BIOINFORMATICS GENOME ANALYSIS

Novel SARS-CoV-2 encoded small RNAs in the passage to humans

Gabriela A. Merino[1,2,3], Jonathan Raad[1], Leandro A. Bugnon[1], Cristian Yones[1], Laura Kamenetzky[4,5], Juan Claus[6], Federico Ariel[7], Diego H. Milone[1] and Georgina Stegmayer[1]

[1] Research Institute for Signals, Systems and Computational Intelligence (sinc(i)), FICH-UNL, CONICET, Ciudad Universitaria UNL, Santa Fe, Santa Fe, 3000, Argentina.

[2] Bioengineering and Bioinformatics Research and Development Institute (IBB), FI-UNER, CONICET, Ruta 11 Km 10.5, Oro Verde, Entre Ríos, 3100, Argentina.

[3] European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridgeshire, CB101SD, United Kingdom.

[4] Instituto de Investigaciones en Microbiología y Parasitología Médica (IMPaM), Facultad de Medicina, UBA-CONICET, Ciudad Autónoma de Buenos Aires, 1121, Argentina.

[5] Laboratorio de Genómica y Bioinformática de Patógenos. iB3 | Instituto de Biociencias, Biotecnología y Biología traslacional. Departamento de Fisiología y Biología Molecular y Celular, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, 1121, Argentina.

[6] Laboratorio de Virología, FBCB, Ciudad Universitaria UNL, Santa Fe, Santa Fe, 3000, Argentina

[7] Instituto de Agrobiotecnología del Litoral (IAL), CONICET, FBCB, Universidad Nacional del Litoral, Colectora Ruta Nacional 168 km 0, Santa Fe, Santa Fe, 3000, Argentina.

ABSTRACT

Motivation: The Severe Acute Respiratory Syndrome-Coronavirus 2 (SARS-CoV-2) has recently emerged as the responsible for the pandemic outbreak of the coronavirus disease (COVID-19). This virus is closely related to coronaviruses infecting bats and Malayan pangolins, species suspected to be an intermediate host in the passage to humans. Several genomic mutations affecting viral proteins have been identified, contributing to the understanding of the recent animal-to-human transmission. However, the capacity of SARS-CoV-2 to encode functional putative microRNAs (miRNAs) remains largely unexplored.

Results: We have used deep learning to discover 12 candidate stem-loop structures hidden in the viral protein-coding genome. Among the precursors, the expression of eight mature miRNAs-like sequences was confirmed in small RNA-seq data from SARS-CoV-2 infected human cells. Predicted miRNAs are likely to target a subset of human genes of which 109 are transcriptionally deregulated upon infection. Remarkably, 28 of those genes potentially targeted by SARS-CoV-2 miRNAs are down-regulated in infected human cells. Interestingly, most of them have been related to respiratory diseases and viral infection, including several afflictions previously associated with SARS-CoV-1 and SARS-CoV-2. The comparison of SARS-CoV-2 pre-miRNA sequences with those from bat and pangolin coronaviruses suggests that single nucleotide mutations could have helped its progenitors jumping inter-species boundaries, allowing the gain of novel mature miRNAs targeting human mRNAs. Our results suggest that the recent acquisition of novel miRNAs-like sequences in the SARS-CoV-2 genome may have contributed to modulate the transcriptional reprogramming of the new host upon infection.

Contact: gstegmayer@sinc.unl.edu.ar

Supplementary information: Supplementary Data are available at Bioinformatics online.

1. INTRODUCTION

MicroRNAs (miRNAs) participate in post-transcriptional gene regulation influencing diverse biological processes such as development, proliferation, cell differentiation and metabolism across different cell types [1,2,3]. Several miRNAs have been linked to the regulation of host-pathogen interactions including viral infection, controlling apoptosis, evasion of the immune response and modulation of the viral replication cycle [4,5,6]. Host miRNAs have been associated with antiviral defense mechanisms [7], whereas it has also been reported the production of miRNAs derived from

viral genomes, hinting at a complex crossed post-transcriptional gene silencing interplay underlying the infection progression [8-10]. Most of the known viral miRNAs are encoded in genomes of either DNA viruses or RNA viruses with a nuclear stage, such as retroviruses, which can benefit from the nuclear host miRNA biogenesis machinery [11-13]. However, the production of functional miRNAs derived from cytoplasmic RNA viruses has also been observed [14,15]. In addition, the increasing sensitivity of deep sequencing technologies has also enabled the identification of other small viral non-coding RNAs from those RNA viruses [16]. Identifying both viral miRNAs and their targets in the human genome is of paramount importance, not only for understanding the molecular basis of the related disease but also for contributing to the identification of target molecules for the improvement of diagnostic and treatment strategies [17-20].

The Severe Acute Respiratory Syndrome-Coronavirus 2 (SARS-COV-2), responsible for the recent pandemic outbreak of a novel coronavirus disease (COVID-19), is a positive-single stranded RNA virus with a genome of ~30 kb. It encodes 14 open reading frames (ORFs) for producing up to 16 non-structural proteins (Nsp1-16) and four structural proteins typical of several coronaviruses: Spike (S), Envelope (E), Membrane (M) and Nucleocapsid (N) [7,10]. Interestingly, using deep sequencing data and experimental validation it was shown that the SARS-CoV emerged in China in 2002 encodes a group of small viral RNAs (svRNAs) functionally linked to the related lung pathogenesis [16]. However, the capacity of SARS-CoV-2 to encode functional miRNA-like effectors and their potential impact in the human transcriptome remains largely unexplored.

Here we used machine learning and a new deep learning model to identify miRNA-like precursors (pre-miRNAs) in the SARS-CoV-2 genome. Based on small RNA sequencing (RNA-seq) of infected epithelial cell cultures, we detected six of the identified pre-miRNAs potentially acting as precursors of eight mature miRNA-like molecules differentially accumulated during viral infection. Furthermore, analyzing bulk RNA-seq data of infected cell cultures we found deregulated 109 predicted human targets, linked to the host immune response and viral infection. Twenty-eight of those targets, most of them involved in respiratory diseases and viral infection, were found as down-regulated supporting the silencing effect of predicted mature miRNA-like molecules. We also performed a sequence comparison of pre-miRNAs with closely related coronaviruses. Strikingly, our results suggest that punctual recent mutations were responsible for the occurrence of the novel predicted mature miRNAs against human genes, thus potentially contributing to zoonotic transfer.

2. MATERIALS AND METHODS

2.1 DATA

Genome data of the virus were obtained from NCBI: Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome and its annotations (NCBI Reference Sequence NC_045512.2). In order to identify pre-miRNA sequences candidates, the SARS-CoV-2 genome was segmented with HextractoR [20a] into 600 nt long fragments with 500 nt overlaps, and the following parameters: single loop trimming, minimum sequence length 60, minimum number of base-pairs that must form a sequence 16 and final trimming optimizing the minimum free energy normalized by the sequence length (NMFE). These segments were folded into their secondary structures by using RNAfold [21] with default settings. Hairpin structures were extracted, obtaining 597 hairpins in total, which were considered as unlabeled samples. Pre-miRNA sequences of viruses, considered the positive labeled class, were downloaded from miRBase release 22.

Expression profiling by high throughput bulk RNA sequencing (RNA-seq) of the epithelial lung cancer cell line Calu-3 mock treated or infected with SARS-CoV-2 (USA-WA1/2020) were obtained from NCBI: Gene expression profiling of SARS-CoV-1/2 infected human cell lines at bulk and single-cell level (NCBI Reference series GSE148729), which was made public on May, 4th. Alignments files of small RNA-seq samples against SARS-CoV-1/2 genomes were downloaded from authors repository (https://filetransfer.mdc-berlin.de/) and filtered for keeping only those reads aligned to the SARS-CoV-2 genome.

2.2 PREDICTION OF PRE-MIRNAS IN THE SARS-COV-2 GENOME

A total of 73 structural features from the folded sequences of the virus genome and the well-known virus pre-miRNAs were extracted with miRNAfe [22] as in [23]. For example, features such as length of the sequence, minimum free energy (MFE), cumulative size of internal loops found in the secondary structure, number of loops, amount of GC, proportion of GC, among many others were extracted. Then, all features were normalized with z-score.

