7-19-2013

Mining Cell Data To Answer Cancer's Tough Questions

by ANNA HAENSCH

July 16, 2013   2:43 PM

Chemistry, genetics and computing give us clues to understand cancer cells.

Sometimes a drug hits cancer hard. Sometimes the cancer cells are unfazed. But it's often hard to know which outcome to expect.

A group of scientists at the National Cancer Institute has spent the last three years turning some mathematical algorithms loose on giant sets of data to better understand the relationship between cancerous cells and cancer drugs.

They've decided to make the data public and published their findings this week in *Cancer Research*.

Shots called molecular pharmacologist Yves Pommier, who led the research team, to find out more. "Cancer is more difficult to treat in certain tissues, including the colon, breasts, ovaries, kidneys, lungs, and also leukemia and melanoma," he says. So he and his crew harvested 60 types of human tumor cells for each of these cancer types.

They put the cells in petri dishes and let them split over and over again. After a while, the had created something they called the "NCI-60." Think of it as simplified cancer simulator that could stand in for tumors in the lab.

Pommier explains the genetic side of the story. "For each of the cells in the NCI-60, we've sequenced the coding part of the genome," he says. This coding part of the genome is called the exome, and it contains the information needed to make proteins.

Just like genome sequencing, exome sequencing creates a gigantic set of data that Pommier says, "gives the genetic landscape of each cell." This is held up to the healthy human genome to identify the genetic variations that are specific to cancer.

For an explanation of the drug side of things, Shots called on James Doroshow, the Director of the Divison of Cancer Treatment and Diagnosis at the NCI, and also a member of the research group. "Thousands of chemical compounds have been screened through the NCI-60," he says.

To be screened, a drug is added into the petri dish with each of the cancer cells. "We look for growth inhibition, growth stabilization, or no growth response," he says.

"More drugs have been screened across this panel of cell lines than any other panel in the world, and for the past several years this has been a free service. Send us a group of molecules, and we will

screen it for you and give the results back," Doroshow says.

And the fruits of these efforts? Millions of data points explaining the genetic mutations in cancer cells, plus millions more data points quantifying how cancer cells react to chemical compounds.

The next step was to find connections between the two huge data sets, so they brought in the mathematicians. They came up with a powerful algorithm to predict the effectiveness of an anti-cancer drug. This so-called Super Learner algorithm takes in the chemical formula for a drug, and then spits out some number quantifying how effective it is in the presence of genetic variations. Low numbers are good, high numbers are bad.

So they arrive at a satisfying, and quantified, answer. And better yet, they have some clues as to why, genetically, these numbers are what they are.

Although the notion of mining medical data to determine treatment options is admittedly a bit scary, it can be invaluable as a tool for building hypotheses. "From a discovery perspective, people can mine these mutation analyses, and find potentially novel mechanics of drugs, both old and new," Doroshow says.

Yves Pommier pointed out one success of the NCI-60 panel, an anti-cancer drug called Nutlin which has been in clinical trials for several years. "Using the NCI-60 database, we predict that Nutlin will only work for cells with a normal p53 pathway," he says. He hopes that this better understanding of the drug will bring it closer to FDA approval.

Since most cancer treatments involve a complex drug cocktail, the group has also developed algorithms to explore combinations of drugs. They've just finished analyzing every possible combination of commercial anti-cancer drugs — that's 5,000 unique combinations. "We got some unpredictable combination synergies, and we hope this will drive new therapeutic combinations in patients," says James Doroshow.

The group has decided to make all of their data publicly available on the NIH sponsored website Cell Miner, so that it can be widely used for hypothesis-driven pharmaceutical research, and hopefully speed up our understanding and improvement of cancer treatments.