SARS-CoV-2 and miRNAs

Tetiana Rabiichuk
University of Tartu
tetiana.rabiichuk@ut.ee

Farid Hasanov
University of Tartu
farid.hasanov@ut.ee

Mykyta Baliesnyi
University of Tartu
mykyta.baliesnyi@ut.ee

1 Code Availability

We provide all relevant code and intermediate results of our pipeline [1].

2 Introduction

Infections with the new coronavirus, namely SARS-CoV-2, have led to millions of people suffering around the globe. According to (Worldometer, 2020), Covid-19 has caused 3,816,649 deaths and infected 176,608,059 people, as of 14/06/2021. Coronaviruses, in general, are enveloped viruses with a single positive-stranded RNA genome (26–32 kb in length). They are classified into four genera: Alphacoronaviruses (α), Betacoronaviruses (β), Gammacoronaviruses (γ), and Deltacoronaviruses (δ) (Woo et al., 2009). SARS-CoV-2 is a β coronavirus with a genome size of 29.9 kb. The new coronavirus shares about 80% of the genome sequence with SARS-COV, as per Gralinski and Menachery (2020).

micro-RNAs (miRNAs) are RNAs consisting of approximately 18-25 nucleotides, the precursors of which are transcribed from the genome and then undergo the maturation process. They regulate the expression of target proteins by inhibiting the translation or causing the degradation of mRNAs specific to them, as confirmed by Cullen (2013). In this case, genes that encode this mRNA are silenced, meaning that miRNAs directly influence the expression of genes. Currently more than 2000 miRNAs have been identified, which are involved in the regulation of about one-third of the human genome (McGuire et al., 2015).

The role of miRNAs in various human diseases has been studied extensively by Ardekani and Naeini (2010). Their findings document the participation of miRNAs in many types of cancer, cardiovascular disease, inflammatory disease, and others. Given miRNAs' established role in human diseases, the main question for this project, as posed by our supervisor, was: does the SARS-CoV-2 genome contain potential miRNAs against the human genome?

3 Methods

For the reference genome of SARS-CoV-2 we have used NCBI Reference Sequence: NC_045512.2. At first, we considered the full reference genome. However, as SARS-CoV-2 is a part of the Betacoronoviruses family, the large proportion of its genome is shared with other viruses in the family (see Figure 1). Since we were interested in the distinct behaviour of SARS-CoV-2, for all of our analysis we only considered a region of reference genome starting from the S-gene (a part of the reference genome, starting from 21563-th nucleotide), that is not shared among the viruses.

The first step in the pipeline was identification of hairpin structures in the reference genome. For that, we used 2 tools: Vmir (Grundhoff et al., 2006) and miRNAFold (Tav et al., 2016). Following the recommendations in Grundhoff (2011), we used default parameters for VMir, and selected hairpins with score $>= 115$ and window count $>= 35$. For miRNAFold, we also used the default parameters. As a result, we obtained 139 (vs 519 for the full genome) candidates from miRNAFold and 14 (vs 222 for the full genome) candidates from VMir.

In the next step, we filtered potential pre-miRNAs. From Zhang et al. (2006), we concluded that MFEI values are good indicators for distinguishing real miRNAs. Based on (Verma et al., 2020), we have selected a threshold for MFEI value to be $<= -0.85$. MFEI was calculated from MFE, which, in turn, was calculated from pre-miRNA sequences using ViennaRNA library (Lorenz et al., 2011). Next, we used iMiRNA-SSF tool (Chen et al., 2016) for further filtering (separately for sequences obtained with VMir and miRNAFold), which uses a machine learning algorithm LibSVM,

[1] Our GitHub repository

pre-trained on an annotated dataset. As a result, we classified pre-miRNAs into *real* and *fake*. We have obtained 12 *real* pre-miRNAs in miRNAFold pipeline and 6 in VMir pipeline.
Next, we have used MatureBayes (Gkirtzou et al., 2010) to predict mature miRNAs from pre-miRNAs classified as *real*.
Finally, we predicted targets for obtained miRNA using miRDB tool (Chen and Wang, 2019). As authors suggest in the tool's website FAQ section, for each miRNA we have kept targets with score > 80.
Next, we analyzed the obtained target genes across 3 directions:

- ran gene ontology analysis for targets of each miRNA to see if we obtain some COVID-related biological processes;
- checked whether there is an existing transcriptomic proof for obtained target genes to be down-regulated in COVID-infected cells.
- looked for overlap of obtained miRNAs with existing ones to see whether the existing miRNAs are known to participate in some diseases;

We will discuss the results of these directions in the next section.

