

## GENE EXPRESSION

# Splice and sequence

**A ligation-based sequencing method reveals distant exon connections within long RNA molecules.**

Down syndrome cell adhesion molecule (*Dscam*) is encoded by a complicated gene. By shuffling exons around, the locus can generate over 38,000 different transcript isoforms in the fruit fly, each encoding a unique protein that will help axons to chart a course in the developing brain. *Dscam* may be an Olympian of alternative splicing, but it highlights the potential, with contributions from multiple promoters and polyadenylation sites, to derive enormous transcript diversity from a single genomic locus. In fact, less than 5% of human genes produce only a single transcript, and most generate more than 10.

Research on isoform regulation has been hindered by transcript lengths. *Dscam* transcripts span over 7 kilobases, for example, whereas short-read sequencing is limited to around 500 bases. Single-molecule sequencing can read many kilobases of DNA, but for RNA, it relies

on reverse transcription, which is inefficient on long stretches. Shared sequences between isoforms also promote errors due to template switching, situations in which reverse transcriptase jumps between template molecules while generating cDNA.

To better probe exon connections in distant parts of the same molecule, Melissa Moore and Phillip Zamore at the University of Massachusetts Medical School and their colleagues have developed the SeqZip method. SeqZip links together short single-strand DNA fragments called 'ligamers' that hybridize to the ends of exons or exon blocks in RNA, allowing the internal portions to be ignored during subsequent sequencing. It works a little like splicing: the most informative parts of a transcript are stitched together, whereas constitutive exons or internal exon sequences are looped out.

The researchers first identified a ligase that actively joins pieces of DNA hybridized to RNA but that does not act on free DNA or DNA-

DNA hybrids and avoids linking ligamers from different templates. They used it to accurately quantify exon choice in cell lines expressing different versions of the mouse gene encoding fibronectin. SeqZip interrogates only one or a few genes at a time, and many alternatively spliced genes lack long constitutive regions that can be looped out. But the method avoids reverse transcription and template switching, and it works well on very long genes across a large range of expression levels.

As the ultimate test of SeqZip, the researchers targeted every exon in three different *Dscam1* exon clusters, finding that isoform diversity increases from younger to older fly embryos and that exon choice is largely independent between clusters.

**Tal Nawy**

**RESEARCH PAPERS**

Roy, C.K. *et al.* Assessing long-distance RNA sequence connectivity via RNA-templated DNA-DNA ligation. *eLife* doi:10.7554/eLife.03700 (13 April 2015).