Using Machine Learning to Fight Future Pandemics

Dr. Turck:
Deep machine learning technology to predict the structure of antibodies to fight against emerging antigens? It almost sounds like science fiction. But believe it or not, this technology is being studied now, and if it comes to fruition, it could help us get ready for the next pandemic.

Coming to you from the ReachMD studies, this is *COVID-19: On the Frontlines*. I'm Dr. Charles Turck, and here to talk about his current research on this topic is Dr. Animesh Ray, a Professor of Computational and Molecular Biology at Keck Graduate Institute. Dr. Ray, welcome to the program.

Dr. Ray:
Thank you very much. Thanks for inviting me.

Dr. Turck:
Now before we dive into your study, Dr. Ray, let's start with some background. Can you tell us what we currently know about variable chain sequencing in the context of antibodies and foreign proteins?

Dr. Ray:
Yes. So currently, there are two ways by which this is done. One is purely experimental in which we inject the antigen the viral protein, let's say, the spike protein, into an animal and then raise antibodies against it. And then we sequence the gene that encodes the antibody that's specifically binds the antigen. And that's from animals, let's say. And then that may be humanized by obtaining the structure of the encoding of genes somewhat, and then that is made into a humanized antibody that the humans can tolerate.

The second approach is to find patients who have been infected with a particular virus, and then from the patient's antibody repertoire, one isolates a tight biding antibody, and the corresponding cells, the t-cells or the b-cells, basically very similar to the antibody variable chains and then sequences the gene and thereby gets a handle on the human antibody itself and then makes it into a therapeutic product. So these are the main ways by which one does that. There are, of course, alternate ways by which one can do it, and they are very plausible and the technologies are very developed, but I don't know to what extent these have been applied to therapeutic antibody against a viral disease. And these have something to do with artificially generating short antibody-like sequences in bacterial viruses and then doing an in vitro panning against a viral antigen in vitro that is in a test tube. This is called phage display and then enriching for tight binding structures and then taking antibody-like sequences that bind very strongly to the antigen, taking that and changing it a little bit, putting a real antibody scaffold on it that can be tolerated by humans, optimizing it, making sure that it neutralizes the virus, and that's true of the other two approaches also, you know, the binding is not sufficient, it has to demonstrate neutralization of the virus and then developing that as therapeutics.

So basically, two of them are from either mammals or humans, and the other is purely artificial in the test tube, but it's exactly the same kind of approach.

Dr. Turck:
And where to do some of our knowledge gaps in this area lie?

Dr. Ray:
I think the knowledge gap is classical. And that is we don't know how a protein that is made is going to fold such that it is going to bind tightly to an antigen and then neutralize that antigen, if it is a virus, neutralize that virus, without at least some knowledge of related

antibodies that bind to this virus. And the reason for that is the so called central dogma of molecular biology, which has been proposed many, many years ago in the 1950s first and then written down in 1960s by Francis Crick, simply on the basis of coding theory that you cannot get very easily, how a protein folds in three-dimension, to the DNA sequence that might encode this set of amino acids that were fold in such a manner because there is a problem of many to one encoding. There are many different three-dimensional structures that can be encoded by many different amino acid sequences. Once the amino acid sequences, encoding it into the gene is very easy. And then the direction or flow of information goes from DNA to RNA to protein. But from protein structural features alone, it's not possible to find unique structures that would encode that three-dimensionally folded structure without some guidance. And that's where the b- hole is and that is imposed by the so-called reality of life, the physics of information transfer.

Dr. Turck:
Let's focus in on your research on this topic now. What made you and your colleagues consider using computers and relying on deep machine learning entirely to be able to predict neutralizing antibody structures?

Dr. Ray:
So my colleagues and I have been working on predicting protein-in-protein interaction from already existing data using machine learning for the past approximately twelve to fourteen years. I wanted to make sure that we understand that this work is a teamwork, which includes my colleague, Dr. Jennifer Hernandez at Keck Graduate Institute as well as a physiochemistry professor at Pomona College, Dr. Matt Sazinsky, and a computer scientist at University of California Riverside Dr. Stefano Lonardi. So we noticed that if you take a particular protein that is of interest, for example the Huntington's Disease protein, it makes a toxic protein. The toxic protein, if it is present, kills the certain neural cells. It also interacts with many other proteins. And we were interested in finding genetic modifiers of the Huntington's Disease, that is gene variants that might alleviate the disease intensity or perhaps make the disease intensity even more and thus could have a therapeutic target. So we wanted a computation that predicts such genetic interactions and the paths that we took many years ago was if you can predict additional protein interactors of the mutant Huntington's Disease protein, then we might be able to find genetic modifiers. So we used machine learning and that was quite a bit successful. So this led us to the idea that if you could predict interactors of Huntington's Disease then why can't we predict interactors of an antigen in terms of an antibody structure?

