Post-transcriptional modifications in development and stem cells

Michaela Frye¹ and Sandra Blanco¹,²,*

ABSTRACT
Cells adapt to their environment by linking external stimuli to an intricate network of transcriptional, post-transcriptional and translational processes. Among these, mechanisms that couple environmental cues to the regulation of protein translation are not well understood. Chemical modifications of RNA allow rapid cellular responses to external stimuli by modulating a wide range of fundamental biochemical properties and processes, including the stability, splicing and translation of messenger RNA. In this Review, we focus on the occurrence of N⁶-methyladenosine (m⁶A), 5-methylcytosine (m⁵C) and pseudouridine (Ψ) in RNA, and describe how these RNA modifications are implicated in regulating pluripotency, stem cell self-renewal and fate specification. Both post-transcriptional modifications and the enzymes that catalyse them modulate stem cell differentiation pathways and are essential for normal development.

KEY WORDS: RNA methylation, 5-methylcytosine, N⁶-methyladenosine, Pseudouridylation, Post-transcriptional modifications, Stem cells

Introduction
The development from a zygote into a multicellular adult organism requires the coordination of complex cellular processes. Cell-specific transcription factor networks determine gene expression programmes that orchestrate developmental decisions. In addition, epigenetic mechanisms, such as DNA and histone covalent modifications, regulate gene expression, representing an additional layer of control. Covalent modifications are also commonly found in RNA but whether these also affect gene expression to determine cell fate remains poorly understood.

To date, only a few post-transcriptional modifications, such as 5′ capping of messenger RNA (mRNA), polyadenylation, splicing or RNA editing, have been studied extensively (Bentley, 2014). The existence of multiple covalent RNA modifications was described over five decades ago, and more than 100 chemical modifications have now been identified in virtually all ribonucleotides (Machnicka et al., 2013). Methylation and pseudouridylation are the most common enzyme-catalysed covalent modifications (Machnicka et al., 2013). Studies in the early 1960s had already reported that the RNA post-transcriptional modification pattern changes during oocyte development, yet the relevance of these pathways in biological processes remained largely unknown (Ozban et al., 1964). Elucidating the function and biological relevance of chemical RNA modifications was hampered by the lack of high-throughput approaches to detect the modifications at high resolution in small biological samples. In addition, the RNA-modifying enzymes that catalyse the specific modifications have not received much attention in multicellular organisms.

The recent development of high-throughput sequencing techniques that allow the detection of N⁶-methyladenosine (m⁶A), 5-methylcytosine (m⁵C) or pseudouridine (Ψ) in low input RNA samples has revived interest in the field and identified functional roles in regulating RNA stability, translation and interactions with other molecules (Blanco et al., 2016; Carlile et al., 2014; Dominissini et al., 2012; Hussain et al., 2013a; Meyer et al., 2012; Squires et al., 2012; Van Haute et al., 2016). The importance of RNA modifications is highlighted by recent studies showing that abnormal RNA modification patterns can have devastating physiological consequences: mutations in RNA-modifying enzymes have been associated with human disorders such as intellectual disability, neurouma-degeneration, obesity and diabetes (Table 1).

In this Review, we summarise the dynamic regulation of m⁶A and m⁵C and pseudouridylation pathways, and discuss their newly discovered biological roles in development and stem cell maintenance and differentiation. We describe how methylation and pseudouridylation of RNA can influence development and stem cell properties by regulating RNA stability and processing and protein translation. These findings uncover an unknown layer of complexity in the regulation of gene expression and protein translation pathways that may be influenced by several other post-transcriptional modifications.

Adenosine methylation
m⁶A deposition in mRNA and miRNA: the story so far
m⁶A is the most prevalent internal modification found in mRNA from viruses to mammals, but also occurs in small non-coding RNA (ncRNA) and long non-coding RNA (lncRNA) in many eukaryotic species (Patil et al., 2016; Yue et al., 2015). m⁶A is enriched at 3′ untranslated regions (UTRs) near stop codons, within long internal exons, in intergenic regions and introns and at 5′ UTRs (Fig. 1) (Dominissini et al., 2012; Geula et al., 2015; Meyer et al., 2015, 2012; Schwartz et al., 2014b; Wang et al., 2014a). The correct deposition of N⁶-methyladenosine depends on an orchestrated and dynamic network of cell- and tissue-specific methyltransferases or ‘writers’ (the molecular machinery that deposits the marks) and demethylases or ‘erasers’ (the molecular machinery that removes the marks). How the methylation machinery selectively targets these specific regions of the transcriptome remains to be fully understood.

A multicomponent protein complex mediates adenosine methylation in higher eukaryotes. The stable core complex is formed by methyltransferase-like protein 3 (METTL3) or METTL14 [the S-adenosyl methionine (SAM)-binding subunit] and the auxiliary factor Wilms tumour 1-associating protein (WTAP), which facilitates m⁶A deposition by guiding the location of the complex into nuclear speckles – nuclear domains...
enriched in pre-mRNA splicing factors (Bokar et al., 1997; Liu et al., 2014; Ping et al., 2014; Wang et al., 2014b). Furthermore, recent evidence suggests that microRNAs (miRNAs) act as additional guides to selectively target mRNA for m6A deposition via a sequence pairing mechanism (Chen et al., 2015). Several studies have shown that the selective enrichment of m6A in specific transcriptome regions can also be achieved by dynamic demethylation of m6A (Fu et al., 2013; Jia et al., 2011; Zheng et al., 2013).

