

## TECHNOLOGY FEATURE

# THE RNA CODE COMES INTO FOCUS

*As researchers open up to the reality of RNA modification, an expanded epitranscriptomics toolbox takes shape.*

LAGUNA DESIGN/SPL



A molecular model of a bacterial ribosome bound to messenger RNA, a complex that is formed during protein synthesis.

BY KELLY RAE CHI

In 2004, oncologist Gideon Rechavi at Tel Aviv University in Israel and his colleagues compared all the human genomic DNA sequences then available with their corresponding messenger RNAs — the molecules that carry the information needed to make a protein from a gene. They were looking for signs that one of the nucleotide building blocks in the RNA sequence, called adenosine (A), had changed to another building block called inosine (I). This ‘A-to-I editing’ can alter a protein’s coding sequence, and, in humans, is crucial for keeping the innate immune

response in check. “It sounds simple, but in real life it was really complicated,” Rechavi recalls. “Several groups had tried it before and failed” because sequencing mistakes and single-nucleotide mutations had made the data noisy. But using a new bioinformatics approach, his team uncovered thousands of sites in the transcriptome — the complete set of mRNAs found in an organism or cell population — and later studies upped the number into the millions<sup>1</sup>.

Inosine is something of a special case: researchers can readily detect this chink in the armour by comparing DNA and RNA sequences. But at least one-quarter of our

mRNAs harbour chemical tags — decorations to the A, C, G and U nucleotides — that are invisible to today’s sequencing technologies. (Similar chemical tags, called epigenetic markers, are also found on DNA.) Researchers aren’t sure what these chemical changes in RNA do, but they’re trying to find out.

A wave of studies over the past five years — many of which focus on a specific RNA mark called N<sup>6</sup>-methyladenosine (m<sup>6</sup>A) — have mapped these alterations across transcriptomes and demonstrated their importance to health and disease. But the problem is vast: these marks coat not only mRNA but other RNA transcripts as well, and they cut across ►

► all the domains of life and beyond, marking even viruses with their presence.

The modifications themselves are not new. What has given them meaning and driven epitranscriptomics into the spotlight is the discovery of enzymes that can add, remove and interpret them. In 2010, chemical biologist Chuan He at the University of Chicago, Illinois, proposed that these chemical tags could be reversible and important regulators of gene expression. Not long afterwards, his group demonstrated<sup>2</sup> the first eraser of these marks on mRNA, an enzyme called FTO. That discovery meant that m<sup>6</sup>A wasn't just a passive mark — cells actively controlled it. And this realization came at about the same time that global approaches, harnessing the power of next-generation sequencing, made it possible to map m<sup>6</sup>A and other modifications across the transcriptome.

Today, epitranscriptomics is blossoming. Yet its toolbox remains a work in progress. Current methods lack the sensitivity required for use with rare and precious samples. It's also not possible to quantify the amount of a given modification in the transcriptome, nor to map more than one modification in a single experiment. "There's an urgent and high demand for additional technology developments for all kinds of RNA modifications," says molecular biologist Tao Pan at the University of Chicago, who collaborated on He's FTO studies.

That said, epitranscriptomics researchers are excited about the direction their field is taking. "Just as you wouldn't think of DNA without thinking about how DNA is packaged, or epigenetically modified," says geneticist Chris Mason at Weill Cornell Medical College in New York City, who has led m<sup>6</sup>A-mapping efforts, "I think now and in the future, no one will think of RNA without thinking 'How is it modified?'"

#### MAPPING WITH ANTIBODIES

In the early 1970s, scientists first showed that mRNA was chemically modified by using radioisotope labelling of m<sup>6</sup>A. But because those studies enriched the mRNA transcripts by selecting their 3' ends, which contain strings of adenosines, researchers worried that those preparations might contain trace amounts of other classes of RNA molecules, as well. "People stopped working on this because it was so difficult to get clear insights into whether the m<sup>6</sup>A in mRNA was a contaminant," says Samie Jaffrey, a chemical biologist at Weill Cornell Medical College.

Also difficult was working out where in the transcriptome m<sup>6</sup>A was located, which could provide clues to its function. Conventional sequencing approaches involve reverse transcription — converting RNA into complementary DNA (cDNA), which is then amplified and sequenced. The problem is that the reverse transcriptase enzyme used to make cDNA erases the modifications. "There was no way to see m<sup>6</sup>A," Jaffrey says. "When you reverse-transcribe it, it behaves exactly like an A."