For recognizing pre-miRNAs in the virus genome, three machine learning (ML) methods were used: a deep convolutional neural network (CNN) [24], the deeSOM model [25], and a one-class support vector machine (OC-SVM) [26,27]. The first method, CNN, was trained directly with raw RNA sequences (substrings of the complete genome), their corresponding predicted secondary structure and the MFE of the folding. The first layer of the network developed here is a one-dimensional convolution, followed by eight stacked identity blocks [28] and pooling layers. The identity blocks allow the model to auto-define the number of convolutional layers needed during training, avoiding optimization of this critical hyperparameter. The next layers concatenate the MFE and flatten the output, and finally a fully connected layer generates the score for the input sequence.

The deeSOM model consists of several hierarchical layers with self-organizing maps (SOM) that has proven to be very suited to the pre-miRNA prediction task [25]. This model has an ensemble of unsupervised SOMs that are used in parallel at the first level. The unlabeled samples in the input data are split among the members of the ensemble, which also receive the full set of positive input cases. At each SOM layer, pre-miRNA neurons are identified as those having, at least, one positive sample. Only the sequences that are in pre-miRNA neurons pass to the next level. At each level, an adaptive algorithm determines the map size. Therefore, several deep layers are added with this self-size-adjusting method, until only known pre-miRNA samples remain at the last layer. The best pre-miRNAs candidates are identified as the ones in the pre-miRNA neurons of the last levels. Finally, the OC-SV model is able to learn a decision frontier only from the well-known pre-miRNAs. SVM has been traditionally one of the most used machine learning models for pre-miRNA prediction [23]. Each method provided a list of candidate sequences corresponding to potential pre-miRNAs supported by a score. Potential pre-miRNAs were identified by considering predictions with scores in the top 10th percentile in all three methods.

2.3 PREDICTION OF MATURE MIRNAS AND TARGETS

Mature miRNA-like molecules definition was done combining MatureBayes predictions [29] with the exploration of the profiles of small RNA-seq reads in the regions covered by the predicted precursors. In this sense, in addition to considering that the reads profiles indicated the production of a miRNA-like molecule, coverage higher than the mean was required. Moreover, since the protocol used for small RNA-seq could enrich polyA regions, miRNA-like occurrences in non-rich polyA regions were verified. Potential targets in the host cells for the mature miRNAs were predicted combining Diana (MR-microT software) [30] and miRDB (Custom prediction option) [31]. In both cases, author recommendations were followed and only targets with scores over 70 were considered. A reliable set of targets was obtained keeping only those genes predicted using both Diana and miRDB tools.

2.4 DIFFERENTIAL EXPRESSION AND FUNCTIONAL ENRICHMENT ANALYSES

Expression matrices obtained from the bulk RNA-seq of the Calu-3 cell cultures were analyzed with DESeq2 [32] for determining the set of differentially expressed genes (DEGs) as a response to the viral infection. Low expressed genes were discarded following DESeq2 manual suggestion and DEGs were detected using a false discovery rate (FDR) cutoff of 0.05, with a minimum value of 1 for the absolute log2 fold change. Three differential expression analyses were carried out for comparing infected and mock cultures at 24 hours upon infection (IM24), and for contrasting host gene expression at 12 and 24 hours upon infection against the respective samples at four hours (I12-4 and I24-4, respectively). The lists of DEGs were then combined and the biological impact caused by those DEGs was evaluated by means of an overrepresentation analysis (ORA) of Gene Ontology terms using PANTHER and PANTHER GO-Slim Biological Process (BP) annotation sets [33,34]. Enrichment of metabolic pathways was also evaluated using both, PANTHER pathways and Reactome [34a] databases. Fisher exact tests with Bonferroni correction for multiple testing (cutoff 0.05) were used taking all the genes with reliable expression after DESeq2 filtering as the reference list. The set of DEGs was also compared against the list of miRNA targets predicted in the host genome, identifying a list of DEGs targets. Additionally, the functionality of these genes was analyzed by an ORA but now considering the set of DEGs as the reference list. Enriched GO terms in BP PANTHER GO-Slim were found and targets DEGs responsible for those enrichments were explored. DEGs targets that were down-regulated in at least one comparison were detected and functionally characterized using PANTHER and the GOPlot R package [35].

2.5 MULTIPLE SEQUENCE ALIGNMENT

Multiple sequence alignment was performed for each predicted mature and its precursor with

genomes of the following coronaviruses: HKU1 (NCBI Reference Sequence DQ415897.1), OC43 (NCBI Reference Sequence KY983585.1), Bat-CoV RF1 (NCBI Reference Sequence DQ412042.1), Bat-CoV RATG13 (NCBI Reference Sequence MN996532.1), Malayan Pangolin (NCBI Reference Sequence MT072864.1), SARS-CoV (NCBI Reference Sequence AY278741.1), MERS-CoV (NCBI Reference Sequence MN120514.1), mice coronavirus (NCBI Reference Sequence GB KF294357.1) and ferret coronavirus (NCBI Reference Sequence 1264898). These alignments were performed with MUSCLE [36] from its web access with ClustalW type output format and default settings.

3. RESULTS

3.1 EIGHT POTENTIAL MATURE MIRNAS DERIVE FROM SIX PRECURSOR TRANSCRIPTS ENCODED IN THE SARS-COV-2 GENOME

The identification of potential pre-miRNAs encoded in the SARS-CoV-2 genome was performed by three alternative computational approaches: a deep learning model (mirDNN), a deep self-organizing map with ensemble layers (deeSOM) and a one-class support vector machine (OC-SVM) (details in 2.2). Each method provided a list of candidate sequences corresponding to potential pre-miRNAs supported by a score. Considering predictions with scores in the top 10th percentile in all three methods, 12 highly reliable precursors were identified (Figure 1A, Supplementary File 1). The read profiles in the regions covered by these precursors were explored in the alignments of small RNA-seq samples from a recently released dataset (NCBI, Series GSE148729) involving SARS-CoV-2 (USA-WA1/2020) infected cultures of the epithelial lung cancer cell line Calu-3. Those reads profiles were combined with MatureBayes [29] predictions for defining mature miRNAs-like sequences. Six of the reliable precursors that were predicted by machine learning techniques, all encoded in the positive strand of viral RNA (vRNA, Figure 1B), have shown to have a significant coverage in their respective miRNAs-like regions (Figure 1C, Supplementary File 2). Notably, two pre-miRNAs, SC2V-mir-M-3 and SCV-mir-ORF10, were predicted each to be processed into two alternative mature miRNAs, expanding the repertoire to eight miRNAs-like derived from vRNAs potentially regulating host genes (Supplementary File 3, examples of pre-miRNAs reads profiles and their predicted secondary structure are shown in Figure 1C and Figure1D, respectively; details in Supplementary File 2). Analyzing the accumulation of mature miRNAs from the small RNA-Seq dataset, it was found that it follows the progression of the estimated viral content (Supplementary File 4, panels A-B). Except for the SC2V-mir-ORF1ab-1, detected only at 24 hours, all the pre-miRNAs were detected in Calu-3 infected samples at both 12 and 24 hours upon infection, hinting at a differential timing of production of each mature miRNA throughout the infection cycle.

A search for related human miRNAs deposited in miRBase was done based on the miRNA seed sequence [37], verifying that there is no significant similarity between predicted viral miRNAs and host miRNAs. Expanding the search for matching the identified miRNA-like molecules with miRNAs from any organism, only the SC2V-mir-M-1-5p was found to be similar to one miRNA from *Drosophila Melanogaster* (dme-miR-4982-3p, E-value 4.6). Then, to better understand what processes may affect this subset of SARS-CoV-2 miRNAs and assuming that they are acting as host genes regulators, human genes potentially targeted by them were identified. For obtaining a reliable set of candidate targets, two independent tools were used: Diana (MR-microT software) [30] and miRDB (custom prediction option) [31]. As recommended by the authors of Diana and miRDB, only predicted targets with a score over 70 were considered. In addition, only those genes identified as potential targets using the two algorithms were considered for further analyses. Overall, a total of 725 human genes were identified as potential targets of predicted SARS-CoV-2 miRNAs (Supplementary File 5). The number of potential targets ranges from one (SC2V-mir-M-3-3p) to 234 (SC2V-mir-ORF1ab-1-3p) being 97 the average number of targets per mature miRNA (Figure 2A).