Dr. Turck:
What specific strategies are you using to generate synthetic antibody candidates from virus proteins? And how do you know if they'll work on emerging antigens in the future?

Dr. Ray:
So we use phage display, but we are using a variant of a phage display which allows us to screen very rapidly and in one step quantitatively putative antibody-like structures against the library of antigens. That's one.

The second approach is actually immunizing mice and using a recently developed technique by somebody else; it's called LIBRA-seq, using deep sequencing of single cell sequencing of b-cells that have been induced by a foreign antigen and then antigens attached with a DNA barcode and are incapsulated in lipid droplets along with the antigen, so the single cells are encapsulated and then resequenced the encoding receptor sequences, which encode something similar to the antibody sequences along with the cognate antigen or the epitope. So that's the plan.

The second part of the question is how do we know these will be effective against future viruses? So for every future virus, you would have to think about possible antigens; that's the first part and that may be difficult part if the virus is not very familiar to us. But because of the epidemiological surveillance and the zoonotic virus surveillance activities that are going on throughout the world, which became somewhat mired in controversy a few years ago, the surveillance works are extremely important. So we need to know as many different kinds of virus that are around that could perhaps in the future affect humans but nowadays, only affect mammals or birds or you know some such animals. We need to know them as much as possible ahead of time so that we can identify possible antigens on the surface of these viruses that we might target to neutralize the virus activity. So this is a very important endeavor. We must know that.

But then, once we know this, if our research succeeds, then by knowing the antigen structure, someday we ought to be able to predict the antibody structure computationally by using a machine learning algorithm that has been trained already to do so using the data that we are producing or the rest of the world are producing currently or have already produced.

Dr. Turck:
For those just tuning in, you're listening to *COVID-19: On the Frontlines* on ReachMD. I'm Dr. Charles Turck and I'm speaking with Dr. Animesh Ray about his current study on using deep machine learning to predict antibody sequences against emerging antigens.

So, Dr. Ray, there's so much about your study that we could delve into, but would you share with us what you hope to discover and at a global level how your results could help fight another pandemic?

**Dr. Ray:**

Yes. So before beginning, I have to emphasize that vaccine and vaccination, as widely as possible, are the most effective ways of combating any such pandemic, any such future pandemic. So you have to develop those vaccines and currently the RNA vaccines are perhaps the best way forward because of their acuity with which they can be developed and then spread around the word. I hope in the future more, more readily and cheaply than we can do today because of the various distribution problems.

Having said that, the second line of defense would be monoclonal antibodies that are able to neutralize the virus if a virus infection does occur. So the vaccination is to prevent the infection, per se, but we are developing a second step in the process of combating the pandemic. Can one come up with monoclonal antibody therapy against any unknown virus that would presumably emerge in the future? And of course we can do that. And our objective is using computers, can we shorten it further to three months? Half as long? And that's where we are trying to do this. So once a virus infects a patients, can one treat the patients with a monoclonal antibody that has the ability to neutralize the virus?

**Dr. Turck:**

Now before we close, I'm curious if there are other deep machine learning technologies you're aware of that are being studied that may be exciting for the medical community to look forward to?

**Dr. Ray:**

Most certainly. Google's DeepMind, which has made AlphaFold, now AlphaFold 2, I believe is a wonderful deep learning algorithm that finds protein structure. We don't know how it can be used to predict interacting protein structures, but we will be using DeepMind very heavily in order to predict some of or simulate some of the structures that we are discovering by experiments to do our work. So we will definitely use that. And there are many other machine learning approaches that are trying to predict three-dimensional structures of protein from primary sequences and they're not as successful as DeepMind but they work to some extent, perhaps better for shorter fragments in certain cases. So there are quite a few right now.

**Dr. Turck:**

Well that gives us a lot to look forward to as we continue leveraging new technologies in the fight against emerging infectious diseases. And as that brings us to the end of today's program, I wanna thank you for joining me and for sharing this fascinating research, Dr. Ray. It was great having you on the program.

**Dr. Ray:**

Thank you very much for the chance to speak to you.

**Dr. Turck:**

I'm Dr. Charles Turck. To access this and other episodes in our series, visit ReachMD.com/COVID-19, where you can Be Part of the Knowledge. Thanks for listening.