The functional consequences of m6A deposition

The functional consequences of adenosine methylation are extensive and determined by the type of modified RNA and a dynamic network of cell- and tissue-specific RNA-binding proteins or ‘readers’ (the molecular machinery that interprets the mark). One unifying view is that the presence of m6A acts as a switch to allow the recognition of the modified RNA by distinct effector proteins or ‘readers’ (Fig. 2A) (Liu et al., 2015). Most of these ‘readers’ proteins and their cellular location directly influence the functional outcome of the methylation mark. For instance, m6A in mRNA is recognised by the RNA-binding proteins YTH domain-containing family 1 (YTHDF1) and YTHDF2 (Dominissini et al., 2012; Liu et al., 2015; Wang et al., 2014a, 2015; Zhou et al., 2015). YTHDF1 is cytosolic and promotes efficient cap-dependent translation by facilitating the interaction of translation initiation factors with the methylated mRNA (Wang et al., 2015). YTHDF2 mainly targets the stop codon region, the 3' UTR, and the coding sequence (CDS) in order to regulate the degradation of the methylated transcripts by recruiting them to RNA decay sites in the cytosol (Fig. 2B) (Wang et al., 2014a). YTHDF2 is predominantly cytosolic, but in response to heat shock stress it is relocated into the nucleus and preferentially binds methylated adenosines within the 5' UTR of newly transcribed mRNAs. Mechanistically, YTHDF2 binding limits FTO-mediated demethylation in the 5' UTR of stress-induced transcripts and promotes cap-independent translation initiation, thus providing a mechanism for selective mRNA translation under heat shock stress (Fig. 2B) (Zhou et al., 2015). These findings show that although m6A is less abundant in the 5' UTR than in other regions of the transcripts, the mark is dynamic, inducible and also influences the functional outcome of the methylated targets. Likewise, other findings have shown that as little as one methylated adenosine is sufficient to initiate translation in a cap-independent manner by inducing binding of eukaryotic initiation factor 3 (eIF3), which in turn recruits the 43S complex to initiate translation (Meyer et al., 2015).

In contrast to other RNA species, the lncRNA X inactive specific transcript (XIST) is highly enriched in m6A residues, containing at least 78 throughout its length (Patil et al., 2016). Recent findings from Jaffrey’s lab showed that the abundant deposition of m6A is mediated by the methylation complex WTAP-METTL3, which is recruited to XIST-specific sites by RBM15 and RBM15B RNA-binding proteins (Patil et al., 2016). These findings reveal that this highly abundant post-transcriptional mark ensures YTHDC1 recruitment to at least a few of the methylated XIST residues and functionally enables the transcriptional repression effects of XIST on X chromosome genes. Mechanistically, it is still unclear how YTHDC1 leads to gene silencing; however, it has been suggested that YTHDC1 binding to XIST might recruit additional and specific XIST-interacting gene-silencing proteins to precise locations on XIST.

m6A in pre-mRNAs promotes differential splicing events by influencing binding to nuclear factors. For instance, m6A content in pre-mRNA promotes local structural changes that determine the direct binding of heterogeneous nuclear ribonucleoproteins such as HNRNPA2B1 and HNRNPC, which in turn affect pre-mRNA splicing and the abundance of alternatively spliced variants (Alarcon et al., 2015a; Liu et al., 2015). The recruitment of other splicing factors, such as SRSF3 and SRSF10, to methylated pre-mRNAs can be directed by the nuclear m6A ‘reader’ YTHDC1, leading to exon inclusion (Xiao et al., 2016). Similar to pre-mRNA, the recognition of the modified RNA by distinct effector proteins or ‘readers’ (the molecular machinery that interprets the mark). One unifying view is that the presence of m6A acts as a switch to allow the recognition of the modified RNA by distinct effector proteins or ‘readers’ (Fig. 2A) (Liu et al., 2015). Most of these ‘readers’ proteins and their cellular location directly influence the functional outcome of the methylation mark. For instance, m6A in mRNA is recognised by the RNA-binding proteins YTH domain-containing family 1 (YTHDF1) and YTHDF2 (Dominissini et al., 2012; Liu et al., 2015; Wang et al., 2014a, 2015; Zhou et al., 2015). YTHDF1 is cytosolic and promotes efficient cap-dependent translation by facilitating the interaction of translation initiation factors with the methylated mRNA (Wang et al., 2015). YTHDF2 mainly targets the stop codon region, the 3' UTR, and the coding sequence (CDS) in order to regulate the degradation of the methylated transcripts by recruiting them to RNA decay sites in the cytosol (Fig. 2B) (Wang et al., 2014a). YTHDF2 is predominantly cytosolic, but in response to heat shock stress it is relocated into the nucleus and preferentially binds methylated adenosines within the 5' UTR of newly transcribed mRNAs. Mechanistically, YTHDF2 binding limits FTO-mediated demethylation in the 5' UTR of stress-induced transcripts and promotes cap-independent translation initiation, thus providing a mechanism for selective mRNA translation under heat shock stress (Fig. 2B) (Zhou et al., 2015). These findings show that although m6A is less abundant in the 5' UTR than in other regions of the transcripts, the mark is dynamic, inducible and also influences the functional outcome of the methylated targets. Likewise, other findings have shown that as little as one methylated adenosine is sufficient to initiate translation in a cap-independent manner by inducing binding of eukaryotic initiation factor 3 (eIF3), which in turn recruits the 43S complex to initiate translation (Meyer et al., 2015).

