EPITRANSCRIPTOMICS: THE NOVEL RNA FRONTIER

EDITED BY: Giovanni Nigita, Mario Acunzo, William Chi Shing Cho and Carlo Maria Croce

frontiers Research Topics

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

EPITRANSCRIPTOMICS: THE NOVEL RNA FRONTIER

Topic Editors:
Giovanni Nigita, The Ohio State University, United States
Mario Acunzo, Virginia Commonwealth University, United States
William Chi Shing Cho, Queen Elizabeth Hospital, Hong Kong
Carlo Maria Croce, The Ohio State University, United States

Following the formulation of the central dogma of molecular biology and the later discovery of classes of non-coding RNAs, the primary focus of Genetics was essentially on variation of DNA aiming at elucidating biological pathways perturbed in diseases. Recently, extensive attention has shifted towards the study of posttranscriptional RNA modifications occurring in both protein-coding as well as non-coding RNAs, revealing a novel and finer layer of complexity in gene regulation. This, in turn, has led to the birth of the novel field of 'Epitranscriptomics'.

The recent increase of applications of high-throughput sequencing technology (HTS) has allowed the unprecedented opportunity to identify on a transcriptome-wide scale, millions of RNA modifications in human genes, counting today more than 140 distinct types such as: methylation (e.g. m6A, m1A, m5C, hm5C, 2'OMe) methylation (e.g. m6A, m1A, m5C, hm5C, 2'OMe), pseudourylation (ψ), deamination (e.g. A-to-I RNA editing).

The scope of this Research Topic was to collect both reviews and research articles addressing the wet lab approaches and bioinformatics methodologies necessary to aid in the identification of novel RNA modifications and characterization of their biological functions. Among the articles embracing the aim of the Research Topic, we have collected four original research and methods articles, five reviews, and a technology article.

Table of Contents

frontiers
in Bioengineering and Biotechnology

Editorial: Epitranscriptomics: The Novel RNA Frontier

Giovanni Nigita[1]*, Mario Acunzo[2]*, William Chi Shing Cho[3] and Carlo M. Croce[1]

[1] Department of Cancer Biology and Genetics, The Ohio State University, Columbus, OH, United States, [2] Division of Pulmonary Diseases and Critical Care Medicine, Virginia Commonwealth University, Richmond, VA, United States,
[3] Department of Clinical Oncology, Queen Elizabeth Hospital, Kowloon, Hong Kong

Editorial on the Research Topic

Epitranscriptomics: The Novel RNA Frontier

From the formulation of the Central Dogma of molecular biology to the discovery of novel non-coding RNA (ncRNAs) classes, the focus of Genetics involved DNA variants with the scope of elucidating biological pathways perturbed in disease. Recently, considerable attention has shifted toward the study of RNA modifications dynamically occurring in both protein-coding as well as non-coding RNAs, in both animal and plant kingdoms. This, in turn, has gradually led to the exciting exploration of the novel frontier of "Epitranscriptomics," revealing an additional, finer layer of complexity in gene regulation.

The application of the fast-growing high-throughput sequencing technology (HTS) has allowed the unprecedented opportunity to unveil millions of RNA modifications in human genes, such as resulting from methylation (e.g., m6A, m1A, m5C, hm5C, 2'OMe), pseudourylation (Ψ) and deamination (e.g., A-to-I RNA editing), counting to date more than 140 different forms of RNA modifications.

The aim of this Research Topic was to collect both original research articles and reviews in addressing the bioinformatics approaches and wet lab methodologies necessary to aid in the detection of novel RNA modifications and characterization of their biological functions, as well as contribute to the identification of the molecular protagonists involved in the regulation of such phenomena. Among the articles embracing the scope of the Research Topic, we have collected five reviews, four original research and methods articles, and a technology article.

Two reviews focused on RNA modifications in ncRNAs. Romano et al. surveyed the main types of RNA methylation in the context of ncRNAs with their associated functions. They also described the methodologies to identify and profile such RNA modifications. On the other hand, Zhao et al. instead, focused on the role of the pseudouridylation of ncRNAs in nuclear gene expression events. In particular, they reviewed how, in the nuclear context, ncRNA pseudouridylation contributes to transcription and pre-miRNA splicing.

Another important RNA modification is the non-templated addition of uridine(s) to the terminal end of RNA, termed 3′ RNA uridylation, which has been described having a role in the regulation of both mRNAs and ncRNAs. In this collection, Menezes et al. reviewed current discoveries of the functional roles for 3′ RNA uridylation, particularly focusing on mammalian biology.

RNA modifications add a novel and fine-tuning layer to gene expression regulation, having recently shown how the dysregulation of these phenomena can lead to the onset of disease, such as neurological diseases and cancer. In this collection, Angelova et al. discussed the emerging findings that show indicate the modifications of RNA impact the development of the brain and their involvement in neurological disorders. One of the most abundant internal modifications

in messenger RNAs (mRNAs) is represented by N6-methyladenosine (m6A), which has been discovered to play important roles in multiple biological processes. Liu et al. reviewed how the malfunctions of m6A machineries lead to several cancer types, including solid and non-solid tumors.

RNA modification phenomena are also present in the plant kingdom. It has been observed that post-transcriptional modifications, such as C-to-U RNA editing events, can occur in chloroplasts. In this collection, Rodrigues et al. introduced a method to profile RNA editing from chloroplast by using small RNA (sRNA) libraries form *Arabidopsis*, soybean and rice. The results obtained in this study encourage using sRNA libraries in order to identify novel RNA editing events and confirm previous detected ones, as well as profiling RNA editing in different plants under different biological conditions.

The improvements of high-throughput sequencing (HTS) techniques allowed us to unveil thousands of editing events in plants. In this collection, Lo Giudice et al. presented the third release of REDIdb, a freely available database for RNA editing events in plants organelles. The current release of REDIdb contains more than 26K RNA editing modifications, together with a new web interface allowing users to contextualize editing events in their genomic, biological and evolutionary contexts.

Together with the improvements of HTS technologies, we have observed remarkable progress in the development of algorithms and computational methods able to analyze and study data in the emerging field of Epitranscriptomics. Song et al. presented PEA-m5C, a transcriptome-wide machine learning-based predictor of m^{5}C modification in *Arabidopsis*. PEA-m5C was developed by employing a random forest algorithm with sequence-based features and optimized window, obtaining an average AUC of 0.939 in 10-fold cross-validation experiments.

The collection also includes a study conducted in Soybean seedlings by Chmielowska-Bak et al. concerning the influence of short-term cadmium stress on two RNA oxidation-dependent modifications: 8-hydroxyguanosine and apurinic/apyrimidinic sites. The authors reported an increased level of RNA oxidation in plants in stress conditions.

Finally, the collection also includes a study on the effect of centrifugal force in quantification of colorectal cancer-related mRNA in plasma using targeted sequencing. We know that the pre-analytical factors are critical for the measurement of analytes in the blood. This study investigated two common centrifugation protocols on plasma mRNA quality and quantity. The authors concluded that more targeted mRNAs could be found by double centrifuges of 1,600xg followed by 16,000xg than a single centrifuge of 3,500xg.

Epitranscriptomics represents a thrilling, vast and novel field in Cellular Biology. The articles we finally accepted for our Research Topic address some of its most exciting challenges, with the intention of providing a useful resource that will peak the interest of many researchers investigating the RNA modification phenomena. We would like to take this opportunity to thanks the contributions of the authors and the professional support from the editorial staffs.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

RNA Methylation in ncRNA: Classes, Detection, and Molecular Associations

Giulia Romano[1], Dario Veneziano[2], Giovanni Nigita[2] and Serge P. Nana-Sinkam[1]*

[1] Internal Medicine "Division of Pulmonary and Critical Care Medicine", Virginia Commonwealth University Health System, Richmond, VA, United States, [2] Department of Cancer Biology and Genetics, The Ohio State University, Columbus, OH, United States

Nearly all classes of coding and non-coding RNA undergo post-transcriptional modification, as more than 150 distinct modification types have been reported. Since RNA modifications were first described over 50 years ago, our understanding of their functional relevance in cellular control mechanisms and phenotypes has truly progressed only in the last 15 years due to advancements in detection and experimental techniques. Specifically, the phenomenon of RNA methylation in the context of ncRNA has emerged as a novel process in the arena of epitranscriptomics. Methylated ncRNA molecules may indeed contribute to a potentially vast functional panorama, from regulation of post-transcriptional gene expression to adaptive cellular responses. Recent discoveries have uncovered novel dynamic mechanisms and new layers of complexity, paving the way to a greater understanding of the role of such phenomena within the broader molecular cellular context of human disease.

Keywords: RNA, methylation, epigenetics, non-coding RNAs, RNA methodologies

INTRODUCTION

Up until recently, the central dogma (Crick, 1970) had supported primary focus on the molecular contributions of DNA and protein to human disease. The inability to detect and evaluate RNA with the necessary molecular resolution and precision has limited our understanding of the spectrum of RNA modifications that may drive disease.

Following the discovery of *pseudouridine* (Davis and Allen, 1957), nine additional modifications were identified in 1965 (Holley et al., 1965b). Finally, modification events in nucleotides of mRNA molecules were also uncovered in the 1970s (Desrosiers et al., 1974; Adams and Cory, 1975; Dubin and Taylor, 1975; Perry et al., 1975). Gradually, the "static" interpretation of the cellular role of RNA started to be challenged (Gilbert, 1986). With the discovery of novel species of non-coding RNA (ncRNA) and their mechanisms further investigated (Lee et al., 1993; Fire et al., 1998; Eddy, 2001), RNA biology came to the forefront (Todd and Karbstein, 2007). Along with advancements in experimental and transcriptomics techniques, which enabled a more detailed investigation of the translational control of cellular responses and phenotypes (Chan et al., 2010), interest in RNA modifications also grew, resulting in significant progress in the last 15 years. Recent discoveries, such as the first and second mRNA m[6]A demethylases FTO (Jia et al., 2011) and ALKBH5 (Zheng et al., 2013), as well as the identification of the METTL3/METTL14 methyltransferase complex (Liu et al., 2014), have triggered renewed interest in RNA modifications.

To date, a total of 163 post-transcriptional RNA modifications have been uncovered across all living organisms (Boccaletto et al., 2017) and are among the most evolutionarily conserved

properties of RNAs (Li and Mason, 2014), revealing a "novel," complex layer of biological regulation known as the *epitranscriptome* (Saletore et al., 2012). The functional diversity provided by these phenomena can indeed affect RNA structure, play a fundamental role in their interactions with other molecules and in regulatory networks, such as metabolic changes (Lewis et al., 2017), thus affecting every aspect of cellular physiology.

RNA modifications have been categorized as reversible and non-reversible. Among non-reversible modifications, we find well-studied phenomena such as RNA editing and pseudouridylation (Meier, 2011). Nonetheless, recent focus has shifted to *reversible* modifications, such as cytosine and adenosine methylations (Klungland et al., 2016). However, this classic distinction is being reassessed, in light of the discovery of "erasers" such as FTO and ALKBH5.

The importance of modifications in novel classes of ncRNA transcripts is also becoming relevant. Well-characterized chemical modifications in traditional classes of RNAs such as transfer (tRNAs) and ribosomal (rRNA) RNA, novel detection technologies and deep sequencing analysis (Veneziano et al., 2015, 2016), have paved the way for a fuller assessment of these molecular events also in regulatory ncRNAs, such as microRNA (Alarcon et al., 2015b) and long ncRNAs (Patil et al., 2016).

RNA METHYLATION

RNA methylation is a reversible, post-transcriptional RNA modification, affecting several biological processes, such as RNA stability and mRNA translation (Ji and Chen, 2012; Wang et al., 2014, 2015; Dev et al., 2017), through a variety of RNA methyltransferases, often using distinct catalytic strategies. Furthermore, recent studies have shown how the deregulation of proteins implicated in these modification phenomena is associated to disease (Supplementary Table 1A). In this section, we will review the main types and functions of methylation in ncRNAs (**Figure 1**).

N[6]-Methyladenosine (m[6]A)

N[6]-methyladenosine (m[6]A) is the most abundant internal modification detected to date in mRNA (Roundtree et al., 2017). Discovered in the 1970s, its function has been thoroughly investigated only in the last decade (Rottman et al., 1974; Wang and He, 2014). This was driven by the recent discovery and characterization of evolutionarily conserved proteins able to encode (*writers*), decode (*readers*), and remove (*erasers*) methylation (Lewis et al., 2017). Since 1994, different writers have been identified, including METTL3 and METTL14, proven to regulate the circadian clock, differentiation of embryonic stem cells and primary miRNA processing (Dominissini et al., 2012; Wang et al., 2014; Alarcon et al., 2015b). These enzymes work in complex with proteins essential to the correct processing of RNA methylation (Schwartz et al., 2014): Wilms tumor 1-associated protein (WTAP), RNA-binding motif protein 15 (RBM15) and Protein virilizer homolog (KIAA1429). Additionally, the discovery of ALKBH5 and FTO has revealed the dynamic dimension of

this modification phenomenon for cellular metabolism (Jia et al., 2011; Zheng et al., 2013). Recently, the YTH domain family proteins (YTHDF1–3) and YTH domain-containing protein 1 (YTHDC1) have been characterized as m[6]A readers, providing the first functional evidence of m[6]A (Wang et al., 2014).

The methyl group in m[6]A does not affect the Watson–Crick base-pairing (Liu and Jia, 2014), is highly conserved between human and mice and located in 5′ UTRs, 3′ UTRs, around stop codons, long internal and alternatively spliced exons (Dominissini et al., 2016; Li et al., 2016a; Lewis et al., 2017). It is also found in tRNA, rRNA, and small nuclear RNA (snRNA) as well as several long non-coding RNA, such as Xist (Dominissini et al., 2012). While not completely understood, m[6]A has been shown to play critical roles in the biological regulation of mRNA and ncRNA (Liu and Jia, 2014), particularly splicing, stability, turnover, nuclear export, and mediation of cap-independent translation (Meyer et al., 2015). Recently, Sun et al. (2016) have integrated all m[6]A sequencing data into a novel database, *RMBase*, identifying ~200,000 N[6]-Methyladenosines (m[6]A) sites in human and mouse. Finally, Linder et al. (2015) mapped m6A and m6Am at single-nucleotide resolution and identified small nucleolar RNAs (snoRNAs) as a new class of m6A-containing non-coding RNAs (ncRNAs).

N[1]-Methyladenosine (m[1]A)

Although the first studies on N[1]-methyladenosine (m[1]A) in total RNA date back more than 50 years (Dunn, 1961), only one study in the last decade has shed substantial light on function. m[1]A is a dynamic methylation event at the N[1] position of adenosine, comprising the addition of a methyl group and a positive charge in the base, specifically in the Watson–Crick interface, obviously altering RNA-protein interaction and RNA secondary structures through electrostatic effects (Roundtree et al., 2017). m[1]A is abundant in tRNA and rRNA (El Yacoubi et al., 2012; Sharma et al., 2013) exercising major influence on structure and function (Anderson, 2005). Two groups recently found a strong conservation of the m[1]A pattern in several human and murine cell lines as well as in yeast, affirming the important role of this modification along the evolutionary chain. In particular, m[1]A has been shown to have a role in mRNA translation, *via* unique localization near the translation start site and first splice site (Dominissini et al., 2016; Li et al., 2016a; Roundtree et al., 2017) and by facilitating non-canonical binding of the exon–exon junction complex (Cenik et al., 2017).

2′-O-Methylation (2′OMe/Nm)

2′OMe is a very common RNA modification in abundant RNAs (rRNA, snRNA, tRNA) (Schibler and Perry, 1977; Borges and Martienssen, 2015; Roundtree et al., 2017) as well as in microRNA and it is fundamental for the biogenesis and function of these molecules (Ji and Chen, 2012). It was initially detected at the second and third nucleotide in many mRNA (Schibler and Perry, 1977). Further, it was observed that in rRNA, the loss of an individual modification had no apparent effect, while the
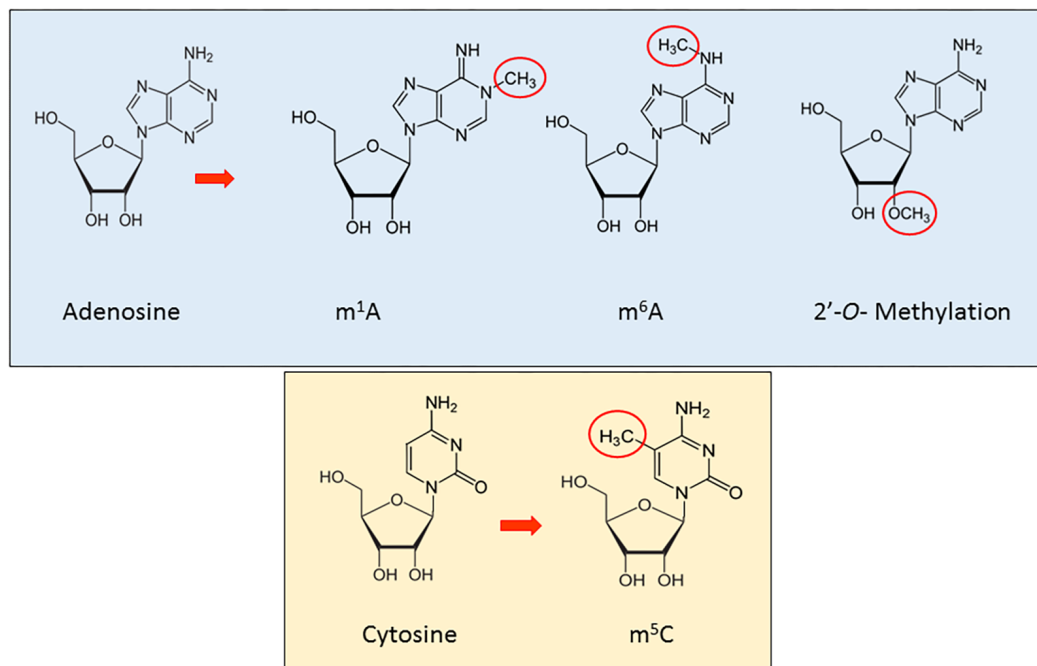
FIGURE 1 | A schematic representation of the principal methylation modifications in eukaryotic RNA.

deletion of 2–3 modifications in A and P site regions impairs translation and strongly delays pre-rRNA processing (Liang et al., 2009).

2′-O-methylation occurs in 3′ termini and is found to be important in plant biogenesis of small RNA, *inter alia* miRNA and siRNAs (Yu et al., 2005). Furthermore 2′-O-methylation plays an important role in protecting against 3′–5′ degradation and 3′ uridylation of some small RNAs as piRNAs in animals and Ago2-associated small RNAs in *Drosophila* (Ji and Chen, 2012). It has been found to be catalyzed by HUA-ENHANCER-1/piwi-methyltransferase (HEN1/piMET) enzyme.

5-Methylcytosine (m⁵C)

5-Methylcytosine (m^{5}C) is an epitranscriptomic modification that involves the 5th carbon atom of cytosine as a target for methylation in poly(A) RNA, rRNA, tRNA, snRNA, and lncRNA (Amort et al., 2013, 2017; Lewis et al., 2017). While some of the proteins regulating m^{5}C in different RNA have been identified, the biological function remains unclear (Nachtergaele and He, 2017). NOL1/NOP2/Sun domain family member 2 (NSUN2) together with DNA methyltransferase-like protein 2 (DNMT2) have been shown to be the writers of m^{5}C, although to date no erasers or readers have been discovered (Lewis et al., 2017), though recently, investigators identified ALYREF as a potential reader of m^{5}C (Yang et al., 2017). Several roles have been suggested for m^{5}C, from the stabilizing of tRNA secondary structure and prevention of degradation or cleavage, to playing a role in translation when in rRNA and increasing the stability of mRNA transcripts (Esteller and Pandolfi, 2017).

METHODOLOGIES FOR THE DETECTION AND PROFILING OF RNA METHYLATION

The recent advent of more sensitive and robust sequencing technologies (Li et al., 2016b), coupled with novel biochemical techniques (Song and Yi, 2017), has greatly improved the characterization and understanding of RNA modifications (Frye et al., 2016). This has allowed us to address challenges such as limitations with reverse transcription (RT) signatures and low transcript expression, as is the case with mRNA and lncRNA. Major advances in high-throughput sequencing methods (Helm and Motorin, 2017) have indeed allowed for the systematic identification of RNA modifications at single-nucleotide resolution, effectively distinguishing their distribution patterns in a transcriptome-wide manner.

Traditional biophysical targeted approaches for the detection and quantification of RNA modifications have further matured and provided the foundation for nearly all current high-throughput techniques (Vandivier and Gregory, 2017). Earlier methodologies relied on chromatography applied to direct sequencing, providing the very first evidence of modifications in RNA (Desrosiers et al., 1974). As these techniques only allowed detection of global patterns of modification, they were soon improved with the application of electrophoresis (Gupta and Randerath, 1979; Sprinzl and Vassilenko, 2005) and mass spectrometry (McCloskey and Nishimura, 1977; Kowalak et al., 1993) attaining for the first time base resolution. Recently other strategies, such as high-resolution melting (Golovina et al., 2014), have been implemented to narrow resolution.

Nonetheless, an important strategy on which several high-throughput techniques were later developed, is based on the detection of variation in RT signatures (Brownlee and Cartwright, 1977; Motorin et al., 2007). As RNA modifications may interfere with the RT enzyme, inducing its arrest and/or the misincorporation of non-complementary deoxyribonucleoside triphosphates (dNTPs), this provided the foundation to several current methodologies exclusively RT-based as well as leveraging on chemical treatment of the RNA pool or the use of antibodies for the enrichment of modified RNA populations. Such is the case of techniques employing methyl RIP-seq (*MeRIP-seq*) (Mishima et al., 2015; Dominissini et al., 2016; Li et al., 2016a) coupled with various crosslinking techniques to improve the resolution window. For instance, in m^1A-*ID-seq*, employ demethylases to generate a m^1A-depleted control library for validation (Li et al., 2016a). In alternative techniques, such as m^1A-*seq*, RNA pools undergo *Dimroth rearrangement* under alkaline conditions, converting m^1A residues to m^6A, thus producing different RT signatures that can validate the MeRIP data (Dominissini et al., 2016). Indeed, certain RNA modifications, such as m^6A and m^5C, are RT-silent. Despite simple antibody pulldown methods have satisfactorily mapped m^6A sites (Dominissini et al., 2012; Meyer et al., 2012) and antibodies highly specific to methylated RNA bases have also been employed (Linder et al., 2015; Li et al., 2016a), most antibody-based methods do not provide nucleotide resolution. For this reason, more recent global approaches have paired antibody binding to covalent crosslinking at specific RNA sites, resulting in RT signatures able to improve resolution (Linder et al., 2015). For instance, after transcripts fragmentation in MeRIP protocols, antibodies forming non-covalent complexes with modified residues are further cross-linked to reactive residues nearby via UV light at distinct frequencies according to the specific techniques (i.e., *miCLIP* and *PA-m^6A-seq*) for m^6A detection (Chen K. et al., 2015). Such induced covalent crosslinks are then the sites at which RT stalls, yielding approximate or precise single-nucleotide resolution. Recently, an innovative detection technique has precisely elucidated m^6A distributions across unknown regions *via* an antibody-independent strategy able to produce abortive cDNA signatures at m^6A sites, greatly increasing resolution (Hong et al., 2018). In the case of m^5C, *bisulfite sequencing* has yielded satisfactory results, although posing a few challenges. As unmodified cytosines are converted to inosines as a result of bisulfite treatment, m^5C residues remain unaffected, providing a signature in cDNA. While this has been effective for highly abundant ncRNA populations (i.e., tRNA and rRNA) (Militello et al., 2014), degradation issues (due to higher pH conditions during treatment) and read mapping challenges have yielded poor results for low-abundance RNA species (Squires et al., 2012; Hussain et al., 2013; Jeltsch et al., 2017). An alternative approach termed "suicide enzyme trap" has been employed to characterize substrates of m^5C-methyltransferases (m^5C - MTases) NSUN2 and NSUN4 (Metodiev et al., 2014; Van Haute et al., 2016). By mutating m^5C-MTases to form irreversible covalent bonds with target residues, the resulting stable enzyme–RNA complexes are suitable for immunoprecipitation and

mapping. Such is also the case of the AZA-seq methodology formalized by Khoddami and Cairns (2014) in which "suicide inhibitor" nucleotide analog *5-azacytidine* is incorporated into cellular RNA and "traps" m5C-MTases for pulldown and sequencing.

Finally, 2′OMe too can be detected at base resolution *via* differential RT profiles, with or without chemical treatment. The *RiboMeth-seq* methodology (Birkedal et al., 2015; Krogh et al., 2016; Marchand et al., 2016, 2017) for instance, leverages on the ability of 2′OMe to preserve adjacent phosphodiester bonds from alkaline cleavage and produces a high-throughput coverage profile of under-represented positions at the extremes of reads. Nonetheless, chemical treatment is not strictly necessary. Indeed, earlier methods relied on the natural ability of 2′OMe to interrupt RT at low dNTP concentrations (Maden et al., 1995). Such principle was recently employed in the development of a high-throughput protocol proven to be more sensitive and specific than methods based on alkaline hydrolysis. These methodologies have specifically been assessed on 2′OMe modifications occurring in ribosomal and transfer RNA, while not as efficiently identifying such phenomena in low abundance RNA molecules such as mRNA and several ncRNAs. To address such deficiency, the recently published *Nm-seq* protocol leverages on the ability of 2′OMe to confer resistance to oxidation by sodium periodate to the ribose backbone of RNA molecules, thus allowing the enrichment and mapping of reads originating from RNA fragments whose internal 2′OMe have been exposed at the 3′ end *via* the elimination of non-modified nucleotides. Such technique has provided a sensitive and precise 2′OMe detection method for rare RNA classes (Dai et al., 2017).

Due to the time-consuming and labor-intensive nature of such techniques, many transcriptomes and potentially novel modifications remain unexplored. For this reason, computational methods have also been developed for the accurate evaluation of modifications events (Zhang et al., 2015; Liu et al., 2017). Moreover, given the error-prone nature of high-throughput techniques, it is strongly suggested that modification sites predicted from big data not be considered as candidates if not validated with at least one additional methodology (Helm and Motorin, 2017). All methodologies described above are summarized in Supplementary Table 1B.

ncRNA SPECIES AND RNA METHYLATION: FUNCTIONAL ASSOCIATIONS

tRNA

tRNA methylations were first identified concurrently with the initial sequencing of the clover-shaped molecule (Holley et al., 1965a). Initially, it was suggested that such phenomena was probably the result of a network of diverse enzymes (Hurwitz et al., 1964). It is now clear that tRNA methylation is highly conserved and that tRNAs are the RNA class containing the majority of modified nucleosides among all discovered RNA

species. With a total of more than 90 modified nucleosides identified (MODOMICS) (Boccaletto et al., 2017), all tRNA molecules from the three domains of life contain 13 methylated nucleosides out of 18 shared (Marck and Grosjean, 2002; Jackman and Alfonzo, 2013). Originally, it was thought that tRNA modifications in general were a straightforward, static process occurring on specific sites of distinct tRNA species. Given the recent characterization of major tRNA modification pathways, along with their associated tRNA methyltransferase enzyme families (Hori, 2014), a relevant diversity has emerged among living organisms. The presence of catalytic interactions, distinct RNA substrate recognition mechanisms and diverse chemical processes, all suggest a complex functional panorama. Generally, four functional categories can be attributed to tRNA methylation phenomena: preservation of secondary and tertiary structures (Helm and Attardi, 2004; Voigts-Hoffmann et al., 2007); thermodynamic stability (Yokoyama et al., 1987); protection from degradation and rapid tRNA decay (Kadaba et al., 2004; Alexandrov et al., 2006; Guy et al., 2014); translation control and fidelity (Anderson et al., 1998, 2000; Chan et al., 2010, 2012). It is thus evident that tRNA methylation contributes to RNA quality control systems, cellular localization (Kaneko et al., 2003), response to stress stimuli (Schaefer et al., 2010; Becker et al., 2012; Muller et al., 2013), proliferation and many other processes (Phizicky and Hopper, 2015). Most importantly, disruption of energy and amino acid metabolism pathways (i.e., depletion of methionine, necessary for methylation) can damage downstream the RNA modification system, resulting in partially modified tRNAs and thus translational errors (explaining why living organisms use the methionine codon as the initiation codon for protein synthesis) (Hori, 2014). Recently, researchers discovered the first tRNA demethylase, ALKBH1, as a novel post-transcriptional gene expression regulation mechanism (Liu et al., 2016). Finally, tRNA methylations and their enzymes may cooperate collectively in functional networks in order to support adaptive cellular responses (Chan et al., 2010; Tomikawa et al., 2010; Ishida et al., 2011).

miRNAs

From transcription to decay, the multi-level process of the biogenesis of miRNAs is regulated by two main actors: processing enzymes such as DROSHA, DICER, and AGO proteins (Ha and Kim, 2014); and post-transcriptional modifications. Established RNA modifications, such as RNA editing events, have been shown to dynamically alter the sequence and/or the structure of miRNAs (Nigita et al., 2015; Nishikura, 2016) and consequently, in some cases, their function (Kawahara et al., 2007; Nigita et al., 2016). Recently, this has been also investigated in the context of miRNAs and RNA methylation.

2′OMe has been detected at the 3′-end of miRNAs (only in plants) and found to confer stability and protection from 3′-uridylation and degradation (Backes et al., 2012; Borges and Martienssen, 2015). m^6A within 3′ UTRs has been generally associated with the presence of miRNA binding sites; roughly 2/3 of mRNAs containing an m^6A site within their 3′ UTR also have at least one microRNA binding site (Meyer et al., 2012). In another study, Alarcon et al. (2015b) described

how miRNAs can undergo N^6-adenosine methylation (m^6A) as a result of the intervention of METTL3 during pri-miRNA processing. The same authors also showed that m^6A marks in pri-miRNAs allow for the RNA-binding protein DGCR8 to identify its specific substrates, promoting the beginning of miRNA biogenesis. Alarcon et al. (2015a) have further hypothesized that the RNA-binding protein HNRNPA2B1 could function as nuclear reader of the m^6A mark, binding to m^6A marks in pri-miRNAs, thus promoting pri-miRNA processing. Additionally, the effects of RNA *demethylation* on miRNA expression have also been investigated. Berulava et al. (2015) reported significant miRNA expression dysregulation as a result of knocking down m^6A demethylase FTO, providing indirect evidence of co-transcriptional processing in the methylation of mRNAs and miRNAs. Finally, Chen T. et al. (2015) discovered that miRNAs positively regulate m^6A installment on mRNAs *via* a sequence pairing mechanism. Methylation events in miRNAs add a new layer of complexity in the regulation of post-transcriptional gene expression and warrant future studies in order to fully elucidate the roles and functions of modified miRNAs.

Long ncRNA

Although the majority of focus has been recently devoted to modifications in mRNA, 1000s of lncRNA transcripts have been detected containing a substantial number of modifications (Shafik et al., 2016). Evidence associating methylation with the most established lncRNA transcripts are just starting to be recognized. MALAT1 has been shown to bind with the m^6A writer METTL16 at its 3′-triple-helical RNA stability element (Brown et al., 2016) specifically in its A-rich portion, after it was previously proven that MALAT1 can carry m^6A (Liu et al., 2013). The presence of m^6A has been further shown to destabilize the hairpin stems in the transcript, making them more flexible and solvent-accessible (Zhou et al., 2016) as well as more accessible for protein binding (Liu et al., 2015). Several putative m^5C sites have also been detected in MALAT1 (Squires et al., 2012), but no enzymes have been identified. lncRNA HOTAIR (Khoddami and Cairns, 2013) possesses a specific m^5C site which has been verified with a 100% modification rate (Amort et al., 2013). Finally, m^6A events have been associated to XIST-mediated transcriptional repression (Patil et al., 2016) while m^5C sites can prevent XIST-protein interactions, although it may not be a conserved mechanism (Amort et al., 2013). More detailed information can be found in Jacob et al. (2017).

AUTHOR CONTRIBUTIONS

SUPPLEMENTARY MATERIAL

REFERENCES

Abbasi-Moheb, L., Mertel, S., Gonsior, M., Nouri-Vahid, L., Kahrizi, K., Cirak, S., et al. (2012). Mutations in NSUN2 cause autosomal-recessive intellectual disability. *Am. J. Hum. Genet.* 90, 847–855. doi: 10.1016/j.ajhg.2012.03.021

Adams, J. M., and Cory, S. (1975). Modified nucleosides and bizarre 5′-termini in mouse myeloma mRNA. *Nature* 255, 28–33. doi: 10.1038/255028a0

Alarcon, C. R., Goodarzi, H., Lee, H., Liu, X., Tavazoie, S., and Tavazoie, S. F. (2015a). HNRNPA2B1 is a mediator of m(6)A-dependent nuclear RNA processing events. *Cell* 162, 1299–1308. doi: 10.1016/j.cell.2015.08.011

Alarcon, C. R., Lee, H., Goodarzi, H., Halberg, N., and Tavazoie, S. F. (2015b). N6-methyladenosine marks primary microRNAs for processing. *Nature* 519, 482–485. doi: 10.1038/nature14281

Alexandrov, A., Chernyakov, I., Gu, W., Hiley, S. L., Hughes, T. R., Grayhack, E. J., et al. (2006). Rapid tRNA decay can result from lack of nonessential modifications. *Mol. Cell* 21, 87–96. doi: 10.1016/j.molcel.2005.10.036

Amort, T., Rieder, D., Wille, A., Khokhlova-Cubberley, D., Riml, C., Trixl, L., et al. (2017). Distinct 5-methylcytosine profiles in poly(A) RNA from mouse embryonic stem cells and brain. *Genome Biol.* 18:1. doi: 10.1186/s13059-016-1139-1

Amort, T., Souliere, M. F., Wille, A., Jia, X. Y., Fiegl, H., Worle, H., et al. (2013). Long non-coding RNAs as targets for cytosine methylation. *RNA Biol.* 10, 1003–1008. doi: 10.4161/rna.24454

Anderson, J., Phan, L., Cuesta, R., Carlson, B. A., Pak, M., Asano, K., et al. (1998). The essential Gcd10p-Gcd14p nuclear complex is required for 1-methyladenosine modification and maturation of initiator methionyl-tRNA. *Genes Dev.* 12, 3650–3662. doi: 10.1101/gad.12.23.3650

Anderson, J., Phan, L., and Hinnebusch, A. G. (2000). The Gcd10p/Gcd14p complex is the essential two-subunit tRNA(1-methyladenosine) methyl-transferase of Saccharomyces cerevisiae. *Proc. Natl. Acad. Sci. U.S.A.* 97, 5173–5178. doi: 10.1073/pnas.090102597

Anderson, J. T. (2005). RNA turnover: unexpected consequences of being tailed. *Curr. Biol.* 15, R635–R638. doi: 10.1016/j.cub.2005.08.002

Backes, S., Shapiro, J. S., Sabin, L. R., Pham, A. M., Reyes, I., Moss, B., et al. (2012). Degradation of host microRNAs by poxvirus poly(A) polymerase reveals terminal RNA methylation as a protective antiviral mechanism. *Cell Host Microbe* 12, 200–210. doi: 10.1016/j.chom.2012.05.019

Becker, M., Muller, S., Nellen, W., Jurkowski, T. P., Jeltsch, A., and Ehrenhofer-Murray, A. E. (2012). Pmt1, a Dnmt2 homolog in *Schizosaccharomyces pombe*, mediates tRNA methylation in response to nutrient signaling. *Nucleic Acids Res.* 40, 11648–11658. doi: 10.1093/nar/gks956

Begley, U., Sosa, M. S., Avivar-Valderas, A., Patil, A., Endres, L., Estrada, Y., et al. (2013). A human tRNA methyltransferase 9-like protein tumour growth by regulating LIN9 and HIF1-α. *EMBO Mol. Med.* 5, 366–383. doi: 10.1002/emmm.201201161

Berulava, T., Rahmann, S., Rademacher, K., Klein-Hitpass, L., and Horsthemke, B. (2015). N6-adenosine methylation in MiRNAs. *PLoS One* 10:e0118438. doi: 10.1371/journal.pone.0118438

Birkedal, U., Christensen-Dalsgaard, M., Krogh, N., Sabarinathan, R., Gorodkin, J., and Nielsen, H. (2015). Profiling of ribose methylations in RNA by high-throughput sequencing. *Angew. Chem. Int. Ed. Engl.* 54, 451–455. doi: 10.1002/anie.201408362

Boccaletto, P., Machnicka, M. A., Purta, E., Piatkowski, P., Baginski, B., Wirecki, T. K., et al. (2017). MODOMICS: a database of RNA modification pathways. 2017 update. *Nucleic Acids Res.* 46, D303–D307. doi: 10.1093/nar/gkx1030

Borges, F., and Martienssen, R. A. (2015). The expanding world of small RNAs in plants. *Nat. Rev. Mol. Cell Biol.* 16, 727–741. doi: 10.1038/nrm4085

Brown, J. A., Kinzig, C. G., Degregorio, S. J., and Steitz, J. A. (2016). Methyltransferase-like protein 16 binds the 3′-terminal triple helix of MALAT1 long noncoding RNA. *Proc. Natl. Acad. Sci. U.S.A.* 113, 14013–14018. doi: 10.1073/pnas.1614759113

Brownlee, G. G., and Cartwright, E. M. (1977). Rapid gel sequencing of RNA by primed synthesis with reverse transcriptase. *J. Mol. Biol.* 114, 93–117. doi: 10.1016/0022-2836(77)90285-6

Cenik, C., Chua, H. N., Singh, G., Akef, A., Snyder, M. P., Palazzo, A. F., et al. (2017). A common class of transcripts with 5′-intron depletion, distinct early coding sequence features, and N(1)-methyladenosine modification. *RNA* 23, 270–283. doi: 10.1261/rna.059105.116

Chan, C. T., Dyavaiah, M., Demott, M. S., Taghizadeh, K., Dedon, P. C., and Begley, T. J. (2010). A quantitative systems approach reveals dynamic control of tRNA modifications during cellular stress. *PLoS Genet.* 6:e1001247. doi: 10.1371/journal.pgen.1001247

Chan, C. T., Pang, Y. L., Deng, W., Babu, I. R., Dyavaiah, M., Begley, T. J., et al. (2012). Reprogramming of tRNA modifications controls the oxidative stress response by codon-biased translation of proteins. *Nat. Commun.* 3:937. doi: 10.1038/ncomms1938

Chen, K., Lu, Z., Wang, X., Fu, Y., Luo, G. Z., Liu, N., et al. (2015). High-resolution N^6-methyladenosine (m^6A) map using photo-crosslinking-assisted m^6A sequencing. *Angew. Chem. Int. Ed. Engl.* 54, 1587–1590. doi: 10.1002/anie.201410647

Chen, T., Hao, Y. J., Zhang, Y., Li, M. M., Wang, M., Han, W., et al. (2015). m6A RNA methylation is regulated by microRNAs and promotes reprogramming to pluripotency. *Cell Stem Cell* 16, 289–301. doi: 10.1016/j.stem.2015.01.016

Crick, F. (1970). Central dogma of molecular biology. *Nature* 227, 561–563. doi: 10.1038/227561a0

Cui, Q., Shi, H., Ye, P., Li, L., Qu, Q., Sun, G., et al. (2017). m6A RNA methylation regulates the self-renewal and tumorigenesis of glioblastoma stem cells. *Cell Rep.* 14, 2622–2634. doi: 10.1016/j.celrep.2017.02.059

Dai, Q., Moshitch-Moshkovitz, S., Han, D., Kol, N., Amariglio, N., Rechavi, G., et al. (2017). Nm-seq maps 2′-O-methylation sites in human mRNA with base precision. *Nat. Methods* 14, 695–698. doi: 10.1038/nmeth.4294

Davis, F. F., and Allen, F. W. (1957). Ribonucleic acids from yeast which contain a fifth nucleotide. *J. Biol. Chem.* 227, 907–915.

Desrosiers, R., Friderici, K., and Rottman, F. (1974). Identification of methylated nucleosides in messenger RNA from Novikoff hepatoma cells. *Proc. Natl. Acad. Sci. U.S.A.* 71, 3971–3975. doi: 10.1073/pnas.71.10.3971

Dev, R. R., Ganji, R., Singh, S. P., Mahalingam, S., Banerjee, S., and Khosla, S. (2017). Cytosine methylation by DNMT2 facilitates stability and survival of HIV-1 RNA in the host cell during infection. *Biochem. J.* 474, 2009–2026. doi: 10.1042/BCJ20170258

Dominissini, D., Moshitch-Moshkovitz, S., Schwartz, S., Salmon-Divon, M., Ungar, L., Osenberg, S., et al. (2012). Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* 485, 201–206. doi: 10.1038/nature11112

Dominissini, D., Nachtergaele, S., Moshitch-Moshkovitz, S., Peer, E., Kol, N., Ben-Haim, M. S., et al. (2016). The dynamic N(1)-methyladenosine methylome in eukaryotic messenger RNA. *Nature* 530, 441–446. doi: 10.1038/nature16998

Dubin, D. T., and Taylor, R. H. (1975). The methylation state of poly A-containing messenger RNA from cultured hamster cells. *Nucleic Acids Res.* 2, 1653–1668. doi: 10.1093/nar/2.10.1653

Dunn, D. B. (1961). The occurrence of 1-methyladenine in ribonucleic acid. *Biochim. Biophys. Acta* 46, 198–200. doi: 10.1016/0006-3002(61)90668-0

Eddy, S. R. (2001). Non-coding RNA genes and the modern RNA world. *Nat. Rev. Genet.* 2, 919–929. doi: 10.1038/35103511

El Yacoubi, B., Bailly, M., and De Crecy-Lagard, V. (2012). Biosynthesis and function of posttranscriptional modifications of transfer RNAs. *Annu. Rev. Genet.* 46, 69–95. doi: 10.1146/annurev-genet-110711-155641

Elhardt, W., Shanmugam, R., Jurkowski, T. P., and Jeltsch, A. (2015). Somatic cancer mutations in the DNMT2 tRNA methyltransferase alter its catalytic properties. *Biochimie* 112, 66–72. doi: 10.1016/j.biochi.2015.02.022

Esteller, M., and Pandolfi, P. P. (2017). The epitranscriptome of noncoding RNAs in cancer. *Cancer Discov.* 7, 359–368. doi: 10.1158/2159-8290.CD-16-1292

Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., and Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391, 806–811. doi: 10.1038/35888

Fischer, J., Koch, L., Emmerling, C., Vierkotten, J., Peters, T., Brüning, J. C., et al. (2009). Inactivation of the Fto gene protects from obesity. *Nature* 458, 894–898. doi: 10.1038/nature07848

Frye, M., Jaffrey, S. R., Pan, T., Rechavi, G., and Suzuki, T. (2016). RNA modifications: what have we learned and where are we headed? *Nat. Rev. Genet.* 17, 365–372. doi: 10.1038/nrg.2016.47

Gilbert, W. (1986). Origin of life: the RNA world. *Nature* 319:618. doi: 10.1038/319618a0

Golovina, A. Y., Dzama, M. M., Petriukov, K. S., Zatsepin, T. S., Sergiev, P. V., Bogdanov, A. A., et al. (2014). Method for site-specific detection of m6A nucleoside presence in RNA based on high-resolution melting (HRM) analysis. *Nucleic Acids Res.* 42:e27. doi: 10.1093/nar/gkt1160

Gupta, R. C., and Randerath, K. (1979). Rapid print-readout technique for sequencing of RNA's containing modified nucleotides. *Nucleic Acids Res.* 6, 3443–3458. doi: 10.1093/nar/6.11.3443

Guy, M. P., Young, D. L., Payea, M. J., Zhang, X., Kon, Y., Dean, K. M., et al. (2014). Identification of the determinants of tRNA function and susceptibility to rapid tRNA decay by high-throughput in vivo analysis. *Genes Dev.* 28, 1721–1732. doi: 10.1101/gad.245936.114

Ha, M., and Kim, V. N. (2014). Regulation of microRNA biogenesis. *Nat. Rev. Mol. Cell Biol.* 15, 509–524. doi: 10.1038/nrm3838

Helm, M., and Attardi, G. (2004). Nuclear control of cloverleaf structure of human mitochondrial tRNA(Lys). *J. Mol. Biol.* 337, 545–560. doi: 10.1016/j.jmb.2004. 01.036

Helm, M., and Motorin, Y. (2017). Detecting RNA modifications in the epitranscriptome: predict and validate. *Nat. Rev. Genet.* 18, 275–291. doi: 10. 1038/nrg.2016.169

Holley, R. W., Apgar, J., Everett, G. A., Madison, J. T., Marquisee, M., Merrill, S. H., et al. (1965a). Structure of a ribonucleic acid. *Science* 147, 1462–1465. doi: 10.1126/science.147.3664.1462

Holley, R. W., Everett, G. A., Madison, J. T., and Zamir, A. (1965b). Nucleotide sequences in the yeast alanine transfer ribonucleic acid. *J. Biol. Chem.* 240, 2122–2128.

