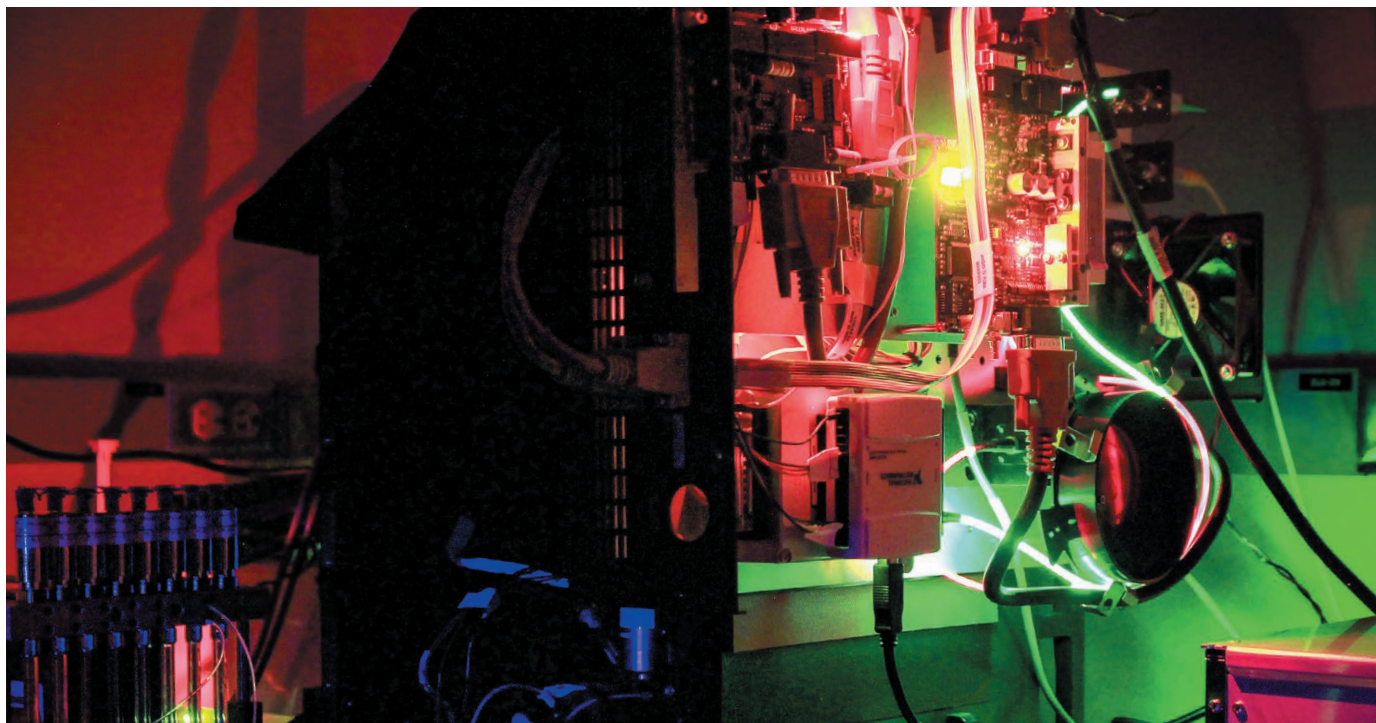


## TECHNOLOGY FEATURE

# HOW TO TEACH AN OLD SEQUENCER NEW TRICKS

*Outdated DNA-sequencing machines need not die — researchers can repurpose them to drive next-generation biochemistry studies.*

CURTIS LAYTON



This next-generation biochemistry platform, dubbed 'Serenity', was assembled from the components of an old Illumina DNA sequencer.

BY JEFFREY M. PERKEL

In a basement storeroom at Stanford University in California, the guts of a dozen DNA sequencers lie exposed — hundreds of thousands of dollars worth of cameras and lasers, optics and fluid controllers, all scavenged from a late-model, next-generation Illumina DNA sequencer called GAIIX. On the floor, the shell of one old instrument sits empty, picked over like a carcass. “I seem like a hoarder,” says Stanford biophysicist William Greenleaf.

But over the past 6 years, this collection has fuelled an effort that has engaged about half of Greenleaf’s 18-member lab team. Whereas most researchers use DNA sequencers to, well, sequence DNA, Greenleaf’s team is one of a small number that has repurposed the devices for an entirely different goal: to study protein and nucleic-acid biochemistry on a massive scale, from macromolecular interactions and

RNA folding to enzyme function.

