Title: Comprehensive analysis of human SARS-CoV-2 infection and host-virus interaction
https://www.iscb.org/ismb2020-submit/covid-19 -- Submitted on May 14th!

Mariana G. Ferrarini[1*], Avantika Lal[2*], Rita Rebollo[1], Andreas Gruber[3], Andrea Guarracino[4], Itziar Martinez Gonzalez[5], Taylor Floyd[6], Daniel Siqueira de Oliveira[7], Taneli Pusa[8], Alexander Kanitz[9], Brett E. Pickett[10#], Vanessa Aguiar-Pulido[6#]

[*] These authors contributed equally
[#] Corresponding authors

[1] University of Lyon, INSA-Lyon, INRA, BF2I, Villeurbanne, France
[2] NVIDIA Corporation, Santa Clara, CA, USA
[3] Oxford Big Data Institute, Nuffield Department of Medicine, University of Oxford, Oxford, UK
[4] Centre for Molecular Bioinformatics, Department of Biology, University Of Rome Tor Vergata, Rome, Italy
[5] Amsterdam UMC, Amsterdam, The Netherlands
[6] Center for Neurogenetics, Weill Cornell Medicine, Cornell University, New York, NY, USA
[7] Laboratoire de Biométrie et Biologie Evolutive, Université de Lyon; Université Lyon 1; CNRS; UMR 5558, Villeurbanne, France
[8] Swiss Institute of Bioinformatics, ELIXIR Switzerland
[9] Swiss Institute of Bioinformatics, ELIXIR Switzerland
[10] Brigham Young University, Provo, UT, USA

Abstract
In December of 2019 a novel betacoronavirus named severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) emerged in Wuhan, China. This virus causes the COVID-19 disease and by May 14th, it had already infected more than four million people worldwide, accounting for 300 thousand deaths. We have performed a comprehensive gene, transcript and transposable element differential expression analysis based on the available dataset of lung cells infected with SARS-CoV-2; identified regulatory motifs that could partially explain these genome-scale expression changes upon virus infection; and predicted putative interaction sites between the viral RNA and human RNA binding proteins, which may play essential roles in regulating viral transcription, replication, and translation. We detected genes involved in general viral response, as well as specific SARS-CoV-2 deregulated genes. Many of the genes identified in this work are interesting and worthy of additional functional analysis. We suggest new avenues for research into the differential susceptibility of humans to COVID-19, and novel insights on the virulence of SARS-CoV-2, which will be helpful to the scientific community to fight this disease in the near future.

Keywords
SARS-CoV-2, COVID-19, gene expression, RNA-seq, RNA-binding proteins, host-pathogen interaction, transcriptomics

Introduction

In December of 2019 a novel betacoronavirus named severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) emerged in Wuhan, China (Huang *et al*., 2020, PMID:31986264). This virus is responsible for causing the COVID-19 disease and by May 14th, it had already infected more than four million people worldwide, accounting for 300 thousand deaths (World Health Organization). Ever since its abrupt emergence, the scientific community has incessantly tried to better understand this virus, with studies ranging from molecular mechanisms, to computational drug discovery and repositioning. The need for an effective treatment for COVID-19 is urgent and, as a result, recent studies have sought to detect proteins or drugs that interact directly with the SARS-CoV-2 viral proteins (Gordon *et al*., 2020, PMID: 32353859; Calligari *et al*. 2020, PMID: 32295237). In this work, we present a comprehensive analysis on human SARS-CoV-2 infection from publicly available datasets along with predictions of host-virus interactions based on available SARS-CoV-2 genomic sequences.

Methods

Viruses are known to trigger a specific but rather drastic transcriptomic response to their infection. These changes in gene expression may be related to viral success and can shed light on potential drug targets. We have performed a general gene and transcript differential expression analysis based on the available dataset of lung cells (NHBE, A549 and Calu-3) infected with SARS-CoV-2, Influenza A virus (IAV), human parainfluenza virus type 3 (HPIV3), and respiratory syncytial virus (RSV) (Blanco-Melo *et al*., 2020, DOI: 10.1016/j.cell.2020.04.026). The list of differentially expressed genes (DEGs) was used as input to pathway and gene ontology (GO) enrichment analyses. In order to detect possible metabolism-related functions, we integrated the RNAseq data with the available human metabolic network (Thiele *et al*., 2013, PMID: 23455439; Pusa *et al*., 2019, PMID: 31504164). Furthermore, transposable element (TE) upregulation has been frequently observed upon viral infection (Machietto *et al*., 2020, PMID: 31964680) and was henceforth studied. To identify regulatory motifs that could partially explain the genome-scale expression changes upon virus infection, we performed a motif activity response analysis (Balwierz et al., 2014, PMID: 24515121). Since coronavirus genomes were described to bind and be regulated by human RNA-binding proteins (RBPs) (Shi & Lai, 2005, PMID: 15609510), we analyzed over 7000 available SARS-CoV-2 genome sequences (from GISAID, Elbe & Buckland-Merrett, 2017, PMID: 31565258) to predict putative interaction sites with human RBPs, and used the gene expression analysis to identify bidirectional regulation between human RBPs and SARS-CoV-2.

