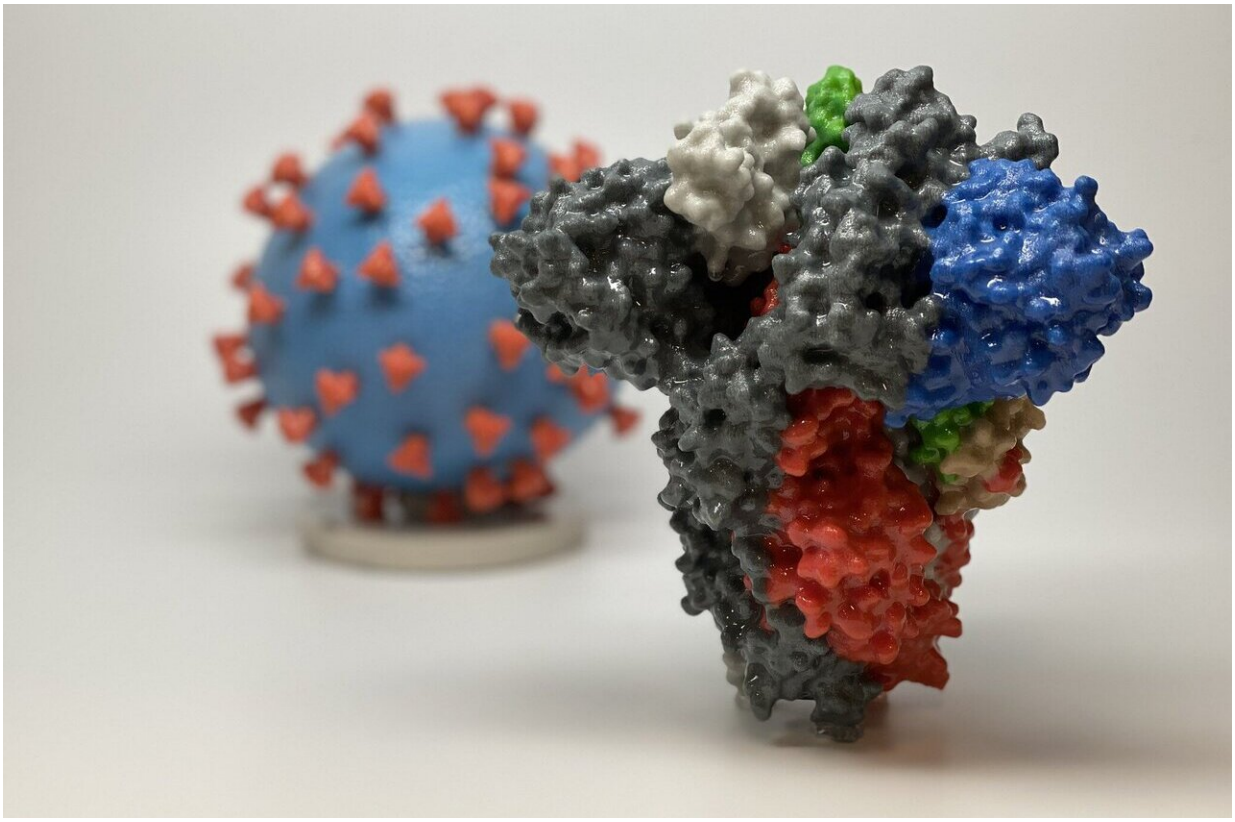
New data analysis tool uncovers important COVID-19 clues

February 28 2022



3D print of a spike protein of SARS-CoV-2, the virus that causes COVID-19—in front of a 3D print of a SARS-CoV-2 virus particle. The spike protein (foreground) enables the virus to enter and infect human cells. On the virus model, the virus surface (blue) is covered with spike proteins (red) that enable the virus to enter and infect human cells. Credit: NIH

A new data analysis tool developed by Yale researchers has revealed the specific immune cell types associated with increased risk of death from COVID-19, they report Feb. 28 in the journal *Nature Biotechnology*.

Immune system cells such as T cells and antibody-producing B cells are known to provide broad protection against pathogens such as SARS-CoV-2, the virus that causes COVID-19. And large-scale data analyses of millions of cells have given scientists a broad overview of the immune system response to this particular virus. However, they have also found that some immune cell responses—including by cell types that are usually protective—can occasionally trigger deadly inflammation and death in patients.

Other data analysis tools that allow for examination down to the level of single cells have given scientists some clues about culprits in severe COVID cases. But such focused views often lack the context of particular cell groupings that might cause better or poorer outcomes.

The Multiscale PHATE tool, a machine learning tool developed at Yale, allows researchers to pass through all resolutions of data, from millions of cells to a single cell, within minutes. The technology builds on an algorithm called PHATE, created in the lab of Smita Krishnaswamy, associate professor of genetics and computer science, which overcomes many of the shortcomings of existing data visualization tools.

"Machine learning algorithms typically focus on a single resolution view of the data, ignoring information that can be found in other more focused views," said Manik Kuchroo, a doctoral candidate at Yale School of Medicine who helped develop the technology and is co-lead author of the paper. "For this reason, we created Multiscale PHATE which allows users to zoom in and focus on specific subsets of their data to perform more detailed analysis."

Kuchroo, who works in Krishnaswamy's lab, used the new tool to analyze 55 million blood cells taken from 163 patients admitted to Yale New Haven Hospital with severe cases of COVID-19. Looking broadly, they found that high levels T cells seem to be protective against poor outcomes while high levels of two white blood cell types known as granulocytes and monocytes were associated with higher levels of mortality.

However, when the researchers drilled down to a more granular level they discovered that TH17, a helper T cell, was also associated with higher mortality when clustered with the immune system cells IL-17 and IFNG.

By measuring quantities of these cells in the blood, they could predict whether the patient lived or died with 83% accuracy, the researchers report.

"We were able to rank order risk factors of mortality to show which are the most dangerous," Krishnaswamy said.

In theory, the new data analytical tool could be used to fine tune risk assessment in a host of diseases, she said.

Jessie Huang in the Yale Department of Computer Science and Patrick Wong in the Department of Immunobiology are co-lead authors of the paper. Akiko Iwasaki, the Waldemar Von Zedtwitz Professor of Immunobiology, is co-corresponding author.

More information: Smita Krishnaswamy, Multiscale PHATE identifies multimodal signatures of COVID-19, *Nature Biotechnology* (2022). DOI: 10.1038/s41587-021-01186-x. www.nature.com/articles/s41587-021-01186-x

Provided by Yale University