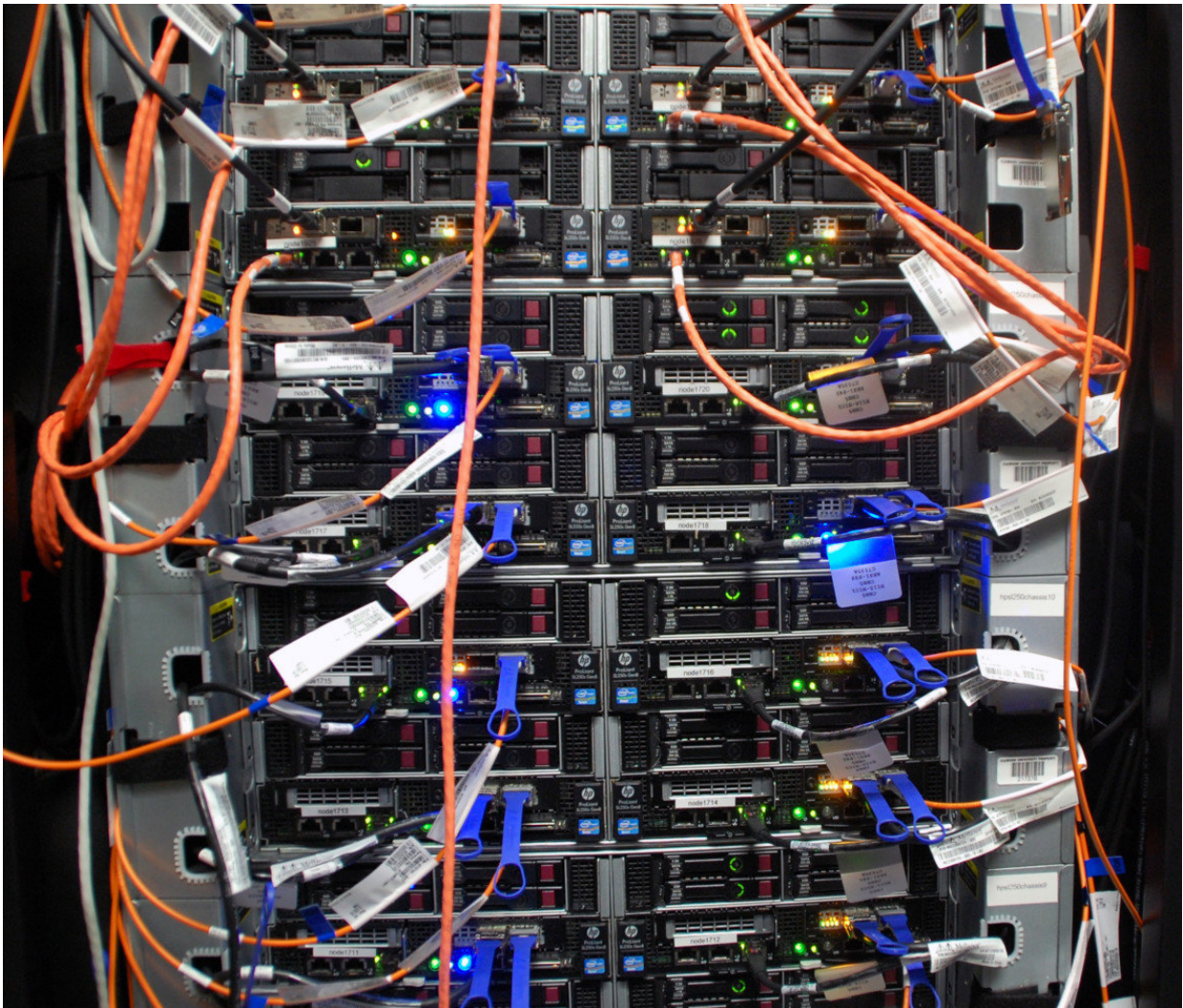


Condition-specific gene co-expression patterns in cancer

October 27 2017, by Hannah Halusker



Clemson's Palmetto Cluster can complete trillions of mathematical operations per second. Credit: College of Science

A Clemson University professor, an alumnus, a former IT staff member and two students have unveiled a computer software that can sort genes to better understand how they interact to cause disease. Published in the journal *Scientific Reports* in August, the software will help researchers examine complex traits controlled by multiple genes.

Known as the Knowledge Independent Network Construction (KINC), the [software](#) package is the culmination of many years of research conducted by professor Alex Feltus in the department of genetics and biochemistry. Feltus' former graduate student, Stephen Ficklin, now an assistant professor at Washington State University, developed the software. Undergraduate and graduate researchers Leland Dunwoodie and Will Poehlman worked with CCIT staff member Kim Roche – presently a doctoral student at Duke – to compile and analyze the data.