Results

SARS-CoV-2 elicited a major activation of transcriptional response in human cells with a total of 229 genes upregulated and only 14 genes downregulated in NHBE cells, with a multiplicity of infection (MOI) of 2. The response of A549 immortalized cells was milder compared to primary cells when the MOI used was ten fold smaller (MOI 0.2: 94 upregulated and 15 downregulated). The infection deregulated a larger number of genes in both A549 and Calu-3 cells when the same MOI from NHBE infection was used. The transcriptional response from SARS-CoV-2 infection was similar to a general viral response common to the other viruses tested, in accordance to what was reported in the original paper (Blanco-Melo *et al*., 2020, DOI:

10.1016/j.cell.2020.04.026). We detected a general signature induced by all three viruses which included *CSF3*, *IL6*, several chemokines and interferon-induced genes, which activate the innate immune system to clear the viruses (Newton et al., 2016). We also looked for genes upregulated uniquely in SARS-CoV-2 infected cells and detected the induction of alarmins *S100A7* and *S100A8*, chemokines *CXCL3* and *CXCL5*, and proinflammatory cytokines that further activate the innate immune cells. *CSF2*, also known as Granulocyte Macrophage-Colony Stimulating Factor (*GM-CSF*) was one of the top up-regulated genes specific to SARS-CoV-2 infection. This particular cytokine promotes differentiation, recruitment and activation of neutrophils and macrophages (Becher *et al.*, 2016, PMID: 27851925). The overactivation of macrophages and neutrophils can lead to what is known as cytokine storm which can lead to acute lung injury and acute respiratory distress syndrome (Confalonieri *et al*., 2017, PMID: 28446599). Although the presence of GM-CSF protein has been shown to be protective against influenza A virus infection in animal models (Huang *et al*., 2011, PMID:21474645), it is possible that the extremely high levels of *GM-CSF* expression during SARS-CoV-2 infection contribute to this cytokine storm. Furthermore, the cytokine storm has been proposed as the underlying cause for the fatal outcome in severe COVID-19 cases.

While SARS-CoV-2 triggers a transcriptional response in the human genome, it has been shown that RNA viruses can bind to host RBPs (Shi & Lai, 2005, PMID: 15609510; Barnhart *et al.*, 2013, PMID: 24210824), resulting in two possible outcomes: (i) some of these interactions may influence SARS-CoV-2 replication, transcription or translation; and (ii) the interaction of human RBPs with viral RNA deviates their availability for human mRNAs, resulting in deregulation of human gene expression. Regarding the identification of genome-scale regulators upon SARS-CoV-2 infection, we detected the RELA and interferon-regulatory factors (IRFs) among the motifs that were most significantly upregulated in activity in both NHBE and A549 cells. Activation of these regulators result in a genome-wide activation of interferon-stimulated genes (ISGs), which are known to restrict viral replication (Chiang & Liu, 2018, PMID: 30671058; Wang *et al*., 2010, PMID: 20610653). Additionally, we predicted 38 human RBPs as binding partners of SARS-CoV-2 RNA. Interestingly, two of these proteins (PABPC1, PABPC4) have already been experimentally shown to interact with the SARS-CoV-2 N protein (Gordon *et al*., 2020, PMID: 32353859). Other interesting candidates include proteins that were previously described to interact with other RNA viruses, such as hnRNPA1 and hnRNP-L Moreover, we detected significant differential expression changes in 4 of the 38 predicted RBPs upon SARS-CoV-2 infection, suggesting a possible mutual regulation between the virus and the human cell. Finally, we identified differences in the enrichment of binding sites for human RBPs in the SARS-CoV-2 3'UTR sequence when compared to that of related coronaviruses. Discovering the roles of these SARS-CoV-2 specific proteins may help understand the remarkable pathogenicity of this virus.

Conclusions

Many of the genes identified in this work are interesting and worthy of additional experiments in the wet lab in order to validate the analyses and predictions presented here. We suggest new avenues for research into the differential susceptibility of humans to COVID-19, and novel insights on the virulence of SARS-CoV-2, which will be helpful to the scientific community to fight this disease in the near future.