Despite the technical challenges, the discovery of unexpected bacterial RNA modifications<sup>3</sup> piqued Jaffrey's interest, and he decided to look for them in mammalian RNA. Working with Mason, his team sheared RNA into tiny pieces, pulled out those that contained m<sup>6</sup>A using antibodies, and sequenced the RNAs<sup>4</sup>. "We were clearly seeing labelling of mRNAs and that was remarkable. It was not a contaminant," Jaffrey says. A similar study<sup>5</sup> by Rechavi's group unearthed a hilly landscape of m<sup>6</sup>A peaks, roughly 12,000 sites in 7,000 human genes. The modifications, Rechavi's team discovered, tended to be concentrated on the protein-coding sequences called exons and on stop codons, the three-letter codes in mRNA that signify the end of the protein-coding sequence.

The methods these researchers used, m<sup>6</sup>A-seq and MeRIP-seq, have since been broadly used to map m<sup>6</sup>A in different disease contexts and organisms. Antibodies and reagents targeting m<sup>6</sup>A are available

from Active Motif ([go.nature.com/2kqgz8](http://go.nature.com/2kqgz8)), MilliporeSigma ([go.nature.com/2kw39m3](http://go.nature.com/2kw39m3)) and New England BioLabs ([go.nature.com/2kqjjaz](http://go.nature.com/2kqjjaz)), among others. Researchers think that the modification could control the way cells develop into different types, a process that goes awry in cancer. Indeed, the first links between the epitranscriptome and cancer have already emerged. He's group, for example, showed that in some forms of acute myeloid leukaemia, FTO is present in higher-than-normal levels and seems to remove m<sup>6</sup>A from certain transcripts<sup>6</sup>, which could spur cells to differentiate.

A parallel line of research has turned that finding on its head. Using an antibody-mapping method called miCLIP, which is higher in resolution than its predecessors, Jaffrey's team showed that its m<sup>6</sup>A antibodies also bind to N<sup>6</sup>, 2'-O-dimethyladenosine

**"There's an urgent demand for additional technology developments for all kinds of RNA modifications."**

(m<sup>6</sup>Am), a modification of the chemical structures that cap the 5' end of mRNAs. At the time, Jaffrey didn't know if m<sup>6</sup>Am carried any biological meaning. But his team has since shown that m<sup>6</sup>Am (and not m<sup>6</sup>A) is in fact the

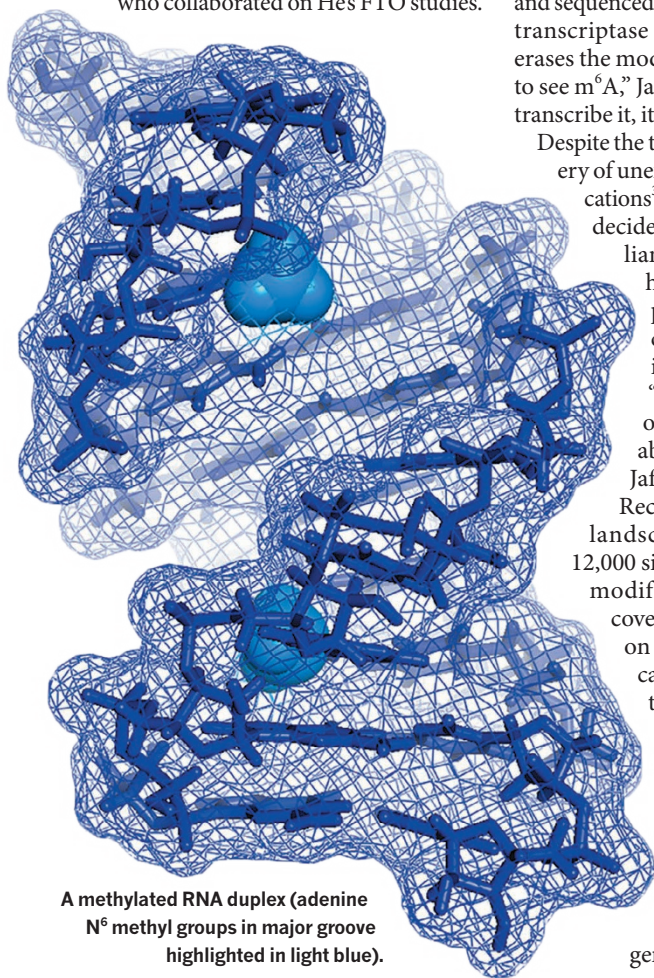
major target of the FTO eraser, and that it affects the stability and subcellular location of mRNAs<sup>7</sup>. To Jaffrey, that suggests that He's findings linking FTO to acute myeloid leukaemia mean that m<sup>6</sup>Am, not m<sup>6</sup>A, is now implicated in the origin and development of cancer.

