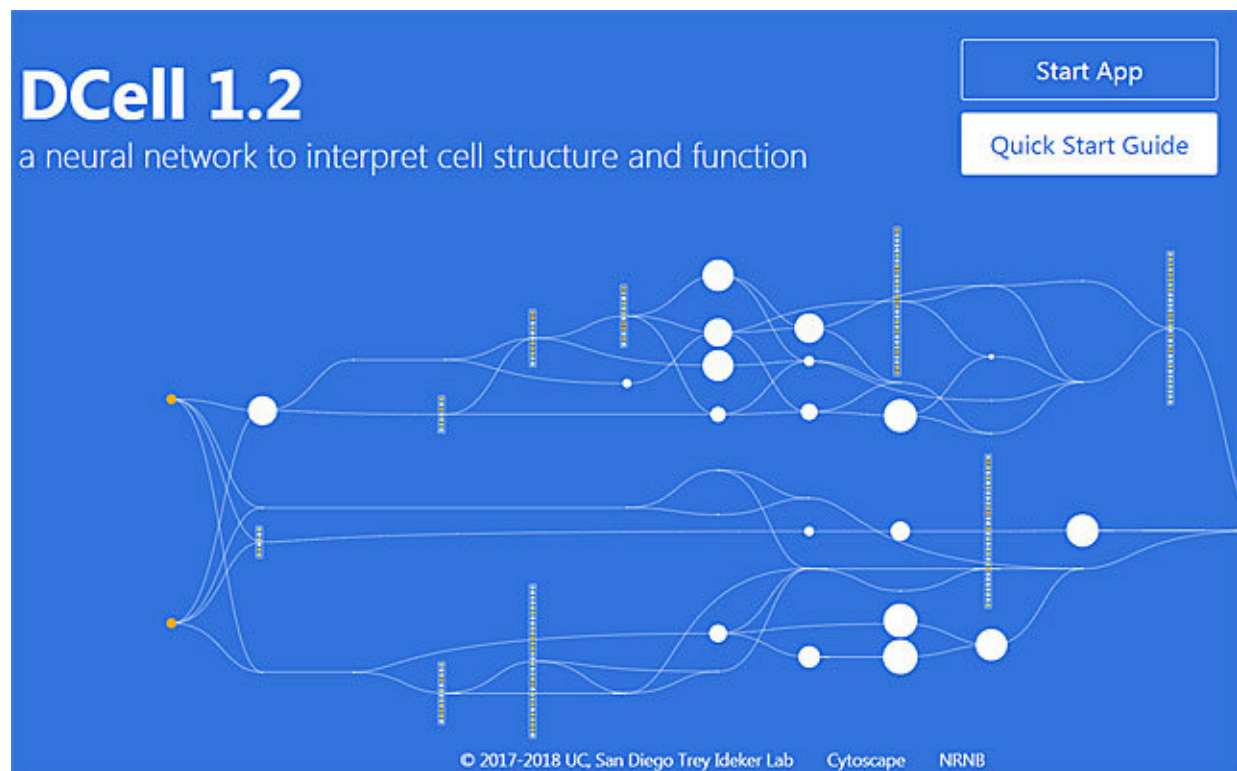# How a yeast cell helps crack open the 'black box' behind artificial intelligence

March 5 2018

Screenshot from d-cell.ucsd.edu, where researchers can use DCell, a new virtual yeast cell developed at UC San Diego School of Medicine. Credit: UC San Diego Health

UC San Diego School of Medicine researchers developed a visible neural network and used it to build DCell, a virtual model of a

functioning brewer's yeast cell.

"It seems like every time you turn around, someone is talking about the importance of artificial intelligence and machine learning," said Trey Ideker, PhD, University of California San Diego School of Medicine and Moores Cancer Center professor. "But all of these systems are so-called 'black boxes.' They can be very predictive, but we don't actually know all that much about how they work."

Ideker gives an example: machine learning systems can analyze the online behaviors of millions of people to flag an individual as a potential "terrorist" or "suicide risk." "Yet we have no idea how the machine reached that conclusion," he said.

For machine learning to be useful and trustworthy in health care, Ideker said, practitioners need to open up the black box and understand how a system arrives at a decision.

Machine learning systems are built on layers of artificial neurons, known as a neural network. The layers are tied together by seemingly random connections between neurons. The systems "learn" by fine-tuning those connections.

Ideker's research team recently developed what they call a "visible" neural network and used it to build DCell, a model of a functioning brewer's yeast cell, commonly used as a model in basic research. To do this, they amassed all knowledge of cell biology in one place and created a hierarchy of these cellular components. Then they mapped standard machine learning algorithms to this knowledgebase.

DCell can be viewed at d-cell.ucsd.edu. The technical details are published on March 5 in *Nature Methods*.

But what excites Ideker the most is that DCell is not a black box; the connections are not a mystery and cannot form by happenstance. Instead, "learning" is guided only by real-world cellular behaviors and constraints coded from approximately 2,500 known cellular components. The team inputs information about genes and genetic mutation and DCell predicts cellular behaviors, such as growth. They trained DCell on several million genotypes and found that the virtual cell could simulate cellular growth nearly as accurately a real cell grown in a laboratory.

"Human knowledge is incomplete," said Jianzhu Ma, PhD, an assistant research scientist in Ideker's lab who led the efforts to build DCell. "We want to complete that knowledge to help guide predictions, in health care and elsewhere."

Ideker and Ma also put DCell to the test. If they deliberately fed the system false information, it wouldn't work. Take ribosomes, for example. Cells use these tiny biological machines to translate genetic information into proteins. But if the researchers instead wired ribosomes to an unrelated process like apoptosis, a system cells use to commit suicide, DCell could no longer predict cell growth. The virtual cell "knows" that the new arrangement isn't biologically possible.

Ideker and his colleagues at the Cancer Cell Map Initiative, which he co-directs, are now generating some of the experimental data they need to build a DCell for human cancer. Then they will determine how best to personalize this virtual cell approach for a patient's unique biology.

"We want one day to be able to input your specific cancer-related genetic mutations and get back a readout on how aggressive your cancer is, and the best therapeutic approach to prevent its growth and metastasis," said Ideker, who is also founder of the UC San Diego Center for Computational Biology and Bioinformatics.

Provided by University of California - San Diego