1. Introduction

In December 2019 a new severe respiratory disease was reported in Wuhan, capital of the Hubei province of China. It was soon discovered that this atypical respiratory infection is caused by a novel coronavirus [1,2]. This new virus was named Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2) because of its homology with the SARS coronavirus (SARS-CoV) [3], which was responsible for acute respiratory disease and high mortality during 2002–2003 [4]. To date, SARS-CoV-2 is the seventh known coronavirus that infects humans [5]. It is known that the virus is transmitted via respiratory droplets and aerosols from person to person and, once inside the body, it can bind specific host receptors and enters host cells through endocytosis or membrane fusion [6,7]. A better understanding of how the virus interacts with the cellular machinery could spur the development of more effective therapies.

Recent studies explored repositioning of FDA-approved drugs in order to identify treatments against SARS-CoV-2 [8,9]. In another study the authors demonstrated the existence of target motifs in the SARS-CoV-2 genome suitable for specific binding with endogenous human micro and long non-coding RNAs (miRNAs and lncRNAs, respectively), which hints at the potential for miRNA-based drugs against the virus [10].

Multiple viruses such as influenza, HIV-1 and dengue are known to produce their own miRNAs. However, the role of miRNAs in the biology of single stranded RNA viruses (e.g. Coronaviridae family) is controversial, as they replicate in the cytoplasm and do not access the nuclear miRNAs’ apparatus. Some viruses may benefit from using their own or the host’s miRNAs to inhibit antiviral responses or to facilitate viral replication [11]. In particular, it was described how the hepatitis C virus is able to de-repress host miR-122 targets by sealing the endogenous miRNA, promoting the virus infection [12]. Moreover, cellular stress responses, cell-type and age can affect host miRNA expression profiles [13–15]. Accumulating evidence suggests that there is a global reduction of miRNA expression following endoplasmic reticulum stress (ER stress) [16,17]. ER stress and the activation of stress response have been associated with coronavirus infection [18,19] and they have been proposed as a mechanism that facilitates SARS-CoV replication [20].

Comparison of peripheral blood miRNA’s expression profiles between SARS-CoV-2 patients and healthy control donors shows the existence of
a class of miRNAs differentially expressed between the two conditions, these miRNAs appear to be involved in protein kinase expression regulation [21]. In another study a set of differentially expressed miRNAs between moderate and severe SARS-CoV-2 patients has been highlighted, functional enrichment analysis for predicted targets of miRNAs shows an involvement in processes such as virus binding and defense response to virus [22]. Given these observations, it is reasonable to think that the miRNA-based interplay between the host and SARS-CoV-2 may be the cause that virus accesses and attacks the host cells [23]. Moreover, it is possible that individual differences in miRNA expression profiles, caused by e.g. age or previous diseases, could affect the severity of infection or the effectiveness of an antiviral treatment. Considering the role that miRNAs can play in the replication and defense evasion mechanisms made by viruses inside the cell, we investigate here the possible role of the SARS-CoV-2 genome and known variants as potential sponge for the miRNAs expressed in the infected cells.

2. Results and discussion

We analyzed and compared the potential sponge effect of seven human coronaviruses (HCoV) on human endogenous miRNAs. Sequencing data for three pathogenic strains (SARS-CoV-2 (NC_045512.2), SARS-CoV (NC_004718.3), and MERS-CoV (NC_019843.3)) and four non-pathogenic strains (HCoV-OC43 (KU131570.1), HCoV-229E (NC_002645.1), HCoV-HKU1 (KF686346.1), and HCoV-NL63 (NC_005831.2)) are currently available and were used in this study. We tested these virus sequences against the set of 940 unique miRNAs known to be expressed in pulmonary epithelial cells, selected from the RNA Atlas database (see Methods). On average we found 122 (±12) miRNAs which can potentially interact with the genome of viruses in the pathogenic group (Fig. 1A blu bars) vs 85 (±13) miRNAs for the non-pathogenic group (Fig. 1A red bars). This analysis shows how all the analyzed virus sequences can potentially act as sponge for hundreds of endogenous human miRNAs and also that there are more predicted binding sites in the pathogenic group compared with the non-pathogenic one (Mann-Whitney P < 0.05).

Despite the high sequence identity between SARS-CoV-2 and SARS-CoV (~80%) only 50% of SARS-CoV-2 miRNA’s binding sites are shared with SARS-CoV (Fig. 1B), while a lower percentage of SARS-CoV-2 miRNA’s binding sites is shared with MERS (Fig. 1B).

Notably, 16 miRNA binding sites appear to be in common among the three pathogenic coronaviruses (Figure 1B), 6 of these are specific for the pathogenic group and do not appear in the non-pathogenic one (miR-10394-5p, miR-103a-1-5p, miR-103a-2-5p, miR-193a-5p, miR-4717-3p, miR-6740-5p) (Fig. 1C red bar).

