls (HC) and 116 SARS-CoV-2 infected individuals (COV) including samples collected at different time-points from some of these individuals.

RNA extraction, library construction and high throughput sequencing

The Norgen Plasma/Serum Circulating and Exosomal RNA Purification Kit (Slurry Format, Cat#51000) was used for isolation of the cell-free RNA from the plasma samples inside the BSL-2 laboratory described above. Approximately 3 μg of total RNA from each sample was used for the library construction by using the NEBNext Multiplex Small RNA Library Prep Set for Illumina (New England Biolabs, Cat. E7300) following the instructions of the manufacturer with an in-house modification, by which the pre-amplified small RNA molecules were labeled with unique molecular identifier (UMI) (LC-Bio Technology Co., LTD, China) to reduce potential duplication bias during the PCR and sequencing steps. The DNA library was purified and quantified by using 6% PAGE gel and the Qubit dsDNA HS Assay Kit (Invitrogen, Cat. Q32854), respectively. The DNA fragments were examined by using Qsep100TM bio-fragment analyser (Bioptic Inc., Taiwan). The high throughput sequencing was performed on the Novaseq 6000 sequencer (PE150 model, Illumina).

QUANTIFICATION AND STATISTICAL ANALYSIS

The miRNA-seq data analysis

The raw sequencing data was firstly filtered by using the fastx toolkit (version: 0.0.13.2, http://hannonlab.cshl.edu/fastx_toolkit/) to remove the low-quality reads, and the adaptor sequences were trimmed by using the cutadapt (version: 1.15) ([Martin, 2011](#)). To eliminate duplication bias introduced in the library

preparation and sequencing, the filtered reads were first clustered based on the UMI sequences and compared to each other by pairwise alignment. The reads with 100% sequence identity were then extracted to a new sub-cluster. Multiple sequence alignments were performed to obtain one consensus sequence for each sub-cluster after all the sub-clusters were generated. The reads were then subjected to an in-house program, ACGT101-miR (LC Sciences, Houston, Texas, USA) to remove common RNA families (rRNA, tRNA, snRNA, snoRNA) and repeats. Subsequently, unique sequences with length in 18–26 nucleotides were mapped to human precursors in miRBase 22.0 by BLAST search to identify known miRNAs and novel 3p- and 5p-derived miRNAs (Kozomara et al., 2019). The length variations at both 3' and 5' ends and one mismatch inside the sequence were accepted in the alignment. The unique sequences mapping to human mature miRNAs in hairpin arms were identified as known miRNAs, whereas the unique sequences mapping to the other arm of known human precursor hairpin opposite to the annotated mature miRNA-containing arm were considered to be novel 5p- or 3p-derived miRNA candidates. The unmapped sequences were BLASTed against the human genomes (ftp://ftp.ensembl.org/pub/release-96/fasta/homo_sapiens/dna/), and the hairpin RNA structures containing sequences were predicated from the flank 80 nt sequences using RNAfold software (<http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi>). The criteria for secondary structure prediction were: (1) number of nucleotides in one bulge in stem (≤ 12); (2) number of base pairs in the stem region of the predicted hairpin (≥ 16); (3) cut-off of free energy (kCal/mol ≤ -15); (4) length of hairpin (up and down stems + terminal loop ≥ 50); (5) length of hairpin loop (≤ 20); (6) number of nucleotides in one bulge in mature region (≤ 8); (7) number of biased errors in one bulge in mature region (≤ 4); (8) number of biased bulges in mature region (≤ 2); (9) number of errors in mature region (≤ 7); (10) number of base pairs in the mature region of the predicted hairpin (≥ 12); (11) percent of mature in stem (≥ 80).

Identification of differentially expressed miRNAs

To identify the differentially expressed miRNAs (DE-miRNAs), miRNA differential expression based on the normalized deep-sequencing counts was analyzed among the eleven groups by using Student's t test (Figure 1A). Nine of the eleven groups (COV vs HC, AS vs HC, SM vs HC, SM vs AS, MD vs AS, SD vs AS, SD vs MD, SD22 vs MD12 and LTNP vs STNP) were analyzed using the independent samples' t-test (Figure 1B), whereas two of them (MD12 vs MD-R and SD22 vs SD-R) were analyzed using the paired t-test. The threshold for significant difference was set to be p value ≤ 0.05 and the fold changes >1.50 or <0.67 . The miRNA with copy number less than 10 was considered to be low expression level. The miRNAs with copy number higher than 10 in at least one sample but less than the average copy number in the dataset (the sum of copy numbers / (sample number \times total number of miRNAs)) was considered to be middle expression level. The miRNA with copy number higher than the average copy number in the dataset was considered to be high expression level.

