Comprehensive analysis of human SARS-CoV-2 infection and host-virus interaction

Mariana Galvao 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 Biom trie et Biologie Evolutive, Universit de Lyon; Universit Lyon 1; CNRS; UMR 5558, Villeurbanne, France
8 Luxembourg Centre for Systems Biomedicine, Belvaux, Luxembourg
9 Swiss Institute of Bioinformatics, ELIXIR Switzerland
10 Brigham Young University, Provo, UT, USA

C46

Introduction and Objectives

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 the etiological agent of COVID-19 disease. This pathogen had already infected more than 10 million people worldwide and caused over 500,000 deaths by June 29th (World Health Organization). Ever since its abrupt emergence, the scientific community has incessantly tried to gain a better understanding of this virus, with studies ranging from laboratory-based mechanistic approaches, 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 identify host proteins or drugs that interact directly with SARS-CoV-2 proteins (Gordon et al., 2020, PMID: 32353859; Calligari et al. 2020, PMID: 32295237). In this work, we present a comprehensive analysis on the human transcriptomic response during SARS-CoV-2 infection from publicly available datasets along with predictions of host-virus interactions based on available SARS-CoV-2 genomic sequences (**Fig. 1**).

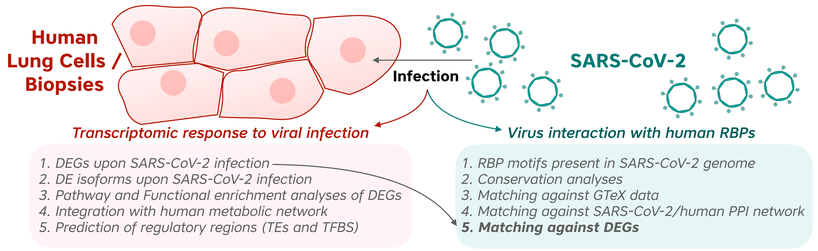


Figure 1. Overview of the approaches and methods used in this study and the two main questions we addressed (i) how does the human host respond to the viral infection and (ii) how does the virus directly manipulate the host. In the first question we performed a general gene and transcript differential expression analysis based on available datasets 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) along with lung biopsies infected with SARS-CoV-2 (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 also studied. In the second question we identified human RNA-binding proteins (RBPs) capable of binding to the viral genome and regulating downstream processes (Shi & Lai, 2005, PMID: 15609510). To do so, we analyzed around 25,000 available SARS-CoV-2 genome sequences (from GISAID, Elbe & Buckland-Merrett, 2017, PMID: 31565258) to predict conserved putative interaction sites with human RBPs, and used the gene expression analysis to identify bidirectional regulation between human RBPs and SARS-CoV-2.

Results

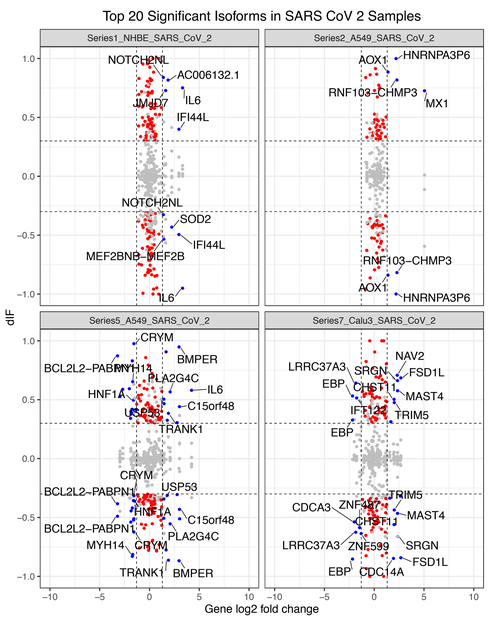


Figure 2. SARS-CoV-2 samples exhibit significant changes in gene expression and isoform usage. Top expressing genes are associated with biological functions such as the anti-viral immune response, RNA splicing and protein processing. Horizontal lines indicate a 30% or greater change in isoform usage between conditions (dIF = |0.3|). dIF = differential isoform fraction. FDR = false discovery rate (p-value).