Interestingly, downregulation of the exosomal miR-103a in SARS-CoV-2 has been observed when comparing SARS-CoV-2 infected patients with high vs low serum level of the D-dimer [24]. Moreover low levels of miR-103 have also been associated with deep vein thrombosis [25,26]. Furthermore downregulation of miR-193a has been observed in SARS-CoV-2 patients [27] but no relationship between miR-193a expression and disease severity has been reported to date.

No associations between the other four miRNAs and SARS-CoV-2 infection have been reported, however their downregulation has been associated with various other diseases. The downregulation of miR-4717 and the resulting upregulation of its target SOCS3 has been associated with type 1 diabetes [28].

Finally, the downregulation of miR-10394 and miR-6740 has been associated with venous thrombosis in colorectal cancer and cardiovascular disease respectively [29].

Given the impact that SARS-CoV-2 variants has on transmissibility, severity, and/or immunity, we investigated the distribution of putative miRNA binding sites in ~100k unique SARS-CoV-2 variant genomes in order to evaluate possible “sponge” effects. Globally, we found 580 binding sites for unique miRNAs across all of the analyzed genomes, 97 unique miRNA binding sites were found to be present in more than 99% of the genomes (Fig. 2).

There are 439 miRNAs that can be targeted by the core set of miRNAs whose binding sites are shared by more than 99% of the SARS-CoV-2 sequences. All the identified miRNAs with the associated miRNAs are listed in the supplementary S1. We then focused our attention on the top 5 miRNAs that are regulated by the highest number of the putative sequenced miRNAs (Fig. 3).

The miRNAs potentially sequenced by the SARS-CoV-2 sequence, could no longer be able to downregulate their targets, resulting in their upregulation. Analyzing the 5 top mRNA we found 7 different miRNAs for VEGF-A, PTEN and 6 different miRNAs for HIF1A, E2F1 and CCND1.

In agreement with our hypothesis, Ackermann and collaborators reported an up-regulation of VEGF-A in patients who died from SARS-CoV-2 [30], data also supported by clinical results demonstrating an
increase in VEGF-A levels in alveolar bronchial lavage fluid from COVID-19 patients [31, 32]. In another study, the PTEN regulation pathway showed the highest association with other activated pathways in human bronchial epithelial cells infected with SARS-CoV, moreover, it is also known how PTEN is involved in the activation of dendritic cells and secretion of proinflammatory mediators promoting the cytokine storm in the SARS-CoV-2 patients [33]. Cytokine storm and hypoxia are a prelude to multiple organ failure and lethality, in this context HIF1A plays a dual role, if on one hand it suppresses the angiotensin-converting enzyme 2 (ACE2) on the other hand it is involved in the activation of pro-inflammatory cytokine expression and the subsequent inflammatory process [34]. Moreover, the vascular endothelial growth factor (VEGF) is transcriptionally upregulated by HIF1A and accumulates under hypoxia enhancing cytokine storm. VEGFA can increase cytokine expression and pulmonary vascular permeability [35] and, notably, the endothelial cells are essential contributors to the initiation and propagation of severe forms of SARS-CoV-2 infection [36]. We also identified the transcription factor E2F1 which is known to be involved in virally-mediated acute respiratory diseases [37]. Finally, there are no studies regarding the involvement of CCND1 in the SARS-CoV-2 infection but its overexpression has been associated many times with bronchial epithelial diseases such as cancer. CCND1 is overexpressed in premalignant versus normal or hyperplastic epithelial cells [38, 39], thus highlighting its contribution to an unhealthy state of the tissue.

Finally, the association between the putative SARS-CoV-2 sponge activity and the severity of the disease was investigated. The miRNAs expression, as well as that of the genes and the proteins, can be influenced by several factors such as sex, ethnicity, health/disease state, or age. Aging is known to be related with changes in pulmonary physiology, pathology and function, during the period of lung infection. Therefore, there are several age-related differences in severity and tolerance that lead to worse clinical outcomes in SARS-CoV-2 infected elderly individuals [40]. To this purpose, the list of age-related miRNAs in bronchial human healthy biopsies released by Brandsma and collaborators has been used [15]. The authors highlighted 27 differentially
expressed miRNAs (13 miRNAs with lower expression levels and 14 with higher expression levels with increasing age). Interestingly, we found that miR-20a-5p, which shows a lower expression level in older subjects, can be sequestered by 99% of all mutated SARS-CoV-2 sequences used in this analysis (Fig. 4). Compared with the healthy controls, miR-20a is downregulated in the peripheral blood of human patients with SARS-CoV-2 [21] and, moreover, its downregulation has been reported to play a role in respiratory diseases influenza A and respiratory syncytial virus [41,42]. miR-20a is involved in the regulation of several targets enriched in biological processes such as: regulation of cell differentiation, cell proliferation and response to endogenous stimulus. Furthermore, the mRNA targets are enriched in pathways such as: cytokine signaling in the immune system, interleukin signaling, cyclin D associated events in G1. Pathways or biological processes that are widely associated with SARS-CoV-2 infection and that could be at the basis of the differences in the degree of the virus lethality in different age groups.