“It’s a revolutionary technology,” says Stanford biochemist Dan Herschlag, who uses it to study interactions between RNAs and other molecules. It provides “deep and broad quantitative information”, he says, “that allows researchers to build more-precise biophysical and cellular models for molecular interactions, and which is also a critical step towards a truly predictive understanding of biological systems”.

Broadly, the work demonstrates what’s possible when scientists look into the guts of their hardware — proof that equipment isn’t necessarily without value just because it is old or outdated.

But there’s a reason such technology development is called bleeding edge: things often go wrong. Sarah Denny, a biophysicist who graduated from Greenleaf’s lab this year, chuckles when asked whether her equipment offers ‘plug-and-play’ simplicity. “Many times

when you did an experiment, something would break and you’d have to figure out how to get it to work again,” she says. But given the volume of data she could extract, the reward was worth the pain. In Denny’s case, her team gained a better understanding of RNA folding. Such is life in the do-it-yourself trenches.

### BIOPHYSICS ON A CHIP

When the GAIIX dropped in 2008, it was a hot commodity. Some sequencing centres had dozens of instruments, costing about US\$600,000 apiece; in 1 week, a machine could push out 30 billion of the lettered bases that make up DNA. But by 2011, when Greenleaf established his lab at Stanford, the industry had migrated to faster hardware — such as the Illumina HiSeq 2000 — that was more efficient and user friendly, and people were giving their old machines away. “They’re basically big paperweights,” says Greenleaf. ►

► Illumina sequencers automate a sequencing-by-synthesis process. A DNA library is randomly arrayed on a device called a flow cell, and amplified in place to create small clusters of about 1,000 molecules, each representing a single fragment of genetic information. The building blocks of DNA, or nucleotides, are then transferred onto a chip, each containing a unique fluorescent signature and a reversible chemical modification. This process ensures that only a single nucleotide can be added at each cluster. The sequencer then images the array, and on the basis of the colour at each position, ‘calls’ or reads, the base that was added. The modification is then removed, and the process repeats, allowing the entire sequence to be identified base by base.

### THE BASICS OF HACKING

At their core, these instruments are high-end microscopes coupled with liquid handlers that help to move reagents around. Some of their components — particularly cameras, lasers and movable stages — can cost tens of thousands of dollars. In 2009, Christopher Burge, an RNA biologist at the Massachusetts Institute of Technology in Cambridge, realized that it might be possible to repurpose that hardware to do something else.

“The GAIi is basically a pumping system that pumps things onto the flow cell and then onto a fancy imaging system,” Burge says. “We realized that, well, maybe you could pump other things onto the flow cell.”

That, he says, is because the GAIi was an ‘open system’, controlled using editable configuration files called recipes and loaded up with reagents that could be changed simply by substituting one tube for another. On the inside, the machine was equally open, with off-the-shelf, third-party components held together with cable ties. “In retrospect, it looks like a high-school science project,” says Gary Schroth, a biochemist who directs the genomics-application group at Illumina in San Diego, California, and who collaborated with Burge on his early studies.

New Illumina instruments, by contrast, are more polished, with custom hardware, hardwired control software and barcoded reagents — features that improve the user experience but preclude hacking.

Jacob Tome helped to hack a GAIi as a graduate student in the lab of molecular biologist John Lis at Cornell University in Ithaca, New York. He recalls “a lot of panicked phone calls” with Illumina representatives, trying to work out the intricacies of the company’s instrument-control software. “There’s no manual to reprogram the instrument,” he says, so working out the logic was largely trial and error.

“One of the most exciting weeks of my PhD was when we just kind of sat there at the sequencer, and we’d write a recipe and then watch the temperature go up and watch what solution it was pumping,” Tome says.

Burge, however, had help. Working with

Schroth, his team modified the GAIi recipes to accept new reagents, introduce pauses, adjust the running temperature and alter the imaging parameters. The researchers then applied that modified system to conduct a comprehensive study<sup>1</sup> of the sequencing preferences of a DNA-binding protein in yeast called Gcn4p.

“We showed for the first time that you could turn a sequencer into a biophysical instrument to measure protein–DNA interactions at very high throughput,” Burge says.

