Assume Nothing: The Tale of Circular RNA

By now, you’d think the research community would stop underestimating RNA. After all, RNA—once considered the boring middle child of molecular biology—is full of surprises.

First, there was the discovery of mRNA splicing. Then came long noncoding RNA. And RNAi. And microRNA. And piwi-interacting RNA. So when researchers started to see that not only could RNA assume yet another form but that this form was also incredibly widespread, you’d think journal editors and scientists would be ready ... but you would be wrong.

In December 2012, Norman (“Ned”) Sharpless, Deputy Director of the Lineberger Comprehensive Cancer Center at the University of North Carolina School of Medicine, published an article in the journal RNA showing that human fibroblasts contain more than 25,000 circular RNA species arising from 14.4% of the expressed genes. That study was not the first to demonstrate the existence of RNA circles. As far back as the 1980s, researchers have observed the occasional ribonucleic acid ring, including hepatitis delta virus RNA and transcripts of the sex-determining Sry gene in mammals. However, most of these observations were dismissed as splicing accidents or artifacts of EST libraries. In the past few years, however, a growing body of research, based both on bioinformatic and biochemical data, suggested that circular RNAs were perhaps more abundant than researchers thought—that is, if they remembered RNA circles existed at all.

Still, circular RNAs appeared to be exceedingly rare. But Sharpless' 2012 data implied that circular RNAs were everywhere, representing a substantial fraction of a cell’s transcriptional output. The response from peer reviewers was less than enthusiastic. “I’ve been running my lab now for 12 years,” Sharpless says. “I’ve published in Cell, Nature, NEJM ... and I’ve never had reviews like this. They were nasty. They were really mean.”

Sharpless submitted his article to five different journals prior to its publication in RNA, but not before Patrick Brown's team at Stanford University beat him to the punch with a similar study in PLoS ONE. Only three months later, in February 2013, back-to-back manuscripts in the journal Nature officially put circular RNAs on the map (and made headline writers’ day): “Circular RNAs throw genetics for a loop,” wrote Nature. “A circuitous route to noncoding RNA,” was Science’s take.

The funny thing is, circular RNAs were hiding in plain sight the whole time. So why were they missed for so long? In a word: Assumptions.

That can’t be right...

Sharpless circular RNA story began several years ago on the short arm of human chromosome 9. There lies a series of genes implicated in cancer, aging, and atherosclerosis. The locus, Sharpless says, “is one of the strangest ... in the human genome.” It encodes 3 tumor suppressor proteins, as well as an exceptionally long 100 kb noncoding RNA transcript called ANRIL that somehow controls expression of the nearby coding genes. Sharpless wanted to know how.

In 2010, his team discovered that ANRIL exists in multiple forms. Some transcripts were in found in the usual linear form, with exons in the expected order: 1, 2, 3, and so on. But a substantial fraction adopted a circular structure—not a splicing lariat, but a fully closed exonic circle that the team called cANRIL. In those molecules, exon 5 followed exon 14, and outward-facing PCR primers positioned at either exon actually yielded a product, which would not happen with a linear molecule.

As Sharpless recalls, at the time even he doubted the truth of those results, attributing the data to sequencing errors. But the postdoc leading the project, Christin Burd (now on the faculty at Ohio State University), persisted, “go[ing] to elaborate lengths to prove to me that they were circles before we could publish.”

Finally, he relented. “I guess I can be just as close-minded as our reviewers were.”

Still, circular or linear, the question remained: What does ANRIL do? Taking a cue from noncoding RNA transcripts of the Xist gene, which bind and recruit epigenetic regulators to the inactive X chromosome, Sharpless guessed ANRIL might serve a similar function at 9p21. But he also wondered if the linear and circular forms might differ in potency—that is, could the circles provide a mechanism to tweak epigenetic regulation? He decided to see if other such circles might also exist in the cell.

It’s relatively easy to show RNA is in a circular form once you’ve found it, but finding a RNA circle amongst the cell’s transcriptional noise isn’t easy. While several researchers adopted bioinformatic strategies, Sharpless preferred a wet-bench approach. “I’m a molecular biologist at heart,” he says. He considered a gel-trap assay, in which an RNA sample is embedded in a polymer and then loaded on a gel. Linear molecules can migrate through such a matrix, but circles cannot, thereby enriching the population. But the gel trap is a complicated “old-school” assay. Sharpless opted for a simpler approach: RNA exonuclease resistance. Since circles have no ends, they should be resistant to digestion with the exonuclease RNase R.