3.2 THE EXPRESSION PROFILE OF A SUBSET OF 109 POTENTIAL SARS-COV-2 MIRNA TARGETS IN THE HUMAN GENOME IS AFFECTED UPON INFECTION

It is known that miRNAs can modulate target mRNA stability or translation [38]. Thus, an assessment of the steady-state RNA levels of the subset of predicted target genes upon infection of cell cultures with SARS-CoV-2 was performed. To this end, the bulk RNA-seq expression matrix of the same samples of infected Calu-3 cell cultures used for identifying the set of reliable miRNAs was analyzed. Differential expression analyses were performed for comparing infected and mock cultures at 24 hours upon infection (IM24), and for contrasting host gene expression at 12 and 24 hours upon infection versus the respective samples at four hours (I12-4 and I24-4, respectively). Based

on the statistical analysis of the provided expression data (see Methods), three lists of differentially expressed genes (DEGs) were determined. The number of DEGs found was 1,877 for the IM24 contrast, 946 for the I12-4 comparison, and 1,955 for the I24-4. By joining these lists, a total of 2,663 deregulated genes were found, with nearly 25% in common across the three subsets of DEG (Figure 2B). Notably, among the 2,663 DEGs, 1,802 (over 67%) genes exhibited an upregulation upon infection. This suggests that the global modulation of the host transcriptome is not only the result of a massive transcriptional shutdown mediated by the viral non-structural protein 1 (Nsp1), an endonuclease capable of impairing host mRNA translation and promoting transcriptional degradation [38a, 38b, 38c].

For evaluating functional enrichment among DEGs, an overrepresentation analysis (ORA) was performed using PANTHER and the PANTHER GO-Slim Biological Process annotation set [33, 34]. Obtained results revealed that the three most enriched categories correspond to the type I interferon signaling pathway (GO:0060337), response to type I interferon (GO:0034340) and to the pyrimidine nucleobase catabolic process (GO:0006208), whereas the highest number of DEG in infected cells are grouped in the defense response term (GO:0006952). In addition, several GO terms also related to immune response (GO:0002456, GO:0030183, GO:0071621, GO:0072676, GO:0042100, GO:0042130, GO:0042742, GO:0050852) and specifics of the response to viral infection (GO:0051607) were underscored (Figure 2C). The enriched Reactome pathways found for DEG genes were seven, five of them related to signaling pathways (Interferon alpha/beta, R-HSA-909733.3; interferon gamma, R-HSA-877300.3; Interleukin 10 R-HSA-6783783.3; Interleukin 4 and Interleukin 13, R-HSA-6785807.6; GPCR downstream, R-HSA-388396.6). The remaining two pathways are Chemokine receptors bind chemokines (R-HSA-380108.3) and Immunoregulatory interactions between a Lymphoid and a non-Lymphoid cell (R-HSA-198933.6). Whereas, p53 (P00059) and apoptosis signaling (P00006) were the over-represented PANTHER pathways.

One hundred and nine DEGs (4.9%) were detected as potential target genes of six of the predicted SARS-CoV-2-derived miRNAs (Figure 3A). Thirty-three of them were found as deregulated in all the evaluated contrasts, whereas 45 targets were uniquely identified in any of the three comparisons (Supplementary File 6). Within those genes, *BIRC3* [39,40], *ACE2* [41], *CXCL3* [42], *JAK2* [43] and *STAT2* [44,45], have been linked to the SARS-CoV-2 infection. Twenty-eight of the 109 deregulated potential targets were significantly down-regulated in at least one of the comparisons evaluating changes caused by viral infection (Figure 3B). Except for *SAMD13*, potentially targeted by both SC2V-mir-ORF1ab-1-3p and SC2V-mir-ORF10-5p miRNAs, those genes were predicted as uniquely targeted by four of the predicted SARS-CoV-2-derived miRNAs. Among the 28 genes, only three were down-regulated for the comparison I12-4, being two of them (*RERG* and *NDRG1*) identified as down-regulated in the other two comparisons too. In addition, seven genes were found silenced in cell cultures after 24 hours upon infection with respect to both, the mock samples at the same time and the infected samples after four hours of the infection. Moreover, 11 genes were uniquely found as significantly down-regulated for the comparison of infected and mock samples at 24 hours. Interestingly, the expression of 18 of these down-regulated genes was found as anticorrelated with the potential miRNAs expression throughout time, i.e. mRNA levels decreased between 4h and 24h after infection (Supplementary File 4, panels B-F). The exploration of the functional annotations of those genes in PANTHER revealed that some of them participate in regulation processes, signaling pathways, and molecules binding, even having a catalytic activity (Figure 3C). ORA of DEG targets over all DEGs revealed a particular enrichment linked to the dendrite morphogenesis process and its regulation, as well as to neuronal development processes, potentially related to the neurological symptoms reported for COVID-19 patients [46] (Figure 3D). Among DEG targets that are down-regulated, genes such as *NTRK2, NAV1, ABHD2, ITGB6, PKIB,* and *ADSSL1* are annotated in the deregulated GO terms. Strikingly, three of these down-regulated target genes (*ITGB6*, *PDK1,* and *ST6GALNAC2*) have been linked to the response to SARS-CoV infections [47-49]. Collectively, our analyses indicate that multiple miRNAs-like predicted from the SARS-CoV-2 genome may dynamically modulate the host transcriptome upon infection.

3.3 SARS-COV-2 PREDICTED MIRNA-LIKE SMALL RNAS ARE A DISTINCTIVE FEATURE RECENTLY ACQUIRED IN COMPARISON WITH RELATED CORONAVIRUSES

Remarkably, the six pre-miRNAs derived from the SARS-CoV-2 genome coincide with ORFs and overlap with protein-coding genes (Figure 1B). Thus, evolutionary pressure exerted by the consequent conservation of protein sequences may

restrict the variability of miRNA precursors. Therefore, for assessing miRNA variability across related genomes, the corresponding regions of the predicted pre-miRNAs were examined in other coronaviruses infecting several organisms:
bat, pangolin, human, mice and ferret. No related sequences were retrieved by sequence identity searches in other human Betacoronaviruses genomes associated with mild symptoms, such as HKU1 (DQ415897.1) and OC43 (KY983585.1) [50]. The comparison with two bat (*Rhinolophus affinis*) coronaviruses RATG13 (MN996532.1) and RF1 (DQ412042.1), the Malayan Pangolin (*Manis javanica*) coronavirus (MT072864.1) and human SARS-CoV (AY278741.1) revealed that most of the predicted mature miRNAs were recently acquired with no or low impact on the corresponding coding capacity of overlapping ORFs. With mice (GB KF294357.1) and ferret (ID 1264898) coronaviruses, no significant conservation of the mature miRNAs across species was found. Overall, sequence alignment results indicated that the predicted mature miRNAs from the SARS-CoV-2 genome are more conserved in the RATG13 than in the Pangolin-CoV (Figure 4A, and two examples given in Figure 4B; details in Supplementary File 7). Notably, none of the identified mutations between RATG13 and SARS-CoV-2 resulted in an amino-acid substitution of the overlapped protein-coding genes, although punctual mutations were found in the seed sequence of three predicted mature miRNAs, likely affecting the recognition of target host mRNAs. Similarly, seven out of the eight miRNA-likes exhibit changes between SARS-CoV-2 and Pangolin-CoV, including four miRNAs with changes in the seed sequence. Remarkably, only one predicted miRNA (SC2V-mir-ORF1ab-1-3p) includes a mutation altering the related protein-coding sequence. Nearly all the predicted miRNAs exhibit differences in the RF1 and SARS-CoV genomes, including six cases with altered seeds sequences, respectively. Remarkably, only two miRNAs suffered changes in the overlapping encoded protein in each of these two latest genomes, being SC2V-mir-ORF1ab-2-5p the least conserved across all genomes assessed (Figure 4). Altogether, our results suggest that, if their function is verified, the acquisition of predicted novel miRNAs potentially targeting human genes may constitute a genome feature that permitted or helped SARS-CoV-2 progenitors to jump inter-species boundaries to infect humans.