4 Results

4.1 Gene Ontology

For each of the mature miRNAs from miRNAFold pipeline we performed enrichment analysis on its targets with miRDB score > 90 using the EnrichR web tool (Chen et al., 2013). We chose EnrichR since it has access to the most recent databases. We retrieved results from the following databases: BioPlanet 2019, WikiPathway 2021 Human, KEGG 2021 Human (pathways); GO Biological Process 2018, GO Molecular Function 2018, GO Cellular Component 2018. We observed enrichment in functions related to *insulin resistance* (consistent with the existing line of research of severity of complications from COVID-19 in patients with diabetes (Santos A, 2021)), protein kinase activity, ribosomal large subunit binding, mRNA 3'-UTR binding and processes related to neurons, to name a few.

4.2 Existing transcriptomic proof

Existing transcriptomic proof would mean that genes that are targeted by the viral miRNAs are found to be empirically down-regulated in COVID-infected cells. This would be the most supportive evidence for the functionality of the predicted v-miRNAs.

To test this, we took the list of 472 predicted genes, targeted by miRNAs with the mRDB cut-off score >95 (stricter than >80, suggested by the authors). We compared this list with the set of 1889 significantly down-regulated genes (Wald test, FDR-corrected p-value < 0.05) from 19 cell types, provided by Delorey et al. (2021).

The overlap between the two sets was 31 genes. Most of them seem not highly relevant to COVID-related complications. However, the one most commonly down-regulated of these genes (in 12 out of 19 cell types, including lung tissue) was ZFP36. Considering that ZFP36 is most highly expressed in lungs[2], it is reasonable to assume that v-miRNAs produced by SARS-CoV-2 may be knocking out this gene in the lung tissue.

ZFP36 regulates the production of inflammatory proteins called cytokines, and its deregulation could lead to a large increase in their levels (Jin et al., 2012; Caracciolo et al., 2018). One of the main causes for lethal outcomes in COVID patients is a syndrome known as the "cytokine storm", or the "cytokine release syndrom", when too many cytokines get produced, over-activating the immune system which starts attacking body's own cells. Thus, one can hypothesize that v-miRNAs produced by SARS-CoV-2 may be knocking out ZFP36 in lungs, which makes the immune system become overly agressive and start attacking its own already virus-burdened lung tissue. This results in the acute respiratory distress syndrome (ARDS), which leads to low oxygen levels in the body and often results in death.

4.3 Existing miRNAs overlap

At this stage, we wanted to check if any of our predicted-to-be mature miRNAs from our two pipelines overlap with already known miRNAs, and if so, check if these miRNAs participate in diseases. To do this, we took mature miRNAs that we got from MatureBayes and checked them for overlap with known miRNAs using miRBase, a tool developed by Kozomara et al. (2018). We found that miRNAs from miRNAFold pipeline overlapped with 17 existing miRNAs , and miRNAs from VMir pipeline overlapped with 11 existing miR-

[2] Tissue expresesion of ZFP36

NAs. Next, we used HMDD (Huang et al., 2018) and miRNASNP-v3 (Liu et al., 2020) to check if any of these miRNAs cause or play a role in human diseases. Full results for this stage are available in a Jupyter notebook [3] . In brief, we observed these miRNAs' participation in different cancer types, such as glioma, carcinoma, melanoma, etc. For one of them, namely bladder cancer, it is argued that miR-576-3p plays a causal role in the development of the disease through targeting the gene cyclin D1. Another miRNA, hsa-mir-1272, seems to be playing a role in the development of lung carcinoma, which is in a way consistent with reports of COVID-19 targeting respiratory system.

5 Discussion

Our pipeline consisted of several tools and approaches at every stage. Filtering and interpreting the results of this pipeline, we were able to answer the initial question, forming a list of predicted viral miRNAs produced by SARS-CoV-2, and providing indirect evidence of functionality of some of these miRNAs in three separate ways: through interpreting the gene ontology of predicted targets, through existing transcriptomic proof, and by showing the overlap with known miRNAs that participate in other diseases.

There were three main challenges along the way: finding working tools, choosing modes, thresholds and cutoff values for these tools, and interpreting the results.

