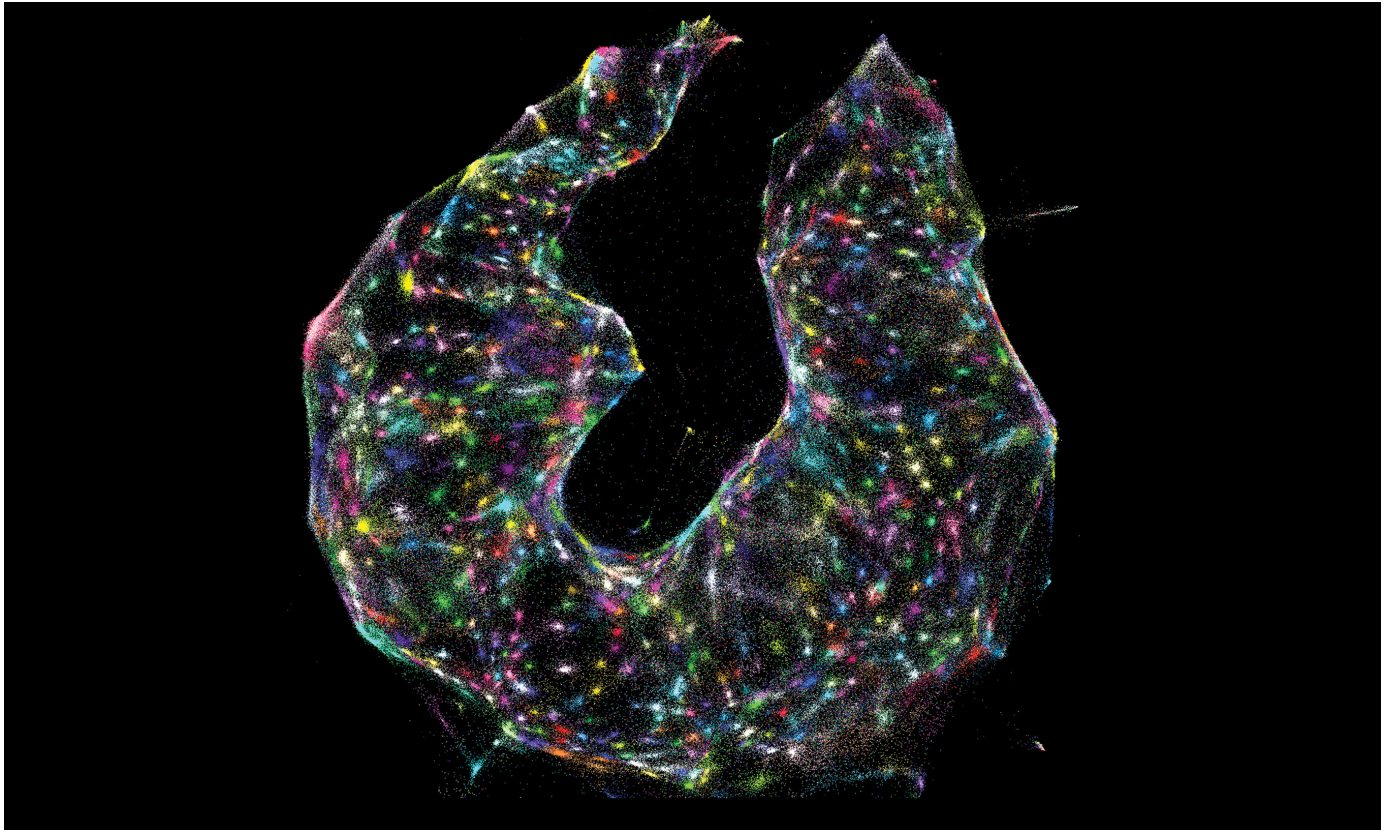


TECHNOLOGY FEATURE

STARFISH ENTERPRISE: RNA GOES SPATIAL

A fast-growing toolkit adds spatial detail to single-cell transcriptomics.

JOSHUA WEINSTEIN/BROAD INST.



Each point is an individual RNA molecule, localized by its proximity to other RNAs. This imaging method is called DNA microscopy, for its use of DNA sequencing.