Such discrepancies are par for the course in an emerging field, says cancer biologist Howard Chang at Stanford University, California. "This particular issue is not that different from the early days of the histone-modification field," he says, referring to the study of chemical alterations to the histone 'spools' around which DNA is tightly packaged in cells.

#### MANY MODIFICATIONS

Other RNA modifications have also attracted researchers' attention. In 2016, teams led by chemist Chengqi Yi at Peking University in Beijing and by Rechavi and He used antibody-based methods to map N<sup>1</sup>-methyladenosine (m<sup>1</sup>A) in mouse and human cell lines and tissues<sup>8,9</sup>. Using different approaches to prevent m<sup>1</sup>A from interfering with reverse transcription, the two teams showed that m<sup>1</sup>A, which was discovered in total RNA in the early 1960s, is present on mRNA at the position at which the translation machinery initiates protein production. Stress conditions alter the maps, suggesting that they are dynamic.

The researchers don't yet know what m<sup>1</sup>A does, but they have a tantalizing clue: most transcripts have only one m<sup>1</sup>A site, and these seem to be translated more often than those that lack the modification. "This is



**A methylated RNA duplex (adenine N<sup>6</sup> methyl groups in major groove highlighted in light blue).**



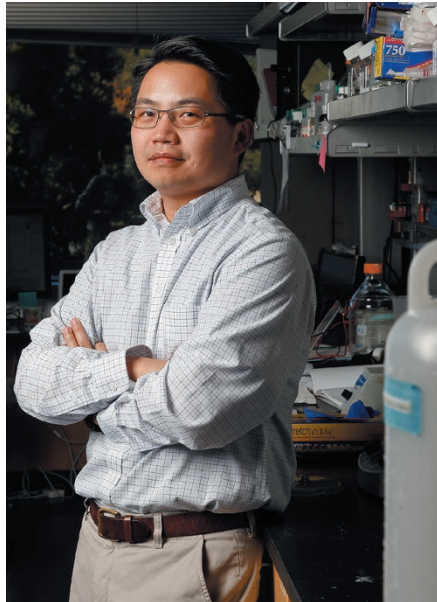
HHMI

very exciting — and of course challenging — because we are dealing with a new regulatory mechanism for translation of messenger RNA,” Rechavi says. An antibody that targets m<sup>1</sup>A is available from MBL International ([go.nature.com/2kvqpf5](http://go.nature.com/2kvqpf5)).

Other global mapping strategies rely on the fact that some RNA modifications can serve as handles for attaching chemical tags. When Yi started his lab in late 2011, modified RNA building blocks called pseudouridines or ‘pseudoUs’ were well known in other classes of RNA but had not been seen on mRNA. In 2015, his group described a chemical labeling and pulldown method for enriching the modification on transcripts<sup>10</sup>. To Yi’s surprise, pseudoUs are much more abundant in human and mouse mRNA than was previously thought. Now his focus is on finding the mark’s function. “I think pseudoU can have multiple functions in mRNA, depending on when, where and how it is installed and regulated on RNA transcripts,” Yi says. Some pseudoU ‘writers’ are already known, he adds, but whether there also are readers or erasers is an open question.

#### MOVING BEYOND MAPPING

Whether using antibodies or chemical approaches, mapping RNA modifications is a tricky business. Antibodies can cross-react with



Cancer biologist Howard Chang.

other modifications, so researchers should use at least two different antibodies and cross-correlate the hits, Mason says. Chemical methods can cut, bind or enrich some areas more frequently than others, producing biased fragmentation patterns. Sequencing depth and choice of bioinformatics algorithms can affect detection of

modification sites, Yi says. Even the length of time for which cells are kept alive in culture could influence modification levels, so it is crucial (and not necessarily trivial) to capture baseline maps for comparison, Mason says.

But in any event, it’s not enough to show that a sample contains a particular modification. Instead, it will be crucial to quantify all RNA modifications, because cells probably rely on just the right amount of a given modification for their proper function, Pan says. Quantification is particularly important for those researchers who want to tune levels of modifications by boosting the enzymes that write and erase them. The mere presence of such enzymes suggests that precisely tuned levels of modification are important, Pan says.

