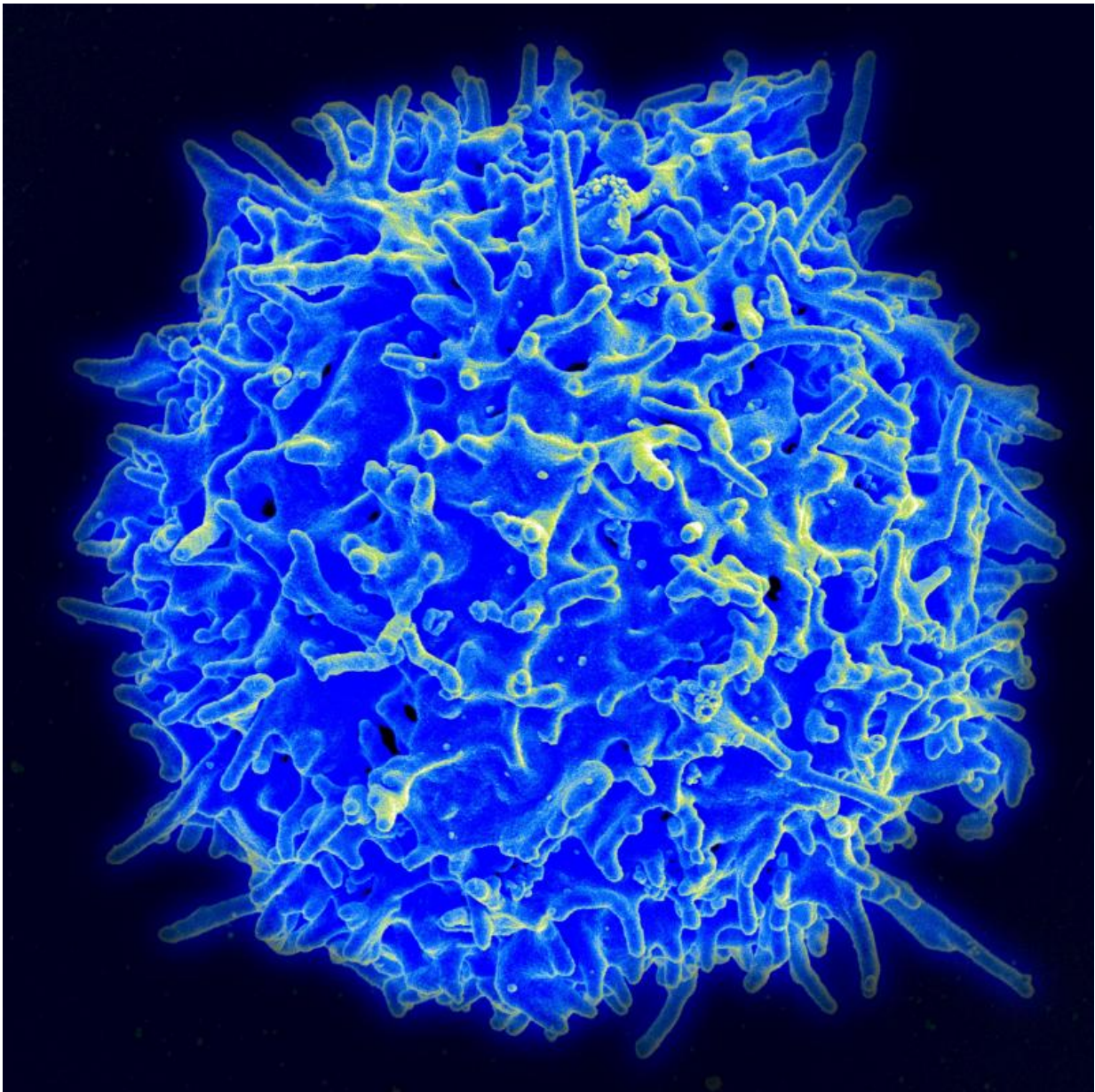# High-performance computing helps chemists sort through cellular statistics

July 24 2015, by Benjamin Recchie

This image shows a healthy T-cell, one of several types of lymphocyte cells that help defend the body from foreign invaders. Researchers led by Prof. Aaron Dinner seek to understand what combinations of up 420,000 genes create proteins that control the development of lymphocytes.

DNA is often referred to as "the blueprint of life." But it's more than just a blueprint—it's also a kind of operations manual for the workings of the cell, telling it what proteins to manufacture and when.

Aaron Dinner, professor in chemistry, and his graduate student Herman Gudjonson are trying to read that manual, as part of the Dinner group's research into bioinformatics—the application of statistics to biological research. To carry out their research, Dinner and Gudjonson turned to Midway, the Research Computing Center's supercomputing cluster.

Gudjonson's project focuses on lymphocytes, a type of white blood cell that's found in vertebrates' immune systems. In addition to the three main types of lymphocytes—T cells, B cells, and NK cells—immunologists have recently identified new types of innate lymphoid cells, each of which is tailored to play a specific role in defending the body from foreign invaders. All these varieties originate from a common progenitor cell and only mature later into their more specialized forms, but exactly how a cell "knows" what form to take is unclear.

Some combination of genes must create the proteins that control the development of the innate lymphoid cells, but which ones? There are up to 420,000 potentially relevant genes in these cells, explained Gudjonson. To make matters worse, the key factor might not necessarily be the protein that's found in the highest concentration in each variant, but instead some combination of proteins found in lower abundances.

It's impossible to track all the proteins in the cell, so instead, Gudjonson and co-workers are measuring concentrations of messenger RNA, an intermediary between the DNA and the final proteins. A cell's mRNA can be extracted and sequenced relatively easily, thanks to recent advances in DNA sequencing technology.

Being one level removed from proteins, it's "not the full story," said Dinner. "It's a shadow in some sense." For one thing, the mRNA can't tell you if the protein is in an active or inactive state. For another, the measurements of the mRNA themselves aren't perfect, he cautioned. "You're effectively doing a massively parallel measurement inside your test tube," Dinner said. "You'll have some variations in each one of those measurements and need a probabilistic model to reconstruct the actual information from that measurement." Despite all this, the researchers say it's the best data they can get for all species in cell at the same time.

The interdependent probabilities and statistical analyses demanded by this project called for high-performance computing. Having access to the Research Computing Center's Midway doesn't just speed up their research, explained Gudjonson. "It gives you more flexibility in the kinds of test you can choose," making more difficult analyses easier and allowing the researchers to approach problems in multiple ways.

In addition, the researchers used Midway to run verifications on their experiments, and RCC set up guest accounts on Midway for the Dinner group's collaborators at Northwestern University. This allowed them to share data with their Chicago counterparts and review the results seamlessly.

Gudjonson's work has already helped to identify key intermediary steps in the development of lymphocytes, showing that certain cells express molecules associated with two or more different types of cells before "deciding" on which immune response specialization to take up. The

next step, said Dinner, is trying to apply machine learning techniques to explore the sequence of the cells' developmental steps, as well as to better understand what fraction of [cells](#) at one stage of development become each type at the next.

"In principle," Dinner said, "these studies could lead to therapies for congenital immune deficiencies, and also for associated cancers." So in addition to be being the blueprint for life, DNA might provide a troubleshooting manual as well.

Provided by University of Chicago