

COMPUTING CANCER'S WEAK SPOTS

An algorithm to unmask tumors' molecular linchpins
is tested in patients *By Roxanne Khamsi*

In 2016, doctors invited Eileen Kapotes to join a clinical trial for a drug that had never been used for her disease. Kapotes, a first grade teacher in her 50s, was fighting an aggressive breast cancer that had spread through her body. She had endured grueling treatments over the previous 4 years, including whole-brain radiation therapy. She had also been taking the breast cancer medication Herceptin, but her tumors were still growing. Now, she had a chance to try something radically different: a drug called ruxolitinib, originally designed to treat cancers affecting the blood and bone marrow.

Kapotes's oncologist, Amy Tiersten at Mount Sinai Hospital, was stunned by how well her patient responded to the new drug. It kept her cancer at bay and she had almost no side effects. "I was amazed," Tiersten says.

The ruxolitinib trial was the product of a decadelong quest by Andrea Califano, a systems biologist at Columbia University. Using sophisticated computing, he models the molecular networks that sustain cancer cells and pinpoints proteins called transcription factors that act as linchpins, controlling the behavior of many genes inside a cell. Califano collaborated with cell biologist José Silva, then also at Columbia, to analyze breast cancer samples in a repository of tissues from other patients who had become resistant to Herceptin. Findings of the analysis suggested a transcription factor called STAT3 plays a critical role in those cancers. And ruxolitinib was known to inhibit STAT3.

Other researchers have focused on identifying genetic mutations that drive the disease in an individual patient. Doing so, the thinking goes, can help identify the best drug for each patient. But because of the diversity of cancer-causing mutations across the population, an arsenal of tens of thousands of drugs might be needed to treat everyone.

Califano's approach, by contrast, is a twist on that idea. He has focused instead on identifying a few transcription factors that act as bottlenecks (see graphic, p. 1176). Target those master regulators, as Califano calls them, and you will stop cancer in its tracks, no matter what mutation initially caused it. Oncologists would still need to analyze

A visualization of 7000 interacting genes in a brain cancer patient shows the handful with the largest number of interactions (red)—which makes them tempting targets for therapies.

each patient's mutations to figure out which regulators are at play in their particular cancer, but instead of tens of thousands of drugs, Califano says, they may only need dozens. It's a depersonalized approach to personalized medicine.

The strategy builds on Califano's computational training as a physicist. "We've built algorithms that can reverse engineer the logic of each different tumor so that we know the targets" for drugs, he says. His algorithms are a prime example of systems biology—which uses complicated math to model intricate biological systems, such as gene interactions. It's a field that has generated tremendous interest, but little real-world clinical success.

In 2015, Califano co-founded a company called DarwinHealth that uses his algorithms to guide doctors by identifying the key transcription factors in a patient's tumor and suggesting drugs to target them. His work has earned praise from other researchers, although some note the approach is only in early stages of human testing, and its clinical usefulness remains uncertain. Ed Liu, president and CEO of the Jackson Laboratory, a nonprofit biomedical institute Califano has collaborated with, is optimistic the method will ultimately pay off. "As we develop more and more precise ways to attack those nodes, then the more useful his algorithms will be."

Califano's approach is about to get its largest test yet. Columbia has allocated \$15 million for a trial of 3000 cancer patients within its hospitals over the next 3 years, using DarwinHealth algorithms to analyze each patient's cancer and recommend treatments. "This is probably one of the most exciting moments in my research," Califano says, "because finally we're able to apply this methodology on a scale that is large enough to be able to really learn something in terms of the response of the patient."

IN FALL OF 1958, a young scientist named François Jacob went to his colleague Jacques Monod at the Pasteur Institute in Paris with a hypothesis about how genetic mechanisms might control cell behavior. Both men had renegade tendencies: Jacob had fought Nazis—and been injured—on behalf of the exiled French government in World War II, and Monod, an accomplished rock climber, had taken part in the guerilla activities of the French Resistance. Over the next few years, the pair worked together, and they were the first to demon-

strate the idea of genetic circuits. The work ultimately won them a share of the 1965 Nobel Prize in Physiology or Medicine.

