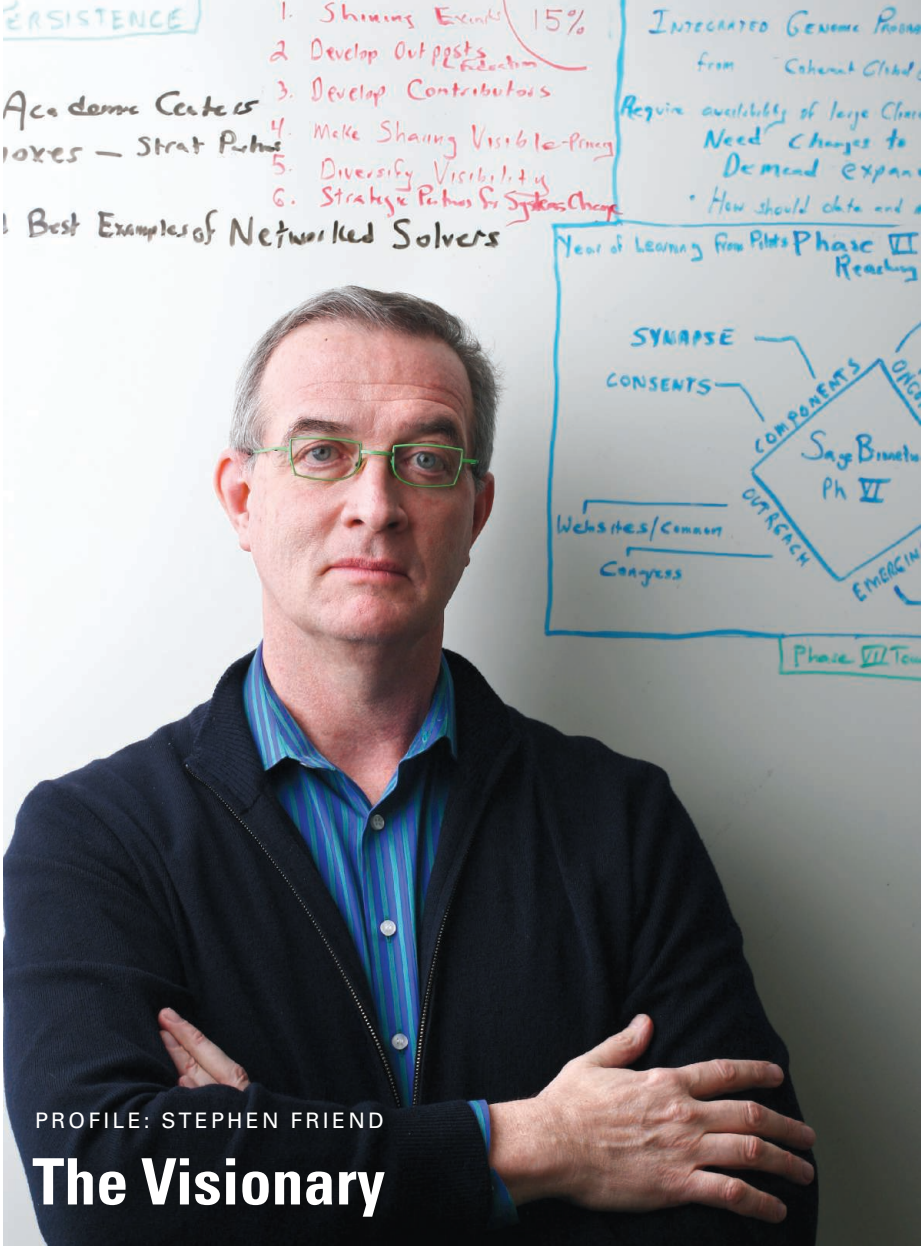


PROFILE: STEPHEN FRIEND

The Visionary

Seeking to spur drug development, Stephen Friend has launched a daring series of initiatives to make biomedical research more open and effective

It's late afternoon and everyone's heading home after an open-access conference in suburban Maryland where Stephen Friend has just delivered yet another of his signature stump speeches. In the past 10 days, his dizzying schedule included stops in Norway, New York City, Boston, and the nearby National Institutes of Health (NIH) in Bethesda, Maryland. But he's more than willing to hop on his soapbox again when a reporter asks. Most biomedical researchers are hunter-gatherers, holding their data close until they've published their findings, Friend says. But that won't lead to the development of new drugs or therapies, he insists: Untangling the biology of diseases such as cancer and diabetes requires finding patterns in enormous amounts of data, such as the torrent of information now pouring out of big genomics projects. Identifying the needles in such haystacks—culprit proteins or genes causing disease—demands a new kind of science. "The scale of and scope of the problem will need to be solved by sharing the data and networking."