To do that, Burge’s team treated the flow cell — which is usually discarded after sequencing — to regenerate double-stranded DNA at each cluster. They then introduced fluorescently tagged Gcn4p protein to the chip at progressively higher concentrations, and used modified software to trick the sequencer into imaging the flow cell as if it had just completed another round of sequencing chemistry. By downloading and processing the final images, rather than the base calls the instrument typically outputs, the team could measure how much Gcn4p was bound at each place. From that, the group could deduce its affinity for every sequence on the chip — some 440 million measurements in all.

“That was a really beautiful, elegant experiment,” says Greenleaf. It showed that the GAIi platform had the potential to provide “all the sorts of measurements I was excited about: kinetic measurements, on-rates and off-rates and equilibrium constants, all the physical measurements that one might need to understand from first principles how DNA–protein interactions work”.

Several researchers have used similar hacks to address questions of their own. At the Massachusetts Institute of Technology, for instance, RNA biologist David Bartel looked at messenger RNA (mRNA), the RNA copy of a gene that is used to produce proteins. His team found a link<sup>2</sup> between the number of ‘A’ bases strung at the end of an mRNA molecule — a feature known as a polyadenylated tail — and how efficiently proteins are produced from genes in animal development.

Ilya Finkelstein, a biophysicist at the University of Texas at Austin, hacked a sequencer flow cell to investigate<sup>3</sup> the DNA-binding preferences of a protein called Cascade, used in one version of the gene-editing technique CRISPR–Cas. And some members of Greenleaf’s team worked with CRISPR pioneer Jennifer Doudna at the University of California, Berkeley, to study<sup>4</sup> why Cas9 — the most commonly used enzyme in CRISPR — sometimes cuts at the wrong place in the DNA.

Researchers have also worked out methods for transcribing on-chip DNA to create ultra-high-throughput RNA arrays — an approach that could be used to study RNA–protein interactions or to screen libraries of folded RNA molecules called aptamers.

The challenge of making such an array is capturing the transcribed RNA. Lis<sup>5</sup> and Greenleaf<sup>6</sup> independently introduced a physical block at the end of double-stranded DNA molecules, stalling RNA polymerase — the enzyme that creates the mRNA molecule — with its dangling transcript. At Weill Cornell Medical College in New York City, RNA biologist Samie Jaffrey’s team — led by chemical biologist Nina Svensen, then a postdoc in the lab — used<sup>7</sup> a special, viral RNA polymerase to chemically link the newly synthesized RNA to the chip, and then added a nuclease to degrade the template DNA.

### HACKERS AGAINST DISEASE

For Stanford biophysicist Rhiju Das, such RNA arrays have served as platforms for identifying molecules that can diagnose active tuberculosis. The project began when a colleague discovered<sup>8</sup> a signature of three RNAs that can distinguish between dormant and active tuberculosis. The challenge was to pinpoint a single molecule — a ‘riboswitch’ — that could actually identify which state the disease was in. Such a diagnostic molecule would have to be able to bind all three RNAs, plus a fluorescent reporter, and change shape depending on which ones were present. “It’s totally bananas, obviously,” Das says.

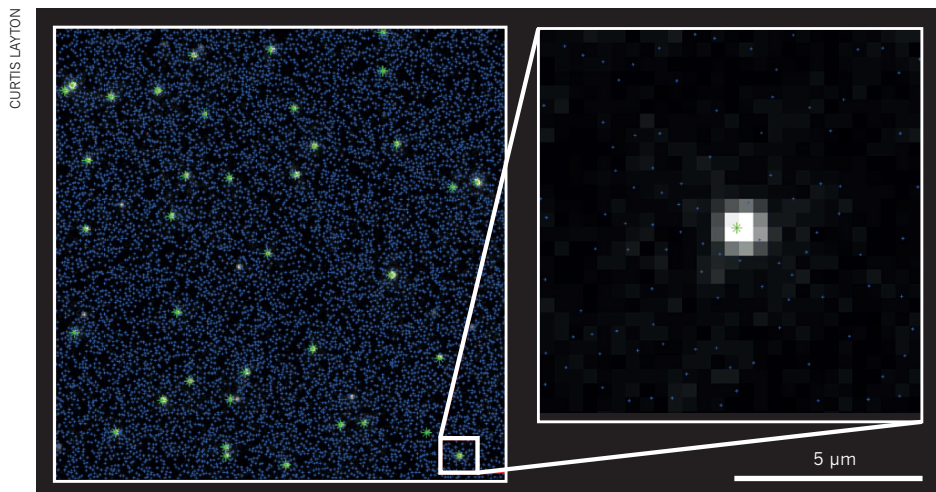
To solve the puzzle, Das recruited the user-base of an online game he co-developed, called eterna. Players are tasked with designing RNAs that can fold in certain ways; the most promising structures get synthesized as part of a DNA-sequencing library, 10,000 at a time, and tested on the GAIi chip. With 10,000 measurements per design, that’s about 100 million data points from each experiment — data that are fed back to the players to hone their skills. Over time, some users have become quite adept, Das says. “These players have now discovered principles that let them get perfect riboswitches, thermodynamically optimal riboswitches for a variety of test cases.”