In contrast to other RNA species, the lncRNA X inactive specific transcript (XIST) is highly enriched in m6A residues, containing at least 78 throughout its length (Patil et al., 2016). Recent findings from Jaffrey’s lab showed that the abundant deposition of m6A is mediated by the methylation complex WTAP-METTL3, which is recruited to XIST-specific sites by RBM15 and RBM15B RNA-binding proteins (Patil et al., 2016). These findings reveal that this highly abundant post-transcriptional mark ensures YTHDC1 recruitment to at least a few of the methylated XIST residues and functionally enables the transcriptional repression effects of XIST on X chromosome genes. Mechanistically, it is still unclear how YTHDC1 leads to gene silencing; however, it has been suggested that YTHDC1 binding to XIST might recruit additional and specific XIST-interacting gene-silencing proteins to precise locations on XIST.

m6A in pre-mRNAs promotes differential splicing events by influencing binding to nuclear factors. For instance, m6A content in pre-mRNA promotes local structural changes that determine the direct binding of heterogeneous nuclear ribonucleoproteins such as HNRNPA2B1 and HNRNPC, which in turn affect pre-mRNA splicing and the abundance of alternatively spliced variants (Alarcon et al., 2015a; Liu et al., 2015). The recruitment of other splicing factors, such as SRSF3 and SRSF10, to methylated pre-mRNAs can be directed by the nuclear m6A ‘reader’ YTHDC1, leading to exon inclusion (Xiao et al., 2016). Similar to pre-mRNA,
m^6A in primary microRNAs (pri-miRNAs) promotes pri-miRNA processing by recruiting the miRNA microprocessor complex protein DGCR8 and the ribonucleoprotein HNRNPA2B1 (Alarcon et al., 2015a,b).

Together, these studies suggest a regulatory network linking the miRNA, lncRNA and m^6A post-transcriptional pathways to a broad range of cellular processes from gene silencing to RNA processing and protein translation.

**Developmental roles for m^6A**

Different patterns of m^6A deposition into RNA occur at distinct developmental stages, indicating that m^6A deposition is tissue specific and dynamically altered in response to external stimuli (Batista et al., 2014; Geula et al., 2015; Meyer et al., 2012; Schwartz et al., 2014b). Deletion of Mettl3 during mouse development is lethal at post-implantation embryonic day (E) 5.5-7.5 (Geula et al., 2015). Knockout embryos at this stage of development retain expression of the pluripotent marker NANOG and fail to upregulate early differentiation markers, suggesting a role for m^6A in the exit from pluripotency (Geula et al., 2015). Deletion of Mettl3 and Mettl14 homologues in flies, zebrafish and plants also point to a role in embryogenesis, development and gametogenesis (Bodi et al., 2012; Hongay and Orr-Weaver, 2011; Ping et al., 2014). WTAP, the non-catalytic subunit of the methyltransferase multicomplex, plays a regulatory role in cell cycle progression and Wtap knockout in mice is also embryonic lethal (Horiuchi et al., 2006).

Genome-wide association studies initially linked genetic variants found in introns of FTO with increased risk of obesity and type 2 diabetes in humans, but the molecular mechanisms linking these non-coding variants with obesity were unclear (Dina et al., 2007; Frayling et al., 2007; Scott et al., 2007). FTO is a member of the alpha-ketoglutarate-dependent dioxygenase (AlkB) subfamily and reverses m^6A to adenine through oxidation (Fu et al., 2013; Jia et al., 2011).

**Fig. 1. Occurrence and function of N6-methyladenosine (m^6A), 5-methylcytosine (m^5C) and pseudouridine (Ψ) in eukaryotic RNAs.** Shown is the occurrence of the m^6A, m^5C and Ψ modifications and their respective modifying enzymes and ‘erasers’, as well as their known function. Enzymes and functions that remain unknown are indicated by a question mark. Transfer RNA (tRNA), mitochondrial RNA for methionine (mt-tRNA^Met^), messenger RNA (mRNA), small nuclear RNA (snRNA), ribosomal RNA (rRNA), mitochondrial RNA (mitrRNA), telomerase RNA component (TERC) and vault RNA (small non-coding RNAs found in cytoplasmic RNP particles) are illustrated.

<table>
<thead>
<tr>
<th>RNA type</th>
<th>Writer</th>
<th>Eraser</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>tRNA</td>
<td>m^5C</td>
<td>m^6A</td>
<td>m^5C</td>
</tr>
<tr>
<td>mt-tRNA^Met^-</td>
<td>NSUN3</td>
<td>?</td>
<td>Required for 5-formylcytosine (f^5C) deposition at the same position, and for efficient mitochondrial translation</td>
</tr>
<tr>
<td>mRNA</td>
<td>m^6A</td>
<td>m^5C</td>
<td>m^6A</td>
</tr>
<tr>
<td>snRNA</td>
<td>m^6A</td>
<td>m^5C</td>
<td>m^6A</td>
</tr>
<tr>
<td>rRNA</td>
<td>m^6A</td>
<td>m^5C</td>
<td>m^6A</td>
</tr>
<tr>
<td>mitrRNA</td>
<td>m^6A</td>
<td>m^5C</td>
<td>m^6A</td>
</tr>
<tr>
<td>TERC</td>
<td>m^6A</td>
<td>m^5C</td>
<td>m^6A</td>
</tr>
<tr>
<td>Vault RNA</td>
<td>m^6A</td>
<td>m^5C</td>
<td>m^6A</td>
</tr>
</tbody>
</table>

**Table 1. Occurrence and function of N6-methyladenosine (m^6A), 5-methylcytosine (m^5C) and pseudouridine (Ψ) in eukaryotic RNAs.**