BY JEFFREY M. PERKEL

For cinephiles, *Space Jam* was a 1996 comedy film pitting cartoon character Bugs Bunny and basketball player Michael Jordan against animated aliens. For neuroscientist Ed Lein, it was the name of a bioinformatics-themed meet-up — a type of ‘hackathon’.

In April, around 40 computational and transcriptional biologists turned up at the Allen Institute for Brain Science in Seattle, Washington, where Lein works. They came for coffee, coding and a common goal: to work out the strengths, weaknesses and analytical challenges of the growing methodological toolset known as *in situ* (or spatial) transcriptomics.

In situ transcriptomics is an alphabet soup of technologies — methods include MERFISH,

seqFISH+, STARmap and FISSEQ — for mapping the gene-expression patterns of cells in their tissue context. Some rely on hybridization — the ability of short nucleic-acid probes to find their complements in the crowded cellular environment — whereas others are based on DNA sequencing. But all produce conceptually similar data — gene-expression values matched to the *x* and *y* coordinates of a cell.

Such data can reveal intercellular relationships that might otherwise be overlooked, such as which cells are talking to which, and their position relative to structural features and cells of interest. As Aviv Regev, a computational and systems biologist, and founding co-chair of the Human Cell Atlas (HCA) project at the Broad Institute of MIT and Harvard in Cambridge, Massachusetts, puts it: “Tell me who your neighbour is and I’ll tell you who you are.”

But so rapid is the field’s growth that researchers might struggle to decide which methods to use. And the plethora of data-analysis algorithms, pipelines and file formats can make it challenging to analyse and compare data. “The state of the field has been one of rampant technology development,” Lein says.

With funding from philanthropic organization the Chan Zuckerberg Initiative (CZI) and under the auspices of the HCA, Lein and others formed a research consortium in 2017 to benchmark the different methods, called SpaceTx — short for spatial transcriptomics. At the same time, programmers at the CZI began building a unified data-analysis tool and file format, called Starfish, to advance the HCA’s efforts and aid the wider transcriptional-biology community. (The name “is a bit of a joke”, explains Jeremy Freeman, who ►

► directs computational-biology efforts at the CZI in Redwood City, California. Many spatial methods rely on FISH, or fluorescence *in situ* hybridization. In programming, an asterisk or star indicates a wildcard. “The joke is that they’re all ‘something-FISH.’”

Starfish is an open-source software suite that can read image files, register and remove the noise from pictures, find spots and identify the RNA molecules that they represent in nine different experimental strategies, with two more in development. The Space Jam event, Lein says, was an effort to bring developers and users — the spatial-transcriptomics specialists themselves — together to talk shop, troubleshoot and advance their methods. In so doing, the team exposed the subtle differences that can trip up those who want, for instance, to compare data across experiments. But it also provided a model for how to navigate a fast-growing technology.

IN SITU TRANSCRIPTOMICS

Researchers studying gene expression have usually done so at the bulk level, extracting RNA from a piece of tissue and then analysing it in its entirety. Over the past decade, single-cell methods such as Drop-seq have allowed researchers to probe the differences between cells at the expense of spatial detail.

That’s where *in situ* transcriptomics comes in. These techniques use mostly fluorescence microscopy and DNA sequencing to reveal the presence and abundance of RNA molecules in cells within the tissues themselves. From there, researchers can work out the types of cell that are present, their spatial arrangement and their relationships to one another.

It’s like a selection of fruity desserts, Regev says. “If all bulk genomics is the fruit smoothie, then single-cell genomics is the fruit salad, and spatial genomics is the fruit tart,” she explains. “If you look at a fruit tart from the top, all the fruits are organized in these really beautiful patterns.”

Depending on the method, such data can resemble stars in a pitch-black sky, or colourful works of art. One study led by Simone Codeluppi, a bioimage informatician in the laboratory of Sten Linnarsson at the Karolinska Institute in Stockholm, for instance, used a cyclic variant of single-molecule FISH, called osmFISH (pronounced ‘awesome fish’), to map the architecture of the mouse somatosensory cortex. The result was an image of the cells coloured on the basis of their gene-expression patterns, a picture that is reminiscent of a stained-glass window¹.