Hong, T., Yuan, Y., Chen, Z., Xi, K., Wang, T., Xie, Y., et al. (2018). Precise antibody-independent m6A identification via 4SedTTP-Involved and FTO-assisted strategy at single-nucleotide resolution. *J. Am. Chem. Soc.* 140, 5886–5889. doi: 10.1021/jacs.7b13633

Hori, H. (2014). Methylated nucleosides in tRNA and tRNA methyltransferases. *Front. Genet.* 5:144. doi: 10.3389/fgene.2014.00144

Hurwitz, J., Gold, M., and Anders, M. (1964). The enzymatic methylation of ribonucleic acid and deoxyribonucleic acid. Iv. The properties of the soluble ribonucleic acid-methylating enzymes. *J. Biol. Chem.* 239, 3474–3482.

Hussain, S., Aleksic, J., Blanco, S., Dietmann, S., and Frye, M. (2013). Characterizing 5-methylcytosine in the mammalian epitranscriptome. *Genome Biol.* 14:215. doi: 10.1186/gb4143

Incarnato, D., Anselmi, F., Morandi, E., Neri, F., Maldotti, M., Rapelli, S., et al. (2017). High-throughput single-base resolution mapping of RNA 2'-O-methylated residues. *Nucleic Acids Res.* 45, 1433–1441. doi: 10.1093/nar/gkw810

Ishida, K., Kunibayashi, T., Tomikawa, C., Ochi, A., Kanai, T., Hirata, A., et al. (2011). Pseudouridine at position 55 in tRNA controls the contents of other modified nucleotides for low-temperature adaptation in the extreme-thermophilic eubacterium *Thermus thermophilus*. *Nucleic Acids Res.* 39, 2304–2318. doi: 10.1093/nar/gkq1180

Jackman, J. E., and Alfonzo, J. D. (2013). Transfer RNA modifications: nature's combinatorial chemistry playground. *Wiley Interdiscip. Rev. RNA* 4, 35–48. doi: 10.1002/wrna.1144

Jacob, R., Zander, S., and Gutschner, T. (2017). The dark side of the epitranscriptome: chemical modifications in long non-coding RNAs. *Int. J. Mol. Sci.* 18:E2387. doi: 10.3390/ijms18112387

Jeltsch, A., Ehrenhofer-Murray, A., Jurkowski, T. P., Lyko, F., Reuter, G., Ankri, S., et al. (2017). Mechanism and biological role of Dnmt2 in Nucleic Acid Methylation. *RNA Biol.* 14, 1108–1123. doi: 10.1080/15476286.2016.1191737

Ji, L., and Chen, X. (2012). Regulation of small RNA stability: methylation and beyond. *Cell Res.* 22, 624–636. doi: 10.1038/cr.2012.36

Jia, G., Fu, Y., Zhao, X., Dai, Q., Zheng, G., Yang, Y., et al. (2011). N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat. Chem. Biol.* 7, 885–887. doi: 10.1038/nchembio.687

Kadaba, S., Krueger, A., Trice, T., Krecic, A. M., Hinnebusch, A. G., and Anderson, J. (2004). Nuclear surveillance and degradation of hypomodified initiator tRNAMet in *S. cerevisiae. Genes Dev.* 18, 1227–1240. doi: 10.1101/gad.1183804

Kaneko, T., Suzuki, T., Kapushoc, S. T., Rubio, M. A., Ghazvini, J., Watanabe, K., et al. (2003). Wobble modification differences and subcellular localization of tRNAs in *Leishmania tarentolae*: implication for tRNA sorting mechanism. *EMBO J.* 22, 657–667. doi: 10.1093/emboj/cdg066

Kawahara, Y., Zinshteyn, B., Sethupathy, P., Iizasa, H., Hatzigeorgiou, A. G., and Nishikura, K. (2007). Redirection of silencing targets by adenosine-to-inosine editing of miRNAs. *Science* 315, 1137–1140. doi: 10.1126/science. 1138050

Khoddami, V., and Cairns, B. R. (2013). Identification of direct targets and modified bases of RNA cytosine methyltransferases. *Nat. Biotechnol.* 31, 458–464. doi: 10.1038/nbt.2566

Khoddami, V., and Cairns, B. R. (2014). Transcriptome-wide target profiling of RNA cytosine methyltransferases using the mechanism-based enrichment procedure Aza-IP. *Nat. Protoc.* 9, 337–361. doi: 10.1038/nprot.2014.014

Klungland, A., Dahl, J. A., Greggains, G., Fedorcsak, P., and Filipczyk, A. (2016). Reversible RNA modifications in meiosis and pluripotency. *Nat. Methods* 14, 18–22. doi: 10.1038/nmeth.4111

Kowalak, J. A., Pomerantz, S. C., Crain, P. F., and Mccloskey, J. A. (1993). A novel method for the determination of post-transcriptional modification in RNA by mass spectrometry. *Nucleic Acids Res.* 21, 4577–4585. doi: 10.1093/nar/21. 19.4577

Krogh, N., Jansson, M. D., Hafner, S. J., Tehler, D., Birkedal, U., Christensen-Dalsgaard, M., et al. (2016). Profiling of 2'-O-Me in human rRNA reveals a subset of fractionally modified positions and provides evidence for ribosome heterogeneity. *Nucleic Acids Res.* 44, 7884–7895. doi: 10.1093/nar/gkw482

Lee, R. C., Feinbaum, R. L., and Ambros, V. (1993). The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* 75, 843–854. doi: 10.1016/0092-8674(93)90529-Y

Lewis, C. J., Pan, T., and Kalsotra, A. (2017). RNA modifications and structures cooperate to guide RNA-protein interactions. *Nat. Rev. Mol. Cell Biol.* 18, 202–210. doi: 10.1038/nrm.2016.163

Li, S., and Mason, C. E. (2014). The pivotal regulatory landscape of RNA modifications. *Annu. Rev. Genomics Hum. Genet.* 15, 127–150. doi: 10.1146/annurev-genom-090413-025405

Li, X., Xiong, X., Wang, K., Wang, L., Shu, X., Ma, S., et al. (2016a). Transcriptome-wide mapping reveals reversible and dynamic N(1)-methyladenosine methylome. *Nat. Chem. Biol.* 12, 311–316. doi: 10.1038/nchembio.2040

Li, X., Xiong, X., and Yi, C. (2016b). Epitranscriptome sequencing technologies: decoding RNA modifications. *Nat. Methods* 14, 23–31. doi: 10.1038/nmeth.4110

Li, Z., Weng, H., Su, R., Weng, X., Zuo, Z., Li, C., et al. (2017). FTO plays an oncogenic role in acute myeloid leukemia as a N(6)-Methyladenosine RNA demethylase. *Cancer Cell* 31, 127–141. doi: 10.1016/j.ccell.2016. 11.017

Liang, X. H., Liu, Q., and Fournier, M. J. (2009). Loss of rRNA modifications in the decoding center of the ribosome impairs translation and strongly delays pre-rRNA processing. *RNA* 15, 1716–1728. doi: 10.1261/rna.1724409

Lichinchi, G., Gao, S., Saletore, Y., Gonzalez, G. M., Bansal, V., Wang, Y., et al. (2016a). Dynamics of the human and viral m(6)A RNA methylomes during HIV-1 infection of T cells. *Nat. Microbiol.* 1:16011. doi: 10.1038/nmicrobiol. 2016.11

Lichinchi, G., Zhao, B. S., Wu, Y., Lu, Z., Qin, Y., He, C., et al. (2016b). Dynamics of human and viral RNA methylation during zika virus infection. *Cell Host Microbe* 20, 666–673. doi: 10.1016/j.chom.2016.10.002

Linder, B., Grozhik, A. V., Olarerin-George, A. O., Meydan, C., Mason, C. E., and Jaffrey, S. R. (2015). Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. *Nat. Methods* 12, 767–772. doi: 10.1038/nmeth. 3453

Little, N. A., Hastie, N. D., and Davies, R. C. (2000). Identification of WTAP, a novel Wilms' tumour 1-associating protein. *Hum. Mol. Genet.* 9, 2231–2239. doi: 10.1093/oxfordjournals.hmg.a018914

Liu, F., Clark, W., Luo, G., Wang, X., Fu, Y., Wei, J., et al. (2016). ALKBH1-mediated tRNA demethylation regulates translation. *Cell* 167, 816–828e816. doi: 10.1016/j.cell.2016.09.038

Liu, J., and Jia, G. (2014). Methylation modifications in eukaryotic messenger RNA. *J. Genet. Genomics* 41, 21–33. doi: 10.1016/j.jgg.2013.10.002

Liu, J., Yue, Y., Han, D., Wang, X., Fu, Y., Zhang, L., et al. (2014). A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. *Nat. Chem. Biol.* 10, 93–95. doi: 10.1038/nchembio. 1432

Liu, L., Zhang, S. W., Huang, Y., and Meng, J. (2017). QNB: differential RNA methylation analysis for count-based small-sample sequencing data with a

quad-negative binomial model. *BMC Bioinformatics* 18:387. doi: 10.1186/s12859-017-1808-4

Liu, N., Dai, Q., Zheng, G., He, C., Parisien, M., and Pan, T. (2015). N(6)-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions. *Nature* 518, 560–564. doi: 10.1038/nature14234

Liu, N., Parisien, M., Dai, Q., Zheng, G., He, C., and Pan, T. (2013). Probing N6-methyladenosine RNA modification status at single nucleotide resolution in mRNA and long noncoding RNA. *RNA* 19, 1848–1856. doi: 10.1261/rna.041178.113

Ma, J. Z., Yang, F., Zhou, C. C., Liu, F., Yuan, J. H., Wang, F., et al. (2017). METTL14 suppresses the metastatic potential of hepatocellular carcinoma by modulating N(6)-methyladenosine-dependent primary microRNA processing. *Hepatology* 65, 529–543. doi: 10.1002/hep.28885

Maden, B. E., Corbett, M. E., Heeney, P. A., Pugh, K., and Ajuh, P. M. (1995). Classical and novel approaches to the detection and localization of the numerous modified nucleotides in eukaryotic ribosomal RNA. *Biochimie* 77, 22–29. doi: 10.1016/0300-9084(96)88100-4

Marchand, V., Blanloeil-Oillo, F., Helm, M., and Motorin, Y. (2016). Illumina-based RiboMethSeq approach for mapping of 2'-O-Me residues in RNA. *Nucleic Acids Res.* 44:e135. doi: 10.1093/nar/gkw547

Marchand, V., Pichot, F., Thuring, K., Ayadi, L., Freund, I., Dalpke, A., et al. (2017). Next-generation sequencing-based RiboMethSeq protocol for analysis of tRNA 2'-O-methylation. *Biomolecules* 7:E13. doi: 10.3390/biom7010013

Marck, C., and Grosjean, H. (2002). tRNomics: analysis of tRNA genes from 50 genomes of Eukarya, Archaea, and Bacteria reveals anticodon-sparing strategies and domain-specific features. *RNA* 8, 1189–1232. doi: 10.1017/S1355838202022021

Martinez, F. J., Lee, J. H., Lee, J. E., Blanco, S., Nickerson, E., Gabriel, S., et al. (2012). Whole exome sequencing identifies a splicing mutation in NSUN2 as a cause of a Dubowitz-like syndrome. *J. Med. Genet.* 49, 380–385. doi: 10.1136/jmedgenet-2011-100686

McCloskey, J. A., and Nishimura, S. (1977). Modified nucleosides in transfer RNA. *Acc. Chem. Res.* 10, 403–410. doi: 10.1021/ar50119a004

Meier, U. T. (2011). Pseudouridylation goes regulatory. *EMBO J.* 30, 3–4. doi: 10.1038/emboj.2010.323

Metodiev, M. D., Spahr, H., Loguercio Polosa, P., Meharg, C., Becker, C., Altmueller, J., et al. (2014). NSUN4 is a dual function mitochondrial protein required for both methylation of 12S rRNA and coordination of mitoribosomal assembly. *PLoS Genet.* 10:e1004110. doi: 10.1371/journal.pgen.1004110

Meyer, K. D., Patil, D. P., Zhou, J., Zinoviev, A., Skabkin, M. A., Elemento, O., et al. (2015). 5' UTR m(6)A promotes cap-independent translation. *Cell* 163, 999–1010. doi: 10.1016/j.cell.2015.10.012

Meyer, K. D., Saletore, Y., Zumbo, P., Elemento, O., Mason, C. E., and Jaffrey, S. R. (2012). Comprehensive analysis of mRNA methylation reveals enrichment in 3'. UTRs and near stop codons. *Cell* 149, 1635–1646. doi: 10.1016/j.cell.2012.05.003

Militello, K. T., Chen, L. M., Ackerman, S. E., Mandarano, A. H., and Valentine, E. L. (2014). A map of 5-methylcytosine residues in *Trypanosoma brucei* tRNA revealed by sodium bisulfite sequencing. *Mol. Biochem. Parasitol.* 193, 122–126. doi: 10.1016/j.molbiopara.2013.12.003

Mishima, E., Jinno, D., Akiyama, Y., Itoh, K., Nankumo, S., Shima, H., et al. (2015). Immuno-northern blotting: detection of RNA modifications by using antibodies against modified nucleosides. *PLoS One* 10:e0143756. doi: 10.1371/journal.pone.0143756

Motorin, Y., Muller, S., Behm-Ansmant, I., and Branlant, C. (2007). Identification of modified residues in RNAs by reverse transcription-based methods. *Methods Enzymol.* 425, 21–53. doi: 10.1016/S0076-6879(07)25002-5

Muller, S., Windhof, I. M., Maximov, V., Jurkowski, T., Jeltsch, A., Forstner, K. U., et al. (2013). Target recognition, RNA methylation activity and transcriptional regulation of the *Dictyostelium discoideum* Dnmt2-homologue (DnmA). *Nucleic Acids Res.* 41, 8615–8627. doi: 10.1093/nar/gkt634

Nachtergaele, S., and He, C. (2017). The emerging biology of RNA post-transcriptional modifications. *RNA Biol.* 14, 156–163. doi: 10.1080/15476286.2016.1267096

Nigita, G., Acunzo, M., Romano, G., Veneziano, D., Lagana, A., Vitiello, M., et al. (2016). microRNA editing in seed region aligns with cellular changes in hypoxic conditions. *Nucleic Acids Res.* 44, 6298–6308. doi: 10.1093/nar/gkw532

Nigita, G., Veneziano, D., and Ferro, A. (2015). A-to-I RNA editing: current knowledge sources and computational approaches with special emphasis on non-coding RNA molecules. *Front. Bioeng. Biotechnol.* 3:37. doi: 10.3389/fbioe.2015.00037

Nishikura, K. (2016). A-to-I editing of coding and non-coding RNAs by ADARs. *Nat. Rev. Mol. Cell Biol.* 17, 83–96. doi: 10.1038/nrm.2015.4

Patil, D. P., Chen, C. K., Pickering, B. F., Chow, A., Jackson, C., Guttman, M., et al. (2016). m(6)A RNA methylation promotes XIST-mediated transcriptional repression. *Nature* 537, 369–373. doi: 10.1038/nature19342

Perry, R. P., Kelley, D. E., Friderici, K., and Rottman, F. (1975). The methylated constituents of L cell messenger RNA: evidence for an unusual cluster at the 5' terminus. *Cell* 4, 387–394. doi: 10.1016/0092-8674(75)90159-2

Phizicky, E. M., and Hopper, A. K. (2015). tRNA processing, modification, and subcellular dynamics: past, present, and future. *RNA* 21, 483–485. doi: 10.1261/rna.049932.115

Reitz, C., Tosto, G., Mayeux, R., and Luchsinger, J. A. (2012). NIA-LOAD/NCRAD family study group, Alzheimer's disease neuroimaging initiative. genetic variants in the fat and obesity associated (FTO) gene and risk of Alzheimer's disease. *PLoS One* 7:e50354. doi: 10.1371/journal.pone.0050354

Rottman, F., Shatkin, A. J., and Perry, R. P. (1974). Sequences containing methylated nucleotides at the 5' termini of messenger RNAs: possible implications for processing. *Cell* 3, 197–199. doi: 10.1016/0092-8674(74)90131-7

Roundtree, I. A., Evans, M. E., Pan, T., and He, C. (2017). Dynamic RNA modifications in gene expression regulation. *Cell* 169, 1187–1200. doi: 10.1016/j.cell.2017.05.045

Saletore, Y., Meyer, K., Korlach, J., Vilfan, I. D., Jaffrey, S., and Mason, C. E. (2012). The birth of the epitranscriptome: deciphering the function of RNA modifications. *Genome Biol.* 13:175. doi: 10.1186/gb-2012-13-10-175

Schaefer, M., Pollex, T., Hanna, K., Tuorto, F., Meusburger, M., Helm, M., et al. (2010). RNA methylation by Dnmt2 protects transfer RNAs against stress-induced cleavage. *Genes Dev.* 24, 1590–1595. doi: 10.1101/gad.586710

Schibler, U., and Perry, R. P. (1977). The 5'-termini of heterogeneous nuclear RNA: a comparison among molecules of different sizes and ages. *Nucleic Acids Res.* 4, 4133–4149. doi: 10.1093/nar/4.12.4133

Schwartz, S., Mumbach, M. R., Jovanovic, M., Wang, T., Maciag, K., Bushkin, G. G., et al. (2014). Perturbation of m6A writers reveals two distinct classes of mRNA methylation at internal and 5' sites. *Cell Rep.* 8, 284–296. doi: 10.1016/j.celrep.2014.05.048

Shafik, A., Schumann, U., Evers, M., Sibbritt, T., and Preiss, T. (2016). The emerging epitranscriptomics of long noncoding RNAs. *Biochim. Biophys. Acta* 1859, 59–70. doi: 10.1016/j.bbagrm.2015.10.019

Sharma, S., Watzinger, P., Kotter, P., and Entian, K. D. (2013). Identification of a novel methyltransferase, Bmt2, responsible for the N-1-methyl-adenosine base modification of 25S rRNA in Saccharomyces cerevisiae. *Nucleic Acids Res.* 41, 5428–5443. doi: 10.1093/nar/gkt195

Song, J., and Yi, C. (2017). Chemical modifications to RNA: a new layer of gene expression regulation. *ACS Chem. Biol.* 12, 316–325. doi: 10.1021/acschembio.6b00960

Sprinzl, M., and Vassilenko, K. S. (2005). Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res.* 33, D139–D140.

Squires, J. E., Patel, H. R., Nousch, M., Sibbritt, T., Humphreys, D. T., Parker, B. J., et al. (2012). Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA. *Nucleic Acids Res.* 40, 5023–5033. doi: 10.1093/nar/gks144

Sun, W. J., Li, J. H., Liu, S., Wu, J., Zhou, H., Qu, L. H., et al. (2016). RMBase: a resource for decoding the landscape of RNA modifications from high-throughput sequencing data. *Nucleic Acids Res.* 44, D259–D265. doi: 10.1093/nar/gkv1036

Tan, A., Dang, Y., Chen, G., and Mo, Z. (2015). Overexpression of the fat mass and obesity associated gene (FTO) in breast cancer and its clinical implications. *Int. J. Clin. Exp. Pathol.* 8, 13405–13410.

Todd, G., and Karbstein, K. (2007). RNA takes center stage. *Biopolymers* 87, 275–278. doi: 10.1002/bip.20824

Tomikawa, C., Yokogawa, T., Kanai, T., and Hori, H. (2010). N7-Methylguanine at position 46 (m7G46) in tRNA from *Thermus thermophilus* is required for cell viability at high temperatures through a tRNA modification network. *Nucleic Acids Res.* 38, 942–957. doi: 10.1093/nar/gkp1059

Van Haute, L., Dietmann, S., Kremer, L., Hussain, S., Pearce, S. F., Powell, C. A., et al. (2016). Deficient methylation and formylation of mt-tRNA(Met) wobble cytosine in a patient carrying mutations in NSUN3. *Nat. Commun.* 7:12039. doi: 10.1038/ncomms12039

Vandivier, L. E., and Gregory, B. D. (2017). Reading the epitranscriptome: new techniques and perspectives. *Enzymes* 41, 269–298. doi: 10.1016/bs.enz.2017.03.004

Veneziano, D., Di Bella, S., Nigita, G., Lagana, A., Ferro, A., and Croce, C. M. (2016). Noncoding RNA: current deep sequencing data analysis approaches and challenges. *Hum. Mutat.* 37, 1283–1298. doi: 10.1002/humu.23066

Veneziano, D., Nigita, G., and Ferro, A. (2015). Computational approaches for the analysis of ncRNA through deep sequencing techniques. *Front. Bioeng. Biotechnol.* 3:77. doi: 10.3389/fbioe.2015.00077

Voigts-Hoffmann, F., Hengesbach, M., Kobitski, A. Y., Van Aerschot, A., Herdewijn, P., Nienhaus, G. U., et al. (2007). A methyl group controls conformational equilibrium in human mitochondrial tRNA(Lys). *J. Am. Chem. Soc.* 129, 13382–13383. doi: 10.1021/ja075520+

Wang, X., and He, C. (2014). Reading RNA methylation codes through methyl-specific binding proteins. *RNA Biol.* 11, 669–672. doi: 10.4161/rna.28829

Wang, X., Lu, Z., Gomez, A., Hon, G. C., Yue, Y., Han, D., et al. (2014). N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* 505, 117–120. doi: 10.1038/nature12730

Wang, X., Zhao, B. S., Roundtree, I. A., Lu, Z., Han, D., Ma, H., et al. (2015). N(6)-methyladenosine modulates messenger RNA translation efficiency. *Cell* 161, 1388–1399. doi: 10.1016/j.cell.2015.05.014

Yang, X., Yang, Y., Sun, B. F., Chen, Y. S., Xu, J. W., Lai, W. Y., et al. (2017). 5-methylcytosine promotes mRNA export - NSUN2 as the methyltransferase and ALYREF as an m(5)C reader. *Cell Res.* 27, 606–625. doi: 10.1038/cr.2017.55

Yi, J., Gao, R., Chen, Y., Yang, Z., Han, P., Zhang, H., et al. (2017). Overexpression of NSUN2 by DNA hypomethylation is associated with metastatic progression in human breast cancer. *Oncotarget* 8, 20751–20765. doi: 10.18632/oncotarget.10612

Yokoyama, S., Watanabe, K., and Miyazawa, T. (1987). Dynamic structures and functions of transfer ribonucleic acids from extreme thermophiles. *Adv. Biophys.* 23, 115–147. doi: 10.1016/0065-227X(87)90006-2

Yu, B., Yang, Z., Li, J., Minakhina, S., Yang, M., Padgett, R. W., et al. (2005). Methylation as a crucial step in plant microRNA biogenesis. *Science* 307, 932–935. doi: 10.1126/science.1107130

Zhang, S., Zhao, B. S., Zhou, A., Lin, K., Zheng, S., Lu, Z., et al. (2017). m6A demethylase ALKBH5 maintains tumorigenicity of glioblastoma stem-like cells by Sustaining FOXM1 expression and cell proliferation program. *Cancer Cell* 31, 591–606. doi: 10.1016/j.ccell.2017.02.013

Zhang, Y. C., Zhang, S. W., Liu, L., Liu, H., Zhang, L., Cui, X., et al. (2015). Spatially enhanced differential RNA methylation analysis from affinity-based sequencing data with hidden markov model. *Biomed Res. Int.* 2015:852070. doi: 10.1155/2015/852070

Zheng, G., Dahl, J. A., Niu, Y., Fedorcsak, P., Huang, C. M., Li, C. J., et al. (2013). ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Mol. Cell* 49, 18–29. doi: 10.1016/j.molcel.2012.10.015

Zhou, K. I., Parisien, M., Dai, Q., Liu, N., Diatchenko, L., Sachleben, J. R., et al. (2016). N^6-Methyladenosine modification in a long noncoding RNA hairpin predisposes its conformation to protein binding. *J. Mol. Biol.* 428, 822–833. doi: 10.1016/j.jmb.2015.08.021

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The Role of Noncoding RNA Pseudouridylation in Nuclear Gene Expression Events

Yang Zhao[1], William Dunker[1], Yi-Tao Yu[2]* and John Karijolich[1,3]*

[1] Department of Pathology, Microbiology, and Immunology, School of Medicine, Vanderbilt University, Nashville, TN, United States, [2] Department of Biochemistry and Biophysics, Center for RNA Biology, School of Medicine and Dentistry, University of Rochester, Rochester, NY, United States, [3] Vanderbilt-Ingram Cancer Center, Nashville, TN, United States

Pseudouridine is the most abundant internal RNA modification in stable noncoding RNAs (ncRNAs). It can be catalyzed by both RNA-dependent and RNA-independent mechanisms. Pseudouridylation impacts both the biochemical and biophysical properties of RNAs and thus influences RNA-mediated cellular processes. The investigation of nuclear-ncRNA pseudouridylation has demonstrated that it is critical for the proper control of multiple stages of gene expression regulation. Here, we review how nuclear-ncRNA pseudouridylation contributes to transcriptional regulation and pre-mRNA splicing.

INTRODUCTION

The proper control of gene expression in the nucleus is achieved by the actions of a diverse set of factors. In addition to proteins, noncoding RNAs (ncRNAs) participate in most, if not all, stages of nuclear gene expression, including both RNA polymerase II (Pol II) transcription and pre-mRNA splicing (Mercer et al., 2009). ncRNAs employ numerous mechanisms to accomplish their function, including binding to and modulating protein function, base pairing with complementary nucleic acids, or directly catalyzing biochemical reactions. Like proteins, proper modification and folding into higher-order structures are a prerequisite for their function.

In addition to the four canonical nucleosides, more than 140 chemically distinct modified RNA nucleosides have been identified in nature (Machnicka et al., 2013). Pseudouridine (ψ), first discovered over 60 years ago (Cohn and Elliot, 1951), is the most abundant internal RNA modification in stable RNAs. ψ, the C5-glycoside isomer of uridine, is formed through an internal transglycosylation reaction in which the N1–C1' bond between the uracil base and the ribose sugar is broken and a C5–C1' glycosidic bond is reformed (**Figure 1**). As a consequence of isomerization, an additional hydrogen bond donor is present at the non-Watson–Crick edge. The distinct structure of ψ increases both the rigidity of the phosphodiester backbone and the thermodynamic stability of ψ–A compared with U–A. This effect is mediated by water-coordinated hydrogen bonding and base stacking (Charette and Gray, 2000).

Initial evidence for a functional role of ψ partially came from the fact that ψ residues are clustered in functionally important and evolutionarily conserved regions of tRNA (Grosjean et al., 1995; Hopper and Phizicky, 2003), rRNA (Branlant et al., 1981; Maden, 1990), and small nuclear RNA (snRNA) (Reddy and Busch, 1988; Massenet et al., 1998; Narlikar et al., 2002; Karijolich and Yu, 2010). Indeed, experimental data have confirmed important roles for pseudouridylation in multiple aspects of gene expression regulation, including spliceosomal small nuclear ribonucleoprotein (snRNP) biogenesis, efficiency of pre-mRNA splicing, and translation fidelity (Karijolich et al., 2010). Here, we will provide an overview of the mechanisms of pseudouridylation and then highlight

several pseudouridylated ncRNAs and their effects on nuclear gene expression events, specifically transcription and pre-mRNA splicing.

FIGURE 1 | Schematic of the pseudouridylation reaction. The isomerization of uridine (U) to pseudouridine (ψ) is mediated by pseudouridine synthases (PUSs). It results in an extra hydrogen bond donor (d) and the same number of hydrogen bond acceptors (a).
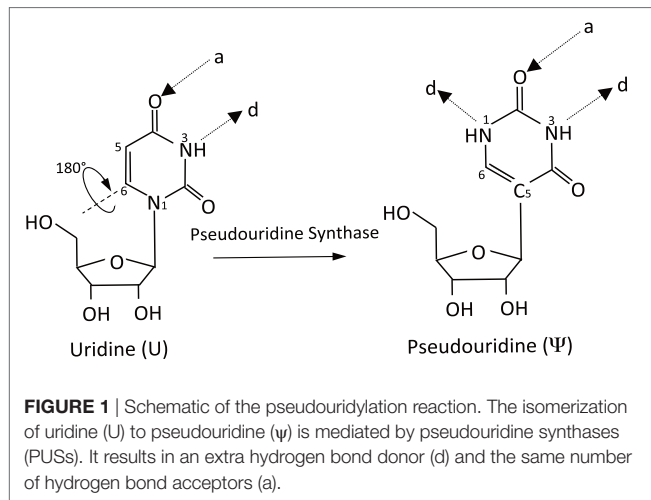
MECHANISMS OF PSEUDOURIDYLATION

The past decades have seen remarkable progress toward defining mechanisms by which pseudouridylation is catalyzed. Pseudouridylation of ncRNA is catalyzed by pseudouridine synthases (PUSs) through two distinct mechanisms, namely RNA-dependent pseudouridylation and RNA-independent pseudouridylation.

RNA-DEPENDENT PSEUDOURIDYLATION

The RNA-dependent pseudouridylation machinery consists of one unique box H/ACA RNA and four core proteins. Box H/ACA RNAs are one of the most evolutionarily conserved families of small ncRNAs and are present in all eukaryotes. They function as guide RNAs to direct pseudouridylation in mRNA, rRNA, spliceosomal snRNAs, and various other types of ncRNAs. With a median length of 133 nucleotides (nts), eukaryotic box H/ACA RNAs adopt a hairpin-hinge–hairpin-tail secondary structure (**Figure 2A**). Within the internal hinge region, there is a H box motif (ANANNA), while an ACA box motif (ACA) is positioned near the 3' end. Within each hairpin structure,

FIGURE 2 | Mechanism of pseudouridylation. **(A)** Schematic of RNA-dependent pseudouridylation by box H/ACA RNP. The box H/ACA RNP is composed of the box H/ACA RNA and four core proteins (Cbf5, Nhp2, Nop10, and Gar1). The secondary structure of a eukaryotic pseudouridylation guide box H/ACA RNA is shown as a blue line. The RNA adopts a hairpin-hinge–hairpin-tail structure. The box H (5'-ANANNA-3') within the hinge region and the box ACA (5'-ACA-3') motif at the 3'-end of the RNA are highlighted in a yellow box. The pseudouridylation pocket (thick-blue line) facilitates substrate recognition *via* complementary base-pairing interactions between the box H/ACA RNA and the substrate RNA (green line). **(B)** The standalone pseudouridine synthase (PUS), PUS7, recognizes the consensus sequence of substrates U2 small nuclear RNA (snRNA) and catalyzes the pseudouridine formation at U35. **(C)** Inducible pseudouridylation of U2 snRNA. Top, the 3' pocket of small nucleolar RNA 81 (snR81) base pairs with the sequence surrounding nucleotide (nt) U93 with two mismatches (denoted as red crosses). Bottom, the sequence surrounding U56 recognized by PUS7 has three nts (highlighted in red), different from the consensus PUS7 recognition sequence.

an internal loop (pseudouridylation pocket) is present, which facilitates substrate recognition *via* complementary base-pairing interactions between the box H/ACA RNA (the loop sequence) and the substrate RNA. Both hairpins of H/ACA RNAs can carry functional pseudouridylation pockets and can thus independently direct pseudouridylation of uridines in separate RNAs or separate uridines located within the same substrate RNA. In the guide RNA–target RNA interaction, the guide sequences hybridize 5' and one nt 3' of the target uridine to the substrate RNA immediately, thereby framing it. The distance between the target uridine and the H or ACA box of the guide RNA is usually 14–15 nts (Kiss et al., 2010).

The four core proteins associated with box H/ACA RNAs are Cbf5 (dyskerin in human and NAP57 in rodents), Nhp2 (L7Ae in archaeal), Gar1, and Nop10 (Yu and Meier, 2014). Cbf5 is the enzymatic component of the RNP and catalyzes the U-to-ψ isomerization reaction (**Figure 2A**). Structures of the enzymes from various species show a high degree of evolutionary conservation, especially in the PUS and Archaeosine transglycosylase (PUA) domain (Hamma et al., 2005; Manival et al., 2006; Rashid et al., 2006). The other three core proteins are also essential, and the depletion of them in yeast, with the exception of GAR1, causes the loss of all H/ACA RNAs (Girard et al., 1992; Bousquet-Antonelli et al., 1997).

Facilitated by the development of *in vitro* systems for the reconstitution of enzymatically active RNP complexes, the crystal structure of the box H/ACA RNP was first solved using archaeal components (Li and Ye, 2006). These experiments demonstrated that the L7Ae, Nop10, and the catalytic domain of Cbf5 bound to the upper stem of the guide RNA, whereas the PUA domain of Cbf5 anchored the lower stem and ACA motif. The substrate RNA is recruited and the target uridine is precisely placed within the active site *via* complementary base-pairing interactions with the bipartite guides, while extensive protein interactions help to stabilize the interaction. In contrast to the other core proteins, Gar1 does not physically interact with the box H/ACA guide RNA or substrate RNA. Instead, Gar1 interacts directly with the thumb loop of Cbf5 and participates in regulating substrate turnover. Studies in yeast have revealed that the structure of the eukaryotic box H/ACA RNP is highly similar to the archaeal one, however, with exceptions. In particular is the independence of the RNP's activity from Nhp2 binding and a novel C-terminal extension in Gar1, which interacts with Cbf5. It is hypothesized that these functional and structural differences reflect the evolutionary adaptations of eukaryotic box H/ACA RNP to the variable RNA structure and moderate temperature range in which eukaryotes live in, respectively (Li et al., 2011).

RNA-INDEPENDENT PSEUDOURIDYLATION

RNA-independent pseudouridylation in eukaryotes acts through standalone enzymes called PUS enzymes. In contrast to the box H/ACA RNP-based mechanism, PUS enzymes carry out both substrate recognition and the internal transglycosylation reaction. Substrate recognition is achieved *via* consensus sequences and/or

secondary structure elements of the substrate RNA (**Figure 2B**). In eukaryotes, there are 10 different PUS enzymes, numbered PUS1 through PUS10. These are classified into five families (TruA, TruB, TruD, RluA, and PUS10) based on their bacterial counterparts (Hamma and Ferre-D'Amare, 2006). Although the primary sequences have diverged, all PUSs, including Cbf5, share a conserved catalytic domain and likely a conserved catalytic mechanism based on the solved crystal structure (Foster et al., 2000; Hoang and Ferre-D'Amare, 2001, 2004; Sivaraman et al., 2002, 2004; Del Campo et al., 2004; Ericsson et al., 2004; Kaya et al., 2004; Mizutani et al., 2004; Hoang et al., 2006; McCleverty et al., 2007). This domain structure is composed predominately of anti-parallel β-sheets, with one face decorated by two groups of α-helices and loops. A forefinger–thumb structure formed by these loops pinches the target RNA, while the strictly conserved catalytic aspartate residue participates in the enzymatic reaction.

Unlike the box H/ACA RNPs, which have been found only to reside within the nucleus, PUS enzymes have been found in the nucleus, cytoplasm, and mitochondria. Each PUS enzyme targets either one specific or multiple uridines in many RNA species, including snRNAs, rRNAs, and tRNAs in both cytoplasm and mitochondria.

PSEUDOURIDYLATION IS INDUCIBLE

Until relatively recently, RNA modifications, including pseudouridylation, were considered constitutive. In 2011, Wu et al. (2011) provided the first evidence that RNA modifications were inducible and demonstrated that yeast U2 snRNA was conditionally pseudouridylated when cells were subjected to nutrient deprivation or heat shock (Wu et al., 2011). In addition to the three constitutive pseudouridines (ψ35, ψ42, and ψ44) of yeast U2 snRNA, two novel pseudouridines (ψ56 and ψ93) were detected in stressed cells. Further detailed analyses revealed that both the RNA-independent PUS (PUS7 catalyzes ψ56 formation) and the box H/ACA RNP-dependent [small nucleolar RNA 81 (snR81)-guided box H/ACA RNP catalyzes ψ93 formation] modification machineries are involved in inducible pseudouridylation. Interestingly, both ψ56 and ψ93 are "imperfect" substrates for their respective modification machineries. The sequences flanking positions U56 and U93 in U2 snRNA are similar, but not identical to the sequences surrounding the constitutively pseudouridylated targets of PUS7 and snR81, respectively. For example, the 3' pocket of snR81 base pairs with the sequence surrounding nt U93 of U2 snRNA with two mismatches. In addition, ψ56 formation is mediated by PUS7 engaging a substrate whose sequence differs by three nts from the consensus PUS7 recognition sequence (**Figure 2C**). Inducible pseudouridylation is also functionally relevant, as demonstrated by the observation that the artificial introduction of ψ93 reduces the efficiency of pre-mRNA splicing. Interestingly, it was recently shown that the TOR-signaling pathway regulates ψ93 formation (Wu et al., 2016b).

Inducible pseudouridylation of other snRNAs has been reported subsequently. U6 snRNA is inducibly pseudouridylated at U28 by PUS1 during the yeast filamentous growth program (Basak and Query, 2014). Further analysis of mutants indicates that U6–ψ28 is functionally relevant, as all U6 snRNA mutations

that resulted in strong pseudouridylation at position U28 also exhibited a pseudohyphal growth phenotype, whereas blocking U6–ψ28 formation prevents filamentous growth.

Recently, transcriptome-wide mapping of pseudouridines in yeast and human cells has revealed its presence in mRNA (Carlile et al., 2014; Lovejoy et al., 2014; Schwartz et al., 2014). While mRNA pseudouridylation appears to be primarily catalyzed by the standalone PUSs (PUS1–PUS4, PUS6, PUS7, and PUS9), several pseudouridine residues are catalyzed by box H/ACA RNPs. Remarkably, mRNA pseudouridylation was also found to be highly inducible and in a stress-specific manner. For example, by comparing mRNA ψ profiles of untreated cells to those of cells exposed to heat shock or H_2O_2, it was found that the inducible pseudouridylation profiles were largely nonoverlapping (Li et al., 2015).

FUNCTION OF ncRNA PSEUDOURIDYLATION IN TRANSCRIPTION

Transcription is the primary control point for gene expression. It therefore determines cellular function and cell identity and is subjected to tight regulation to achieve a high degree of specificity and efficiency. The eukaryotic DNA template is packaged by histone proteins into a highly condensed structure called chromatin. The chromatin structure is dynamically regulated by both histone modifications and chromatin-remodeling factors (Narlikar et al.,

2002). Promoters contain elements that bind to transcriptional activators and repressors, as well as the transcription machinery (Smale and Kadonaga, 2003; Kadonaga, 2004). RNA Pol II is the enzyme to catalyze the transcription reaction of mRNA from DNA. Pol II is recruited to promoters by transcriptional activators in a holoenzyme form together with general transcription factors and a multiprotein complex called the Srb/Mediator (Bjorklund and Kim, 1996). Following transcription initiation, Pol II transits to a productive elongation status through interactions with multiple elongation factors (Zhou et al., 2012). Given the central role of transcription in gene expression, it is not surprising that transcription is subject to diverse steps of regulation. Here, we will discuss two pseudouridylated ncRNAs and their function in regulating RNA Pol II transcription (**Figure 3**).

STEROID RECEPTOR RNA ACTIVATOR (SRA) AND TRANSCRIPTION PREINITIATION

One layer of transcriptional control comes from the binding of activator and repressor proteins to the promoters of target genes in a sequence-specific manner (Smale and Kadonaga, 2003). Coactivators and corepressors, which interact with activators and repressors, are required to achieve optimal transcriptional regulation in cells (Kadonaga, 2004). The SRA was first identified as a transcriptional coactivator for several steroid-hormone receptors, including receptors for androgens (ARs), estrogens

FIGURE 3 | Functions of noncoding RNA (ncRNA) pseudouridylation during transcription. The secondary structures of ncRNAs are shown, and the pseudouridylation sites are denoted as a red star. The green arrow and red-blocking arrow highlight the ncRNAs that are known to regulate polymerase II (Pol II) transcription, the factors they target and whether the effect of the ncRNA on transcription is stimulatory or inhibitory, respectively. HIV-1 LTR, human immunodeficiency virus-1 long terminal repeat; HRE, hormone response element; HIV-1 TAR, human immunodeficiency virus-1 transactivation response; NHR, nuclear hormone receptor; P, phosphoryl group; P-TEFb, positive transcription–elongation factor-b; SRA; steroid receptor RNA activator; SRC-1, steroid receptor activator-1.

(ERs), glucocorticoids (GRs), and progestins (PRs) (Lanz et al., 1999). Interestingly, while it was initially presumed that a protein encoded by a specific 5'-spliced variant of the SRA gene was the functional factor, subsequent experiments demonstrated that the factor was an ncRNA.

Steroid receptor RNA activator operates as part of a ribonucleoprotein complex containing steroid receptor activator-1, which is an AF-2 coactivator (Lanz et al., 1999). Computer-assisted modeling suggests that SRA adopts a highly complex secondary structure containing 11 topological substructures (STRs). Mutagenesis of each STR indicated that 5 of 11 STRs are required for SRA to coactivate transcription, and STR7 is the most important one for SRA function (Lanz et al., 2002).

In a study to identify coactivators for retinoic acid receptor (RAR) in mouse S91 melanoma cells, mPUS1p was unexpectedly identified. In addition, SRA turned out to be a substrate of mPUS1p (Zhao et al., 2004). Using chromatin immunoprecipitation, RAR, PUS1p, and SRA were found to cooccupy the retinoic acid response promoter in a ligand-independent complex. PUS1-mediated pseudouridylation of SRA promotes the formation of an "active" structure and aids in establishing the transcription preinitiation complex upon ligand binding. Further supporting a role of SRA pseudouridylation in transcriptional regulation, mutations in PUS1p that disrupt its interaction with RAR or its pseudouridylation activity attenuate the activation of RAR-dependent transcription. mPUS1p also significantly augmented transactivation by other nuclear receptors (NRs) including thyroid hormone receptor (TR), GR, AR, PR, and ER, illustrating that this mechanism likely applies universally to the regulation of NR-dependent transcription.

In addition to mPUS1p, mPUS3p also modifies SRA and serves as an NR coactivator (Zhao et al., 2007). Unlike mPUS1p, mPUS3p does not enhance sex steroid receptor activity, suggesting that substrate-site specificity may have distinct roles. Indeed, in vitro mPUS1p and mPUS3p generally modify different positions in SRA, with a few positions commonly targeted. Intriguingly, the order of modification of SRA by mPUS1p and mPUS3p determines the positions within SRA that are required to be pseudouridylated. However, it is important to note that the only in vivo-pseudouridylated site identified in SRA is U206. Interestingly, a U206A mutation, which promotes hyperpseudouridylation of SRA in vitro, switches SRA from a coactivator to a molecule with dominant-negative activity in vivo.

Pseudouridylation of SRA by mPUS1p and mPUS3p is a highly complex posttranscriptional mechanism that controls a coactivator–corepressor switch in SRA with major consequences for NR signaling (Zhao et al., 2007). Unexpectedly, pseudouridylation of SRA occurs in a stem-loop structure STR5 (at position U206), whose secondary structure was not shown to be important for SRA function. Moreover, the thermodynamic and secondary structure differences between STR5 in hSRA-WT and hSRA-U206A are relatively minor, which further suggest that the secondary structure remodeling is unlikely to explain the large biochemical and functional effects observed. Instead, it is proposed that ψ206 stabilizes stems I and II of STR5 in a higher-order conformation through the base-stacking-enhancing properties of ψ, resulting in masking new sites and preventing hyperpseudouridylation by

mPUS1p and mPUS3p. This in turn may interfere with the binding of SRA to other proteins that define its function as a scaffold for both repressors and activators.

The physiological importance of PUS1p-mediated SRA pseudouridylation is illustrated by a disorder known as mitochondrial myopathy and sideroblastic anemia (MLASA). It is caused by an inactivating mutation in human PUS1p (Bykhovskaya et al., 2004). Some abnormalities in these patients, such as facial dysmorphisms, are suggested to be the consequence of defective hSRA–NR signaling (Fernandez-Vizarra et al., 2007). In addition, ERs and ARs are important targets for cancer therapy. Given the importance of SRA in NR-transcriptional regulation, coupled with the functional significance of PUS1p-mediated SRA pseudouridylation, recent work has suggested that the disruption of SRA pseudouridylation could serve as a novel RNA-based cancer therapeutic (Ghosh et al., 2012).

7SK snRNA AND TRANSCRIPTION ELONGATION

For the past three decades, most of the attention has been put on the early stages of the transcription cycle involving the recruitment of Pol II to gene promoters and assembly of active preinitiation complexes, which were thought to be the principal points where transcription was controlled (Kuras and Struhl, 1999; Ptashne, 2005). In recent years, however, accumulating evidences indicate that the subsequent stages of the transcription cycle are also highly regulated (Guenther et al., 2007; Muse et al., 2007). Notably, the promoter-proximal pausing and release of Pol II has been identified as a major rate-limiting step for controlling the expression of many metazoan genes and plays a critical role in cell growth, renewal, and differentiation (Levine, 2011; Zhou et al., 2012).

