RNA PROCESSING AND MODIFICATIONS

Charting the unknown epitranscriptome

Eva Maria Novoa[1,2], Christopher E. Mason[3] and John S. Mattick[1,2]

Abstract | RNA modifications can alter RNA structure–function relationships and various cellular processes. However, the genomic distribution and biological roles of most RNA modifications remain uncharacterized. Here, we propose using phage display antibody technology and direct sequencing through nanopores to facilitate systematic interrogation of the distribution, location and dynamics of RNA modifications.

In contrast to DNA, which contains relatively few nucleobase variations, RNA is decorated with over 140 different chemical modifications. For decades, RNA modifications were tacitly considered to be passive 'structural' features, present only on, and required for the function in gene expression of RNA species such as tRNAs and rRNAs, and hence unlikely to have an active and direct role in regulating the process. This view was changed dramatically by the discovery that the demethylase fat mass and obesity-associated protein could reverse the RNA modification N^6-methyladenosine (m^6A)[1], leading to the birth of the term 'epitranscriptome' to describe the collective of labile and regulated RNA modifications[2].

The significance of the epitranscriptome is largely unknown but may be profound, as a means for the superimposition of plasticity on an otherwise genomically hard-wired transcriptome. The epitranscriptome 'code' may enable or enhance not only specific chemistries in RNA-catalysed or RNA-dependent reactions, but also changes to RNA structure–function relationships, which offers an additional layer of gene regulation in a spatiotemporal- and signal-dependent manner.

The first assays for studying RNA modifications — methylated RNA immunoprecipitation followed by sequencing (MeRIP–seq)[3], and specifically for m^6A, m^6A–seq[4] — appeared in 2012, and used methylation-specific antibodies for immunoprecipitation coupled to next-generation sequencing. These pioneering studies revealed that m^6A modifications are far more widespread than previously thought, occur in thousands of mRNAs, are subjected to dynamic regulation in different growth conditions and have major impact on mRNA processing, stability, translation and localization. Subsequent studies have found that m^6A modifications have pivotal roles in sex determination, in the regulation of the maternal-to-zygotic transition in vertebrates and in the DNA damage response. These findings highlight the necessity of mapping and studying other RNA modifications, whose combined importance could be analogous to that of DNA epigenetic modifications.

Hurdles in mapping modifications

Currently, sequencing-by-synthesis (SBS) technologies dominate genomic research; they operate by copying a DNA template (in the case of RNA sequencing, a prior reverse transcription step is required) to incorporate fluorescently-labelled nucleotides, the identity of which is determined by an optical device. The major limitation of SBS is that DNA polymerases are relatively blind to almost all DNA and RNA modifications, although some modifications may lead to an increase in nucleotide misincorporation, which can be used to infer their presence[5]. Consequently, indirect methods are being used to identify RNA modifications, which involve either detection by antibodies, such as in the case of m^6A, or chemical detection, using compounds that selectively react with modified ribonucleotides and block the progression of the reverse transcriptase.

During the past decade, the coupling of immunoprecipitation with next-generation sequencing has generated vast datasets of the genomic location of transcription factor-binding sites and of histone modification marks in different tissues, developmental stages and cell types. Unfortunately, however, suitable and specific antibodies that target RNA modifications are currently available only for m^6A, m^1A, 5-methylcytosine and 5-hydroxymethylcytosine, thus greatly limiting our understanding of the epitranscriptome.

Expanding the antibody repertoire

Antibody production is typically achieved by animal immunization that induces expression of antigen-specific polyclonal antibodies, which are recovered from serum, or requires obtaining hybridomas for the production of monoclonal antibodies. However, antibody development *in vivo* does not allow for counter-selection and requires extensive screening to identify modification-specific antibodies against a large background of less-specific antibodies. An alternative is phage display technology (PDT), which is an *in vitro* selection methodology whereby immunoglobulin variable domains are displayed on the surface of bacteriophages, which are then selected

[1]*Garvan Institute of Medical Research, Darlinghurst, NSW 2010, Australia.*
[2]*University of New South Wales, Sydney, NSW 2052, Australia.*
[3]*Weill Cornell Medicine, New York, NY 10021, USA.*
e.novoa@garvan.org.au; chm2042@med.cornell.edu; j.mattick@garvan.org.au

