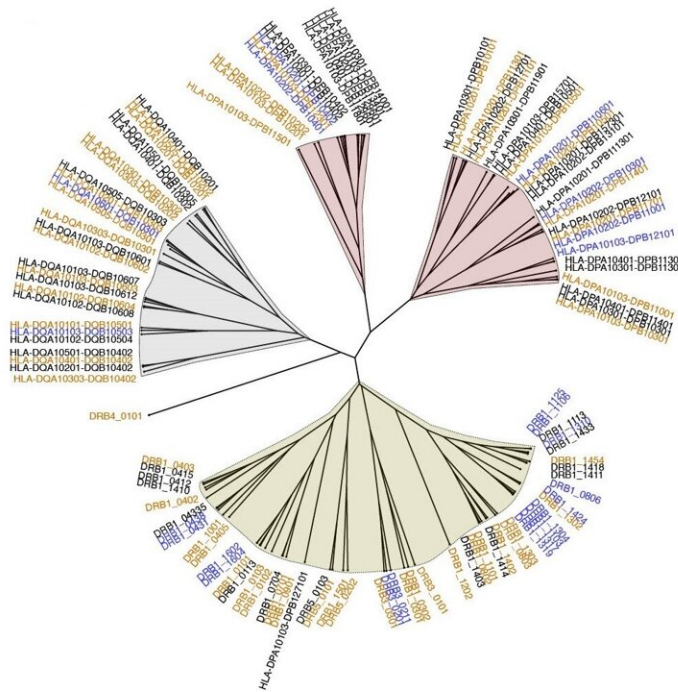# Scientists map the antigenic landscape

November 27 2023

Combined specificity tree for HLA-DR, HLA-DP, and HLA-DQ. Orange molecules have peptide coverage corresponding to at least 50 high-confidence ligands, and blue molecules have a pseudo-sequence distance of at most 0.05 to an orange molecule. Logos in red frames correspond to noncovered molecules. Credit: Jonas Birkelund Nilsson

A new paper in *Science Advances* details how scientists have succeeded in mapping a central part of the immune system—the HLA class II molecules—while accurately predicting how they display fragments of pathogens on the surface of cells.

When we are sick, the immune system relies on signs on cell surfaces that something foreign is present inside. Immune cells—specifically T-cells—latch onto the cell's surface and kill the cancer, virus or whichever pathogen is there, so long as they can determine the threat.

Cells alert the immune system of the intruder through special proteins called human leukocyte antigen (HLA) molecules. They are responsible for letting the immune system know that something is amiss.

"When a cell becomes infected, whatever is inside it is hidden from the immune system, which lives outside the cells," says Morten Nielsen, a professor from DTU Health Technology and corresponding author of the paper in *Science Advances* announcing the mapping of more than 96% of the entire HLA class II landscape.

"The reason the body can detect that something is hiding inside the cell is HLA class molecules and the fact that they take fragments of proteins from the pathogen inside the cell, transport them to the surface, and display them. If the fragments have properties that aren't recognizably yours, the immune system starts a reaction which kills the cell."

"But the rules for which protein fragments are displayed and which are not, and what other properties it had, have been very unclear for many years because there are many different HLA variants. You could say there are more than 50.000 ways to display our protein fragments."

Nielsen has been working on HLA for the past 20 years and has made significant contributions to the process behind developing treatments aimed at assisting and training the immune system in combating diseases. Much of the progress made within immunotherapy against cancer has some connections to tools developed by Nielsen.

In the paper, titled "Accurate prediction of HLA class II antigen

[presentation across all loci using tailored data acquisition and refined machine learning](link)," scientists from DTU, University of Oklahoma, Leiden University and the company pureMHC successfully complete the mapping of the entire system, or, as it is called in the paper the "specificity tree" of HLA class II.

20 years in the making

It has taken 20 years to complete the HLA class specificity landscape map for several reasons. For one, they are never the same from person to person. Their genes differ widely, so different people have different kinds of HLA that recognize different parts of a pathogen.

While they all play a pivotal part in the function of the immune system by displaying the protein fragments, they affect health in different ways. Some make us more likely to get autoimmune diseases, where the immune system attacks the body. Some increase the likelihood of rejecting an organ transplant. Some affect how well the immune system responds to treatments such as vaccines or drugs.

Also, there are two parts to each HLA class II molecule: an alpha part and a beta part. They, in turn, come from three different groups of genes: DR, DP, and DQ. The DR group has one primary gene, DRB1, and three other genes, DRB3, DRB4, and DRB5. The DP and DQ groups have two genes, DPA and DPB and DQA and DQB. The alpha and beta parts can come from the same gene or different chromosomes.

At times, it has been stipulated that knowledge of DRB1 was sufficient or that other combinations were less important when characterizing the functional HLA class II space. It turns out, however, that several other HLA class II play an essential role in, for example, autoimmune disorders and with regards to not repelling transplanted organs. They may also be vital in treating other diseases, so the interest in creating

immunotherapy treatments that recognize them is rising.

In any case, there are many possible combinations in the HLA class II system, and since only the DRB1 molecules have been investigated and mapped extensively, the understanding of the entire complex of HLA class II has been lacking.

Large-scale datasets and machine learning

To understand how the myriad HLA class II genes affect health, Nielsen and his colleagues needed to know what kinds of pathogens they recognize and how they present them to our immune system. To make this final push and figure out the rules defining HLA class II, they integrated large-scale, high-quality datasets covering a wide variety of HLA class II molecules and their specificities. They used tailored machine learning frameworks, thereby improving the ability to accurately predict how they function.

"Twenty years ago, we were looking at 500 data points from one molecule, but we soon learned that there were rules to this. We didn't have to measure everything. So, gradually, our understanding grew, as well as the available technology. We have gone from our first paper with one molecule to our latest paper, which covers 50,000 molecules. All of which are described in detail," says Nielsen.

"We have overcome every hurdle and completely understand what every HLA class II molecule does. For instance, our tools have been used for the past 15 years in developing cancer immunotherapy, and they have served as cornerstones for many companies developing cancer vaccines. And our tools are the most used."

"With the current paper, we now offer the full toolbox, a toolbox that may also be used for viral infections or autoimmune diseases. There will

still be a lot of research in this field, but conceptually, I believe the journey is complete, and I don't believe anything more is going to happen."

More information: Jonas B. Nilsson et al, Accurate prediction of HLA class II antigen presentation across all loci using tailored data acquisition and refined machine learning, *Science Advances* (2023).

Provided by Technical University of Denmark