Shortly after initiation, Pol II is paused at a promoter-proximal region by negative elongation factors NELF and DSIF, resulting in a short nascent transcript ~12 nts in length. Promoter-proximately paused Pol II is the major form of Pol II found on metazoan chromosomes and is poised for entry into productive elongation. The positive transcription elongation factor-b (P-TEFb) is the major factor required to overcome this restriction. P-TEFb, composed of cyclin-dependent kinase 9 (CDK9) and cyclin T1, phosphorylates and thereby antagonizes the inhibitory actions of NELF and DSIF, triggering the release of Pol II from promoter-proximal pausing. In addition to this, P-TEFb also phosphorylates the C-terminal domain of the largest subunit of Pol II, which then serves as a platform for assembling key transcription and RNA-processing factors that promote transcriptional elongation, cotranscriptional processing of pre-mRNA, and lastly termination (Zhou et al., 2012).

Under normal growth conditions, up to 90% of cellular P-TEFb is sequestered in an inactive complex called the 7SK RNP (Yang et al., 2001). Within this complex, a noncoding RNA, namely 7SK snRNA, functions as a scaffold and mediates the interaction of P-TEFb with HEXIM1 or -2 and thus inhibits CDK9's kinase activity (Yik et al., 2003). 7SK snRNA, transcribed by RNA Pol III, is an abundant noncoding RNA and highly conserved in

higher eukaryotes (Wassarman and Steitz, 1991). Although its levels remain relatively constant, several genome-wide studies suggested that 7SK was pseudouridylated (Kishore et al., 2013; Carlile et al., 2014). Indeed, site-specific and quantitative pseudouridylation assays demonstrated that up to 94% of 7SK snRNA in HeLa cells is pseudouridylated at residue U250. Although the guide RNA has not been identified, it is clear that the box H/ACA RNP machinery catalyzes this modification as the depletion of DKC1 significantly reduces 7SK snRNA pseudouridylation. In addition, 7SK snRNA pseudouridylation was demonstrated to play a critical role in regulating the formation of the 7SK-P-TEFb snRNP, as mutation of U250, or depletion of DKC1, reduced the binding of CDK9 and HEXIM to 7SK snRNA (Zhao et al., 2016).

The identification of a key role for pseudouridylation in 7SK snRNP stability has potential clinical relevance. For instance, one emerging strategy in curing human immunodeficiency virus (HIV) infection is to "shock," or reactivate, the latent HIV reservoirs for their subsequent "kill" by Highly Active Anti-Retroviral Therapy (Richman et al., 2009; Deeks, 2012). Along this line, Zhao et al. demonstrated that reducing 7SK snRNA pseudouridylation

destabilized the 7SK snRNP and activated Tat-dependent HIV-1 transcription. Furthermore, reduction in 7SK snRNA pseudouridylation in combination with latency reversal agents (LRAs) significantly increased the reversal of HIV latency, implicating the DKC1-box H/ACA RNP as a promising new target to eradicate latent viral reservoirs (Zhao et al., 2016). It will be interesting to determine whether the widely used chemotherapeutic agent 5-fluorouracil, which is an inhibitor of PUS enzymes, can act synergistically with current LRAs to further enhance the reversal of HIV latency.

FUNCTION OF SPLICEOSOMAL snRNA PSEUDOURIDYLATION IN PRE-mRNA SPLICING

Most genes in eukaryotes are not transcribed in mature but rather as pre-mRNA, containing coding exons as well as non-coding introns. Therefore, further pre-mRNA splicing is needed to remove intronic sequences and assemble exons into mature
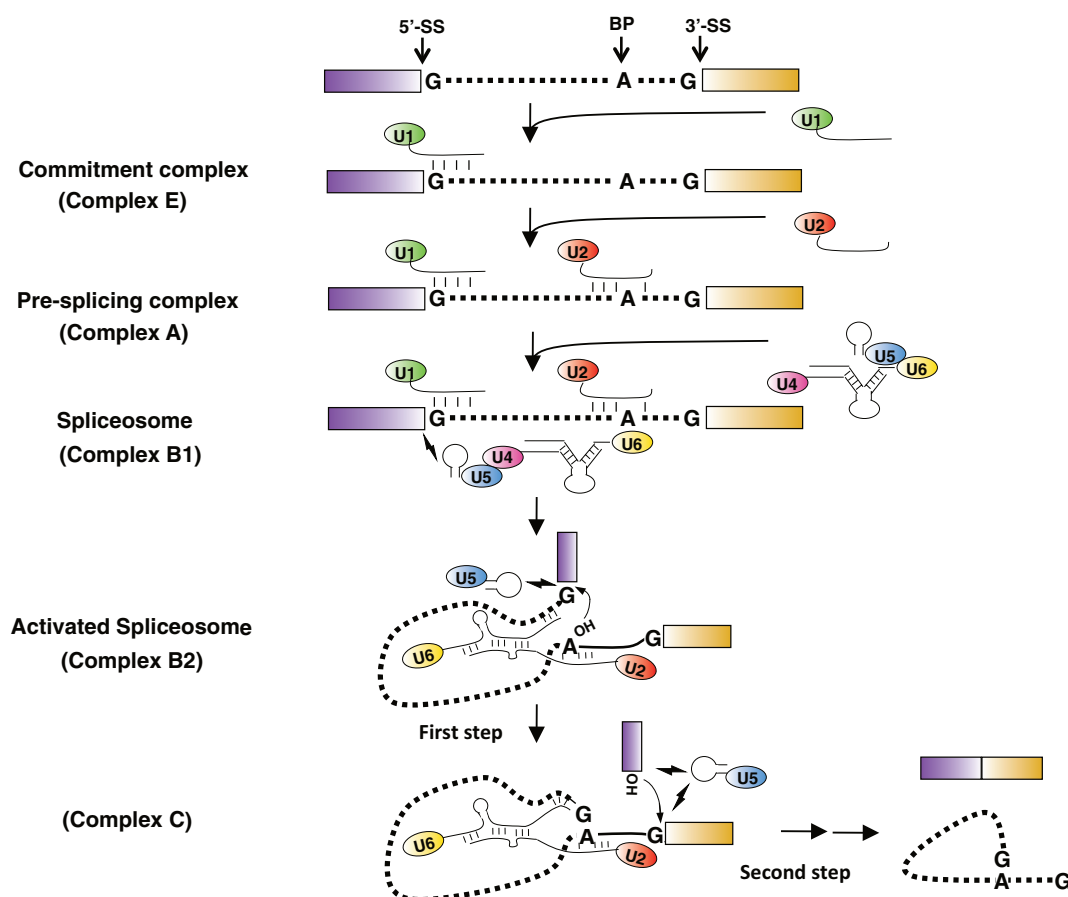
FIGURE 4 | Major spliceosome assembly and pre-mRNA splicing. The dotted lines represent the intron, and the boxes are exons. The 5'-splice site (5'-SS), the 3'-splice site (3'-SS), and the branch point adenosine (BP) are indicated in the pre-mRNA. The conserved residues at the branch site and the 5'- and 3'-SS are shown. The thin lines are snRNAs with their names in the ellipses. The short thick lines between RNA strands represent Watson–Crick base-pairing interactions. The lightning symbols depict non-Watson–Crick base-pairing interactions. The 2'-OH groups of branch point adenosine and the cutoff 5' exon are pictured in the activated spliceosome. The small arrows near 2'-OH group indicate the nucleophilic chemical reactions.

mRNA (Green, 1986; Newman, 1994). Splicing is catalyzed by the spliceosome, a massively large complex consisting of snRNAs and numerous protein components (Wahl et al., 2009; Matera and Wang, 2014). There are five snRNAs within the major spliceosome—U1, U2, U4, U5, and U6 that participate in the splicing reaction as snRNP complexes.

Pre-mRNA splicing is initiated by the recognition of the 5'-splice site (5'-SS) by the U1 snRNP *via* complementary base-pairing interactions (Kramer et al., 1984; Bindereif and Green, 1987; Ruby and Abelson, 1988; Seraphin and Rosbash, 1989). The branch-site sequence is then engaged by the U2 snRNP *via* complementary base-pairing interactions, resulting in the bulging out of the branch point adenosine of the pre-mRNA and the formation of the spliceosomal complex A (Zhuang and Weiner, 1989; Michaud and Reed, 1991; Wassarman and Steitz, 1992). Subsequently, the tri-snRNP, a complex of U4 snRNP, U6 snRNP, and U5 snRNP, is recruited, creating a fully assembled spliceosome (complex B1). Following a series of RNA–RNA rearrangements, U1 and U4 snRNPs are destabilized and released, resulting in the formation of an active spliceosome (complex B2) (Sawa and Abelson, 1992; Lesser and Guthrie,

1993). This complex catalyzes the first step of pre-mRNA splicing in which the 2'-OH group of the bulged-out branch point adenosine nucleophilically attacks the phosphate at the 5'-SS. The result of the first-step reaction is the generation of a lariat 2/3 intermediate and a cutoff 5' exon intermediate. After additional conformational changes, complex B2 is converted to complex C, and the second step of splicing is catalyzed, resulting in the production of mature mRNA and lariat intron products. The U2, U5, and U6 snRNPs are recycled for new rounds of pre-mRNA splicing (**Figure 4**) (Burge et al., 1999).

Interestingly, all of the major spliceosomal snRNAs are post-transcriptionally pseudouridylated (**Figure 5A**). U2 snRNA is the most extensively pseudouridylated snRNA, and unsurprisingly investigations into snRNA pseudouridylation have primarily focused on U2 snRNA. The functional study of U2 snRNA pseudouridylation was initiated in the early 1990s by Patton, who prevented U2 snRNA pseudouridylation in HeLa cell S100 and nuclear extracts by the incorporation of 5-fluorouridine (5-FU) (Patton, 1993a,b). Although 5-FU-substituted U2 snRNA was able to form a U2 snRNP, the snRNP was more vulnerable to dissociation by salt.

FIGURE 5 | Continued

FIGURE 5 | Pseudouridines in human spliceosomal small nuclear RNA (snRNAs). **(A)** Primary and secondary structures of human major spliceosomal snRNAs (U1, U2, U4, U5, and U6) are shown. **(B)** Primary and secondary structures of human minor spliceosomal snRNAs (U11, U12, U4atac, and U6atac) are shown. U5 snRNA is shared by both the major and minor spliceosomes. Pseudouridines (ψ) are in red. The thick lines indicate the nucleotides participating in RNA–RNA interactions or involved in catalysis during pre-mRNA splicing. The rectangles indicate the Sm-binding sites.

As experimental systems and assays developed, a more detailed description of the effects of U2 snRNA pseudouridylation on premRNA splicing emerged. For example, Yu et al. demonstrated that while *in vitro*-transcribed U2 snRNA, which lacks modification, was unable to rescue a splicing defect in U2 snRNA-depleted oocytes, following prolonged reconstitution periods, U2 snRNA was pseudouridylated and able to reconstitute splicing activity. In addition, anti-snRNP immunoprecipitation coupled with glycerol-gradient sedimentation demonstrated that U2 snRNA lacking pseudouridine was unable to form functional 17 S U2 snRNP. Thus, a good correlation between modification status, U2 snRNP biogenesis, and pre-mRNA splicing was established. In addition, by creating chimeric U2 snRNAs between cellular-derived and *in vitro*-transcribed U2, Yu et al. demonstrated that the functionally important modifications primarily resided within the first 27 nts of U2 snRNA (Yu et al., 1998).

Pseudouridylation of residues within the branch-site recognition region is also functionally important for pre-mRNA

splicing. Zhao and Yu demonstrated that pseudouridine residues within the branch-site recognition region of Xenopus U2 snRNA are required for U2 snRNP assembly and spliceosome assembly (Zhao and Yu, 2007). In addition, an NMR structure of yeast U2 snRNA:pre-mRNA branch-site helix demonstrated that ψ35 induces a dramatic structural alternation, which is required for the bulging out of the branch point adenosine and nucleophilic attack on the 5'-SS (Newby and Greenbaum, 2002). Consistent with this, a yeast knockout of PUS7, which catalyzes ψ35 of U2 snRNA, exhibited reduced fitness under conditions of high-salt media, or when in competition with a wild-type strain (Ma et al., 2003). The other pseudouridines within this region, ψ42 and ψ44, are also functionally relevant (Wu et al., 2016a). The deletion of both SNR81, responsible for ψ42, and PUS1, which catalyzes ψ44 formation, reduces the efficiency of pre-mRNA splicing, leading to growth defect. Further genetic and biochemical analyses demonstrated that U2 snRNA ψ42 and ψ44 facilitate the interaction with Prp5, a U2-dependent ATPase known to play

an important role in monitoring U2-intron branch-site interactions at early stages of spliceosome assembly (Xu and Query, 2007). Furthermore, Prp5 has reduced ATPase activity on U2 snRNA lacking ψ42 and ψ44, suggesting that these modifications regulate Prp5 enzymatic activity. Collectively, these data indicate that pseudouridylation within the branch-site recognition region plays a role not only in the biogenesis of functional snRNP but also in spliceosome assembly by influencing the enzymatic activity of Prp5.

In contrast to U2 snRNA, pseudouridylations within the other spliceosomal U snRNAs have received much less attention. However, this is not to say that they have not been investigated. For instance, the functional importance of two ψs within the 5' end of U1 snRNA has been investigated. Adopting an *in vitro*-splicing system in which two 5'-SS are in competition with each other, Roca et al. suggested that U1 snRNA pseudouridylation participates in 5'-SS discrimination (Roca et al., 2005). In addition, Freund et al. demonstrated that a ψ–G base pair, between the U1 snRNA and the substrate pre-mRNA, respectively, stabilized the U1 snRNA interaction with the 5'-SS of HIV-1 RNA (Freund et al., 2003). Lastly, as described earlier, the inducible pseudouridylation of U6 snRNA at U28 during the yeast filamentous growth program is also functionally relevant, as shown by differential growth phenotypes that are dependent on its pseudouridylation (Basak and Query, 2014).

CONCLUSION

Remarkable progress has been made toward elucidating the mechanism and function of RNA pseudouridylation in various cellular processes. However, a function for nc RNA pseudouridylation in transcriptional regulation has only just begun to emerge. Our limited understanding of the impact of pseudouridylation on transcription is partially due to the limited number of ncRNAs that were known to be pseudouridylated. However, recent efforts of transcriptome-wide mapping of RNA pseudouridylation have greatly expanded the catalog of known pseudouridylated RNAs and have identified novel modification sites within ncRNAs participating in transcriptional regulation.

For instance, metastasis-associated lung adenocarcinoma transcript 1 (MALAT1), which is involved in the epigenetic modulation of gene expression, as well as alternative splicing, is pseudouridylated at two distinct sites (Carlile et al., 2014). How these modifications contribute to MALAT1 function remains unclear and certainly worth investigating. As more ψs in ncRNA transcriptional regulators are identified, a better understanding of how these ncRNAs impact eukaryotic mRNA transcription will be achieved. Since many of these ncRNAs are associated with diseases, determining how the functional impact of their pseudouridylation may open the door to novel therapeutic strategies.

Although the mechanism and function of snRNA pseudouridylation in splicing is considerably well studied, there are still many unanswered questions. For example, in contrast to U2 snRNA, the functional significance of pseudouridylation within U1, U4, U5, and U6 is not clear. In addition, within higher eukaryotes, there exists a second spliceosome, the minor spliceosome, which consists of U5, in addition to four distinct U snRNAs, U11, U12, U4atac, and U6atac, and all of them are pseudouridylated (**Figure 5B**). Interestingly, the positions of minor spliceosomal snRNA pseudouridylation are homologous to those within major spliceosomal snRNAs, suggesting their importance in minor intron splicing (Massenet and Branlant, 1999). Detailed functional analysis of these pseudouridylations is required if we seek to understand the mechanism of minor spliceosome biogenesis and minor intron splicing. In addition, these studies may provide a better understanding of how introns are selectively recognized by the two distinct spliceosomes.

AUTHOR CONTRIBUTIONS

All authors contributed to writing the manuscript.

FUNDING

REFERENCES

Basak, A., and Query, C. C. (2014). A pseudouridine residue in the spliceosome core is part of the filamentous growth program in yeast. *Cell Rep.* 8, 966–973. doi:10.1016/j.celrep.2014.07.004

Bindereif, A., and Green, M. R. (1987). An ordered pathway of snRNP binding during mammalian pre-mRNA splicing complex assembly. *EMBO J.* 6, 2415–2424.

Bjorklund, S., and Kim, Y. J. (1996). Mediator of transcriptional regulation. *Trends Biochem. Sci.* 21, 335–337. doi:10.1016/S0968-0004(96)10051-7

Bousquet-Antonelli, C., Henry, Y., G'Elugne, J. P., Caizergues-Ferrer, M., and Kiss, T. (1997). A small nucleolar RNP protein is required for pseudouridylation of eukaryotic ribosomal RNAs. *EMBO J.* 16, 4770–4776. doi:10.1093/emboj/16.15.4770

Branlant, C., Krol, A., Machatt, M. A., Pouyet, J., Ebel, J. P., Edwards, K., et al. (1981). Primary and secondary structures of *Escherichia coli* MRE 600 23S ribosomal RNA. Comparison with models of secondary structure for maize chloroplast 23S rRNA and for large portions of mouse and human 16S mitochondrial rRNAs. *Nucleic Acids Res.* 9, 4303–4324. doi:10.1093/nar/9.17.4303

Burge, C. B., Tuschl, T., and Sharp, P. A. (1999). "Splicing of precursors to mRNAs by the spliceosomes," in *The RNA World*, 2nd Edn (Cold Spring Harbor Laboratory Press), 525–560.

Bykhovskaya, Y., Casas, K., Mengesha, E., Inbal, A., and Fischel-Ghodsian, N. (2004). Missense mutation in pseudouridine synthase 1 (PUS1) causes mitochondrial myopathy and sideroblastic anemia (MLASA). *Am. J. Hum. Genet.* 74, 1303–1308. doi:10.1086/421530

Carlile, T. M., Rojas-Duran, M. F., Zinshteyn, B., Shin, H., Bartoli, K. M., and Gilbert, W. V. (2014). Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature* 515, 143–146. doi:10.1038/nature13802

Charette, M., and Gray, M. W. (2000). Pseudouridine in RNA: what, where, how, and why. *IUBMB Life* 49, 341–351. doi:10.1080/152165400410182

Cohn, W. E. V., and Elliot, V. (1951). Nucleoside-5'-phosphates from ribonucleic acid. *Nature* 167, 483–484. doi:10.1038/167483a0

Deeks, S. G. (2012). HIV: shock and kill. *Nature* 487, 439–440. doi:10.1038/487439a

Del Campo, M., Ofengand, J., and Malhotra, A. (2004). Crystal structure of the catalytic domain of RluD, the only rRNA pseudouridine synthase required for normal growth of *Escherichia coli*. *RNA* 10, 231–239. doi:10.1261/rna.5187404

Ericsson, U. B., Nordlund, P., and Hallberg, B. M. (2004). X-ray structure of tRNA pseudouridine synthase TruD reveals an inserted domain with a novel fold. *FEBS Lett.* 565, 59–64. doi:10.1016/j.febslet.2004.03.085

Fernandez-Vizarra, E., Berardinelli, A., Valente, L., Tiranti, V., and Zeviani, M. (2007). Nonsense mutation in pseudouridylate synthase 1 (PUS1) in two brothers affected by myopathy, lactic acidosis and sideroblastic anaemia (MLASA). *J. Med. Genet.* 44, 173–180. doi:10.1136/jmg.2006.045252

Foster, P. G., Huang, L., Santi, D. V., and Stroud, R. M. (2000). The structural basis for tRNA recognition and pseudouridine formation by pseudouridine synthase I. *Nat. Struct. Biol.* 7, 23–27. doi:10.1038/71219

Freund, M., Asang, C., Kammler, S., Konermann, C., Krummheuer, J., Hipp, M., et al. (2003). A novel approach to describe a U1 snRNA binding site. *Nucleic Acids Res.* 31, 6963–6975. doi:10.1093/nar/gkg901

Ghosh, S. K., Patton, J. R., and Spanjaard, R. A. (2012). A small RNA derived from RNA coactivator SRA blocks steroid receptor signaling *via* inhibition of PUS1p-mediated pseudouridylation of SRA: evidence of a novel RNA binding domain in the N-terminus of steroid receptors. *Biochemistry* 51, 8163–8172. doi:10.1021/bi300602r

Girard, J. P., Lehtonen, H., Caizergues-Ferrer, M., Amalric, F., Tollervey, D., and Lapeyre, B. (1992). GAR1 is an essential small nucleolar RNP protein required for pre-rRNA processing in yeast. *EMBO J.* 11, 673–682.

Green, M. R. (1986). Pre-mRNA splicing. *Annu. Rev. Genet.* 20, 671–708. doi:10.1146/annurev.ge.20.120186.003323

Grosjean, H., Sprinzl, M., and Steinberg, S. (1995). Posttranscriptionally modified nucleosides in transfer RNA: their locations and frequencies. *Biochimie* 77, 139–141. doi:10.1016/0300-9084(96)88117-X

Guenther, M. G., Levine, S. S., Boyer, L. A., Jaenisch, R., and Young, R. A. (2007). A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* 130, 77–88. doi:10.1016/j.cell.2007.05.042

Hamma, T., and Ferre-D'Amare, A. R. (2006). Pseudouridine synthases. *Chem. Biol.* 13, 1125–1135. doi:10.1016/j.chembiol.2006.09.009

Hamma, T., Reichow, S. L., Varani, G., and Ferre-D'Amare, A. R. (2005). The Cbf5-Nop10 complex is a molecular bracket that organizes box H/ACA RNPs. *Nat. Struct. Mol. Biol.* 12, 1101–1107. doi:10.1038/nsmb1036

Hoang, C., Chen, J., Vizthum, C. A., Kandel, J. M., Hamilton, C. S., Mueller, E. G., et al. (2006). Crystal structure of pseudouridine synthase RluA: indirect sequence readout through protein-induced RNA structure. *Mol. Cell* 24, 535–545. doi:10.1016/j.molcel.2006.09.017

Hoang, C., and Ferre-D'Amare, A. R. (2001). Cocrystal structure of a tRNA Psi55 pseudouridine synthase: nucleotide flipping by an RNA-modifying enzyme. *Cell* 107, 929–939. doi:10.1016/S0092-8674(01)00618-3

Hoang, C., and Ferre-D'Amare, A. R. (2004). Crystal structure of the highly divergent pseudouridine synthase TruD reveals a circular permutation of a conserved fold. *RNA* 10, 1026–1033. doi:10.1261/rna.7240504

Hopper, A. K., and Phizicky, E. M. (2003). tRNA transfers to the limelight. *Genes Dev.* 17, 162–180. doi:10.1101/gad.1049103

Kadonaga, J. T. (2004). Regulation of RNA polymerase II transcription by sequence-specific DNA binding factors. *Cell* 116, 247–257. doi:10.1016/S0092-8674(03)01078-X

Karijolich, J., Kantartzis, A., and Yu, Y. T. (2010). RNA modifications: a mechanism that modulates gene expression. *Methods Mol. Biol.* 629, 1–19. doi:10.1007/978-1-60761-657-3_1

Karijolich, J., and Yu, Y. T. (2010). Spliceosomal snRNA modifications and their function. *RNA Biol.* 7, 192–204. doi:10.4161/rna.7.2.11207

Kaya, Y., Del Campo, M., Ofengand, J., and Malhotra, A. (2004). Crystal structure of TruD, a novel pseudouridine synthase with a new protein fold. *J. Biol. Chem.* 279, 18107–18110. doi:10.1074/jbc.C400072200

Kishore, S., Gruber, A. R., Jedlinski, D. J., Syed, A. P., Jorjani, H., and Zavolan, M. (2013). Insights into snoRNA biogenesis and processing from PAR-CLIP of snoRNA core proteins and small RNA sequencing. *Genome Biol.* 14, R45. doi:10.1186/gb-2013-14-5-r45

Kiss, T., Fayet-Lebaron, E., and Jady, B. E. (2010). Box H/ACA small ribonucleoproteins. *Mol. Cell* 37, 597–606. doi:10.1016/j.molcel.2010.01.032

Kramer, A., Keller, W., Appel, B., and Luhrmann, R. (1984). The 5' terminus of the RNA moiety of U1 small nuclear ribonucleoprotein particles is required for the splicing of messenger RNA precursors. *Cell* 38, 299–307. doi:10.1016/0092-8674(84)90551-8

Kuras, L., and Struhl, K. (1999). Binding of TBP to promoters *in vivo* is stimulated by activators and requires Pol II holoenzyme. *Nature* 399, 609–613. doi:10.1038/21239

Lanz, R. B., McKenna, N. J., Onate, S. A., Albrecht, U., Wong, J., Tsai, S. Y., et al. (1999). A steroid receptor coactivator, SRA, functions as an RNA and is present in an SRC-1 complex. *Cell* 97, 17–27. doi:10.1016/S0092-8674(00)80711-4

Lanz, R. B., Razani, B., Goldberg, A. D., and O'Malley, B. W. (2002). Distinct RNA motifs are important for coactivation of steroid hormone receptors by steroid receptor RNA activator (SRA). *Proc. Natl. Acad. Sci. U.S.A.* 99, 16081–16086. doi:10.1073/pnas.192571399

Lesser, C. F., and Guthrie, C. (1993). Mutations in U6 snRNA that alter splice site specificity: implications for the active site. *Science* 262, 1982–1988. doi:10.1126/science.8266093

Levine, M. (2011). Paused RNA polymerase II as a developmental checkpoint. *Cell* 145, 502–511. doi:10.1016/j.cell.2011.04.021

Li, L., and Ye, K. (2006). Crystal structure of an H/ACA box ribonucleoprotein particle. *Nature* 443, 302–307. doi:10.1038/nature05151

Li, S., Duan, J., Li, D., Yang, B., Dong, M., and Ye, K. (2011). Reconstitution and structural analysis of the yeast box H/ACA RNA-guided pseudouridine synthase. *Genes Dev.* 25, 2409–2421. doi:10.1101/gad.175299.111

Li, X., Zhu, P., Ma, S., Song, J., Bai, J., Sun, F., et al. (2015). Chemical pulldown reveals dynamic pseudouridylation of the mammalian transcriptome. *Nat. Chem. Biol.* 11, 592–597. doi:10.1038/nchembio.1836

Lovejoy, A. F., Riordan, D. P., and Brown, P. O. (2014). Transcriptome-wide mapping of pseudouridines: pseudouridine synthases modify specific mRNAs in *S. cerevisiae*. *PLoS ONE* 9:e110799. doi:10.1371/journal.pone.0110799

Ma, X., Zhao, X., and Yu, Y. T. (2003). Pseudouridylation (Psi) of U2 snRNA in *S. cerevisiae* is catalyzed by an RNA-independent mechanism. *EMBO J.* 22, 1889–1897. doi:10.1093/emboj/cdg191

Machnicka, M. A., Milanowska, K., Osman Oglou, O., Purta, E., Kurkowska, M., Olchowik, A., et al. (2013). MODOMICS: a database of RNA modification pathways—2013 update. *Nucleic Acids Res.* 41, D262–D267. doi:10.1093/nar/gks1007

Maden, B. E. (1990). The numerous modified nucleotides in eukaryotic ribosomal RNA. *Prog. Nucleic Acid Res. Mol. Biol.* 39, 241–303. doi:10.1016/S0079-6603(08)60629-7

Manival, X., Charron, C., Fourmann, J. B., Godard, F., Charpentier, B., and Branlant, C. (2006). Crystal structure determination and site-directed mutagenesis of the *Pyrococcus abyssi* aCBF5-aNOP10 complex reveal crucial roles of the C-terminal domains of both proteins in H/ACA sRNP activity. *Nucleic Acids Res.* 34, 826–839. doi:10.1093/nar/gkj482

Massenet, S., and Branlant, C. (1999). A limited number of pseudouridine residues in the human atac spliceosomal UsnRNAs as compared to human major spliceosomal UsnRNAs. *RNA* 5, 1495–1503. doi:10.1017/S1355838299991537

Massenet, S., Mougin, A., and Branlant, C. (1998). "Posttranscriptional modifications in the U small nuclear RNAs," in *Modification and Editing of RNA*, ed. H. Grosjean (Washington, DC: ASM Press), 201–228.

Matera, A. G., and Wang, Z. (2014). A day in the life of the spliceosome. *Nat. Rev. Mol. Cell Biol.* 15, 108–121. doi:10.1038/nrm3742

McCleverty, C. J., Hornsby, M., Spraggon, G., and Kreusch, A. (2007). Crystal structure of human PUS10, a novel pseudouridine synthase. *J. Mol. Biol.* 373, 1243–1254. doi:10.1016/j.jmb.2007.08.053

Mercer, T. R., Dinger, M. E., and Mattick, J. S. (2009). Long non-coding RNAs: insights into functions. *Nat. Rev. Genet.* 10, 155–159. doi:10.1038/nrg2521

Michaud, S., and Reed, R. (1991). An ATP-independent complex commits pre-mRNA to the mammalian spliceosome assembly pathway. *Genes Dev.* 5, 2534–2546. doi:10.1101/gad.5.12b.2534

Mizutani, K., Machida, Y., Unzai, S., Park, S. Y., and Tame, J. R. (2004). Crystal structures of the catalytic domains of pseudouridine synthases RluC and RluD from *Escherichia coli*. *Biochemistry* 43, 4454–4463. doi:10.1021/bi036079c

Muse, G. W., Gilchrist, D. A., Nechaev, S., Shah, R., Parker, J. S., Grissom, S. F., et al. (2007). RNA polymerase is poised for activation across the genome. *Nat. Genet.* 39, 1507–1511. doi:10.1038/ng.2007.21

Narlikar, G. J., Fan, H. Y., and Kingston, R. E. (2002). Cooperation between complexes that regulate chromatin structure and transcription. *Cell* 108, 475–487. doi:10.1016/S0092-8674(02)00654-2

Newby, M. I., and Greenbaum, N. L. (2002). Sculpting of the spliceosomal branch site recognition motif by a conserved pseudouridine. *Nat. Struct. Biol.* 9, 958–965. doi:10.1038/nsb873

Newman, A. J. (1994). Pre-mRNA splicing. *Curr. Opin. Genet. Dev.* 4, 298–304. doi:10.1016/S0959-437X(05)80057-7

Patton, J. R. (1993a). Multiple pseudouridine synthase activities for small nuclear RNAs. *Biochem. J.* 290(Pt 2), 595–600. doi:10.1042/bj2900595

Patton, J. R. (1993b). Ribonucleoprotein particle assembly and modification of U2 small nuclear RNA containing 5-fluorouridine. *Biochemistry* 32, 8939–8944. doi:10.1021/bi00085a027

Ptashne, M. (2005). Regulation of transcription: from lambda to eukaryotes. *Trends Biochem. Sci.* 30, 275–279. doi:10.1016/j.tibs.2005.04.003

Rashid, R., Liang, B., Baker, D. L., Youssef, O. A., He, Y., Phipps, K., et al. (2006). Crystal structure of a Cbf5-Nop10-Gar1 complex and implications in RNA-guided pseudouridylation and dyskeratosis congenita. *Mol. Cell* 21, 249–260. doi:10.1016/j.molcel.2005.11.017

Reddy, R., and Busch, H. (1988). "Small nuclear RNAs: RNA sequences, structure, and modifications," in *Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles*, ed. M. Birnsteil (Heidelberg: Springer-Verlag Press), 1–37.

Richman, D. D., Margolis, D. M., Delaney, M., Greene, W. C., Hazuda, D., and Pomerantz, R. J. (2009). The challenge of finding a cure for HIV infection. *Science* 323, 1304–1307. doi:10.1126/science.1165706

Roca, X., Sachidanandam, R., and Krainer, A. R. (2005). Determinants of the inherent strength of human 5' splice sites. *RNA* 11, 683–698. doi:10.1261/rna.2040605

Ruby, S. W., and Abelson, J. (1988). An early hierarchic role of U1 small nuclear ribonucleoprotein in spliceosome assembly. *Science* 242, 1028–1035. doi:10.1126/science.2973660

Sawa, H., and Abelson, J. (1992). Evidence for a base-pairing interaction between U6 small nuclear RNA and 5' splice site during the splicing reaction in yeast. *Proc. Natl. Acad. Sci. U.S.A.* 89, 11269–11273. doi:10.1073/pnas.89.23.11269

Schwartz, S., Bernstein, D. A., Mumbach, M. R., Jovanovic, M., Herbst, R. H., Leon-Ricardo, B. X., et al. (2014). Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. *Cell* 159, 148–162. doi:10.1016/j.cell.2014.08.028

Seraphin, B., and Rosbash, M. (1989). Identification of functional U1 snRNA-pre-mRNA complexes committed to spliceosome assembly and splicing. *Cell* 59, 349–358. doi:10.1016/0092-8674(89)90296-1

Sivaraman, J., Iannuzzi, P., Cygler, M., and Matte, A. (2004). Crystal structure of the RluD pseudouridine synthase catalytic module, an enzyme that modifies 23S rRNA and is essential for normal cell growth of *Escherichia coli*. *J. Mol. Biol.* 335, 87–101. doi:10.1016/j.jmb.2003.10.003

Sivaraman, J., Sauve, V., Larocque, R., Stura, E. A., Schrag, J. D., Cygler, M., et al. (2002). Structure of the 16S rRNA pseudouridine synthase RsuA bound to uracil and UMP. *Nat. Struct. Biol.* 9, 353–358. doi:10.1038/nsb788

Smale, S. T., and Kadonaga, J. T. (2003). The RNA polymerase II core promoter. *Annu. Rev. Biochem.* 72, 449–479. doi:10.1146/annurev.biochem.72.121801.161520

Wahl, M. C., Will, C. L., and Luhrmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. *Cell* 136, 701–718. doi:10.1016/j.cell.2009.02.009

Wassarman, D. A., and Steitz, J. A. (1991). Structural analyses of the 7SK ribonucleoprotein (RNP), the most abundant human small RNP of unknown function. *Mol. Cell. Biol.* 11, 3432–3445. doi:10.1128/MCB.11.7.3432

Wassarman, D. A., and Steitz, J. A. (1992). Interactions of small nuclear RNA's with precursor messenger RNA during *in vitro* splicing. *Science* 257, 1918–1925. doi:10.1126/science.1411506

Wu, G., Adachi, H., Ge, J., Stephenson, D., Query, C. C., and Yu, Y. T. (2016a). Pseudouridines in U2 snRNA stimulate the ATPase activity of Prp5 during spliceosome assembly. *EMBO J.* 35, 654–667. doi:10.15252/embj.201593113

Wu, G., Radwan, M. K., Xiao, M., Adachi, H., Fan, J., and Yu, Y. T. (2016b). The TOR signaling pathway regulates starvation-induced pseudouridylation of yeast U2 snRNA. *RNA* 22, 1146–1152. doi:10.1261/rna.056796.116

Wu, G., Xiao, M., Yang, C., and Yu, Y. T. (2011). U2 snRNA is inducibly pseudouridylated at novel sites by PUS7p and snR81 RNP. *EMBO J.* 30, 79–89. doi:10.1038/emboj.2010.316

Xu, Y. Z., and Query, C. C. (2007). Competition between the ATPase Prp5 and branch region-U2 snRNA pairing modulates the fidelity of spliceosome assembly. *Mol. Cell* 28, 838–849. doi:10.1016/j.molcel.2007.09.022

Yang, Z., Zhu, Q., Luo, K., and Zhou, Q. (2001). The 7SK small nuclear RNA inhibits the CDK9/cyclin T1 kinase to control transcription. *Nature* 414, 317–322. doi:10.1038/35104575

Yik, J. H., Chen, R., Nishimura, R., Jennings, J. L., Link, A. J., and Zhou, Q. (2003). Inhibition of P-TEFb (CDK9/Cyclin T) kinase and RNA polymerase II transcription by the coordinated actions of HEXIM1 and 7SK snRNA. *Mol. Cell* 12, 971–982. doi:10.1016/S1097-2765(03)00388-5

Yu, Y. T., and Meier, U. T. (2014). RNA-guided isomerization of uridine to pseudouridine—pseudouridylation. *RNA Biol.* 11, 1483–1494. doi:10.4161/15476286.2014.972855

Yu, Y. T., Shu, M. D., and Steitz, J. A. (1998). Modifications of U2 snRNA are required for snRNP assembly and pre-mRNA splicing. *EMBO J.* 17, 5783–5795. doi:10.1093/emboj/17.19.5783

Zhao, X., Patton, J. R., Davis, S. L., Florence, B., Ames, S. J., and Spanjaard, R. A. (2004). Regulation of nuclear receptor activity by a pseudouridine synthase through posttranscriptional modification of steroid receptor RNA activator. *Mol. Cell* 15, 549–558. doi:10.1016/j.molcel.2004.06.044

Zhao, X., Patton, J. R., Ghosh, S. K., Fischel-Ghodsian, N., Shen, L., and Spanjaard, R. A. (2007). PUS3p- and PUS1p-dependent pseudouridylation of steroid receptor RNA activator controls a functional switch that regulates nuclear receptor signaling. *Mol. Endocrinol.* 21, 686–699. doi:10.1210/me.2006-0414

Zhao, X., and Yu, Y. T. (2007). Incorporation of 5-fluorouracil into U2 snRNA blocks pseudouridylation and pre-mRNA splicing *in vivo*. *Nucleic Acids Res.* 35, 550–558. doi:10.1093/nar/gkl1084

Zhao, Y., Karijolich, J., Glaunsinger, B., and Zhou, Q. (2016). Pseudouridylation of 7SK snRNA promotes 7SK snRNP formation to suppress HIV-1 transcription and escape from latency. *EMBO Rep.* 17, 1441–1451. doi:10.15252/embr.201642682

Zhou, Q., Li, T., and Price, D. H. (2012). RNA polymerase II elongation control. *Annu. Rev. Biochem.* 81, 119–143. doi:10.1146/annurev-biochem-052610-095910

Zhuang, Y., and Weiner, A. M. (1989). A compensatory base change in human U2 snRNA can suppress a branch site mutation. *Genes Dev.* 3, 1545–1552. doi:10.1101/gad.3.10.1545

Check for updates

3′ RNA Uridylation in Epitranscriptomics, Gene Regulation, and Disease

*Miriam R. Menezes†, Julien Balzeau† and John P. Hagan**

Department of Neurosurgery, University of Texas Health Science Center at Houston, Houston, TX, United States

Emerging evidence implicates a wide range of post-transcriptional RNA modifications that play crucial roles in fundamental biological processes including regulating gene expression. Collectively, they are known as epitranscriptomics. Recent studies implicate 3′ RNA uridylation, the non-templated addition of uridine(s) to the terminal end of RNA, as a key player in epitranscriptomics. In this review, we describe the functional roles and significance of 3′ terminal RNA uridylation that has diverse functions in regulating both mRNAs and non-coding RNAs. In mammals, three Terminal Uridylyl Transferases (TUTases) are primarily responsible for 3′ RNA uridylation. These enzymes are also referred to as polyU polymerases. TUTase 1 (TUT1) is implicated in U6 snRNA maturation via uridylation. The TUTases TUT4 and/or TUT7 are the predominant mediators of all other cellular uridylation. Terminal uridylation promotes turnover for many polyadenylated mRNAs, replication-dependent histone mRNAs that lack polyA-tails, and aberrant structured noncoding RNAs. In addition, uridylation regulates biogenesis of a subset of microRNAs and generates isomiRs, sequent variant microRNAs that have altered function in specific cases. For example, the RNA binding protein and proto-oncogene LIN28A and TUT4 work together to polyuridylate pre-let-7, thereby blocking biogenesis and function of the tumor suppressor let-7 microRNA family. In contrast, monouridylation of Group II pre-miRNAs creates an optimal 3′ overhang that promotes recognition and subsequent cleavage by the Dicer-TRBP complex that then yields the mature microRNA. Also, uridylation may play a role in non-canonical microRNA biogenesis. The overall significance of 3′ RNA uridylation is discussed with an emphasis on mammalian development, gene regulation, and disease, including cancer and Perlman syndrome. We also introduce recent changes to the HUGO-approved gene names for multiple terminal nucleotidyl transferases that affects in part TUTase nomenclature (TUT1/TENT1, TENT2/PAPD4/GLD2, TUT4/ZCCHC11/TENT3A, TUT7/ZCCHC6/TENT3B, TENT4A/PAPD7, TENT4B/PAPD5, TENT5A/FAM46A, TENT5B/FAM46B, TENT5C/FAM46C, TENT5D/FAM46D, MTPAP/TENT6/PAPD1).

Keywords: RNA epitranscritpomics, 3′ terminal RNA uridylation, TUTase, LIN28/let-7 pathway, DIS3L2, cancer, perlman syndrome, Wilms tumor

INTRODUCTION

Epitranscriptomics refer to a diverse set of RNA chemical modifications and post-transcriptional nucleotide additions that play central roles in pre-mRNA splicing, translation, and regulation of gene expression. For example, ribosomal and transfer RNAs undergo extensive chemical modifications that are required for function through stabilization of their RNA secondary structure, while non-templated nucleotide additions are critical for polyadenylated mRNAs and aminoacylation of tRNAs. Recent evidence implicates 3′ RNA uridylation, the addition of non-templated uridine(s) to the RNA end, in several biological processes. This review presents our current understanding of the biochemical and functional roles for 3′ terminal RNA uridylation with a focus on mammalian biology. Over the past decade, 3′ RNA uridylation has emerged to be functionally significant for multiple RNA types such as mRNAs, microRNAs, and structured non-coding RNAs. Terminal Uridylyl Transferase (TUTases) that are also known as polyU polymerases are the enzymes responsible for 3′ RNA uridylation.

NON-CANONICAL TERMINAL RIBONUCLEOTIDYL TRANSFERASES

TUTases fall within a class of seven non-canonical terminal ribouncleotidyl transferases that contain a DNA polymerase β-like nucleotidyltransferase domain (**Figure 1**). Unfortunately, each member of the non-canonical polymerase family that includes TUTases has multiple names, creating considerable confusion in relation to their ribonucleotidyl specificity (**Table 1**). Here, we will refer to each enzyme using their HUGO-approved nomenclature that was recently updated by the Human Gene Nomenclature Committee to address several concerns. Two genes TUT4 (aka ZCCHC11) and TUT7 (aka ZCCHC6) encode terminal uridylyl transferases that are primarily cytoplasmic and are quite similar structurally, having arisen from a gene duplication event. These enzymes interact transiently with RNA and are slow polymerases (e.g., the TUT4 uridylation rate is ∼0.2 nucleotides per second; Yeom et al., 2011). As such, they typically add one or very few uridines to their RNA substrates, except in cases where TUTase interaction with RNA is stabilized by other factors such as the RNA binding protein LIN28A. TUT1/STAR-PAP has dual specificity for UTP and ATP with respective k_{cat} values of 0.059 and 0.002 s^{-1} (Yamashita et al., 2017). The remaining non-canonical transferases primarily add non-templated adenosines. Phylogenetic and biochemical analyses reveal that the insertion of a histidine at the active site alters the ribonucleotidyl specificity of the polymerase from ATP to UTP (Munoz-Tello et al., 2012; Yates et al., 2015; Chung et al., 2016; Yamashita et al., 2017). Consistent with this idea, human TENT2 (aka GLD2/TUT2) lacks this critical histidine and prefers ATP over UTP by >80-fold (Chung et al., 2016). Furthermore, human TENT2 mutated by histidine insertion within the active site converts the mutant protein to a TUTase. In this review, we describe in detail the functional roles of three

mammalian TUTases, TUT1, TUT4, and TUT7. Human TUT4 and TUT7 have numerous roles in regulating both mRNAs and non-coding RNAs. TUT1 has dual functions where it is critical for uridylation and maturation of the spliceosomal U6 snRNA in the nucleus and in the adenylation of select mRNAs, including HO-1 (Heme Oxygenase 1), BIK (BCL2-interacting killer), PTEN, WIF1, and CDH1 (Gonzales et al., 2008; Mellman et al., 2008; Li et al., 2012, 2017; Kandala et al., 2016; Sudheesh and Laishram, 2017). Lastly, we will discuss the roles of TUTases in mammalian development, physiology, and diseases including cancer and Perlman Syndrome.

SPLICEOSOMAL U6 SNRNA AND URIDYLATION

Transcription of eukaryotic protein coding genes, especially in multicellular organisms, generates a pre-mRNA that typically undergoes splicing to remove introns and join successive exons. The major spliceosome, a large ribonucleoprotein (RNP) complex, is required for splicing and contains five essential snRNPs (U1, U2, U4, U5, and U6). U6 small nuclear RNA (U6 snRNA) has a unique maturation mechanism (reviewed in Mroczek et al., 2012). As illustrated in **Figure 2**, U6 snRNA is the only known RNA substrate where uridylation occurs within the nucleus and this uridylation is essential for its splicing function (Trippe et al., 1998, 2003, 2006). U6 snRNA is transcribed by RNA polymerase III (Kunkel et al., 1986) where transcription termination occurs within a short (∼5–6 nucleotide) oligo(dT) DNA stretch. The original U6 snRNA transcript contains four uridines at the 3′ end. After transcription and initial 3′ end formation, the chaperon-like La protein can bind this polyU tail, favoring stabilization of the U6 snRNA (Wolin and Cedervall, 2002). To continue the maturation process, La protein is removed and TUT1 adds up to 20 uridines to the polyU tail. This longer polyU tail is a signal for USB1, a 3′→ 5′ exoribonuclease, to remove uridines, leaving only five of them. USB1 then catalyzes the formation of a 2′,3′-cyclic phosphate at the 3′ end. This maturation process allows the LSm2-8 complex to bind the 3′ extremity, facilitates the proper assembly of mature snRNPs, and is important for nuclear retention (Licht et al., 2008).