And that includes the OpenTB Challenge. Over 3 rounds of game play, Das’s team has tested nearly 27,000 riboswitches from 187 players, identifying several that seem to fit the bill. One, submitted by a retired engineer in Falmouth, Massachusetts, looks like the floor plan of a castle in one state, and like an anchor in the other. Now, with funding from the Bill & Melinda Gates Foundation, Das plans to begin testing some of these designs for use as a pregnancy-test-like diagnostic for tuberculosis.

Meanwhile, a few scientists have taken the next logical step, translating on-chip RNAs into protein. In 2016, Svensen showed<sup>7</sup> she could synthesize short protein sequences called peptides on the array, by exploiting the antibiotic puromycin to tether the peptide to the surface. And in June this year, Greenleaf-lab postdoc Curtis Layton reported<sup>9</sup> developing an assay called Protein display on a Massively Parallel Array to profile 156,140 variants of an enzyme called SNAP-tag. He wanted to learn the subtle rules that relate amino-acid sequences to protein function. “It’s been so exciting to see just a

**“When it actually works, boy, it really cooks. Nobody can generate data like that.”**





Fluorescence imaging of a sequencing flow cell reveals binding interactions at specific clusters.

whole new scale of protein functional analysis that we're able to do on these machines," he says.

Exciting, but not easy: Layton spent four or five years developing his method and hardware. Svensen spent two years developing her method, which Jaffrey's lab is now applying to create a whole-human-proteome chip. "To buy a sequencing kit for the GAIIX back then was \$5,000. You can't really make many rounds of optimization if you have to spend that for every experiment you do," she says. As a result, Svensen, now at the Wellcome Centre for Anti-Infectives Research at the University of Dundee, UK, did much of her testing on microscope slides, or by using discarded flow cells from her department's core facility.

### A SPACE ODYSSEY

Most GAIIX hackers use the machine largely as is, tweaking the instrument software to make it do what they want. According to Jaffrey, that makes it accessible to anyone with an old GAIIX lying around. "We don't want something only an engineer can use," he says. "Our protocols can be used by any researcher with that machine."

And for its initial RNA work on the GAIIX, the Greenleaf lab was no different. Donated by Stanford geneticist Patrick Brown, the system used in that work, Layton says, was named in honour of the *Star Wars* spacecraft Millennium Falcon — specifically, pilot Han Solo's plea to the ship: "You hear me, baby? Hold together!"

"It's this old machine that we're still trying to run for something it wasn't intended to do," Layton says. "Just don't give us any errors, don't give us any problems. But, when it actually works, boy, it really cooks. Nobody can generate data like that."

Ultimately, the team decided the approach was unsustainable. For one thing, says Layton, unlike modern sequencers with their push-button interfaces, the GAIIX was less Millennium Falcon and more "like flying an Apollo spacecraft". Newer sequencers are also faster, and produce longer reads. Plus, Illumina announced in 2015 that it would be 'end-of-life-ing' the GAIIX, meaning that flow cells, reagents and parts

would no longer be available as of July 2017.

The team extracted the system's optics, stage, camera, pumping system and lasers; supplemented the sequencer's built-in illumination system — which lights up the sample at an acute angle from below — with downward-facing widefield illumination; designed a new printed circuit board to operate the lasers; cobbled an autosampler out of GAIIX parts to automate reagent addition; and added an on-board temperature controller.

Most significantly, the team designed new sample holders, so the GAIIX imaging system could accommodate Illumina's new, smaller MiSeq chips. "That gives us the ability to decouple our biophysical measurements of fluorescence and the sequencing," Greenleaf says. (Finkelstein also runs his assays on MiSeq chips, imaging them on a conventional microscope.)