4. DISCUSSION

The presence of numerous miRNAs in viral genomes has been previously described [11]. Although most of them were found in genomes of nuclear-replicating DNA viruses and in retroviral genomes, in recent years the existence of miRNA-like sequences has been described in genomes of RNA viruses that replicate in the cytoplasm, such as Ebola [51,52], dengue-2 [15], hepatitis A [53] and West Nile [54]. Their functionality is still a matter of debate. It has been speculated that the miRNAs encoded into viral RNA genomes could modulate the expression of host genes involved in the antiviral response [51,52] as well as of the own viral genes. Thus it can be stated that they are contributing to the regulation of the replication cycle [15,53,54]. Interestingly, it was recently reported that three small RNAs encoded into the SARS-CoV-1 genome are functional and play a role in lung pathology [16]. The mechanisms involved in the biogenesis of miRNAs from the RNA virus genomes whose replication is confined into the cytoplasm remain unclear, as the vRNA could not have access to the nuclear enzymatic complex necessary to initiate the canonical process of miRNA generation [12,55]. Notably, some proteins encoded in coronavirus genomes could be involved, directly or indirectly, in the cytoplasmic production of viral miRNAs. Among them, Nsp15 (EndoU) has a demonstrated activity as endoribonuclease, and it was shown that EndoU-deficient CoVs are highly attenuated [56]. Although it was proposed that EndoU may target vRNA [57], its potential role in the biogenesis of functional svRNAs remains unexplored. Besides, the Nsp1 protein, which also has the ability to induce cytoplasmic endonucleolytic activity and participates in the transcriptional shutdown of host mRNAs [58,59, 59a], causes disruption of the nuclear pore [60], that could be associated with the eventual relocation of targets or enzymes necessary for the coronavirus miRNA biogenesis process.
In humans, most miRNA genes are located in intergenic or intronic regions. However, several miRNA genes have also been found in exons and intron-exon junctions [61-63]. It has been proposed that the processing of exon-encoded pre-miRNAs could serve to destabilize the parent gene transcript [64]. Combining machine learning approaches and small RNA-seq data analysis, we identified six precursors of eight potentially active miRNAs in the SARS-CoV-2 genome. We found that all the predicted pre-miRNAs are overlapped to protein-coding genes. In agreement with the observations made in humans, it has been suggested that the processing of miRNA encoded in an RNA virus would destroy the parent genome, or replicative intermediates, conspiring against its own replication [12]. However, according to the estimation of Rouha

et al. (2010), less than 1% of the synthesized viral genome in cells infected with the flavivirus TBEV is needed to generate a functional amount of miRNAs, and therefore the productivity of the viral replication would be only marginally affected by the generation of miRNAs.

We found that several genes predicted as targets of our miRNA candidates were transcriptionally deregulated upon infection. It was proposed that between 10% and nearly 40% of human miRNA target genes are impaired in translation without altering RNA steady-state levels [65], suggesting that the effect of several viral miRNAs upon their host targets could remain undetectable by RNA-Seq approaches. Remarkably, among DEG targets we found 81 genes upregulated upon infection for at least one of the comparisons between the time-points assessed. For miRNAs blocking mRNA translation, this could be the result of negative feedback regulatory loops exerted by proteins over the transcription of their parent genes. Alternatively, this may indicate that viral miRNAs would be produced in different phases of the viral replication cycle, or depending on the viral charge in the cell, likely triggering post-transcriptional gene silencing under specific circumstances. A better understanding of miRNA biogenesis throughout the infection may allow the use of circulating viral miRNAs as biomarkers of the progression of the associated disease [66].

Interestingly, 28 genes potentially targeted by predicted SARS-CoV-2 miRNAs resulted down-regulated upon infection (Figure 3B). Remarkably, three down-regulated genes have been linked to SARS-CoV infection. The first one of them, *ST6GALNAC2*, encodes a sialyltransferase and its transcriptional accumulation was found highly associated with SARS pathogenesis [49]. The second one gene, *ITGB6*, encodes an integrin first linked to the infection by herpes simplex virus [67]. It has been recently found that the expression of *ITGB6* may be modulated by SARS-CoV-2 infection of human myocardium [47]. Notably, the induction of *ITGB6* by TFGβ was proposed to be related with pulmonary fibrosis [68], whereas genetic variants in this gene have been associated with the risk of pneumonitis in lung cancer patients treated with thoracic radiation therapy [69]. Furthermore, it has been suggested that *ITGB6* participates in cytomegalovirus-associated pneumonia [70]. The third coronavirus-related downregulated target is *PDK1*. Strikingly, SARS-CoV-1 membrane protein M induces apoptosis via interfering with the kinase PDK1 [48]. Here we showed that the M-locus derived miRNA, SC2V-mir-M-1-5p, would be able to target *PDK1* mRNA, whose accumulation significantly decreases upon infection (Figure 3B). Interestingly, the over-expression of the M protein in transgenic Drosophila has been shown to trigger mitochondrial-mediated apoptosis, together with the down-regulation of Akt protein phosphorylation in a PDK1-dependent manner. However, *PDK1* transcript levels were not quantified in M-expressing flies [71]. Our findings suggest that the *M* transgene would potentially produce active miRNAs against endogenous *PDK1* transcripts, thus affecting the PDK1-Akt survival pathway. Notably, the progression of different tumors, including non-small cell lung cancer, has been associated with the activity of several human endogenous miRNAs targeting *PDK1* [72-74], hinting at general post-transcriptional regulation of the *PDK1* gene expression in disease.

Additionally, 19 of the 28 down-regulated target genes have been previously linked to respiratory diseases, whereas eight have been related to the response to virus, being seven genes in common. *ABHD2* participates in virus propagation and the modulation of the immune response, for which it was identified as a potential target of anti-hepatitis B drugs [75]. Interestingly, genetic variations in this gene have been associated with obstructive pulmonary disease in a Chinese Han population and age-related pulmonary emphysema in mice [76]. Moreover, the expression of *ABHD2* was identified as a discriminating feature between smokers and non-smokers [77]. The deregulation of the tumor suppressor *NDRG1* has been associated with the progression of lung cancer [78]. Strikingly, infection by Merkel Cell Polyomavirus leads to *NDRG1* down-regulation and cellular proliferation and migration [79]. Further, the Porcine Reproductive and Respiratory Syndrome virus was found to facilitate viral replication through the down-regulation of *NDRG1* [80]. It was also suggested that this gene plays an important role in the epithelial physical barrier defects in chronic rhinosinusitis [81]. Notably, *NDRG1* is subjected to post-transcriptional gene silencing upon Influenza A infection [82]. The RNA stability of the chromatin-related gene *CBX5/HP1* has also been linked to the inflammatory response and apoptosis [83]. The chromatin disorganization suffered upon *CBX5* deregulation has been proposed to modulate lung tumor progression [84] as well as the infection by HIV [85-87]. In addition, the transcript levels of the *FUT9* gene encoding a fucosyltransferase have been associated with the modulation of the immune response to pathogens [88] and the progression to apoptosis [89]. Remarkably, it has been shown that the intraperitoneal administration of FUT9 in neonatal mice alleviates experimental

bronchopulmonary dysplasia, resulting in significant decrease in inflammation, alveolar-capillary leakage, alveolar simplification, and cell death [90]. The gene *NBEA* has been related to oropharyngeal carcinoma [91], and *RERG* behavior was considered a biomarker in nasopharyngeal Epstein-Barr virus-associated carcinoma [92]. Further potential SARS-CoV-2 miRNA targets were found implicated in lung development (*DNAJC22*, [93]) and function (*CTSH*, [94], and in the progression of lung cancer and disease (*NTRK2*, [95]; *SYT7*, [96,97]; *TMEM106B*, [98,99]; *PKIB*, [100]).

