Genes that might explain why some healthy young people develop severe COVID-19 symptoms

November 1 2021, by Bob Yirka

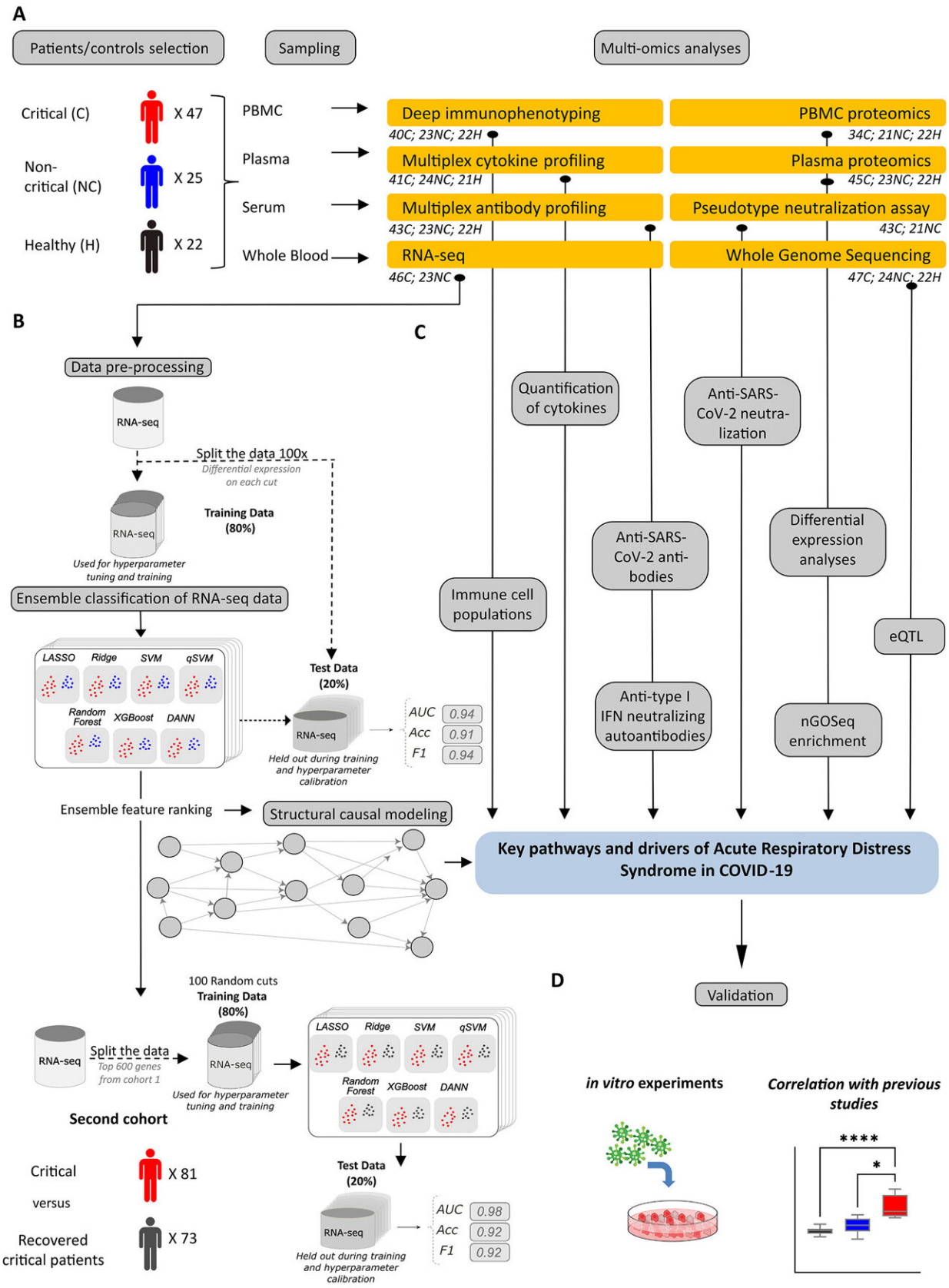
A
Patients/controls selection
Sampling
Multi-omics analyses
Critical (C)
X 47
Non-critical (NC)
X 25
Healthy (H)
X 22
PBMC
Plasma
Serum
Whole Blood
Deep immunophenotyping
40C; 23NC; 22H
Multiplex cytokine profiling
41C; 24NC; 21H
Multiplex antibody profiling
43C; 23NC; 22H
RNA-seq
46C; 23NC
PBMC proteomics
34C; 21NC; 22H
Plasma proteomics
45C; 23NC; 22H
Pseudotype neutralization assay
43C; 21NC
Whole Genome Sequencing
47C; 24NC; 22H
B
Data pre-processing
RNA-seq
Split the data 100x
Differential expression on each cut
Training Data (80%)
Used for hyperparameter tuning and training
Ensemble classification of RNA-seq data
LASSO
Ridge
SVM
qSVM
Random Forest
XGBoost
DANN
Test Data (20%)
Held out during training and hyperparameter calibration
AUC 0.94
Acc 0.91
F1 0.94
Ensemble feature ranking
Structural causal modeling
C
Quantification of cytokines
Anti-SARS-CoV-2 neutra-lization
Anti-SARS-CoV-2 anti-bodies
Differential expression analyses
Immune cell populations
eQTL
Anti-type I IFN neutralizing autoantibodies
nGOSeq enrichment
Key pathways and drivers of Acute Respiratory Distress Syndrome in COVID-19
D
Validation
100 Random cuts
Training Data (80%)
Split the data
Top 600 genes from cohort 1
Second cohort
Critical
versus
Recovered critical patients
X 81
X 73
AUC 0.98
Acc 0.92
F1 0.92
in vitro experiments
Correlation with previous studies
****
*