The result is a device to warm an engineer's heart: a compact black cube adorned with what looks like a ripped-open desktop computer on one face. The hardest part, says Layton, was coding the software required to control the hardware, and working out which signals would drive what action. "That was a Herculean effort," he says. But overall, the project appealed to his hobbyist interests in engineering and software design, not to mention biochemistry. Winston Becker, a dual PhD and medical student in the lab, with a master's degree in engineering mechanics, was drawn to the project for similar reasons. "I wanted to do something very physical, very quantitative. And obviously an opportunity to use instrumentation skills and build things was really exciting, too."

But wouldn't it have been easier to just buy a microscope instead? According to Greenleaf, that wasn't really an option — or at least, not an affordable one. "There's really nothing out there," he says, "nothing that will do the sorts of things that we want to do." A custom instrument, he estimates, would probably have cost hundreds of thousands of dollars.

Das, who built his own instrument off Greenleaf's design (and with Greenleaf-lab assistance), says the experience was "so fun".

But with no how-to manual, "you have to be willing to go bug people," he says. Still, vexing issues can arise. Das's instrument wouldn't focus properly, for instance. After two weeks of troubleshooting, Das realized he was missing a lens. "It was so dumb," he says. (He has since compiled a 64-page instruction manual. Another researcher, Nick Kaplinsky at Swarthmore College in Pennsylvania, has published detailed blog commentary (see [go.nature.com/2l4fdme](http://go.nature.com/2l4fdme)) describing his work turning a GAIIX into 'RootScope', a microscope for imaging plant roots.)

Sticking with the tradition of naming instruments after fictional starships, Das christened his instrument Red Dwarf, after the British science-fiction series; this team then added a sister ship, Nostromo (*Alien*). Other Greenleaf imagers include Heart of Gold (*Hitchhiker's Guide to the Galaxy*), Borg Cube (*Star Trek: The Next Generation*) and Serenity (*Firefly*). "If you've built an instrument, you will get the glory of the name," Greenleaf says. His lab actually has a dozen sequencers in various states of disassembly throughout the building, most of them acquired for the cost of shipping.

If things go according to plan, that fleet will become ever more integral to Greenleaf's work. His lab is perhaps most famous for developing ATAC-seq<sup>10</sup>, a method for assessing how accessible chromatin is to the proteins that bind it, and half the lab is dedicated to that work. Now, Greenleaf hopes to merge the teams, using his hacked sequencers to build mathematical models that accurately reflect macromolecular interactions in the cell itself.

The question is, will other researchers follow suit? Schroth says that between the reagents, the hardware and the analysis tools, these assays are so customized that they could prove difficult to reproduce in other labs, at least without a knowledgeable insider to help.

But there's an easy fix, hackers say: open up modern platforms to exploration. "Then the scientific community could push the instruments in new directions that could benefit everybody," says Burge. Finkelstein says: "There's still a lot of space to play with these sequencers off-label, if you will." ■

Jeffrey M. Perkel is technology editor for Nature.

1. Nutiu, R. *et al. Nature Biotechnol.* **29**, 659–664 (2011).
2. Subtelny, A. O., Eichhorn, S. W., Chen, G. R., Sive, H. & Bartel, D. P. *Nature* **508**, 66–71 (2014).
3. Jung, C. *et al. Cell* **170**, 35–47 (2017).
4. Boyle, E. A. *et al. Proc. Natl Acad. Sci. USA* **114**, 5461–5466 (2017).
5. Tome, J. M. *et al. Nature Meth.* **11**, 683–688 (2014).
6. Buenrostro, J. *et al. Nature Biotechnol.* **32**, 562–568 (2014).
7. Svensen, N., Peersen, O. B. & Jaffrey, S. R. *ChemBioChem* **17**, 1628–1635 (2016).
8. Sweeney, T. E., Braviak, L., Tato, C. M. & Khatri, P. *Lancet Respir. Med.* **4**, 213–224 (2016).
9. Layton, C. J., McMahon, P. L. & Greenleaf, W. J. Preprint at bioRxiv <https://doi.org/10.1101/342808> (2018).
10. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. *Nature Meth.* **10**, 1213–1218 (2013).

**CORRECTION**

The referencing in the Technology feature 'How to teach an old sequencer new tricks' (*Nature* **559**, 643–645; 2018) was incorrect. The correct version can be found online at [go.nature.com/2wmnhgc](https://go.nature.com/2wmnhgc).