In experiments with *Escherichia coli*, Jacob and Monod showed that the gene networks in those bacteria can alter the production of certain enzymes depending on the type of food available. When the sugar lactose was abundant, the bacteria turned on genes that code for the enzymes to metabolize it. But with access only to glucose, a different sugar, the microbes shut down those genes. It was a pioneering demonstration that the activity of individual genes could be either boosted or repressed.

Experiments in later decades helped explain how the cell machinery exerts that control. One key player is transcription factors, proteins that boost or inhibit the activity of other genes. The gene-regulating network of a single cell is far more elaborate than Jacob and Monod had the tools to uncover. The human genome contains 20,000 genes, and an estimated 1500 of those produce transcription factors. That system creates a complex web of on and off switches.

Califano thought that if he could identify the key switches in cancer, he might be able to shut down the catastrophic genetic changes that drive its growth. But after he finished his training as a physicist in 1986, IBM recruited him to spearhead projects in computer vision and artificial intelligence. The building codes at the IBM facility prevented Califano from having an experimental lab to pursue his interests in biology. He left in 2000 and landed at Columbia in 2003. He started to write code to solve the riddle of cancer on the day he arrived.

Nowadays, the data underlying his algorithms come from a method called RNA sequencing (RNA-seq). The method gauges gene activity within cells by sequencing RNA molecules, which act as a proxy for which genes are turned on and off. Algorithms crunch the massive amount of RNA-seq data to reveal which genes are overactive or underactive in cancer compared with healthy tissue. The algorithms then use complex equations to infer patterns of gene interactions and zero in on the transcription factors with the largest influence.

The search for key drivers of cancer isn't easy. Consider a 2018 analysis of more than 9000 samples that reported almost 1.5 million mutations. Genes influence one another in intricate webs and feedback

"You have more potential combinations of cancerous mutations than atoms in the universe."

Andrea Califano,
Columbia University

loops, so the number of ways those genetic perturbations might interact in a tumor is vast. “There’s, say, 1000 genes that are recurrently mutated across all tumors that may drive cancer, so you have more potential combinations of cancerous mutations than atoms in the universe,” Califano says.

The pharmaceutical industry can’t make a new drug for each of those unique combinations. (By comparison, 126 new cancer drugs received approval from the Food and Drug Administration [FDA] between 1980 and 2018.) That’s why pinpointing the master regulators that are common culprits across cancers is so crucial, Califano says.

At Columbia, he has worked with his former postdoctoral researcher Mariano Alvarez to develop more efficient algorithms for sorting through those networks. The current one, called VIPER—short for virtual inference of protein activity by enriched regulon analysis—has been used in dozens of studies of how vast, interconnected genetic networks have gone awry in bladder, prostate, and lung cancers.

Califano and colleagues recently used the VIPER algorithm to look across RNA-seq data from more than 10,000 individual tumor samples in the Cancer Genome Atlas, a database sponsored by the U.S. government. The team found that different types of cancer have more in common than previously thought. The analysis, now under review for publication, identified 407 transcription factor genes that acted as suspected linchpins across all the cancer samples. Only 20 to 25 of them were implicated in any given cancer—and Califano says fighting the cancer might not require knocking out all those transcription factors: Toppling just a few nodes might be enough.

Califano “was among the first to put the complex algorithms out there, and then others have followed,” Liu says. A strength of Califano’s algorithms is that they look at an entire network of gene products, including RNA and proteins, adds David Tuveson, director of the Cold Spring Harbor Laboratory Cancer Center. Tuveson uses VIPER in his own search for treatments for pancreatic cancer.

Califano, too, hopes to put his algorithms to work for patients. The idea to commercialize that approach began in 2013, on a beach in the British Virgin Islands. There he met a fellow vacationer, Gideon Bosker, a physician who had gotten his start in emergency medicine and later launched a successful medical education company. The pair hit it off, and 2 years later they decided to form DarwinHealth. Columbia licensed the VIPER technology

to DarwinHealth, and Bosker put \$1.4 million of his money into the venture to get it off the ground. Since then, industry collaborators have sponsored more than a dozen research projects with the algorithms from the company.