Fig. 1. A multi-omics analytical strategy identifies key pathways and drivers of Acute Respiratory Distress Syndrome in COVID-19. (A) Forty-seven critical patients (C), 25 non-critical patients (NC) and 22 healthy controls (H) were enrolled in the study. PBMCs were isolated by density gradient and frozen until utilization for mass cytometry and whole proteomics. Plasma was used for cytokine profiling and whole proteomics. Serum was used to measure anti-type I IFN neutralizing antibodies, anti-SARS-CoV-2 neutralizing antibodies and multi-target antiviral serology. Whole blood was used for RNA-seq and whole-genome sequencing (WGS). The number of treated samples per group and per omics is indicated below each omics designation. (B) The RNA-seq pipeline is shown based on the NC versus C comparison. To increase robustness of downstream analyses, an ensemble intelligence approach with seven algorithms was applied to multiple partitions of the RNA-seq data (see Methods) to classify NC versus C patients, performing differential analysis on each partition of the data. An ensemble ranking score across six of the seven algorithms and all partitions of the data was then determined, and the top 600 of those genes were used as the input for structural causal modeling to derive a putative causal network. To support the key findings from the first patient cohort, RNA-seq data from a second patient cohort consisting of 81 critical and 73 recovered critical patients were used. The data was partitioned analogously to the first patient cohort, but only the top 600 features from the first patient cohort were used to assess the informativeness of the gene signature. (C) Cytokines and immune cells were quantified. WGS data were used for eQTL analysis together with the gene counts from the RNA-seq. Proteomics data were subjected to differential protein expression and nGOseq enrichment analyses. (D) The key pathways and drivers resulting from the omics analyses in (B and C) were validated in a second cohort of 81 critical and 73 recovered critical patients. The differential expression of ADAM9, the main driver gene, was compared to publicly available bulk RNA-seq data. Finally, ex vivo infection experiments with SARS-CoV-2 were conducted to validate a driver gene candidate. Credit: DOI: 10.1126/scitranslmed.abj7521

An international team of researchers has isolated five genes that are

more active in young people with severe COVID-19 symptoms than in those with less severe symptoms. In their paper published in the journal *Science Translational Medicine*, the group describes their genetic analysis of young infected people with no known contributing factors for the disease.

Some healthy young people, despite having no apparent underlying conditions, still develop very serious COVID symptoms. Why this happens remains a mystery. In this new effort, the researchers found a possible clue—genes that appear to reduce the body's ability to fight the disease when activated.

The work involved collecting plasma samples from 72 young, hospitalized COVID patients, 47 of whom were critically ill. None of those sampled had any underlying conditions that might explain their symptoms. The team also collected samples from 22 infected young people who had few or no symptoms to serve as a control group. The researchers noted at the outset that those patients who were the most ill also suffered from increased inflammation and coagulation. They then conducted whole-genome sequencing on the samples, along with RNA sequencing, cell proteomics, cytokine profiling and immunophenotyping. They also used a machine-learning application to spot patterns in the genes. Their analysis revealed five genes that were more active in the patients with severe symptoms. One, ADAM9, was found to be the most prevalent.

The team found that when ADAM9 was blocked in human lung tissue in vitro, the virus was less efficient at duplicating itself. The researchers suggest that all five of the genes they identified need to be studied in more depth, particularly ADAM9. Doing so could lead to therapeutics to disable the activity of such genes, and in so doing, reduce symptoms of COVID-19.

More information: Raphael Carapito et al, Identification of driver genes for critical forms of COVID-19 in a deeply phenotyped young patient cohort, *Science Translational Medicine* (2021). DOI: 10.1126/scitranslmed.abj7521