- **m^6A**: Protects from RNA cleavage, stress-induced selective protein translation changes.
- **m^5C**: Required for 5-formylcytosine (f^5C) deposition at the same position, and for efficient mitochondrial translation.
- **Ψ**: mRNA-protein interaction, mRNA stability, alternative splicing, miRNA binding, nuclear export, translation.
- **DKC1**: mRNA stability?, translation accuracy?
- **NSUN2**: mRNA stability?, translation accuracy?
- **NSUN2**: mRNA stability, translation accuracy?
- **NSUN3**: mRNA stability, translation accuracy?
- **PUS**: mRNA stability, translation accuracy?
- **PUS**: mRNA stability, translation accuracy?
- **NSUN1**: mRNA stability, translation accuracy?
- **NSUN2**: mRNA stability, translation accuracy?
- **NSUN4**: mRNA stability, translation accuracy?
- **DKC1**: mRNA stability, translation accuracy?
- **NSUN2**: mRNA stability, translation accuracy?
- **NSUN4**: mRNA stability, translation accuracy?
- **NSUN1**: mRNA stability, translation accuracy?
- **NSUN2**: mRNA stability, translation accuracy?
- **NSUN4**: mRNA stability, translation accuracy?
- **NSUN1**: mRNA stability, translation accuracy?
- **NSUN2**: mRNA stability, translation accuracy?
- **NSUN4**: mRNA stability, translation accuracy?
- **NSUN1**: mRNA stability, translation accuracy?
- **NSUN2**: mRNA stability, translation accuracy?
- **NSUN4**: mRNA stability, translation accuracy?
- **NSUN1**: mRNA stability, translation accuracy?
- **NSUN2**: mRNA stability, translation accuracy?
- **NSUN4**: mRNA stability, translation accuracy?
- **NSUN1**: mRNA stability, translation accuracy?
- **NSUN2**: mRNA stability, translation accuracy?
- **NSUN4**: mRNA stability, translation accuracy?
- **NSUN1**: mRNA stability, translation accuracy?
- **NSUN2**: mRNA stability, translation accuracy?
- **NSUN4**: mRNA stability, translation accuracy?
nascent polypeptides are depicted as orange lines. Cap-independent translation initiation. Orange circles depict ribosomes and nascent polypeptides are depicted as orange lines.

2011). Recent evidence showed that the obesity-associated single-nucleotide polymorphisms within the FTO intronic region directly interact with the promoter of IRX3 and functionally regulate IRX3 but not FTO expression (Smemo et al., 2014). The transcription factor IRX3 is linked to the regulation of body mass and composition in mice, thus revealing IRX3 as a novel determinant of obesity. Besides the controversial association of FTO with increased risk of obesity, inactivating mutations of FTO have been shown to lead to an autosomal-recessive lethal syndrome in humans (Boissel et al., 2009). Affected individuals show postnatal growth retardation, microcephaly, functional brain deficits and cardiac defects and die shortly after birth (Boissel et al., 2009). In mice, loss of Fto is not embryonic lethal but leads to postnatal growth retardation and a significant reduction of body mass (Fischer et al., 2009; McMurray et al., 2013). Another m6A demethylase in mammals is AlkB homologue 5 (ALKBH5). Alkbh5 knockout mice are viable and anatomically normal but exhibit impaired male fertility in adulthood (Zheng et al., 2013). Studies such as these have identified essential roles of m6A demethylation pathways in CNS and cardiac system development, as well as in gonad development in mammals. The very broad developmental defects caused by impairment of m6A demethylating enzymes might suggest very low substrate specificity for each demethylase.

mRNA methylation regulates embryonic stem cell fate

The biological impact of m6A in embryonic stem cell (ESC) self-renewal and differentiation was initially controversial. Early studies showed that partial depletion of Mettl3 or Mettl14 by shRNAs reduced self-renewal in mouse ESCs (Wang et al., 2014b). However, a later study showed that complete Mettl3 knockout in mouse and human ESCs increased self-renewal but impaired differentiation into the neural lineage (Batista et al., 2014). The discrepancy among these studies might reflect, at least in part, the degree or time-dependent effect of m6A depletion. For instance, the selection and expansion of ESCs in the absence of METTL3 might select for colonies resistant to m6A depletion with increased self-renewal capacity. Recent studies have confirmed that Mettl3- and Mettl14-deficient mouse ESCs exist in a preserved state of naïve pluripotency, failing to progress through pluripotency to the primed state upon induction (Fig. 3A). In the same study, the authors showed that ablation of Mettl3 in epiblast stem cells – which exist in a differentiation-prepared, primed state of pluripotency – produces the opposite effect and strongly induces the expression of lineage commitment markers (Fig. 3B) (Geula et al., 2015). This contradictory effect was explained by the fact that in these experiments the majority of RNAs are methylated and thus depletion of METTL3 further enhances highly expressed genes; for ESCs in the naïve state, this would mean pluripotency genes, but in the primed state it would mean lineage commitment genes. Differential alternative splicing of METTL3-targeted transcripts and a modest increase in protein translation were also observed (Fig. 3) (Geula et al., 2015). Depletion of m6A increased the half-life of mRNAs and elevated the translation of a set of transcripts that were sufficient to sustain the distinct stem cell state (Batista et al., 2014; Geula et al., 2015). In conclusion, these studies show that the m6A modification plays a major role in ESC fate regulation and early development, yet the precise underlying molecular
mechanisms that enable its control over lineage fate decisions remain elusive.

In summary, dynamic cell- and tissue-specific deposition of m6A is achieved by a network of coordinated ‘writers’, ‘erasers’ and ‘readers’, and controls essential processes in stem cells and development. Thus, RNA methylation pathways add a significant layer of complexity to gene expression and protein translational regulation. The importance of RNA methylation pathways for physiological processes is underscored by the finding that their disruption can cause complex human diseases, including neurological and cardiac disorders.