In 2015, Pan’s team described a potential way to quantify modifications, at least in another class of RNA called transfer RNAs<sup>11</sup>. The approach uses a reverse transcriptase that can read through the modifications efficiently, thereby capturing sequences downstream of the first modification it encounters. Now Pan says his group is working to apply the same strategy to mRNA modifications such as m<sup>1</sup>A.

But perhaps the fastest way to define function is to identify the readers, writers and erasers of these tags. In their 2012 m<sup>6</sup>A-mapping study<sup>4</sup>, Rechavi’s group created short stretches of RNA with and without m<sup>6</sup>A modifications,

using these fragments as bait to pull out any proteins that might be bound to the RNA. In 2014, He's group discovered several m<sup>6</sup>A readers using a comparable strategy<sup>12</sup>. Other studies have implicated other cellular functions as well. Now Rechavi plans to try a similar baiting approach to pull out readers of m<sup>1</sup>A, although that may prove more difficult because the site at which m<sup>1</sup>A concentrates (where protein translation begins) is more highly structured than are the sequences that contain m<sup>6</sup>A.

Once enzyme readers are identified for a particular chemical tag, gene-editing technologies should make it easy to tune their expression, allowing researchers to glean some insight from global changes to a modification. For instance, Chang's work deleting a writer of m<sup>6</sup>A showed the importance of this modification in determining cell fate<sup>13</sup>.

But as with all things epigenetic, interpreting such findings may also prove complicated, as the same enzyme may well work across broad swathes of RNA species, and a given modification can have different functions in each type of RNA, says Chang. "It wouldn't surprise me if some of the effects happen through other species of RNAs that aren't on people's radars but they're still prevalent in the cell and very important," he says.

#### ENHANCING THE TOOLBOX

In the past few years, He's group has discovered evidence<sup>14</sup> suggesting that RNA modifications provide a way to regulate transcripts involved in broad cellular roles, such as switching on cell-differentiation programs. Researchers need better technologies to explore these links; and, in October 2016, the US National Institutes of Health awarded He and Pan a 5-year, US\$10.6-million grant to establish a

centre to develop methods for identifying and mapping RNA modifications. One big focus is to come up with a way to generate mutations at the sites of a modification, and to amplify those, He says.

With new imaging techniques, it might eventually be possible to resolve single marks on a given transcript for visual inspection. "I'm dying to say that someone developed a technology to image the m<sup>6</sup>A modification in mRNA," He says, but at the moment, that is not the case. Ye Fu, a former graduate student in his lab, is exploring this approach in biophysicist Xiaowei Zhuang's lab at Harvard University in Cambridge, Massachusetts. Fu is combining super-resolution microscopy with another method Zhuang has pioneered for visualizing RNA in single cells, called MERFISH (multiplexed error-robust fluorescence *in situ* hybridization). Fu says that he has made progress in the past two years, but the data are noisy and need to be optimized for these modifications to be detected efficiently.

Others are working to circumvent the problems associated with conventional sequencing by sequencing RNA directly. Scientists at Oxford Nanopore Technologies in the United Kingdom reported one such method<sup>15</sup>, which extends to RNA the company's capability to thread DNA and other polymers through a nanopore embedded in a membrane. Researchers at Pacific Biosciences in Menlo Park, California, have also demonstrated direct RNA sequencing using the company's single-molecule real-time (SMRT) sequencing chemistry<sup>16</sup>. "The idea is certainly as old as SMRT sequencing itself," says Jonas Korlach, chief scientific officer. SMRT sequencing uses an enzyme called DNA polymerase to replicate strands of DNA, capturing the addition of

fluorescent nucleotides in real time. To adapt the technology to RNA, Korlach, Mason and their collaborators substituted the polymerase with the reverse transcriptase from HIV. As with the DNA platform, this enzyme incorporates fluorescent molecules across modified bases more slowly than it does across unmodified ones, giving each modification its own 'kinetic signature'.

