

# Deep Learning-Based Cell Model May Predict Patient Tumor Cell Growth Someday

Mar 06, 2018 | [Uduak Grace Thomas](#)

## *Premium*

NEW YORK (GenomeWeb) – Researchers from the University of California, San Diego and Tel Aviv University have developed an *in silico* model of a yeast cell that simulates cellular growth from genotype information that they hope can eventually be used to make predictions about tumor cell behavior.

The so-called DCell, a visible neural network, uses existing deep learning techniques to learn and model cellular behavior almost as accurately as has been observed in the lab. Unlike traditional neural networks, which, once trained, offer little insight into exactly how they predict behavior, the mechanisms by which DCell works are "visible" and can be manipulated because the model was trained using information about cellular subsystems and hierarchies, according to its developers.

Essentially, DCell is "constrained" to model cell biology, meaning that its developers only allowed connections between network structures that conform to known cellular mechanisms, Trey Ideker, a professor of medicine and bioengineering at UCSD and one of the developers of DCell, explained. What this means is that on one hand, researchers can use the model to predict cellular growth based on genomic information, and on the other hand, they can identify the specific mechanisms that govern the predictions that the model is making, he said.

This is crucial because "in biology and medicine, simple prediction is not enough, you also have to know the mechanism," Ideker said, and it has historically been one of the challenges of using neural networks to predict phenotype from genotype. "Although there's been some promising work in making predictions from the genome using artificial intelligence [techniques], the problem with all of these deep learning systems is that they are what we call 'black boxes' ... you don't really know why they are making the predictions," he said. "So how do we get a deep learning system not just to make good predictions from your genome but [also] accurately capture the mechanisms for why those predictions are made? That's what this system does."

According to a paper published in *Nature Methods* this week that offers technical details of the model's development, the researchers used molecular data from Gene Ontology and the Clique-extracted Ontology to generate the relationships between different subsystems of the cell model. They then trained DCell to predict phenotypes related to cellular fitness using several million genotype-phenotype training examples. Specifically, they trained the model to predict cell growth and to calculate a genetic interaction score for double gene deletions, according to the paper.

DCell not only simulates cellular growth almost as accurately as real yeast cells grown in the lab, the researchers reported, but its predictions are as good as those provided by a similar neural network that did not have any constraints. Moreover, DCell outperformed existing predictors based on metabolic models and protein-protein interaction networks, according to the paper. The model is also able to correctly respond to false data inputs — when it receives information that contradicts known biological reality, it does not work.

Furthermore, "mechanistically, we can look at every system in the cell and say 'what are the systems in the cell that explain why we got this bad growth rate or this accelerated growth rate for this particular genetic perturbation or genome?'" Ideker said. "That's exciting because we have a deep learning system that's not a black box but is visible [and] models not just function but structure."

For their next steps, the researchers are exploring the possibility of using the model to make predictions about disease. In *Nature Methods*, they suggest some possible research scenarios where DCell could be used. For instance, the model could be used to study genotype-phenotype associations to identify the specific cellular subsystems associated with growth defects. Researchers could also use the model to assess the contributions of different cellular systems to a given phenotype and prioritize those that are more important. Finally, DCell could be used to identify previously unknown connections between genotypes and phenotypes.

Meanwhile, Ideker and his colleagues have their sights set on creating a version of DCell that can model tumor cell growth. The idea would be to build a model that would allow scientists to input information about specific cancer mutations and get information about how aggressively the cancer is going to grow, providing clinicians with additional information that they could use to select treatments for their patients. "If I can simply predict the growth rate of the tumor from its set of tumor mutations, the genotype of the tumor, that would be incredibly powerful and that would be a huge coup for precision medicine," Ideker said.

But providing personalized models of tumor cell growth is a far more challenging prospect than modeling the behavior of simple yeast cells. To that end, Ideker and colleagues at the [Cancer Cell Map Initiative](#) (CCMI), which he co-directs, have begun generating experimental data that they believe they will need to build a DCell model for human cancers. According to a [separate paper](#) published in 2015, the CCMI aims to understand the complex interactions between cancer genes and how these interactions differ between diseased and healthy states.

"The idea there is really to ask, 'what do we need to build one of these models for cancer?'" Ideker said. "It involves new experimental data to better understand the structure of a cancer cell and it involves informatics advances like the one we are talking about for how you process those data." Initially, the researchers are focusing on head, neck, and breast cancers because of the expertise of the scientists involved in the initiative at UCSD and the University of California, San Francisco.

As part of their efforts to apply DCell to cancer, Ideker and his colleagues will work with existing datasets from large-scale projects such as the Cancer Genome Atlas and the International Cancer Genome Consortium. Collectively, these and other cancer-centric resources provide access to several thousand exomes from breast cancer samples and several hundred head and neck cancer datasets, providing ample training data for the model. What is lacking "is the structure of the model that would turn it from a black box into an open visible system," he said. "In yeast, we had a lot of prior datasets that let us structure that."

Specifically, the researchers had data on some 2,500 known cellular components, which they used to generate the thousands of connections that comprise their yeast model. For cancer cells, "we don't have that kind of rich data," he said. To that end, Ideker's team will need to collect data on how genes and proteins in cancer cells connect to one another to create protein complexes and pathways. "We are doing protein-protein interaction mapping [and] we are doing genetic interaction mapping," he said.

They will also work with imaging datasets, including data on different cell organelles and their components. "This is an ambitious project," he said. "My lab is focusing on generating interaction mapping datasets ... [but] there's a whole network that we are trying to create of different pieces." For example, they are partnering with the lab of David Agard, a professor of biochemistry and biophysics and of pharmaceutical chemistry at UCSF and one of the CCMI consortium members, to use cryogenic electron microscopy techniques to study protein complexes in cancer cells, Ideker said.

Furthermore, the researchers may explore commercialization options for DCell down the road. Ideker is one of the founders of [Data4Cure](#), a California-based company that has built a business around a proprietary cloud-based platform that combines and analyzes various kinds of omics data to help users identify molecular markers associated with specific disease types, including cancer.

Whether or not it will be possible to use the same model for multiple patients will depend on how much the structure and function of the cells varies from patient to patient, according to Ideker. However, "our [focus] right now should be on the first model – 'can we come up with a complete simulation of a cancer cell?'"

**Filed Under**   [Informatics](#)   [Research & Discovery](#)   [Cancer](#)   [breast cancer](#)   [head and neck cancer](#)  
[bioinformatics](#)   [computational biology](#)   [imaging](#)   [machine learning](#)