William Jeck, a graduate student in Sharpless’ lab, digested total cellular RNA, tweaking conditions to optimize cANRIL recovery as a positive control. “Then we just took that gambish of [exonuclease-enriched] RNA and next-gen sequenced it,” Sharpless says. That was when things got surprising.

Instead of a “couple of circles”, they found more than 25,000. In many cases, circles were far more abundant than related linear RNAs—perhaps not surprising, given their higher nuclease stability. Circular RNAs were detected in both mice and man, suggesting these molecules are not molecular mistakes, but evolutionarily conserved functional elements.

It was then that Jeck and Sharpless tried to publish their findings. The pair sent their manuscript to a “famous high-profile journal. Rejected. Sent it to the next branch on the journal tree, same deal: Got reviewed, can’t be right. And worked our way down until we finally got reviewers who really knew RNA
features

biom. And they immediately said, you know, this is really interesting, we should publish it right away.”

All told, it took the team 14 months to publish their findings. “If you really had something that assaulted peoples’ worldview, that was really different from the way they were thinking, then they’re not gonna go down without a fight.”

To be fair, Sharpless concedes, that viewpoint, at least in part, is based on well established, but incorrect, assumptions. Early RNA-Seq efforts used reads that were too short to reliably detect circular transcripts. Thus, they missed circular signatures that newer sequencing data sets, with their longer reads, might not. But even more importantly, he says, until recently most alignment algorithms assumed exons could only appear in increasing order. That is, exon 2 must always precede exon 3, or (by alternative splicing) 4 or 5. “If they found a read that said exon 3 came before exon 2, they would throw it out as a non-mapping read.”

It took a new, less biased alignment program called MapSplice, created by University of North Carolina researcher Jan Prins and University of Kentucky researcher Jinze Liu, to finally break out of that intellectual box.

The reviewers, he says, “didn’t realize that their mapping algorithm had made an important choice for them not to look for these circles, and so I think that’s where there was confusion.”

From professor to post-doc

Julia Salzman had an easier time with her circular RNA story. But Salzman, a postdoctoral researcher in Patrick Brown’s lab at Stanford, didn’t set out to be a transcriptome biologist. In fact, she wasn’t, technically speaking, a biologist at all.

With a bachelor’s degree in math and a PhD in statistics, Salzman was an assistant professor of statistics at Columbia University when she decided she needed a change. “I loved biology so much that I really wanted to get a better understanding of it and have experience as an experimentalist,” she says.

In Brown’s lab, Salzman started looking for evidence of genomic rearrangement in cancer, not at the DNA level but in the transcriptome. To do that, she wrote software to probe RNA-Seq data sets for so-called “scrambled” exons. “I don’t have a degree in computer science,” she concedes, “but I know enough to make trouble.” To her surprise, she found quite a bit.

“Thousands of genes had evidence of [scrambling],” she says, but not only in the cancer samples; exons were rearranged in normal controls too. She found the same thing in her own RNA-Seq data sets.

At first, she attributed these scrambled genes to genomic rearrangements, such as tandem duplications (e.g., 1, 2, 3, 2, 3, 4). But the statistical models she was developing didn’t support that theory.

“It took a lot of head-scratching, but eventually we realized that the data was consistent with a model where the exons were circularized.”

Lambates largely ignored her findings, she says, suspecting she had made a programming error. But when she showed the results to Brown, he was hooked. “Many people might have discouraged us from pursuing it, but he always thought that it was a very interesting and worthwhile phenomenon to get to the bottom of—which I just think reflects Brown’s amazing scientific intuition.”

John Rinn, a noncoding RNA expert at Harvard Medical School, seconds that point. “That guy’s a Jedi. He is Yoda… Pat Brown has constantly had a profound impact in understanding gene regulation.”

Salzman and colleagues Chuck Gawad and Peter Wang eventually demonstrated that hundreds of genes produce circularized isoforms, some of which they validated experimentally using RNome R digestion and outward-directed PCR. The team submitted their findings to Proc ONE in November 2011, and the article was published the following February.

For Salzman, circular RNAs are not merely molecular curiosities. Their circular structure, she says, “has implications for essentially everything that we know or believe about RNA transcripts,” from how they are spliced and exported from the nucleus, to how and when transcripts are translated. “It makes a big difference.”

Everything old is new again

Yet it was back-to-back papers in Nature that really elevated circular RNAs into the scientific community’s consciousness in early 2013. Why? These studies were the first to ascribe an actual function to one of these molecules.

The first report, from Nikolaus Rajewsky’s lab at the Max Delbrück Center for Molecular Medicine in Berlin, used bioinformatics to identify 1500 circular RNAs in human leukocytes, 1903 in mouse brains (81 of which were also found in the human data set), and 724 from C. elegans. Rajewsky’s group then biochemically characterized a subset of these molecules, including a circular form antisense to the human CDR1 gene, which they called CDR1as. A second team, led by Jørgen Kjems at Aarhus University, Denmark, focused specifically on this molecule.

