Special News Feature

EPITRANSCRIPTOMICS: MAPPING METHODS AND BEYOND

THE FIELD OF EPIGENETICS HAS CHANGED THE WAY WE LOOK AT BOTH DNA AND PROTEINS. COULD THE SAME BE TRUE FOR RNA? JOSEPH MARTIN DISCUSSES

One of the most exciting developments to come out of the completion of the Human Genome Project in 2003 was the emergence of the new field of epigenomics.

The blueprint of our genes may have been defined by the Human Genome Project but the rules governing them are prescribed by the science of epigenomics. Therefore, the origins of health and susceptibility to disease are, in part, the result of epigenetic regulation of the genetic blueprint.

The role of DNA modification in gene regulation is well studied; however, much less is known about how RNA modification affects RNA fate and influences the way genes are expressed.

In 2006, the Nobel Prize in Physiology and Medicine was awarded to Andrew Fire and Craig Mello for their discovery of RNA interference [1]. This sparked an abundance of research into the function of RNAs.

Adenosine, guanosine, cytidine and uridine determine the fate of messenger RNA (mRNA). However, these molecular building blocks can be chemically modified in ways that convey additional information to mRNAs. The collection of such modifications within the transcriptome comprises the epitranscriptome.

Transcriptome-wide technologies for mapping these modified nucleotides have pointed towards important roles for the epitranscriptome. There have been six modifications in mRNA and long non-coding RNA (lncRNA) that have been mapped to date (Figure 1). Analyzing these modifications will provide key insights into whether epitranscriptomic regulation is involved in important signalling pathways or disease processes. Moreover, these maps could tell us if mRNAs of interest are controlled by specific nucleotide modifications.

Mapping the epitranscriptome

The emergence of epitranscriptomics has been enabled by new technologies that allow nucleotide modifications to be mapped. There are three main approaches used to map modified nucleotides, which will be discussed here.

In 2012, research was published that many describe as the study that launched the resurgence of interest in epitranscriptomics. A team from Weill Cornell Medical College (NY, USA) advanced the concept of mRNA regulation by epitranscriptomic modification by developing methylated RNA immunoprecipitation and sequencing (MeRIP-Seq) technology, which can be used to globally profile m⁶A [2]. This method involves RNA fragmentation and immunoprecipitation with a modification-specific antibody. The antibody-bound RNA fragments are then subjected to RNA sequencing.

The modified nucleotides are marked throughout the transcriptome by accumulations of aligned sequence reads, also known as peaks (Figure 2). The distribution of these peaks is then analyzed to reveal the identities of modified transcripts, alongside their location within RNAs. It is important to note that peaks only represent assumed m⁶A sites. Another step is required to confirm that these peaks reflect m⁶A, rather than noise that occurs during mapping, thus making it difficult to confirm the location or number of modified residues in a transcript.

It is worth mentioning that antibodies can bind to structurally similar nucleotides; for instance, the highly prevalent nucleotide, m⁶Am [3]. The question remains, how can we be certain that the peaks established using these methods are attributed to m⁶A?

Samie Jaffrey of Weill Cornell Medical College comments on the development and future of m⁶A mapping: “*The development*

of MeRIP-Seq has opened up the field of epitranscriptomics by revealing transcriptome-wide patterns of m⁶A in diverse cells, tissues and disease contexts. However we still can't tell the stoichiometry of m⁶A modification at any specific site in the transcriptome. A method to map the stoichiometry of m⁶A at every site in the transcriptome will be the next advance in m⁶A mapping."

Another approach used to chart modified nucleotides exploits the chemical properties of the modified nucleotide to interfere with reverse transcription. Using aspects of this approach, N3-CMC-enriched pseudouridine sequencing (CeU-seq) was developed, which enhanced the sensitivity of pseudouridine mapping and led to an almost tenfold increase in the number of residues detectable [4].

The final approach involves the covalent trapping of proteins that interact with modified nucleotides, providing adducts that can be detected by reverse transcription. This is utilized by Aza-IP, which was recently used to map m⁵C [5]. These methods involve covalently trapping the modification-synthesizing enzyme on RNA.

