Software package enables deeper understanding of cancer immune responses
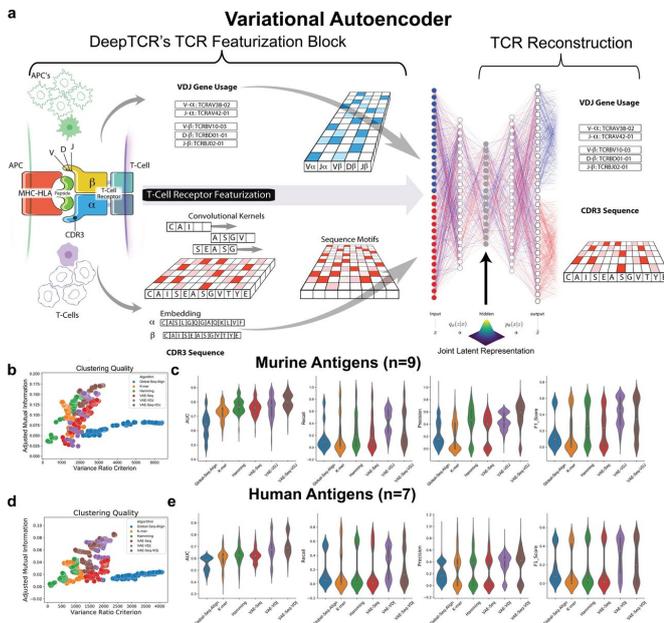
2 April 2021, by Amy Mone

Fig. 1: Unsupervised TCR sequence representation. Credit: *Nature Communications* (2021). DOI: 10.1038/s41467-021-21879-w

Researchers at the Bloomberg~Kimmel Institute for Cancer Immunotherapy at the Johns Hopkins Kimmel Cancer Center have developed DeepTCR, a software package that employs deep-learning algorithms to analyze T-cell receptor (TCR) sequencing data. T-cell receptors are found on the surface of immune T cells. These receptors bind to certain antigens, or proteins, found on abnormal cells, such as cancer cells and cells infected with a virus or bacteria, to guide the T cells to attack and destroy the affected cells.

"DeepTCR is an open-source software that can be used to answer questions in research into infectious disease, cancer immunology and autoimmune disease; any place where the immune system has a role through its T-cell receptors," said lead study author John-William Sidhom, an

M.D./Ph.D. student at the Johns Hopkins University School of Medicine and Department of Biomedical Engineering working in the Bloomberg~Kimmel Institute for Cancer Immunotherapy.

The research was published March 11 in *Nature Communications.*

Sidhom was inspired to develop the software after attending a presentation on the use of deep learning for the medical sciences at the 2017 meeting of the American Association for Cancer Research. "I was doing research on T-cell receptor sequencing, and it struck me that this was the right technology to better analyze T-cell sequencing data," he says.

Deep learning is a form of artificial intelligence that roughly mimics the workings of the human brain in terms of pattern recognition. "Deep learning is a very flexible and powerful way to do pattern recognition on any kind of data. In this paper, we use deep learning to identify patterns in sequencing data of the T-cell receptor," says Sidhom, adding that the way the software explores T-cell receptors is analogous to an internet search. "When someone performs an internet search for an image of cats or dogs, the query doesn't involve looking for images that have a caption that labels the image as a cat or dog, but rather applies an algorithm that explores the features of the images and recognizes patterns that identify the images as a cat or dog. This is deep learning."

DeepTCR is a comprehensive deep-learning framework that includes both unsupervised and supervised deep learning models that can be applied at the sequence and sample level. Sidhom says the unsupervised approaches allow investigators to analyze their data in an exploratory fashion, where there may not be known immune exposures, and the supervised approaches will allow investigators to leverage known exposures to improve the learning of the models. As a result, he

says, DeepTCR will enable investigators to study the function of the T-cell [immune response](link) in basic and clinical sciences by identifying the patterns in the [receptors](link) that confer the function of the T cell to recognize and kill pathological cells.

One of the main challenges of analyzing TCR sequencing data is distinguishing meaningful sequencing data from inconsequential data, and DeepTCR helps perform this analysis. "There are a lot of sequences in someone's immune repertoire. There are a lot of pathogens that someone can be infected by, so the immune response is very broad. As a result, there is a sea of noise in the immune response, and only parts of it are important at a certain time for a certain infection," Sidhom explains. "I may have T-cell responses to a thousand different viruses, but when the flu impacts me, I only need to utilize a small subset of those T cells to fight the flu. The main thing that the algorithm can do is isolate and match the right T cells to specific responses."

The [software package](link), which employs a type of deep-learning architecture called a convolutional neural network, provides users the ability to find T-cell sequencing patterns that are relevant to a specific exposure, like a flu infection, a cancer or an autoimmune disease.

"When presented with a lot of data, our algorithms can learn rules of these TCR sequence patterns. For example, we may not know the rules for how the body responds to flu, but with enough data, our software can learn those rules and then teach us what they are," says Sidhom. "It is very well-suited to identify complex patterns in a very, very large immune repertoire to identify the interacting partners between a T-cell receptor and its antigen."

Provided by Johns Hopkins University