Transcriptomic response to SARS-CoV-2 infection

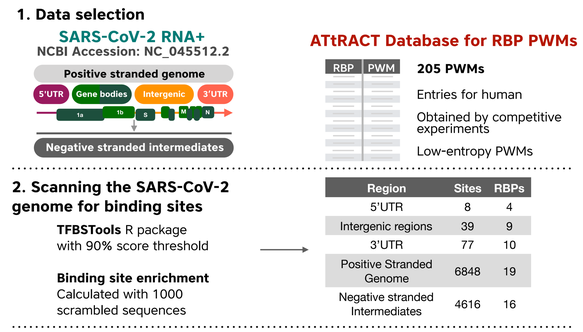
From the transcriptomic analyses, we tried to pinpoint genes specific to SARS-CoV-2 infection.

- We detected 33 up- and 8 down-regulated genes specifically differentially expressed (DE) in 3 out of 4 SARS-CoV-2 datasets.
- Isoform switch analysis detected interesting isoform differential usage such as IL-6 and NHRNPA3P6 (**Fig. 2**).
- Transcriptomic data integration with human metabolic network detected common decreased fluxes in inositol phosphate metabolism (PIP3). In A549 cells, we also predicted decreased fluxes in several lipid related pathways: fatty acid, cholesterol, sphingolipid, and glycerophospholipid.
- Signaling pathway enrichment revealed that type I interferon signaling was inhibited in both A549 and Calu-3 cells at MOIs of 0.2 and 2, respectively.
- We detected 16 DE transposable element (TE) families upregulated specifically in SARS-CoV-2 infected lung cells (**Fig. 3**).

SARS-CoV-2 interaction with human RNA bidning proteins (RBPs)

Although SARS-CoV-2 infection triggers a transcriptional response in human cells, it has been shown that viral RNA can associate with host human binding proteins 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 affects their availability for human mRNAs, resulting in dysregulation of human gene expression. The overview of these analyses and results are depicted in **Fig. 4** and **Table 1**.

Selected and SARS-CoV-2 specific genes from all analyses are depicted in **Fig. 5**.



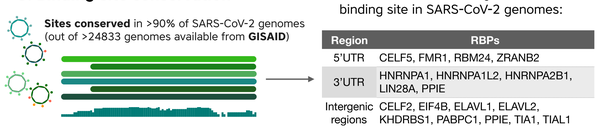
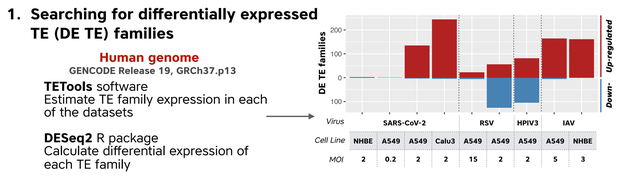
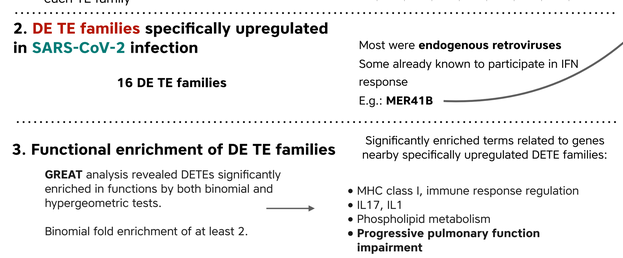


Figure 4. General pipeline for RBP putative binding site search within SARS-CoV-2 genome. We searched the SARS-CoV-2 genome for known RBP binding motifs detecting whether these hits were enriched inside specific viral regions: 5'UTR, 3'UTR, intergenic, positive genome or negative stranded intermediates. Once we had a list of enriched RBPs, we looked at the conservation in these sites on ~25k SARS-CoV-2 sequences available in GISAID. This analysis resulted in 18 human RBPs whose binding motifs are enriched and conserved within SARS-CoV-2 genomes. Interestingly, all proteins predicted to interact with 5'UTR are associated with viral RNA stability in the context of other viral infections.





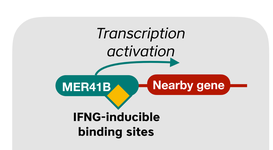


Figure 3. Overview of transposable element analyses. We first used TEtools (Lerat et al., 2017, PMID:28204592) to detect differential expression of families. TE expression can be cell line specific, but to detect SARS-CoV-2 related TE activation, we selected upregulated families in Calu-3 and A549 cell lines infected with SARS-CoV-2, and excluded all DE TE families detected in A549 infected with other viruses. Functional enrichment analysis was performed with GREAT (McLean et al., 2010, PMID: 20436461).