MICRORNAS AND URIDYLATION

The following sections describe the biochemical functions that uridylation has in relation to microRNAs, a recently identified class of negative regulators of gene expression. MicroRNA expression is most frequently regulated transcriptionally; however, both microRNA biogenesis as well as generation of isomiRs, sequence variant microRNAs, can be regulated by TUTases. Briefly, uridylation plays both positive and negative roles in regulating canonical microRNA biogenesis for a small, albeit important set of microRNAs, most notably the tumor suppressor let-7 microRNA family. Uridylation is also implicated in two distinct non-canonical microRNA biogenesis pathways that differ on their reliance on Drosha and Dicer. Lastly,
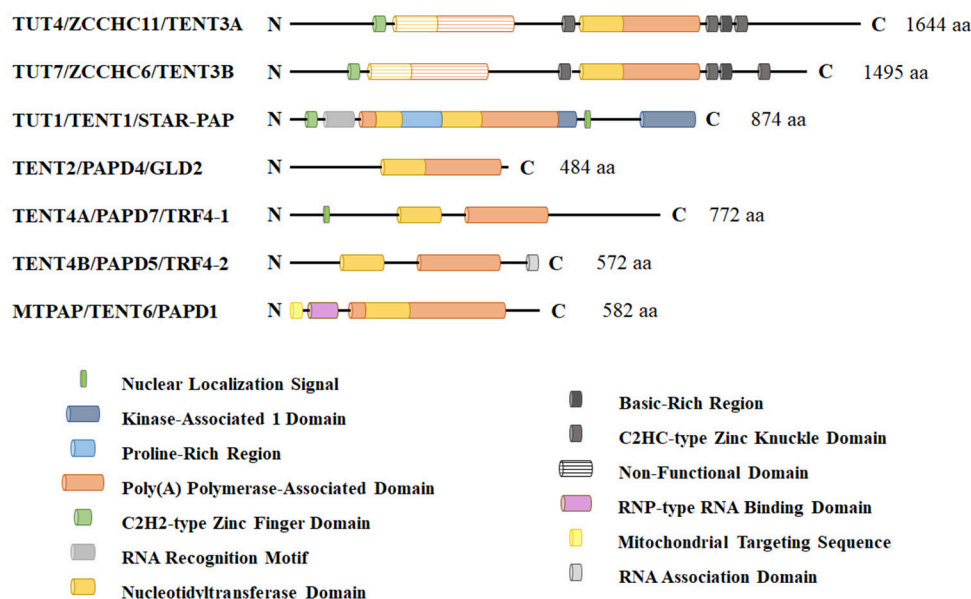
FIGURE 1 | Schematic representation of non-canonical terminal ribonucleotdyl transferases in humans. These enzymes contain a DNA polymerase β-like nucleotidyltransferase domain and a more C-terminal polyA polymerase-like domain that is required for enzymatic activity. TUT1, TUT4, and TUT7 are the primary polymerases responsible for terminal 3′ RNA uridylation in mammalian cells. Other family members lack a critical histidine that confers UTP specificity and are thought to be primarily adenyltransferases. TUT1 has a nuclear localization signal and functions in part in nuclear U6 snRNA maturation and recycling. TUT4 and TUT7 are predominantly cytoplasmic proteins where they function in noncoding RNA quality control, mRNA turnover (both polyadenylated mRNAs and histone mRNAs), and regulation of let-7 microRNA biogenesis with LIN28A.

uridylation of mature microRNAs can generate isomiRs with altered activity.

Canonical MicroRNA Biogenesis and Function

MicroRNAs are a class of small noncoding RNAs (∼19–23 nucleotides) that primarily function as negative regulators of their mRNA targets (reviewed in Hagan and Croce, 2007; Bartel, 2009; Lin and Gregory, 2015b; Balzeau et al., 2017). Mature microRNAs are generated through two step-wise cleavage reactions from the primary microRNA (pri-miRNA) transcripts. Pri-miRNAs are produced by RNA polymerase II (Cai et al., 2004; Lee et al., 2004) and to a much lesser extent by RNA polymerase III (Borchert et al., 2006). The microRNA host gene can be either a protein coding gene or a long non-coding RNA (lncRNA). For protein coding host genes, the microRNA almost exclusively exists in the intron as expected, since microRNA processing would disrupt the mRNA. The few exceptions to this rule are several microRNAs that lie either in alternatively spliced exons or between alternative polyA sites. For host genes that are lncRNAs, microRNA lie in both introns and exons at significant frequencies.

Canonical microRNA biogenesis occurs by two sequential endonucleolytic reactions to produce a mature, functional microRNA. The pri-miRNA is first cleaved by the nuclear RNase III enzyme Drosha in the Microprocessor complex (Lee et al., 2003; Denli et al., 2004; Gregory et al., 2004; Han et al., 2004). The resultant cleavage product is termed precursor microRNA

(pre-miRNA) and has a characteristic hairpin structure of ∼50–75 nucleotides. The pre-miRNA is transported to the cytoplasm by the Exportin-5/Ran GTP complex (Yi et al., 2003; Bohnsack et al., 2004). In the cytoplasm, the pre-miRNA is processed by a second RNase III enzyme Dicer to release a small RNA duplex that is subsequently loaded onto Argonaute (Bernstein et al., 2001; Hutvagner et al., 2001; Ketting et al., 2001; Chendrimada et al., 2005; Haase et al., 2005). Typically, one strand is preferentially incorporated into the miRISC (microRNA Induced Silencing Complex) where it base pairs with imperfect complementarity to the 3′ UTR of its target mRNAs, resulting in mRNA destabilization and/or translational inhibition. As such, microRNAs may potentially regulate 30% of all mammalian genes (Lewis et al., 2005; Lim et al., 2005).

Monouridylation Promotes Dicer Processing of Group II Precursor MicroRNAs

Addition of non-templated uridine(s) to the 3′ end of microRNAs is an important post-transcriptional modification that has been shown to impact activity and biogenesis of miRNAs. Specific pre-miRNAs are substrates for monouridylation and/or oligouridylation that can impact microRNA biogenesis positively or negatively, respectively. As the name indicates, monouridylation is the addition of a single non-templated uridine, while oligouridylation and polyuridylation refer to the addition of polyU-tails that can be as long as 30 nucleotides.

For the vast majority of microRNAs, nuclear Drosha cleavage of the pri-miRNA results in a pre-miRNA with a 2 nucleotide

TABLE 1 | Nomenclature of Non-canonical terminal ribonucleotidyl transferases.

New HUGO-approved symbol*	Approved name	Old HUGO-approved symbol*	Other alternate symbols	Subcellular localization	Disease	Ribonucleotide specificity
TUT4	Terminal Uridylyl Transferase 4	ZCCHC11	TENT3A, PAPD3, KIAA0191	Cytoplasm (Heo et al., 2009; Schmidt et al., 2011) Nucleus>Cytoplasm (Minoda et al., 2006) Cytoplasm/Nucleolus (Barbe et al., 2008)	Numerous Cancer Types including breast, ovarian, and colon (Hallett and Hassell, 2011; Piskounova et al., 2011; Balzeau et al., 2017) Myotonic dystrophy (Rau et al., 2011)	U
TUT7	Terminal Uridylyl Transferase 7	ZCCHC6	TENT3B, PAPD6, KIAA1711		In Critical region for Cutaneous Malignant Melanoma Predisposition Locus (Cannon-Albright et al., 2013) Closest ORF to rs952834 Autism susceptibility loci (Weiss et al., 2009)	U
TUT1	Terminal Uridylyl Transferase 1, U6 snRNA-specific	TUT1	PAPD2, STAR-PAP, RBM21	Nucleolus (Trippe et al., 2006) Colocalizes with mRNA cleavage and polyadenylation specificity factor complex (Gonzales et al., 2008) Nuclear Speckles (Mellman et al., 2008)		U and A
TENT2	Terminal Nucleotidyltransferase 2	PAPD4	GLD2, TUT2	Cytoplasm/Nucleus (Nakanishi et al., 2006)		A or A, U, G
TENT4A	Terminal Nucleotidyltransferase 4A	PAPD7	POLS, TRF4-1, POLK, LAK-1	Nucleoplasm > Cytoplasm (Ogami et al., 2013)		A
TENT4B	Terminal Nucleotidyltransferase 4B	PAPD5	TRF4-2	Nucleus/Nucleolar (Lubas et al., 2011; Rammelt et al., 2011; Ogami et al., 2013) Cytoplasm, some cells have additional nuclear staining (Mullen and Marzluff, 2008)		A
MTPAP	Mitochondrial poly(A) polymerase	MTPAP	TENT6, PAPD1, SPAX4	Mitochondrion (Tomecki et al., 2004; Nagaike et al., 2005; Xiao et al., 2006) Nucleus/Cytoplasm/Mitochondrion/Golgi Apparatus (Barbe et al., 2008) Cytoplasm (Mullen and Marzluff, 2008)	spastic ataxia autosomal recessive type 4 (SPAX4) [OMIM:613672] (Crosby et al., 2010)	A

*Effective date for HUGO-approved change in gene nomenclature was April 12, 2018.
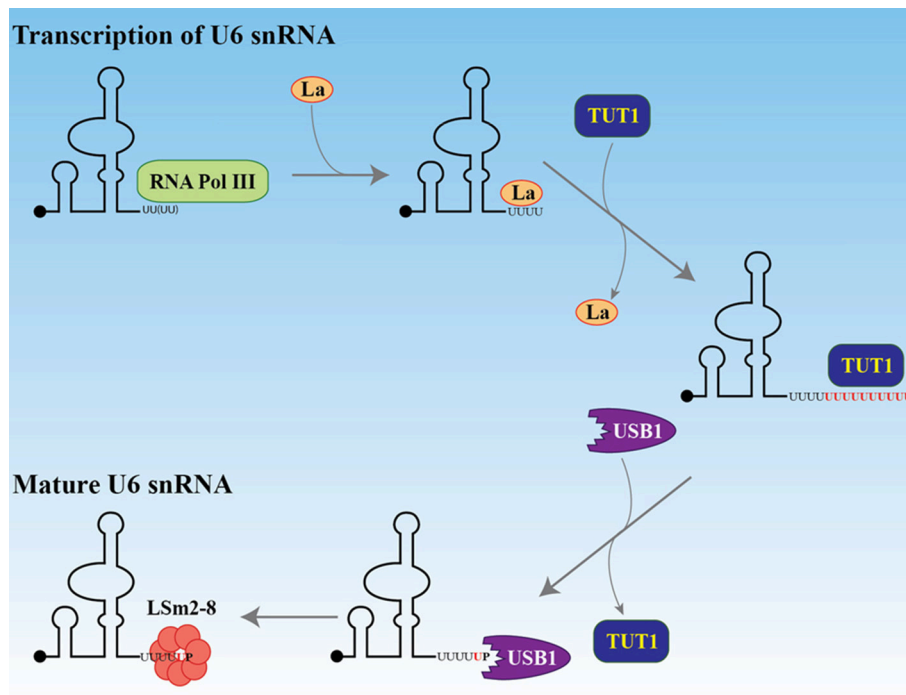
FIGURE 2 | Maturation of U6 snRNA. After transcription by RNA polymerase III, the La protein binds the four uridines at the 3′ end. The black dot at the 5′ end represents the gamma-monomethylguanosine triphosphate cap. The La protein is replaced by TUT1 which adds about 20 uridines to the encoded four. The exonuclease and phosphodiesterase USB1 (MPN1) removes the uridines, keeping five of them and adding a terminal 2′, 3′ cyclic phosphate. This 3′ structure facilitates the recruitment of the LSm2-8 complex and protects the U6 snRNA from degradation by the exosome.

3′ overhang that is ultimately recognized in the cytoplasm by the Dicer/TRBP complex that directs the second cleavage event (Zhang et al., 2004; Park et al., 2011). As depicted in **Figure 3**, Narry Kim's lab through sequencing and bioinformatic analyses discovered that a small number of pre-miRNAs are characterized by a single non-templated uridine at the 3′ end that creates a 2 nucleotide overhang (Heo et al., 2012). These atypical precursor microRNAs have been termed Group II and include most let-7 family members (7a-1, 7a-3, 7b, 7d, 7f-1, 7f-2, 7g, 7i, and miR-98) as well as miR-105. Monouridylation of these precursors promotes their subsequent Dicer processing and increases the levels of mature microRNA levels *in vivo*. Sequencing of pre-let-7 revealed that roughly 20% of all pre-let-7 microRNAs were monouridylated at their 3′ ends in HeLa cells that lack expression of both LIN28A and LIN28B. In contrast, all other single nucleotide additions to pre-let-7 combined represent only 1%. Analysis of numerous deep sequencing libraries indicated that the majority of mature Group II let-7 microRNAs are likely processed from monouridylated precursors, as evidenced by their let-7-3p reads. *In vitro* uridylation assays identified three terminal ribonucleotidyl transferases (TUT4/ZCCHC11, TUT7/ZCCHC6, and TENT2/PAPD4/TUT2/GLD2) competent for mono-uridylating pre-let-7a-1 *in vitro* and each had a preference for pre-let-7a-1 with a 1nt vs. 2nt 3′ overhang. The NTP specificity was determined for these enzymes, revealing that TUT4 and TUT7 utilized UTP, while TENT2 could use

UTP, GTP, or ATP. *In vivo* knockdowns performed singly or in combination revealed that these enzymes function redundantly in HeLa cells where triple knockdowns resulted in the most significant loss of Group II mature let-7 and mono-uridylated pre-let-7. For single knockdown, TUT7 had the most prominent effect. Similar results were observed for miR-105 in the triple knockdowns. It remains unclear if TENT2 is an *in vivo* TUTase and whether its effects are mediated by monoadenylation rather than monouridylation, given the reported affinity of TENT2 for ATP over UTP (Chung et al., 2016). These results highlight the surprising complexity in how let-7 microRNA biogenesis is controlled either by monouridylation that promotes microRNA maturation versus LIN28A-directed polyuridylation that blocks biogenesis that is described in the next section.

Polyuridylation of Select Precursor MicroRNAs Blocks Their Biogenesis

To date, the best-characterized example of 3′ RNA polyuridylation blocking microRNA biogenesis involves LIN28A-mediated repression of the let-7 microRNA family (**Figure 3**) (reviewed in Lee et al., 2016; Balzeau et al., 2017). Briefly, the developmentally regulated RNA binding protein and proto-oncogene Lin28A binds to the terminal loop of pre-let-7 where it recruits the TUT4 to add a short polyU-tail to pre-let-7, blocking Dicer cleavage (Hagan et al., 2009; Heo
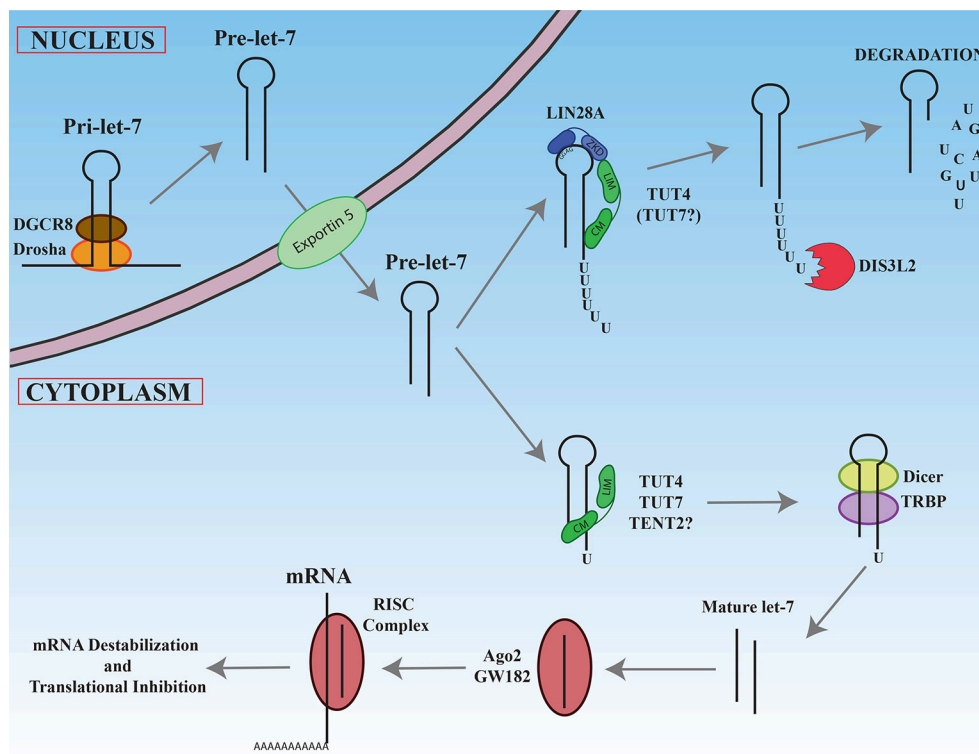
FIGURE 3 | Maturation of let-7 miRNAs. Newly transcribed pri-let-7 is cleaved first in the nucleus by the Drosha/DGCR complex to generate pre-let-7. After export ot the cytoplasm, pre-let-7 can be recognize by the LIN28A protein by its GGAG sequence in the loop. LIN28A recruits via its Zinc knuckle domain (ZKD), TUT4 on its LIN28 interacting module (LIM). The catalytic module (CM) of TUT4 adds a polyU tail to pre-let-7. The resultant this polyU tail blocks Dicer cleavage and is recognized by the exonuclease DIS3L2 that degrades polyuridylated pre-let-7. In the absence of LIN28A, TUT4 and/or TUT7 add only one uridine at the 3′ end, promoting recognition and cleavage by Dicer/TRBP complex. The mature let-7 produced interacts with Ago2, enters in the RISC complex, and base-pairs with its mRNA targets to promote mRNA degradation and/or translational inhibition.

et al., 2009; Piskounova et al., 2011). Polyuridylated pre-let-7 is rapidly degraded by the exonuclease Dis3L2 that recognizes the polyU-tail (Chang et al., 2013; Ustianenko et al., 2013). Let-7 microRNAs have gained considerable attention due to their prominent roles as tumor suppressors. Notably, loss of let-7 expression and miR-21 overexpression are the most commonly dysregulated microRNAs delineating poor clinical prognosis in cancers (Yang et al., 2010; Wang et al., 2011; Nair et al., 2012). Consistent with this finding, let-7 microRNAs act as tumor suppressors by negatively regulating numerous oncogenes such as MYC, RAS, HMGA2, YAP1, CDK6, CCND1, and BLIMP1. In cancer, reactivation of either LIN28A or LIN28B expression accounts for the widespread repression of the tumor suppressor let-7 microRNA family. The significance of the LIN28/let-7 pathway is further highlighted by the fact that LIN28A is a Thomson reprogramming factor whose expression on its own is necessary and sufficient to make induced pluripotent stem cells (Yu et al., 2007; Buganim et al., 2012).

Let-7 is an evolutionarily ancient microRNA family whose founding member was discovered in *C. elegans* as a heterochronic gene that regulates developmental timing during the transition from late larval to adult cell fates. In humans, there are 12 let-7 family members encoded by eight chromosomes. These microRNAs are quite abundant in most cell types and often account for >10% of the entire cellular microRNA population (Griffiths-Jones et al., 2006). In both embryonic stem cells and many cancer cells, multiple let-7 microRNA primary transcripts are actively generated; however, mature let-7 microRNAs are remarkably low, due to post-transcriptional gene regulation mediated by LIN28 paralogs.

LIN28A and LIN28B are two closely related and developmentally regulated RNA binding proteins and are implicated directly in negatively regulating let-7 microRNA biogenesis (Heo et al., 2008, 2009; Newman et al., 2008; Rybak et al., 2008; Viswanathan et al., 2008; Hagan et al., 2009; Piskounova et al., 2011). These proteins contain two RNA binding domains whose interactions with the loop of immature let-7 is required for high affinity binding. Specifically, an N-terminal cold shock domain binds a stem-loop structure in the pre-E element while the C-terminal Zinc knuckles bind to a conserved GRAG in the 3′ of the loop (Newman et al., 2008; Piskounova et al., 2008; Nam et al., 2011). Let-7 microRNAs negatively regulate both LIN28A and LIN28B mRNAs by direct binding to their 3′ UTRs, thereby establishing a double negative feedback loop and self-enforcing binary switch. Although both

LIN28 proteins were initially thought to block Microprocessor cleavage of nuclear pri-let-7, the subcellular localization of these protein suggested an alternate model where LIN28 functions in the cytoplasm to block the conversion of pre-let-7 to its mature and functional form. Reports at the time showed that LIN28A is almost exclusively cytoplasmic (Balzer and Moss, 2007; Polesskaya et al., 2007), while LIN28B exists in the cytoplasm throughout the cell cycle with some nuclear accumulation during S/G2 phases (Guo et al., 2006).

Narry Kim's and our research subsequently discovered that Lin28A recruits TUT4/Zcchc11 that adds an oligouridine tail to the 3′ end of the pre-let-7, blocking Dicer cleavage and provoking the degradation of polyuridylated pre-let-7 (Hagan et al., 2009; Heo et al., 2009). To reach this conclusion, we both used biochemical and cell culture studies to demonstrate that TUT4 is a TUTase and that knockdown of TUT4 elevated mature let-7 levels in Lin28A-expressing cells. Knockdown of no other terminal ribonucleotidyl transferase tested affected mature let-7 levels. Mechanistically, the Zinc Knuckle Domain (ZKD) of LIN28A and the N-terminal region of TUT4 (LIM: LIN28-Interacting Module) is thought to be critical for mediating their protein-protein interactions (Faehnle et al., 2017; Wang et al., 2017). The RNase II exonuclease DIS3L2, a gene whose germline mutation causes Perlman syndrome (Astuti et al., 2012; Higashimoto et al., 2013), has recently been implicated in the degradation of polyuridylated pre-let-7 (Chang et al., 2013; Ustianenko et al., 2013). As expected, Dis3l2 loss does not affect the levels of mature let-7 microRNAs in Lin28A-expressing cells, since polyuridylated pre-let-7 is no longer a viable Dicer substrate.

The Gregory lab has shown that Lin28A and Lin28B *in vitro* can promote uridylation of pre-let-7 by both TUT4 and TUT7, raising the possibility that the LIN28 paralogs can use alternate or redundant TUTases (Thornton et al., 2012). They reported that double knockdown of both TUT4 and TUT7 increased let-7 more than TUT4 knockdown alone in mouse embryonic stem (ES) and P19 embryonal carcinoma cells that express Lin28A (Thornton et al., 2012). Furthermore, this work confirmed earlier studies where TUT7 knockdown on its own does not elevate mature let-7 levels (Hagan et al., 2009; Heo et al., 2009). Our research demonstrated that LIN28A and LIN28B regulate let-7 microRNA biogenesis by distinct mechanisms with differential reliance on the TUT4 (Piskounova et al., 2011). For LIN28B, it remains unresolved the precise mechanism(s) responsible for let-7 regulation, even though in each scenario direct binding of LIN28B to the terminal loop of immature let-7 is involved. Four mechanisms separately or in combination may be important. Specifically, LIN28B may act in the nucleus to block Microprocessor cleavage of pre-let-7, may sequester pri-let-7 in the nucleolus, may block Dicer cleavage of pre-let-7 in the cytoplasm, and lastly, may work with the alternate TUTase TUT7 to block biogenesis. Since the subcellular localization of LIN28B appears cell type dependent, diverse mechanisms of action may be responsible (Guo et al., 2006; Hafner et al., 2010; Piskounova et al., 2011; Molenaar et al., 2012; Suzuki et al., 2015).

In addition to the LIN28A/let-7 pathway, polyuridylation of pre-miRNA has been proposed to repress other microRNAs.

Initially, the Kim lab reported that miR-107, miR-143, and miR-200c are regulated by Lin28A and polyuridylation via TUT4 (Heo et al., 2009); however, their followup study argues against this conclusion (Cho et al., 2012). To define RNAs bound to Lin28A, they performed CLIP-Seq, a genome wide method that incorporates UV crosslinking of live cells to capture RNAs bound to a protein of interest that is subsequently enriched via immunoprecipitation. Following protease digestion, high-throughput sequencing of cDNA libraries is performed that correspond to the co-purified bound RNA. Among pre-microRNAs that bind Lin28A, pre-let-7 was discovered but not pre-miR-107, pre-miR-143, or pre-miR-200c. Moreover, knockdown studies confirmed that Lin28A knockdown only upregulates mature let-7 microRNAs, confirming an earlier report (Hagan et al., 2009).

Polyuridylation has also been proposed to regulate miR-1 biogenesis in myotonic dystrophy patients (Rau et al., 2011); however, this work did not report *in vivo* TUTase loss-of-function or gain-of-function experiments. Myotonic dystrophy is caused by expansion of CTG and CCTG repeats in the DMLK and ZNF9 genes, respectively, that create aberrant RNAs that functionally sequester the splicing regulator MBNL1. In heart samples from myotonic dystrophy patients, microRNA profiling revealed that mature miR-1 levels were markedly reduced. In normal cardiac cells, the RNA binding protein MBNL1 binds to the terminal loop of pre-miR-1-1 and pre-miR-1-2 as determined by UV crosslinking. Rau and colleagues hypothesized that in myotonic dystrophy, MBNL1 is sequestered, thereby permitting LIN28 to bind to the pre-miR-1 terminal loop and block miR-1 maturation (Rau et al., 2011). In HeLa cells that do not express either LIN28 paralog, co-transfection experiments showed that LIN28A reduced mature miR-1 when pri-miR-1 was ectopically expressed. *In vitro* uridylation assays showed that Lin28A promotes uridylation of pre-miR-1 by TUT4. Loss-of-function of MBNL1 and Lin28B in H9C2 rat cardiomyoblasts decreased and increased mature miR-1 levels, respectively, consistent with them having antagonistic functions. In myotonic dystrophy patients, miR-1 loss is predicted to cause upregulation of its target genes such as GJA1 (Connexin 43) and CACNA1C (Cav 1.2), leading to dysregulation of important gap junction and calcium channel proteins in the heart.

Non-canonical MicroRNA Biogenesis

Uridylation also plays roles in non-canonical microRNA biogenesis. In contrast to canonical microRNA biogenesis, a small subset of microRNAs are made by alternative, non-canonical pathways that bypass the requirement for either Drosha or Dicer (reviewed in Miyoshi et al., 2010; Abdelfattah et al., 2014). Dicer-independent and Drosha-independent microRNAs are rare in mammals. The following sections review our current knowledge.

Ago2-Dependent and Dicer-Independent MicroRNAs

For canonical microRNAs, one strand of the microRNA duplex after Dicer cleavage typically is loaded onto an Argonaute protein (Ago1-4), leading to incorporation into the RNA-induced silencing complex (RISC). Although Ago1-4 are individually

competent to bind and load microRNAs and siRNAs into RISC, Ago2 is unique in having slicer activity that is responsible for siRNA-mediated (and to a much lesser extent miRNA-directed) cleavage of their target mRNAs.
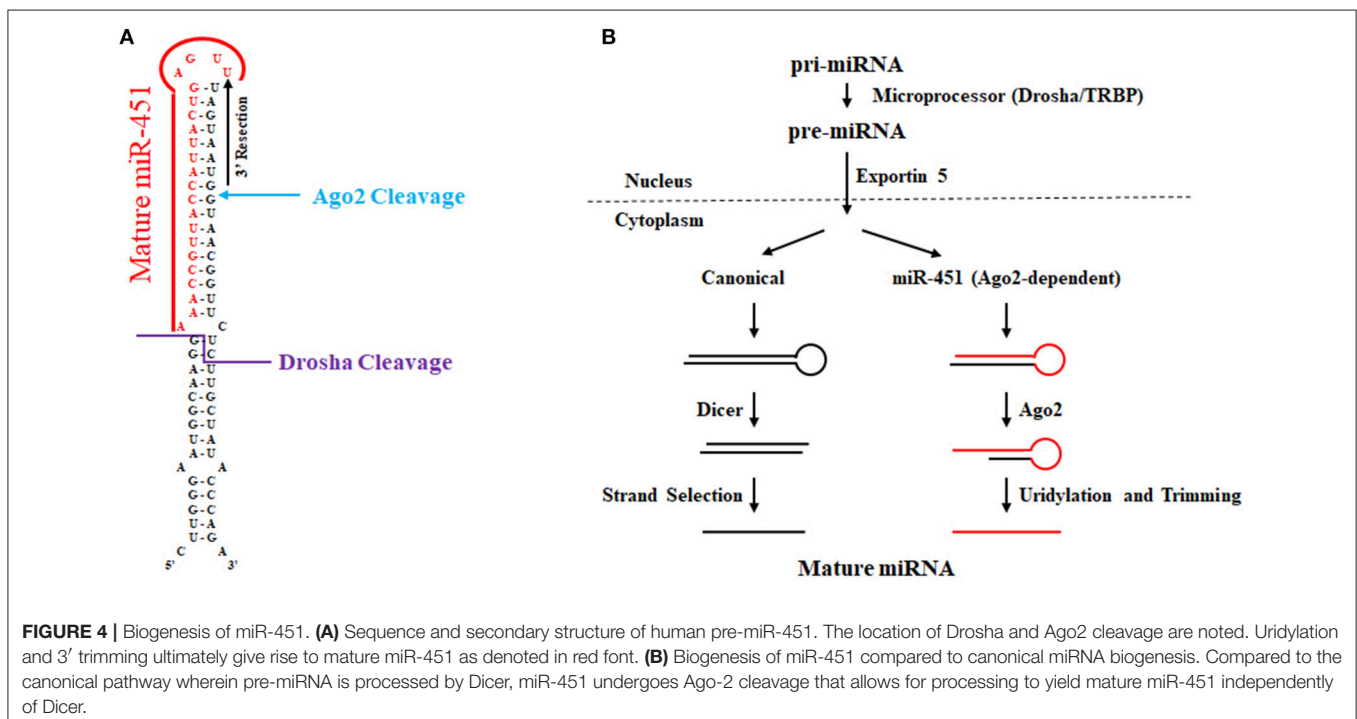
Research initially in zebrafish implicated miR-451 as a key erythropoiesis gene (Dore et al., 2008; Pase et al., 2009) whose function is evolutionarily conserved in mammals (Patrick et al., 2010; Rasmussen et al., 2010). In comparison to other microRNAs as shown in **Figure 4**, the conserved microRNA miR-451 is rather unusual, since Drosha cleavage yields a short (42 nt) hairpin with only a 17 nucleotide stem that is too short for Dicer cleavage (Siolas et al., 2005). In addition, the mature miR-451 sequence includes the entire hairpin loop. Research using zebrafish and mouse models ultimately led to the discovery that miR-451 biogenesis requires Ago2 rather than Dicer (Cheloufi et al., 2010; Cifuentes et al., 2010). In the zebrafish model, microRNA expression was interrogated in control, maternal-zygotic *dicer* mutants, and maternal-zygotic *ago2* mutants at 48 h post fertilization (Cifuentes et al., 2010). Intriguingly, several microRNAs such as miR-451 and miR-735-5p were unaffected by *dicer* loss. In contrast, mature miR-451 expression was abolished in *ago2* mutants. For miR-451, many reads in wild-type and dicer-deficient zebrafish cells contained the first 30 nucleotides of pre-miR-451 with 1–5 non-templated uridines. The last templated base in this RNA species base-pairs with position 10 in pre-miR-451, a location that is reminiscent of how Ago2 slices mRNA targets using siRNA as a guide in RISC. Further biochemical studies revealed that Ago2 associates with pre-miR-451 and is responsible for pre-miR-451 using *in vitro* processing assays. Their data shows that trimming of Ago2-sliced pre-miR-451 gives rise to mature miR-451. Mouse miR-451

was shown by the Hannon lab to undergo a similar process and was initially discovered using genetically engineered mice where Ago2's slicer activity was abolished by a targeted point mutation that resulted in alanine replacing an essential catalytic aspartate. Their mouse work in conjunction with additional *in vitro* and cell culture assays showed that the biogenesis of miR-451 is Ago2-dependent and Dicer-independent. Recently, another study using the catalytically dead Ago2 mice revealed that miR-486 requires both the catalytic activity of Ago2 as well as Dicer cleavage of its precursor. Specifically, Dicer cleavage of pre-miR-486 produces duplex miRNA-5p/miRNA-3p; however, this duplex is arrested and Ago2 is required to cleave the passenger strand to liberate the guide for its incorporation into the RISC complex.

Mirtrons

Another alternative microRNA biogenesis pathway termed "mirtrons" was recently discovered wherein, splicing generates pre-miRNA hairpin mimics, thereby bypassing the requirement for Drosha (reviewed in Rissland, 2015). Mirtrons have been identified in diverse animals including worms, flies, and humans (Berezikov et al., 2007; Okamura et al., 2007; Ruby et al., 2007; Bortolamiol-Becet et al., 2015; Reimao-Pinto et al., 2015). Typical mirtrons have both hairpin ends generated by splicing forming a short 3′ overhang for nuclear export and cleavage (**Figure 5**). In mammals, short introns are relatively uncommon that could produce a canonical mirtron. A limited number of microRNAs are now definitively known to be mirtrons and include miR-877 and miR-1224 in mammals and miR-1225 and miR-1226 in primates (Berezikov et al., 2007).

Bioinformatic studies have identified hundreds of additional candidate murine and human mirtrons that are predominantly

FIGURE 4 | Biogenesis of miR-451. **(A)** Sequence and secondary structure of human pre-miR-451. The location of Drosha and Ago2 cleavage are noted. Uridylation and 3′ trimming ultimately give rise to mature miR-451 as denoted in red font. **(B)** Biogenesis of miR-451 compared to canonical miRNA biogenesis. Compared to the canonical pathway wherein pre-miRNA is processed by Dicer, miR-451 undergoes Ago-2 cleavage that allows for processing to yield mature miR-451 independently of Dicer.

FIGURE 5 | Pathways leading to mirtron biogenesis. Splicing can generate three types of mirtrons. Introns are spliced and the lariat is debranched to allow folding into pre-miRNA-like hairpins. Splicing also leads to tailed mirtrons which undergo additional trimming step after splicing and debranching. For 5′ mirtrons, trimming is carried out by an unknown nuclease while for 3′ mirtrons, trimming is carried out by the exosome. After export to the cytoplasm, the pre-miRNA hairpins are cleaved by Dicer and loaded onto Argonaute complexes.

comprised of endogenous Dicer substrates with unique patterns of ordered 5′ and 3′ heterogeneity (Ladewig et al., 2012; Westholm et al., 2012; Wen et al., 2015). In most cases, the end of mammalian mirtrons are characterized by the presence of oligouridine tails due to untemplated uridylation whose functional significance remains unclear. In *Drosophila*, the TUTase Tailor uridylates mirtrons and is linked to mirtron turnover (Bortolamiol-Becet et al., 2015; Reimao-Pinto et al., 2015). It remains unclear how uridylation of mirtrons affects their biogenesis in mammals.

IsomiRs and Uridylation

Mature microRNAs can differ in sequence and length from their canonical miRNA sequence reported in miRBase. These variants, known as isomiRs, are of different types: there is heterogeneity in Drosha and Dicer cleavage site selection, templated additions, or deletions at the 5′ and/or 3′ termini of the miRNA, non-templated additions at the 3′ end and substitutions within the sequence (Morin et al., 2008). Non-templated variations have been attributed to the activity of ribonucleotidyl transferases, posttranscriptional modifications or the presence of genetic variants within the miRNA transcript (Morin et al., 2008; Cloonan et al., 2011; Wyman et al., 2011; Neilsen et al., 2012; Lee et al., 2013; Ma et al., 2013; Starega-Roslan et al., 2015; McCall et al., 2017). IsomiRs due to altered target specificity may regulate different subset of genes compared to the canonical miRNA and

may play a role in disease pathogenesis (Gong et al., 2012; Tan et al., 2014). Although this review focuses on uridylation, it is known that other terminal ribouncucleotidyl transferases are significant players in isomiR formation. For example, TENT2 (aka GLD-2) through monoadenylation stabilizes miR-122, a microRNA that is highly expressed in the liver (Katoh et al., 2009; Burns et al., 2011; D'Ambrogio et al., 2012).

To identify and quantify isomiRs, small-RNA-Seq has been routinely used. In this method, sequencing libraries are prepared from size fractionated RNAs to enrich for small RNAs that are used as the initial template. For the high-throughput sequencing itself, a single read is sufficient and the overall number of required reads is less in contrast to mRNA libraries that typically use paired-end reads and require far more sequencing depth. A handful of bioinformatic algorithms exist to analyze small-RNA-Seq data for isomiRs (miraligner, isomiRex, isomiRID, IsomiRage, DeAnnIso, mir-isomiRExp, isomiR-SEA; Pantano et al., 2010; de Oliveira et al., 2013; Sablok et al., 2013; Muller et al., 2014; Guo et al., 2016; Urgese et al., 2016; Zhang et al., 2016). Of interest, a recent report by Garalde and colleagues suggests that direct RNA sequencing using nanopore technology has several beneficial attributes and has the ability to detect specific nucleotide modifications in RNA (Gyarfas et al., 2009). Since the RNA is sequenced directly, nanopore does not lose information relative to post-transcriptional RNA

modification in comparison to traditional RNA-Seq. RNA-Seq suffers from additional technical liabilities associated with the library preparation that includes PCR amplification and the short reads make definitive analysis of alternative splicing events difficult.

A striking example of the complexity of post-transcriptional regulation linked to microRNA uridylation involves TUT4, miR-26a, miR-26b, and interleukin-6 (IL-6) (Jones et al., 2009). In LIN28B-expressing A549 cells, Jones and colleagues showed that knockdown of TUT4 decreased the inflammatory cytokine IL-6 at both the mRNA and protein levels. IL-6 mRNA in TUT4 knockdowns had reduced half-life and a shorter polyA-tail length. The IL-6 3′ UTR was sufficient to convey TUT4-dependence on a chimeric luciferase construct. Deep sequencing of A549 cells revealed that miR-26a in control cells exists largely (>80%) as monouridylated 22 nucleotide microRNA, while in TUT4-deficient cells two populations predominate: a 21 nucleotide microRNA missing the 3′ non-templated U (~50%) or as a 23 nucleotide miRNA ending with a non-templated UA dinucleotide (~28%). In contrast, miR-26b reads are increased in TUT4 knockdowns by >20-fold. Luciferase assays showed using microRNA mimics that the ability of miR-26b to repress IL-6 decreases as the length of non-templated Us increase. Overall, their data suggests that TUT4 in part promotes IL-6 expression by decreasing miR-26b levels as well as shifting the isomiR population of miR-26a. A subsequent study has implicated miR-26a in negatively regulating both LIN28B and TUT4, leading to let-7 upregulation (Fu et al., 2013). It remains unknown if TUT4 can regulate directly IL-6 mRNA turnover via microRNA-independent mechanisms.

Recent work from the Gregory and Zon labs have interrogated the consequence of simultaneous knockdowns of both TUT4 and TUT7 on isomiR generation (Thornton et al., 2014). In HeLa cells that do not express the LIN28 paralogs, microRNA sequencing revealed that overall abundance of individual microRNAs was largely unaffected by TUTase loss. Depending on the particular microRNA, up to 8% were determined to be uridylated relative to total reads. TUTase depletion reduced mature microRNA uridylation for multiple microRNAs as expected and resulted in a concomitant gain in adenylation.
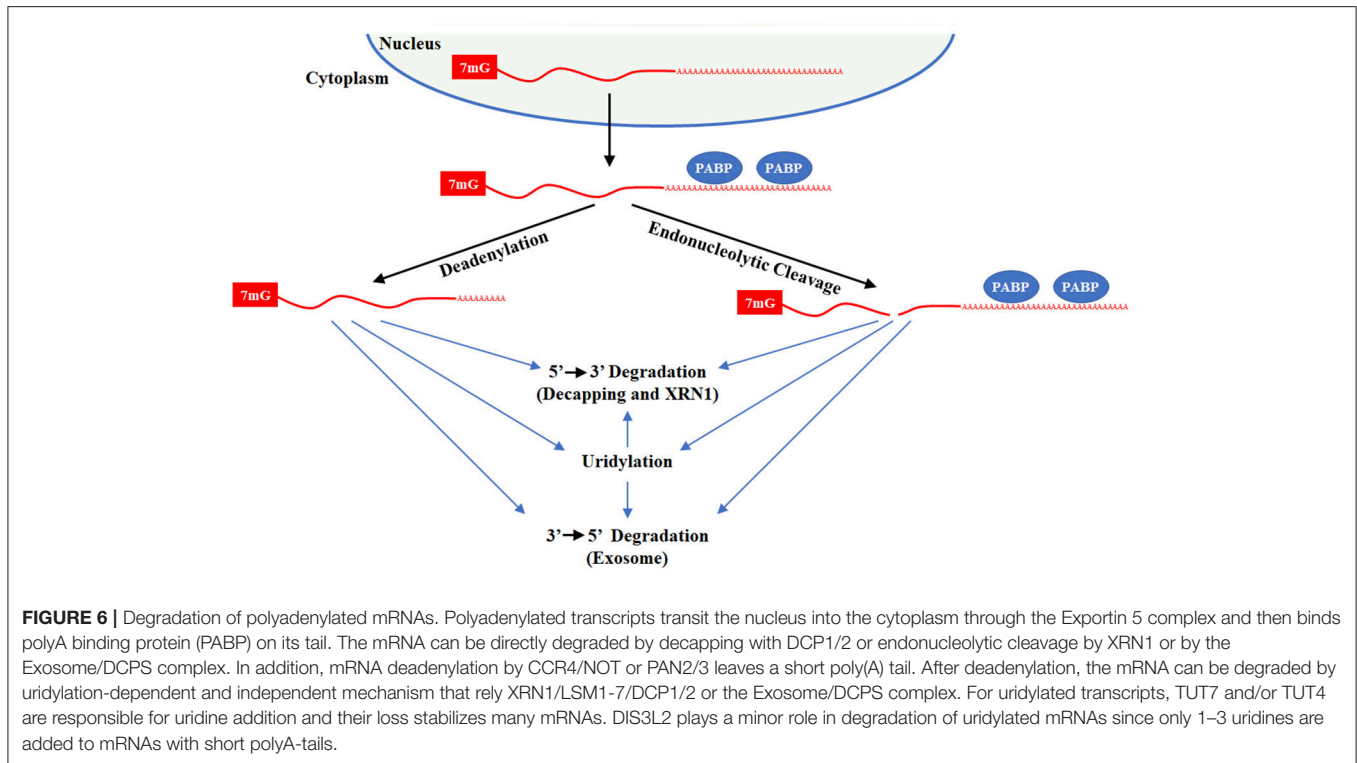
URIDYLATION IN MRNA TURNOVER

Transcription and degradation are both critical for controlling mRNA abundance. Like transcription, mRNA turnover is a regulated process that integrates a multitude of factors including cell-type specificity, timing within the cell cycle, and response to intrinsic and extrinsic signals. Multiple mechanisms exist that are important for mRNA turnover that have differential reliance on deadenylation and uridylation. In the forthcoming sections, we will elaborate on the functions of uridylation in degradation of polyadenylated mRNAs, histone mRNAs that lack polyA-tails, and cleaved mRNAs.

Uridylation in Polyadenylated mRNA Degradation

Almost all mRNAs end with a polyA-tail that promotes transcript stability and translational efficiency. Over time, polyA-tails become shorter, decreasing translation and enhancing degradation. Mechanistically, mRNA can be actively degraded by multiple mechanisms where there is significant commonality in the critical enzymes (**Figure 6**). Decapping and 5′→3′ exonuclease, typically XRN1, degrade mRNA from the 5′ end while the exosome degrades mRNA 3′→5′. Emerging evidence demonstrates that uridylation contributes to mRNA turnover. The first evidence that uridylation plays a role in mRNA turnover was derived from research on caffeine-induced death suppressor (Cid1) in the fission yeast *Schizosaccharomyces pombe* (Rissland et al., 2007). In *S. pombe*, caffeine in combination with hydroxyurea treatment leads to genomic instability and cell death due to failure in S-M checkpoint control. In a suppressor screen of this phenotype, Cid1 was identified. Bacterially expressed recombinant Cid1 utilizes both UTP and ATP while the Cid1 complex purified for fission yeast acts predominantly as a polyU polymerase that can uridylate multiple mRNAs, targeting them for degradation (Rissland et al., 2007; Rissland and Norbury, 2009).

In mammals, it was widely assumed that polyA tails end in adenine for obvious reasons. In groundbreaking studies by Narry Kim's lab, a novel technique that they termed TAIL-Seq was employed to interrogate the terminal ends of transcripts, revealing that a subset of mRNA transcripts in end U (Chang et al., 2014; Lim et al., 2014). TAIL-Seq is an innovative method to capture the true 3′ end of RNAs. In this method, rRNA are first depleted and a 3′ adapter with biotin is ligated to the 3′ end of the RNA. RNA is fragmented by modest digestion with RNase T1. A streptavidin purification is then used, the 5′ end of RNA is phosphorylated, and a 5′ adapter is added. At this point, library preparation and paired-end high throughput sequencing follows traditional protocols.