Several of the above-mentioned genes have been demonstrated to be regulated by human miRNAs. *ABHD2* encodes a serine hydrolase whose regulation by endogenous miRNAs may contribute to the progression of different tumors [101,102]. *REPS2*, involved in esophageal carcinoma [103] and apoptosis [104], is regulated by an endogenous miRNA enhancing tumorigenesis. Also, the miRNA-mediated regulation of *LZTS3* expression promotes lung cancer cell proliferation and metastasis [105]. Interestingly, we showed here that the SC2V-mir-ORF1ab-1-3p miRNA likely targets the *VAV3* circular RNA (circRNA). CircRNAs are highly stable covalently bound circular molecules capable of sequestering miRNAs, among other regulatory functions. Furthermore, several circRNAs have been related to antiviral immunity [106]. It was shown that the interplay between *VAV3* circRNA and associated miRNAs can modulate the progression of lung tumors in humans [107]. Moreover, chicken *VAV3* sponges gga-miR-375 promoting the epithelial-mesenchymal transition [108]. Notably, it was proposed that gga-miR-375 plays a key role in tumorigenesis post avian leukosis virus infection in chicken [109]. Finally, it is worth mentioning that the SARS-CoV-2 potential target *LTBP1* is regulated by the human miRNA-155, activating the expression of downstream oncogenes and repressing the TGF-β signaling pathway [110]. Strikingly, it was shown that the double-stranded DNA virus causing Marek's disease encodes an analog of miR-155. The viral miRNA promotes the proliferation of chicken embryo fibroblasts by targeting endogenous mRNAs [111]. Collectively, our observations suggest that SARS-CoV-2-derived miRNA-like sequences may modulate the host transcriptome promoting the infection progression.

Although several efforts are being carried out to prove the origin of the SARS-CoV-2, it is a matter still under debate. A comparison with closely related coronaviruses served to propose the bat *Rhinolophus affinis* and the pangolin *Manis javanica* as potential reservoir hosts of SARS-CoV-2 progenitors [112, 113]. However, the divergences detected here between the predicted mature miRNAs of SARS-CoV-2 and Pangolin CoV seem to support other evidences that deny the role of the latter as a direct ancestor of the former [114]. The identification of punctual mutations in the Spike receptor-binding domain from these coronaviruses served to explain the SARS-CoV-2 improved binding to the human cell receptor ACE2 and its rapid human-to-human transmission [115], whereas changes in their structure due to post-translational modifications have been recently linked to differential protein stability and interaction with the host receptors [116]. Here, we identified a group of accumulated punctual mutations, many of which do not alter the corresponding encoded proteins, notably between RATG13 and SARS-CoV-2, albeit affecting the potential activity of the novel derived miRNAs in the modulation of the human transcriptome. Thus, these genomic variations may have contributed to this or another as yet unknown ancestor towards overcoming the barriers between species and acquiring the ability to replicate in humans. Notably, it has been shown that a variety of model organisms and cell lines exhibit a differential susceptibility to SARS-CoV-2 [117,118]. Although it has been proposed that these differences are likely dependent on species specific make-up for ACE2 [119], we cannot exclude that the emergence of novel active miRNAs derived from the viral genome may participate in the host-virus interplay restricting the range of susceptible animals.

The international scientific community is devoting enormous efforts to cope with COVID-19. The experimental validation of the activity of these potential miRNA encoding genes, their role in the recent evolution of SARS-CoV-2, and their underlying molecular mechanisms will likely help to push back the frontiers of human therapeutics in the context of the actual sanitary crisis worldwide.

ACKNOWLEDGMENTS

Authors would like to thank M. Cucher for her contributions in the identification of pre-miRNAs; and Emanuel Wyler and Prof. M. Landthaler for providing the alignments of the small RNA-seq data of SARS-CoV-1/2 infected human cells, which were used for validating our predictions. We thank NVIDIA Corporation for the donation of the GPUs used for this project.

FUNDING

This work was supported by ANPCyT (PICT 2016 #0007, PICT 2018 #3384), UNL (CAI+D 2016 #082), UNER (PID 2019 #6204).

CONFLICT OF INTEREST. *None to declare.*

REFERENCES

1. Bartel, D. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116, 281–297 (2004)
2. Hausser, J., & Zavolan, M. Identification and consequences of miRNA–target interactions—beyond repression of gene expression. Nature Reviews Genetics 15(9), 599-612 (2014)
3. Gurtan, A. M., & Sharp, P. A. The role of miRNAs in regulating gene expression networks. Journal of molecular biology 425(19), 3582-3600 (2013)
4. Kincaid, R.P. & Sullivan, C.S. Virus-Encoded microRNAs: An Overview and a Look to the Future. PLoS Pathog 8(12), e1003018 (2012)
5. Bruscella, P. et al. Viruses and miRNAs: More Friends than Foes. Frontiers in microbiology 8, 824 (2017)
6. Isaac, W. & Renne, R. Viral miRNAs: tools for immune evasion. Current opinion in microbiology 13(4), 540-545 (2010)
7. Kemp, V. et al. miRNA repertoire and host immune factor regulation upon avian coronavirus infection in eggs. Arch Virol 165, 835–843(2020)
8. Li, X. & Zou, X. An overview of RNA virus-encoded microRNAs. ExRNA 1, 37 (2019)
9. Grundhoff, A. & Sullivan, C.S. Virus-encoded microRNAs. Virology 411(2), 325–343 (2011)
10. Guzzi, P.H., Mercatelli, D., Ceraolo, C., & Giorgi, F.M. Master Regulator Analysis of the SARS-CoV-2/Human interactome. J. Clin. Med. 9(4), 982 (2020)
11. Bernier, A. & Sagan, S.M. The Diverse Roles of microRNAs at the Host?Virus Interface. Viruses 10(8), 440 (2018)
12. Cullen, B.R. Five questions about viruses and microRNAs. PLoS Pathog 6(2), e1000787 (2010)
13. Chi, J-Q et al. Marek?s disease virus-encoded analog of microRNA-155 activates the oncogene c-Myc by targeting LTBP1 and suppressing the TGF-ß signaling pathway. Virology 476, 72-84 (2015)
14. Rouha, H., Thurner, C., Mandl, C.W. Functional microRNA generated from a cytoplasmic RNA virus. Nucleic Acids Res. 38(22), 8328–8337 (2010)
15. Hussain, M. & Asgari, S. MicroRNA-like viral small RNA from Dengue virus 2 autoregulates its replication in mosquito cells. Proc Natl Acad Sci U S A 111(7), 2746–2751 (2014)
16. Morales, L. et al. SARS-CoV-Encoded Small RNAs Contribute to Infection-Associated Lung Pathology. Cell Host Microbe 21(3),344–355 (2017)
17. Ivashchenko, A., Rakhmetullina, A. & Aisina, D. How miRNAs can protect humans from coronaviruses COVID-19, SARS-CoV, and MERS-CoV. Preprint at https://www.researchsquare.com/article/rs-16264/v1 (2020)
18. Huan, T. et al. Genome-wide identification of microRNA expression quantitative trait loci. Nature Communications 6(1), 6601 (2015).
19. Takahashi, R., Miyazaki, H., & Takeshita, F. Loss of microRNA-27b contributes to breast cancer stem cell generation by activating ENPP1. Nature Communications 6(1), 7318 (2015).
20. Cheng, C. et al. MicroRNA silencing for cancer therapy targeted to the tumour microenvironment. Nature 518(1), 107–110 (2015).
20a. Yones, C., Macchiaroli, N., Kamenetzky, L., Stegmayer, G. & Milone, D.H. HextractoR: an R package for automatic extraction of hairpins from genome-wide data. bioRxiv 2020.10.09.333898 (2020)
21. Hofacker, I.L. Vienna RNA secondary structure server. Nucleic Acids Res. 31, 3429–3431 (2003)
22. Yones, C., Stegmayer, G., Kamenetzky, L. & Milone, D.H. miRNAfe: A comprehensive tool for feature extraction in microRNA prediction, Biosystems 138,1-5 (2015)
23. Stegmayer, G. et al. Predicting novel microRNA: a comprehensive comparison of machine learning approaches. Briefings in Bioinformatics 20, 1607-1620 (2019)
24. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. Nature 521, 436–444 (2015).
25. Bugnon, L., Yones, C., Milone D.H. & Stegmayer, G. Deep neural architectures for highly imbalanced data in bioinformatics. IEEE Trans. on Neural Networks and Learning Systems, in press, 10.1109/TNNLS.2019.2914471 (2019)
26. Tran, D., Pham, T., Satou, K. & Ho, T. Prediction of microRNA Hairpins using One-Class Support Vector Machines. Int. Conf. Bioinformatics and Biomedical Engineering, ICBBE 2008, 1, 33-36, DOI: 10.1109/ICBBE.2008.15 (2008)
27. Yousef, M., Najami, N. & Khalifa, W. A comparison study between one-class and two-class machine learning for MicroRNA target detection. J. Biomedical Science and Engineering. 333033. 247-252 (2010)
28. He K., Zhang X., Ren S. & Sun J. Identity Mappings in Deep Residual Networks. In: Leibe B., Matas J., Sebe N., Welling M. (eds) Computer Vision .Lecture Notes in Computer Science, vol 9908 (2016)
29. Gkirtzou, K., Tsamardinos, I., Tsakalides, P. & Poirazi, P. MatureBayes: a probabilistic algorithm for identifying the mature miRNA within novel precursors. PLoS One 5(8), e11843 (2010)
30. Reczko, M. et al. Functional microRNA targets in protein coding sequences. Bioinformatics 28(6), 771-776 (2012)
31. Chen, Y., & Wang, X. miRDB: an online database for prediction of functional microRNA targets. Nucleic Acids Research 48(D1), D127-D131 (2020)
32. Love, M.I., Huber, W., & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology 15(12), 550 (2014)
33. Thomas, P. et al. PANTHER: a library of protein families and subfamilies indexed by function. Genome Res. 13, 2129-2141 (2003)
34. Mi, H. et al. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. Nucleic Acids Research 47(D1), D419-D426 (2019)
34a. Jassal, B. et al. The REACTOME pathway knowledgebase. Nucleic Acids Research 48(D1), D498-D503 (2020)
35. Walter, W., Sanches Cabo, F. & Ricote, M. GOplot: an R package for visually combining expression data with functional analysis. Bioinformatics 31(17), 2912-2914 (2015)
36. Madeira, F. et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Research 47(W1):W636-W641 (2009)
37. Kehl, T. et al. About miRNAs, miRNA seeds, target genes and target pathways. Oncotarget 8(63), 107167 (2017)
38. Bartel, D.P. MicroRNAs: target recognition and regulatory functions. Cell 136(2), 215–233 (2009)
38a. Kamitani, W., Huang, C., Narayanan, K. et al. A two-pronged strategy to suppress host protein synthesis by SARS