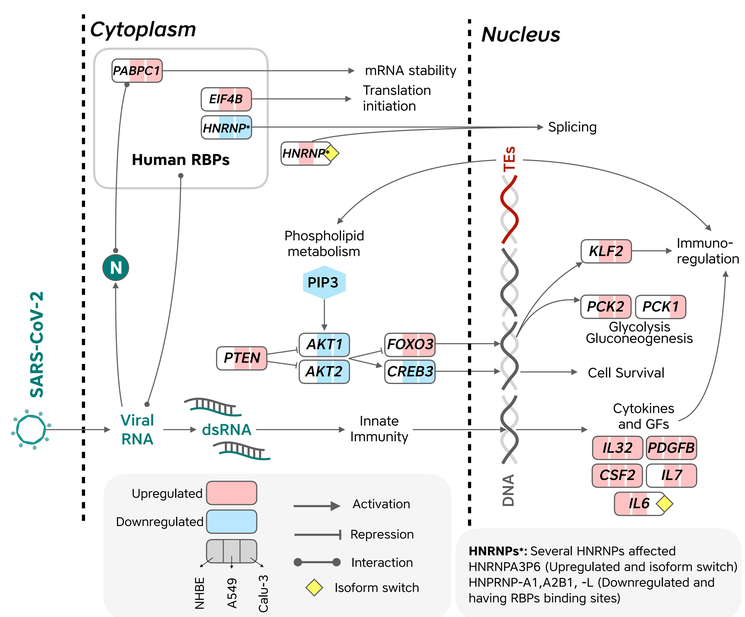


Figure 5. Schematic representation of possible specific SARS-CoV-2 mechanisms and host cellular response to infection. In this work we tried to pinpoint genes and proteins regulated specifically in the context of SARS-CoV-2 when compared to the other viruses tested. The green circle represents the **N protein** from the viral genome. The blue hexagon **PIP3** represents the metabolite phosphatidyl inositol triphosphate, which was predicted to be less available in the metabolic network integration. Even though many other proteins in each of the pathways are impacted, we show here only those that were not dysregulated by other viruses in the datasets tested.

RPB	DE Analysis: A549 LogFC	DE Analysis: Calu-3 LogFC	DE Analysis: SARS-CoV-2 Specific DEG	Expression in GTex Datasets: Lung Tissue (TPM)	Expression in GTex Datasets: scRNA*1	PPI Map*2: Viral Protein	RBP binding site prediction: Type	RBP binding site prediction: Conserved*3
PABPC1	0.72	0.44	✓	448.025	✓	N	Positive Strand	✓
PABPC4	0.30	-0.28		103.082	✓	N		
HNRNPL	-0.97	-0.30	✓	208.765	✓			
PABPC1	0.72	0.44	✓	448.025	✓	N	3'UTR	✓
PABPC4	0.30	-0.28		103.082	✓	N		
HNRNPA1	0.00	-0.32		331.336	✓			✓
HNRNPA2B1	-1.08	-0.29		539.829	✓			✓
CELF5	0.56	0.00		0.079	✓		5'UTR	✓
RBM24	0.34	0.00		1.412				✓
PABPC1	0.72	0.44	✓	448.025	✓	N	Intergenic	✓
EIF4B	0.53	0.64		170.303	✓			✓
TIA1	0.34	0.41	✓	46.934	✓			✓
TIA1	0.34	0.41	✓	46.934	✓		Negative Strand	✓
HNRNPL	-0.97	-0.30	✓	208.765	✓			

Table 1. Detailed expression information for selected RBPs. Information reagarding this work (DE Analysis and RBP binding sire prediction) along with literature data (GTex datasets and PPI Map).

*1 scRNA expression in ACE+ and TMPRSS2+ lung cells, dataset GSE12960;

*2 PPI Map : Gordon et al, 2020, PMID 32353859;

*3 Conserved in SARS-CoV-2 genomes.

Conclusions

Many of the genes identified in this work are interesting and worthy of additional experiments to validate the analyses and predictions presented here (**Fig. 3**). 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.

Financial Support

This collaborative team started in the context of the virtual biohackathon for COVID-19 which took place from April 5 to April 11 2020.