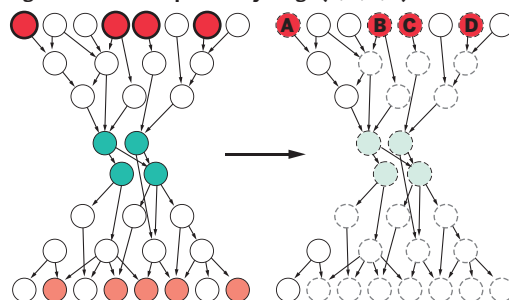
DarwinHealth combines Califano’s algorithms with a database of information from experiments about how drugs affect multiple genes, compiled through the company’s review of the literature and other sources. As long as they follow established rules

Hitting cancer’s choke points

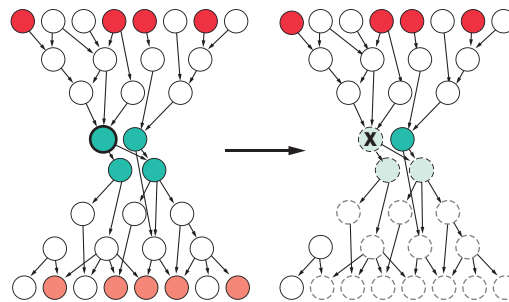
Shutting down a cancer cell’s malfunctioning gene network is a tall order if you target mutations too far upstream in the network. But disabling a transcription factor that acts as a master regulator in the cell’s genetic circuitry can cause its demise even with just one drug.

○ Gene ○ Corrected gene behavior ● Master regulator
● Initial mutation ● Malfunctioning gene

Targeting all mutations that can activate cancer’s master regulators would require many drugs (A, B, C, D).



A single drug (X) that targets a master regulator can have the same effect.



for patient tissue transfer, doctors around the world can now send a tissue sample to Columbia’s pathology department, where RNA is extracted from cells. For \$1600, the company generates an “OncoTarget” readout of the individual master regulator that seems to have the biggest sway in the patient’s cancer, as well as a more sophisticated “OncoTreat” readout of existing drugs that tamp down the tumor’s 25 most activated transcription factors and boost the 25 that are most turned down. The products launched in 2018.

GORDON MILLS, who serves as director of precision oncology for the Knight Cancer Institute at Oregon Health & Science University and helped pioneer the field of systems biology, notes that Califano’s cancer-fighting algorithms still have to overcome a lot of obstacles. “There’s skepticism that we’re far enough along to be able to predict the complexity of human disease,” says Mills, who has applied Califano’s algorithms to data in his own research. “There have been hundreds of algorithms that have failed to truly capture the complexity and heterogeneity of cancer and have not panned out in attempts to move them through to the clinic.”

But Califano sees a sign that his search for cancer linchpins will pay off in an unlikely cancer success story: thalidomide. In the mid-1950s, the drug arrived on the market as a sedative to help with sleep and anxiety. Doctors also prescribed it to pregnant women for nausea—which proved devastating because it caused massive birth defects, including missing limbs and heart problems. But thalidomide has made a comeback as a medicine for diseases such as leprosy. In 1997, doctors began to test the drug against multiple myeloma, a cancer affecting white blood cells.

Scientists have since learned more about how thalidomide works. In 2018, they found it prompts a protein complex called cereblon inside cells to mark certain transcription factors for disposal. In multiple myeloma, those transcription factors, IKZF1 and IKZF3, act as linchpins in the genetic network that allows the cancer to thrive. To Califano, thalidomide’s success shows the value of finding existing drugs that can target cancer’s master regulators.

Candidates are scarce. Whereas many drugs go after proteins that act as enzymes and have easy-to-find active sites to target, transcription factors lack such readily targetable spots, and many researchers have considered them undruggable.

But Califano’s Columbia lab is trying to add to the list of potential drugs. HULKING machines with robotic arms process tumor cell samples to do high-throughput screening that looks at how candidate drugs alter the cells’ RNA-seq profile and whether the drugs reverse the activity of master regulators. A \$12 million supercomputer in the basement of the building, a shared resource for university researchers, analyzes the data.

Bristol Myers Squibb, which manufactures the FDA-approved version of thalidomide for multiple myeloma, has joined the

hunt. It has contracted DarwinHealth to systematically search the pharmaceutical giant's library of compounds for other compounds that might target master regulators.

