GENOMICS

Spatial transcriptomics

By capturing and barcoding RNA in its native tissue location, researchers can visualize and quantify gene expression *in situ*.

The project started at the cross-section of two disciplines. Jonas Frisén from the Karolinska Institute, who was using microscopy to study neurogenesis, joined forces with Joakim Lundeberg, head of the Genomics Core Facility at the Science for Life Laboratory (SciLifeLab) in Sweden, to bring together the worlds of imaging and sequencing.

Traditional RNA sequencing experiments provide quantitative information on expression levels but lose spatial information. The goal of Frisén and Lundeberg was to retain this information, not only for a few transcripts but for close to the entire transcriptome. The idea behind spatial transcriptomics was conceptually simple, but the approach proved challenging in practice. The protocol begins with easy steps: a tissue is sectioned and immobilized on a chip prior to staining and imaging. The tissue is then permeabilized to release RNA, which is captured by poly-dT oligos immobilized on the array. Once bound to the chip, the RNA is reverse-transcribed and imaged or sequenced. It is the capture step that was difficult to perfect. "If you permeabilize too long," says Lundeberg, "you lose the morphology and then you lose the purpose of the technology."

After a lot of optimization, Patrik Ståhl from the Frisén lab and Fredrik Salmén from Lundeberg's group, who spearheaded the project, found optimal conditions for several tissues. To prove the validity of their approach, the researchers used fluorescent nucleotides during cDNA synthesis to monitor the position of the RNA, and they observed the same tissue structure that is revealed in a histological stain with hematoxylin and eosin. "That proved that we could capture mRNA without any diffusion," recalls Lundeberg. The researchers compared sections from different tumors and saw different areas of gene activity.

To generate a permanent spatial record of the mRNA molecules, the team incorporated positional barcodes and unique molecular tags into their capture oligos, reverse-transcribed the RNA and cleaved it off the array for amplification and sequencing. The correlation with

EPIGENETICS

Tracing cell lineage with 5hmC

Single-cell 5hmC sequencing uncovers cell-to-cell differences in the two DNA strands of a given chromosome.

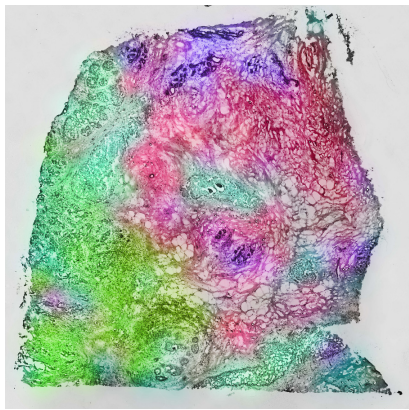
It is no secret that DNA methylation, the addition of a 5′ methyl group to cytosine, is important for the regulation of gene expression and development. But the roles of the 5′ hydroxymethylcytosine (5hmC) mark on DNA are far less well understood. Now, a method by Alexander van Oudenaarden and his team at the Hubrecht Institute in the Netherlands enables genome-wide 5hmC sequencing at the single-cell level. The approach is not just a step forward in the elucidation of 5hmC functions; it also enabled the researchers to demonstrate that differences in 5hmC can be used to trace cell lineages.

The discovery a few years ago that 5-methylcytosine (5mC) can be converted to 5hmC by proteins of the TET family sparked much interest among epigeneticists. Methods to sequence the 5hmC modification at individual target sites and at the whole-genome level soon followed. Whether 5hmC encodes information or is simply an intermediate in the removal of 5mC is not entirely resolved, but it does appear to have important roles in gene regulation and pluripotency. Technologies enabling single-cell transcriptomics and genomics have made possible important discoveries in a range of fields, including cellular heterogeneity and differentiation. In epigenetics, methods for bisulfite sequencing of 5mC sites at the single-cell level have been described, but achieving this resolution with 5hmC sequencing had not been possible.

To tackle this challenge, van Oudenaarden and his team adapted a method for bulk 5hmC sequencing. The approach is based on AbaSI, a restriction endonuclease that recognizes glucosylated 5hmC sites and cuts 11–13 nucleotides downstream. Glucosylation with T4 phage β-glucosyltransferase followed by AbaSI treatment results

standard RNA-seq was high, and the spatial barcodes allowed them to trace each transcript back to its spatial origin in the tissue section. Comparison of gene expression in spatially defined domains in brain sections revealed genes specific for certain regions and allowed the researchers to compare adjacent sections, as well as the same section in different animals.

The resolution of spatial transcriptomics is limited by the density of features (areas with the same spatial barcode), which currently stands at 100 micrometers. If they could get this density down to 10 micrometers, says Lundeberg, they would have single-cell resolution. Efforts are under way to increase the resolution.



J. Lundeberg

Gene expression heterogeneity in breast cancer: clustered spatial transcriptomic data overlap with morphological features (violet, normal mammary glands; red, fat tissue; green, cancerous regions).

Lundeberg is happy about the convergence of disciplines, which, he says, was the idea behind the founding of the SciLifeLab in 2010. He hopes to use the technology to compare protein maps, partly generated at SciLifeLab as part of the Human Protein Atlas, to a molecular atlas.

Another goal is to get a 3D view of cancer. The researchers are taking sequential sections through a tumor to determine its spatial transcriptome. The heterogeneity they see between sections underscores how difficult it is to get a representative example of a tumor and how much is missed by sequencing of pooled RNA from the entire tumor. "Down the line," says Lundeberg, "I hope that this can be a complement to histology in clinical labs."

Nicole Rusk

RESEARCH PAPERS

Ståhl, P.L. *et al*. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78–81 (2016).

in two-nucleotide overhangs. The researchers used the overhangs to add a cell-specific barcode, a sequencing adaptor and a T7 promoter. The T7 promoter enabled them to amplify the DNA fragments in a strand-specific orientation by *in vitro* transcription, and they then sequenced the RNA library to yield the desired readout.

The researchers applied their method (scAba-seq) to study mouse embryonic stem cells and corroborate that their analysis is representative of what had previously been observed in bulk. But looking at single cells showed what bulk 5hmC sequencing could only hint at: direct evidence of substantial differences from cell to cell in the number of 5hmC sites between the two strands of a chromosome. Using mathematical modeling, van Oudenaarden and colleagues discovered that this strand bias represents strand age—the difference between a 5hmC-modified strand from the mother cell and a mirror version, without any 5hmC marks, that is newly synthesized before cell division. The new strand gradually accumulates 5hmC modifications until both strands arrive at steady state. Further evidence of this included sharp transitions in strand bias, a consequence of sister chromatid exchange during the G2 phase of the cell cycle. The results suggested that 5hmC strand bias might contain sufficient information about previous cell divisions to enable reconstruction of cellular lineages. Indeed, the scientists traced cell origins in two-cell and four-cell mouse embryos.

Our current understanding of the biological implications of epigenetic marks is already informing therapeutic approaches in cancer and beyond. If applied to fundamental and disease-related questions of gene regulation and cellular differentiation, single-cell 5hmC sequencing will enable important new discoveries.

Irene Jarchum

RESEARCH PAPERS

Mooijman, D. *et al.* Single-cell 5hmC sequencing reveals chromosome-wide cell-to-cell variability and enables lineage reconstruction. *Nat. Biotechnol.* **34**, 852–856 (2016).