**Pseudouridylation in coding and non-coding RNAs**

**Prevalence and distribution of pseudouridine**

In addition to m6A, recent studies have characterised the global distribution of pseudouridine (also known as 5-ribofuranose or Ψ), another widespread and dynamic modification of mRNA (Carlile et al., 2014; Schwartz et al., 2014a). Ψ is one of the most abundant and conserved post-transcriptional modifications present in RNA. Ψs represent 1.4% of total nucleotides in mammalian ribosomal RNAs (rRNAs) and are present in other stable non-coding RNAs such as transfer RNAs (tRNAs), spliceosomal small nuclear RNAs (snRNAs), small nucleolar RNAs (snoRNAs) and telomerase RNA (Fig. 1) (Ge and Yu, 2013; Machnicka et al., 2013). Global Ψ profiling techniques identified a large number of inducible pseudouridilatory sites in mRNA in response to heat shock or nutrient deprivation (Carlile et al., 2014; Schwartz et al., 2014a). Together, these studies revealed remarkable complexity in the global landscape of RNA pseudouridilation.

Since Ψs promote base stacking interactions and acquire chemical properties to form hydrogen bridges with bases other than adenine, their presence facilitates base pairing, influences folding and increases duplex stability (Davis, 1995). All these parameters have a strong impact on the structure, potential interactions and activity of the targeted RNAs. For instance, Ψ is found in almost all tRNAs and occurs mostly at position 55, but also within the D stem as well as the anti-codon stem and loop (Fig. 1) (Charrette and Gray, 2000). At all positions, Ψ contributes to the stabilisation of the structural motif in which it occurs. Accordingly, in the anti-codon loop Ψ is crucial for the proper binding of tRNA to the ribosome, which increases translation accuracy (Charrette and Gray, 2000). In tRNA, Ψs are concentrated in the peptidyl transferase centre, the decoding centre, the A-site finger region, and sites where ribosomal subunits interact (Fig. 1) (King et al., 2003; Liang et al., 2007; Piekna-Przybylska et al., 2008). All these sites are functionally important, and pseudouridylation at these sites can enhance ribosome biogenesis, polysome formation and translation. The highly conserved Ψ site in TERC (the telomerase RNA component) is located in a region essential for telomerase activity and TERT binding, suggesting that it might influence telomerase activity (Chen et al., 2002; Schwartz et al., 2014a).

**Catalysis of pseudouridylation**

Ψ is catalysed in a guide RNA-independent or -dependent manner. Guide RNA-independent pseudouridylation is catalysed by pseudouridine synthases (PUSs), which carry out both substrate recognition and catalysis. The guide RNA-dependent mechanism relies on RNA-protein complexes known as box H/ACA small RNPs, which consist of a box H/ACA snoRNA and four core proteins: centromere-binding factor 5 (CBF5, or dyskerin in mammals), non-histone protein 2 (NHP2), nucleolar protein 10 (NOP10) and glycine-arginine-rich protein 1 (GAR1) (Hamma and Ferre-D’Amare, 2006). The pseudouridylation pocket is guided through H/ACA-box snoRNAs and is based on a 10-12 nucleotide stretch of complementarity to the target sequence. The catalytic activity is provided by dyskerin (Duan et al., 2009; Liang et al., 2009). Dyskerin modifies mainly rRNAs and snoRNAs and it also participates in the active telomerase complex and thus plays a well-established role in the maintenance of telomere integrity (Hamma and Ferre-D’Amare, 2006; Mitchell et al., 1999). By contrast, PUSs directly catalyse Ψ formation at specific sites in tRNA and mitochondrial RNA (Hamma and Ferre-D’Amare, 2006). PUSs can also modify spliceosomal snRNAs in yeast, although there is no evidence to show that this mechanism exists in higher eukaryotes (Yu et al., 2011). Both stand-alone PUSs and H/ACA-box RNPs catalyse mRNA pseudouridylation, although a clear consensus sequence has yet to be identified (Fig. 1) (Carlile et al., 2014).

**Functional consequences of pseudouridylation**

The isomerisation of uridines into Ψs in non-coding RNAs can influence diverse cellular processes, including gene expression regulation, translation efficiency and accuracy, splicing or telomere length. This is because the Ψs are located in areas that are important for these processes. However, the precise functional consequences of pseudouridylation in mRNA remains unclear. Ψ residues were detected within 5′ UTRs, CDSs and 3′ UTRs with no clear positional bias (Carlile et al., 2014; Schwartz et al., 2014a). mRNAs containing heat shock-induced Ψs were more abundant in wild-type S. cerevisiae than in Pseudouridylate synthase 7 (PUS7)-deficient cells, suggesting that Ψ enhances mRNA stability (Fig. 4) (Schwartz et al., 2014a). Other studies reported an impact on protein translation, where the presence of Ψ in CDS of mRNA increased translation efficiency (Anderson et al., 2010; Kariko et al., 2008). Mass spectrometry sequencing as well as genetic ablation experiments in yeast showed that Ψ in stop codons decreases the efficiency of translation termination (Fig. 4) (Karijolich and Yu, 2011).
**Pseudouridylation in stem cell maintenance**

Given that pseudouridylation can modify the structure and activity of targeted RNAs that are important for translation, gene regulation and telomere maintenance, it is perhaps not surprising that they are involved in the regulation of stem cells. Indeed, mutations in dyskerin – the catalytic protein that facilitates guide RNA-dependent pseudouridylation – have been shown to affect stem cell functions in mice and humans. Dyskerin is highly expressed in pluripotent ESCs and is crucial for telomere elongation. Interestingly, dyskerin also regulates expression of the pluripotency factors OCT4 (POU5F1) and SOX2, raising the possibility that the abundance of dyskerin might regulate pluripotency (Agarwal et al., 2010; Fong et al., 2014). Mutations in dyskerin 1 (DKC1) cause the X-linked form of dyskeratosis congenita (X-DC) and Hoyeraal-Hreidarsson syndrome (HH) (Heiss et al., 1998; Knight et al., 1999) (see Table 1). X-DC patient-derived fibroblasts show disrupted telomerase activity during reprogramming into induced pluripotent stem cells (iPSCs), leading to progressive telomere shortening and loss of self-renewal capacity in long-term cultures (Batista et al., 2011). Mutations in dyskerin also lead to the aberrant differentiation of hematopoietic stem cells (Bellodi et al., 2013), while deletion of Dkc1 causes early embryonic lethality and leads to an age-dependent decrease in hematopoietic stem cells due to DNA damage and growth defects in adult mice (Gu et al., 2011; He et al., 2002).