Using TAIL-Seq, the Kim lab identified 1–3 uridine addition as a common modification at the end of mRNAs with short polyA-tails (<25 As) and guanylation at the end of mRNAs with longer polyAs (>40 nts) at the downstream of the poly(A) tail. Roughly 50 and 80% of polyadenylated transcripts are uridylated at a frequency of >5 and >2%, respectively, with specific transcript as high as 40%. Considering that uridylation targets mRNAs for degradation, the levels of uridylated transcripts at steady state are rather remarkable. Furthermore, they showed that the TUT4 and TUT7 are the responsible TUTases for mRNA uridylation and their knockdown increased the half-life of numerous transcripts. Overall, they discovered more than >600 genes upregulated in HeLa cells upon TUTase loss. Their data supports the idea that uridylated mRNAs are degraded by canonical pathways where DIS3L2 only has a minor role as anticipated given the limited number of Us added to mRNA. The further importance of TUTases in regulation of the maternal transcriptome during oocyte development and in the maternal-to-zygotic transition that occurs shortly after fertilization is

FIGURE 6 | Degradation of polyadenylated mRNAs. Polyadenylated transcripts transit the nucleus into the cytoplasm through the Exportin 5 complex and then binds polyA binding protein (PABP) on its tail. The mRNA can be directly degraded by decapping with DCP1/2 or endonucleolytic cleavage by XRN1 or by the Exosome/DCPS complex. In addition, mRNA deadenylation by CCR4/NOT or PAN2/3 leaves a short poly(A) tail. After deadenylation, the mRNA can be degraded by uridylation-dependent and independent mechanism that rely XRN1/LSM1-7/DCP1/2 or the Exosome/DCPS complex. For uridylated transcripts, TUT7 and/or TUT4 are responsible for uridine addition and their loss stabilizes many mRNAs. DIS3L2 plays a minor role in degradation of uridylated mRNAs since only 1–3 uridines are added to mRNAs with short polyA-tails.

discussed in section TUTases in Vertebrate Development and Physiology.

During apoptosis, uridylation plays a critical role in the rapid degradation of global mRNAs that relies prominently on DIS3L2 (Thomas et al., 2015). For this process, TUT4 and TUT7 are responsible for widespread mRNA uridylation. Knockdown of TUTases or DIS3L2 impaired apoptotic mRNA decay and decreased programmed cell death, while DIS3L2 overexpression promotes apoptosis. When taken together, these results suggest that dysregulated uridylation may play multiple roles in disorders involving apoptotic cell death and/or aberrant mRNA turnover, leading to dysregulated gene expression and cell survival.
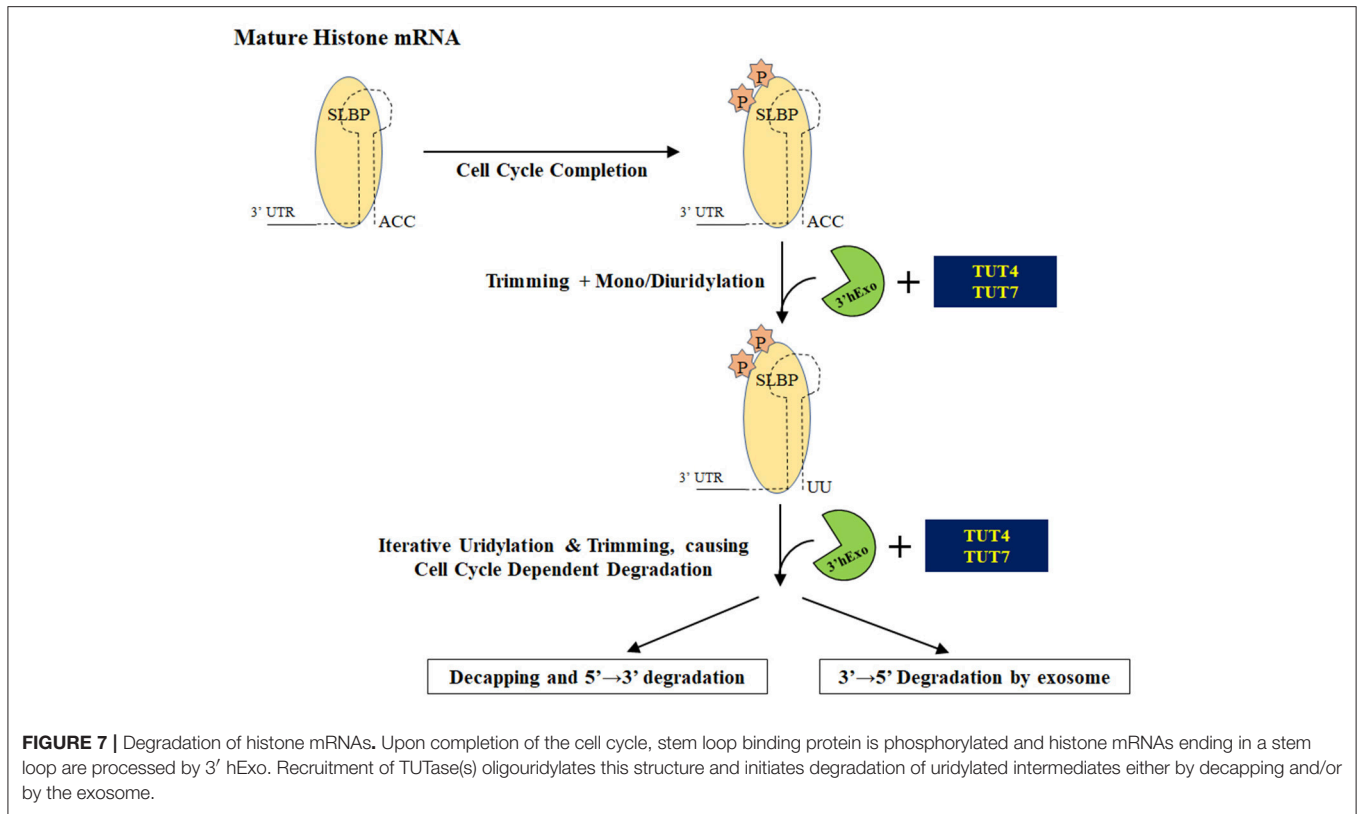
Uridylation and Replication-Dependent Histone mRNA Turnover

Replication-dependent histone mRNAs are exclusively present during the S phase of the cell cycle to encode histone proteins required for packaging of newly synthesized DNA (reviewed in Hoefig and Heissmeyer, 2014; Marzluff and Koreski, 2017). In eukaryotes, this restricted expression is accomplished through regulation of both histone mRNA transcription and degradation (**Figure 7**). In metazoans, histone mRNAs are unique because they end in a highly conserved stem-loop and are not polyadenylated like other mRNAs. The stem-loop structure interacts with LSm1-7 and Stem-Loop Binding Protein (SLBP) to form a complex that recruits 3′ hExo to bring about the degradation of histone mRNAs at the end of S phase or during the inhibition of DNA replication (Mullen and Marzluff,

2008; Hoefig et al., 2013; Lyons et al., 2014; Brooks et al., 2015).

A pioneering study by Mullen and Marzluff demonstrated that histone mRNA turnover is initiated by 3′ terminal oligouridylation that leads to either 3′→5′ degradation by the exosome and/or by decapping and subsequent 5′→3′ degradation by the exonuclease Xrn1 (Mullen and Marzluff, 2008). A subsequent collaborative study by the Rhoads and Marzluff labs showed that uridylated histone mRNAs are predominantly degraded by decapping and then 5′→3′ degradation (Su et al., 2013). The TUTase(s) involved in oligouridylation of histone mRNAs has not been definitively identified. Different terminal ribonucleotidyl transferases have been proposed as important for histone uridylation: MTPAP, TENT4B, TUT4, and TUT7. The emerging consensus is that TUT4 and/or TUT7 are the responsible TUTase(s). Both MTPAP and TENT4B have been shown to be polyA polymerases and lack the critical histidine that confers UTP specificity (Tomecki et al., 2004; Nagaike et al., 2005; Xiao et al., 2006; Rammelt et al., 2011; Berndt et al., 2012; Boele et al., 2014; Yamashita et al., 2017). Furthermore, MTPAP protein is almost exclusively mitochondrial where it is responsible for polyadenylation on mRNA transcripts encoded by mitochondrial DNA (Tomecki et al., 2004; Nagaike et al., 2005; Xiao et al., 2006). Subsequent work showed that MTPAP likely has an indirect effect on histone mRNA levels, since MTPAP knockdown impairs cell growth and reduces the number of cells actively going through S phase (Su et al., 2013).

In 2011, the Norbury lab showed that knockdown of the cytoplasmic TUT4 in HeLa cells blocked histone mRNA

FIGURE 7 | Degradation of histone mRNAs. Upon completion of the cell cycle, stem loop binding protein is phosphorylated and histone mRNAs ending in a stem loop are processed by 3′ hExo. Recruitment of TUTase(s) oligouridylates this structure and initiates degradation of uridylated intermediates either by decapping and/or by the exosome.

degradation and subsequently stalled DNA replication, accompanied by a proportional reduction in uridylated histone mRNA transcripts (Schmidt et al., 2011). The significance of TUT4 was independently confirmed in a subsequent study (Su et al., 2013). These data suggest that TUT4 is the terminal U-transferase responsible for directing histone mRNAs for degradation upon the inhibition or completion of DNA replication. Lackey and colleagues found through high-throughput sequencing that knockdown of TUT7 reduces uridylation at the 3′ end of histone mRNA transcripts as well as uridylation of degradation intermediates in the stem loop (Lackey et al., 2016). In contrast to prior reports, a knockdown of TUT4 did not produce an effect on the 3′ uridylation pattern or on the stem-loop, suggesting that TUT7 in concert with 3′hExo play a major role in trimming and uridylation of histone mRNAs while TUT4 plays a minor role in this process.

Uridylation Promotes Degradation of Cleaved mRNAs

The discovery of RNA interference and the experimental utility of shRNA/siRNAs has revolutionized how gene function is interrogated (Mello and Conte, 2004). In cases of perfect (or near perfect) complementarity, the Ago2-RISC complex directs siRNA-directed cleavage of their targeted mRNA. In contrast, miRNAs in the RISC complex due to their imperfect complementarity to their mRNA targets usually result in mRNA destabilization and/or translational inhibition. When

Ago2-RISC does cleave the target mRNA, the resulting 5′ mRNA fragment is often characterized by the addition of 1–24 uridines that promotes degradation (Shen and Goodman, 2004).

URIDYLATION IN NON-CODING RNA QUALITY CONTROL

Non-coding RNAs such as rRNAs, tRNAs, snRNAs, snoRNAs are involved in diverse cellular processes ranging from chromosome replication to mRNA translation. In certain cases, these ncRNAs may be non-functional due to mutations in their genetic sequence, transcriptional errors, premature termination, and misprocessing. Therefore, it is imperative to eliminate aberrant ncRNAs in order to ensure proper cellular function and avoidance of diseases. Elimination of such misprocessed ncRNA precursors is mediated in part by uridylation of ncRNA termini.

Haas and colleagues used a proteomics approach to identify protein complexes involved in miRNA tailing and trimming and identified TUT1 and DIS3L2 as likely candidates to be involved in ncRNA quality control (Haas et al., 2016). Studies with DIS3L2 and its catalytically inactive variant in Perlman syndrome revealed that DIS3L2 recognizes uridylated ncRNAs in the cytoplasm and subsequently degrades them (Labno et al., 2016). In this case, uridylation was carried out by TUT4 and TUT7 regardless of the origin of the ncRNAs. CLiP-Seq studies revealed that DIS3L2 has affinity for various uridylated

ncRNAs: unprocessed tRNAs, vault RNAs, Y-YNAs, 5S RNA, snoRNAs transcribed by RNA pol III, transcription start site-associated short RNAs (TSSas), extended snRNAs and a FTL short RNA from the ferritin mRNA 5′ UTR transcribed by RNA polymerase II (Pirouz et al., 2016; Ustianenko et al., 2016). Also, polyuridylation of aberrant precursor snRNAs targets them for degradation by DIS3L2, an exonuclease that recognizes polyU-tails (Faehnle et al., 2014; Labno et al., 2016; Ishikawa et al., 2018). Also, TUT7 and to a lesser extent TUT4 is thought to be critical for elimination of trimmed pre-miRNAs (Kim et al., 2015). These findings suggest that the TUT-DIS3L2 surveillance (TDS) is responsible for monitoring and elimination of aberrantly structured ncRNAs.

TUTASES IN VERTEBRATE DEVELOPMENT AND PHYSIOLOGY

The *in vivo* functions of TUTases have been elucidated in morphant zebrafish and in genetically engineered knockout mice. In zebrafish, knockdown of either *tut4* or *tut7* via morpholino injection into fertilized eggs causes the majority of fish to die within 5 days post-fertilization (Thornton et al., 2014). These morphant zebrafish are characterized by developmental delay, failure in tail elongation, and somite degeneration. *In situ* hybridization revealed that several Hox genes had aberrant expression patterns in the morphant zebrafish. Of note, the morpholinos used in this experiment block pre-mRNA splicing and therefore do not interrogate the role of the spliced maternal TUTase transcript that exist in the unfertilized egg.

In mammals, TUT4 works with Lin28A to block let-7 microRNA biogenesis (Hagan et al., 2009; Heo et al., 2009; Piskounova et al., 2011), while Lin-28 utilizes the TUTase PUP-2 to suppress let-7 in the nematode *C. elegans* (Lehrbach et al., 2009), implicating uridylation as an ancient mechanism important for Lin28 repression of let-7. Similar to TUTase-deficient fish, loss of either *lin-28a* or *lin-28b* in zebrafish caused multiple phenotypes, including developmental delay (Ouchi et al., 2014). These embryos were characterized by severe gastrulation defects that led to ∼40% lethality by 12 h post fertilization and those that survived longer had reduced body lengths and head size.

Knockout mice have elucidated the function of the TUTases TUT4/Zcchc11 and TUT7/Zcchc6 in mammals. TUT7-deficient mice are born in expected Mendelian ratios and appear overtly normal (Kozlowski et al., 2017). TUT4 and TUT7 are broadly expressed at the RNA level where TUT7 is relatively higher in the liver, lungs, and alveolar macrophages. To determine if TUT7 is involved in the defense of inhaled pathogens, wild-type and TUT7 knockout mice were challenged with *S. pneumoniae* through intratracheal exposure. TUT7 nulls had increase neutrophil recruitment during early infection and elevated several cytokine mRNAs such as CXCL5, IL-6, and CXCL1. This finding suggests a role for TUT7 in altering the innate immune response. Organism-wide double knockout of both TUT4 and TUT7 starting at 2 months of age are reported to be overtly healthy for several months, indicating that these genes are likely not essential in adult mammals (Morgan et al., 2017); however, detailed analyses of these mice remain to understand how uridylation may contribute in more subtle ways to adult physiology across multiple tissues.

TUT4 knockout mice are born normal in size and at expected Mendelian ratios (Jones et al., 2012); however, null mice are characterized by postnatal growth retardation that persists throughout life and roughly 50% of nulls die within a week of birth. For knockout mice that survive to weaning age, they are long lived. Reduced levels of insulin-like growth factor 1 (Igf-1) in the liver and the blood were reported to contribute to the observed growth retardation and lethality. Igf-1 is a potent mitogen previously implicated in organismal growth (Baker et al., 1996). Mechanistically, the 3′ UTR of Igf-1 was critical for TUT4-mediated regulation as shown by luciferase reporter assays, suggesting possible microRNA involvement. Of note, deep sequencing of microRNAs in neonatal livers revealed that several mature microRNA species in TUT4-deficient mice had reduced levels of non-templated uridine additions while the overall abundance of individual microRNAs was largely unaffected. For mature microRNAs characterized by differential uridylation, several are predicted to repress Igf-1. The uridylated forms of miR-126-5p and miR-379 have diminished ability to repress luciferase expression constructs that harbor the IGF-1 3′ UTR. Therefore, loss of uridylation of these microRNAs is thought to enhance their ability to repress Igf-1. One caveat to this study is that it was conducted prior to the realization that TUTases promote degradation of mRNAs with short polyA-tails (Chang et al., 2014; Lim et al., 2014).

Regulation of miRNA expression is necessary for CD4 T-cell maturation. Using deep sequencing, Vasquez et al. demonstrated upon T-cell activation, terminally uridylated miRNA sequences decreased, accompanied by a proportional decrease in the levels of TUT4 and TUT7 (Gutierrez-Vazquez et al., 2017). Furthermore, analysis of TUT4 deficient T lymphocytes, demonstrated that TUT4 is essential for the maintenance of miRNA uridylation in steady-state T lymphocytes. This study underscores the role of post-transcriptional uridylation in regulating miRNA levels during T-cell activation.

In mammals, the program for early development is established during oogenesis through the maternal transcriptome which is achieved by stabilization of RNA binding proteins and suppression of RNA degradation pathways. However, during oocyte growth, limited degradation is required to confer distinctiveness to the maternal transcriptome. In a recent study by Morgan et al. TAIL-Seq was used to demonstrate that during oocyte growth in mice, 3′ terminal oligo-uridylation of mRNA that is mediated by TUT4 and TUT7 facilitates the elimination of certain transcripts (Morgan et al., 2017). These events play a critical role in defining the maternal transcriptome. They generated mice in which Zp3-Cre was used to knockout conditionally both TUT4 and TUT7 during oogenesis starting at the secondary stage. TUT4/TUT7 conditional knockout mice were infertile but underwent ovulation normally. Further studies revealed that this phenotype was due to the inability

of these animals to support early embryonic development due to meiosis failure. This was mainly attributed due to a lack of elimination of TUT4 uridylated transcripts that resulted in an inaccurate maternal transcriptome. This landmark study revealed that uridylation is imperative for accurate development of the maternal transcriptome, which in turn regulates oocyte maturation and fertility (Morgan et al., 2017).

After fertilization, TUT4 and TUT7 are implicated in the targeted degradation of maternal transcripts during the maternal-to-zygotic transition (MZT) in multiple vertebrate species that include frogs, fish, and mice (Chang et al., 2018). To reach this conclusion, the Kim lab profiled mRNAs by TAIL-Seq at fertilization and immediately thereafter. They discovered that a marked increase in uridylation during MZT that occurs 4–6 h post fertilization. Using TUT4 and TUT7 morpholinos targeting the start codon that blocks translation of the pre-existing spliced maternal transcripts for these two genes, they found that MZT uridylation was impaired, a subset of maternal transcripts was stabilized, and that TUTase loss cause significant gastrulation defects in both zebrafish and frogs. TUT7 was found to be primarily responsible for mRNA uridylation prior to the blastula stage and its loss recapitulated the observed failures in gastrulation. In addition, morpholinos that block splicing of the TUTase pre-mRNAs did not cause MZT defects indicating that new zygotic expression of the TUTases is not required for this process and confirmed the significance of maternal TUTase mRNAs. Altogether, these results highlight that TUTases play critical roles in the elimination of maternal transcript that are required during MZT. It should be noted that TUT7 null mice are born in appropriate Mendelian ratios and as such, its loss in insufficient to cause embryonic lethality (Kozlowski et al., 2017).

To elucidate the developmental roles of Lin28 in mammals and by extension shed light on potential functions for TUTases, we generated and interrogated conditional mouse knockouts for both Lin28A and Lin28B in collaboration with George Daley's group (Zhu et al., 2011; Shinoda et al., 2013a,b). Lin28B nulls are overtly normal except for a modest impairment in male postnatal growth. Lin28A mouse knockouts are born in the expected Mendelian ratios; however, these mice are growth impaired and the vast majority die perinatally. Mechanistically, loss of Lin28A during embryogenesis leads to lifelong defects in glucose metabolism that likely contribute to slowed postnatal growth in survivors. Additional analysis of these conditional knockout as well as gain-of-function Lin28 mice reveal a direct role for Lin28 in regulating glucose metabolism and cellular bioenergetics, in part by let-7 mediated regulation of the Insulin-PI3K-mTOR pathway (Zhu et al., 2010, 2011). Specifically, muscle specific loss of Lin28A or overexpression of let-7 resulted in insulin resistance and impaired glucose tolerance. Double knockout of both Lin28A and Lin28B results in lethality by embryonic day 12.5, with numerous defects including developmental delay and neural tube closure defects, suggesting that the paralogs have partially redundant developmental functions. Conditional knockout of either Lin28A or Lin28B at 6 weeks of age yield no overt phenotypes, demonstrating that these genes do not play essential functions shortly after birth. This result is not surprising

as expression of both genes is largely restricted to embryonic development.

TUTASES IN DISEASE

Given the relative infancy of the 3′ RNA uridylation field, our understanding of how uridylation contributes to disease is rather limited. Recent and provocative evidence implicates uridylation as a critical gene regulator and driver of tumorigenesis. These data indicate dysregulated TUTase activity alone and in concert with the onco-fetal LIN28/let-7 pathway as hallmarks of poor prognosis in multiple cancer types. In addition, Perlman syndrome and potentially a subset of Wilms tumors are caused by mutations in DIS3L2, a 3′→5′ exonuclease that is responsible for degrading numerous polyuridylated RNA substrates. As discussed earlier, polyuridylation may play a role in myotonic dystrophy. As research continues, it is likely that defective uridylation will be discovered that has significance in additional diseases.

TUTases and the LIN28/let-7 Pathway in Cancer

The global significance of the onco-fetal LIN28/let-7 pathway has been the subject of multiple reviews, given its importance in ES biology, iPSC reprogramming, and cancer (reviewed in Bussing et al., 2008; Peter, 2009; Viswanathan and Daley, 2010; Lee et al., 2016; Balzeau et al., 2017). The LIN28 paralogs, LIN28A and LIN28B, are RNA binding proteins and proto-oncogenes whose expression occurs primarily during embryonic development. In multiple cancers, one of the LIN28 paralogs will be reactivated, blocking biogenesis of the tumor suppressor let-7 microRNA family. For LIN28A in particular, it recruits the TUT4 to polyuridylate pre-let-7, blocking let-7 maturation and function. Let-7 exerts its tumor suppressor function in part by inhibiting the expression of numerous oncogenes such as MYC, RAS, YAP1, HMGA2, and CDK6 and by altering cellular bioenergetics, including glucose metabolism. LIN28A and LIN28B expression in cancer is typically mutually exclusive and expression of either gene is almost invariably associated with poor prognosis. Since there are 12 let-7 family members encoded by eight human chromosomes, the activation of the LIN28/let-7 pathway explains the global post-transcriptional let-7 repression observed in many human cancers.

The LIN28/let-7 pathway is directly implicated in cancer using genetically engineered mouse models where ectopic LIN28A and/or LIN28B expression cause and/or enhance progression of neuroblastoma, Wilms tumor, mast cell leukemia, hepatocellular carcinoma, and colorectal adenocarcinoma (Molenaar et al., 2012; Nguyen et al., 2014; Urbach et al., 2014; Tu et al., 2015; Wang et al., 2015). Reintroduction of let-7 expression effectively impedes tumor growth or even causes tumor regression in several mouse models (Esquela-Kerscher et al., 2008; Viswanathan et al., 2009; Trang et al., 2010; Piskounova et al., 2011). For example, our research demonstrated that TUT4-depletion or let-7 reintroduction caused established xenograft tumors to regress in mice (Piskounova et al., 2011). LIN28 paralog expression

confers resistance via a let-7 dependent mechanism to ionizing radiation and several chemotherapies (Weidhaas et al., 2007; Blower et al., 2008; Yang et al., 2008; Chen et al., 2009; Oh et al., 2010; Boyerinas et al., 2012). Altogether, these data provide strong evidence implicating the onco-fetal LIN28/let-7 pathway in cancer and as an attractive target for new therapies. This view is further reinforced by our conditional mouse knockout studies that demonstrate that loss of either Lin28A or Lin28B proto-oncogenes at 6 weeks of age cause no overt phenotypes (Shinoda et al., 2013b), consistent with predominant embryonic expression of these genes. In addition, conditional knockout of both TUT4 and TUT7 in adult mice are overtly healthy (Morgan et al., 2017). Therefore, targeted inhibition of LIN28 is less likely to cause deleterious side effects in cancer patients. Recently, the Gregory lab has performed a high throughput biochemical screen to identify mouse TUT4 inhibitors (Lin and Gregory, 2015a). Although the majority of initially identified compounds were thiol-containing false positives due to lack of a reducing agent in the screen, several compounds were ultimately identified that could block pre-let-7 uridylation *in vitro*.

In addition to the role of uridylation in LIN28A-expressing cancers, TUT4 is implicated in both gliomas and breast cancer, independently of the LIN28/let-7 pathway. In breast cancer, Hallett and Hassell defined a dual gene classifier to predict breast cancer survival using E2F1 and TUT4 (i.e., KIAA0191; Hallett and Hassell, 2011). Other proliferation-related genes (e.g., BUB1 or AURKA) could substitute for E2F1 in their 2-gene signature and TUT4 was not prognostic on its own. When taken together, these results suggest that TUT4 overexpression in certain contexts may contribute to poor clinical prognosis in breast cancer. Of note, this classifier was independent of breast cancer subtype and the LIN28/let-7 pathway. Proteomic studies have also revealed that overexpression of TUT4 is common in high grade gliomas in comparison to low grade (Gerth et al., 2013). The potential significance of this association remains to be elucidated.

DIS3L2 in Perlman Syndrome and Wilms Tumor

Perlman syndrome is a rare, autosomal recessive congenital overgrowth and cancer susceptibility disorder (Neri et al., 1984, 1985). In 2012, germline mutations in the DIS3L2 exonuclease was discovered by a team working on Perlman syndrome (Astuti et al., 2012). Subsequently, two DIS3L2 heterozygous missense changes were identified in sporadic Wilms tumor, a tumor highly associated Perlman syndrome. Astuti and colleagues showed that the loss of DIS3L2 is associated with mitotic abnormalities and the abnormal expression of mitotic proteins, introducing the tumor suppressor activity of DIS3L2 (Astuti et al., 2012). A case of homozygous deletion of DIS3L2 exon 9 in a Japanese patient with Perlman syndrome was described later (Higashimoto et al., 2013). DIS3L2 is a cytoplasmic RNA exonuclease that is important for degrading numerous polyuridylated non-coding RNAs through recognition of their polyU tails (reviewed in Morris et al., 2013; Pashler et al., 2016).

Recent work has identified multiple non-coding RNAs as direct targets for Dis3l2-mediated decay where the underlying theme is terminal polyuridylation is the key specificity determinant. Chang and colleagues showed that Dis3l2 knockdown in mouse embryonic stem cells does not affect mature let-7 levels, consistent with polyuridylated pre-let-7 no longer being a Dicer substrate; however, an increase in polyuridylated pre-let-7 levels was observed. They normalized their data to U6 snRNA, a key spliceosomal component that undergoes oligouridylation and trimming. *In vitro*, Dis3l2 preferentially interacts with and degrades polyuridylated precursor microRNAs (e.g., pre-miR-21 and pre-let-7g) in comparison to their non-uridylated counterparts, highlighting that a polyU-tail directs at least in part the activity of this enzyme. In the study of Ustianenko and colleagues, they identified Dis3L2 initially as an oligoU binding protein (Ustianenko et al., 2013). They further showed Dis3l2 associates and degrades polyuridylated pre-let-7. Interestingly, in LIN28A- and LIN28B-negative HeLa cells, they found that Dis3L2 loss reduced mature let-7 levels. The potential significance of let-7 microRNAs in Perlman syndrome remains to be resolved. Also, the precise defects that leads to Perlman syndrome is unclear as one can envision that disease is caused by the toxic accumulation of defective RNA species and/or by loss of specific RNA(s) that require DIS3L2 activity for function.

FUTURE PERSPECTIVES

Although the last decade has revealed the central importance of uridylation in mammals, much remains to be discovered that has significant implications for basic biology and translational medicine. For example, the LIN28/let-7 pathway is directly implicated in numerous poor prognosis cancers where expression of either LIN28 paralog may confer resistance to ionizing radiation and several frontline chemotherapies such as doxorubicin, 5-fluorouracil, taxanes, and platinum-based drugs (reviewed in Lee et al., 2016; Balzeau et al., 2017). Moreover, cancer stem cells that are often responsible for recurrent and more deadly disease are characterized by LIN28A or LIN28B reactivation even when the tumor bulk lacks their expression. Both the LIN28/let-7 pathway and the TUT4 and TUT7 appear to be *bona fide* molecular targets for therapeutic intervention in cancer where their specific inhibition may be well-tolerated in patients (Piskounova et al., 2011; Shinoda et al., 2013b; Morgan et al., 2017). Altogether, these results suggest that LIN28 and/or TUTase inhibitors may represent a novel drug class whose development is worthwhile.

Further research is also needed to define the precise molecular mechanisms by which TUT4 promotes the development of poor prognosis breast cancers and high grade gliomas, independently of the LIN28/let-7 pathway. Ongoing research continues to unravel how mutations in DIS3L2 contribute to Perlman syndrome and Wilms tumors. Also, advances in next generation sequencing technology such as nanopore sequencing may enhance our ability to look at terminal RNA uridylation that

does not involve cumbersome sequencing library construction that is used for small-RNA-Seq or TAIL-Seq. In terms of more basic biology, uridylation is likely regulated in cell-type and disease-specific manners. Therefore, the complete repertoire of functionally significant uridylation targets remains to be elucidated. Numerous biological questions remain about RNA uridylation. What regulates TUTase expression and function? Are there other tissue specific RNA binding proteins that function like LIN28A to target specific microRNAs or other RNA substrates for uridylation-dependent degradation? How functionally significant are isomiRs generated by uridylation in development, normal physiology, and disease biology? What are the precise roles that uridylation plays in non-canonical microRNA biogenesis? Are there cis- and trans-acting factors that modulate uridylation activity on specific RNA substrates such as mRNAs with short polyA-tails? Does dysregulated uridylation contribute to pathology in other diseases and by what mechanisms? The answers to these questions as well as others should provide critical new insights into how uridylation contributes to post-transcriptional gene regulation in the emerging field of RNA epitranscriptomics.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication. MRM and JB contributed equally to this work.

FUNDING

REFERENCES

Abdelfattah, A. M., Park, C., and Choi, M. Y. (2014). Update on non-canonical microRNAs. *Biomol. Concepts* 5, 275–287. doi: 10.1515/bmc-2014-0012

Astuti, D., Morris, M. R., Cooper, W. N., Staals, R. H., Wake, N. C., Fews, G. A., et al. (2012). Germline mutations in DIS3L2 cause the Perlman syndrome of overgrowth and Wilms tumor susceptibility. *Nat. Genet.* 44, 277–284. doi: 10.1038/ng.1071

Baker, J., Hardy, M. P., Zhou, J., Bondy, C., Lupu, F., Bellvé, A. R., et al. (1996). Effects of an Igf1 gene null mutation on mouse reproduction. *Mol. Endocrinol.* 10, 903–918. doi: 10.1210/mend.10.7.8813730

Balzeau, J., Menezes, M. R., Cao, S., and Hagan, J. P. (2017). The LIN28/let-7 pathway in cancer. *Front. Genet.* 8:31 doi: 10.3389/fgene.2017.00031

Balzer, E., and Moss, E. G. (2007). Localization of the developmental timing regulator Lin28 to mRNP complexes, P-bodies and stress granules. *RNA Biol.* 4, 16–25. doi: 10.4161/rna.4.1.4364

Barbe, L., Lundberg, E., Oksvold, P., Stenius, A., Lewin, E., Björling, E., et al. (2008). Toward a confocal subcellular atlas of the human proteome. *Mol. Cell Proteomics* 7, 499–508. doi: 10.1074/mcp.M700325-MCP200

Bartel, D. P. (2009). MicroRNAs: target recognition and regulatory functions. *Cell* 136, 215–233. doi: 10.1016/j.cell.2009.01.002

Berezikov, E., Chung, W. J., Willis, J., Cuppen, E., and Lai, E. C. (2007). Mammalian mirtron genes. *Mol. Cell* 28, 328–336. doi: 10.1016/j.molcel.2007.09.028

Berndt, H., Harnisch, C., Rammelt, C., Stöhr, N., Zirkel, A., Dohm, J. C., et al. (2012). Maturation of mammalian H/ACA box snoRNAs: PAPD5-dependent adenylation and PARN-dependent trimming. *RNA* 18, 958–972. doi: 10.1261/rna.032292.112

Bernstein, E., Caudy, A. A., Hammond, S. M., and Hannon, G. J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* 409, 363–366. doi: 10.1038/35053110

Blower, P. E., Chung, J. H., Verducci, J. S., Lin, S., Park, J. K., Dai, Z., et al. (2008). MicroRNAs modulate the chemosensitivity of tumor cells. *Mol. Cancer Ther.* 7, 1–9. doi: 10.1158/1535-7163.MCT-07-0573

Boele, J., Persson, H., Shin, J. W., Ishizu, Y., Newie, I. S., Sokilde, R., et al. (2014). PAPD5-mediated 3′ adenylation and subsequent degradation of miR-21 is disrupted in proliferative disease. *Proc. Natl. Acad. Sci. U.S.A.* 111, 11467–11472. doi: 10.1073/pnas.1317751111

Bohnsack, M. T., Czaplinski, K., and Gorlich, D. (2004). Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. *RNA* 10, 185–191. doi: 10.1261/rna.5167604

Borchert, G. M., Lanier, W., and Davidson, B. L. (2006). RNA polymerase III transcribes human microRNAs. *Nat. Struct. Mol. Biol.* 13, 1097–1101. doi: 10.1038/nsmb1167

Bortolamiol-Becet, D., Hu, F., Jee, D., Wen, J., Okamura, K., Lin, C. J., et al. (2015). Selective suppression of the splicing-mediated microRNA

pathway by the terminal uridyltransferase tailor. *Mol. Cell* 59, 217–228. doi: 10.1016/j.molcel.2015.05.034

Boyerinas, B., Park, S. M., Murmann, A. E., Gwin, K., Montag, A. G., Zillhardt, M., et al. (2012). Let-7 modulates acquired resistance of ovarian cancer to Taxanes via IMP-1-mediated stabilization of multidrug resistance 1. *Int. J. Cancer* 130, 1787–1797. doi: 10.1002/ijc.26190

Brooks, L., Lyons, S. M., Mahoney, J. M., Welch, J. D., Liu, Z., Marzluff, W. F., et al. (2015). A multiprotein occupancy map of the mRNP on the 3′ end of histone mRNAs. *RNA* 21, 1943–1965. doi: 10.1261/rna.053389.115

Buganim, Y., Faddah, D. A., Cheng, A. W., Itskovich, E., Markoulaki, S., Ganz, K., et al. (2012). Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. *Cell* 150, 1209–1222. doi: 10.1016/j.cell.2012.08.023

Burns, D. M., D'Ambrogio, A., Nottrott, S., and Richter, J. D. (2011). CPEB and two poly(A) polymerases control miR-122 stability and p53 mRNA translation. *Nature* 473, 105–108. doi: 10.1038/nature09908

Bussing, I., Slack, F. J., and Grosshans, H. (2008). let-7 microRNAs in development, stem cells and cancer. *Trends Mol. Med.* 14, 400–409. doi: 10.1016/j.molmed.2008.07.001

Cai, X., Hagedorn, C. H., and Cullen, B. R. (2004). Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *RNA* 10, 1957–1966. doi: 10.1261/rna.7135204

Cannon-Albright, L. A., Teerlink, C. C., Farnham, J. M., Thomas, A. W., Zone, J. J., and Leachman, S. A. (2013). Linkage analysis of extended high-risk pedigrees replicates a cutaneous malignant melanoma predisposition locus on chromosome 9q21. *J. Invest. Dermatol.* 133, 128–134. doi: 10.1038/jid.2012.271

Chang, H., Lim, J., Ha, M., and Kim, V. N. (2014). TAIL-seq: genome-wide determination of poly(A) tail length and 3′ end modifications. *Mol. Cell* 53, 1044–1052. doi: 10.1016/j.molcel.2014.02.007

Chang, H., Yeo, J., Kim, J. G., Kim, H., Lim, J., Lee, M., et al. (2018). Terminal uridylyltransferases execute programmed clearance of maternal transcriptome in vertebrate embryos. *Mol. Cell* 70, 72–82 e7. doi: 10.1016/j.molcel.2018.03.004

Chang, H. M., Triboulet, R., Thornton, J. E., and Gregory, R. I. (2013). A role for the Perlman syndrome exonuclease Dis3l2 in the Lin28-let-7 pathway. *Nature* 497, 244–248. doi: 10.1038/nature12119

Cheloufi, S., Dos Santos, C. O., Chong, M. M., and Hannon, G. J. (2010). A dicer-independent miRNA biogenesis pathway that requires Ago catalysis. *Nature* 465, 584–589. doi: 10.1038/nature09092

Chen, G. Q., Zhao, Z. W., Zhou, H. Y., Liu, Y. J., and Yang, H. J. (2009). Systematic analysis of microRNA involved in resistance of the MCF-7 human breast cancer cell to doxorubicin. *Med. Oncol.* 27, 406–415. doi: 10.1007/s12032-009-9225-9

Chendrimada, T. P., Gregory, R. I., Kumaraswamy, E., Norman, J., Cooch, N., Nishikura, K., et al. (2005). TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. *Nature* 436, 740–744. doi: 10.1038/nature03868

Cho, J., Chang, H., Kwon, S. C., Kim, B., Kim, Y., Choe, J., et al. (2012). LIN28A is a suppressor of ER-associated translation in embryonic stem cells. *Cell* 151, 765–777. doi: 10.1016/j.cell.2012.10.019

Chung, C. Z., Jo, D. H., and Heinemann, I. U. (2016). Nucleotide specificity of the human terminal nucleotidyltransferase Gld2 (TUT2). *RNA* 22, 1239–1249. doi: 10.1261/rna.056077.116

Cifuentes, D., Xue, H., Taylor, D. W., Patnode, H., Mishima, Y., Cheloufi, S., et al. (2010). A novel miRNA processing pathway independent of Dicer requires Argonaute2 catalytic activity. *Science* 328, 1694–1698. doi: 10.1126/science.1190809

Cloonan, N., Wani, S., Xu, Q., Gu, J., Lea, K., Heater, S., et al. (2011). MicroRNAs and their isomiRs function cooperatively to target common biological pathways. *Genome Biol.* 12:R126. doi: 10.1186/gb-2011-12-12-r126

Crosby, A. H., Patel, H., Chioza, B. A., Proukakis, C., Gurtz, K., Patton, M. A., et al. (2010). Defective mitochondrial mRNA maturation is associated with spastic ataxia. *Am. J. Hum. Genet.* 87, 655–660. doi: 10.1016/j.ajhg.2010.09.013

D'Ambrogio, A., Gu, W., Udagawa, T., Mello, C. C., and Richter, J. D. (2012). Specific miRNA stabilization by Gld2-catalyzed monoadenylation. *Cell Rep.* 2, 1537–1545. doi: 10.1016/j.celrep.2012.10.023

Denli, A. M., Tops, B. B., Plasterk, R. H., Ketting, R. F., and Hannon, G. J. (2004). Processing of primary microRNAs by the microprocessor complex. *Nature* 432, 231–235. doi: 10.1038/nature03049

de Oliveira, L. F., Christoff, A. P., and Margis, R. (2013). isomiRID: a framework to identify microRNA isoforms. *Bioinformatics* 29, 2521–2523. doi: 10.1093/bioinformatics/btt424

Dore, L. C., Amigo, J. D., Dos Santos, C. O., Zhang, Z., Gai, X., Tobias, J. W., et al. (2008). A GATA-1-regulated microRNA locus essential for erythropoiesis. *Proc. Natl. Acad. Sci. U.S.A.* 105, 3333–3338. doi: 10.1073/pnas.0712312105

Esquela-Kerscher, A., Trang, P., Wiggins, J. F., Patrawala, L., Cheng, A., Ford, L., et al. (2008). The let-7 microRNA reduces tumor growth in mouse models of lung cancer. *Cell Cycle* 7, 759–764. doi: 10.4161/cc.7.6.5834

Faehnle, C. R., Walleshauser, J., and Joshua-Tor, L. (2014). Mechanism of Dis3l2 substrate recognition in the Lin28-let-7 pathway. *Nature* 514, 252–256. doi: 10.1038/nature13553

Faehnle, C. R., Walleshauser, J., and Joshua-Tor, L. (2017). Multi-domain utilization by TUT4 and TUT7 in control of let-7 biogenesis. *Nat. Struct. Mol. Biol.* 24, 658–665. doi: 10.1038/nsmb.3428

Fu, X., Meng, Z., Liang, W., Tian, Y., Wang, X., Han, W., et al. (2013). miR-26a enhances miRNA biogenesis by targeting Lin28B and Zcchc11 to suppress tumor growth and metastasis. *Oncogene* 33, 4296–306. doi: 10.1038/onc.2013.385

Gerth, A. M. J., Deighton, R. F., McCulloch, J., and White, I. R. (2013). "Proteomic studies in low-grade gliomas: what have they informed about pathophysiology?," in *Diffuse Low-Grade Gliomas in Adults*, ed H. Duffau (London: Springer), 117–136

Gong, J., Tong, Y., Zhang, H. M., Wang, K., Hu, T., Shan, G., et al. (2012). Genome-wide identification of SNPs in microRNA genes and the SNP effects on microRNA target binding and biogenesis. *Hum. Mutat.* 33, 254–263. doi: 10.1002/humu.21641

Gonzales, M. L., Mellman, D. L., and Anderson, R. A. (2008). CKIalpha is associated with and phosphorylates star-PAP and is also required for expression of select star-PAP target messenger RNAs. *J. Biol. Chem.* 283, 12665–12673. doi: 10.1074/jbc.M800656200

Gregory, R. I., Yan, K. P., Amuthan, G., Chendrimada, T., Doratotaj, B., Cooch, N., et al. (2004). The microprocessor complex mediates the genesis of microRNAs. *Nature* 432, 235–240. doi: 10.1038/nature03120

Griffiths-Jones, S., Grocock, R. J., van Dongen, S., Bateman, A., and Enright, A. J. (2006). miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.* 34, D140–D144. doi: 10.1093/nar/gkj112

Guo, L., Yu, J., Liang, T., and Zou, Q. (2016). miR-isomiRExp: a web-server for the analysis of expression of miRNA at the miRNA/isomiR levels. *Sci. Rep.* 6:23700 doi: 10.1038/srep23700

Guo, Y., Chen, Y., Ito, H., Watanabe, A., Ge, X., Kodama, T., et al. (2006). Identification and characterization of lin-28 homolog B (LIN28B) in human hepatocellular carcinoma. *Gene* 384, 51–61. doi: 10.1016/j.gene.2006.07.011

Gutierrez-Vazquez, C., Enright, A. J., Rodriguez-Galan, A., Perez-Garcia, A., Collier, P., Jones, M. R., et al. (2017). 3′ Uridylation controls mature microRNA turnover during CD4 T-cell activation. *RNA* 23, 882–891. doi: 10.1261/rna.060095.116

Gyarfas, B., Olasagasti, F., Benner, S., Garalde, D., Lieberman, K. R., and Akeson, M. (2009). Mapping the position of DNA polymerase-bound DNA templates in a nanopore at 5 A resolution. *ACS Nano* 3, 1457–1466. doi: 10.1021/nn900303g

Haas, G., Cetin, S., Messmer, M., Chane-Woon-Ming, B., Terenzi, O., Chicher, J., et al. (2016). Identification of factors involved in target RNA-directed microRNA degradation. *Nucleic Acids Res.* 44, 2873–2887. doi: 10.1093/nar/gkw040

Haase, A. D., Jaskiewicz, L., Zhang, H., Laine, S., Sack, R., Gatignol, A., et al. (2005). TRBP, a regulator of cellular PKR and HIV-1 virus expression, interacts with Dicer and functions in RNA silencing. *EMBO Rep.* 6, 961–967. doi: 10.1038/sj.embor.7400509

Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., et al. (2010). Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* 141, 129–41. doi: 10.1016/j.cell.2010.03.009

Hagan, J. P., and Croce, C. M. (2007). MicroRNAs in carcinogenesis. *Cytogenet. Genome Res.* 118, 252–259. doi: 10.1159/000108308

Hagan, J. P., Piskounova, E., and Gregory, R. I. (2009). Lin28 recruits the TUTase Zcchc11 to inhibit let-7 maturation in mouse embryonic stem cells. *Nat. Struct. Mol. Biol.* 16, 1021–1025. doi: 10.1038/nsmb.1676

Hallett, R. M., and Hassell, J. A. (2011). E2F1 and KIAA0191 expression predicts breast cancer patient survival. *BMC Res. Notes* 4:95. doi: 10.1186/1756-0500-4-95

Han, J., Lee, Y., Yeom, K. H., Kim, Y. K., Jin, H., and Kim, V. N. (2004). The Drosha-DGCR8 complex in primary microRNA processing. *Genes Dev.* 18, 3016–3027. doi: 10.1101/gad.1262504

Heo, I., Ha, M., Lim, J., Yoon, M. J., Park, J. E., Kwon, S. C., et al. N. (2012). Mono-uridylation of pre-microRNA as a key step in the biogenesis of group II let-7 microRNAs. *Cell* 151, 521–532. doi: 10.1016/j.cell.2012.09.022

Heo, I., Joo, C., Cho, J., Ha, M., Han, J., and Kim, V. N. (2008). Lin28 mediates the terminal uridylation of let-7 precursor MicroRNA. *Mol. Cell* 32, 276–84. doi: 10.1016/j.molcel.2008.09.014

Heo, I., Joo, C., Kim, Y. K., Ha, M., Yoon, M. J., Cho, J., et al. (2009). TUT4 in concert with Lin28 suppresses microRNA biogenesis through pre-microRNA uridylation. *Cell* 138, 696–708. doi: 10.1016/j.cell.2009.08.002

Higashimoto, K., Maeda, T., Okada, J., Ohtsuka, Y., Sasaki, K., Hirose, A., et al. (2013). Homozygous deletion of DIS3L2 exon 9 due to non-allelic homologous recombination between LINE-1s in a Japanese patient with Perlman syndrome. *Eur. J. Hum. Genet.* 21, 1316–1319. doi: 10.1038/ejhg.2013.45

Hoefig, K. P., and Heissmeyer, V. (2014). Degradation of oligouridylated histone mRNAs: see UUUUU and goodbye. *Wiley Interdisc. Rev. RNA* 5, 577–589. doi: 10.1002/wrna.1232

Hoefig, K. P., Rath, N., Heinz, G. A., Wolf, C., Dameris, J., Schepers, A., et al. (2013). Eri1 degrades the stem-loop of oligouridylated histone mRNAs to induce replication-dependent decay. *Nat. Struct. Mol. Biol.* 20, 73–81. doi: 10.1038/nsmb.2450

Hutvagner, G., McLachlan, J., Pasquinelli, A. E., Balint, E., Tuschl, T., and Zamore, P. D. (2001). A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science* 293, 834–838. doi: 10.1126/science.1062961

Ishikawa, H., Nobe, Y., Izumikawa, K., Taoka, M., Yamauchi, Y., Nakayama, H., et al. (2018). Truncated forms of U2 snRNA (U2-tfs) are shunted toward a novel uridylylation pathway that differs from the degradation pathway for U1-tfs. *RNA Biol.* 15, 261–268. doi: 10.1080/15476286.2017.1408766

Jones, M. R., Blahna, M. T., Kozlowski, E., Matsuura, K. Y., Ferrari, J. D., Morris, S. A., et al. (2012). Zcchc11 uridylates mature miRNAs to enhance neonatal IGF-1 expression, growth, and survival. *PLoS Genet.* 8:e1003105. doi: 10.1371/journal.pgen.1003105

Jones, M. R., Quinton, L. J., Blahna, M. T., Neilson, J. R., Fu, S., Ivanov, A. R., et al. (2009). Zcchc11-dependent uridylation of microRNA directs cytokine expression. *Nat. Cell Biol.* 11, 1157–1163. doi: 10.1038/ncb1931

Kandala, D. T., Mohan, N., A, V., A, P. S., G, R., and Laishram, R. S. (2016). CstF-64 and 3′-UTR cis-element determine Star-PAP specificity for target mRNA selection by excluding

PAPalpha. *Nucleic Acids Res.* 44, 811–823. doi: 10.1093/nar/gkv 1074

Katoh, T., Sakaguchi, Y., Miyauchi, K., Suzuki, T., Kashiwabara, S., Baba, T., et al. (2009). Selective stabilization of mammalian microRNAs by 3′ adenylation mediated by the cytoplasmic poly(A) polymerase GLD-2. *Genes Dev.* 23, 433–438. doi: 10.1101/gad.1761509

Ketting, R. F., Fischer, S. E., Bernstein, E., Sijen, T., Hannon, G. J., and Plasterk, R. H. (2001). Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in *C. elegans*. *Genes Dev.* 15, 2654–2659. doi: 10.1101/gad.927801

Kim, B., Ha, M., Loeff, L., Chang, H., Simanshu, D. K., Li, S., et al. (2015). TUT7 controls the fate of precursor microRNAs by using three different uridylation mechanisms. *EMBO J.* 34, 1801–1815. doi: 10.15252/embj.201590931

Kozlowski, E., Wasserman, G. A., Morgan, M., O'Carroll, D., Ramirez, N.P., et al. (2017). The RNA uridyltransferase Zcchc6 is expressed in macrophages and impacts innate immune responses. *PLoS ONE* 12:e0179797. doi: 10.1371/journal.pone.0179797

Kunkel, G. R., Maser, R. L., Calvet, J. P., and Pederson, T. (1986). U6 small nuclear RNA is transcribed by RNA polymerase III. *Proc. Natl. Acad. Sci. U.S.A.* 83, 8575–8579.