This was not Kjems’ first time working on CDR1as; he actually first described the molecule (which he calls ciRS-7) in 2011. At the time, his team was exploring the hypothesis that microRNAs could repress transcription by acting at the DNA level. In their search for candidate genes, they discovered that the CDR1 promoter contained a perfect match to miR-671. “But the funny thing was, it was a match to the antisense direction,” Kjems recalls—that is, to the strand opposite the CDR1 coding sequence. “But that of course doesn’t matter if you look at DNA effects.”

The team spent months chasing down leads, looking at DNA methylation and histone modification patterns. Everything came up negative. So, they shifted tack; maybe miR-671 worked in the usual way, but on an antisense transcript. That transcript was incredibly easy to find. “This transcript could be the highest expressed gene in all the neurons

.numbers

Vol. 55 | No. 2 | 2013

www.BioTechniques.com
in the brain,” Kjems explains, “it’s extremely highly expressed.”

But as they studied it, Kjems also realized the transcript was odd. It “seemed to behave really kind of weirdly on a gel,” running slower than its size suggested. For many, maybe most transcription biologists, the significance of that finding might have been lost. But Kjems was in a unique position to recognize its import.

Thirty years earlier, as a graduate student, he had discovered one of the first circular RNAs, an intron from a thermophilic archaeabacterium. He wrote a couple of papers on the topic, left for a postdoc with Phil Sharp at MIT, and never looked at circular transcripts again. But when the ciRS-7 data started rolling in, he says it was like déjà vu. “My senses were open for the possibility that it could be a circular RNA.”

In 2011, the team published that fact in *EMBO*, along with the observation that ciRS-7 could be cleaved and degraded by miR-671, the first documented case of such a cleavage event in a mammalian transcript.

ciRS-7’s story doesn’t end here though. ciRS-7 enhances *CDR1* expression, probably by aiding in nuclear export, according to Kjems. Unusual among eukaryotic genes, *CDR1* contains no introns, which can complicate nuclear export. But it has a second function, which both Kjems and Rajewsky’s teams explored. The 1500 bp molecule contains some 70 partial binding sites for a second microRNA, miR-7. Initially, Kjems says, they noticed the repetitive sequence in the antisense message, but attributed that to a repeating amino acid sequence in *CDR1* itself. Now he believes he had the story backwards; that is, the protein sequence repeats because the antisense message has multiple microRNA binding sites. “This is how we can actually be a bit fooled,” he says.

Expression of the circle “sponges up” miR-7 in the cell, effectively canceling out its expression and boosting miR-7 regulated genes. Kjems team showed that at least one other circular RNA can also do this: *Sry* contains 16 binding sites for miR-138, and its expression dampens miR-138 activity.

These papers provide the first evidence of function for a circular RNA, the first proof that the molecules are not mere molecular accidents. But that doesn’t mean it’s case closed. Rinn, for one, notes that if the circle can sop up microRNAs, so too can the linear molecules. “We’re pinning function way too fast,” Rinn says, rattling off a list of other possibilities that includes transcriptional memory, translational inhibition, and nuclear import/export. Maybe, he suggests, RNA circles could be packed into vesicles as a kind of transmissible transcriptome, “like little escape pods.”

Sharpless has a six-pack riding on the idea that some endogenous RNA circle will be shown to encode a novel run-off protein—an idea that Peter Sarnow at Stanford University showed to be possible in the 1990s using synthetic circular RNAs. Basically, if the number of base pairs in a circle is divisible by three, “you can make essentially a never-ending polypeptide,” Sharpless explains. “Even cooler, if [the transcript length] is not divisible by three, but does not have a stop codon in any reading frame, it will be translated around and around again, alternating between all three reading frames”—a possibility Sharpless and Jeck dub a “Möbius protein.”

The point, Rinn concludes, is that “the transcriptome continually surprises us with new forms of regulation in higher and higher complexity.” It won’t be easy to identify them, of course—ascribing function to noncoding RNAs is always tough, and distinguishing between circles and their linear partners compounds the problem. But the fact is, microRNA inhibition could turn out to be RNA circles’ least interesting function.

Bottom line: When it comes to RNA, assume nothing.

Written by Jeffrey M. Perkel, Ph.D.

*Bio Techniques* 55:55-57 (August 2013)
doi 10.2144/000114061

To purchase reprints of this article, contact: biotechniques@fosterprinting.com