Technical hurdles remain, however, as RNA poses challenges not seen in DNA. One is that RNA readily folds in on itself to form loops and knots, so it is highly structured. In its study<sup>15</sup>,

**"It wouldn't surprise me if some of the effects happen through other species of RNAs that aren't on people's radars"**

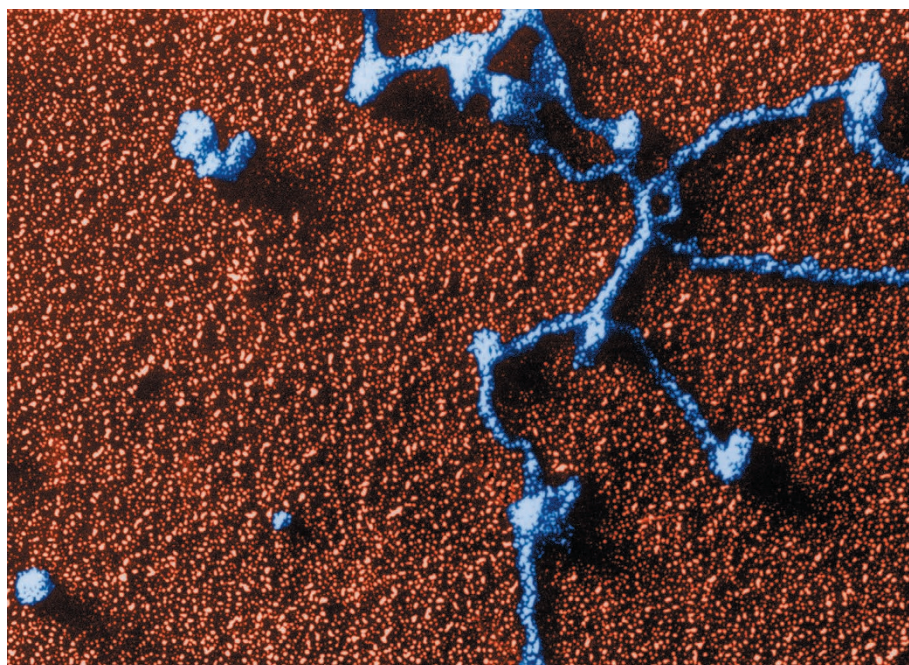
Oxford Nanopore attached RNA to cDNA, which helps 'iron out' the secondary structures and move the RNA through the pore. A second challenge is that RNA degrades more easily than DNA, a problem that

may stymie long-read sequencing approaches.

Yet another challenge, on the data side, is the sheer number of RNA modifications. Recognition of multiple different modifications on the same RNA would require massive training sets to teach the detection software to distinguish one modification from another. Winston Timp, a biomedical engineer at Johns Hopkins University in Baltimore, Maryland, has been using Oxford Nanopore technology to develop new methods for detecting specific DNA modifications<sup>17</sup>. He now plans to move into RNA modifications, developing a training set that will help recognize m<sup>6</sup>A modifications. "The problem is, we don't know how diverse on a given molecule the modifications are," he says. "But this is something we can probe. It's an exciting area of research." ■ **SEE NEWS FEATURE P.406**

**Kelly Rae Chi** is a freelance science writer in Cary, North Carolina.

1. Levanon, E. Y. *et al. Nature Biotechnol.* **22**, 1001–1005 (2004).
2. Jia, G. *et al. Nature Chem. Biol.* **7**, 885–887 (2011).
3. Kowtoniuk, W. E., Shen, Y., Heemstra, J. M., Agarwal, I. & Liu, D. R. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 7768–7773 (2009).
4. Meyer, K. D. *et al. Cell* **149**, 1635–1646 (2012).
5. Dominissini, D. *et al. Nature* **485**, 201–206 (2012).
6. Li, Z. *et al. Cancer Cell* **31**, 127–141 (2017).
7. Mauer, J. *et al. Nature* **541**, 371–375 (2017).
8. Dominissini, D. *et al. Nature* **530**, 441–446 (2016).
9. Li, X. *et al. Nature Chem. Biol.* **12**, 311–316 (2016).
10. Li, X. *et al. Nature Chem. Biol.* **11**, 592–597 (2015).
11. Zheng, G. *et al. Nature Methods* **12**, 835–837 (2015).
12. Wang, X. *et al. Nature* **505**, 117–120 (2014).
13. Batista, P. J. *et al. Cell Stem Cell* **15**, 707–719 (2014).
14. Zhao, B. S., Roundtree, I. A. & He, C. *Nature Rev. Mol. Cell Biol.* **18**, 31–42 (2017).
15. Garalde, D. R. *et al. Preprint at* <http://biorxiv.org/content/early/2016/08/12/068809> (2016).
16. Vilfan, I. D. *et al. J. Nanobiotechnol.* **11**, 8 (2013).
17. Simpson, J. T. *et al. Nature Methods* <http://dx.doi.org/10.1038/nmeth.4184> (2017).
18. Roost, C. *et al. J. Am. Chem. Soc.* **137**, 2107–2115 (2015).



Electron micrograph showing characteristic attachment of RNA polymerase molecules to DNA strands.