

How computational immunology changed the face of COVID-19 vaccine development

15 July 2020

Image credit: alashi/DigitalVision Vectors/Getty

When Chinese officials posted the sequence of the coronavirus SARS-CoV-2 on 10 January 2020, it triggered a race among vaccine manufacturers. Historically, this process has taken years, even decades. The vaccine against the Ebola virus, which zoomed through human trials in a record-breaking 5 years, took more than twice as long in preclinical development. However, SARS-CoV-2 was different. Within a few hours, several companies had developed potential vaccine targets.

Understanding the immune system is a tall order, says Maggie Ackerman, an immunologist at Dartmouth College, in Hanover, New Hampshire. It is why so many scientists like her have turned to computational and informatics approaches, mathematical models that have become so sophisticated they can predict which parts of a novel pathogen will be recognized by B cells and T cells or create targeted immunotherapies against tumor cells.

When the novel coronavirus struck, years of work on these models meant that scientists were poised to respond immediately. By being able to predict the exact parts of SARS-CoV-2 that would elicit an immune response, scientists have been able to sprint through the early stages of vaccine development and into animal trials.

"These approaches offer incredible speed at getting from genetic sequence to a candidate vaccine. Nothing can compete with that," says Ackermann.

It all started with a copy–paste

One of the main reasons scientists first turned to computational tools to understand the immune system is that they needed to. Millions of years of natural selection have meant that the immune system has created multiple layers of defense with the redundancy and adaptability to meet nearly any threat it is faced with. The end result is an organ system that, even today, researchers still do not fully understand.

Because most of the early work in immunology was in mice, there is detailed understanding of the mouse immune system, explains Mark Davis, a computational immunologist at Stanford University, in California. However, understanding what is going on in humans is another story.

“Mathematical models, statistical approaches, and machine learning are the only way you can make meaning out of so much data,” says Maia Smith, a bioinformatics engineer at AbCellera in Vancouver, Canada.

As the field of bioinformatics grew, immunologists began borrowing some techniques used by geneticists and systems biologists, which ultimately created the subfield of computational or systems immunology.

These mathematical models started off with relatively simple ordinary and partial differential equations that let scientists describe and predict how a system changes over time and space, says Filippo Castiglione, a computational immunologist at the Institute for Computing Applications at the National Research Council of Italy.

The mid-1990s were a turning point for the field of immunology. At the time, the escalating human immunodeficiency virus (HIV)–AIDS crisis and early DNA sequencing results from the Human Genome Project had begun to generate large datasets on immune functioning. This gave immunologists a newfound urgency for addressing the deadly pandemic, as well as a starting point for developing their models, according to Annie De Groot, chief

executive officer and founder of the computational immunology company EpiVax in Providence, Rhode Island.

The problem was that when immunologist started in this field, it was very difficult to get funded, because people just did not believe that computers could do this, De Groot says.

To be fair, compared with the sophistication of current tools, De Groot's initial computational algorithms were rather simple. At the time, she was a postdoc in the lab of Jay Berzofsky and was working to understand how T cells recognize pathogens. As soon as *Helicobacter pylori* was first sequenced, in 1997, she simply copied and pasted the sequence into a word processor and ran some macro, looking to identify peptides recognized by major histocompatibility complex (MHC) class II proteins — and it worked.

De Groot continued to search the literature for more peptides and MHC class II proteins to be able to improve the precision of her computations as she started her own lab working on tuberculosis and HIV at Brown University. What resulted was EpiMatrix, a computer algorithm that breaks down a pathogen's protein sequences into chunks that are ten amino acids in length, and then ranks them by their likelihood of binding to a given MHC protein. The output is an estimated binding probability that compares the algorithm's predictions with its score of known MHC binders and non-binders.

When De Groot tested the program in *Mycobacterium tuberculosis* to identify proteins that might make good vaccine candidates, she was able to reduce the number of epitopes by 99.8%, from 1.6 million to 3,000. The algorithm also identified conserved HIV epitopes that are recognized by MHC proteins.

Making sense of all the data

As De Groot toiled away to find MHC class II epitopes, other researchers were using differential equations to identify a target for a universal vaccine

against influenza, a task that became more urgent with the 2009 H1N1 influenza pandemic. Both HIV and the influenza virus mutate at staggering rates, which means that vaccinologists had to try to find epitopes that remain constant over time and elicit a strong and durable immune response. It was a tall order, since those epitopes most recognized by the immune system were also the most variable.

However, scientists at Oxford University created a set of equations to model the evolution of seasonal influenza virus from year to year and found that they could identify epitopes that would fit both those criteria, which made them good candidates for a universal vaccine against influenza. Equations predicted that these epitopes would wax and wane over time as populations developed immunity, something the researchers verified in human samples in a 2018 paper in *Nature Communications*. A US-based startup, Blue Water Vaccines, licensed this strategy and is using it to develop a universal vaccine against influenza.