When studying similar research, we discovered in total more than twenty tools for different stages of miRNA prediction and analysis. Despite the high number, the field is moving very rapidly, and so many of these tools were no longer supported or available, and we had to look for newer tools, or use fewer tools at certain stages.

Once we had working tools for a stage in the pipeline, the question that came up every time was how to choose cutoff values for the tools that determine which results to use further. For example, MatureBayes outputs predictions of two types: a duplex of miRNA predictions using both stems, and single top-scoring prediction for each stem separately, which can result in 2 to 3 unique mature miRNA predictions. This required thinking of which ones to keep, and how to justify this decision. In other case, using mRDB for target prediction with default cutoff score (>80, as suggested by the authors) gave us in total more than 8000 unique genes. As explained by our supervisor, this was too many to be realistic, so we had to to reduce this number by choosing a stricter cutoff score or keeping only those miRNAs that have a lower number of target predictions. Again, we had to make decisions in cases for which there are no established guidelines, and yet keep the results sensible.

Finally, after all stages of filtering, we had to assess the functionality of predicted miRNAs by interpreting the results. This was challenging because none of our team members had a strong background in biology at the start of the project, and we had to read a lot of literature to understand how things work. Even then, we would not be able to produce a comprehensive analysis without the help of our supervisor.

In general, we enjoyed the project and did not come up with things we would do differently.

6 Author contribution

Authors contributed equally to the project. Initial identification of hairpin structures in the reference genome, target prediction and running gene ontology were done by Tetiana. Filtering of hairpin structure predictions and analyzing predicted miRNA overlap with known miRNAs were done by Farid. Pre-processing output from hairpin identification tools, predicting mature miRNAs, and searching for existing transcriptomic proof were done by Mykyta.

7 Acknowledgments

The authors are grateful to Erik Abner for the guidance along the way and tremendous help with interpreting the results. Authors are also grateful for the financial support by the European Regional Development Fund and the University of Tartu.

[3] Overlap with existing miRNAs - Jupyter notebook

8 Supplementary material

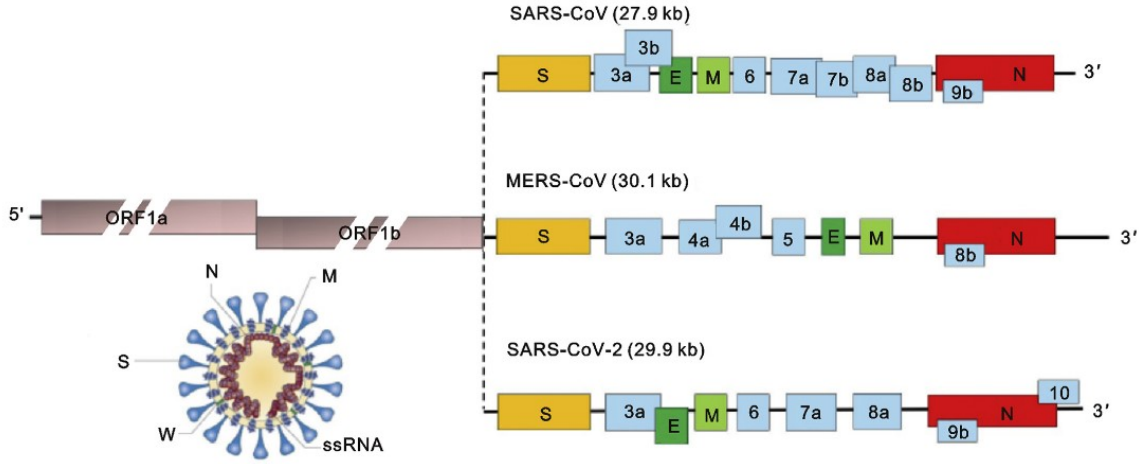

Figure 1: Coronavirus genomes (Li et al., 2020)

	Pre-miRNA	Mutation ID	Position	Ref/Alt	Disease	Region	Mature miRNA
0	hsa-mir-1272	COSN1081476	chr15:64762412	G/A	endometrium;carcinoma	pre-miRNA	-
1	hsa-mir-1272	COSN1081475	chr15:64762480	T/C	endometrium;carcinoma	Mature	hsa-miR-1272
2	hsa-mir-1272	COSN19395611	chr15:64762483	C/T	prostate;carcinoma	Mature	hsa-miR-1272
3	hsa-mir-1272	COSN26991829	chr15:64762497	G/A	large intestine;carcinoma(PMID:27149842)	pre-miRNA	-
4	hsa-mir-1272	COSN4710431	chr15:64762497	G/A	endometrium;carcinoma	pre-miRNA	-
5	hsa-mir-1272	COSN2378591	chr15:64762506	T/C	lung;carcinoma	pre-miRNA	-

Figure 2: Diseases overlap from one of the pipelines

References

A. M. Ardekani and M. M. Naeini. 2010. The Role of MicroRNAs in Human Diseases. *Avicenna J Med Biotechnol*, 2(4):161–179.