Additional support for the DarwinHealth approach comes from a recent study by Samir Parekh at the Icahn School of Medicine at Mount Sinai and a team of international collaborators, who recently completed a clinical trial to test a combination of two drugs, dexamethasone and selinexor, for multiple myeloma. The combination only worked in about one-quarter of patients, reducing levels of a myeloma protein in their blood. In a retrospective analysis, the DarwinHealth tools predicted which patients would respond. By assessing RNA-seq data from 12 patients, the tools identified four of the five patients who benefited from the drugs and six of the seven who did not, the researchers reported last year in *The New England Journal of Medicine*.

Morgan Craig, who uses computational approaches to identify new drugs at the University of Montreal, says efforts to understand molecular networks in cancer have the potential to improve personalized medicine. Algorithms like those used by DarwinHealth “may not take over clinical approaches right away,” Craig explains. “But it's definitely a step toward trying to do this target identification in a more systematic way.”

DARWINHEALTH DOESN'T RUN clinical trials, but for the past 3 years, Califano's lab has tested the company's algorithms in experiments at Columbia. The researchers analyzed RNA-seq data from biopsy samples from more than 100 cancer patients to identify master regulators and suggest drugs that might not normally be considered (much as DarwinHealth's commercial service does). In a few dozen cases, the researchers later tested the drug in mice with a grafted version of a patient's tumor to confirm the drug affected the master regulators as predicted. For five of those patients, doctors felt bold enough to try the algorithm's suggested drug. Each patient had late-stage cancer that had already stopped responding to available treatments.

Four of those five patients responded to the drugs given, at least for a time. For one patient with a meningioma, a tumor that can exert fatal pressure on the brain, the algorithm pointed to etoposide—a drug originally designed to treat lung and ovarian cancer. His tumor stopped growing for more than a year; it then started to slightly rebound and he was put in a different clinical trial. After that, his tumor started to rapidly grow again.

Califano hopes to build on those anecdotal results with the formal clinical trial, now underway at Columbia. The Oncotarget and Oncotreat tests from DarwinHealth will be used with 3000 patients in the Columbia system. Ultimately, the drugs they receive will be chosen by a board of doctors on the basis of readouts either from mutations detected by traditional sequencing or from the VIPER-based algorithm. DarwinHealth will receive no money for the tests to avoid

clinical studies will all focus on gastrointestinal tumors, particularly gastric and esophageal cancer,” Zhang says.

As the DarwinHealth approach goes into more clinical testing sites, more patients like Kapotes will receive drugs never intended to fight their particular cancers. For some people, like her, it might buy precious time. For more than 2 years after she enrolled in the ruxolitinib trial, Kapotes's cancer remained stable. When scans eventually showed her tumor had



Andrea Califano (right) and Gideon Bosker (left) founded DarwinHealth to apply systems biology to cancer.

any conflict of interest, given that Califano is part of the company's leadership while working at the university.

Califano and Bosker are also licensing the DarwinHealth tools to other researchers around the world to test against cancer. In January, Beijing Cancer Hospital confirmed it would be using DarwinHealth's tools to guide treatment for patients in clinical trials there. Xiaotian Zhang, an oncologist leading the new study, says that if early results look promising, the research will be expanded to other hospitals. “These

started to grow again, Tiersten switched her to another medication, which had just received FDA approval. These days, Kapotes is taking time to enjoy her retirement and her family. The newly approved drug she now takes works through a different mechanism, but Kapotes never would have had the chance to take it if not for ruxolitinib. “She hung on long enough because she was in the trial,” Tiersten says. ■

Roxanne Khamsi is a science journalist based in Montreal.

Science

Computing cancer's weak spots

Roxanne Khamsi

Science **368** (6496), 1174-1177.
DOI: 10.1126/science.368.6496.1174

ARTICLE TOOLS

<http://science.sciencemag.org/content/368/6496/1174>

RELATED CONTENT

<http://stm.sciencemag.org/content/scitransmed/9/392/eaal5148.full>
<http://stm.sciencemag.org/content/scitransmed/10/453/eaan0941.full>
<http://stm.sciencemag.org/content/scitransmed/5/205/205rv1.full>

PERMISSIONS

<http://www.sciencemag.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of Service](#)

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. The title *Science* is a registered trademark of AAAS.

Copyright © 2020 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works