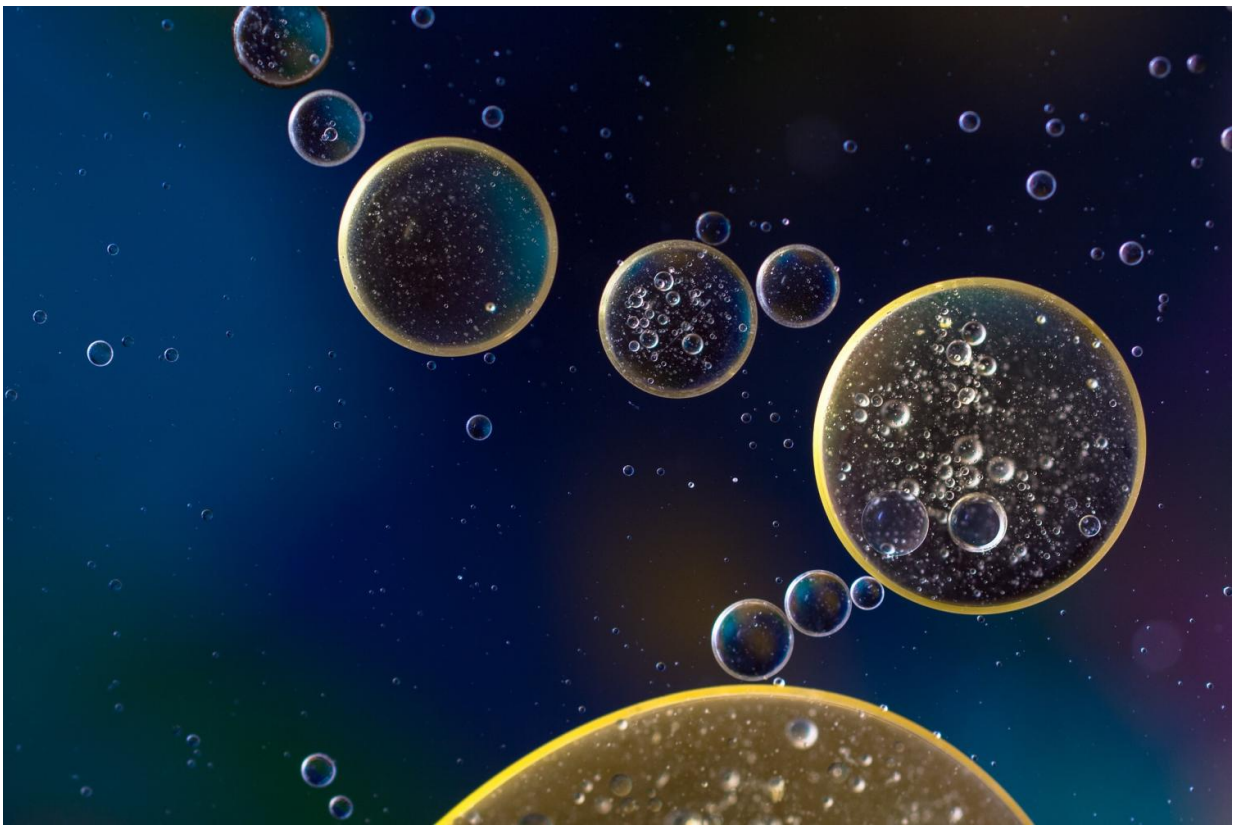# Medical Xpress

Researchers post genetic profiles of a half-million human immune cells on Human Cell Atlas online portal

April 16 2018, by Jenny Rood

Credit: CC0 Public Domain

A team of postdoctoral and research scientists at the Broad Institute has made a data set of half a million human immune cells openly accessible

on a preview site that provides initial access to data for the Human Cell Atlas initiative.

The data set, one of the largest of its kind, includes primary data and associated metadata from nearly 530,000 immune cells from umbilical cord blood of newborns and bone marrow of adults. Additional data sets were also provided by Wellcome Sanger Institute and collaborators.

"This is a wonderful example of science at its most open and collaborative," said team co-leader Orit Rozenblatt-Rosen, an Institute Scientist at the Broad and director of the Klarman Cell Observatory (KCO).

This data lays the foundation for an immune cell atlas, an important first step in the Human Cell Atlas consortium's goal of an initial draft atlas of 30 million cells covering many tissues. "The immune system is deeply complex, involved in many diseases, and distributed throughout our body. This data set will be critical to help unlock its secrets," said Monika Kowalczyk, a hematologist who led the experimental team while a postdoctoral researcher in the lab of Broad Core Institute Member Aviv Regev.

By making the data openly accessible before drafting their manuscript for publication the researchers have provided the broader scientific community with a valuable resource. The data set can reveal basic biology, provide a reference for studying disease, and allow computational biologists to test new analysis tools on a large data set that would be hard for smaller labs to generate.

"Collecting and processing half a million immune cells was a Herculean feat, involving tightly coordinated teamwork across many areas of expertise," said team member Danielle Dionne of the KCO at the Broad.

First, Kowalczyk and her KCO colleagues Dionne, Michal Slyper, and Julia Waldman isolated single cells from human cord blood and bone marrow samples and prepared them for sequencing. This required meticulous advanced planning since the team was handling 224,000 cells from four patients in a 20-minute window—up to 100-fold more cells than in a typical experiment.

Computational biologists on the team then needed to determine how to assess quality and analyze a batch of data large enough that they couldn't be analyzed with existing computational tools. To handle the data, the trio of Orr Ashenberg of the KCO and Bo Li and Marcin Tabaka of the Regev lab built new computational methods, working from code that was either openly available (such as SCANPY) or provided by their colleague Karthik Shekhar. These tools identified for example cell types from the sequencing data, found signature genes that characterize them and showed how particular cell types developed from others.

Next, before releasing the massive data set, the team worked with other Broad colleagues—Jane Lee, who coordinated logistics for the entire project, Stacey Donnelly, and Andrea Saltzman—to ensure that each sample had appropriate patient consent for data release. In the process, they set up an approach applicable to future samples—including an additional set of 1.08 million cord blood, bone marrow, and white blood cells that the team, in collaboration with Broad Institute Member Nir Hacohen and Alexandra-Chloe Villani, has already processed and will release once all approvals are confirmed.

   More information: The data is now available at preview.data.humancellatlas.org

Provided by Broad Institute of MIT and Harvard