Labno, A., Warkocki, Z., Kulinski, T., Krawczyk, P. S., Bijata, K., Tomecki, R., et al. (2016). Perlman syndrome nuclease DIS3L2 controls cytoplasmic non-coding RNAs and provides surveillance pathway for maturing snRNAs. *Nucleic Acids Res.* 44, 10437–10453. doi: 10.1093/nar/gkw649

Lackey, P. E., Welch, J. D., and Marzluff, W. F. (2016). TUT7 catalyzes the uridylation of the 3′ end for rapid degradation of histone mRNA. *RNA* 22, 1673–1688. doi: 10.1261/rna.058107.116

Ladewig, E., Okamura, K., Flynt, A. S., Westholm, J. O., and Lai, E. C. (2012). Discovery of hundreds of mirtrons in mouse and human small RNA data. *Genome Res.* 22, 1634–1645. doi: 10.1101/gr.133553.111

Lee, H., Han, S., Kwon, C. S., and Lee, D. (2016). Biogenesis and regulation of the let-7 miRNAs and their functional implications. *Protein Cell* 7, 100–113. doi: 10.1007/s13238-015-0212-y

Lee, H. Y., Zhou, K., Smith, A. M., Noland, C. L., and Doudna, J. A. (2013). Differential roles of human Dicer-binding proteins TRBP and PACT in small RNA processing. *Nucleic Acids Res.* 41, 6568–6576. doi: 10.1093/nar/gkt361

Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., et al. (2003). The nuclear RNase III Drosha initiates microRNA processing. *Nature* 425, 415–419. doi: 10.1038/nature01957

Lee, Y., Kim, M., Han, J., Yeom, K. H., Lee, S., Baek, S. H., et al. (2004). MicroRNA genes are transcribed by RNA polymerase II. *EMBO J.* 23, 4051–4060. doi: 10.1038/sj.emboj.7600385

Lehrbach, N. J., Armisen, J., Lightfoot, H. L., Murfitt, K. J., Bugaut, A., Balasubramanian, S., et al. (2009). LIN-28 and the poly(U) polymerase PUP-2 regulate let-7 microRNA processing in *Caenorhabditis elegans*. *Nat. Struct. Mol. Biol.* 16, 1016–1020. doi: 10.1038/nsmb.1675

Lewis, B. P., Burge, C. B., and Bartel, D. P. (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* 120, 15–20. doi: 10.1016/j.cell.2004.12.035

Li, W., Laishram, R. S., Ji, Z., Barlow, C. A., Tian, B., and Anderson, R. A. (2012). Star-PAP control of BIK expression and apoptosis is regulated by nuclear PIPKIalpha and PKCdelta signaling. *Mol. Cell* 45, 25–37. doi: 10.1016/j.molcel.2011.11.017

Li, W., Li, W., Laishram, R. S., Hoque, M., Ji, Z., Tian, B., et al. (2017). Distinct regulation of alternative polyadenylation and gene expression by nuclear poly(A) polymerases. *Nucleic Acids Res.* 45, 8930–8942. doi: 10.1093/nar/gkx560

Licht, K., Medenbach, J., Luhrmann, R., Kambach, C., and Bindereif, A. (2008). 3′-cyclic phosphorylation of U6 snRNA leads to recruitment of recycling factor p110 through LSm proteins. *RNA* 14, 1532–1538. doi: 10.1261/rna.1129608

Lim, J., Ha, M., Chang, H., Kwon, S. C., Simanshu, D. K., Patel, D. J., et al. (2014). Uridylation by TUT4 and TUT7 marks mRNA for degradation. *Cell* 159, 1365–1376. doi: 10.1016/j.cell.2014.10.055

Lim, L. P., Lau, N. C., Garrett-Engele, P., Grimson, A., Schelter, J. M., Castle, J., et al. (2005). Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* 433, 769–773. doi: 10.1038/nature03315

Lin, S., and Gregory, R. I. (2015a). Identification of small molecule inhibitors of Zcchc11 TUTase activity. *RNA Biol.* 12, 792–800. doi: 10.1080/15476286.2015.1058478

Lin, S., and Gregory, R. I. (2015b). MicroRNA biogenesis pathways in cancer. *Nat. Rev. Cancer* 15, 321–333. doi: 10.1038/nrc3932

Lubas, M., Christensen, M. S., Kristiansen, M. S., Domanski, M., Falkenby, L. G., Lykke-Andersen, S., et al. (2011). Interaction profiling identifies the human nuclear exosome targeting complex. *Mol. Cell* 43, 624–637. doi: 10.1016/j.molcel.2011.06.028

Lyons, S. M., Ricciardi, A. S., Guo, A. Y., Kambach, C., and Marzluff, W. F. (2014). The C-terminal extension of Lsm4 interacts directly with the 3′ end of the histone mRNP and is required for efficient histone mRNA degradation. *RNA* 20, 88–102. doi: 10.1261/rna.042531.113

Ma, H., Wu, Y., Choi, J. G., and Wu, H. (2013). Lower and upper stem-single-stranded RNA junctions together determine the Drosha cleavage site. *Proc. Natl. Acad. Sci. U.S.A.* 110, 20687–20692. doi: 10.1073/pnas.1311639110

Marzluff, W. F., and Koreski, K. P. (2017). Birth and death of histone mRNAs. *Trends Genet.* 33, 745–759. doi: 10.1016/j.tig.2017.07.014

McCall, M. N., Kim, M. S., Adil, M., Patil, A. H., Lu, Y., Mitchell, C. J., et al. (2017). Toward the human cellular microRNAome. *Genome Res.* 27, 1769–1781. doi: 10.1101/gr.222067.117

Mellman, D. L., Gonzales, M. L., Song, C., Barlow, C. A., Wang, P., Kendziorski, C., et al. (2008). A PtdIns4,5P2-regulated nuclear poly(A) polymerase controls expression of select mRNAs. *Nature* 451, 1013–1017. doi: 10.1038/nature06666

Mello, C. C., and Conte, D. Jr. (2004). Revealing the world of RNA interference. *Nature* 431, 338–342. doi: 10.1038/nature02872

Minoda, Y., Saeki, K., Aki, D., Takaki, H., Sanada, T., Koga, K., et al. (2006). A novel Zinc finger protein, ZCCHC11, interacts with TIFA and modulates TLR signaling. *Biochem. Biophys. Res. Commun.* 344, 1023–1030. doi: 10.1016/j.bbrc.2006.04.006

Miyoshi, K., Miyoshi, T., and Siomi, H. (2010). Many ways to generate microRNA-like small RNAs: non-canonical pathways for microRNA production. *Mol. Genet. Genomics* 284, 95–103. doi: 10.1007/s00438-010-0556-1

Molenaar, J. J., Domingo-Fernandez, R., Ebus, M. E., Lindner, S., Koster, J., Drabek, K., et al. (2012). LIN28B induces neuroblastoma and enhances MYCN levels via let-7 suppression. *Nat. Genet.* 44, 1199–1206. doi: 10.1038/ng.2436

Morgan, M., Much, C., DiGiacomo, M., Azzi, C., Ivanova, I., Vitsios, D. M., et al. (2017). mRNA 3′ uridylation and poly(A) tail length sculpt the mammalian maternal transcriptome. *Nature* 548, 347–351. doi: 10.1038/nature23318

Morin, R. D., O'Connor, M. D., Griffith, M., Kuchenbauer, F., Delaney, A., Prabhu, A., et al. (2008). Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. *Genome Res.* 18, 610–621. doi: 10.1101/gr.7179508

Morris, M. R., Astuti, D., and Maher, E. R. (2013). Perlman syndrome: overgrowth, wilms tumor predisposition and DIS3L2. *Am. J. Med. Genet. C Semin. Med. Genet.* 163C, 106–113. doi: 10.1002/j.1552-4876.2013.31358.x

Mroczek, S., Krwawicz, J., Kutner, J., Lazniewski, M., Kucinski, I., Ginalski, K., et al. (2012). C16orf57, a gene mutated in poikiloderma with neutropenia, encodes a putative phosphodiesterase responsible for the U6 snRNA 3′ end modification. *Genes Dev.* 26, 1911–1925. doi: 10.1101/gad.193169.112

Mullen, T. E., and Marzluff, W. F. (2008). Degradation of histone mRNA requires oligouridylation followed by decapping and simultaneous degradation of the mRNA both 5′ to 3′ and 3′ to 5′. *Genes Dev.* 22, 50–65. doi: 10.1101/gad.16 22708

Muller, H., Marzi, M. J., and Nicassio, F. (2014). IsomiRage: from functional classification to differential expression of miRNA isoforms. *Front. Bioeng. Biotechnol.* 2:38. doi: 10.3389/fbioe.2014.00038

Munoz-Tello, P., Gabus, C., and Thore, S. (2012). Functional implications from the Cid1 poly(U) polymerase crystal structure. *Structure* 20, 977–986. doi: 10.1016/j.str.2012.04.006

Nagaike, T., Suzuki, T., Katoh, T., and Ueda, T. (2005). Human mitochondrial mRNAs are stabilized with polyadenylation regulated by mitochondria-specific poly(A) polymerase and polynucleotide phosphorylase. *J. Biol. Chem.* 280, 19721–19727. doi: 10.1074/jbc.M500804200

Nair, V. S., Maeda, L. S., and Ioannidis, J. P. (2012). Clinical outcome prediction by microRNAs in human cancer: a systematic review. *J. Natl. Cancer Inst.* 104, 528–540. doi: 10.1093/jnci/djs027

Nakanishi, T., Kubota, H., Ishibashi, N., Kumagai, S., Watanabe, H., Yamashita, M., et al. (2006). Possible role of mouse poly(A) polymerase mGLD-2 during oocyte maturation. *Dev. Biol.* 289, 115–126. doi: 10.1016/j.ydbio.2005.10.017

Nam, Y., Chen, C., Gregory, R. I., Chou, J. J., and Sliz, P. (2011). Molecular basis for interaction of let-7 microRNAs with Lin28. *Cell* 147, 1080–1091. doi: 10.1016/j.cell.2011.10.020

Neilsen, C. T., Goodall, G. J., and Bracken, C. P. (2012). IsomiRs–the overlooked repertoire in the dynamic microRNAome. *Trends Genet.* 28, 544–549. doi: 10.1016/j.tig.2012.07.005

Neri, G., Martini-Neri, M. E., Katz, B. E., and Opitz, J. M. (1984). The Perlman syndrome: familial renal dysplasia with Wilms tumor, fetal gigantism and multiple congenital anomalies. *Am. J. Med. Genet.* 19, 195–207. doi: 10.1002/ajmg.1320190120

Neri, G., Martini-Neri, M. E., Opitz, J. M., and Freed, J. J. (1985). The Perlman syndrome: clinical and biological aspects. *Prog. Clin. Biol. Res.* 200, 269–276.

Newman, M. A., Thomson, J. M., and Hammond, S. M. (2008). Lin-28 interaction with the Let-7 precursor loop mediates regulated microRNA processing. *RNA* 14, 1539–1549. doi: 10.1261/rna.1155108

Nguyen, L. H., Robinton, D. A., Seligson, M. T., Wu, L., Li, L., Rakheja, D., et al. (2014). Lin28b is sufficient to drive liver cancer and necessary for its maintenance in murine models. *Cancer Cell* 26, 248–261. doi: 10.1016/j.ccr.2014.06.018

Ogami, K., Cho, R., and Hoshino, S. (2013). Molecular cloning and characterization of a novel isoform of the non-canonical poly(A) polymerase PAPD7. *Biochem. Biophys. Res. Commun.* 432, 135–140. doi: 10.1016/j.bbrc.2013.01.072

Oh, J. S., Kim, J. J., Byun, J. Y., and Kim, I. A. (2010). Lin28-let7 modulates radiosensitivity of human cancer cells with activation of K-Ras. *Int. J. Radiat. Oncol. Biol. Phys.* 76, 5–8. doi: 10.1016/j.ijrobp.2009.08.028

Okamura, K., Hagen, J. W., Duan, H., Tyler, D. M., and Lai, E. C. (2007). The mirtron pathway generates microRNA-class regulatory RNAs in Drosophila. *Cell* 130, 89–100. doi: 10.1016/j.cell.2007.06.028

Ouchi, Y., Yamamoto, J., and Iwamoto, T. (2014). The heterochronic genes lin-28a and lin-28b play an essential and evolutionarily conserved role in early zebrafish development. *PLoS ONE* 9:e88086. doi: 10.1371/journal.pone.0088086

Pantano, L., Estivill, X., and Marti, E. (2010). SeqBuster, a bioinformatic tool for the processing and analysis of small RNAs datasets, reveals ubiquitous miRNA modifications in human embryonic cells. *Nucleic Acids Res.* 38:e34. doi: 10.1093/nar/gkp1127

Park, J. E., Heo, I., Tian, Y., Simanshu, D. K., Chang, H., Jee, D., et al. and Kim, V., N. (2011). Dicer recognizes the 5' end of RNA for efficient and accurate processing. *Nature* 475, 201–205. doi: 10.1038/nature10198

Pase, L., Layton, J. E., Kloosterman, W. P., Carradice, D., Waterhouse, P. M., and Lieschke, G. J. (2009). miR-451 regulates zebrafish erythroid maturation *in vivo* via its target gata2. *Blood* 113, 1794–1804. doi: 10.1182/blood-2008-05-155812

Pashler, A. L., Towler, B. P., Jones, C. I., and Newbury, S. F. (2016). The roles of the exoribonucleases DIS3L2 and XRN1 in human disease. *Biochem. Soc. Trans.* 44, 1377–1384. doi: 10.1042/BST20160107

Patrick, D. M., Zhang, C. C., Tao, Y., Yao, H., Qi, X., Schwartz, R. J., et al. (2010). Defective erythroid differentiation in miR-451 mutant mice mediated by 14-3-3zeta. *Genes Dev.* 24, 1614–1619. doi: 10.1101/gad.1942810

Peter, M. E. (2009). Let-7 and miR-200 microRNAs: guardians against pluripotency and cancer progression. *Cell Cycle* 8, 843–852. doi: 10.4161/cc.8.6.7907

Pirouz, M., Du, P., Munafo, M., and Gregory, R. I. (2016). Dis3l2-mediated decay is a quality control pathway for noncoding RNAs. *Cell Rep.* 16, 1861–1873. doi: 10.1016/j.celrep.2016.07.025

Piskounova, E., Polytarchou, C., Thornton, J. E., LaPierre, R. J., Pothoulakis, C., Hagan, J. P., et al. (2011). Lin28A and Lin28B inhibit let-7 microRNA biogenesis by distinct mechanisms. *Cell* 147, 1066–1079. doi: 10.1016/j.cell.2011.10.039

Piskounova, E., Viswanathan, S. R., Janas, M., LaPierre, R. J., Daley, G. Q., Sliz, P., et al. (2008). Determinants of microRNA processing inhibition by the developmentally regulated RNA-binding protein Lin28. *J. Biol. Chem.* 283, 21310–21314. doi: 10.1074/jbc.C800108200

Polesskaya, A., Cuvellier, S., Naguibneva, I., Duquet, A., Moss, E. G., and Harel-Bellan, A. (2007). Lin-28 binds IGF-2 mRNA and participates in skeletal myogenesis by increasing translation efficiency. *Genes Dev.* 21, 1125–1138. doi: 10.1101/gad.415007

Rammelt, C., Bilen, B., Zavolan, M., and Keller, W. (2011). PAPD5, a noncanonical poly(A) polymerase with an unusual RNA-binding motif. *RNA* 17, 1737–1746. doi: 10.1261/rna.2787011

Rasmussen, K. D., Simmini, S., Abreu-Goodger, C., Bartonicek, N., Di Giacomo, M., Bilbao-Cortes, D., et al. (2010). The miR-144/451 locus is required for erythroid homeostasis. *J. Exp. Med.* 207, 1351–1358. doi: 10.1084/jem.20100458

Rau, F., Freyermuth, F., Fugier, C., Villemin, J. P., Fischer, M. C., Jost, B., et al. (2011). Misregulation of miR-1 processing is associated with heart defects in myotonic dystrophy. *Nat. Struct. Mol. Biol.* 18, 840–845. doi: 10.1038/nsmb.2067

Reimao-Pinto, M. M., Ignatova, V., Burkard, T. R., Hung, J. H., Manzenreither, R. A., Sowemimo, I., et al. (2015). Uridylation of RNA hairpins by tailor confines the emergence of microRNAs in drosophila. *Mol. Cell* 59, 203–216. doi: 10.1016/j.molcel.2015.05.033

Rissland, O. S. (2015). Tailoring microRNA function: the role of uridylation in antagonizing mirtron expression. *Mol. Cell* 59, 141–143. doi: 10.1016/j.molcel.2015.07.001

Rissland, O. S., Mikulasova, A., and Norbury, C. J. (2007). Efficient RNA polyuridylation by noncanonical poly(A) polymerases. *Mol. Cell Biol.* 27, 3612–3624. doi: 10.1128/MCB.02209-06

Rissland, O. S., and Norbury C. J. (2009). Decapping is preceded by 3' uridylation in a novel pathway of bulk mRNA turnover. *Nat Struct Mol Biol*, 16, 616–623. doi: 10.1038/nsmb.1601

Ruby, J. G., Jan, C. H., and Bartel, D. P. (2007). Intronic microRNA precursors that bypass Drosha processing. *Nature* 448, 83–86. doi: 10.1038/nature05983

Rybak, A., Fuchs, H., Smirnova, L., Brandt, C., Pohl, E. E., Nitsch, R., et al. (2008). A feedback loop comprising lin-28 and let-7 controls pre-let-7 maturation during neural stem-cell commitment. *Nat. Cell Biol.* 10, 987–993. doi: 10.1038/ncb1759

Sablok, G., Milev, I., Minkov, G., Minkov, I., Varotto, C., Yahubyan, G., et al. (2013). isomiRex: web-based identification of microRNAs, isomiR variations and differential expression using next-generation sequencing datasets. *FEBS Lett.* 587, 2629–2634. doi: 10.1016/j.febslet.2013.06.047

Schmidt, M. J., West, S., and Norbury, C. J. (2011). The human cytoplasmic RNA terminal U-transferase ZCCHC11 targets histone mRNAs for degradation. *RNA* 17, 39–44. doi: 10.1261/rna.2252511

Shen, B., and Goodman, H. M. (2004). Uridine addition after microRNA-directed cleavage. *Science* 306:997 doi: 10.1126/science.1103521

Shinoda, G., De Soysa, T. Y., Seligson, M. T., Yabuuchi, A., Fujiwara, Y., Huang, P. Y., et al. Q. (2013a). Lin28a regulates germ cell pool size and fertility. *Stem Cells* 31, 1001–1009. doi: 10.1002/stem.1343

Shinoda, G., Shyh-Chang, N., Soysa, T. Y., Zhu, H., Seligson, M. T., Shah, S. P., et al. Q. (2013b). Fetal deficiency of lin28 programs life-long aberrations in growth and glucose metabolism. *Stem Cells* 31, 1563–1573. doi: 10.1002/stem.1423

Siolas, D., Lerner, C., Burchard, J., Ge, W., Linsley, P. S., Paddison, P. J., et al. (2005). Synthetic shRNAs as potent RNAi triggers. *Nat. Biotechnol.* 23, 227–231. doi: 10.1038/nbt1052

Starega-Roslan, J., Witkos, T. M., Galka-Marciniak, P., and Krzyzosiak, W. J. (2015). Sequence features of Drosha and Dicer cleavage sites affect the complexity of isomiRs. *Int. J. Mol. Sci.* 16, 8110–8127. doi: 10.3390/ijms16048110

Su, W., Slepenkov, S. V., Slevin, M. K., Lyons, S. M., Ziemniak, M., Kowalska, J., et al. (2013). mRNAs containing the histone 3' stem-loop are degraded primarily by decapping mediated by oligouridylation of the 3' end. *RNA* 19, 1–16. doi: 10.1261/rna.034470.112

Sudheesh, A. P., and Laishram, R. S. (2017). Nuclear phosphatidyl-inositol-phosphate type I kinase alpha coupled Star-PAP polyadenylation regulates cell invasion. *Mol. Cell Biol.* 38:e00457-17. doi: 10.1128/MCB.00457-17

Suzuki, H. I., Katsura, A., and Miyazono, K. (2015). A role of uridylation pathway for blockade of let-7 microRNA biogenesis by Lin28. *Cancer Sci.* 106, 1174–1181. doi: 10.1111/cas.12721

Tan, G. C., Chan, E., Molnar, A., Sarkar, R., Alexieva, D., Isa, I. M., et al. (2014). 5' isomiR variation is of functional and evolutionary importance. *Nucleic Acids Res.* 42, 9424–9435. doi: 10.1093/nar/gku656

Thomas, M. P., Liu, X., Whangbo, J., McCrossan, G., Sanborn, K. B., Basar, E., et al. (2015). Apoptosis triggers specific, rapid, and global mRNA decay with 3' uridylated intermediates degraded by DIS3L2. *Cell Rep.* 11, 1079–1089. doi: 10.1016/j.celrep.2015.04.026

Thornton, J. E., Chang, H. M., Piskounova, E., and Gregory, R., I. (2012). Lin28-mediated control of let-7 microRNA expression by alternative TUTases Zcchc11 (TUT4) and Zcchc6 (TUT7). *RNA* 18, 1875–1885. doi: 10.1261/rna.034538.112

Thornton, J. E., Du, P., Jing, L., Sjekloca, L., Lin, S., Grossi, E., et al. (2014). Selective microRNA uridylation by Zcchc6 (TUT7) and Zcchc11 (TUT4). *Nucleic Acids Res*. 42, 11777–11791. doi: 10.1093/nar/gku805

Tomecki, R., Dmochowska, A., Gewartowski, K., Dziembowski, A., and Stepien, P. P. (2004). Identification of a novel human nuclear-encoded mitochondrial poly(A) polymerase. *Nucleic Acids Res*. 32, 6001–6014. doi: 10.1093/nar/gkh923

Trang, P., Medina, P. P., Wiggins, J. F., Ruffino, L., Kelnar, K., Omotola, M., et al. (2010). Regression of murine lung tumors by the let-7 microRNA. *Oncogene* 29, 1580–1587. doi: 10.1038/onc.2009.445

Trippe, R., Guschina, E., Hossbach, M., Urlaub, H., Luhrmann, R., and Benecke, B. J. (2006). Identification, cloning, and functional analysis of the human U6 snRNA-specific terminal uridylyl transferase. *RNA* 12, 1494–1504. doi: 10.1261/rna.87706

Trippe, R., Richly, H., and Benecke, B. J. (2003). Biochemical characterization of a U6 small nuclear RNA-specific terminal uridylyltransferase. *Eur. J. Biochem*. 270, 971–980. doi: 10.1046/j.1432-1033.2003.03466.x

Trippe, R., Sandrock, B., and Benecke, B. J. (1998). A highly specific terminal uridylyl transferase modifies the 3'-end of U6 small nuclear RNA. *Nucleic Acids Res*. 26, 3119–3126.

Tu, H. C., Schwitalla, S., Qian, Z., LaPier, G. S., Yermalovich, A., Ku, Y. C., et al. (2015). LIN28 cooperates with WNT signaling to drive invasive intestinal and colorectal adenocarcinoma in mice and humans. *Genes Dev*. 29, 1074–1086. doi: 10.1101/gad.256693.114

Urbach, A., Yermalovich, A., Zhang, J., Spina, C. S., Zhu, H., Perez-Atayde, A. R., et al. (2014). Lin28 sustains early renal progenitors and induces Wilms tumor. *Genes Dev*. 28, 971–982. doi: 10.1101/gad.237149.113

Urgese, G., Paciello, G., Acquaviva, A., and Ficarra, E. (2016). isomiR-SEA: an RNA-Seq analysis tool for miRNAs/isomiRs expression level profiling and miRNA-mRNA interaction sites evaluation. *BMC Bioinformatics* 17:148. doi: 10.1186/s12859-016-0958-0

Ustianenko, D., Hrossova, D., Potesil, D., Chalupnikova, K., Hrazdilova, K., Pachernik, J., et al. (2013). Mammalian DIS3L2 exoribonuclease targets the uridylated precursors of let-7 miRNAs. *RNA* 19, 1632–1638. doi: 10.1261/rna.040055.113

Ustianenko, D., Pasulka, J., Feketova, Z., Bednarik, L., Zigackova, D., Fortova, A., et al. (2016). TUT-DIS3L2 is a mammalian surveillance pathway for aberrant structured non-coding RNAs. *Embo J*. 35, 2179–2191. doi: 10.15252/embj.201694857

Viswanathan, S. R., and Daley, G. Q. (2010). Lin28: a microRNA regulator with a macro role. *Cell* 140, 445–449. doi: 10.1016/j.cell.2010.02.007

Viswanathan, S. R., Daley, G. Q., and Gregory, R. I. (2008). Selective blockade of microRNA processing by Lin28. *Science* 320, 97–100. doi: 10.1126/science.1154040

Viswanathan, S. R., Powers, J. T., Einhorn, W., Hoshida, Y., Ng, T. L., Toffanin, S., et al. (2009). Lin28 promotes transformation and is associated with advanced human malignancies. *Nat. Genet*. 41, 843–848. doi: 10.1038/ng.392

Wang, L. D., Rao, T. N., Rowe, R. G., Nguyen, P. T., Sullivan, J. L., Pearson, D. S., et al. (2015). The role of Lin28b in myeloid and mast cell differentiation and mast cell malignancy. *Leukemia* 29, 1320–1330. doi: 10.1038/leu.2015.19

Wang, L., Nam, Y., Lee, A. K., Yu, C., Roth, K., Chen, C., et al. (2017). LIN28 zinc knuckle domain is required and sufficient to induce let-7 oligouridylation. *Cell Rep*.18, 2664–2675. doi: 10.1016/j.celrep.2017.02.044

Wang, W., Peng, B., Wang, D., Ma, X., Jiang, D., Zhao, J., et al. (2011). Human tumor microRNA signatures derived from large-scale oligonucleotide microarray datasets. *Int. J. Cancer* 129, 1624–1634. doi: 10.1002/ijc.25818

Weidhaas, J. B., Babar, I., Nallur, S. M., Trang, P., Roush, S., Boehm, M., et al. (2007). MicroRNAs as potential agents to alter resistance to cytotoxic anticancer therapy. *Cancer Res*. 67, 11111–11116. doi: 10.1158/0008-5472.CAN-07-2858

Weiss, L. A., Arking, D. E., Gene Discovery Project of Johns the Autism, C., Daly, M. J., and Chakravarti, A. (2009). A genome-wide linkage and association

scan reveals novel loci for autism. *Nature* 461, 802–808. doi: 10.1038/nature 08490

Wen, J., Ladewig, E., Shenker, S., Mohammed, J., and Lai, E. C. (2015). Analysis of nearly one thousand mammalian mirtrons reveals novel features of dicer substrates. *PLoS Comput. Biol*. 11:e1004441. doi: 10.1371/journal.pcbi.1004441

Westholm, J. O., Ladewig, E., Okamura, K., Robine, N., and Lai, E. C. (2012). Common and distinct patterns of terminal modifications to mirtrons and canonical microRNAs. *RNA* 18, 177–192. doi: 10.1261/rna.030627.111

Wolin, S. L., and Cedervall, T. (2002). The La protein. *Annu. Rev. Biochem*. 71, 375–403. doi: 10.1146/annurev.biochem.71.090501.150003

Wyman, S. K., Knouf, E. C., Parkin, R. K., Fritz, B. R., Lin, D. W., Dennis, L. M., et al. (2011). Post-transcriptional generation of miRNA variants by multiple nucleotidyl transferases contributes to miRNA transcriptome complexity. *Genome Res*. 21, 1450–1461. doi: 10.1101/gr.118059.110

Xiao, Q., Wu, X. L., Michal, J. J., Reeves, J. J., Busboom, J. R., Thorgaard, G. H., et al. (2006). A novel nuclear-encoded mitochondrial poly(A) polymerase PAPD1 is a potential candidate gene for the extreme obesity related phenotypes in mammals. *Int. J. Biol. Sci*. 2, 171–178. doi: 10.7150/ijbs.2.171

Yamashita, S., Takagi, Y., Nagaike, T., and Tomita, K. (2017). Crystal structures of U6 snRNA-specific terminal uridylyltransferase. *Nat Commun*. 8:15788. doi: 10.1038/ncomms15788

Yang, N., Kaur, S., Volinia, S., Greshock, J., Lassus, H., Hasegawa, K., et al. (2008). MicroRNA microarray identifies Let-7i as a novel biomarker and therapeutic target in human epithelial ovarian cancer. *Cancer Res*. 68, 10307–10314. doi: 10.1158/0008-5472.CAN-08-1954

Yang, Z., Ren, F., Liu, C., He, S., Sun, G., Gao, Q., et al. (2010). dbDEMC: a database of differentially expressed miRNAs in human cancers. *BMC Genomics* 11(Suppl. 4):S5. doi: 10.1186/1471-2164-11-S4-S5

Yates, L. A., Durrant, B. P., Fleurdepine, S., Harlos, K., Norbury, C. J., and Gilbert, R. J. (2015). Structural plasticity of Cid1 provides a basis for its distributive RNA terminal uridylyl transferase activity. *Nucleic Acids Res*. 43, 2968–2979. doi: 10.1093/nar/gkv122

Yeom, K. H., Heo, I., Lee, J., Hohng, S., Kim, V. N., and Joo, C. (2011). Single-molecule approach to immunoprecipitated protein complexes: insights into miRNA uridylation. *EMBO Rep*. 12, 690–696. doi: 10.1038/embor.2011.100

Yi, R., Qin, Y., Macara, I. G., and Cullen, B. R. (2003). Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev*. 17, 3011–3016. doi: 10.1101/gad.1158803

Yu, J., Vodyanik, M. A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J. L., Tian, S., et al. (2007). Induced pluripotent stem cell lines derived from human somatic cells. *Science* 318, 1917–1920. doi: 10.1126/science.1151526

Zhang, H., Kolb, F. A., Jaskiewicz, L., Westhof, E., and Filipowicz, W. (2004). Single processing center models for human Dicer and bacterial RNase III. *Cell* 118, 57–68. doi: 10.1016/j.cell.2004.06.017

Zhang, Y., Zang, Q., Zhang, H., Ban, R., Yang, Y., Iqbal, F., et al. (2016). DeAnnIso: a tool for online detection and annotation of isomiRs from small RNA sequencing data. *Nucleic Acids Res*. 44, W166–W175. doi: 10.1093/nar/gkw427

Zhu, H., Shah, S., Shyh-Chang, N., Shinoda, G., Einhorn, W. S., Viswanathan, S. R., et al. (2010). Lin28a transgenic mice manifest size and puberty phenotypes identified in human genetic association studies. *Nat. Genet*. 42, 626–630. doi: 10.1038/ng.593

Zhu, H., Shyh-Chang, N., Segre, A. V., Shinoda, G., Shah, S. P., Einhorn, W. S., et al. Q. (2011). The Lin28/let-7 axis regulates glucose metabolism. *Cell* 147, 81–94. doi: 10.1016/j.cell.2011.08.033

The Emerging Field of Epitranscriptomics in Neurodevelopmental and Neuronal Disorders

Margarita T. Angelova[1†], Dilyana G. Dimitrova[1†], Nadja Dinges[2†], Tina Lence[2†], Lina Worpenberg[2†], Clément Carré[1] and Jean-Yves Roignant[2*]*

[1] Drosophila Genetics and Epigenetics, Sorbonne Université, Centre National de la Recherche Scientifique, Biologie du Développement—Institut de Biologie Paris Seine, Paris, France, [2] Laboratory of RNA Epigenetics, Institute of Molecular Biology, Mainz, Germany

Analogous to DNA methylation and histone modifications, RNA modifications represent a novel layer of regulation of gene expression. The dynamic nature and increasing number of RNA modifications offer new possibilities to rapidly alter gene expression upon specific environmental changes. Recent lines of evidence indicate that modified RNA molecules and associated complexes regulating and "reading" RNA modifications play key roles in the nervous system of several organisms, controlling both, its development and function. Mutations in several human genes that modify transfer RNA (tRNA) have been linked to neurological disorders, in particular to intellectual disability. Loss of RNA modifications alters the stability of tRNA, resulting in reduced translation efficiency and generation of tRNA fragments, which can interfere with neuronal functions. Modifications present on messenger RNAs (mRNAs) also play important roles during brain development. They contribute to neuronal growth and regeneration as well as to the local regulation of synaptic functions. Hence, potential combinatorial effects of RNA modifications on different classes of RNA may represent a novel code to dynamically fine tune gene expression during brain function. Here we discuss the recent findings demonstrating the impact of modified RNAs on neuronal processes and disorders.

Keywords: RNA modification, m^5C, Nm, pseudouridine, m^6A, neurons, disease

INTRODUCTION

An estimated 1–2% of all genes in a given organism contribute to nucleic acid modification systems, suggesting biological importance of modified nucleotides (Grosjean, 2009). A classic example is the methylation of cytosine on DNA, which acts as a critical epigenetic regulator of gene expression (Bird, 2002). Additionally, current advances in RNA modification research report over 140 distinct post-transcriptional RNA modifications (Cantara et al., 2011; Machnicka et al., 2013). Initial knowledge has been derived from studies on abundant non-coding RNAs (ncRNAs), such as transfer RNAs (tRNAs) and ribosomal RNAs (rRNAs), in prokaryotes and simple eukaryotes. These pioneer investigations described a diverse, chemically complex, and strongly conserved nature of RNA nucleotide modifications (Cantara et al., 2011; Machnicka et al., 2013). The most heavily modified RNAs in any cell type and organism are tRNAs. Up to 20% of nucleotides

in mammalian cytoplasmic tRNAs carry modifications (Motorin and Helm, 2011; Pan, 2018). Modified nucleotides outside the anticodon loop of tRNAs occur non-randomly at conserved positions across diverse species and affect in general its stability (Helm, 2006; Motorin and Helm, 2010). In addition, modifications in the anticodon loop can contribute to optimize mRNA decoding by directly affecting codon-anticodon interactions (Agris, 2008).

Aberrant tRNA and rRNA modifications have been linked to various human disease syndromes and the phenotypes are often observed in specific tissues such as the gonads and the nervous system (Torres et al., 2014). Notably, increasing number of predicted human transfer RNA (tRNA) modification genes have been associated with neurological disorders, in particular with intellectual disability (ID) (for recent review see Bednárová et al., 2017). ID, or previously known as Mental Retardation (MR), is characterized by non-progressive cognitive impairment and affects 1–3% of the general population (Daily et al., 2000). It is presently unclear whether all observed phenotypes are caused by aberrant tRNA modifications, by effects on unidentified other RNA substrates (see below) and/or by a modification-independent function of the involved enzymes (Guo and Schimmel, 2013; Genenncher et al., 2018). Likewise, it is unknown why some tissues, in particular the brain, are more sensitive to the loss of these modifications.

Importantly, besides the heavily modified tRNAs and rRNAs, mRNAs, small and long non-coding RNAs were also found to harbor post-transcriptional modifications. Recent technological advances that allowed mapping of selected RNA modifications on a transcriptome-wide scale revealed widespread distribution of N6-methyladenosine (m^6A), pseudouridine (Ψ) and ribose 2′-O-methylation (Nm) on mRNA (Dominissini et al., 2012; Meyer et al., 2012; Carlile et al., 2014; Schwartz et al., 2014a; Dai et al., 2017). The prevalence of some others, including N1-methyladenine (m^1A) and 5-methylcytidine (m^5C) is still debated (Dominissini et al., 2016; Li et al., 2016, 2017c; Dominissini and Rechavi, 2017; Legrand et al., 2017; Safra et al., 2017). m^6A, the most abundant mRNA modification, was shown to affect almost every step of mRNA biogenesis, including splicing, export, translation, and mRNA decay (Lence et al., 2017; Roignant and Soller, 2017). It is thus not surprising that misregulation of m^6A results in several physiological defects, including brain development abnormalities, obesity, cancer, and other diseases (Batista, 2017; Dai et al., 2018). In addition, the discovery of m^6A RNA demethylases (Jia et al., 2011; Zheng et al., 2013; Jacob-Hirsch et al., 2018) and the identification of m^6A-binding proteins (Dominissini et al., 2012) indicated that similarly to DNA modification, RNA methylation can be reversible and convey information *via* recognition of effector proteins.

Altogether these recent studies revealed an entire new layer of regulation of gene expression, which has been central to the development of a novel concept called "RNA epigenetics or epitranscriptomics" (He, 2010; Meyer et al., 2012). However, the exact biological function of the majority of modified RNA nucleotides remains to be discovered. In this review, we will focus

on several RNA modifications and will discuss their involvement in the development of the brain and neurological disorders.

5-METHYLCYTOSINE (m^5C)

Cytosine can be methylated at the 5th position of the pyrimidine ring to form 5-methylcytosine (m^5C) (**Figure 1**). Various eukaryotic cytosine-5-RNA methyltransferases catalyze the formation of m^5C at specific positions (Motorin and Grosjean, 1999; Brzezicha et al., 2006; Sharma et al., 2013; Metodiev et al., 2014; Haag et al., 2015; Schosserer et al., 2015). The analysis of genetic mutations in two particular RNA cytosine-5 methyltransferase family members (Dnmt2/Trdmt and NCL1/TRM4/NSun2) has provided important insights into the biological effects of aberrant m^5C deposition.

Dnmt2

Dnmt2 is a member of the most widely conserved eukaryotic cytosine-5-DNA methyltransferase protein family (Goll and Bestor, 2005). Despite this classification, only few studies reported Dnmt2-mediated DNA methylation (Hermann et al., 2003; Kunert et al., 2003; Phalke et al., 2009) and it is today acknowledged that Dnmt2 functions mainly as a tRNA methylase (Okano et al., 1998; Schaefer and Lyko, 2010a,b; Raddatz et al., 2013). Dnmt2 methylation activity on position C38 of three tRNAs, which include tRNA^{Asp}, tRNA^{Val}, and tRNA^{Gly}, has been described in yeast, *Drosophila*, mouse, and human cells (Goll et al., 2006; Jurkowski et al., 2008; Schaefer et al., 2010). Knockdown of Dnmt2 in zebrafish embryos leads to differentiation defects in some organs, and notably, to abnormal neurogenesis in the hypothalamus and diencephalon (Rai et al., 2007). In *Dnmt2* mutant flies, reduced viability under stress conditions was observed (Schaefer et al., 2010). This is in accordance with previous studies that suggest an increased tolerance for stress in *Drosophila* and *Entamoeba* upon Dnmt2 overexpression (Lin et al., 2005; Fisher et al., 2006). Nevertheless, the majority of studies suggests that *Dnmt2* mutation does not trigger strong detrimental phenotypes in yeast, *Drosophila* and mice (Wilkinson et al., 1995; Kunert et al., 2003; Goll et al., 2006; Schaefer et al., 2010), which raises the question why zebrafish relies on Dnmt2 for proper development, whereas mice and flies do not. One possible explanation is that these organisms have redundant mechanisms that compensate for the loss of Dnmt2, which may be absent or less robust in zebrafish. Consistent with this possibility, it was shown that *Dnmt2* mutant mice exhibit lethal phenotypes in the absence of a second m^5C methyltransferase, NSun2 (Tuorto et al., 2012). In human, polymorphisms in *DNMT2* have been associated with spina bifida, a congenital malformation of the central nervous system (Franke et al., 2009).

