

A dangerous stage in the evolution of the novel coronavirus is upon us with the discovery of “escape mutations”. Artificial intelligence may be our best response

Real life with COVID-19 is now scarier than anything a sci-fi writer could envision. So-called “escape mutations” that can turn the virus into an out-of-control shape-shifter that hides from the immune system are now a frightening reality. And they can’t be totally stopped with masks or social distancing, lockdowns or travel restrictions. Even if we could keep all viruses out, the ones already here are mutating in a direction that keeps them infectious and deadly. The battle between us and this often-lethal virus has just jumped to a new level.

While it may take awhile to see whether these escape mutations will evade the *vaccines* approved or in the pipeline, Tyler Starr from the Fred Hutchinson Cancer Research Center and colleagues report in a new study in [Science](#) an effect on two already available *treatments* — monoclonal antibodies. They’ve identified an escape mutation with a single glitch that enables the virus to evade Regeneron’s double-antibody REGN-COV2 “[cocktail](#)” (which Trump took) and a third antibody in Eli Lilly’s LY-CoV016. The researchers found the escapee using a new lab mapping technique that displays viruses contorted with mutation, and then they found it in a patient who was still testing positive, 145 days after the first test.

What does this mean? The discovery of escape mutations derailing antibody treatments means that the companies’ initial tests hadn’t caught them all. And the escape mutations — the new mapping revealed three others — are already in circulation.

The researchers conclude, “Ultimately, it will be necessary to wait and see what mutations spread as SARS-CoV-2 circulates in the human population. Our work will help with the ‘seeing,’ by enabling immediate interpretation of the effects of the mutations cataloged by viral genomic surveillance.”

A scientific crystal ball to predict escape mutations to come, is already in the works: Artificial Intelligence

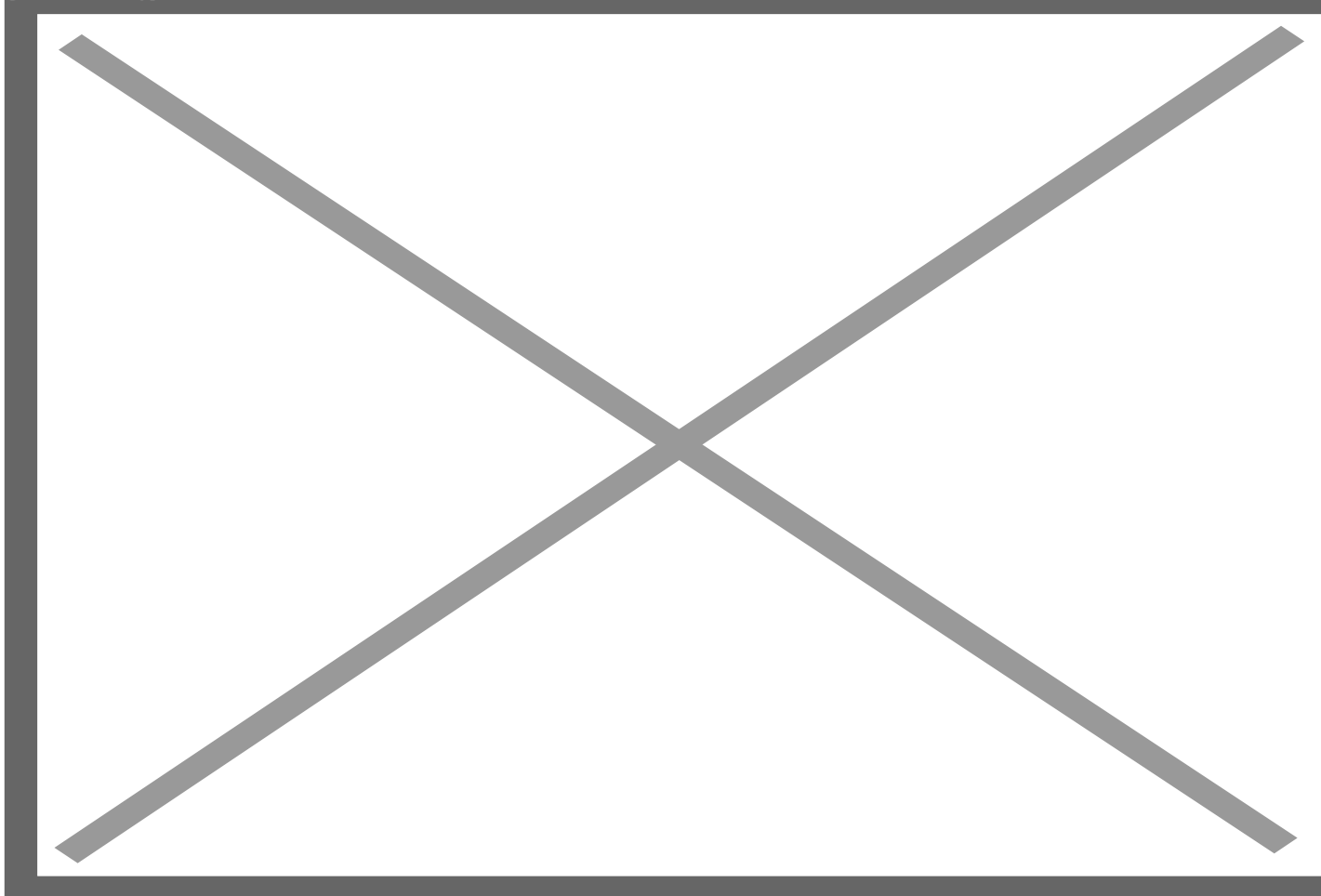
“Toto, I have a feeling we’re not in Oz anymore.”