Friend should know something about what's needed for biomedical science: He's treated children with cancer, helped discover a new class of cancer genes while an academic at Harvard University, and co-founded a biotech company that Merck bought for more than $600 million and, as part of the deal, hired him to head its cancer research. Now he's trying to do nothing less than change the competitive culture of science.

Three years ago, frustrated by the constraints of working in a big company, Friend negotiated a friendly separation from Merck and co-launched a nonprofit called Sage Bionetworks, based at the Fred Hutchinson Cancer Research Center in Seattle, Washington. Among its goals, Sage wants to persuade drug companies, academics, clinicians, and patients to share genomic and other biomedical information freely in a huge database. Other researchers would download and work on it together, mapping out the intracellular pathways that contribute to human diseases and building computational models of those conditions that would be better than the animal models now used in preclinical drug development. In short, Friend wants to bring to biomedical science the same open-source ethos embraced by the computer programmers who wrote the Linux operating system and by the people who contribute information to Wikipedia.

"What I realized was that drug discovery would continue to be consistently hampered by the lack of good models of disease. And to build those models was going to take massive amounts of data being shared over many iterations, over decades," says Friend, 58, who enlivens his usual semicasual attire—sport coat, button-down shirt, no tie—with green metal-rimmed glasses.

Friend isn't the only researcher who wants colleagues to take a more open approach to biomedical problems, but he is one of the most driven and persuasive. Thanks to his legendary persistence, dozens of big names in biomedical research have signed on to various Sage projects. "He's able to galvanize people to do things," says cancer biologist Robert Weinberg of the Whitehead Institute for Biomedical Research in Cambridge, Massachusetts, who once trained Friend and has followed his career. In his new position, Friend has also forged ties beyond the scientific community: He works with open-access activists on data-sharing rules, gets free cloud-computing space from Amazon, and collaborates with major patient advocacy groups.

Even some of the scientists who have signed on, however, are skeptical. Several point out that systems biology, the disease-modeling approach that Sage promotes, is still a young and evolving field. Others are confused by Sage itself, which Friend admits is "big, ill-defined," and has

"Stephen has a very idealistic vision of how we as scientists in the bioarena should be working together and why that hasn't happened."

—ERIC SCHADT, MOUNT SINAI MEDICAL CENTER

CREDIT: JOHN LOK

Published by AAAS

yet to hit a home run that might clarify its purpose.

But many believe Friend is on to something—or at least trust his track record enough to give him the benefit of the doubt. "He's an incredible innovator. He's always at the forefront of new things," says William Chin, executive dean for research at Harvard Medical School in Boston and a former senior vice president at Eli Lilly. Friend is a "machine on ideas," adds Eric Schadt, a longtime scientific collaborator. "Stephen has a very idealistic vision of how we as scientists in the bioarena should be working together and why that hasn't happened."

A change in philosophy

Friend didn't expect to become a scientist; he majored in philosophy in college, veered toward medical ethics, and wound up in medical school and graduate school as part of an M.D.-Ph.D. program. As a pediatric oncologist, Friend was moved by the patients he saw, including a boy and his father who had both lost an eye to a rare inherited cancer called retinoblastoma. Frustrated by how little was known about the disease, he accepted a postdoc position in Weinberg's lab with the goal of finding the gene behind retinoblastoma.

One small problem: "He had no idea how to do it," Weinberg says—Friend's doctorate was in biophysics. "I taught him, basically," says René Bernards of the Netherlands Cancer Institute, a fellow postdoc in the Weinberg lab. Yet within 2 years, Friend was first author of a *Nature* paper on the cloning of *RB1*, the first tumor suppressor gene—a major milestone in cancer research.

At the time, Bernards was struck by Friend's sense of direction. "He said he would divide his life into three parts: basic science, drug development, and art collecting," Bernards recalls. "I thought he was bullshitting." (Friend now collects art and creates it himself, working on a new project each year with artists in the Seattle area.)

After starting his own lab at the Harvard-affiliated Massachusetts General Hospital, Friend later moved to Seattle to work with geneticist Leland Hartwell of the Hutchinson Center, who later shared the Nobel Prize for work on the cell cycle. They eventually teamed up with University of Washington, Seattle, immunologist Leroy Hood, one of the early pioneers of systems biology, to make inexpensive gene expression arrays—chiplike devices that were becoming increasingly popular for measuring gene activity—and look for differences in gene expression in healthy and diseased tissue. That led the three to found the biotech company Rosetta Inpharmatics and, in collaboration with Bernards, to develop the first commercial gene expression test, MammaPrint, which was approved by U.S. regulators in 2007 for predicting whether breast cancer will recur.

