Deconstructing the tenacity of cancer

How could your research help fight cancer?

The goal of my research is to find drugs that target harmful proteins in cancer cells. Processes in cells, such as growth and metabolism, are carried out by proteins, and cancer-causing mutations in genes can give rise to proteins that behave abnormally. I study how changes in protein function affect cellular processes and lead to cancer.

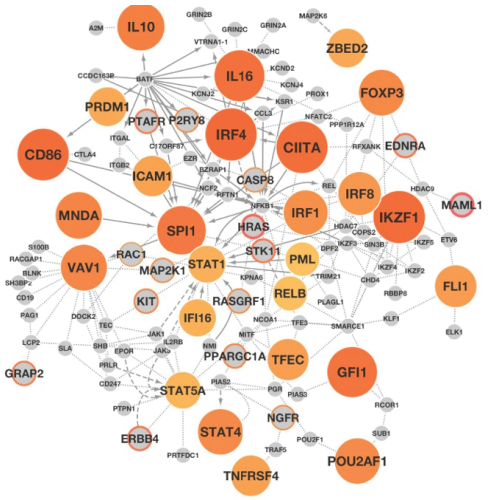
Evan Paull

As a Research Scientist and Bioinformatician in Systems Biology at Columbia University Medical Center, I design computer programs to find patterns among proteins from cancer patients to guide the discovery of cancer-fighting drugs. I work in the lab of Professor Andrea Califano where we use systems biology principles along with physics, computer science, and laboratory experiments to investigate proteins that interfere with cellular processes in cancer cells. Much of the data I use comes from The Cancer Genome Atlas (TCGA). TCGA researchers recently organized a "pan-cancer" dataset of genomic and clinical information on 10,000 patient tumors from 33 forms of cancer. The Pan-Cancer Atlas provides an unprecedented resource that scientists and doctors can leverage to understand how, where, and why tumors arise.

I implement machine learning algorithms to predict proteins that make tumors so deadly. In the Califano lab, we call these proteins "master regulators" because we have found that they drive cell processes from normal to abnormal during the development and maintenance of cancer. Machine learning algorithms use computational methods to "learn" information directly from data, and they perform better as the amount of data available for learning increases. This makes the Pan-Cancer Atlas an excellent dataset for my application.

One of my novel contributions has been to find groups of master regulators that form meaningful biological patterns called pan-cancer "checkpoints." For example, the "immune checkpoint" consists of about 30 master regulators that help cancer cells avoid detection and destruction by immune cells from tissue surrounding the tumor. Overall, I discovered 400 proteins that group into 24 pan-cancer checkpoints. Accounting for a large proportion of TCGA data with only 400 proteins is an astounding finding given that the number of human proteins is greater than 20,000!

We recently published my research as part of a large collaboration with scientists across the country. The network in the figure shows that the immune checkpoint proteins (circles) are highly connected and interactive (lines). I see big differences in survival between patients whose proteins form the immune checkpoint pattern versus those whose proteins do not. We hypothesize that disrupting the function of even one or a few of these proteins may collapse the ability of cancer cells to circumvent the immune response. A way to disrupt the function of a protein is to find drugs that bind to it in a way that prevents it from interacting with its network partners. Therefore, the immune checkpoint can guide us to identify such drugs and, importantly, to identify patients who are likely to benefit from the therapy. But what if a patient's proteins lack the immune checkpoint pattern? The immune checkpoint is only one of many ways cancer co-opts cellular processes. I compare each patient's network of proteins against all 24 checkpoints to elucidate that patient's particular pattern profile. A tremendous benefit of my approach lies in the possibility of developing personalized treatments for almost all cancer patients.



The immune checkpoint is represented by a "network" of protein interactions in cancer. Orange circles are master regulator proteins, such as IRF4, which plays a key role in leukemia. Gray circles are proteins that link master regulators. Many "linker proteins" are mutated (orange outline). For example, HRAS, a key protein in many processes, is mutated in 25% of cancers.

What sparked your interest in science?

I got my undergraduate degree in theoretical mathematics because I liked the creativity involved in constructing proofs and the process of working through the logic required to find an answer. Computational biology and bioinformatics are even more interesting because I apply that logical thinking to real data. In science, one starts with a mental model that states the world looks like A; the creative and analytical process consists of converting that into an experiment that will let the data decide if it really does look like A or if it's more like some alternative. The great thing about studying nature is that the answer is so often surprising. Biology, in particular, is so complex that we rarely get it right the first time. Studying how cells work is like peering into some alien technology that's thousands of years ahead of our own. It always leaves me amazed and curious to learn more.