Replacing just one word in a sentence can profoundly alter the meaning, even if the result is grammatically correct. Dorothy, of course, had been whisked out of Kansas and not Oz.

The same can be true for a virus. And this natural changeability may pose serious challenges to health officials everywhere. Small alterations in the viral genome can have no effect at all — or be so profound as to ultimately shift the course of the pandemic. The mutation that enhances transmissibility in the [UK, South African, and Brazilian](#) variants now spreading around the globe began as just a single RNA base change. Particularly distressing are the escape mutations that enable a virus to essentially hide in plain sight. A mutant virus then can more easily replicate and jump from person to person, spreading infection.

A tool from a team of computer scientists thinking about the nuances of language may be able to alert us to such future mutations. In “Learning the language of viral evolution and escape,” a recent article in [Science](#), Brian Hie and colleagues at the Computer Science and Artificial Intelligence Laboratory at MIT described how they are harnessing machine learning to predict rogue viruses that hide from our immune responses. The researchers “have uncovered a parallel between the properties of a virus and its interpretation by the host immune system and the properties of a sentence in natural language and its interpretation by a human,” wrote Yoo-Ah Kim and Teresa M. Przytycka, from the National Library of Medicine, in an accompanying [Perspective](#).

Image not found or type unknown



Depending on how the COVID-19 virus evolves, the vaccines could become less effective. Credit: Volanthevist/Getty Image

A mutation changes the sequence of the amino acid building blocks that build a protein. An *escape* mutation alters that sequence in a way that makes the protein, and the pathogen it's part of, invisible to the immune system. And that enables the virus to go about the business of making more of itself unchecked.

The researchers equated a change in the sequence of amino acids of a protein to a change in the sequence of letters of a sentence. Then they adapted artificial intelligence methods developed for linguistics applications to recognize the nuances of known escape mutations and then use the information to identify new ones, testing the approach on familiar, well-studied viruses.

Virology meets linguistics

Machine learning is an AI technique in which “training” on one dataset is used to extract meaning from new information, enabling predictions. Applications are familiar and eclectic: voice recognition, predicting financial meltdowns, developing new programming for streaming services, finding hidden patterns in fine art, and filling Facebook feeds with those annoyingly spot-on ads.

For analyzing the virus behind COVID, machine learning “trains” on databases of thousands of known mutations, seeking single-RNA-base substitutions that enhance the ability of the virus to hide from the immune system. Presumably natural selection would have weeded out mutations that didn’t help the virus. Identifying health-relevant mutations would be faster than sequencing entire viral genomes. That’s TMI.

The researchers turned to an AI machine learning technique originally developed to train computers to understand human language using sequences of words that have distinct meanings. The “words” of a virus are the amino acids that build its proteins.

“Toto, I have a feeling we’re not in Kansas anymore” has a literal meaning between Dorothy and her dog, but the deeper meaning is the emotion of being suddenly flung far away from home. Similarly, certain amino acid changes in certain parts of a key protein – the spike – spell “escape,” and the virus instantaneously has an advantage.

Validation on influenza and HIV

To test this linguistic approach to predicting scary new COVID mutations, the researchers looked at proteins from two viruses that have plagued humans for decades: influenza A and HIV.

For influenza, the algorithm compared the “head” and “stalk” regions of sugar-tipped proteins that jut from the viral surface like Tootsie roll pops. Escape mutants arise from the head region of the proteins because antibodies glom onto the stalk, preventing those viruses from replicating. In fact, the instability of the heads is what has stymied attempts to develop a [universal flu vaccine](#).

(Flu viruses change too often and in too many ways for any one vaccine to fight off all variants. That’s why we need new flu shots every year. But flu virus surfaces are not at all like those of the spiky coronaviruses.)

human immunodeficiency virus hiv aids stem cell shutterstock

Image not found or type unknown

HIV (Credit: Kateryna Kon/Shutterstock)

Similarly, for HIV, the AI algorithm zeroed in on part of a sugar-tipped protein that forms a [highly variable](#) part of the virus's outermost layer, the envelope. The protein was already known to spawn new mutations, so the AI is accurate.