Valentina Caracciolo, Jeanette Young, Donna Gonzales, Yingchun Ni, Stephen J. Flowers, Ross Summer, Scott A. Waldman, Jason K. Kim, Dae Young Jung, Hye Lim Noh, Taekyoon Kim, Perry J. Blackshear, Danielle O'Connell, Robert C. Bauer, and Caleb B. Kallen. 2018. Myeloid-specific deletion of zfp36 protects against insulin resistance and fatty liver in diet-induced obese mice. *American Journal of Physiology-Endocrinology and Metabolism*, 315(4):E676–E693. PMID: 29509432.

Edward Chen, Christopher Tan, Yan Kou, Qiaonan Duan, Zichen Wang, Gabriela Meirelles, Neil Clark, and Avi Ma'ayan. 2013. Enrichr: Interactive and collaborative html5 gene list enrichment analysis tool. *BMC bioinformatics*, 14:128.

Junjie Chen, Xiaolong Wang, and Bin Liu. 2016. imirna-ssf: Improving the identification of microrna precursors by combining negative sets with different distributions. *Scientific Reports*, 6:19062.

Yuhao Chen and Xiaowei Wang. 2019. miRDB: an online database for prediction of functional microRNA targets. *Nucleic Acids Research*, 48(D1):D127–D131.

Bryan R Cullen. 2013. MicroRNAs as mediators of viral evasion of the immune system. *Nature Immunology*, 14(3):205–210.

Toni M. Delorey, Carly G. K. Ziegler, Graham Heimberg, Rachelly Normand, Yiming Yang, Asa Segerstolpe, Domenic Abbondanza, Stephen J. Fleming, Ayshwarya Subramanian, Daniel T. Montoro, Karthik A. Jagadeesh, Kushal K. Dey, Pritha Sen, Michal Slyper, Yered H. Pita-Juárez, Devan Phillips, Zohar Bloom-Ackerman, Nick Barkas, Andrea Ganna, James Gomez, Erica Normandin, Pourya Naderi, Yury V. Popov, Siddharth S. Raju, Sebastian Niezen, Linus T.-Y. Tsai, Katherine J. Siddle, Malika Sud, Victoria M. Tran, Shamsudheen K. Vellarikkal, Liat Amir-Zilberstein, Deepak S. Atri, Joseph Beechem, Olga R. Brook, Jonathan Chen, Prajan Divakar, Phylicia Dorceus, Jesse M. Engreitz, Adam Essene, Donna M. Fitzgerald, Robin Fropf, Steven Gazal, Joshua Gould, John Grzyb, Tyler Harvey, Jonathan Hecht, Tyler Hether, Judit Jane-Valbuena, Michael Leney-Greene, Hui Ma, Cristin McCabe, Daniel E. McLoughlin, Eric M. Miller, Christoph Muus, Mari Niemi, Robert Padera, Liuliu Pan, Deepti Pant, Carmel Pe'er, Jenna Pfiffner-Borges, Christopher J. Pinto, Jacob Plaisted, Jason Reeves, Marty Ross, Melissa Rudy, Erroll H. Rueckert, Michelle Siciliano, Alexander Sturm, Ellen Todres, Avinash Waghray, Sarah Warren, Shuting Zhang, Daniel R. Zollinger, Lisa Cosimi, Rajat M. Gupta, Nir Hacohen, Winston Hide, Alkes L. Price, Jayaraj Rajagopal, Purushothama Rao Tata, Stefan Riedel, Gyongyi Szabo, Timothy L. Tickle, Deborah Hung, Pardis C. Sabeti, Richard Novak, Robert Rogers, Donald E. Ingber, Z. Gordon Jiang, Dejan Juric, Mehrtash Babadi, Samouil L. Farhi, James R. Stone, Ioannis S. Vlachos, Isaac H. Solomon, Orr Ashenberg, Caroline B.M. Porter, Bo Li, Alex K. Shalek, Alexandra-Chloé Villani, Orit Rozenblatt-Rosen, and Aviv Regev. 2021. A single-cell and spatial atlas of autopsy tissues reveals pathology and cellular targets of sars-cov-2. *bioRxiv*.