Target genes prediction

We employed the following two computational target prediction algorithms, TargetScan (http://www.targetscan.org/vert_71/docs/help.html) and miRanda (<http://cbio.mskcc.org/miRNA2003/miranda.html>) to identify the binding sites and predict the genes targeted by the differentially expressed miRNAs (Enright et al., 2003; Agarwal et al., 2015). TargetScan removes target genes with context score percentile <90 , whereas miRanda removes target genes with score cut-off <140 and the maximum energy > -20 kcal/mol. Finally, the data obtained by both algorithms were combined and the overlaps were calculated.

The comparison of sequences between the SARS-CoV-2 genome and miRNAs

To examine whether SARS-CoV-2 may generate virus-encoded miRNAs, the candidate miRNAs detected only in the SARS-CoV-2 infected subjects but not in the healthy controls were aligned to the SARS-CoV-2 genome (NC_045,512.2, a viral genome from Wuhan) with bowtie software (version:1.0.0). The allowed maximum number of mismatches was 3. On the other hand, TargetScan and miRanda were used to check whether the cellular miRNAs may target the SARS-CoV-2 genome. For all of miRNAs in our miRNA library, target gene prediction was performed to target at SARS-CoV-2 gene using miRanda and TargetScan with TargetScan score ≥ 90 , miRanda score ≥ 140 and maximum energy < -20 kcal/mol. In addition, miRTarBase (https://mirtarbase.cuhk.edu.cn/~miRTarBase/miRTarBase_2019/php/index.php) was used to search the miRNAs that target the cellular receptors associated with SARS-CoV-2.

Gene ontology (GO) and KEGG pathway enrichment analysis

For significant target genes, functional annotation analysis was performed to identify the most relevant biological terms, including GO biological process, GO molecular function, GO cellular component and

Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathway. For a GO term or pathway, P was calculated

$$\text{as } P = 1 - \sum_{i=0}^{S-1} \frac{\binom{B}{i} \binom{TB-B}{TB-i}}{\binom{TB}{TS}}. S \text{ denotes the number of significant genes that were mapped to a specific GO term or pathway, } B \text{ denotes the number of genes mapped to a specific GO or pathway, } TB \text{ represents the total number of genes mapped to all GO or pathways, } TS \text{ represents the total number of the significant differentially expressed genes. The flowchart for our analysis was showed in Figure 1A.}$$

Machine learning

The machine learning was performed using the expression data of the differentially expressed miRNAs for each of the compared groups. The framework of XGBoost (Extreme Gradient Boosting) was derived from the published study (Chen and Guestrin, 2016). The package includes efficient linear model solver and tree learning algorithms. In brief, 75% of the modeling data was selected as the training set, whereas the remaining 25% data was used as the testing set. From the training set, we selected significant microRNAs by machine learning using XGBoost method with the R package “xgboost” (version 1.4.1.1) (Chen and Guestrin, 2016). Then, the training sets served as the input to perform 5-fold-cross validation to get best parameters. In XGBoost model, L1 Regularization were considered by adjusting “reg_alpha” parameter in XGBoost. Besides, all parameters of these models were added in Data S10. Once the models were built, we applied the model to the test sets and used R package “InformationValue” (version:1.2.3) to get AUC curve. Besides, the performance of the XGBoost model in the training and test set were analyzed by “ggplot2” (version:3.3.5).

Statistical analysis

The two-sided Student’s t test and paired t-test were used to calculate the differentially expressed miRNAs in different compared groups. The threshold for significant difference was set to be $p \leq 0.05$, Fold changes >1.50 or <0.67 . The miRNAs with low expression levels were excluded for further analysis. For the differentially expressed miRNAs in LTNP vs STNP, the correlation analysis was performed based on the number of days in hospital with nucleic acid test positive and the normalized copy number of the significant differentially miRNA in the symptomatic patient samples. GraphPad Prism 8.0.2 and OmicStudio tools (<https://www.omicstudio.cn/tool>) were used to perform our analysis.