But such data can also reveal insights. At the University of Cambridge, UK, neurobiologist and physician David Rowitch has used a method called RNAscope to study the spatial diversity and organization of astrocytes in the mouse brain². Astrocytes, Rowitch found, “adopt layer patterns in the cortex similar to, but out of register with, neurons”. Long Cai, who studies single-cell biology at the California

Institute of Technology in Pasadena, and his team used a strategy called seqFISH+ to identify transcripts encoding interacting proteins on the surfaces of adjacent cells³.

PROVIDING CLARITY

Both seqFISH+ and RNAscope rely on nucleic-acid hybridization; they leverage short, fluorescently labelled molecules to light up their target sequences in the cell. Other methods use DNA sequencing or even mass spectrometry (see ‘Alphabet soup’).

More than a dozen spatial-transcriptomics methods have been described, including six in 2019 (refs 3–8). They differ in the number of RNAs that they can detect, their spatial resolution and the number of cells they can probe, but all provide the spatial localization detail that single-cell transcriptomics cannot. But spatial methods have shortcomings too, says Regev. Microscopy, for instance, is slow (sometimes involving weeks of continuous imaging), expensive and technically demanding. Many methods can access only a predefined fraction of the cellular transcriptome, and practical considerations can limit the number of cells that can be probed.

In trying to choose the right method for their work, researchers could become overwhelmed. SpaceTx aims to provide some clarity.

The project was funded as part of some US\$100 million that the CZI has spent over the past two-and-a-half years on the HCA and ancillary projects, a CZI spokesperson says. Each team — there were 19 in all — applied its own method to identical samples of human and mouse brain, which were prepared at the Allen Institute. Now Lein and his colleagues, as well as the broader computational-biology community, are crunching the numbers to see how the methods compare, and which is best for a given set of circumstances.

“This is actually quite unusual,” Lein says. Normally, researchers work to develop the best method, publish and move on. But with SpaceTx, “we’re trying to bring everyone together and say, these methods are all useful, but we need to understand what you can use each method for, and how they really quantitatively compare to one another.”

But doing that presents a computational problem, because different methods produce different data types. Some hybridization-based methods, for instance, assign each transcript a different colour, whereas others use multiple colours as a barcode. Some labs identify RNAs by tracking fluorescent spots in each image and then monitoring their intensity between imaging rounds, whereas others measure intensity at every pixel, correlating those intensity values with the list of possible barcodes to determine whether an RNA was present. How those

images are organized on disk, and the ‘metadata’ used to annotate them, can also vary.

Such incompatibilities can stymie research, says Matthew Green, a bioinformatician at the European Bioinformatics Institute in Hinxton, UK. Even if they do not, researchers often struggle to install their colleagues’ analysis software (thanks to the complex computational requirements and dependencies such software entails). And the sheer volume of data that spatial studies produce can be intimidating — Linnarsson’s automated osmFISH rig churns out 2 terabytes of images per day, he says; for SpaceTx, his team produced some 25 terabytes.

SPACE JAM

A team of computational biologists and software engineers led by Deep Ganguli and Ambrose Carr at the CZI set out to create a standard file format and pipeline for *in situ* transcriptome analysis — a way to mix and match different computational and wet-lab methods, whether on a laptop or in the cloud. The team even went on the road, visiting labs and talking to bioinformaticians to understand their workflows. “At least one of the graduate students in one of the labs told Deep: ‘This is so wonderful, because no one’s ever looked at my code before,’” Freeman says.

April’s hackathon gave the Starfish team a chance to let biologists take the software for a test drive, providing an opportunity for researchers and coders to learn from each other face-to-face, rather than through bug reports on the code-sharing platform GitHub.

“We were able to help all of those folks get their data-processing pipelines implemented in Starfish to help with their scientific efforts,” says Justin Kiggins, who leads Starfish development at the CZI. “And it gave our team critical insight into the gaps and challenges.”

Matt Cai, a bioengineering graduate student at the University of California, San Diego, who developed a method called DARTFISH, says he had two goals at the Seattle hackathon: to exchange ideas with other spatial-transcriptomics groups, and to get up to speed with Starfish. “We have our own in-house analysis methods, but they’re not written in a way that’s easy for people to use,” he says. “Starfish is being written for the scientific community.”