NSun2

Unlike Dnmt2, it has been established that mammalian NSun2 does not only modify tRNAs (Blanco et al., 2011; Tuorto et al., 2012) but also other small ncRNAs such as 7SK, vault, and Y-RNAs (Hussain et al., 2013; Khoddami and Cairns, 2013). *Dnmt2* and *Nsun2* double knockout mice showed a lethal

FIGURE 1 | Some of the most common modifications in RNA.
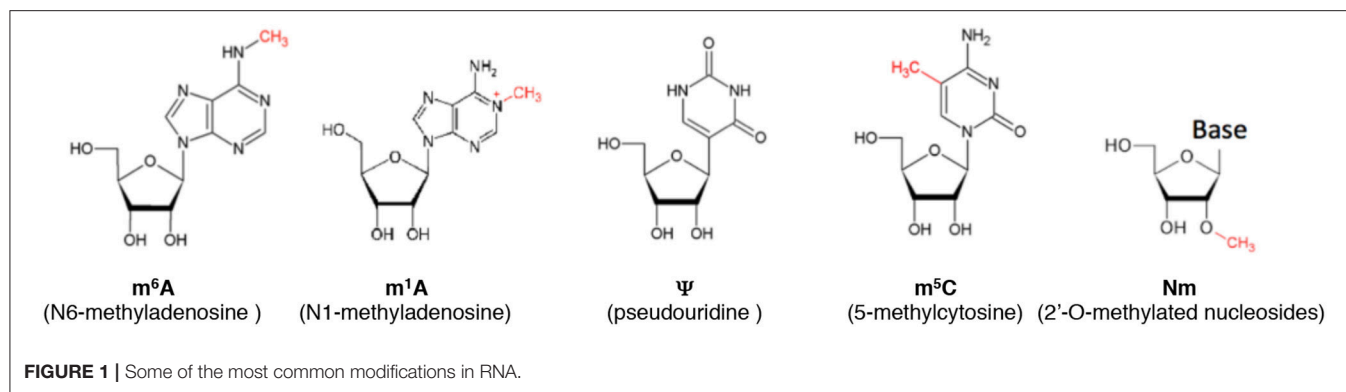
phenotype. However, deletion of *NSun2* alone (Blanco et al., 2011; Tuorto et al., 2012; Hussain et al., 2013) or in combination with *Dnmt2* (Rai et al., 2007; Tuorto et al., 2012) in specific tissues impairs cellular differentiation pathways in mammalian skin, testes, and brain. The function in the brain appears conserved as *Nsun2* mutations are associated with ID and Dubowitz-like syndrome in humans (Abbasi-Moheb et al., 2012; Khan et al., 2012; Martinez et al., 2012), as well as with microcephaly in human and mice (Blanco et al., 2014). In addition, a recent study in human and mice neuron precursor cells showed that m^5C deposited by Nsun2 regulates neural stem cell (NSC) differentiation and motility (Flores et al., 2017). This study thus provides some links between the failure of RNA m^5C deposition and the associated brain development diseases.

It is intriguing that patient fibroblasts and *Nsun2*-deficient mice (Blanco et al., 2014), as well as *Dnmt2* mutant flies (Durdevic et al., 2013), exhibit increase cleavage of tRNA, and elevated production of tRNA fragments (tRFs). This accumulation of tRFs reduces protein translation rates and increases oxidative stress as well as neuronal apoptosis. Interestingly, reducing tRNA cleavage in *Nsun2*-deficient brains is sufficient to rescue sensitivity to oxidative stress, implying that tRFs play a role in Nsun2-mediated defects.

(2′-O)-METHYLATION (Nm)

2′-O-methylation (Nm) is a common nucleoside modification of RNA, where a methyl group is added to the 2′ hydroxyl of the ribose moiety (**Figure 1**). Nm increases hydrophobicity, protects RNAs from nuclease attacks and stabilizes helical structures (Kurth and Mochizuki, 2009; Byszewska et al., 2014; Kumar et al., 2014; Yildirim et al., 2014). Nm is predominantly found internally in ribosomal RNAs and small nuclear RNAs as well as in tRNAs and in a number of sites on mRNA (Darzacq et al., 2002; Rebane et al., 2002; Kurth and Mochizuki, 2009; Zhao et al., 2012; Somme et al., 2014; Dai et al., 2017). This modification is also present at the 3′-end of miRNAs and siRNAs in plants (Li et al., 2005; Yu et al., 2005), as well as in siRNAs and piRNAs in animals (Horwich et al., 2007; Saito et al., 2007). Nm methyltransferases acting on tRNAs are highly conserved from bacteria and archaea to humans (Somme et al., 2014) and usually target positions in the anticodon loop. For instance, TRM7 in

S. cerevisiae modifies positions 32 and 34 of selected tRNA, amongst which is tRNA^{Phe} (Pintard et al., 2002; Guy et al., 2012). Strikingly, FTSJ1, the TRM7 ortholog in human, methylates the exact same positions of the exact same tRNAs (Guy and Phizicky, 2015; Guy et al., 2015). Consistent with a conserved function, expression of human FTSJ1 can suppress the severe growth defect of *S. cerevisiae Δtrm7* mutants (Pintard et al., 2002; Guy and Phizicky, 2015). Reduction of the modification level in tRNA^{Phe} was reported in carcinoma and neuroblastoma in mice (Pergolizzi and Grunberger, 1980; Kuchino et al., 1982) and is associated with ID in human (see below).

FTSJ1

One of the best characterized associations between ID in human and mutations in a gene encoding for Nm is the one between non-syndromic X-linked ID (NSXLID) and mutations in the *FTSJ1* gene (OMIM:300499) (Guy et al., 2015). One third of the X-linked ID (XLID) conditions are syndromic (S-XLID) and the other two thirds are non-syndromic (NS-XLID) (Lubs et al., 2012). NSXLID is associated with no obvious and consistent phenotype other than mental retardation (IQ < 70), indeed NSXLID disorders are clinically diverse and genetically heterogeneous. *FTSJ1* loss of function causes NSXLID retardation in males (Froyen et al., 2007; Takano et al., 2008). Heterozygous loss of function mutations in females do not cause the disease, which is probably due to inactivation of the affected X chromosome. Several alleles of *FTSJ1* from six independent families correlate with NSXLID. All of these alleles lead to a reduction in mRNA levels and/or protein function (Willems et al., 1993; Hamel et al., 1999; Freude et al., 2004; Ramser et al., 2004; Froyen et al., 2007; Takano et al., 2008; Guy et al., 2015; **Table 1**). Consistently with the 2′-O-methyltransferase activity of FTSJ1 on tRNAs, Guy and Phizicky reported that two genetically independent lymphoblastoid cell lines (LCLs) of NSXLID patients with *FTSJ1* loss of function mutations nearly completely lack Cm_{32} and Gm_{34} on tRNA^{Phe} (Guy et al., 2015). Additionally, tRNA^{Phe} from a patient carrying an *FTSJ1-p.A26P* missense allele specifically lacks Gm_{34}, but has normal levels of Cm_{32}. tRNA^{Phe} from the corresponding *Saccharomyces cerevisiae TRM7-A26P* mutant also specifically lacks Gm_{34}. Altogether, these findings strongly suggest that the absence of Gm_{34}, but not Cm_{32} modification on tRNA^{Phe} causes NSXLID in patients

TABLE 1 | Proteins required for writing, reading, or removal of different RNA modifications and their mutations associated with altered brain functions.

Modification	Gene	Organism	Defect	Effect/Disease	References
m^{5}C	Dnmt2/Trdmt	*Dr*	KD	Abnormal neurogenesis in the hypothalamus and diencephalon. Defects in retina and liver.	Rai et al., 2007
		Dm	Lof	Decreased tolerance to stress (reduced viability under stress conditions).	Schaefer et al., 2010
	NSun2/NCL1/TRM4	*Hs*	Lof	Autosomal-recessive ID, facial dysmorphism, microcephaly and Dubowitz-like syndrome.	Abbasi-Moheb et al., 2012; Khan et al., 2012; Martinez et al., 2012
		Mm	Lof	Impaired cortical, hippocampal and striatal expansion during development Microcephaly. Decrease in neural stem cell (NSC) differentiation and motility.	Blanco et al., 2014; Flores et al., 2017
		Dm	Lof	Severe short-term memory deficits.	Abbasi-Moheb et al., 2012
	Dnmt2, NSUN2 double mutant	*Ms*	KO	Reduced proliferation rates, underdeveloped pheno- type in several tissues, including thickness and organization of the cerebral cortex.	Tuorto et al., 2012
Nm	FTSJ1	*Hs*	Lof	Nonsyndromic X-linked ID (NSXLID) in males.	Willems et al., 1993; Hamel et al., 1999; Freude et al., 2004; Guy et al., 2015
			SNPs	Impact on general cognitive ability, verbal comprehension, and perceptual organization in males.	Gong et al., 2008
			Gof	ID	Giorda et al., 2009; Honda et al., 2010
	TRMT44	*Hs*	SNPs	Partial epilepsy with pericentral spikes (PEPS).	Leschziner et al., 2011
	C/D box snoRNAs SNORD115 (HBII-52); SNORD116 (HBII-85) and others in the 15q11-q13 region	*Hs*	Lof	Prader-Willi syndrome (PWS).	Cavaillé et al., 2000; Kishore and Stamm, 2006; Peters, 2008; Sahoo et al., 2008; Sridhar et al., 2008; Doe et al., 2009; Duker et al., 2010
	15q11-q13 region	*Hs*	Gof	Autism	Bolton et al., 2004; Cook and Scherer, 2008
	Hen1	*Dm*	Lof	Accelerated neurodegeneration-related phenotypes (brain vacuolization, memory defaults and shorter life span).	Abe et al., 2014
Ψ	Unknown	*Hs*	Gof	Alzheimer's disease	Lee et al., 2007
		Hs	Lof	Myotonic dystrophy type 2 (DM2).	Delorimier et al., 2017
	Pus1 (TruA family member)	*Hs*	Lof	Mild-cognitive impairment, mitochondrial myopathy and sideroblastic anemia.	Cao et al., 2016
	Pus3 (TruA family member)	*Hs*	Lof	ID	Shaheen et al., 2016
	DKC1 (dyskerin)	*Hs*	Lof	X-linked recessive dyskeratosis congenita (DKC)	Heiss et al., 1998
m^{6}A	METTL3 (m^{6}A writer)	*Hs*	Lof	Impaired neuronal differentiation and formation of mature neurons from embryoid bodies.	Batista et al., 2014; Geula et al., 2015
		Dm	SNPs	Severe locomotion defects due to altered neuronal functions.	Haussmann et al., 2016; Lence et al., 2016; Kan et al., 2017
	Mettl14 (m^{6}A writer)	*Mm*	cKO	Delayed specification of different neuronal subtypes during brain development. Altered axon regeneration.	Yoon et al., 2017; Wang et al., 2018; Weng et al., 2018
	Wtap (m^{6}A writer)	*Dr*	KD	Smaller brain ventricles and curved notochord.	Ping et al., 2014
	ZC3H13	*Hs*	SNP	Schizophrenia	Oldmeadow et al., 2014
	Spenito (m^{6}A writer)	*Dm*	Lof	Control axon outgrowth, branching and synaptic bouton formation.	Gu et al., 2017
	Ythdc1 (m^{6}A reader)	*Dm*	KD	Enhancement of SCA1-induced neurodegeneration	Fernandez-Funez et al., 2000
	ALKBH5 (m^{6}A eraser)	*Hs*	SNP	Major depressive disorder (MDD).	Du et al., 2015

(Continued)

Modification	Gene	Organism	Defect	Effect/Disease	References
	FTO (m^{6}A eraser)	*Hs*	SNP	Decreased brain volume, increased risk for attention-deficit/hyperactivity disorder (ADHD) and Alzheimer's disease.	Keller et al., 2011; Reitz et al., 2012; Choudhry et al., 2013; Melka et al., 2013; Li et al., 2017a
		Mm	KO	Altered behavior (e.g., locomotion defects), abnormal electrophysiological response to cocaine (impaired dopamine type 2 and 3 receptor response) and enhanced consolidation of cued fear memory.	Hess et al., 2013; Widagdo et al., 2016

ID, intellectual disability; KD, knockdown; KO, knock out; cKO, conditional knock out; Lof, reduced function or loss of function; Gof, gain of function; SNP, single nucleotide polymorphism; Hs, Homo sapiens; Mm, Mus musculus; Dr, Danio rerio; Dm, Drosophila melanogaster N6-methyladenosine (m^{6}A); pseudouridine (Ψ), 5-methylcytosine (m^{5}C); and 2′-O-methylation (Nm).

carrying distinct *FTSJ1* alleles. Nevertheless, the molecular consequences arising from the loss of this 2′-O-methylation are not yet determined. Furthermore, it is noteworthy to mention two additional studies involving families from the Chinese Han population (Dai et al., 2008; Gong et al., 2008), where three single nucleotide polymorphisms (SNPs) in the *FTSJ1* gene were analyzed. Authors found a positive association with occurrence of NSXLID (Dai et al., 2008) as well as with general cognitive ability, verbal comprehension, and perceptual organization in male individuals (Gong et al., 2008). Although it seems tempting to link the variance of *FTSJ1* gene to general human cognitive ability, more profound studies are needed to support this idea.

TRMT44

TRMT44 is a putative 2′-O-methyluridine methyltransferase predicted to methylate residue 44 in tRNA^{Ser} (Leschziner et al., 2011). Mutations in this gene were identified as a causative mutation in partial epilepsy with pericentral spikes (PEPS), a novel mendelian idiopathic epilepsy (Leschziner et al., 2011). However, the underlying mechanisms are currently unknown.

Small Nucleolar RNAs (snoRNAs)

snoRNAs are a class of regulatory RNAs responsible for post-transcriptional modification of ribosomal RNAs (rRNAs). Two families of snoRNAs have been described, based on their structure and function: C/D box snoRNAs are responsible for 2′-O-methylation (Cavaillé et al., 1996), whereas H/ACA box snoRNAs mediate pseudouridylation (Ganot et al., 1997 and see the following chapter in this review). In zebrafish, loss of three snoRNAs results in impaired rRNA modifications, causing severe developmental defects including growth delay and deformations in the head region (Higa-Nakamine et al., 2012). In human, C/D box snoRNAs have been implicated in Prader-Willi syndrome (PWS), a complex neurological disease characterized with mental retardation, low height, obesity, and muscle hypotonia (Sridhar et al., 2008; Doe et al., 2009). In several independent studies, PWS was shown to be caused by the loss of imprinted snoRNAs in locus 15q11-q13. Large deletions of this region underlie about 70% of cases of PWS (Peters, 2008), whereas duplication of the same region is associated with autism (Belmonte et al., 2004; Bolton et al., 2004; Cook and Scherer, 2008). Locus 15q11–q13 contains numerous copies of two C/D box snoRNAs—SNORD115 (HBII-52), and SNORD116 (HBII-85) (Cavaillé et al., 2000). SNORD115 is believed to play key roles in the fine-tuning of serotonin receptor (5-HT2C) by influencing its pre-mRNA splicing (Vitali et al., 2005; Kishore and Stamm, 2006; Falaleeva et al., 2017), whereas SNORD116 loss is thought to contribute to the etiology of the PWS (Cavaillé et al., 2000; Sahoo et al., 2008; Duker et al., 2010).

Hen1/Pimet

Hen1/Pimet is a conserved enzyme, which adds 2′-O-methyl group to 3′-terminal nucleotides of miRNAs and siRNAs in plants, and of siRNAs and piRNAs in animals. Addition of this modification protects these small non-coding RNAs (sncRNAs) from 3′ → 5′ exonuclease degradation (Li et al., 2005; Horwich et al., 2007; Saito et al., 2007; Terrazas and Kool, 2009; Ross et al., 2014). In the absence of Hen1/Pimet, piRNA, and siRNA are destabilized and sncRNA silencing activities are compromised. Surprisingly, *Hen1* mutant flies display neither increased lethality nor sterility under normal laboratory conditions but show however accelerated neurodegeneration (brain vacuolization), memory default, and shorter lifespan (Abe et al., 2014). This suggests a protective effect of Nm and small RNA pathways against age-associated neurodegenerative events. Accordingly, *Drosophila* lacking the siRNA effector, Argonaute 2 (Ago2), are viable but exhibit memory impairment and shortened lifespan (Li et al., 2013).

PSEUDOURIDINE (ψ)

Pseudouridine (also known as 5-ribosyluracil or ψ) is the first discovered (Cohn and Volkin, 1951) and most abundant RNA modification, present in a broad range of non-coding RNA, and was also recently detected in coding mRNA (Carlile et al., 2014; Lovejoy et al., 2014; Schwartz et al., 2014a; Li et al., 2015). The isomerization of uridine into ψ improves the base stacking in RNAs by the formation of additional hydrogen bonds, which influences RNA secondary structure and increases the stability of RNA duplexes (Arnez and Steitz, 1994; Davis, 1995). Pseudouridylation was shown to have a strong impact on different aspects of cellular processes, including translation efficiency, splicing, telomere maintenance, and the

regulation of gene expression (Mochizuki et al., 2004; Carlile et al., 2014; Schwartz et al., 2014a). This base modification is catalyzed by pseudouridine synthases (Pus) that act on their substrates by two distinct mechanisms. One of those mechanisms is the guide RNA-dependent pseudouridylation, in which H/ACA box snoRNAs target RNAs for pseudouridylation *via* specific sequence interactions between the snoRNAs and the target RNA. A specific enzyme present in the snoRNP (sno-ribonucleoprotein) particle catalyzes the uridine modification (dyskerin in human, Cbf5 in yeast; Duan et al., 2009; Liang et al., 2009). Alternatively, RNA-independent pseudouridylation requires stand-alone pseudouridine synthases (Pus) that directly catalyze ψ formation at particular target RNA (Yu et al., 2011; Carlile et al., 2014; Rintala-Dempsey and Kothe, 2017). Each enzyme has a unique specificity for its target RNA and modifies uridine in a certain consensus sequence. Pus enzymes are present in all kingdoms of life, evolutionary conserved and are categorized into six families, based on their consensus sequences: TruA, TruB, TruD, RluA, and RsuA. The sixth family member, Pus10, is exclusive to eukaryotes and archaea (Koonin, 1996; Kaya and Ofengand, 2003; Fitzek et al., 2018).

Several pieces of evidence hint toward an implication of ψ in regulating neuronal functions. For instance, patients with mild-to-moderate severity of Alzheimer's disease show significantly elevated levels of urinal ψ (Lee et al., 2007) but it is currently unknown whether there is a link between this increase and the Alzheimer's disease etiology. Furthermore, it has been suggested that pseudouridylation can serve as a direct indicator of oxidative stress, which in turn has been linked to an increasing risk of neurodegeneration (Roth et al., 1999; Uttara et al., 2009). Accordingly, in cells exposed to acute oxidative stress by H_2O_2 treatment, Li and colleagues detected an elevation by ∼40–50% in mRNA ψ levels, demonstrating that mRNA pseudouridylation acts as a direct response to cellular stress (Li et al., 2015).

A recent report demonstrated a direct implication of ψ in neuronal disorders from patients with myotonic dystrophy type 2 (DM2) (Delorimier et al., 2017). DM2 is a neuromuscular disease characterized by severe gray matter changes, including neuronal loss and global neuronal impairment (Minnerop et al., 2011; Meola and Cardani, 2015). DM2 patients have an increased binding of Muscleblind-like 1 protein (MBNL1) to CCUG repeats in an intron of the *CNBP* gene (Cho and Tapscott, 2007). Interestingly, it was recently reported that pseudouridylation within CCUG repeats reduces RNA flexibility and thus modestly inhibits MBNL1 binding (Delorimier et al., 2017). Similarly, ψ modification of a minimally structured model RNA resulted in an even more drastic reduction of MBNL1 binding to CCUG repeats. This study shows that ψ can reduce the disease-causing binding of MBNL1 at extended CCUG repeats and offers a basis for future research in treating neurodegenerative diseases.

Pus1

Pus1 is a member of the TruA family that typically pseudouridylates tRNA but was also recently found to act on rRNA, snRNA, and mRNA (Schwartz et al., 2014a; Carlile et al., 2015). Mutations of *Pus1* in human lead to mitochondrial

myopathy and sideroblastic anemia (Bykhovskaya et al., 2004; Fernandez-Vizarra et al., 2007; Bergmann et al., 2010). Recently, a mild cognitive impairment was also characterized in a long-surviving patient with two novel *Pus1* mutations (Cao et al., 2016). A different study demonstrated a Pus1-dependent pseudouridylation of the steroid RNA activator (SRA). Pseudouridylated SRA acts as a co-activator of the nuclear estrogen receptor α (ERα) (Zhao et al., 2004; Leygue, 2007). Given that ERα was shown to regulate neuronal survival (Gamerdinger et al., 2006; Foster, 2012), it is conceivable that one of the functions of Pus1 in brain activity is mediated *via* the control of the ER pathway.

Pus3

Pus3 is another member of the TruA family, which has a strong sequence homology to Pus1 but acts on distinct target RNA. *In situ* hybridization showed accumulation of *Pus3* mRNA in the nervous system of mice embryos, suggesting a role of Pus3 in neural development (Diez-Roux et al., 2011). Accordingly, a truncated form of Pus3 accompanied by reduced levels of ψ U39 in tRNA was detected in patients with ID (Shaheen et al., 2016). Taken together, the discovery of impaired cognition caused by mutations of TruA enzymes emphasizes their importance in neuronal development and maintenance regulation.

Dyskerin and RluA-1

The pseudouridine synthase dyskerin is essential for the H/ACA-box mediated pseudouridylation in human (Heiss et al., 1998; Lafontaine et al., 1998). Mutations of the dyskerin-encoding gene DKC1 causes X-linked recessive dyskeratosis congenita (DKC), a rare progressive congenital disorder that mostly affects highly regenerative tissues, such as the skin and bone marrow (Heiss et al., 1998; Mochizuki et al., 2004). Cells of the affected patients have decreased telomerase activity and thus reduced telomere length, which may be responsible for the disease (Mitchell et al., 1999). Interestingly, expression analysis of *Dyskerin 1* showed a high level in embryonic neural tissue, as well as in specific subsets of neurons in the cerebellum and olfactory bulb of adult brains (Heiss et al., 2000). While the function of dyskerin in hematopoiesis has been studied intensively, we are yet lacking a detailed understanding about its potential nervous system function in adult brains. In *Drosophila melanogaster*, RluA enzymes modify uridines in rRNA and tRNA. *In situ* hybridization in embryos revealed a specific RluA-1 mRNA localization to dendrites of a subset of peripheral neurons, which also raises the question about the molecular function of RluA-1 and its target RNA in the peripheral nervous system during embryonic development (Wang et al., 2011).

N6-METHYLADENOSINE (m⁶A)

m⁶A is an abundant mRNA modification that regulates nearly all aspects of mRNA processing including splicing, export, translation, stability, and decay (Meyer and Jaffrey, 2017; Roignant and Soller, 2017; Roundtree et al., 2017). This modification is catalyzed by a stable protein complex composed of two methyltransferases, Methyltransferase like-3 (Mettl3) and

Methyltransferase like-14 (Mettl14) (Sledz and Jinek, 2016; Wang et al., 2016a,b; Schöller et al., 2018). Additional proteins required for m^{6}A deposition are Wilms' tumor 1-associating protein (Wtap) (Liu et al., 2014; Ping et al., 2014; Wang et al., 2014), Vir like m^{6}A methyltransferase associated (Virma) (Schwartz et al., 2014b; Yue et al., 2018), Zinc finger CCCH domain-containing protein 13 (Zc3h13) (Guo et al., 2018; Knuckles et al., 2018; Wen et al., 2018), RNA binding protein 15 (Rbm15) and its paralog Rbm15B (Patil et al., 2016). Mettl3 has the catalytic activity and can accommodate the SAM substrate, while Mettl14 serves to stabilize the binding to RNA (Sledz and Jinek, 2016; Wang et al., 2016a,b; Schöller et al., 2018). In vertebrates, m^{6}A modification is dynamically regulated and can be reversed by two demethylases belonging to the family of α-ketoglutarate dependent dioxygenases, Fat mass and obesity associated protein (FTO) and ALKBH5 (Jia et al., 2011; Zheng et al., 2013). Recent advances in techniques to map m^{6}A modification in a transcriptome wide manner enabled identification of thousands of modified mRNAs and lncRNAs (Dominissini et al., 2012; Meyer et al., 2012). While m^{6}A has been involved in many physiological processes, increasing evidence suggests an importance of m^{6}A modification in brain development and in the function of the nervous system.

m^{6}A Writer Complex

m^{6}A levels are particularly high in the nervous system, as shown in the developing mouse brain (Meyer et al., 2012), and in heads of adult flies (Lence et al., 2016). Furthermore, a recent study detected higher m^{6}A content in the mouse cerebellum and in neurons compared to glia (Chang et al., 2017). Using in situ hybridization in zebrafish embryos, Ping et al. showed that Wtap is ubiquitously expressed at 36 h post-fertilization with enrichment in the brain region (Ping et al., 2014). Consistently, Wtap depletion using morpholino treatment resulted in severe developmental defects, including appearance of smaller brain ventricles and curved notochord at 24 h post-fertilization. Importance of m^{6}A during neuronal development was further demonstrated by depletion of METTL3 in human embryonic stem cells (hESC), which strongly impaired neuronal differentiation (Batista et al., 2014), as well as the formation of mature neurons from embryoid bodies (Geula et al., 2015). Notably, m^{6}A mRNA modification is essential for mouse survival as mice lacking Mettl3 die at E6.5 (Geula et al., 2015). However, two recent studies performed a conditional KO (cKO) of Mettl14 specifically in neurons and revealed an essential role of m^{6}A in embryonic cortical neurogenesis (Yoon et al., 2017; Wang et al., 2018). Mettl14 cKO animals showed a decreased NSC proliferation and premature differentiation of NSCs (Wang et al., 2018), as well as delayed specification of different neuronal subtypes during brain development (Yoon et al., 2017). Yoon at al. further demonstrated that m^{6}A modification is required for timely decay of transcripts involved in stem cell maintenance and cell cycle regulation in cortical neuronal progenitors. This allows accurate progression of the cell cycle and in turn induces the spatiotemporal formation of different neuronal subtypes. Interestingly, the authors also observed that many transcripts linked to mental disorders (autism, schizophrenia) are m^{6}A

modified in human, but not in mouse cultures of neuronal progenitor cells (NPC), raising the possibility that m^{6}A regulates specifically these human diseases (Yoon et al., 2017). Consistent with this hypothesis, polymorphisms in ZC3H13 have been associated with schizophrenia (Oldmeadow et al., 2014).

Beyond the role in neuronal development, m^{6}A modification also plays a critical role in the process of axon regeneration in mature mouse neurons (Weng et al., 2018). Weng et al. showed that upon peripheral nerve injury m^{6}A levels of many transcripts in dorsal root ganglion (DRG) were elevated, which led to increased translation during the time of axon regeneration, via the specific m^{6}A reader protein Ythdf1. Mettl14 and Ythdf1 conditional KO mice displayed strong reduction of sensory axon regeneration, resulting from reduced protein synthesis, revealing the critical role of m^{6}A modification in response to injury (Weng et al., 2018).

In Drosophila, loss of components of the methyltransferase complex results in severe locomotion defects due to altered neuronal functions (Haussmann et al., 2016; Lence et al., 2016; Kan et al., 2017). Mettl3 mutants display alterations in walking speed and orientation, which can be rescued by ectopic expression of Mettl3 cDNA in neurons. Whether a particular subset of neurons is responsible for the observed alterations awaits further investigations. Interestingly, another member of the m^{6}A methyltransferase complex, Nito (RBM15 in human), was recently shown to control axon outgrowth, branching and to regulate synaptic bouton formation via the activity of the CCAP/bursicon neurons (Gu et al., 2017), providing first insights toward addressing this question.

m^{6}A Readers

As mentioned above, most characterized functions of m^{6}A rely on the direct binding of the so-called m^{6}A "reader" proteins to the modified site. The best-studied m^{6}A readers are the YTH-domain containing proteins, which can specifically bind m^{6}A via their YTH domain (Luo and Tong, 2014; Theler et al., 2014; Xu et al., 2014). RNA in situ hybridization of rat brain sections showed that one particular member of the YTH family, Ythdc1, is enriched in specific cells in the brain (Hartmann et al., 1999). Interestingly, in a yeast two-hybrid screen to identify Ythdc1-interacting proteins, libraries from P5 and E16 brains were screened and the rat homolog of Sam68 was found as the main interactor (Hartmann et al., 1999). Sam68 is known to regulate neuronal activity-dependent alternative splicing events (i.e., of neurexin-1) (Iijima et al., 2011). In line with this function, in situ hybridization assays showed that Drosophila Ythdc1 localizes in the ventral neurectoderm and central nervous system of Drosophila embryos (Lence et al., 2016) and a reduced level of Ythdc1 was found to enhance SCA1-induced neurodegeneration (Fernandez-Funez et al., 2000).

Apart from the YTH-protein family of conventional m^{6}A reader proteins, a number of other proteins that bind RNA in m^{6}A-dependent fashion has been recently identified (Edupuganti et al., 2017). Among them, Fragile X mental retardation protein (FMRP, also known as POF; FMR1; POF1; FRAXA) was shown to preferentially bind an RNA probe containing m^{6}A sites (Edupuganti et al., 2017). FMRP plays critical roles in synaptic

plasticity and neuronal development. Its loss of function in human leads to the Fragile X syndrome, which is the most prevalent form of inherited ID and the foremost monogenic cause of autism (Bardoni et al., 2001; Lubs et al., 2012; Wang et al., 2012; Hagerman and Polussa, 2015). FMRP has a central role in neuronal development and synaptic plasticity through the regulation of alternative mRNA splicing, mRNA stability, mRNA dendritic transport and postsynaptic local protein synthesis of a subset of mRNAs (Antar et al., 2006; Didiot et al., 2008; Bechara et al., 2009; Ascano et al., 2012; Guo et al., 2015). Moreover, it represses mRNA translation during the transport of dendritic mRNAs to postsynaptic dendritic spines and activates mRNA translation of a subset of dendritic mRNAs at synapses (Bechara et al., 2009; Fähling et al., 2009). Consistent with a potential interplay between FMRP and m^{6}A, a recent study found that m^{6}A is present on many synaptic mRNAs that are known targets of FMRP protein (Chang et al., 2017). Future research will seek to further illuminate the potential role of FMRP within the m^{6}A pathway. It is interesting to note that Nm also appears to contribute to FMRP-mediated translation regulation at synapses. FMRP can form a complex with the non-coding RNA, *brain cytoplasmic RNA* (*BC1*), to repress translation of a subset of FMRP target mRNAs (Zalfa et al., 2003). This interaction is modulated by the Nm status of *BC1* RNA. In both nucleus and cytoplasm in the cell body, Nm is present on *BC1*, but it is virtually absent at synapses (Lacoux et al., 2012). The authors suggested that changes in the 2′-O-methylation status of *BC1* RNA contribute to the fine-tuned regulation of gene expression at synapses and consequently to neuronal plasticity by influencing FMRP local translational control. This example supports a likely combinatorial role of RNA modifications in the regulation of similar targets and/or processes during brain function.

m^{6}A Erasers

Two proteins in humans were reported to act as m^{6}A erasers: (FTO) and AlkB homolog 5 (ALKBH5) (Jia et al., 2011; Zheng et al., 2013). Both belong to the family of Fe^{2+}-α-ketoglutarate-dependent deoxygenases and catalyze the removal of the methyl group of m^{6}A by oxidation. Interestingly, even though ALKBH5 is only moderately expressed in the brain, it has been associated with mental disorders. Du et al. found that certain polymorphisms within the *ALKBH5* gene correlate with the major depressive disorder (MDD), suggesting an involvement of ALKBH5 in conferring risk of MDD (Du et al., 2015).

In comparison, FTO is highly expressed in the human brain, especially in the hypothalamus and the pituitary gland and displays dynamic expression during postnatal neurodevelopment. Polymorphic alleles of *FTO* in human were identified to increase the risk for hyperactive disorder (Choudhry et al., 2013), for Alzheimer's disease (Keller et al., 2011; Reitz et al., 2012) and to affect the brain volume (Melka et al., 2013; Li et al., 2017a). Molecular and functional studies have shown that *FTO* knockout mice display altered behavior, including locomotion defects, but also influences learning and memory. For instance, fear conditioned mice showed a significant increase in m^{6}A intensity on several neuronal targets, and knockdown of FTO

further enhanced consolidation of cued fear memory (Widagdo et al., 2016). In line with this finding, *FTO* deficiency reduces the proliferation and neuronal differentiation of adult NSCs, which leads to impaired learning and memory (Li et al., 2017a). Electrophysiological tests in *FTO* knockout mice demonstrated an impaired dopamine type 2 and 3 receptor response, resulting in an abnormal response to cocaine (Hess et al., 2013). Strikingly, a recent study in mouse embryonic dorsal root ganglia found that FTO is enriched and specifically expressed in axons, influencing translation of axonal mRNAs (Yu et al., 2018). This demonstrates the dynamic role of m^{6}A modification in regulating local translation. However, it is important to stress that FTO was also demonstrated to additionally demethylate m^{6}Am, which is present next to the 7mG cap modification (Mauer et al., 2017). In comparison to m^{6}A, m^{6}Am also contains a methyl group on the ribose. As available antibodies recognize both modifications indistinguishably, it is currently difficult to assign their respective contribution in the context of brain activity.
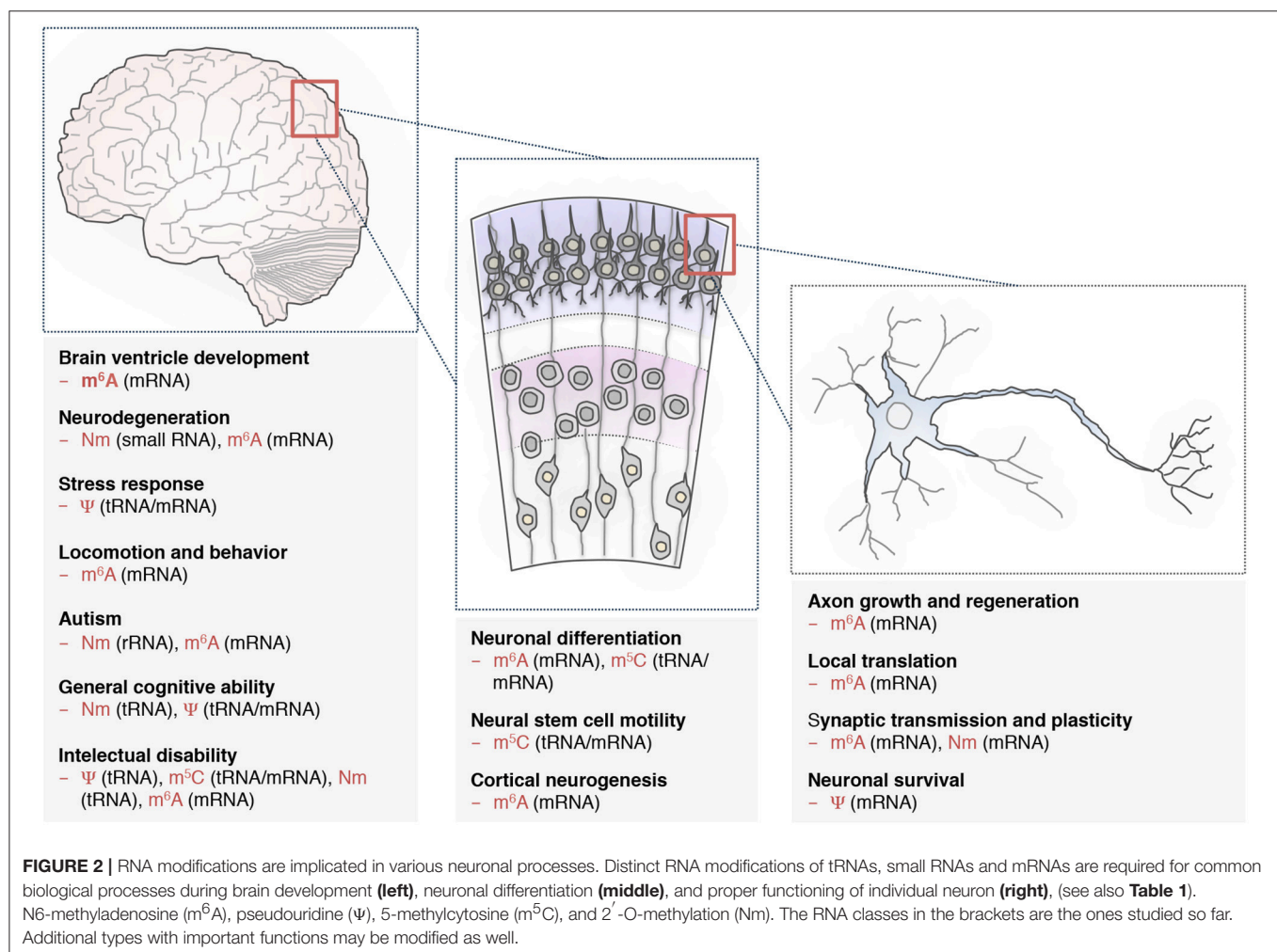
CONCLUSION

The association of aberrant RNA modifications with various neurological disorders highlights the importance of these chemical moieties for proper brain development and cognition. However, today the role of RNA modifications in these processes is not completely understood. One of the current challenges lies in identifying the class and identity of RNAs that are targeted by RNA modification enzymes and are causative of the neurological defects. Common targets for many of these enzymes are tRNAs and rRNAs, thus it is likely that in many cases their dysfunction plays an important role in the etiology of the disease. Yet, this does not explain why the phenotypes observed upon mutations of these enzymes are often restricted to the brain. Some of these enzymes or snoRNAs are predominantly expressed in the nervous system, which indicates their importance in this tissue and suggests the existence of differentially modified ribosomes (also recently called specialized ribosomes) that may carry distinct functions (Briggs and Dinman, 2017; Sloan et al., 2017). However, other snoRNA or enzymes exhibit wider tissue distribution, suggesting that brain specific phenotypes may reflect a higher sensitivity to altered translation in this tissue compared to others organs. This would be consistent with other disease-causing mutations in ribosomal proteins or tRNA synthetase genes that also manifest their effect specifically in the nervous system (Antonellis et al., 2003; Antonellis and Green, 2008; Yao and Fox, 2013; Brooks et al., 2014). What could be the reason(s) behind this increased sensitivity? It has been observed that tRNA level is in general higher in the brain compared to other tissues, suggestive of a bigger translational demand (Dittmar et al., 2006). This neuronal specificity may arise from the local translation that occurs at synapses upon environmental changes. In this context, RNA modifications represent an attractive system to regulate this acute need in a dynamic and flexible manner. However, other examples point toward a translation-independent mechanism. For instance, in the case of *FTSJ1* mutations and the associated NSXLID, neither

the amount, nor the charging of the concerned tRNAs appear to be affected (Guy et al., 2012, 2015), seemingly ruling out a general translation defect. Perhaps in this case, the absence of modification can stimulate tRNA cleavage and generate tRFs that can also interfere with translation. This increase in tRFs would not necessarily be associated with a corresponding reduction of the uncleaved tRNA since tRNA levels are tightly regulated (Wilusz, 2015). Alternatively, the absence of modification on the tRNA may affect its interaction with the ribosome and thus influence translation efficiency or fidelity. Therefore, careful examination at different molecular levels is required to appreciate the effect of tRNA and rRNA modification enzyme mutations on translation and their consequence on the neurological phenotype.

That being said, the general picture is probably more complex. For instance, recent reports show that tRFs not only interfere with translation but can also affect transposon regulation and genome stability (Durdevic et al., 2013; Martinez et al., 2017; Schorn et al., 2017; Zhang et al., 2017). This activity could in principle also contribute to neurological disorders as growing evidence suggests associations between (re)expression

of transposable elements and the occurrence of neuropathies (Perrat et al., 2013; Krug et al., 2017; Zahn, 2017; Jacob-Hirsch et al., 2018). In addition, beyond tRNA and tRF, the recent studies on m^{6}A clearly demonstrate the involvement of mRNA modification in different aspects of neuronal development and regulation. The large diversity of RNA processing events in the brain, including the high rates of alternative splicing and recursive splicing (Duff et al., 2015; Sibley et al., 2015), the inclusion of microexons (Irimia et al., 2014) and the biogenesis of circular RNAs (cirRNAs) can all in principle be affected by RNA modifications. For instance, m^{6}A modification was recently found on circRNAs, and enables their translation (Yang et al., 2017; Zhou et al., 2017). Given that misexpression of circRNAs has been associated with neurological disorders (Shao and Chen, 2016; van Rossum et al., 2016; Li et al., 2017b), some of m^{6}A brain functions may rely on circRNA-mediated regulation. Thus, because neurons face distinct challenges with regards to localisation of RNAs to distal processes and to localized translation it is likely that in the brain, more than in any other tissues, a combinatorial effect of RNA modifications on different classes of RNAs represents a critical informational

FIGURE 2 | RNA modifications are implicated in various neuronal processes. Distinct RNA modifications of tRNAs, small RNAs and mRNAs are required for common biological processes during brain development **(left)**, neuronal differentiation **(middle)**, and proper functioning of individual neuron **(right)**, (see also **Table 1**). N6-methyladenosine (m^{6}A), pseudouridine (Ψ), 5-methylcytosine (m^{5}C), and 2′-O-methylation (Nm). The RNA classes in the brackets are the ones studied so far. Additional types with important functions may be modified as well.

layer that dynamically fine-tunes gene regulation (**Figure 2**). Exciting discoveries are lying ahead for deciphering this intricate epitranscriptomics code.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

REFERENCES

Abbasi-Moheb, L., Mertel, S., Gonsior, M., Nouri-Vahid, L., Kahrizi, K., Cirak, S., et al. (2012). Mutations in NSUN2 cause autosomal-recessive intellectual disability. *Am. J. Hum. Genet.* 90, 847–855. doi: 10.1016/j.ajhg.2012.03.021

Abe, M., Naqvi, A., Hendriks, G. J., Feltzin, V., Zhu, Y., Grigoriev, A., et al. (2014). Impact of age-associated increase in 2′-O-methylation of miRNAs on aging and neurodegeneration in Drosophila. *Genes Dev.* 28, 44–57. doi: 10.1101/gad.226654.113

Agris, P. F. (2008). Bringing order to translation, the contributions of transfer RNA anticodon-domain modifications. *EMBO Rep.* 9, 629–635. doi: 10.1038/embor.2008.104

Antar, L. N., Li, C., Zhang, H., Carroll, R. C., and Bassell, G. J. (2006). Local functions for FMRP in axon growth cone motility and activity-dependent regulation of filopodia and spine synapses. *Mol. Cell. Neurosci.* 32, 37–48. doi: 10.1016/j.mcn.2006.02.001

Antonellis, A., and Green, E. D. (2008). The role of aminoacyl-tRNA synthetases in genetic diseases. *Annu. Rev. Genomics Hum. Genet.* 9, 87–107. doi: 10.1146/annurev.genom.9.081307.164204

Antonellis, A., Ellsworth, R. E., Sambuughin, N., Puls, I., Abel, A., Lee-Lin, S. Q., et al. (2003). Glycyl tRNA synthetase mutations in Charcot-Marie-Tooth disease type 2D and distal spinal muscular atrophy type V. *Am. J. Hum. Genet.* 72, 1293–1299. doi: 10.1086/375039

Arnez, J. G., and Steitz, T. A. (1994). Crystal-structure of unmodified Trna(Gln) complexed with glutaminyl-transfer-rna synthetase and atp suggests a possible role for pseudo-uridines in stabilization of Rna structure. *Biochemistry* 33, 7560–7567. doi: 10.1021/bi00190a008

Ascano, M., Mukherjee, N., Bandaru, P., Miller, J. B., Nusbaum, J. D., Corcoran, D. L., et al. (2012). FMRP targets distinct mRNA sequence elements to regulate protein expression. *Nature* 492, 382–386. doi: 10.1038/nature11737

Bardoni, B., Schenck, A., and Mandel, J. L. (2001). The fragile X mental retardation protein. *Brain Res. Bull.* 56, 375–382. doi: 10.1016/S0361-9230(01)00647-5

Batista, P. J. (2017). The RNA modification N(6)-methyladenosine and Its implications in human disease. *Genomics Proteomics Bioinformatics* 15, 154–163. doi: 10.1016/j.gpb.2017.03.002

Batista, P. J., Molinie, B., Wang, J., Qu, K., Zhang, J., Li, L., et al. (2014). m(6)A RNA modification controls cell fate transition in mammalian embryonic stem cells. *Cell Stem Cell* 15, 707–719. doi: 10.1016/j.stem.2014.09.019

Bechara, E. G., Didiot, M. C., Melko, M., Davidovic, L., Bensaid, M., Martin, P., et al. (2009). A novel function for fragile x mental retardation protein in translational activation. *PLoS Biol.* 7:e16. doi: 10.1371/journal.pbio.1000016

Bednárová, A., Hanna, M., Durham, I., VanCleave, T., England, A., Chaudhuri, A., et al. (2017). Lost in translation: defects in transfer rna modifications and neurological disorders. *Front. Mol. Neurosci.* 10:135. doi: 10.3389/fnmol.2017.00135

Belmonte, M. K., Cook, E. H., Anderson, G. M., Rubenstein, J. L. R., Greenough, W. T., Beckel-Mitchener, A., et al. (2004). Autism as a disorder of neural information processing: directions for research and targets for therapy. *Mol. Psychiatry* 9, 646–663. doi: 10.1038/sj.mp.4001499

Bergmann, A. K., Campagna, D. R., McLoughlin, E. M., Agarwal, S., Fleming, M. D., Bottomley, S. S., et al. (2010). Systematic molecular genetic analysis of congenital sideroblastic anemia: evidence for genetic heterogeneity and identification of novel mutations. *Pediatr. Blood Cancer* 54, 273–278. doi: 10.1002/pbc.22244

Bird, A. (2002). DNA methylation patterns and epigenetic memory. *Genes Dev.* 16, 6–21. doi: 10.1101/gad.947102

Blanco, S., Dietmann, S., Flores, J. V., Hussain, S., Kutter, C., Humphreys, P., et al. (2014). Aberrant methylation of tRNAs links cellular stress to neuro-developmental disorders. *EMBO J.* 33, 2020–2039. doi: 10.15252/embj.201489282