doi:10.1038/nrm.2017.49
Published online 10 May 2017

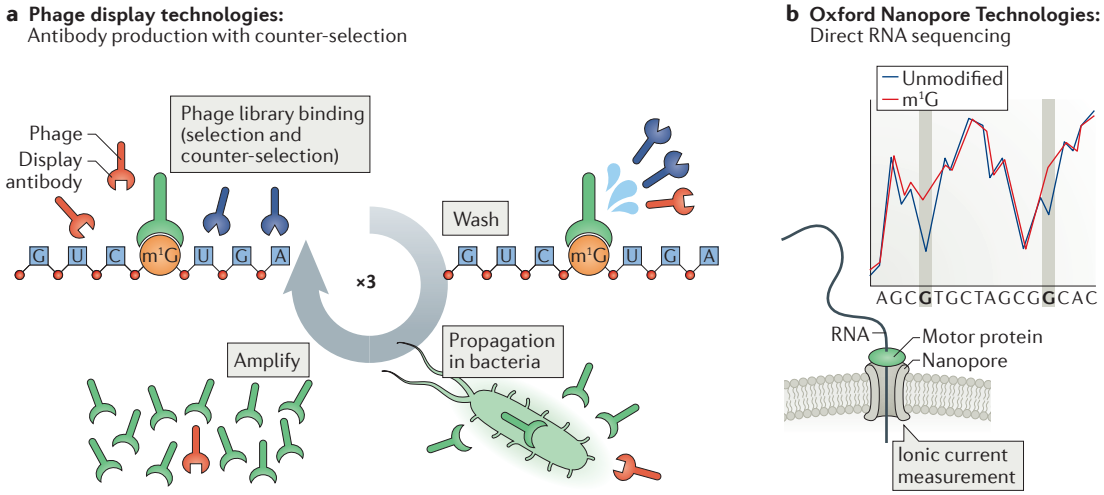

Figure 1. **The two proposed approaches to map RNA modifications. a** | In phage display technologies, highly specific and selective antibodies against RNA modifications can be produced using *in vitro* selection and counter-selection. **b** | RNA sequencing using Oxford Nanopore Technologies can directly detect modifications on RNA molecules that pass through the pore. The modified nucleotides are shaded in grey. m^1G, N^1-methylguanosine.

and counter-selected for binding to specific antigens by progressive rounds of biopanning, thereby enabling the selection of antibodies with desired properties from a large repertoire of variants[6] (FIG. 1a). Furthermore, PDT allows the derivation and improvement of antibodies that recognize specific epitopes, and counter-selection against competitor binders, thereby reducing the cross-reactivity of antibodies that target similar RNA modifications.

Direct sequencing of RNA

The field of epitranscriptomics would be transformed by the ability to perform direct sequencing of RNA without prior conversion to cDNA. A new alternative to SBS technologies is the direct RNA sequencing platform that is being developed by Oxford Nanopore Technologies (ONT), which comprises thousands of membrane-embedded protein nanopores coupled to highly sensitive ammeters that measure ionic current passing through the pore (FIG. 1b). As a DNA or RNA molecule is propelled through the pore, the sensor measures disruptions in current, the characteristics of which can be used to identify the transiting nucleotides — including modified nucleotides — using hidden Markov models or recurrent neural networks previously trained with known sequences. Notably, ONT can produce sequencing reads from single molecules and thus is potentially capable of producing quantitative transcriptome-wide maps with single nucleotide resolution and phased (coincidence) data, with read lengths limited primarily by sample preparation.

ONT was initially limited by its poor base-calling accuracy, but in the past year it has become a competitor of SBS technologies after its base-calling accuracy for single strands of DNA increased from <70% to >98%, owing to improvements in the pore technology and the base-calling algorithms. Indeed, ONT has been applied successfully in a wide range of settings, including the detection of DNA structural variants in cancerous samples, rapid detection of antibiotic resistance, sequencing of RNA genomes of Ebola and Zika viruses during outbreaks, and even metagenomic studies in space. Although to date the use of ONT has focused on conventional DNA sequencing, it recently proved successful in detecting DNA modifications[7]. Furthermore, a direct RNA sequencing protocol, which was commercially released in April 2017, has been used previously to detect m^6A RNA modifications[8], becoming the second single-molecule platform to detect m^6A in RNA[2].

ONT sequencing can directly detect modifications from a single RNA molecule during their transit through the nanopore, thereby enabling the detection of RNA modifications with minimal sample manipulation, at single nucleotide resolution and in an isoform-specific, quantitative and cost-effective manner. Therefore, in contrast to indirect methods, this technology could provide quantitative transcriptome-wide maps of RNA modifications and potentially be applicable to detecting any RNA modification. Once PDT and ONT sequencing are widely applicable in the nascent field of epitranscriptomics, we will be able to incorporate another important layer of gene regulation into our understanding of human development and physiology.

1 Jia, G. *et al.* N^6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat. Chem. Biol.* **7**, 885–887 (2011).
2 Saletore, Y. *et al.* The birth of the Epitranscriptome: deciphering the function of RNA modifications. *Genome Biology* **13**, 175 (2012).
3 Meyer, K. D. *et al.* Comprehensive analysis of mRNA methylation reveals enrichment in 3′ UTRs and near stop codons. *Cell* **149**, 1635–1646 (2012).
4 Dominissini, D. *et al.* Topology of the human and mouse m^6A RNA methylomes revealed by m^6A-seq. *Nature* **485**, 201–206 (2012).
5 Hauenschild, R. *et al.* The reverse transcription signature of *N*-1-methyladenosine in RNA-Seq is sequence dependent. *Nucleic Acids Res.* **43**, 9950–9964 (2015).
6 Lee, C. M., Iorno, N., Sierro, F. & Christ, D. Selection of human antibody fragments by phage display. *Nat. Protoc.* **2**, 3001–3008 (2007).
7 Simpson, J. T. *et al.* Detecting DNA cytosine methylation using nanopore sequencing. *Nat. Methods* **14**, 407–410 (2017).
8 Garalde, D. R. *et al.* Highly parallel direct RNA sequencing on an array of nanopores. *bioRxiv* https://doi.org/10.1101/068809 (2016).

Competing interests statement
The authors declare no competing interests.