coronavirus Nsp1 protein. Nature Struct Mol Biol 16, 1134–1140 (2009)
38b. Kamitani, K. et al. Severe acute respiratory syndrome coronavirus nsp1 protein suppresses host gene expression by promoting host mRNA degradation. Proceedings of the National Academy of Sciences 103(34), 12885-12890 (2006)
38c. Lokugamage, KG. et al. Middle East Respiratory Syndrome Coronavirus nsp1 Inhibits Host Gene Expression by Selectively Targeting mRNAs Transcribed in the Nucleus while Sparing mRNAs of Cytoplasmic Origin. Journal of Virology 89(21), 10970-81 (2015)
39. Fagone, P. et al. Transcriptional landscape of SARS-CoV-2 infection dismantles pathogenic pathways activated by the virus, proposes unique sex-specific differences and predicts tailored therapeutic strategies. Autoimmunity Reviews (19) 7, 102571 (2020)
40. Karakurt, H.U. & Plr, P. Integration of transcriptomic profile of SARS-CoV-2 infected normal human bronchial epi-thelial cells with metabolic and protein-protein interaction networks. Turkish journal of biology 44(3), 168–177 (2020)
41. Ziegler, C. et al. SARS-CoV-2 Receptor ACE2 Is an Interferon-Stimulated Gene in Human Airway Epithelial Cells and Is Detected in Specific Cell Subsets across Tissues. Cell 181(5), 1016-1035 (2020)
42. Blanco-Melo, D. et al. Imbalanced Host Response to SARS-CoV-2 Drives Development of COVID-19. Cell 181(5), 1036-1045 (2020)
43. Bourgonje, AR. et al. Angiotensin-converting enzyme 2 (ACE2), SARS-CoV-2 and the pathophysiology of coronavirus disease 2019 (COVID-19). J Pathol. 10.1002/path.5471 (2020)
44. Song, P. et al. Cytokine Storm Induced by SARS-CoV-2. Clinica Chimica Acta (2020).
45. Boudewijns, R. et al. STAT2 signaling as double-edged sword restricting viral dissemination but driving severe pneumonia in SARS-CoV-2 infected hamsters. BioRxiv (2020).
46. Romero-Sánchez, C. M. et al. Neurologic manifestations in hospitalized patients with COVID-19: The ALBACOVID registry. Neurology 77(6), 683-690 (2020)
47. Bristow, M. et al. Dynamic Regulation of SARS-CoV-2 Binding and Cell Entry Mechanisms in Remodeled Human Ventricular Myocardium. JACC: Basic to Translational Science. doi.org/10.1016/j.jacbts.2020.06.007 (2020)
48. Tsoi, H., Li, L., Chen, Z.S., Lau, K.F., Tsui, S.K., Chan, H.Y. The SARS-coronavirus membrane protein induces apoptosis via interfering with PDK1-PKB/Akt signalling. Biochem J. 464(3), 439-447 (2014)
49. Nima, H. et al. Neutrophils, Crucial, or Harmful Immune Cells Involved in Coronavirus Infection: A Bioinformatics Study. Frontiers in Genetics 11, 641 (2020)
50. Corman, V. M., Muth, D., Niemeyer, D. & Drosten, C. Hosts and Sources of Endemic Human Coronaviruses. Adv. Virus Res., 100, 163–188 (2018).
51. Duy, J. et al. Virus-encoded miRNAs in Ebola virus disease. Sci Rep. 8(1):6480 (2018)
52. Liang, H. et al. Identification of Ebola virus microRNAs and their putative pathological function. Sci China Life Sci. 57(10):973–981 (2014)
53. Shi, J., Sun, J., Wu, M., Hu, N. & Hu, Y. Hepatitis A virus-encoded miRNAs attenuate the accumulation of viral genomic RNAs in infected cells. Virus Genes 52(3):317–324 (2016)
54. Hussain, M. et al. West Nile virus encodes a microRNA-like small RNA in the 3' untranslated region which up-regulates GATA4 mRNA and facilitates virus replication in mosquito cells. Nucleic Acids Res. 40(5):2210–2223 (2012)
55. Harwig, A., Das, A.T. & Berkhout, B. Retroviral microRNAs. Curr Opin Virol 7:47–54 (2014)
56. Deng, X., et al. Coronavirus nonstructural protein 15 mediates evasion of dsRNA sensors and limits apoptosis in macrophages. Proc. Natl. Acad. Sci. USA 114 (21), E4251–E4260 (2017)
57. Deng, X. & Baker, S.C. An "Old" protein with a new story: Coronavirus endoribonuclease is important for evading host antiviral defenses. Virology 5, 517:157–163 (2018)
58. Huang, C. et al. SARS coronavirus nsp1 protein induces template-dependent endonucleolytic cleavage of mRNAs: viral mRNAs are resistant to nsp1-induced RNA cleavage. PLoS Pathog 7(12):e1002433 (2011)
59. Nakagawa, K. et al. The Endonucleolytic RNA Cleavage Function of nsp1 of Middle East Respiratory Syndrome Coronavirus Promotes the Production of Infectious Virus Particles in Specific Human Cell Lines. J Virol 92(21):e01157-18 (2018)
59a. Huang, C., Lokugamage, K.G., Rozovics, J.M., Narayanan, K., Semler, B.L., Makino, S. SARS Coronavirus nsp1 Protein Induces Template-Dependent Endonucleolytic Cleavage of mRNAs: Viral mRNAs Are Resistant to nsp1-Induced RNA Cleavage. PLoS Pathog 7(12), e1002433 (2011)
60. Gomez, G.N., Abrar, F., Dodhia, M.P., Gonzalez, F.G. & Nag A. SARS coronavirus protein nsp1 disrupts localization of Nup93 from the nuclear pore complex. Biochem Cell Biol. 97(6):758–766 (2019)
61. Kim, Y.K. & Kim, V.N. Processing of intronic microRNAs. EMBO J. 26, 775–783 (2007)
62. Morin, R.D. et al. Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. Genome Res. 18, 610–621 (2008)
63. Rodriguez, A., Griffiths-Jones, S., Ashurst, J.L. & Bradley, A. Identification of mammalian microRNA host genes and transcription units. Genome Res. 14, 1902–1910 (2004)
64. Han, J et al. Post Transcriptional cross regulation between drosha and DGCR8, Cell 136 (1) 75–84 (2009)
65. Eichhorn, S.W. et al. mRNA destabilization is the dominant effect of mammalian microRNAs by the time substantial repression ensues. Mol Cell. 56(1):104–115 (2014)
66. Correia C. et al. Circulating microRNAs as Potential Biomarkers of Infectious Disease. Frontiers in Immunology 8, 118 (2017)
67. Gianni, T., Salvioli, S., Chesnokova, L.S., Hutt-Fletcher, L.M., Campadelli-Fiume, G. avb6- and avb8-Integrins Serve As Interchangeable Receptors for HSV gH/gL to Promote Endocytosis and Activation of Membrane Fusion. PLoS Pathog 9(12), e1003806 (2013)
68. Tatler, A,L, et al. Amplification of TGFß Induced ITGB6 Gene Transcription May Promote Pulmonary Fibrosis. PLoS One 11(8), e0158047 (2016)
69. Yi, M. et al. Genetic variants in the ITGB6 gene is associated with the risk of radiation pneumonitis in lung cancer patients treated with thoracic radiation therapy. Tumour Biol 37(3), 3469-3477 (2016)
70. Arai, Y. et al. Effects of intrapulmonary viral tropism and