**Cytosine-5 methylation in non-coding RNA**

Pathways of m^5^C deposition

Cytosine-5 methylation is most prevalent in tRNA and rRNA and was first described in the 1970s (Gambaryan et al., 1976). Thanks to the development of novel methods based on RNA-methylase complex precipitation or bisulphite conversion of cytosines into uracils followed by next-generation sequencing, it is now known that m^5^C is a widespread mark in the transcriptome of multicellular organisms (Hussain et al., 2013a; Khoddami and Cairns, 2013; Schaefer et al., 2010; Squires et al., 2012). To date, seven enzymes have been confirmed to methylate cytosine-5 in mammalian RNA (see Fig. 1); the DNA methyltransferase DNMT2 (TRDMT1) and several members of the NOP2/Sun domain RNA methyltransferase family (NSUN1-6). DNMT2, NSUN2 and NSUN6 all target cytoplasmic tRNA, yet in a non-overlapping and strictly site- and structure-specific manner (Blanco et al., 2014; Goll et al., 2006; Haag et al., 2015; Tuorto et al., 2012). NSUN2 targets most tRNAs at the variable loop (C48-50) (Blanco et al., 2014; Hussain et al., 2013b; Tuorto et al., 2012). DNMT2-mediated methylation is restricted to only three tRNAs: Gly^CCC^, Asp^GTC^ and Val^AAC^ at the anti-codon loop (C38) (Tuorto et al., 2012). NSUN6 targets the acceptor stem (C72) of CCA-tagged cysteine and threonine tRNAs (Haag et al., 2015). NSUN2 is currently the only methylase with a confirmed broader substrate specificity, as it also methylates other non-coding RNAs and a small number of mRNAs (Hussain et al., 2013b; Khoddami and Cairns, 2013; Squires et al., 2012).

NSUN3 and NSUN4 methylate mitochondrial tRNA and rRNA, respectively, and NSUN1 and NSUN5 target cytoplasmic rRNA (Fig. 1) (Metodiev et al., 2014; Nakano et al., 2016; Schosserer et al., 2015; Sharma et al., 2013; Van Haute et al., 2016). The methylase activity of NSUN7 has not been confirmed; however, its interaction with enhancer RNAs (eRNAs) can influence the transcription of the eRNA-targeted genes (Agulio et al., 2016). Similar to m^5^C on DNA, m^5^C in RNA can also be further modified by oxidation processes. Methylation of mitochondrial tRNA^Met^ by NSUN3 is required for the subsequent formation of 5-formylcytosine (f^5^C) at the same tRNA position (Nakano et al., 2016; Van Haute et al., 2016). Recent findings also suggest that TET-mediated oxidation reverses cytosine-5 methylation in vitro and 5-hydroxymethylcytosine (hm^5^C) has been detected in polyA-enriched RNA in vivo (Delatte et al., 2016; Fu et al., 2014; Huber et al., 2015).

**Mechanism of action of cytosine-5 methylation**

Similar to other post-transcriptional modifications, the presence of m^5^C alters the affinity of specific RNA-binding proteins and influences the activity of the targeted RNAs. For instance, loss of NSUN1- and NSUN4-mediated methylation of rRNA affects pre-rRNA processing, ribosome biogenesis and polysome assembly (Metodiev et al., 2014; Sharma et al., 2013). By contrast, lack of NSUN5-dependent methylation relaxes rRNA structure in the vicinity of the peptidyltransferase centre and alters the translation fidelity of the ribosome during stress responses in yeast (Fig. 1) (Schosserer et al., 2015; Sharma et al., 2013). Other modifications can also fine-tune the function of ribosomes or tRNAs; for instance, 2′-O-methylation and pseudouridylation enhance the rigidity and stability of certain tRNA secondary structures (Kawai et al., 1992). Thus, tRNA and rRNA modifications are crucial for the accuracy and efficiency of protein translation.

tRNAs are the most heavily modified types of RNA and these modifications can influence different processes: ribosylation at A64 influences tRNA discrimination (Shin et al., 2011) and m^1^A58 methylation is required for tRNA stability (Anderson et al., 1998). m^5^C stabilises the secondary structure of tRNA and also affects translation fidelity. For instance, changes in m^5^C levels at the wobble position of tRNA^1^eu^CAA^ in yeast cause the selective translation of specific transcripts during stress responses (Chan et al., 2012). In mammals, deposition of m^5^C by DNMT2 or NSUN2 protects tRNAs from endonucleolytic cleavage by preventing binding of the endonuclease angiogenin and cleavage into two small non-coding RNA fragments (Fig. 5A,B) (Blanco et al., 2014; Schaefer et al., 2010; Tuorto et al., 2015). Lack of NSUN2-mediated methylation causes accumulation of 5′ tRNA-derived fragments and decreases global protein translation in cells (Blanco et al., 2016, 2014). By contrast, loss of DNMT2-mediated methylation of tRNA^Asp^GTC^, Gly^CCC^ and Val^AAC^ causes tRNA-specific fragmentation patterns, which leads to specific codon mistranslation (Fig. 5A,B) (Tuorto et al., 2015). Cytosine-5 methylation of vault non-coding RNAs alters their processing into Argonaute-associated small RNA fragments that can function as miRNAs and regulate the translation of targeted mRNAs (Fig. 1) (Hussain et al., 2013b). Together, these studies reveal m^5^C modification as a common link between tRNA and other non-coding RNA processing pathways.