In 1999, Friend hired Schadt, a young, maverick mathematician, to expand Rosetta's bioinformatics efforts. Schadt worked on combining DNA mutation data with gene expression levels in particular tissues from healthy animals or people and those with a disease. Massive computations using a supercomputer revealed networks of genetic interactions that pointed to genes that play a key role in the disease. These could then become potential targets for drugs or indicate which patients would not benefit from a treatment. In an influential 2005 *Nature Genetics* paper, for example, the Schadt team used their approach to find a network of genes active in obese mice.

Hoping to use Rosetta's gene expression analysis for drug development, Merck bought the company in 2001 for $620 million, netting Friend over $10 million. The company also made Friend its director of oncology drug development. The network analysis yielded a string of papers in top journals and new drug candidates for diseases such as diabetes and heart disease.

Friend and Schadt saw the next step as taking their network analyses from animals—the source of much of their data—into humans. But because people are much more genetically diverse than strains of mice, and diseases such as depression come in many subtypes, Friend realized they would need a wide variety of data from tens or hundreds of thousands of patients to reach the needed statistical strength for human studies. "It became apparent that you would need these megascale analyses," he says. So Friend began several Merck collaborations with academic cancer centers to build massive databases, eventually to be open to all, that could be used to identify which subsets of cancer patients might benefit from a treatment.

Friend says he and Schadt realized, however, that "no one company" could do what was needed. So 3 years ago, when Merck decided to close down much of Rosetta as part of a companywide downsizing, Friend, according to Hartwell, now at Arizona State University in Tempe, "talked Merck into letting him" take Rosetta's bioinformatics division and an estimated $150 million in computers, software, data, and intellectual property and create Sage Bionetworks.

An uncommon approach

Sage, which has grown from 15 to 35 staff members and has a modest $5 million budget, has some characteristics of a conventional nonprofit research organization. Its team of computational biologists have NIH and foundation funding and collaborate with academics and companies on network modeling studies. Schadt, now at Pacific Biosciences in Menlo Park, California, and Mount Sinai Medical Center in New York City, is one such partner. And Sage has inked a deal with Pfizer to use its modeling approach to develop cancer drugs.

But Sage is also part think tank, one with a grand vision of creating what Friend calls a "commons" where researchers come together to share and analyze vast amounts of biomedical data. At Sage's invitation-only annual meetings, participants refine their ideas for breaking down cultural barriers. For example, Sage and outside experts are working on data-sharing rules that Friend calls "a set of principles about what is good behavior, what is bad behavior."

Indeed, those who collaborate with Sage, including companies, agree to share data and models online within a year of the end of a project. The mandatory policy is reminiscent of the so-called Bermuda rules developed to compel DNA sequencing centers to release their data within a reasonable period, says Robert Cook-Deegan of the Duke Institute for Genome Sciences & Policy in Durham, North Carolina, who collaborates with Sage. "I think the way they're doing it is really smart."

As scaffolding for the commons, Sage is building an online repository for data sets. It differs from existing databases such as

FRIEND'S PRESCRIPTION FOR A BIOMEDICAL REVOLUTION

1 Synapse
Online platform for sharing data and disease models

2 Federation
Pilot studies on collaborative creation of disease models

3 Portable Legal Consent
People in clinical studies control who uses their data

4 Clinical Trials Comparator Arms
Shared genomic data on people in control arms of trials

5 Arch2POCM
Open-access drug development through midstage trials

Published by AAAS

NIH's GenBank because the data sets are more extensively curated, Friend says, and the repository will also host the software underlying their computer models of diseases. That is important and "much more complicated than genomic data," says David Haussler of the University of California (UC), Santa Cruz, who led the way in developing tools for sharing human genome data. "I applaud Sage for taking on this task."

But building the repository has gone slowly. When Sage staffers set out to gather 100 or so data sets that they had their eye on, such as gene expression and proteomics data from cancer patients, more than half were not available or were missing key underlying data such as annotations describing how and when clinical measurements were collected. Often the consent forms signed by patients did not allow even data stripped of personal identifiers to be shared publicly. "I've been disappointed in how difficult sharing data is," says Sage staffer Lara Mangravite. Sage is now working on a tiered access system to its gathered data sets that would comply with privacy rules.

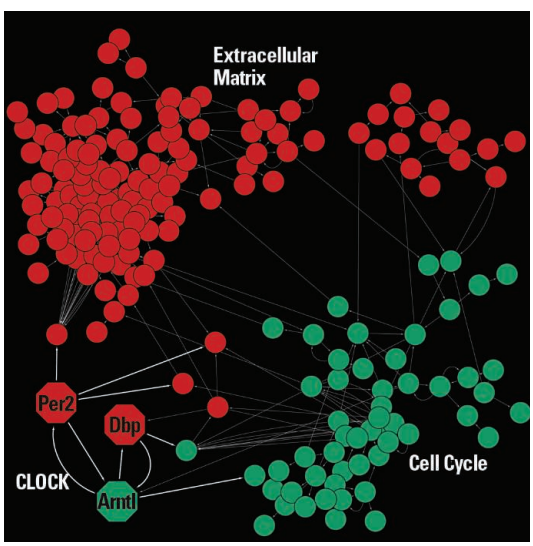
Data-sharing frustrations have driven another project that Sage supports: an online consent form that allows people to control which scientific studies use their genomic and health data. "It's a shift from giving data to institutions where they really can't share it to shifting control to the patient," Friend says.