Justin Bahl, an epidemiologist at the University of Georgia, in Athens, Georgia, also took clues from evolutionary biology to try and trace the evolution of influenza virus strains and see if he could identify a common ancestor that might be useful in vaccine design. Rather than modeling the future evolution of the genetic code of influenza virus, Bahl instead tried to predict what the virus's RNA sequences and protein structures looked like in the past.

"If we know what's conserved in all of these different influenza viruses, we can combine that with what's conserved in the human body response," Bahl says.

What these volumes of data perhaps illustrated best was the heterogeneity of immune responses. The vaccine against hepatitis B virus is often given at birth, and while some people need only one dose to be fully protected, others need two or three, says Richard Scheuermann of the J. Craig Venter Institute in San Diego, California. He and his colleagues studied immune cells

from samples collected before and after vaccination and used single-cell RNA sequencing followed by machine learning to identify what contributes to different vaccine responses. Computational methods, Scheuermann says, helped them narrow down a number of candidate genes expressed specifically by dendritic cells. "We ended up with only a dozen genes to evaluate, which is a much more specific signal," he says.

The answer, Scheuermann's team found, was the number of myeloid dendritic cells expressing a gene called NDRG2, according to a 2018 study in the *Journal of Immunology*. With these results, Scheuermann says, he can investigate whether adjuvants designed to boost the immune response to a vaccine will affect the activity of NDRG2 in myeloid dendritic cells.

Instead of relying on mathematical models based on differential equations, Castiglione and other scientists have begun using agent-based models. These models treat each cell or other entity as an agent that is defined by a set of rules that also incorporate some randomness, with the goal of determining their effects on the system as a whole. The result is a distribution of probabilities for a variety of outcomes, instead of an estimate of average behavior. Combining this with neural networks and other machine-learning techniques has allowed Castiglione to predict the existence of a phenomenon called 'memory anti-naïve', in which cross-reactive memory T cells inadvertently inhibit the formation of a more effective T cell response to a secondary infection.

Computational immunology can fast-forward the research, but is not a shortcut

These computational approaches are especially good at generating hypotheses, according to bioinformatician Sagi Shapira of Columbia University, in New York, New York. He points to work by Peter Howley of Harvard University, in Cambridge, Massachusetts, who asked why certain strains of human papillomavirus are associated with cervical cancer and

others are not. Bioinformatics data showed that equivalent proteins in different human papillomavirus strains bind to a different constellation of host proteins in the cell. Howley hypothesized that these differences could explain why some strains cause cancer. It allowed Howley and other immunologists to hone their hypotheses more finely before diving into expensive and time-consuming 'wet lab work'.

When it comes to designing vaccines and antibody therapies, building a viable candidate can take years and cost tens of millions of dollars. By developing and investing in the advanced computational tools used by scientists at EpiVax, Moderna, AbCellera and other companies, this process can be compressed into hours instead of years. Although De Groot's databases and mathematical modeling have grown exponentially more sophisticated, compared with her original Microsoft Word-based endeavors, she still keeps the recognition of peptides by MHC class II proteins at the center of her analysis. However, she also looks for sequences that may alert regulatory T cells and dial back the immune response, as well as modeling how the use of two different peptides might affect immunogenicity. At AbCellera, scientists have whittled down the billions of antibodies found in a sample of blood from someone who has recovered from COVID-19 to a top few candidate antibodies from which an effective blocking epitope could be identified. The key to this process has been the start-up's immunoinformatics tools that can link antibody to antigen. Developing their antibody therapy even further will occur in collaboration with Eli Lilly.

Shapira also cautions that no in silico analysis, no matter how high-quality the input and how exacting the computational algorithms, will ever be a substitute for experimental data. Many hypotheses and many vaccines against malaria and HIV, and even universal vaccines against influenza, have looked good on paper but been a flop when tested in humans. There is also the ongoing issue of reproducibility, an issue that has plagued much of science but has hit many '-omics' studies especially hard.

"There's no shortcut to actually doing the experiment," Shapira says.

The testing stage is where both EpiVax and AbCellera are at with their COVID-19 response. AbCellera began phase 1 human trials of a SARS-CoV-2-neutralizing monoclonal antibody called 'LY-CoV555' that targets the virus's spike protein. Also, EpiVax is collaborating with several labs to develop candidate vaccines, some of which have already begun preclinical studies. Moreover, AbCellera has teamed up with Eli Lilly to identify the best antibodies from the blood of patients who have recovered from SARS-CoV-2 infection to create a monoclonal antibody therapeutic. To De Groot, more than two decades of work in computational immunology was what enabled her company to be able to develop a vaccine candidate in just a few hours.

"With the tools that we have, we can pivot to whatever seems to be capturing public interest at the moment," De Groot says, "and eventually we can address those really big problems."

doi: <https://doi.org/10.1038/d41591-020-00027-9>