**Cytosine-5 methylation of tRNA in stem cells, development and cancer**

All members of the NSUN protein family are expressed during mouse embryogenesis and are enriched in the developing brain (Chi and Delgado-Olguin, 2013). Consistent with this expression pattern, loss-of-function of NSUN2 is not lethal but causes neurodevelopmental deficits in the fly, mouse and human (Abassi-Moheb et al., 2012; Blanco et al., 2014; Khan et al., 2012; Komara et al., 2015; Martinez et al., 2012). Lack of NSUN2 results in complete loss of m^5^C in the vast majority of transcribed tRNAs and causes increased angiogenin-induced tRNA cleavage leading to the accumulation of 5′ tRNA-derived small non-coding RNAs (Blanco et al., 2016, 2014). The molecular function of 5′ tRNA fragments is to repress cap-dependent translation by
displacing translation initiation and elongation factors from mRNAs (Gebetsberger et al., 2012; Ivanov et al., 2011; Spriggs et al., 2010). Accordingly, global protein synthesis is reduced in **NSUN2**-deficient cells (Blanco et al., 2016, 2014).

Several lines of evidence indicate that cell type-specific differences in protein synthesis levels are required to establish and maintain stem cell identity and function in tissue homeostasis, development and cancer (Blanco et al., 2016; Buszczak et al., 2014;
Signer et al., 2014). The function of tRNA methylation is to determine protein translation rates in those processes. Lack of m^5C facilitates tRNA cleavage, which in turn determines the identity and concentration of distinct sets of tRNA-derived non-coding RNAs. The accumulation of 5’ tRNA fragments shapes a distinct translational programme in which the translation of most proteins is repressed, with the exception of those involved in apoptosis, stress response, cell motility and cell morphogenesis. This inhibitory translational programme is sufficient to maintain stem cell functions, but it needs to be overcome to allow cell differentiation (Fig. 5A) (Blanco et al., 2016). Similarly, activation of cytosine-5 RNA methylation is also required to upregulate protein synthesis after cytotoxic stress to promote cell survival (Fig. 5B) (Blanco et al., 2016, 2014). Thus, the regulation of global protein synthesis by balancing loss and gain of tRNA methylation is an integral part of the cellular response to cytotoxic stress (Fig. 5C). The same principle applies to tumorigenesis: inhibition of tRNA methylation by ablation of NSUN2 in skin tumours locks the tumour-initiating cells in an undifferentiated and proliferative state; however, these NSUN2-deficient cells are also highly sensitive to stress and promptly undergo cell death upon treatment with cytotoxic drugs (Fig. 5D) (Blanco et al., 2016).

Loss-of-function studies of NSUN2 and DNMT2 in multicellular organisms confirm that cytosine-5 methylation of tRNA is required for tissue development and homeostasis. For instance, simultaneous deletion of Dnmt2 and Nsun2 in mice specifically affects brain, liver and adipose tissue development due to impaired differentiation programmes (Tuorto et al., 2012). Single deletion of Dnmt2 in mice, flies and plants does not impair viability or fertility (Goll et al., 2006). However, loss of Dnmt2 in new-born mice affects the differentiation of haematopoietic stem and progenitor cell populations (Tuorto et al., 2015). Likewise, morpholino-mediated loss of Dnmt2 in zebrafish specifically impairs liver, retina and brain differentiation (Rai et al., 2007).

Loss of Nsun2 causes significant delays in cell differentiation in skin and testis in mice (Blanco et al., 2011; Hussain et al., 2013c). In adult mouse skin, NSUN2 is most highly expressed in a restricted population of committed progenitor cells in growing hair follicles, and loss of Nsun2 delays stem cell differentiation (Blanco et al., 2011). Likewise, NSUN2 is expressed in an undifferentiated and proliferative population of committed progenitor cells in mouse skin tumours, and loss of Nsun2 expands the tumour-initiating cell population at the expense of differentiation, resulting in enhanced tumour growth (Blanco et al., 2016). In male gonads, loss of Nsun2 specifically blocks the meiotic progression of germ cells into the pachytene stage (Hussain et al., 2013c). Loss-of-function mutations of NSUN7 have also been linked to male infertility in humans, indicating that cytosine-5 methylation pathways are essential for germ cell differentiation (Harris et al., 2007; Khosronezhad et al., 2014).

Although the exact biological role of other tRNA chemical modifications in development and/or stem cell biology is currently unknown, it is clear that many of these can cause complex phenotypes when disrupted, similar to the lack of cytosine-5 methylation (Table 1). These chemical modifications in tRNA might be dynamically regulated in order to fine-tune protein translation according to cellular requirements. In line with this hypothesis, transcription of mRNAs and their codon-enriched tRNAs is coordinated during proliferation and differentiation processes, suggesting that the concomitant tRNA transcription coordination adapts the translation machinery to cope with the differential codon-biased usage in the new cellular state (Gingold et al., 2014).

### Cytosine-5 methylation of rRNA in development and lifespan

Individual ribosome components exhibit substantial differences in expression among cell types, which is suggestive of specialised and cell type-specific ribosomes. Indeed, mutations in genes encoding ribosome components often result in cell type-specific phenotypes (Buszczak et al., 2014). Similarly, the occurrence of m^5C at distinct sites in rRNAs may directly influence ribosomal activity to favour cell type-specific translational programmes. Deletion of the mitochondrial tRNA methylase gene Nsun4 is lethal in mice, and embryos at E8.5 show severe growth retardation and lack of visibly discernible anatomical structures (Metodiev et al., 2014). Lack of m^5C in mitochondrial rRNA impairs the assembly of mito-ribosomes, causing inhibition of mitochondrial translation and respiratory chain deficiencies (Metodiev et al., 2014). Loss of Nop2 causes defects in pre-rRNA processing, 60S biogenesis and polysome assembly (Hong et al., 1997). NSUN5 methylates rRNA in a region close to the peptidyltransferase centre and its loss relaxes the rRNA structure and favours translation of stress-responsive mRNAs (Schosserer et al., 2015; Sharma et al., 2013). Ablation of Nsun5 increases resistance to stress stimuli and extends lifespan inworms, flies and yeast (Schosserer et al., 2015).