The project's leader is Sage board member John Wilbanks, formerly with Creative Commons, the nonprofit group that works to eliminate copyright issues and other factors limiting the exchange of information. Friend is convinced that only "patient-oriented, citizen-oriented projects" will generate enough data to build disease models complete enough to help suggest or test treatments. Volunteers are now testing a draft version of the form, and several patient groups will begin using it this spring.

Sage is also experimenting with a version of "networked" science. Friend recruited five research teams to work together on "superhard problems" in aging, cancer, and diabetes, such as finding genes driving the Warburg effect, a metabolic process used by tumor cells. "It was a pilot to see whether science got done faster and whether people felt as though their ideas were benefiting by sharing them or whether they felt robbed," Friend says. After more than a year of work, the so-called Federation has submitted two papers to journals—one, for example, found patterns in methyl groups attached to DNA that correlate with a person's age. Friend says the result "would have taken much longer and may never have come about" without the Federation.

But Stanford University bioinformatics researcher Atul Butte, a Federation member, says that although members "learned to trust each other," the lack of specific funding for the collaboration slowed progress. "The jury is out on whether we did more than we could have as individual labs collaborating" with an NIH grant, he says.

A successor project will use a new software platform under development at Sage to make it easier to build disease models together in real time the way software engineers now do, Friend says. Other efforts at Sage aim to coax data out of drug companies. One that is furthest along invites companies to contribute genetic and clinical data on people with a disease who were enrolled in the control arms of drug clinical trials. The companies have little incentive to keep these data proprietary, yet they could be a gold mine as a shared resource for studying mechanisms of disease, Friend says. Seven companies initially agreed to submit data sets, and two, GlaxoSmithKline and Johnson & Johnson, have begun to do so, but the rest realized that the consent forms signed by the patients wouldn't allow the companies to share the data. That experience, too, has underscored the need for a new kind of consent, Friend says.



Muscular model. Sage wants to spur modeling of cell networks, such as this one showing how in muscle cells, circadian clock genes modulate expression of other genes, such as those involved in the cell cycle and extracellular matrix (red indicates genes with increased activity, green decreased).

CREDIT: JONATHAN DERRY/SAGE BIONETWORKS

Climbing a big hill

Friend's most ambitious project attempts to speed drug development and reduce its financial risk to companies by bringing companies and academics together for the initial testing of potential drugs. Dubbed Arch2POCM and led by Friend, Aled Edwards of the University of Toronto, and Chas Bountra of the University of Oxford, participants will freely share compounds and data until a potential drug has shown safety and efficacy in a phase II clinical trial. (The unwieldy acronym stands for Archipelago to Proof of Clinical Mechanism, with the initial word symbolizing groups collaborating.) The reasoning is that companies don't need intellectual property protection up to that point because they will want to modify and improve any compounds that work in this initial efficacy trial, says Harvard's Chin, who is not directly involved. But compared with other public-private partnerships, Arch2POCM would reach much further into the drug-development pipeline. "It's a very good idea" but "risky," Chin says.

Arch2POCM's leaders are talking to five companies and hope to have commitments by midsummer for projects involving drug candidates for cancer, autism, and schizophrenia. "Stephen has a big hill to climb," says molecular biologist Keith Yamamoto of UC San Francisco, who is helping run Arch2POCM. "But I think we're at a stage where we need to be trying new experiments."

Some researchers who aren't disease modelers themselves but would contribute data to Sage caution that its systems biology approach is still experimental. Cardiologist Eric Topol, director of the Scripps Translational Science Institute in San Diego, California, says that until one of the drugs Merck found using network analysis reaches the market, the approach "isn't validated yet." Still, Topol calls systems biology "increasingly important" in human genomics. And although the diagrams in a disease-modeling paper can be bewildering to a clinician, he expects "that the juice we'll get out of it will be useful."

If that happens, Friend may have already moved on to something else equally ambitious. Sage's mission is deliberately fuzzy so that it can "evolve," he says. "We are not under the illusion that our solution has to be what happens." And claiming success for specific projects is not the objective but rather being a "catalyst for others doing it," Friend says. Once projects are set up, "our hope is that Sage drops out of the picture, and in 5 to 10 years no one knows what Sage is."

–JOCELYN KAISER

Published by AAAS