Katerina Gkirtzou, Ioannis Tsamardinos, Panagiotis Tsakalides, and Panayiota Poirazi. 2010. Maturebayes: A probabilistic algorithm for identifying the mature mirna within novel precursors. *PloS one*, 5:e11843.

Lisa E. Gralinski and Vineet D. Menachery. 2020. Return of the coronavirus: 2019-nCoV. *Viruses*, 12(2):135.

Adam Grundhoff. 2011. Computational prediction of viral mirnas. *Methods in molecular biology (Clifton, N.J.)*, 721:143–52.

Adam Grundhoff, CS Sullivan, and D Ganem. 2006. A combined computational and microarray-based approach identifies novel micrornas encoded by human gamma-herpesviruses. *RNA*, 12:733–750.

Zhou Huang, Jiangcheng Shi, Yuanxu Gao, Chunmei Cui, Shan Zhang, Jianwei Li, Yuan Zhou, and Qinghua Cui. 2018. HMDD v3.0: a database for experimentally supported human microRNA–disease associations. *Nucleic Acids Research*, 47(D1):D1013–D1017.

Wen-Jing Jin, Cai-Feng Chen, Hui-Yu Liao, Lu-Lu Gong, Xiao-Hui Yuan, Bin-Bin Zhao, Ding Zhang, Xia Feng, Jing-Jun Liu, Yu Wang, Guo-Feng Chen, Hui-Ping Yan, and You-Wen He. 2012. Downregulation of the au-rich rna-binding protein zfp36 in chronic hbv patients: Implications for anti-inflammatory therapy. *PLOS ONE*, 7(3):1–11.

Ana Kozomara, Maria Birgaoanu, and Sam Griffiths-Jones. 2018. miRBase: from microRNA sequences to function. *Nucleic Acids Research*, 47(D1):D155–D162.

Xiaowei Li, Manman Geng, Yizhao Peng, Liesu Meng, and Shemin Lu. 2020. Molecular immune pathogenesis and diagnosis of covid-19. *Journal of Pharmaceutical Analysis*, 10(2):102–108.

Chun-Jie Liu, Xin Fu, Mengxuan Xia, Qiong Zhang, Zhifeng Gu, and An-Yuan Guo. 2020. miRNASNP-v3: a comprehensive database for SNPs and disease-related variations in miRNAs and miRNA targets. *Nucleic Acids Research*, 49(D1):D1276–D1281.

Ronny Lorenz, Stephan Bernhart, Christian Höner zu Siederdissen, Hakim Tafer, Christoph Flamm, Peter Stadler, and Ivo Hofacker. 2011. Viennarna package 2.0. *Algorithms for molecular biology : AMB*, 6:26.

Andrew McGuire, James A. L. Brown, and Michael J. Kerin. 2015. Metastatic breast cancer: the potential of miRNA for diagnosis and treatment monitoring. *Cancer and Metastasis Reviews*, 34(1):145–155.

Evangelista-Poderoso R Saad MJA Santos A, Magro DO. 2021. Diabetes, obesity, and insulin resistance in covid-19: molecular interrelationship and therapeutic implications. *Diabetology metabolic syndrome*.

Christophe Tav, Sébastien Tempel, Laurent Poligny, and Fariza Tahi. 2016. miRNAFold: a web server for fast miRNA precursor prediction in genomes. *Nucleic Acids Research*, 44(W1):W181–W184.

Sonia Verma, Abhisek Dwivedy, Neeraj Kumar, and Bichitra K. Biswal. 2020. Computational prediction of sars-cov-2 encoded mirnas and their putative host targets. *bioRxiv*.

Patrick C. Y. Woo, Susanna K. P. Lau, Yi Huang, and Kwok-Yung Yuen. 2009. Coronavirus diversity, phylogeny and interspecies jumping. *Experimental Biology and Medicine*, 234(10):1117–1127.

Coronavirus Cases Worldometer. 2020. https://www.worldometers.info/coronavirus/.

Baohong Zhang, Xiaoping Pan, Stephen Cox, George Cobb, and Todd Anderson. 2006. Evidence that mirnas are different from other rnas. *Cellular and molecular life sciences : CMLS*, 63:246–54.