Reduced protein synthesis has been shown to extend lifespan and increase resistance to stress in an evolutionarily conserved manner (Dubnikov and Cohen, 2015). ESCs or adult haematopoietic and skin stem cells also display low levels of protein synthesis compared with their differentiating immediate progeny (Blanco et al., 2016; Sampath et al., 2008; Signer et al., 2014). Why reduced protein synthesis extends the lifespan of an organism is currently unknown.

### Conclusions

Current studies on post-transcriptional modifications have revealed another layer of complexity in the regulation of cellular processes. A consensus is emerging that the fine-tuning of both gene expression and protein translation is crucial in order to regulate stem cell function and developmental processes. In this Review, we have summarised current knowledge on m^6A, m^Ψ and m^5C and described their roles in stem cell self-renewal, development and tissue homeostasis. To conclude, we offer our thoughts on the current state of the field and suggest some possible research goals to pursue in the coming years.

m^6A is found predominantly in mRNAs and non-coding RNAs and its deposition is regulated by a coordinated activity of ‘writers’ and ‘erasers’. The m^6A mark is dynamic, reversible, and inducible in stress conditions and cell and tissue specific. There is no consensus as to the exact function of the m^6A mark, yet it is known to be dictated by ‘reader’ proteins that bind the methylated RNA and execute a broad range of activities from RNA decay to cap-independent translation. Despite the progress made in recent years, there are still important goals to achieve. The next steps include determining the additional factors that coordinate m^6A deposition or removal, and understanding how these factors are regulated and how the specificity of the methylated sites is achieved. Finding the factors involved in this pathway will provide crucial information on the mechanism and functional role that the m^6A modification accomplishes in stem cell self-renewal and early development.

RNA pseudouridylation has been extensively studied and we now know that Ψ is found in a broad range of non-coding RNAs. With the recent development of new transcriptome-wide approaches it has also been found in mRNAs. Its deposition is dynamic and inducible in stress conditions and is coordinated by two types of
‘writers’: RNA-guided and RNA-independent enzymes. The molecular function of RNA pseudouridylation is very broad and varies from mRNA splicing to protein translation regulation, depending on the type of targeted RNA. Although we have accumulated a wealth of knowledge on the mechanisms regulating Ψ deposition and the molecular function of Ψ, there are still many questions to be answered. For instance, how exactly is pseudouridylation induced in stress and which factors coordinate the deposition? It will also be important to understand whether it can be induced in other cellular processes, or whether pseudouridylation is reversible and, if so, what are the ‘erasers’. Another interesting issue is how the pseudouridylated codons or tRNAs are recognised by the ribosome, and whether the cell uses this strategy to modulate protein diversity to adapt to complex cellular responses, such as to stress. These results will significantly advance our knowledge of the scope of Ψ function in stem cell and early embryonic development.

m5C is found predominantly in tRNA and rRNA. As for m6A and Ψ, m5C deposition is dynamic and cell/tissue specific. Its role is to regulate stem cell functions in tissue homeostasis, in cancer and in stress. However, the exact regulation of m5C deposition remains poorly understood. One important goal is to determine all cytosine-5 RNA methylases and their specific RNA targets to gain insight into the scope of m5C function. Other interesting avenues for exploration include the reversibility of cytosine-5 methylation – whether it is possible and the identity of the regulatory pathways that control the activity of both the m5C ‘writers’ and ‘erasers’. It will be equally important to identify how the m5C mark is interpreted by ‘readers’, and which ‘readers’ determine the outcome of methylated RNAs. Revealing the molecular mechanisms underlying m5C deposition and its function in stem cell, cancer and developmental biology will significantly increase our understanding of its association in human diseases and might open the door to novel therapeutic strategies.

Altogether, RNA post-transcriptional modification pathways overlap substantially to functionally regulate stress responses and stem cell fate decisions in development and adult tissues. Given the close association with the stress response, environmental signals might be the main trigger to modulate the levels of chemical modifications in RNA. The functional consequences of these dynamic and distinct RNA modifications converge mostly into regulating protein synthesis. Thus, a coordinated network of post-transcriptional modification pathways may ultimately ensure cellular adaptation to stress or lineage choice by co-ordinating the abundance, splicing and translation efficiency of transcripts that maintain the cell type-specific proteome. Ultimately, by understanding fundamental aspects of RNA modifications we will be able to develop small-molecule inhibitors or gene therapy tools for targeting proteins that could lead to new ways of controlling gene expression or protein translation. Such discoveries might lead to the development of novel therapeutic strategies to treat complex diseases, including developmental disorders and cancer.

Acknowledgements
We gratefully acknowledge the support of the Cambridge Stem Cell Initiative and Stephen Evans-Freke.

Competing interests
The authors declare no competing or financial interests.

Funding
We thank our funders Cancer Research UK (CR-UK), Worldwide Cancer Research, the Medical Research Council (MRC), the European Research Council (ERC) and European Molecular Biology Organization (EMBO). Research in the M.F. laboratory was supported by a core support grant from the Wellcome Trust and MRC to the Wellcome Trust-Medical Research Council Cambridge Stem Cell Institute.

References


