“Meta-analysis of SARS COV public data” by Mehdi Ahmadi1**,** James Flynn2, Joe Delaney2, Enrique Torchia1 and Michael Edwards3.

1. Department of Dermatology, University of Colorado Anschutz Medical Campus, Aurora,CO. 2. Illumina Corporation, San Diego, CA. 3. Bioinfo Solutions LLC, Parker, CO

**Background**

To date, little is known about the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) that is at the center of the current global epidemic. (Lai *et al.*, 2020; Shereen *et al.*, 2020). SARS-CoV-2 belongs to a new strain of human viruses that emerged from Wuhan providence in China in December 2019. There are currently no animal studies addressing the host biology of SARS-CoV-2 infection. However, the virus does shares 79% homology to previously identified human infecting coronavirus (e.g., SARS CoV1) including binding to ACE2 receptor on susceptible cells (Lu *et al.*, 2020; Wu *et al.*, 2020). These observations provide the rationale for the mining of previous studies and their public available datasets on coronavirus infection to understand the current viral outbreak.

**Data Collection**

BaseSpace Correlation Engine (BSCE, Illumina Inc.) was used to identify and analyze publicly accessible data from the Gene Expression Omnibus repository (GEO) (www.ncbi.nlm.nih.gov/geo/). Relevant and curated biosets were identified by the correlation engine app using the search term “Coronavirus infection”. We focused on the host gene expression responses during viral infection because of the abundance of available published data and the potential to identify possible therapeutic interventions. We identified 39 biosets (comparisons) from 14 different studies that measured changes in global gene expression in mouse and monkey lungs collected at 1, 4 and 7 days post-infection relative to mock infection controls. A list of biosets used for this analysis can be found in an excel file. Associated experimental and study information can also be found on the second tab of the file. This sheet contains the meta-data for each time point and the first 17 columns on the text file containing the combined meta-dataset.

**Data Processing**

A meta-dataset was created with the BSCE app using a ranked-based gene score for each bioset and time point. The data was then downloaded from the BSCE software and further processed by a python-based algorithm that allowed the conversion of the ranked Illumina score data into positive and negative values based on the fold change. See Cheema et al. (Cheema *et al.*, 2020) and Kupershmidt et al. (Kupershmidt *et al.*, 2010) for further details on data processing and statistical methods used by correlation engine.

**References**

Cheema SK, Isesele PO, Marchando S*, et al.* (2020) The suppression of very long chain fatty acids is associated with skin carcinogenesis. *Journal of Investigative Dermatology*. In press.

Kupershmidt I, Su QJ, Grewal A*, et al.* (2010) Ontology-based meta-analysis of global collections of high-throughput public data. *PLoS One* 5.

Lai C-C, Liu YH, Wang C-Y*, et al.* (2020) Asymptomatic carrier state, acute respiratory disease, and pneumonia due to severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2): Facts and myths. *Journal of Microbiology, Immunology and Infection* <https://doi.org/10.1016/j.jmii.2020.02.012>.

Lu R, Zhao X, Li J*, et al.* (2020) Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *The Lancet* 395:565-74.

Shereen MA, Khan S, Kazmi A*, et al.* (2020) COVID-19 infection: Origin, transmission, and characteristics of human coronaviruses. *Journal of Advanced Research* 24:91-8.

Wu A, Peng Y, Huang B*, et al.* (2020) Genome Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China. *Cell Host & Microbe* 27:325-8.