Q&A: Stephen Friend on a Bioinformation Commons

Sharing data and tools is the only effective way to beat cancer, says Sage Bionetworks founder

After making the first discovery of a human tumor suppressor gene in 1986, Stephen Friend, MD, PhD, went on to teach at Harvard Medical School, cofound Rosetta InPharmatics and sell it to Merck for more than $620 million, and lead Merck’s basic cancer research efforts. In 2009, he founded Sage Bionetworks, a Seattle nonprofit that aims to rewire many of the rules by which data-intensive biomedical research is pursued. Friend talked recently with Cancer Discovery’s Eric Bender about how Sage hopes to speed cancer research.

What is Sage’s mission?

Sage Bionetworks has as its goal building better models of disease together. That has 2 subthemes.

One, we think that the current models for cancer are not resolved at a molecular level sufficient to drive personal therapy and don’t have enough granularity to show what targets are likely to be new avenues for therapeutics.

Two, the “together” part: we believe that there is no one institution, no one company, no one foundation that by itself can solve the cancer that it works on. When data have more eyeballs on them, when more people can work on them, we are more likely to get to higher-resolution maps of disease that evolve over time.

Over the next decades, we need to begin aggregating data and building models together that will increasingly have the resolving power to identify which patients should get what therapy, and what are the good combinations of targets. The approach we have taken for this very large task is to build a biomedical information commons incubator for doing pilot tests of the tools, various ways of working together, and reward structures.

Where are you starting?

We are focusing concurrently on 5 areas.

The first is genomic data. It’s very hard to broadly share the most valuable data because of issues around privacy. We are working with John Wilbanks to develop what’s called portable legal consent. Currently, consent forms are almost all based within the institution, the institution acts as the holder of the data, and the Institutional Review Board gives approvals. We need to look at alternatives that give some control to the patient in a dynamic way.

The second is that although groups such as the Avon Foundation and the Multiple Myeloma Foundation are building awesome sets of patients and data, for them to create their own cancer disease models, they need ways to share that data and work with tools that help them to build networks. In April, we piloted Bridge, a tool to do that.

Third, if someone at the University of California, San Francisco, is working with someone at Harvard to figure out a model for a pancreatic cancer, they’d like a place where they can leave the data and build models to share with each other prepublication. Synapse is a compute space in the “cloud” that we have launched that allows you to do this. It has pipelines for building the models, curating the data, and versioning the models. We’re trying to create a compute space where you can post the model and have people react there rather than publishing papers.

The benefit of that is all about speed. Today, we fund people in cancer biology to generate data, analyze it, and then validate the findings. Because people are developing their data to be published, they are hesitant to share it with people who are not on the team. Confirming findings requires 2 iterations of that cycle, which ends up taking on the order of 2 to 8 years. Other fields such as physics, astronomy, and synthetic biology have developed ways where credit can be given for posting, not publishing, findings. If credit is given for posting a hypothesis or posting a validation, then that 2- to 8-year cycle literally can be reduced to weeks.

Fourth is a group that is building new models of disease, examining questions such as what does it mean to add an activated RAF pathway to a molecular network.

And fifth, we’ve found that you have to develop rewards and incentives that encourage people to share. Academia is so steeped in the nonsharing of access that you’ve got to come up with incentives to get people to work together. In a project that we launched in April, we and IBM are hosting a challenge competition to generate models of breast cancer classifiers, using data from 2,000 patients recently published in Nature. A number of journals are willing to give the winners of such competitions an automatic publication because the challenge will be more rigorous than a peer review.

In general, we’re building examples that will scare people that they’re being left behind if they do things the way they always have.

In this hugely ambitious task, what worries you most?

I worry about complacency and a lack of urgency. If funding is not in question, people settle back into ways of not sharing and the reality of patients dying is not brought up enough. I also worry that we’ve got a bravado right now, that people working with next-generation sequencing (NGS) are conveying to politicians and funders that we’re on the verge of understanding everything. NGS and all our new tools are not so powerful that solo efforts will break through to cures for diseases. We really need to use these tools jointly, because the diseases are more complex than we’re imagining.