Machine learning could help scientists design better viral diagnostics
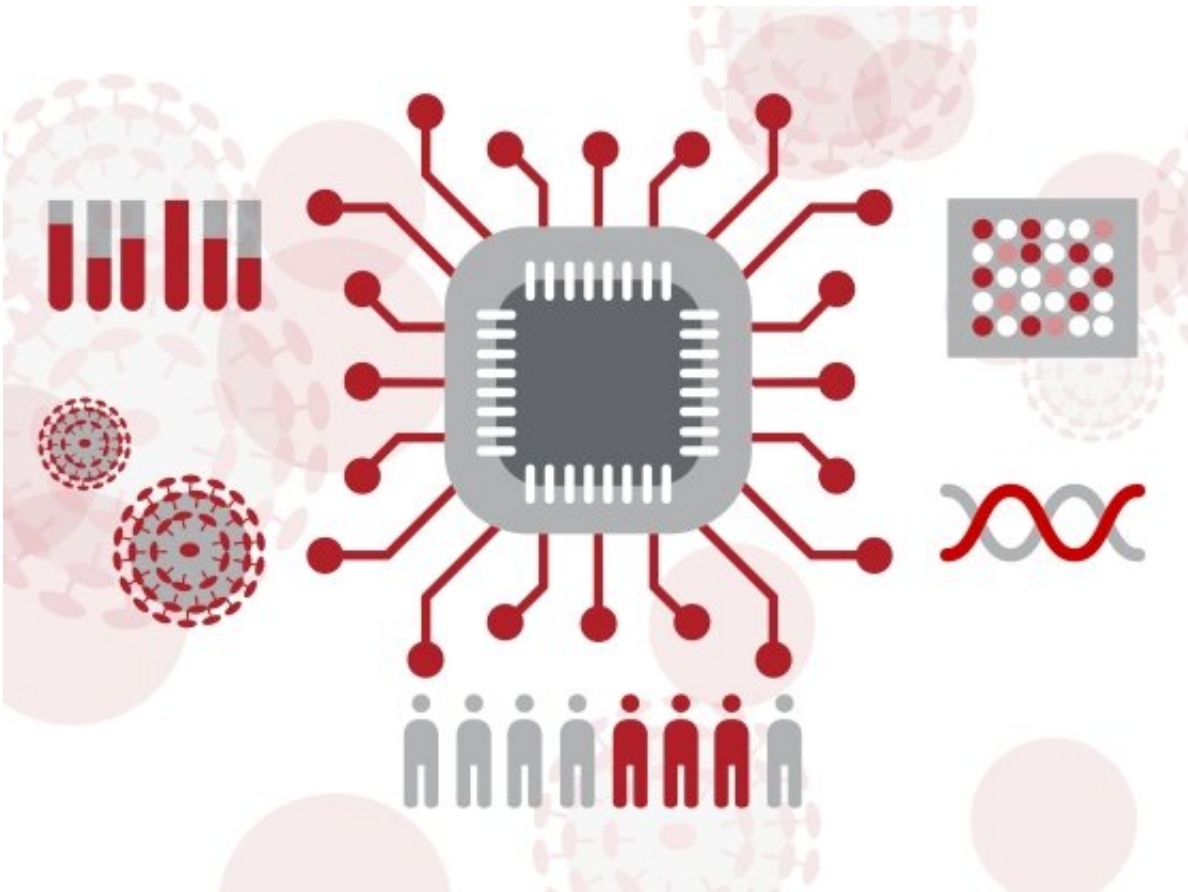
March 4 2022, by Allessandra Dicorato

Credit: Colette A. Zylstra, Broad Communications

Researchers have developed an automated method that predicts the effectiveness of viral diagnostic tests and designs optimized ones.

The surge of the omicron variant has highlighted an urgent need for [diagnostic tests](#) that accurately detect viruses, even when they mutate. Now, scientists at the Broad Institute of MIT and Harvard have developed the first fully automated system that uses machine learning to design viral diagnostics.

The method, called ADAPT, helps scientists create diagnostics that are highly sensitive (can detect low levels of [virus](#)) and specific, meaning that they detect only the virus of interest and not others. The researchers describe in *Nature Biotechnology* how they used their approach to create diagnostics for each of the nearly 2,000 viruses known to infect vertebrates, including SARS-CoV-2.