For Green, the meeting was unlike any other he had been to. Although he’s attended multiple conferences over the years, “I’ve never been to a first meeting,” he says. “Literally every conversation was like a massive exchange of information. And it felt quite exciting.”

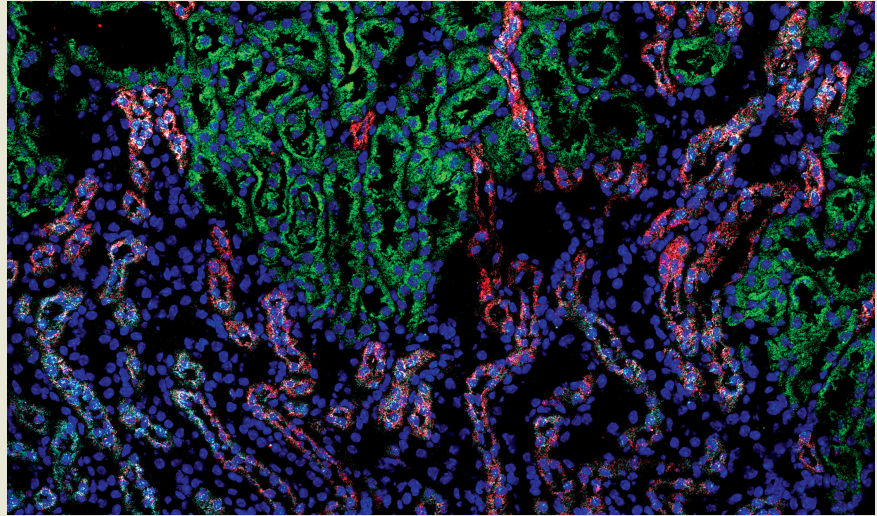
Every team has successfully converted a sample data set into the Starfish format, Lein says, and data generation is ongoing. But the software itself remains a work in progress. Aleksandra Tarkowska, a programmer at the Wellcome Sanger Institute in Hinxton, UK, says she was unable to convert her data sets into the Starfish format and align different fields of view into a unified image, owing to “the complexity of the data”. And Nico Pierson,

“Literally every conversation was like a massive exchange of information. And it felt quite exciting.”

Alphabet soup

At least a dozen *in situ* transcriptomics methods have been described, including:

- **APEX-Seq.** Unless a cell is specially stained, it appears featureless under a microscope. It is thus difficult to determine where a particular RNA is located. APEX-Seq localizes the enzyme APEX2 to a specific cellular ‘address’ and uses it to tag nearby RNAs. By isolating and sequencing those RNAs, researchers can profile the transcriptomes of individual cellular domains⁶.
- **DNA microscopy.** Blending molecular and computational techniques, DNA microscopy infers each molecule’s position on the basis of its neighbours, like creating a map of the United States using the coverage of radio-station transmitters. “We’re taking a biomolecular sample, and turning every single RNA into a radio tower,” says lead developer Joshua Weinstein, a postdoctoral researcher in Aviv Regev’s lab at the Broad Institute of MIT and Harvard in Cambridge, Massachusetts. RNAs in intact tissue are amplified in place, creating ever-larger ‘diffusion clouds’ of nucleic acid. As the clouds come into contact with neighbouring ones, a unique signature is created, which researchers can then ‘read’ using DNA sequencing to recreate the sample’s molecular architecture⁷.
- **IMC.** Imaging mass cytometry, normally used for mapping proteins in the cell, can also be used to pinpoint a handful of RNAs. The method blends a technique called RNAscope with mass spectrometry to reveal such things as growth-factor signalling by immune cells, which cannot be detected by protein-based approaches alone⁹.



An image of the mouse kidney, captured using fluorescent probes that attach to single molecules.