A related approach, m⁶A individual nucleotide crosslinking and immunoprecipitation (miCLIP), is the preferred method for mapping m⁶A. In this method, UV light covalently cross-links m⁶A-specific antibodies at m⁶A residues in purified RNA. Crosslinked antibody–RNA complexes are then proteolyzed to generate peptide–nucleotide adducts that induce termina-

tions or nucleotide misincorporations adjacent to m⁶A during reverse transcription [6].

Discrepancies in m¹A maps

A number of modifications that have been mapped are suggested to be highly abundant in mRNA; however, subsequent studies have challenged these conclusions and propose that these modifications are rare or nonexistent.

One modification at the center of this debate is N1-methyladenosine (m¹A). The addition of a methyl group is what separates the m¹A modification from the nucleoside adenosine. The presence of this group adds a positive charge, preventing the formation of standard pairing interactions with other bases.

In 2016, two independent groups employed first-generation m¹A-sequencing methods, m¹A methylated RNA immunoprecipitation and sequencing (m¹A-seq) [7] and m¹A-specific RNA immunoprecipitation with demethylase-assisted RNA sequencing (m¹A-ID-seq) [8]. The procedures used were largely similar, both using the m¹A antibody to induce a mismatch or truncation during the reverse transcription process. These studies report that there are more than 7000 m¹A modifications across the diverse mRNA transcripts in the cell.

Interestingly, another group challenged the original finding, suggesting that m¹A is not an abundant epitranscriptomic regulatory mark but rather a rare modification in human mRNAs [9]. Safra *et al.* used a reverse transcriptase enzyme that synthesizes a complementary DNA strand from RNA, but which incorporates mutations precisely at m¹A locations. By using the same antibody as the previous studies, the group could not only judge whether the RNA fragments that bind the antibody contained m¹A, but also their location.

Is there an explanation for the inconsistencies in reported m¹A sites?

Schwartz *et al.* [10] recently reanalyzed the experimental dataset produced by Li *et al.* [8] and suggested that the computational analysis was not careful enough, with nearly 90% of the identified sites likely erroneous. Moreover, the research team applied their computational pipeline to filter such events. This produced identical results to their original study, revealing that with a few interesting and consistently identified exceptions, m¹A within mRNA are very rare.

The human genome is also complex and internally redundant, which leads to systematic biases in sequencing and alignment, all of which can lead to spurious identification of modification sites. Problems can also arise when searching for a particular misincorporation pattern due to various artifacts such as misannotations, duplicated events, sequencing errors and nontemplated reverse transcriptase incorporation.

Other explanations for the discrepancies come from research suggesting that the m¹A antibody has the potential to bind to unmodified regions of mRNA [11]. Distinguishing between RNA fragments that bind to antibodies through a modification and fragments that bind nonspecifically is a very difficult task. Dominissini *et al.* used a reverse transcriptase that is blocked by m¹A, leading to the absence of complementary DNA opposite the m¹A in RNA [7]. Importantly, such absences are a common,

Figure 1. Mapped modifications in mRNA and lncRNA. Six modifications (marked in red) have been mapped to date. m⁶A, N6: Methyladenosine; m⁶Am: N6,2'-O-dimethyladenosine; m¹A, N1: Methyladenosine; m⁵C: 5-methylcytosine; hm⁵C: 5-hydroxymethylcytosine. Reproduced with permission from [12].

Figure 2. Outline of MeRIP-Seq protocol and distribution of sequencing reads. (A) Schematic representation of MeRIP-Seq. RNAs containing m⁶A are immunoprecipitated. m⁶A-containing RNAs are then eluted and subjected to a second round of m⁶A immunoprecipitation. The resulting RNA pools are then subjected to next-generation sequencing. **(B)** Schematic of sequencing reads and their alignment to locations in the genome surrounding an m⁶A site. Top: An mRNA that contains a single m⁶A residue along its length. Middle: Individual 100 nt-wide mRNA fragments which are isolated following m⁶A immunoprecipitation. Bottom: Histogram showing predicted frequency of MeRIP-Seq reads obtained by sequencing individual immunoprecipitated fragments. **(C)** Sequencing reads from MeRIP-Seq converge over m⁶A sites. Reproduced with permission from [2].

nonspecific feature of next-generation sequencing, potentially giving the impression of widespread modifications in mRNA, another possible explanation for the large discrepancy between research groups.