Designing a viral diagnostic involves carefully selecting the best places in a virus's DNA or RNA for the test to target. Researchers choose those sequences mostly by hand, guided by some rules, but there is also a lot of trial and error. ADAPT, which uses trained algorithms to predict the best sequences for a diagnostic, promises to help scientists rapidly design tests that are more effective for a large number of different viruses and can be quickly modified and scaled as viruses evolve.

"ADAPT is really about developing countermeasures that target the virus that's circulating right now and being prepared to move with the virus as it changes," said Pardis Sabeti, senior author of the study and an institute member at the Broad. Sabeti is also a Howard Hughes Medical Institute investigator, a professor at the Center for Systems Biology and the Department of Organismic and Evolutionary Biology at Harvard University, and a professor in the Department of Immunology and Infectious Disease at the Harvard T. H. Chan School of Public Health.

"As we've watched SARS-CoV-2 adapt in real time, we've learned just how much we need to change with it and other viruses."

Building a better model

In 2018, a team led by then-graduate student Hayden Metsky in the Sabeti lab began developing a machine learning model to analyze the wealth of viral sequence data being generated by labs around the world.

"Current techniques in machine learning and optimization are really well suited to making sense of all this data," Metsky said. "Our goal was to better leverage the diverse sequencing data out there to design more effective diagnostics."

To develop ADAPT, the team first focused their efforts on CRISPR-based tests, which use programmable "guide RNAs" and CRISPR enzymes that find specific viral sequences and generate a fluorescent signal.

The scientists then designed a large number of these tests, each to look for a different target from viral genomes. They used a recently developed Broad technology called CARMEN to measure the effectiveness of thousands of combinations of guide RNAs and viral targets simultaneously.

Using this large trove of test efficiency data, the researchers then trained a machine learning model to predict which guide RNAs would generate strong signals in a diagnostic test across different viral strains and variants. Metsky says this means that a diagnostic will be likely to detect different lineages—known and even novel ones—as a virus evolves. ADAPT also automatically incorporates new viral genomes from public databases into the design process so that it stays up-to-date as new variants emerge.

"At the core of building good diagnostics is knowing what to target and how to target it," Sabeti said. "We spend a lot of time building

technologies to do that, but we've shown that with thoughtful algorithmic work, we can get these methods to work much, much better."

Detecting SARS-CoV-2 and beyond

Early in 2020, when COVID-19 was beginning its march around the world, Sabeti and Metsky, by then a postdoctoral fellow, quickly refocused their efforts.

"When we concentrated on COVID in mid-January 2020, it was remarkable how quickly the global community was generating genomic data on the virus, with 20 genomes at the time and that number growing exponentially," Metsky said. "We had been building machine learning models and algorithms that accounted for viral variation based on genomic data, and wanted to apply our work to rapidly generate highly sensitive assays for SARS-CoV-2 that maintained that sensitivity as the virus evolves."

Metsky and the team used ADAPT to create diagnostics for SARS-CoV-2 and 66 other viruses that are genetically related or cause similar symptoms. When they tested four of ADAPT's designs in the lab, they found that the tests were more sensitive than diagnostics developed according to more traditional rules.

Though the team first used their approach to create CRISPR-based diagnostics, they say ADAPT can be applied to other sequence-based tests as well, and are already adapting it for qPCR, the most widely used viral diagnostic tool.

Metsky and Broad software engineer Priya Pillai also built a website where researchers can find and visualize diagnostics the team designed for known viruses, or run ADAPT on new data to develop their own. As ADAPT and its user base grow, the team will continue to improve their

website to make it easy to use for labs with little in-house computing power or bioinformatics expertise.

Ultimately, the team says other researchers could use ADAPT to create new, highly effective diagnostics for known or emerging viruses. In the meantime, Metsky says tests that distinguish between SARS-CoV-2 and other respiratory viruses that cause similar symptoms will continue to be critical, and ADAPT could be useful in developing those tests. "If COVID becomes endemic, we'll need to do a better job identifying the wide swath of respiratory viruses that are circulating, including their vast and ever-changing variation," he said.

  More information: Hayden C. Metsky et al, Designing sensitive viral diagnostics with machine learning, *Nature Biotechnology* (2022). [DOI: 10.1038/s41587-022-01213-5](https://doi.org/10.1038/s41587-022-01213-5)

Provided by Broad Institute of MIT and Harvard