3. Materials and methods

3.1. Genome sequences collection and pre-processing

SARS-CoV-2 genome sequences and metadata were downloaded from the COVID-19 Data Portal (last access: 2021 June 11). Only the genomes with Homo sapiens host and coverage 100% were downloaded in FASTA format. To discard incomplete genomes, only sequences longer than 25kbp and without unresolved nucleotides were considered and deduplicated, obtaining a total of 103763 unique genomes. Each genome was aligned to the SARS-CoV-2 reference genome (NCBI accession ID: NC_045512.2) by applying the Multiple Alignment using Fast Fourier Transform (MAFFT) sequence aligner [44]. Variants were called by parsing the MAFFT alignment with a custom Python script.

3.2. Pulmonary epithelial cell’s miRNA and gene selection

The experimentally verified mature miRNAs and genes expressed in the pulmonary epithelial cells have been extracted from the RNA Atlas database [45].

3.3. Age-related miRNA in human bronchial biopsies

miRNAs whose expression was shown to be significantly related to age were selected from the work of Brandsma and collaborators [15]. The Authors highlighted 27 miRNAs differentially expressed between young and elder subjects. These included 13 miRNAs with lower expression levels and 14 with higher expression levels with increasing age.

3.4. RNA-RNA hybridization

The putative RNA-RNA hybridizations between the host miRNAs and
virus genomic sequence were calculated using the RNA hybrid tool [46]. We selected bulged hybridizations, i.e. hybridizations showing a perfect complementarity between the miRNA seed sequence and the viral genome, followed by 3 or 4 mismatches and then again followed by another stretch of perfect complementarity. This model was proposed and tested by Lavenniah and collaborators who have shown how the bulged hybridization can evade the degradation via RISC-mediated endonucleolytic cleavage upon miRNA binding [47].

3.5. miRNA-target interaction

The miRNA-mRNA target interactions were selected using the RAID v2.0 database; only human experimentally validated interactions were used for this analysis.

3.6. Functional enrichment analysis

Gene Ontology (GO) classification and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis were performed using the online tools Gene Set Enrichment Analysis (GSEA) [48].

4. Conclusions and future perspectives

In this study, we performed a bioinformatic analysis to investigate the potential sponge role of SARS-CoV-2 sequence on the miRNAs expressed in human pulmonary epithelial cells. We first compared putative binding sites of these human miRNA within a set of seven HCoV RNA genomes. The pathogenic group (MERS, SARS-CoV, SARS-CoV-2) showed a higher number of binding sites compared with the non-pathogenic one (OC43, HKU1, 229E, NL63). In this analysis, we identified 6 miRNA binding sites present only in the pathogenic group, namely miR-103a-5p, miR-103a-1-5p, miR-193a-5p, miR-4717-3p, miR-6740-5p. Interestingly, the down-regulation of these miRNAs has been seen to be associated with deep vein thrombosis and cardiovascular diseases. We then analysed the potential sponge effect of about 100 thousand SARS-CoV-2 genome variants and we identified a total of 439 miRNA binding sites, 97 of which are in common among more than 99% of the genomes analyzed. These miRNAs, potentially seized by the virus genome, may not carry out their regulatory action. The physiological expression level of a selected number of miRNAs shows a correlation with age and can also be influenced by previous pathologies. For this reason, the possible SARS-CoV-2 sponge effect on these miRNAs could be related with the disease severity. SARS-CoV-2 could seize the low expressed miRNAs causing a more severe disease in a certain portion of the population. This is the case of the miR-20a-5p which shows a lower expression level in older subjects and it is known to be involved in pathways such as cytokine and interleukin signaling.

These results could be a starting point for the development of new strategies for the treatment of SARS-CoV-2 infection, based on an enhanced autogenous miRNA production or exogenous administration. Moreover, the altered concentration in the bloodstream of specific miRNAs, caused by the SARS-CoV-2 sponge activity inside the cells, could be helpful to follow the disease progression. Finally, understanding the biological differences underlying the severity of SARS-CoV-2 is a key point for developing more accurate antiviral therapies.

Funding

AIRC project [to MHC] (grant number IG 23539).

Data availability statement

SARS-CoV-2 genome identifiers (NCBI accession ID: NC_045512.2), datasets and scripts used in this manuscript are available upon request.