- **INSTA-Seq.** *In situ* transcriptome accessibility sequencing is a variant of fluorescent *in situ* sequencing, or FISSEQ. The method uses sequencing-by-ligation to identify short barcodes of RNA molecules *in situ*; those RNAs are then extracted and sequenced again using proprietary Illumina chemistry to read their full length. Because the synthesis step required to produce those longer reads can be blocked by factors such as binding to proteins or other RNA molecules, the method can provide insight into the ‘spatial epitranscriptome’, says molecular geneticist Je H. Lee at Cold Spring Harbor Laboratory in New York, who developed the method⁸.
- **RNA scope.** Commercialized by Advanced Cell Diagnostics in Newark, California, RNAscope is an *in situ* hybridization-based approach that uses signal amplification to boost the brightness of each target RNA. Twelve RNA species can be differentiated over three imaging rounds.
- **seqFISH+.** Sequential FISH+ combines fluorescent barcodes, a ‘pseudocolouring’ scheme and multiple rounds of hybridization to ‘dilute’ cellular RNAs and make them easier to resolve. Up to 10,000 different RNAs can be detected in each cell.
- **Slide-seq & HDST.** Tissue samples on arrays of spatially resolved, barcoded beads allow each cell’s RNAs to be associated with a cellular ‘postcode’. Slide-seq⁵ uses 10-micrometre beads (small enough to resolve a one-cell-thick feature of a mouse brain). High-density spatial transcriptomics⁴ uses 2-µm beads for subcellular resolution.
- **STARmap.** Spatially resolved transcript amplicon readout mapping is a blend of tissue-clearing technology, RNA amplification and DNA sequencing that can identify up to 1,020 RNA species in otherwise-opaque tissues. Each RNA is assigned a five-base gene-identification barcode, which is read out using sequencing-by-ligation¹⁰. **J.M.P.**

a software engineer in Long Cai’s lab, reported issues with the software’s ‘spot decoder’, the algorithm that matches fluorescence patterns to barcodes, because it was unable to handle the density of seqFISH+ data. “With our data, the efficiency is very low, it’s probably only 10%,” he says.

Still, attendees praised the event for getting programmers and biologists talking. The programmers came away with a pile of bug reports and feature requests, some of which could be solved on the spot. And researchers returned to their labs armed with new and better ideas for data analysis. Codeluppi, for instance, discovered a ‘segmentation’ strategy for computationally identifying cell boundaries in his image data, particularly for small-volume cells.

Matt Cai says his lab now routinely runs Starfish alongside its own computational

pipeline to compare performance. But others might be reluctant to abandon the in-house pipelines they have so meticulously crafted. Starfish could, therefore, find its biggest adoption among labs that are trying to implement the methods that others have developed.

“With all the different approaches, it’s really valuable to have something that can tie them all together in a computational format, and Starfish I think will allow that to happen,” says Abbas Rizvi, a molecular neuroscientist at the Zuckerman Mind Brain Behavior Institute at Columbia University in New York City. Rizvi is a member of the HCA project who is building an atlas of the human spinal cord using, in part, spatial methods.

“It reminds me of the earliest stages of single-cell transcriptomics,” he says. “It was tough enough to get the experiments to work, but it was

also kind of exciting to look at the data and to try to find ways to extract real meaning from them. And that’s where I see the field right now.” ■

Jeffrey M. Perkel is technology editor at Nature.

1. Codeluppi, S. *et al. Nature Methods* **15**, 932–935 (2018).
2. Bayraktar, O. A. *et al.* Preprint at bioRxiv <https://doi.org/10.1101/432104> (2018).
3. Eng, C.-H. L. *et al. Nature* **568**, 235–239 (2019).
4. Vickovic, S. *et al.* Preprint at bioRxiv <https://doi.org/10.1101/563338> (2019).
5. Rodrigues, S. G. *et al. Science* **363**, 1463–1467 (2019).
6. Fazal, F. M. *et al. Cell* **178**, 473–490 (2019).
7. Weinstein, J. A., Regev, A. & Zhang, F. *Cell* **178**, 229–241 (2019).
8. Fürth, D., Hatini, V. & Lee, J. H. Preprint at bioRxiv <https://doi.org/10.1101/722819> (2019).
9. Schulz, D. *et al. Cell Syst.* **6**, 25–36 (2018).
10. Wang, X. *et al. Science* **361**, eaat5691 (2018).