SEQUENCING

The big RNA picture

Researchers sequence RNA directly within tissue.

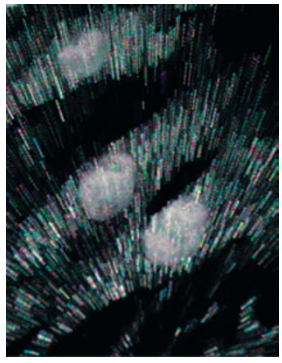
What's in a pixel? It is a question that George Church and his team at Harvard University have been asking. The Church lab is not an imaging lab; rather, it is well known for pioneering modern sequencing methods. But lately, researchers in the lab have been using fluorescence *in situ* sequencing (FISSEQ) to load each pixel of an imaged tissue with sequence information, expanding the range of biological questions that scientists can pose in a spatial context.

Current methods such as *in situ* hybridization and ligation-based *in situ* sequencing with 'padlock probes' are effective at pinning sequences to a cellular address, but they capture only targeted sequences. An unbiased survey of transcripts within tissue has been elusive.

FISSEQ debuted in 2003, based on generating polymerase colonies of copied DNA molecules on a glass slide coated with polyacrylamide matrix. The matrix prevents products from diffusing from their source and mixing signals, allowing many templates to be queried in parallel and forming the basis for high-throughput 'next-generation' DNA sequencing two years later.

Moving FISSEQ into tissue was shelved until 2008, when postdoc Je Hyuk Lee joined the lab to study allele-specific expression—differences in the activity of gene copies inherited from one's mother and father. "We knew that there had to be spatial mosaicism in heritable gene expression," says Lee. Sequencing in tissue could expose these spatial differences, but it meant overcoming major hurdles in both creating and imaging sequencing libraries.

To create FISSEQ libraries, researchers reverse-transcribe RNA in fixed cells with short random primers, and the resulting DNA is then circularized and subjected to rolling-circle amplification to create 'nanoballs' containing around 1,000 copies of concatenated template. The DNA is synthesized using a modified base that can be chemically cross-linked to fix the products securely in the tissue. The result is a three-dimensional structured library that can be sequenced by ligating fluorescence-labeled nucleotides and imaging each incorporation with a confocal microscope.



J.H. Lee

RNA barcodes can now be sequenced *in situ*.

The cell is a far noisier place than a sequencing lane: autofluorescence can be strong, and it varies among tissues. "Sometimes it's impossible to quantitate, when you have so much debris, what's a dot and what's not," says Lee. Realizing that subjective intensity thresholds lose or misinterpret signal, the researchers decided to call a base from every pixel or voxel (three-dimensional pixel). For noise filtering, the sequence of colors at a single cluster, regardless of intensity, can be compared to known transcript sequences because the chance of a spurious match vanishes as sequence length increases.

Another problem is signal density. Lee points out that regardless of the microscope, the presence of up to 200,000 mRNA molecules per cell means that you will reach a resolution limit. The team's solution was to develop 'partition sequencing', in which primers are lengthened by one, two or three bases to drastically reduce the number of targets that are primed in each set of sequencing cycles.

Yet another issue is imaging at scales that reveal biologically relevant expression patterns. Images from different cycles need to match at the pixel level, making it difficult to scan across different fields of view. Going wide field also makes the method slow and requires a dedicated four-color microscope. Applying faster imaging techniques such as spinning disk microscopy may dramatically improve speed.

The researchers used FISSEQ to sequence RNA in human stem cells, skin cells subjected to wounding, whole fruit fly and mouse embryos, and adult mouse brains. They ran up to 30 cycles, enough to uniquely identify over 4,000 genes in a single skin cell sample. Results largely correlated with traditional RNA sequencing across most RNA classes but were blind to some species that may be less physically accessible. FISSEQ captures an estimated 0.1% of a given cell's RNA. There is clearly a long way to go, but, fortunately, most cell-type–specific transcripts were sequenced, which suggests that FISSEQ can be used to classify tissues by gene expression.

The potential for *in situ* sequencing is enormous. The researchers are interested in the spatial distribution of allele-specific expression and post-transcriptional differences in RNA. They are sequencing differentiating neuroprogenitors and have set their sights on using FISSEQ to map gene activity onto neural connectivity in brain tissue.

With the ability to conjugate nucleotide barcodes to other molecules in the cell, Church and Lee are also thinking well beyond RNA. "It's really about tagging individual pixels to track all the objects," says Lee. "With a barcode length of just 20 bases, you can identify 1 trillion unique cells or proteins or any other thing." FISSEQ can be combined with immunostaining, or RNA-barcoded membrane proteins can be used to highlight cell outlines in sequencing libraries, for example.

"You can write a book in a single pixel if you have long enough sequence information," says Lee.

Tal Nawy

RESEARCH PAPERS
Lee, J.H. *et al.* Highly multiplexed subcellular RNA sequencing *in situ*. *Science* **343**,1360–1363 (2014).