THE AUTHOR FILE

Robert F. Murphy

Creating algorithms to turn images into cell models

At a Gordon conference in 1998, Robert F. Murphy presented data showing that automated analysis could recognize how proteins are arranged within a cell. The system had performed well. It classified 85% of patterns correctly, much better than the 10% expected from random guessing. Still, of the 20 or so scientists who came up to Murphy after the talk, 18 told him that automated systems would never work. "They said, 'You have to go to grad school in cell biology to know what a Golgi looks like.'"

Robert F. Murphy

That was typical of the early days of bioimage informatics, says Murphy, now head of the computational biology center at Carnegie Mellon University. "Most of that time was writing papers and giving talks to convince people of the validity of the approach," he says. "Far less time was spent actually doing it."

He remembers when attitudes started to change. Four years after the Gordon conference, he addressed a general cell biology meeting. Unknown to him, a program officer for the US National Institutes of Health (NIH) was in the audience. A few days later, the officer phoned Murphy to say that he had looked in the database of NIH-funded research and was disappointed Murphy's work was not supported. The officer wanted to let Murphy know that the NIH was reconsidering how grants were evaluated.

Eventually Murphy became the first chair of an NIH study section on biodata management and analysis. It was created because many proposals were being dismissed because innovations did not fit neatly into areas such as genetics or cell or molecular biology. "There was a realization that we needed study sections that could look not at primarily biological methods but computational methods," Murphy says. When the NIH began organizing *ad hoc* sections, he was asked to join in part because he'd been submitting so many proposals that fell between traditional topics.

Murphy first became interested in using computers to automate analysis while in graduate school at the California Institute of Technology. "It was a whole new world," says Murphy. "It was the kind of world I'd been looking for when I was deciding what to major in." As he moved on to postdoctoral studies, flow cytometers were becoming popular. These cell-sorting machines could measure multiple parameters on thousands of cells per second, and Murphy began writing programs to better understand cell populations. It was a good fit for his interests, he says. "There was a biological question and a computational problem for how to analyze that data."

But every time Murphy's team submitted their work for publication, reviewers asked for images of the cells. "I was struck by the fact that these images were not proof," he recalls. "There was an extremely subjective process by which people would say what was in those images." Murphy waited for other researchers to apply computational tools to images to support the kind of analysis he was doing with cytometry. "It didn't happen, and at a certain point, I decided it would be me." Since then Murphy has been working on automated ways to estimate protein content in organelles and other subcellular locations. Now there are multiple systems that do this, such as FarSight and CellProfiler, he explains.

In this issue, Murphy describes an image-based search system. The key factor is that the search is based on patterns, not words. For example, if scientists have an image of a marker that clusters in a certain way in the endoplasmic reticulum, they can search for images with similar patterns using the image database system OMERO. Such searches can work in situations where searches relying on written descriptions would fail, says Murphy. "What if there's a pattern and no one knows what to call it?"

> "What if there's a pattern and no one knows what to call it?"

Most recently, Murphy has been working on what he calls 'generative models', systems that do not simply recognize patterns but "construct *in silico* cells from images." This work is essential, he says, because biology has moved beyond a reductionist system in which scientists could readily know which experiments should come next. Now researchers are considering myriad interacting parts with models "too complex to build in your own head," he says. "We need some assistance. That's where computer models come in."

In fact, Murphy believes that biomedical research will come to be driven by automated modeling that can incorporate images. Images, after all, can provide richer and more complex data than many 'omics' technologies. "It's one of the last frontiers to be computerized," he says. "I see this as perhaps the dominant mode by which we study biological systems."

Monya Baker

Cho, B.H., Cao-Berg, I., Bakal, J.A. & Murphy, R.F. *Nat. Methods* **9**, 633–634 (2012).