

Master thesis in Bioinformatics
Bioinformatics Research Center (BIRC)
Molekylær Medicinsk Afdeling (MOMA)
Århus University

Investigating regulatory motif perturbation during SARS-CoV-2 infection in a single-cell setup



Paula Rodrigo Martín

Supervisor: Jakob Skou Pedersen

June 2024

Contents

1	Introduction	6
1.1	Post-transcriptional regulation	6
1.1.1	miRNA	7
1.1.2	RBPs	8
1.2	SARS-CoV-2 virus	8
1.2.1	SARS-CoV-2 life cycle	8
1.2.2	Immune response to SARS-CoV-2	9
2	Current evidence on post-transcriptomic deregulation during SARS-CoV-2	11
2.1	Proposed hypotheses for this study based on current evidence	13
3	Data	15
3.1	SARS-CoV-2 (single-cell transcriptome) atlas	15
3.1.1	Sample type	16
3.1.2	Processing of scRNAseq data (done by Ren et al. [1])	16
3.1.3	Detection of viral RNA (done by Ren et al. [1])	16
3.2	Cohort used for this study	17
4	Methods	20
4.1	Workflow	20
4.2	miReact	21

4.3	miRNA annotations	24
4.4	RBP annotations	24
4.5	Gene ontology annotations	24
4.6	SARS-CoV-2 genome	25
4.7	Differential activity of motifs between infected and uninfected conditions	25
4.8	Correlation with viral load	26
4.9	Motif ranking	26
4.9.1	Hypothesis 1: host miRNA deregulation by infection	27
4.9.2	Hypothesis 2: v-miRNA impacts gene expression	27
4.9.3	Hypothesis 3: competition of host miRNAs with endogenous viral RNA molecules	27
4.10	Assembly of longer kmers after miReact analysis	27
4.10.1	Overlapping selected kmers into longer sequences [Figure 4.3 step 2]	27
4.10.2	Position Frequency Matrix	28
4.10.3	Regular Expressions	28
4.10.4	Filtering steps	29
5	Results	32
5.1	Inferred miRNA activity	32
5.2	Summary statistics of annotated variables in epithelial cells	32
5.2.1	Differential activity analysis	32
5.2.2	Correlation with viral load	33
5.2.3	Motif hits in viral genome	33
5.3	Ranked lists for each hypothesis	33
5.4	Hypothesis 1: host miRNA deregulation	37
5.4.1	GCGCGGT motif	37
5.4.2	miR-4684-5p: downregulated in epithelial cells and upregulated in immune cells	37

5.5 Hypothesis 2: v-miRNA	39
5.5.1 potential v-miRNA analogous to human miR-20b-3p	39
5.6 Hypothesis 3: competition between host mRNAs and viral endogenous RNA for one miRNA	40
5.6.1 miR-640 shows evidence of sponging by the viral genome	40
5.7 Comparison to available findings from Ren et al.	40
5.8 <i>longermer</i> pipeline: VHL complex expression is disregulated in squamous cells	42
5.9 RBPs	45
6 Discussion	46
7 Conclusions	50
7.1 Code availability	51

ABSTRACT

Post-transcriptional regulation is a cellular mechanism for gene expression control that constitutes a complex network of many factors. SARS-CoV-2 is known to take over the host's transcriptional and translational machinery leading to perturbations in normal gene expression. We therefore ask whether post-transcriptional regulators are also perturbed under SARS-CoV-2 infection. The overall study objective is to investigate the perturbation of microRNAs (miRNAs) and RNA binding proteins (RBPs) during SARS-CoV-2 infection. For that purpose, we analyze a single-cell atlas, comprising approximately 1.4 million healthy and SARS-CoV-2 infected cells. We aim to identify regulatory molecules with known target motifs that substantially up- or downregulate target genes. By using the miReact pipeline, we systematically screen all 7-mer motifs found in 3'UTRs for significant association with target gene perturbation during infection. We have developed a framework that, through systematic annotation with signals indicative of biological function, generates catalogs of motifs that can be ranked by biological interest and hence allow hypothesis generation and identification of candidates for further study.

This approach identifies mechanisms in epithelial cells that are related to viral infection. We found the miR-20b-3p activity to be perturbed during SARS-CoV-2 infection of epithelial cells. We hypothesize the presence of an analogous viral miRNA (v-miRNA) to miR-20b-3p to be targetting genes related to transcription. miR-640 shows potential sponging by the viral genome and we discover a 13 nucleotide long motif associated with Von-Hippel-Lindau (VHL) complex deregulation in squamous cells. Our analysis points to significant post-transcriptional deregulation of specific miRNAs during SARS-CoV-2 infection. The results show-case the usefulness of expression-coupled motif analysis in detecting differential miRNA activities, which are otherwise unobserved.

Keywords Post-transcriptomic deregulation · SARS-CoV-2 · miR-20b-3p · motif · miR-640 · miReact · longermer