By pulling more than 2,000 tumor gene expression datasets from The Cancer Genome Atlas – a public repository for genomic information relating to 33 different types of cancer – the team was able to statistically and visually organize the [genes](#) based on their shared functions.

"The idea is that if two genes are on at the same time, they might be working together," Feltus said. "So we weaved together a network of genes that interacted with each other and with this we can actually find five, 20, sometimes thousands of genes that have to rise and fall together in their expression levels. Then what we can do is say, 'Well, what are those gene packages doing?' "

With the KINC software, researchers don't need to have prior knowledge of what the gene network should look like because the software will run the data to determine how genes should be categorized. This knowledge-independent method reduces the amount of "noise" – from laboratory protocols or from natural variation between cells – that can prevent

genetic interactions from being discovered.

"Sometimes the software sorts samples into groups of the same kinds of tumors, which for us was thyroid, bladder, ovarian and two kinds of brain tumors: glioma and glioblastoma," Feltus said. "But sometimes, for each pair of genes, the software looks at how they're related to each other in different groups of samples. So maybe you find these two genes only interact with each other in thyroid cancer, so they're more likely to be thyroid-cancer-specific gene interactions. But the software can also get specific to the grade of the tumor by identifying interactions that happen in early stage tumors, advanced stage tumors, male tumors, female tumors or even ethnicity."

Once KINC has sorted the genes into groups, researchers can conduct deeper statistical tests to uncover correlations between genes and cellular pathways in the body. In the resulting gene co-expression network (GCN), two genes that have a high likelihood of interacting with one another will be connected by a line, called an edge.

Not uncommon in the construction of gene networks, which tend to have large datasets, the team's case study resulted in a GCN with so many gene interactions – so many edges – that it resembled a hairball.

"In science, we're always trying to reduce a system down to one or two variables. But using bioinformatics, we're reducing hundreds of thousands of variables down to just hundreds," Feltus said. "We embrace the complexity of the system, but we want it to be meaningful by cutting out the noise."

To process and analyze these complex genomic data, the field of bioinformatics requires high-speed, large-scale computing abilities. Considering that just one experiment generates 700 terabytes of data – enough to fill more than 700 laptops with data files – it's easy to see why

bioinformatics and supercomputing go hand-in-hand.

Fortunately, Clemson University has the Palmetto Cluster, one of the top 100 supercomputers in the world, located in Pendleton. Because it's operated on a democratized condominium model system, any Clemson faculty, staff or student can register for an account and use the supercomputer free. But given that it's shared among so many people, the amount of storage available for the team's cancer study was limited.

"Compared to traditional co-expression analyses, running our KINC software was a large computational challenge," Poehlman said. "We quickly realized that we couldn't generate results in a reasonable amount of time using only the supercomputer here at Clemson, so I spent a lot of time working with the Open Science Grid to develop workflows that enabled us to tap into computing resources across the country to complete this experiment."

The development of KINC came as part of a broader investigation by Feltus and colleague Melissa Smith of the department of electrical and computer engineering called "Scientific Data Analysis at Scale," or SciDAS. Funded by a \$2.95 million National Science Foundation grant, the team intends to build a national computing system to make data processing more efficient.

"With SciDAS, we're now generating supercomputers dynamically by taking one supercomputer and mapping a new supercomputer to it through advanced networks, and we're using KINC as a way to process data from many, many species. Then, we put the data out to these distributing computer systems, opening up the ability to process these large datasets for people," Feltus said.

Dunwoodie, who uncovered 22 genes specific to glioblastoma while conducting the team's study, said he's honored to have connected the

KINC algorithm to cancer biology. He currently has a paper in preparation that is analyzing how those genes affect glioblastoma tumor development.

The team's software is free and available to the public. Those who download KINC can even modify it for their research needs by changing its code so long as they keep their modified version publicly available, too.

More information: Stephen P. Ficklin et al. Discovering Condition-Specific Gene Co-Expression Patterns Using Gaussian Mixture Models: A Cancer Case Study, *Scientific Reports* (2017). [DOI: 10.1038/s41598-017-09094-4](https://doi.org/10.1038/s41598-017-09094-4)

Provided by Clemson University

Citation: Condition-specific gene co-expression patterns in cancer (2017, October 27) retrieved 27 April 2024 from <https://medicalxpress.com/news/2017-10-condition-specific-gene-co-expression-patterns-cancer.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.