What does the future hold for epitranscriptomics?

Despite the doubts surrounding the validation of mapping technologies, the field of epitranscriptomics does seem to be heading in the right direction.

Technological and methodological advancements have transformed our understanding of the epitranscriptome. However, there is still room for improvement when it comes to transcriptome-wide mapping technologies. It is still not determined which modifications are abundant in mRNA and which previously described modifications are technical artifacts. Emphasis is needed on validating modification sites that have been predicted by mapping studies. Understanding and overcoming these artifacts is vital in order for us to explore this largely uncharted field. The continuous improvements in sequencing approaches will only further our understanding of the role unseen mRNA modifications play in biological outcomes.

Author of several of the abovementioned research articles, Schraga Schwartz of the Weizmann Institute of Science (Rehovot, Israel), is hopeful that enthusiasm within the field

alongside new technologies will help expand our knowledge of this vastly unexplored field: “These are exciting days in which emerging technologies are allowing us to interrogate the fundamental building blocks of mRNA at unprecedented scale and sensitivity. The enthusiasm accompanying these efforts must be checked by utmost scientific rigor, to ensure obtaining an accurate representation of the repertoire and whereabouts of mRNA modifications. Such maps are critical for guiding future studies into functions and mechanisms of action.”

References

1. Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391(6669), 806 (1998).
2. Meyer KD, Saletore Y, Zumbo P, Elemento O, Mason CE, Jaffrey SR. Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell* 149(7), 1635–1646 (2012).
3. Munns TW, Oberst RJ, Sims HF, Liszewski MK. Antibody-nucleic acid complexes. Immuno-specific recognition of 7-methylguanine- and N6-methyladenine-containing 5'-terminal oligonucleotides of mRNA. *J. Biol. Chem.* 254(11), 4327–4330 (1979).
4. Li X, Zhu P, Ma S *et al.* Chemical pulldown reveals dynamic pseudouridylation of the mammalian transcriptome. *Nat. Chem. Biol.* 11(8), 592–597 (2015).
5. Khoddami V, Cairns BR. Identification of direct targets and modified bases of RNA cytosine methyltransferases. *Nat. Biotechnol.* 31(5), 458–464 (2013).
6. Linder B, Grozhik AV, Olarerin-George AO, Meydan C, Mason CE, Jaffrey SR. Single-nucleotide-resolution mapping of m⁶A and m⁶Am throughout the transcriptome. *Nat. Methods* 12(8), 767–772 (2015).
7. Dominissini D, Nachtergaele S, Moshitch-Moshkovitz S *et al.* The dynamic N1-methyladenosine methylome in eukaryotic messenger RNA. *Nature* 530(7591), 441–446 (2016).
8. Li X, Xiong X, Wang K *et al.* Transcriptome-wide mapping reveals reversible and dynamic N1-methyladenosine methylome. *Nat. Chem. Biol.* 12(5), 311–316 (2016).
9. Safra M, Sas-Chen A, Nir R *et al.* The m⁶A landscape on cytosolic and mitochondrial mRNA at single-base resolution. *Nature* 551(7679), 251–255 (2017).
10. Scharitz S. m⁶A within cytoplasmic mRNAs at single nucleotide resolution: A reconciled transcriptome-wide map. *RNA* (2018) (In Press).
11. Schwartz S, Agarwala SD, Mumbach MR *et al.* High-resolution mapping reveals a conserved, widespread, dynamic meiotically regulated mRNA methylation program. *Cell* 155(6), 1409–1421 (2013).
12. Grozhik AV, Jaffrey SR. Distinguishing RNA modifications from noise in epitranscriptome maps. *Nat. Chem. Biol.* 14(3), 215–225 (2018).

Written by Joseph Martin

BioTechniques 65: 121–124 (September 2018)
10.2144/btn-2018-0117