cytokine expression on the histological patterns of cytomegalovirus pneumonia. Pathol Int 62(9), 628-639 (2012)
71. Chan, C.M., Ma, C., Chan, W., Edwin Chan, H.Y. The SARS-Coronavirus Membrane protein induces apoptosis through modulating the Akt survival pathway. Archives of Biochemistry and Biophysics 459(2), 197-207 (2007)
72. Wang, Y., Zhang, Z., Wang, H. et al. miR-138-1* regulates aflatoxin B1-induced malignant transformation of BEAS-2B cells by targeting PDK1. Arch Toxicol 90, 1239–1249 (2016)
73. Chen, G-M., Zheng, A-J., Cai, J., Han, P., Ji, H-B., Wang, L-L. microRNA-145-3p inhibits non-small cell lung cancer cell migration and invasion by targeting PDK1 via the mTOR signaling pathway. J Cell Biochem. 119, 885–895 (2018)
74. Wei, Y., Liao, Y., Deng, Y., Zu, Y., Zhao, B., Li, F. MicroRNA-503 Inhibits Non-Small Cell Lung Cancer Progression By Targeting PDK1/PI3K/AKT Pathway. Onco Targets Ther. 12, 9005-9016 (2018)
75. Ding, X., Yang, J., Sun, D. et al. Whole genome expression profiling of hepatitis B virus-transfected cell line reveals the potential targets of anti-HBV drugs. Pharmacogenomics J 8, 61–70 (2008)
76. Liu, L. et al. Associations of ABHD2 Genetic Variations with Risks for Chronic Obstructive Pulmonary Disease in a Chinese Han Population. PLoS ONE 10(4), e0123929 (2015)
77. Shahdoust, M., Hajizadeh, E., Mozdarani, H., Chehrei, A. Finding genes discriminating smokers from non-smokers by applying a growing self-organizing clustering method to large airway epithelium cell microarray data. Asian Pac J Cancer Prev. 14(1), 111-116 (2013)
78. Wang, Y. et al. N-myc downstream regulated gene 1(NDRG1) promotes the stem-like properties of lung cancer cells through stabilized c-Myc. Cancer Letters 401, 53-62 (2017)
79. Gupta, P. et al. Merkel Cell Polyomavirus Downregulates N-myc Downstream-Regulated Gene 1, Leading to Cellular Proliferation and Migration. Journal of Virology 94 (3), e00899-19 (2020)
80. Wang, J. et al. Porcine Reproductive and Respiratory Syndrome Virus Activates Lipophagy To Facilitate Viral Replication through Downregulation of NDRG1 Expression. Journal of Virology 93(17), e00526-19 (2019)
81. Jiao, J. Wang, C. & Zhang, L. Epithelial physical barrier defects in chronic rhinosinusitis. Expert Review of Clinical Immunology 15:6, 679-688 (2019)
82. Gao, J et al. Integrated analysis of microRNA-mRNA expression in A549 cells infected with influenza A viruses (IAVs) from different host species. Virus Research 263, 34-46 (2019)
83. Kloetgen, A. et al. YBX1 Indirectly Targets Heterochromatin-Repressed Inflammatory Response-Related Apoptosis Genes through Regulating CBX5 mRNA. Int. J. Mol. Sci. 21, 4453 (2020)
84. Yu, Y., Chiou, G., Huang, P. et al. Network Biology of Tumor Stem-like Cells Identified a Regulatory Role of CBX5 in Lung Cancer. Sci Rep 2, 584 (2012)
85. Shimura, M., Toyoda, Y., Iijima, K., et al. Epigenetic displacement of HP1 from heterochromatin by HIV-1 Vpr causes premature sister chromatid separation. J Cell Biol 194(5),721-735 (2011)
86. Marban, C., Suzanne, S., Dequiedt, F. et al. Recruitment of chromatin-modifying enzymes by CTIP2 promotes HIV-1 transcriptional silencing. EMBO J. 26(2), 412-423 (2007)
87. Rohr, O., Lecestre, D., Chasserot-Golaz, S. et al. Recruitment of Tat to heterochromatin protein HP1 via interaction with CTIP2 inhibits human immunodeficiency virus type 1 replication in microglial cells. J Virol. 77(9), 5415-5427 (2003)
88. Kashiwazaki, H., Kakizaki, M., Ikehara, Y., et al. Mice lacking a1,3-fucosyltransferase 9 exhibit modulation of in vivo immune responses against pathogens. Pathology International 64(5), 199-208 (2014)
89. Azuma, Y., Kurusu, Y., Sato, H., Higai, K., Matsumoto, K. Increased expression of Lewis X and Y antigens on the cell surface and FUT 4 mRNA during granzyme B-induced Jurkat cell apoptosis. Biol Pharm Bull 30(4), 655-660 (2007)
90. Chaubey, S., Nader, Y.M., Shah, D. et al. a1,3-Fucosyltransferase-IX, an enzyme of pulmonary endogenous lung stem cell marker SSEA-1, alleviates experimental bronchopulmonary dysplasia. Pediatr Res. 10.1038/s41390-020-0891-9 (2020)
91. Gao, G., Kasperbauer, J.L., Tombers, N.M., Wang, V., Mayer, K., Smith, D.I. A selected group of large common fragile site genes have decreased expression in oropharyngeal squamous cell carcinomas. Genes Chromosomes Cancer 53(5), 392-401 (2014)
92. Zhao, W., Mo, Y., Wang, S. et al. Quantitation of DNA methylation in Epstein-Barr virus–associated nasopharyngeal carcinoma by bisulfite amplicon sequencing. BMC Cancer 17, 489 (2017)
93. Wingen, C., Aschenbrenner, A.C., Stümpges, B., Hoch, M., Behr, M. The Wurst protein: a novel endocytosis regulator involved in airway clearance and respiratory tube size control. Cell Adh Migr 3(1),14-18 (2009)
94. Bühling, F., Kouadio, M., Chwieralski, C.E. et al. Gene targeting of the cysteine peptidase cathepsin H impairs lung surfactant in mice. PLoS One 6(10),e26247 (2011)
95. Richard Gomez, D. et al. Integrative proteomic and transcriptomic analysis provides evidence for TrkB (NTRK2) as a therapeutic target in combination with tyrosine kinase inhibitors for non-small cell lung cancer. Oncotarget 9, 14268-14284 (2018)
96. Fei, Z., Gao, W., Xie, R., Feng, G., Chen, X., Jiang, Y. Synaptotagmin-7, a binding protein of P53, inhibits the senescence and promotes the tumorigenicity of lung cancer cells. Biosci Rep 39(2), BSR20181298 (2019)
97. Peng, X., Moore, M., Mathur, A., et al. Plexin C1 deficiency permits synaptotagmin 7-mediated macrophage migration and enhances mammalian lung fibrosis. FASEB J. 30(12).4056-4070 (2016)
98. Ou, S. et al. Identification of a novel TMEM106B-ROS1 fusion variant in lung adenocarcinoma by comprehensive genomic profiling. Lung Cancer 88(3), 352-354 (2015)
99. Kundu, S.T., Grzeskowiak, C., Fradette, J. et al. TMEM106B drives lung cancer metastasis by inducing TFEB-dependent lysosome synthesis and secretion of cathepsins. Nat Commun 9, 2731 (2018)
100. Dou, P., Zhang, D., Cheng, Z., Zhou, G., Zhang, L. PKIB promotes cell proliferation and the invasion-metastasis cascade through the PI3K/Akt pathway in NSCLC cells. Exp Biol Med 241(17),1911-1918 (2016)
101. He, Y. et al. miR-140-3p Inhibits Cutaneous Melanoma Progression by Disrupting AKT/p70S6K and JNK Pathways