AI also painted what the investigators call "semantic landscapes" that illuminated viral evolution. For example, one part of the 2009 avian influenza A surface matched a telltale amino acid sequence from the virus that caused the 1918 flu.

The AI predictions for SARS-CoV-2 are consistent with those for the better known viruses: the immune system "sees" surfaces, such as the tootsie-roll pops and spikes.

It's also intriguing to compare the amino acid sequences of spike proteins from the new coronavirus to the older ones that cause SARS and MERS. Matching spikes among the viruses and 22 types of mammals showed that bats and pangolins gave us SARS-CoV-2, camels transmitted MERS, and SARS came from civets and bats. (One theory of viral origin is that they came from the genomes of their hosts.)

The spike spawns escapees

Trios of spikes crown the coronaviruses, hence their name. Each spike has two parts. Subunit 1 (S1) binds the spike to the receptor (ACE2) on cells of the lungs and elsewhere, attaching like a fighter plane landing on an aircraft carrier. Subunit 2 (S2) then fuses with the cell membrane, creating an entryway for

the virus into the cell.

The central role of the spike in infection is why it is the target of treatments for COVID-19 and the basis of vaccines (see [“COVID-19 Vaccine Will Close in on the Spikes”](#)).

Using the AI technique described in the new study revealed precisely where in the spike protein’s amino acid sequence escape mutations are most likely to arise, and where they’re not . And that could be important intel for predicting the next threatening viral variant.

The escapees come from two specific parts of the spike, and are much less likely to come from a third place.

ace

Image not found or type unknown

Molecular model of coronavirus spike (S) protein (red) bound to angiotensin-converting enzyme 2 (ACE) human cell.

The two parts of the spike most likely to spawn escapees are where subunit 1 grabs the receptor on our cells and at one end of the protein. That's also where the new mapping technique that discovered the escape mutations that evade antibody treatments zeroed in – the receptor binding domain.

The part of the spike least likely to give rise to escape mutations, according to the AI algorithm, is the

smaller subunit 2, the part that ushers the impinging virus across the cell membrane into the cell. The fact that the doorway into our cells doesn't change much means that it's helping the virus – that's natural selection at work.

The bottom line is that we should focus on subunit 1 to catch future escapees, hopefully before they surreptitiously spread around the planet. We must get the spike where it binds – that's the virus's Achilles heel.

Even though viruses have points of vulnerability, they don't mutate on purpose; they're not 'trying' to make more of us sick or sabotage our vaccines. Instead, they are in a Darwinian battle for survival. Mutations are a consequence of errors when genetic material copies itself, like a biochemical typo. If a mutation brings a survival advantage, it persists. Mutation and natural selection fuel evolution.

Getting ahead of the virus

Instead of discovering the latest variation on the coronavirus theme weeks or months after people have already exhaled it on international flights or other forms of [traveling](#), virus trackers can use the new tool to actively watch for specific escape mutations, as well as for combinations of mutations into novel variants. Perhaps a catalog of hotspots in the SARS-CoV-2 genome, derived from many iterations of machine learning surveillance, can be translated into a rapid test applied to COVID test swabs, instead of sequencing entire viral genomes.

Follow the latest news and policy debates on sustainable agriculture, biomedicine, and other 'disruptive' innovations. Subscribe to our newsletter.

[SIGN UP](#)

One could envision deploying such a rapid test in the aftermath of events expected to have superspreader potential, like people packing airports during a holiday, fans crowding spontaneously onto a football field at a key game's end, or people rioting in a nation's capitol. If viral escapees can be picked up sooner, we'll be better able to successfully halt their spread through more conventional approaches, such as contact tracing, quarantine, and isolation.

As time goes on and the number of sequenced SARS-CoV-2 genomes continues to climb — it's currently nearing half a million at the database [GISAID.org](#) — the machine learning tool can track viral evolution and make predictions, at the same time that the new mapping tool keeps up with ongoing mutations. Mutation surveillance tools can monitor new sets of mutations accruing into [variants](#), and variants begetting strains when their component mutations interact and bestow broader traits.

Ultimately, AI machine learning may be invaluable in tracking how the no- longer-so-“novel” coronavirus changes in response to reinfections and to vaccines and in helping scientists to find new ways to slow or stop its deadly global march.

Ricki Lewis has a PhD in genetics and is a science writer and author of several human genetics books. She is an adjunct professor for the Alden March Bioethics Institute at Albany Medical College. Follow her at her [website](#) or on Twitter [@rickilewis](#)