through ABHD2. Molecular Therapy - Oncolytics 17, 83-93 (2020)
102. Yu, M., Zhang, L., Liu, Y., Liu, D., Guo, Z. Retinoic Acid Induces Differentiation of Mouse F9 Embryonic Carcinoma Cell by Modulating the miR-485 Targeting of Abhd2. Int J Mol Sci. 20(9), 2071 (2019)
103. Zhang, H., Duan, C.J., Zhang, H., Cheng, Y.D., Zhang, C.F. Expression and clinical significance of REPS2 in human esophageal squamous cell carcinoma. Asian Pac J Cancer Prev 14(5),2851-2857 (2013)
104. Yadav, S., Zajac, E., Singhal ET AL. POB1 over-expression inhibits RLIP76-mediated transport of glutathione-conjugates, drugs and promotes apoptosis. Biochemical and Biophysical Research Communications 328(4),1003-1009 (2005)
105. He, J., Yu, L., Wang, C.M., Zhou, X.F. MiR-1275 promotes non-small cell lung cancer cell proliferation and metastasis by regulating LZTS3 expression. Eur Rev Med Pharmacol Sci 22(9),2680-2687 (2018)
106. Wang, M., Yu, F., Wu, W., et al. Circular RNAs: A novel type of non-coding RNA and their potential implications in antiviral immunity. Int J Biol Sci 13(12),1497-1506 (2017)
107. Li, M., Zhang, S., Wu, N., Wu, L., Wang, C., Lin, Y. Overexpression of miR-499-5p inhibits non-small cell lung cancer proliferation and metastasis by targeting VAV3. Sci Rep 6, 23100 (2016)
108. Zhang, X., Yan, Y., Lin, W. et al. Circular RNA Vav3 sponges gga-miR-375 to promote epithelial-mesenchymal transition. RNA Biol 16(1),118-132 (2019)
109. Li, H., Shang, H., Shu, D. et al. gga-miR-375 plays a key role in tumorigenesis post subgroup J avian leukosis virus infection. PLoS One 9(4),e90878 (2014)
1104. Sarrazin, C., Wedemeyer, H., Cloherty, G. et al. Importance of very early HCV RNA kinetics for prediction of treatment outcome of highly effective all oral direct acting antiviral combination therapy. J Virol MethodS 214,29-32 (2015)
111. Dang, L., Teng, M., Li, H.Z. et al. Marek's disease virus type 1 encoded analog of miR-155 promotes proliferation of chicken embryo fibroblast and DF-1 cells by targeting hnRNPAB. Vet Microbiol 207,210-218 (2017)
112. Andersen, K.G. et al. The proximal origin of SARS-CoV-2. Nature Medicine 26, 450–452 (2020).
113. Zhang, T., Wu, Q. & Zhang, Z. Probable Pangolin Origin of SARS-CoV-2 Associated with the COVID-19 Outbreak. Curr Biol. 30(7),1346-1351(2020)
114. Li, X., Zai, J., Zhao, Q. et al. Evolutionary history, potential intermediate animal host, and cross-species analyzes of SARS-CoV-2. J Med Virol, 92(6), 602–611 (2020)
115. Lan, J. et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. Nature 581, 215–220(2020)
116. Wrobel, A.G., Benton, D.J., Xu, P. et al. SARS-CoV-2 and bat RaTG13 spike glycoprotein structures inform on virus evolution and furin-cleavage effects. Nat Struct Mol Biol, doi.org/10.1038/s41594-020-0468-7 (2020)
117. Takayama, K. In Vitro and Animal Models for SARS-CoV-2 research. Trends in Pharmacological Sciences 41(8), 513-517 (2020)
118. Yuan, L., Tang, Q., Cheng, T., Xia, N. Animal models for emerging coronavirus: progress and new insights. Emerg Microbes Infection 9(1), 949-961 (2020)
119. Johansen, M.D., Irving, A., Montagutelli, X. et al. Animal and translational models of SARS-CoV-2 infection and COVID-19. Mucosal Immunol doi 10.1038/s41385-020-00340-z (2020).

FIGURE CAPTIONS

Figure 1: Identification of potential miRNA from six precursors hidden in the protein-coding genome of SARS-CoV-2. A) Venn diagram showing the number of candidate sequences to pre-miRNAs in the SARS-CoV-2 virus found by three machine learning methods: deeSOM (blue), mirDNN (orange) and OC-SVM (green). Deep models are schematically represented. On top is the deeSOM, several layers with ensembles of self-organizing maps, with elastic map size and automatic depth. On bottom is the mirDNN, a novel convolutional neural network with several layers and identity blocks. B) Genome location of the six predicted pre-miRNAs in the SARS-CoV-2 virus. C) Two examples of the read profiles in small RNA-seq samples of Calu-3 cell cultures at 24 hours upon infection. The horizontal blue line indicates the average reads counts for the whole genome. D) Predicted hairpin structures of the two SARS-CoV-2 pre-miRNAs shown in C).

Figure 2: Predicted viral miRNA targets in the human genome and differentially expressed genes upon infection. A) Number of targets predicted for each SARS-CoV-2 miRNA candidate. Colors indicate the corresponding pre-miRNA. B) Venn diagram of differential expression results from RNA-seq experiments of Calu-3 cell cultures infected by SARS-CoV-2. Performed tests compared: i) Infected (I) versus mock (M) samples at 24 hours upon infection; ii) Infected samples at 4 and 12 hours upon infection; iii) Infected samples at 4 and 24 hours upon infection. C) Enrichment results for the top 30th terms of the Biological Process PANTHER GO-Slim caused by differentially expressed genes (DEGs). Terms are grouped based on the ontology hierarchy being the term in bold the deepest.

Figure 3: A subset of potential SARS-CoV-2 miRNA targets deregulated upon infection are related to specific biological and molecular functions. A) Number of targets predicted for each SARS-CoV-2 miRNA candidate that were detected as differentially expressed genes (DEGs) in RNA-seq experiments of Calu-3 cultures infected with the virus. Colors indicate the corresponding pre-miRNA. B) Fold change, in log-2 scale, of those predicted SARS-CoV-2 miRNAs targets that were detected as down-regulated in at least one of the contrasts evaluated. Marks indicate significant changes. C) GOChord plot showing Biological Process (BP) and Molecular Function (MF) GO terms of the targets DEGs down-regulated. D) Enrichment results for the overrepresentation analysis of targets DEGs over all the DEGs. Only those BP PANTHER GO-Slim terms with enrichment scores higher than 4.2 (average enrichment score) are shown. Terms are grouped based on the ontology hierarchy being the term in bold the deepest.

Figure 4: Punctual mutations detected across related coronaviruses hint at the acquisition of potential mature miRNAs targeting human genes. A) Graphical representation of alignment results of the predicted miRNAs in SARS-CoV-2 along with several coronaviruses genomes. Each box contains the number of changes detected in the genomic sequence of a particular miRNA and the number of changes detected in the sequence of the overlapping protein. The color of each box indicates the number of changes in the genomic sequence of the miRNA that are in its seed. B) Two examples of the alignment results of predicted miRNAs in SARS-CoV-2 against other coronaviruses. Positions corresponding to miRNAs seeds are marked and highlighted boxes represent mismatches. C) Illustration of the changes in the corresponding coding sequences (CDSs) overlapped with the two mature miRNAs shown in B).