Blanco, S., Kurowski, A., Nichols, J., Watt, F. M., Benitah, S. A., and Frye, M. (2011). The RNA-methyltransferase Misu (NSun2) poises epidermal stem cells to differentiate. *PLoS Genet.* 7:e1002403. doi: 10.1371/journal.pgen.1002403

Bolton, P. F., Veltman, M. W., Weisblatt, E., Holmes, J. R., Thomas, N. S., Youings, S. A., et al. (2004). Chromosome 15q11-13 abnormalities and other medical conditions in individuals with autism spectrum disorders. *Psychiatr. Genet* 14, 131–137. doi: 10.1097/00041444-200409000-00002

Briggs, J. W., and Dinman, J. D. (2017). Subtractional heterogeneity: a crucial step toward defining specialized ribosomes. *Mol. Cell* 67, 3–4. doi: 10.1016/j.molcel.2017.06.022

Brooks, S. S., Wall, A. L., Golzio, C., Reid, D. W., Kondyles, A., Willer, J. R., et al. (2014). A novel ribosomopathy caused by dysfunction of RPL10 disrupts neurodevelopment and causes X-linked microcephaly in humans. *Genetics* 198, 723–733. doi: 10.1534/genetics.114.168211

Brzezicha, B., Schmidt, M., Makalowska, I., Jarmolowski, A., Pienkowska, J., and Szweykowska-Kulinska, Z. (2006). Identification of human tRNA:m5C methyltransferase catalysing intron-dependent m5C formation in the first position of the anticodon of the pre-tRNA Leu (CAA). *Nucleic Acids Res.* 34, 6034–6043. doi: 10.1093/nar/gkl765

Bykhovskaya, Y., Casas, K., Mengesha, E., Inbal, A., and Fischel-Ghodsian, N. (2004). Missense mutation in pseudouridine synthase 1 (PUS1) causes mitochondrial myopathy and sideroblastic anemia (MLASA). *Am. J. Hum. Genet.* 74, 1303–1308. doi: 10.1086/421530

Byszewska, M., Smietanski, M., Purta, E., and Bujnicki, J. M. (2014). RNA methyltransferases involved in 5′ cap biosynthesis. *RNA Biol.* 11, 1597–1607. doi: 10.1080/15476286.2015.1004955

Cantara, W. A., Crain, P. F., Rozenski, J., McCloskey, J. A., Harris, K. A., Zhang, X., et al. (2011). The RNA modification database, RNAMDB: 2011 update. *Nucleic Acids Res.* 39, D195–D201. doi: 10.1093/nar/gkq1028

Cao, M., Dona, M., Valentino, M. L., Semplicini, C., Maresca, A., Cassina, M., et al. (2016). Clinical and molecular study in a long-surviving patient with MLASA syndrome due to novel PUS1 mutations. *Neurogenetics* 17, 143–143. doi: 10.1007/s10048-016-0475-3

Carlile, T. M., Rojas-Duran, M. F., and Gilbert, W. V. (2015). Transcriptome-wide identification of pseudouridine modifications using Pseudo-seq. *Curr. Protoc. Mol. Biol.* 112, 4. 25. 21-24. doi: 10.1002/0471142727.mb0425s112

Carlile, T. M., Rojas-Duran, M. F., Zinshteyn, B., Shin, H., Bartoli, K. M., and Gilbert, W. V. (2014). Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature* 515, 143–146. doi: 10.1038/nature13802

Cavaillé, J., Buiting, K., Kiefmann, M., Lalande, M., Brannan, C. I., Horsthemke, B., et al. (2000). Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization. *Proc. Natl. Acad. Sci. U.S.A.* 97, 14311–14316. doi: 10.1073/pnas.250426397

Cavaillé, J., Nicoloso, M., and Bachellerie, J. P. (1996). Targeted ribose methylation of RNA *in vivo* directed by tailored antisense RNA guides. *Nature* 383, 732–735. doi: 10.1038/383732a0

Chang, M., Lv, H., Zhang, W., Ma, C., He, X., Zhao, S., et al. (2017). Region-specific RNA m(6)A methylation represents a new layer of control in the gene regulatory network in the mouse brain. *Open Biol.* 7:170166. doi: 10.1098/rsob.170166

Cho, D. H., and Tapscott, S. J. (2007). Myotonic dystrophy: emerging mechanisms for DM1 and DM2. *Biochim Biophys. Acta* 1772, 195–204. doi: 10.1016/j.bbadis.2006.05.013

Choudhry, Z., Sengupta, S. M., Grizenko, N., Thakur, G. A., Fortier, M. E., Schmitz, N., et al. (2013). Association between obesity-related gene FTO and ADHD. *Obesity* 21, E738–E744. doi: 10.1002/oby.20444

Cohn, W. E., and Volkin, E. (1951). Nucleoside-5′-phosphates from ribonucleic acid. *Nature* 167, 483–484. doi: 10.1038/167483a0

Cook, E. H., and Scherer, S. W. (2008). Copy-number variations associated with neuropsychiatric conditions. *Nature* 455, 919–923. doi: 10.1038/nature07458

Dai, D., Wang, H., Zhu, L., Jin, H., and Wang, X. (2018). N6-methyladenosine links RNA metabolism to cancer progression. *Cell Death Dis.* 9:124. doi: 10.1038/s41419-017-0129-x

Dai, L., Xing, L., Gong, P., Zhang, K., Gao, X., Zheng, Z., et al. (2008). Positive association of the FTSJ1 gene polymorphisms with nonsyndromic X-linked mental retardation in young Chinese male subjects. *J. Hum. Genet.* 53, 592–597. doi: 10.1007/s10038-008-0287-x

Dai, Q., Moshitch-Moshkovitz, S., Han, D., Kol, N., Amariglio, N., Rechavi, G., et al. (2017). Nm-seq maps 2'-O-methylation sites in human mRNA with base precision. *Nat. Methods* 14, 695–698. doi: 10.1038/nmeth.4294

Daily, D. K., Ardinger, H. H., and Holmes, G. E. (2000). Identification and evaluation of mental retardation. *Am. Fam. Physician* 61, 1059–1067. Available online at: https://www.aafp.org/afp/2000/0215/p1059.html

Darzacq, X., Jády, B. E., Verheggen, C., Kiss, A. M., Bertrand, E., and Kiss, T. (2002). Cajal body-specific small nuclear RNAs: a novel class of 2'-O-methylation and pseudouridylation guide RNAs. *EMBO J.* 21, 2746–2756. doi: 10.1093/emboj/21.11.2746

Davis, D. R. (1995). Stabilization of RNA stacking by pseudouridine. *Nucleic Acids Res.* 23, 5020–5026. doi: 10.1093/nar/23.24.5020

Delorimier, E., Hinman, M. N., Copperman, J., Datta, K., Guenza, M., and Berglund, J. A. (2017). Pseudouridine modification inhibits muscleblind-like 1 (MBNL1) binding to CCUG repeats and minimally structured RNA through reduced RNA flexibility. *J. Biol. Chem.* 292, 4350–4357. doi: 10.1074/jbc.M116.770768

Didiot, M. C., Tian, Z., Schaeffer, C., Subramanian, M., Mandel, J. L., and Moine, H. (2008). The G-quartet containing FMRP binding site in FMR1 mRNA is a potent exonic splicing enhancer. *Nucleic Acids Res.* 36, 4902–4912. doi: 10.1093/nar/gkn472

Diez-Roux, G., Banfi, S., Sultan, M., Geffers, L., Anand, S., Rozado, D., et al. (2011). A high-resolution anatomical atlas of the transcriptome in the mouse embryo. *PLoS Biol.* 9:e1000582. doi: 10.1371/journal.pbio.1000582

Dittmar, K. A., Goodenbour, J. M., and Pan, T. (2006). Tissue-specific differences in human transfer RNA expression. *PLoS Genet.* 2:e221. doi: 10.1371/journal.pgen.0020221

Doe, C. M., Relkovic, D., Garfield, A. S., Dalley, J. W., Theobald, D. E. H., Humby, T., et al. (2009). Loss of the imprinted snoRNA mbii-52 leads to increased 5htr2c pre-RNA editing and altered 5HT(2C)R-mediated behaviour. *Hum. Mol. Genet.* 18, 2140–2148. doi: 10.1093/hmg/ddp137

Dominissini, D., and Rechavi, G. (2017). Loud and clear epitranscriptomic m(1)A signals: now in single-base resolution. *Mol. Cell* 68, 825–826. doi: 10.1016/j.molcel.2017.11.029

Dominissini, D., Moshitch-Moshkovitz, S., Schwartz, S., Salmon-Divon, M., Ungar, L., Osenberg, S., et al. (2012). Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* 485, 201–206. doi: 10.1038/nature11112

Dominissini, D., Nachtergaele, S., Moshitch-Moshkovitz, S., Peer, E., Kol, N., Ben-Haim, M. S., et al. (2016). The dynamic N(1)-methyladenosine methylome in eukaryotic messenger RNA. *Nature* 530, 441–446. doi: 10.1038/nature16998

Du, T., Rao, S., Wu, L., Ye, N., Liu, Z., Hu, H., et al. (2015). An association study of the m6A genes with major depressive disorder in Chinese Han population. *J. Affect. Disord.* 183, 279–286. doi: 10.1016/j.jad.2015.05.025

Duan, J., Li, L., Lu, J., Wang, W., and Ye, K. (2009). Structural mechanism of substrate RNA recruitment in H/ACA RNA-guided pseudouridine synthase. *Mol. Cell* 34, 427–439. doi: 10.1016/j.molcel.2009.05.005

Duff, M. O., Olson, S., Wei, X., Garrett, S. C., Osman, A., Bolisetty, M., et al. (2015). Genome-wide identification of zero nucleotide recursive splicing in Drosophila. *Nature* 521, 376–379. doi: 10.1038/nature14475

Duker, A. L., Ballif, B. C., Bawle, E. V., Person, R. E., Mahadevan, S., Alliman, S., et al. (2010). Paternally inherited microdeletion at 15q11.2 confirms a significant role for the SNORD116 C/D box snoRNA cluster in Prader-Willi syndrome. *Eur. J. Hum. Genet.* 18, 1196–1201. doi: 10.1038/ejhg.2010.102

Durdevic, Z., Mobin, M. B., Hanna, K., Lyko, F., and Schaefer, M. (2013). The RNA methyltransferase Dnmt2 is required for efficient Dicer-2-dependent siRNA pathway activity in Drosophila. *Cell Rep.* 4, 931–937. doi: 10.1016/j.celrep.2013.07.046

Edupuganti, R. R., Geiger, S., Lindeboom, R. G. H., Shi, H., Hsu, P. J., Lu, Z., et al. (2017). N(6)-methyladenosine (m(6)A) recruits and repels proteins to regulate mRNA homeostasis. *Nat. Struct. Mol. Biol.* 24, 870–878. doi: 10.1038/nsmb.3462

Fähling, M., Mrowka, R., Steege, A., Kirschner, K. M., Benko, E., Förstera, B., et al. (2009). Translational regulation of the human achaete-scute homologue-1 by fragile X mental retardation protein. *J. Biol. Chem.* 284, 4255–4266. doi: 10.1074/jbc.M807354200

Falaleeva, M., Welden, J. R., Duncan, M. J., and Stamm, S. (2017). C/D-box snoRNAs form methylating and non-methylating ribonucleoprotein complexes: old dogs show new tricks. *BioEssays* 39:1600264. doi: 10.1002/bies.201600264

Fernandez-Funez, P., Nino-Rosales, M. L., de Gouyon, B., She, W. C., Luchak, J. M., Martinez, P., et al. (2000). Identification of genes that modify ataxin-1-induced neurodegeneration. *Nature* 408, 101–106. doi: 10.1038/35040584

Fernandez-Vizarra, E., Berardinelli, A., Valente, L., Tiranti, V., and Zeviani, M. (2007). Nonsense mutation in pseudouridylate synthase 1 (PUS1) in two brothers affected by myopathy, lactic acidosis and sideroblastic anaemia (MLASA). *J. Med. Genet.* 44, 173–180. doi: 10.1136/jmg.2006.045252

Fisher, O., Siman-Tov, R., and Ankri, S. (2006). Pleiotropic phenotype in Entamoeba histolytica overexpressing DNA methyltransferase (Ehmeth). *Mol. Biochem. Parasitol.* 147, 48–54. doi: 10.1016/j.molbiopara.2006.01.007

Fitzek, E., Joardar, A., Gupta, R., and Geisler, M. (2018). Evolution of eukaryal and archaeal pseudouridine synthase pus10. *J. Mol. Evol.* 86, 77–89. doi: 10.1007/s00239-018-9827-y

Flores, J. V., Cordero-Espinoza, L., Oeztuerk-Winder, F., Andersson-Rolf, A., Selmi, T., Blanco, S., et al. (2017). Cytosine-5 RNA methylation regulates neural stem cell differentiation and motility. *Stem Cell Rep.* 8, 112–124. doi: 10.1016/j.stemcr.2016.11.014

Foster, T. C. (2012). Role of estrogen receptor alpha and beta expression and signaling on cognitive function during aging. *Hippocampus* 22, 656–669. doi: 10.1002/hipo.20935

Franke, B., Vermeulen, S. H., Steegers-Theunissen, R. P., Coenen, M. J., Schijvenaars, M. M., Scheffer, H., et al. (2009). An association study of 45 folate-related genes in spina bifida: Involvement of cubilin (CUBN) and tRNA aspartic acid methyltransferase 1 (TRDMT1). *Birth Defects Res. A Clin. Mol. Teratol.* 85, 216–226. doi: 10.1002/bdra.20556

Freude, K., Hoffmann, K., Jensen, L. R., Delatycki, M. B., des Portes, V., Moser, B., et al. (2004). Mutations in the FTSJ1 gene coding for a novel S-adenosylmethionine- binding protein cause nonsyndromic X-linked mental retardation. *Am. J. Hum. Genet.* 75, 305–309. doi: 10.1086/422507

Froyen, G., Bauters, M., Boyle, J., Van Esch, H., Govaerts, K., van Bokhoven, H., et al. (2007). Loss of SLC38A5 and FTSJ1 at Xp11.23 in three brothers with non-syndromic mental retardation due to a microdeletion in an unstable genomic region. *Hum. Genet.* 121, 539–547. doi: 10.1007/s00439-007-0343-1

Gamerdinger, M., Manthey, D., and Behl, C. (2006). Oestrogen receptor subtype-specific repression of calpain expression and calpain enzymatic activity in neuronal cells - implications for neuroprotection against Ca^{2+-} mediated excitotoxicity. *J. Neurochem.* 97, 57–68. doi: 10.1111/j.1471-4159.2006.03675.x

Ganot, P., Bortolin, M. L., and Kiss, T. (1997). Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs. *Cell* 89, 799–809. doi: 10.1016/S0092-8674(00)80263-9

Genenncher, B., Durdevic, Z., Hanna, K., Zinkl, D., Mobin, M. B., Senturk, N., et al. (2018). Mutations in cytosine-5 tRNA methyltransferases impact mobile element expression and genome stability at specific DNA repeats. *Cell Rep.* 22, 1861–1874. doi: 10.1016/j.celrep.2018.01.061

Geula, S., Moshitch-Moshkovitz, S., Dominissini, D., Mansour, A. A., Kol, N., Salmon-Divon, M., et al. (2015). m(6)A mRNA methylation facilitates resolution of naive pluripotency toward differentiation. *Science* 347, 1002–1006. doi: 10.1126/science.1261417

Giorda, R., Bonaglia, M. C., Beri, S., Fichera, M., Novara, F., Magini, P., et al. (2009). Complex segmental duplications mediate a recurrent dup(X)(p11.22-p11.23) associated with mental retardation, speech delay, and EEG anomalies in males and females. *Am. J. Hum. Genet.* 85, 394–400. doi: 10.1016/j.ajhg.2009.08.001

Goll, M. G., and Bestor, T. H. (2005). Eukaryotic cytosine methyltransferases. *Annu. Rev. Biochem.* 74, 481–514. doi: 10.1146/annurev.biochem.74.010904.153721

Goll, M. G., Kirpekar, F., Maggert, K. A., Yoder, J. A., Hsieh, C. L., Zhang, X., et al. (2006). Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2. *Science* 311, 395–398. doi: 10.1126/science.1120976

Gong, P., Li, J., Dai, L., Zhang, K., Zheng, Z., Gao, X., et al. (2008). Genetic variations in FTSJ1 influence cognitive ability in young males in the chinese han population. *J. Neurogenet.* 22, 277–287. doi: 10.1080/01677060802337299

Grosjean (2009). *DNA and RNA Modification Enzymes.* Austin, TX: Landes Bioscience. 289–302.

Gu, T., Zhao, T., Kohli, U., and Hewes, R. S. (2017). The large and small SPEN family proteins stimulate axon outgrowth during neurosecretory cell remodeling in Drosophila. *Dev. Biol.* 431, 226–238. doi: 10.1016/j.ydbio.2017.09.013

Guo, J., Tang, H. W., Li, J., Perrimon, N., and Yan, D. (2018). Xio is a component of the *Drosophila* sex determination pathway and RNA N^6-methyladenosine methyltransferase complex. *Proc. Natl. Acad. Sci. U.S.A.* 115, 3674–3679. doi: 10.1073/pnas.1720945115

Guo, M., and Schimmel, P. (2013). Essential nontranslational functions of tRNA synthetases. *Nat. Chem. Biol.* 9, 145–153. doi: 10.1038/nchembio.1158

Guo, W., Polich, E. D., Su, J., Gao, Y., Christopher, D. M., Allan, A. M., et al. (2015). Fragile X proteins FMRP and FXR2P control synaptic GluA1 expression and neuronal maturation via distinct mechanisms. *Cell Rep.* 11, 1651–1666. doi: 10.1016/j.celrep.2015.05.013

Guy, M. P., and Phizicky, E. M. (2015). Conservation of an intricate circuit for crucial modifications of the tRNAPhe anticodon loop in eukaryotes. *RNA* 21, 61–74. doi: 10.1261/rna.047639.114

Guy, M. P., Podyma, B. M., Preston, M. A., Shaheen, H. H., Krivos, K. L., Limbach, P. A., et al. (2012). Yeast Trm7 interacts with distinct proteins for critical modifications of the tRNAPhe anticodon loop. *RNA* 18, 1921–1933. doi: 10.1261/rna.035287.112

Guy, M. P., Shaw, M., Weiner, C. L., Hobson, L., Stark, Z., Rose, K., et al. (2015). Defects in tRNA Anticodon Loop 2'-O-methylation are implicated in nonsyndromic X-linked intellectual disability due to mutations in FTSJ1. *Hum. Mutat.* 36, 1176–1187. doi: 10.1002/humu.22897

Haag, S., Warda, A. S., Kretschmer, J., Günnigmann, M. A., Höbartner, C., and Bohnsack, M. T. (2015). NSUN6 is a human RNA methyltransferase that catalyzes formation of m5C72 in specific tRNAs. *RNA* 21, 1532–1543. doi: 10.1261/rna.051524.115

Hagerman, R. J., and Polussa, J. (2015). Treatment of the psychiatric problems associated with fragile X syndrome. *Curr. Opin. Psychiatry* 28, 107–112. doi: 10.1097/YCO.0000000000000131

Hamel, B. C., Smits, A. P., van den Helm, B., Smeets, D. F., Knoers, N. V., van Roosmalen, T., et al. (1999). Four families (MRX43, MRX44, MRX45, MRX52) with nonspecific X-linked mental retardation: clinical and psychometric data and results of linkage analysis. *Am. J. Med. Genet.* 85, 290–304. doi: 10.1002/(SICI)1096-8628(19990730)85:3<290::AID-AJMG21>3.0.CO;2-H

Hartmann, A. M., Nayler, O., Schwaiger, F. W., Obermeier, A., and Stamm, S. (1999). The interaction and colocalization of Sam68 with the splicing-associated factor YT521-B in nuclear dots is regulated by the Src family kinase p59(fyn). *Mol. Biol. Cell* 10, 3909–3926. doi: 10.1091/mbc.10.11.3909

Haussmann, I. U., Bodi, Z., Sanchez-Moran, E., Mongan, N. P., Archer, N., Fray, R. G., et al. (2016). m(6)A potentiates Sxl alternative pre-mRNA splicing for robust Drosophila sex determination. *Nature* 540, 301–304. doi: 10.1038/nature20577

He, C. (2010). Grand challenge commentary: RNA epigenetics? *Nat. Chem. Biol.* 6, 863–865. doi: 10.1038/nchembio.482

Heiss, N. S., Bächner, D., Salowsky, R., Kolb, A., Kioschis, P., and Poustka, A. (2000). Gene structure and expression of the mouse dyskeratosis congenita gene, Dkc1. *Genomics* 67, 153–163. doi: 10.1006/geno.2000.6227

Heiss, N. S., Knight, S. W., Vulliamy, T. J., Klauck, S. M., Wiemann, S., Mason, P. J., et al. (1998). X-linked dyskeratosis congenita is caused by mutations in a highly conserved gene encoding a protein with putative nucleolar functions. *Nat. Genet.* 19, 32–38. doi: 10.1038/ng0598-32

Helm, M. (2006). Post-transcriptional nucleotide modification and alternative folding of RNA. *Nucleic Acids Res.* 34, 721–733. doi: 10.1093/nar/gkj471

Hermann, A., Schmitt, S., and Jeltsch, A. (2003). The human Dnmt2 has residual DNA-(cytosine-C5) methyltransferase activity. *J. Biol. Chem.* 278, 31717–31721. doi: 10.1074/jbc.M305448200

Hess, M. E., Hess, S., Meyer, K. D., Verhagen, L. A., Koch, L., Brönneke, H. S., et al. (2013). The fat mass and obesity associated gene (Fto) regulates activity of the dopaminergic midbrain circuitry. *Nat. Neurosci.* 16, 1042–1048. doi: 10.1038/nn.3449

Higa-Nakamine, S., Suzuki, T., Uechi, T., Chakraborty, A., Nakajima, Y., Nakamura, M., et al. (2012). Loss of ribosomal RNA modification causes developmental defects in zebrafish. *Nucleic Acids Res.* 40, 391–398. doi: 10.1093/nar/gkr700

Honda, S., Hayashi, S., Imoto, I., Toyama, J., Okazawa, H., Nakagawa, E., et al. (2010). Copy-number variations on the X chromosome in Japanese patients with mental retardation detected by array-based comparative genomic hybridization analysis. *J. Hum. Genet.* 55, 590–599. doi: 10.1038/jhg.2010.74

Horwich, M. D., Li, C., Matranga, C., Vagin, V., Farley, G., Wang, P., et al. (2007). The Drosophila RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC. *Curr. Biol.* 17, 1265–1272. doi: 10.1016/j.cub.2007.06.030

Hussain, S., Sajini, A. A., Blanco, S., Dietmann, S., Lombard, P., Sugimoto, Y., et al. (2013). NSun2-mediated cytosine-5 methylation of vault noncoding RNA determines its processing into regulatory small RNAs. *Cell Rep.* 4, 255–261. doi: 10.1016/j.celrep.2013.06.029

Iijima, T., Wu, K., Witte, H., Hanno-Iijima, Y., Glatter, T., Richard, S., et al. (2011). SAM68 regulates neuronal activity-dependent alternative splicing of neurexin-1. *Cell* 147, 1601–1614. doi: 10.1016/j.cell.2011.11.028

Irimia, M., Weatheritt, R. J., Ellis, J. D., Parikshak, N. N., Gonatopoulos-Pournatzis, T., Babor, M., et al. (2014). A highly conserved program of neuronal microexons is misregulated in autistic brains. *Cell* 159, 1511–1523. doi: 10.1016/j.cell.2014.11.035

Jacob-Hirsch, J., Eyal, E., Knisbacher, B. A., Roth, J., Cesarkas, K., Dor, C., et al. (2018). Whole-genome sequencing reveals principles of brain retrotransposition in neurodevelopmental disorders. *Cell Res.* 28, 187–203. doi: 10.1038/cr.2018.8

Jia, G., Fu, Y., Zhao, X., Dai, Q., Zheng, G., Yang, Y., et al. (2011). N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat. Chem. Biol.* 7, 885–887. doi: 10.1038/nchembio.687

Jurkowski, T. P., Meusburger, M., Phalke, S., Helm, M., Nellen, W., Reuter, G., et al. (2008). Human DNMT2 methylates tRNA(Asp) molecules using a DNA methyltransferase-like catalytic mechanism. *RNA* 14, 1663–1670. doi: 10.1261/rna.970408

Kan, L., Grozhik, A. V., Vedanayagam, J., Patil, D. P., Pang, N., Lim, K. S., Huang, Y. C., et al. (2017). The m(6)A pathway facilitates sex determination in Drosophila. *Nat. Commun.* 8:15737. doi: 10.1038/ncomms15737

Kaya, Y., and Ofengand, J. (2003). A novel unanticipated type of pseudouridine synthase with homologs in bacteria, archaea, and eukarya. *RNA* 9, 711–721. doi: 10.1261/rna.5230603

Keller, L., Xu, W., Wang, H. X., Winblad, B., Fratiglioni, L., and Graff, C. (2011). The obesity related gene, FTO, interacts with APOE, and is associated with Alzheimer's disease risk: a prospective cohort study. *J. Alzheimer's Dis.* 23, 461–469. doi: 10.3233/JAD-2010-101068.

Khan, M. A., Rafiq, M. A., Noor, A., Hussain, S., Flores, J. V., Rupp, V., et al. (2012). Mutation in NSUN2, which encodes an RNA methyltransferase, causes autosomal-recessive intellectual disability. *Am. J. Hum. Genet.* 90, 856–863. doi: 10.1016/j.ajhg.2012.03.023

Khoddami, V., and Cairns, B. R. (2013). Identification of direct targets and modified bases of RNA cytosine methyltransferases. *Nat. Biotechnol.* 31, 458–464. doi: 10.1038/nbt.2566

Kishore, S., and Stamm, S. (2006). The snoRNA HBII-52 regulates alternative splicing of the serotonin receptor 2C. *Science* 311, 230–232. doi: 10.1126/science.1118265

Knuckles, P., Lence, T., Haussmann, I. U., Jacob, D., Kreim, N., Carl, S. H., Masiello, I., et al. (2018). Zc3h13/Flacc is required for adenosine methylation by bridging the mRNA-binding factor Rbm15/Spenito to the m(6)A machinery component Wtap/Fl(2)d. *Genes Dev.* 32, 415–429. doi: 10.1101/gad.309146.117

Koonin, E. V. (1996). Pseudouridine synthases: four families of enzymes containing a putative uridine-binding motif also conserved in dUTPases and dCTP deaminases. *Nucleic Acids Res.* 24, 2411–2415. doi: 10.1093/nar/24.12.2411

Krug, L., Chatterjee, N., Borges-Monroy, R., Hearn, S., Liao, W. W., Morrill, K., et al. (2017). Retrotransposon activation contributes to neurodegeneration in a Drosophila TDP-43 model of ALS. *PLoS Genet.* 13:e1006635. doi: 10.1371/journal.pgen.1006635

Kuchino, Y., Borek, E., Grunberger, D., Mushinski, J. F., and Nishimura, S. (1982). Changes of post-transcriptional modification of wye base in tumor-specific tRNAPhe. *Nucleic Acids Res.* 10, 6421–6432. doi: 10.1093/nar/10.20.6421

Kumar, S., Mapa, K., and Maiti, S. (2014). Understanding the effect of locked nucleic acid and 2'-O-methyl modification on the hybridization thermodynamics of a miRNA-mRNA pair in the presence and absence of AfPiwi protein. *Biochemistry* 53, 1607–1615. doi: 10.1021/bi401677d

Kunert, N., Marhold, J., Stanke, J., Stach, D., and Lyko, F. (2003). A Dnmt2-like protein mediates DNA methylation in Drosophila. *Development* 130, 5083–5090. doi: 10.1242/dev.00716

Kurth, H. M., and Mochizuki, K. (2009). 2'-O-methylation stabilizes Piwi-associated small RNAs and ensures DNA elimination in Tetrahymena. *RNA* 15, 675–685. doi: 10.1261/rna.1455509

Lacoux, C., Di Marino, D., Boyl, P. P., Zalfa, F., Yan, B., Ciotti, M. T., et al. (2012). BC1-FMRP interaction is modulated by 2'-O-methylation: RNA-binding activity of the tudor domain and translational regulation at synapses. *Nucleic Acids Res.* 40, 4086–4096. doi: 10.1093/nar/gkr1254

Lafontaine, D. L., Bousquet-Antonelli, C., Henry, Y., Caizergues-Ferrer, M., and Tollervey, D. (1998). The box H + ACA snoRNAs carry Cbf5p, the putative rRNA pseudouridine synthase. *Genes Dev.* 12, 527–537. doi: 10.1101/gad.12.4.527

Lee, S. H., Kim, I., and Chung, B. C. (2007). Increased urinary level of oxidized nucleosides in patients with mild-to-moderate Alzheimer's disease. *Clin. Biochem.* 40, 936–938. doi: 10.1016/j.clinbiochem.2006.11.021

Legrand, C., Tuorto, F., Hartmann, M., Liebers, R., Jacob, D., Helm, M., et al. (2017). Statistically robust methylation calling for whole-transcriptome bisulfite sequencing reveals distinct methylation patterns for mouse RNAs. *Genome Res.* 27, 1589–1596. doi: 10.1101/gr.210666.116

Lence, T., Akhtar, J., Bayer, M., Schmid, K., Spindler, L., Ho, C. H., et al. (2016). m(6)A modulates neuronal functions and sex determination in Drosophila. *Nature* 540, 242–247. doi: 10.1038/nature20568

Lence, T., Soller, M., and Roignant, J. Y. (2017). A fly view on the roles and mechanisms of the m(6)A mRNA modification and its players. *RNA Biol.* 14, 1232–1240. doi: 10.1080/15476286.2017.1307484

Leschziner, G. D., Coffey, A. J., Andrew, T., Gregorio, S. P., Dias-Neto, E., Calafato, M., et al. (2011). Q8IYL2 is a candidate gene for the familial epilepsy syndrome of partial epilepsy with pericentral spikes (PEPS). *Epilepsy Res.* 96, 109–115. doi: 10.1016/j.eplepsyres.2011.05.010

Leygue (2007). Steroid receptor RNA activator (SRA1): unusual bifaceted gene products with suspected relevance to breast cancer. *Nucl. Recept. Signal.* 5:e006. doi: 10.1621/nrs.05006

Li, J., Yang, Z., Yu, B., Liu, J., and Chen, X. (2005). Methylation protects miRNAs and siRNAs from a 3'-end uridylation activity in Arabidopsis. *Curr. Biol.* 15, 1501–1507. doi: 10.1016/j.cub.2005.07.029

Li, L., Zang, L., Zhang, F., Chen, J., Shen, H., Shu, L., et al. (2017a). Fat mass and obesity-associated (FTO) protein regulates adult neurogenesis. *Hum. Mol. Genet.* 26, 2398–2411. doi: 10.1093/hmg/ddx128

Li, T. R., Jia, Y. J., Wang, Q., Shao, X. Q., and Lv, R. J. (2017b). Circular RNA: a new star in neurological diseases. *Int. J. Neurosci.* 127, 726–734. doi: 10.1080/00207454.2016.1236382

Li, W., Prazak, L., Chatterjee, N., Grüninger, S., Krug, L., Theodorou, D., et al. (2013). Activation of transposable elements during aging and neuronal decline in Drosophila. *Nat Neurosci* 16, 529–531. doi: 10.1038/nn.3368

Li, X., Zhu, P., Ma, S. , Song, J., Bai, J., Sun, F., et al. (2015). Chemical pulldown reveals dynamic pseudouridylation of the mammalian transcriptome. *Nat. Chem. Biol.* 11, 592–597. doi: 10.1038/nchembio.1836

Li, X., Xiong, X., Wang, K., Wang, L., Shu, X., Ma, S., et al. (2016). Transcriptome-wide mapping reveals reversible and dynamic N(1)-methyladenosine methylome. *Nat. Chem. Biol.* 12, 311–316. doi: 10.1038/nchembio.2040

Li, X., Xiong, X., Zhang, M., Wang, K., Chen, Y., Zhou, J., et al. (2017c). Base-resolution mapping reveals distinct m(1)A methylome in nuclear- and mitochondrial-encoded transcripts. *Mol. Cell* 68, 993.e9–1005. e9. doi: 10.1016/j.molcel.2017.10.019

Liang, B., Zhou, J., Kahen, E., Terns, R. M., Terns, M. P., and Li, H. (2009). Structure of a functional ribonucleoprotein pseudouridine synthase bound to a substrate RNA. *Nat. Struct. Mol. Biol.* 16, 740–785. doi: 10.1038/nsmb.1624

Lin, M. J., Tang, L. Y., Reddy, M. N., and Shen, C. K. (2005). DNA methyltransferase gene dDnmt2 and longevity of Drosophila. *J. Biol. Chem.* 280, 861–864. doi: 10.1074/jbc.C400477200

Liu, J. Z., Yue, Y. N., Han, D. L., Wang, X., Fu, Y., Zhang, L., et al. (2014). A METTL3-METTL14 complex mediates mammalian nuclear RNA N-6-adenosine methylation. *Nat. Chem. Biol.* 10, 93–95. doi: 10.1038/nchembio.1432

Lovejoy, A. F., Riordan, D. P., and Brown, P. O. (2014). Transcriptome-wide mapping of pseudouridines: pseudouridine synthases modify specific mRNAs in S. cerevisiae. *PLoS ONE* 9:e110799. doi: 10.1371/journal.pone.0110799

Lubs, H. A., Stevenson, R. E., and Schwartz, C. E. (2012). Fragile X and X-linked intellectual disability: four decades of discovery. *Am. J. Hum. Genet.* 90, 579–590. doi: 10.1016/j.ajhg.2012.02.018

Luo, S., and Tong, L. (2014). Molecular basis for the recognition of methylated adenines in RNA by the eukaryotic YTH domain. *Proc. Natl. Acad. Sci. U.S.A.* 111, 13834–13839. doi: 10.1073/pnas.1412742111

Machnicka, M. A., Milanowska, K., Osman Oglou, O., Purta, E., Kurkowska, M., Olchowik, A., et al. (2013). MODOMICS: a database of RNA modification pathways−2013 update. *Nucleic Acids Res.* 41, D262–D267. doi: 10.1093/nar/gks1007

Martinez, F. J., Lee, J. H., Lee, J. E., Blanco, S., Nickerson, E., Gabriel, S., et al. (2012). Whole exome sequencing identifies a splicing mutation in NSUN2 as a cause of a Dubowitz-like syndrome. *J. Med. Genet.* 49, 380–385. doi: 10.1136/jmedgenet-2011-100686

Martinez, G., Choudury, S. G., and Slotkin, R. K. (2017). tRNA-derived small RNAs target transposable element transcripts. *Nucleic Acids Res.* 45, 5142–5152. doi: 10.1093/nar/gkx103

Mauer, J., Luo, X., Blanjoie, A., Jiao, X., Grozhik, A. V., Patil, D. P., et al. (2017). Reversible methylation of m(6)Am in the 5' cap controls mRNA stability. *Nature* 541, 371–375. doi: 10.1038/nature21022

Melka, M. G., Gillis, J., Bernard, M., Abrahamowicz, M., Chakravarty, M. M., Leonard, G. T., et al. (2013). FTO, obesity and the adolescent brain. *Hum. Mol. Genet.* 22, 1050–1058. doi: 10.1093/hmg/dds504

Meola, G., and Cardani, R. (2015). Myotonic dystrophies: an update on clinical aspects, genetic, pathology, and molecular pathomechanisms. *Biochim. Biophys. Acta* 1852, 594–606. doi: 10.1016/j.bbadis.2014.05.019

Metodiev, M. D., Spåhr, H., Loguercio Polosa, P., Meharg, C., Becker, C., Altmueller, J., et al. (2014). NSUN4 is a dual function mitochondrial protein required for both methylation of 12S rRNA and coordination of mitoribosomal assembly. *PLoS Genet.* 10:e1004110. doi: 10.1371/journal.pgen.1004110

Meyer, K. D., and Jaffrey, S. R. (2017). Rethinking m(6)A readers, writers, and erasers. *Annu. Rev. Cell Dev. Biol.* 33, 319–342. doi: 10.1146/annurev-cellbio-100616-060758

Meyer, K. D., Saletore, Y., Zumbo, P., Elemento, O., Mason, C. E., and Jaffrey, S. R. (2012). Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell* 149, 1635–1646. doi: 10.1016/j.cell.2012.05.003

Minnerop, M., Weber, B., Schoene-Bake, J. C., Roeske, S., Mirbach, S., Anspach, C., et al. (2011). The brain in myotonic dystrophy 1 and 2: evidence for a predominant white matter disease. *Brain* 134, 3527–3543. doi: 10.1093/brain/awr299

Mitchell, J. R., Wood, E., and Collins, K. (1999). A telomerase component is defective in the human disease dyskeratosis congenita. *Nature* 402, 551–555. doi: 10.1038/990141

Mochizuki, Y., He, J., Kulkarni, S., Bessler, M., and Mason, P. J. (2004). Mouse dyskerin mutations affect accumulation of telomerase RNA and small nucleolar RNA, telomerase activity, and ribosomal RNA processing. *Proc. Natl. Acad. Sci. U.S.A.* 101, 10756–10761. doi: 10.1073/pnas.0402560101

Motorin, Y., and Grosjean, H. (1999). Multisite-specific tRNA:m5C-methyltransferase (Trm4) in yeast *Saccharomyces cerevisiae*: identification of the gene and substrate specificity of the enzyme. *RNA* 5, 1105–1118. doi: 10.1017/S1355838299982201

Motorin, Y., and Helm, M. (2010). tRNA stabilization by modified nucleotides. *Biochemistry* 49, 4934–4944. doi: 10.1021/bi100408z

Motorin, Y., and Helm, M.(2011). RNA nucleotide methylation. *Wiley Interdiscip. Rev. RNA* 2, 611–631. doi: 10.1002/wrna.79

Okano, M., Xie, S., and Li, E. (1998). Dnmt2 is not required for de novo and maintenance methylation of viral DNA in embryonic stem cells. *Nucleic Acids Res.* 26, 2536–2540. doi: 10.1093/nar/26.11.2536

Oldmeadow, C., Mossman, D., Evans, T.-J., Holliday, E. G., Tooney, P. A., Cairns, M. J., et al. (2014). Combined analysis of exon splicing and genome wide polymorphism data predict schizophrenia risk loci. *J. Psychiatr. Res.* 52, 44–49. doi: 10.1016/j.jpsychires.2014.01.011

Pan, T. (2018). Modifications and functional genomics of human transfer RNA. *Cell Res*. doi: 10.1038/s41422-018-0013-y. [Epub ahead of print].

Patil, D. P., Chen, C. K., Pickering, B. F., Chow, A., Jackson, C., Guttman, M., et al. (2016). m(6)A RNA methylation promotes XIST-mediated transcriptional repression. *Nature* 537, 369–373. doi: 10.1038/nature19342

Pergolizzi, R. G., and Grunberger, D. (1980). The effect of exogenous nutrients on the biosynthesis of Y base in tRNAPHe from Ehrlich ascites carcinoma. *Cancer Lett.* 8, 329–333. doi: 10.1016/0304-3835(80)90149-4

Perrat, P. N., DasGupta, S., Wang, J., Theurkauf, W., Weng, Z., Rosbash, M., et al. (2013). Transposition-driven genomic heterogeneity in the Drosophila brain. *Science* 340, 91–95. doi: 10.1126/science.1231965

Peters, J. (2008). Prader-willi and snoRNAs. *Nat. Genet.* 40, 688–689. doi: 10.1038/ng0608-688

Phalke, S., Nickel, O., Walluscheck, D., Hortig, F., Onorati, M. C., and Reuter, G. (2009). Retrotransposon silencing and telomere integrity in somatic cells of Drosophila depends on the cytosine-5 methyltransferase DNMT2. *Nat. Genet.* 41, 696–702. doi: 10.1038/ng.360

Ping, X. L., Sun, B. F., Wang, L., Xiao, W., Yang, X., Wang, W. J., et al. (2014). Mammalian WTAP is a regulatory subunit of the RNA N6-methyladenosine methyltransferase. *Cell Res.* 24, 177–189. doi: 10.1038/cr.2014.3

Pintard, L., Lecointe, F., Bujnicki, J. M., Bonnerot, C., Grosjean, H., and Lapeyre, B. (2002). Trm7p catalyses the formation of two 2'-O-methylriboses in yeast tRNA anticodon loop. *EMBO J.* 21, 1811–1820. doi: 10.1093/emboj/21.7.1811

Raddatz, G., Guzzardo, P. M., Olova, N., Fantappié, M. R., Rampp, M., Schaefer, M., et al. (2013). Dnmt2-dependent methylomes lack defined DNA methylation patterns. *Proc. Natl. Acad. Sci. U.S.A.* 110, 8627–8631. doi: 10.1073/pnas.1306723110

Rai, K., Chidester, S., Zavala, C. V., Manos, E. J., James, S. R., Karpf, A. R., et al. (2007). Dnmt2 functions in the cytoplasm to promote liver, brain, and retina development in zebrafish. *Genes Dev.* 21, 261–266. doi: 10.1101/gad.1472907

Ramser, J., Winnepenninckx, B., Lenski, C., Errijgers, V., Platzer, M., Schwartz, C. E., et al. (2004). A splice site mutation in the methyltransferase gene FTSJ1 in Xp11.23 is associated with non-syndromic mental retardation in a large Belgian family (MRX9). *J. Med. Genet.* 41, 679–683. doi: 10.1136/jmg.2004. 019000

Rebane, A., Roomere, H., and Metspalu, A. (2002). Locations of several novel 2'-O-methylated nucleotides in human 28S rRNA. *BMC Mol. Biol.* 3:1. doi: 10.1186/1471-2199-3-1

Reitz, C., Tosto, G., Mayeux, R., Luchsinger, J. A., Group, N.-L. N. F. S., and Alzheimer's Disease Neuroimaging, I. (2012). Genetic variants in the Fat and Obesity Associated (FTO) gene and risk of Alzheimer's disease. *PLoS ONE* 7:e50354. doi: 10.1371/journal.pone.0050354

Rintala-Dempsey, A. C., and Kothe, U. (2017). Eukaryotic stand-alone pseudouridine synthases - RNA modifying enzymes and emerging regulators of gene expression? *RNA Biol.* 14, 1185–1196. doi: 10.1080/15476286.2016.1276150

Roignant, J. Y., and Soller, M. (2017). m(6)A in mRNA: an ancient mechanism for fine-tuning gene expression. *Trends Genet.* 33, 380–390. doi: 10.1016/j.tig.2017.04.003

Ross, R. J., Weiner, M. M., and Lin, H. (2014). PIWI proteins and PIWI-interacting RNAs in the soma. *Nature* 505, 353–359. doi: 10.1038/nature12987

Roth, A., Schaffner, W., and Hertel, C. (1999). Phytoestrogen kaempferol (3,4',5,7-tetrahydroxyflavone) protects PC12 and T47D cells from beta-amyloid-induced toxicity. *J. Neurosci. Res.* 57, 399–404. doi: 10.1002/(SICI)1097-4547(19990801)57:3<399::AID-JNR12>3.0.CO;2-W

Roundtree, I. A., Evans, M. E., Pan, T., and He, C. (2017). Dynamic RNA modifications in gene expression regulation. *Cell* 169, 1187–1200. doi: 10.1016/j.cell.2017.05.045

Safra, M., Sas-Chen, A., Nir, R., Winkler, R., Nachshon, A., Bar-Yaacov, D., et al. (2017). The m1A landscape on cytosolic and mitochondrial mRNA at single-base resolution. *Nature* 551, 251–255. doi: 10.1038/nature24456

Sahoo, T., del Gaudio, D., German, J. R., Shinawi, M., Peters, S. U., Person, R. E., et al. (2008). Prader-Willi phenotype caused by paternal deficiency for the HBII-85C/D box small nucleolar RNA cluster. *Nat. Genet.* 40, 719–721. doi: 10.1038/ng.158

Saito, K., Sakaguchi, Y., Suzuki, T., Suzuki, T., Siomi, H., and Siomi, M. C. (2007). Pimet, the Drosophila homolog of HEN1, mediates 2'-O-methylation of PIWI-interacting RNAs at their 3' ends. *Genes Dev.* 21, 1603–1608. doi: 10.1101/gad.1563607

Schaefer, M., and Lyko, F. (2010a). Lack of evidence for DNA methylation of Invader4 retroelements in Drosophila and implications for Dnmt2-mediated epigenetic regulation. *Nat. Genet.* 42, 920–921. doi: 10.1038/ng1110-920

Schaefer, M., and Lyko, F. (2010b). Solving the Dnmt2 enigma. *Chromosoma* 119, 35–40. doi: 10.1007/s00412-009-0240-6

Schaefer, M., Pollex, T., Hanna, K., Tuorto, F., Meusburger, M., Helm, M., et al. (2010). RNA methylation by Dnmt2 protects transfer RNAs against stress-induced cleavage. *Genes Dev.* 24, 1590–1595. doi: 10.1101/gad.586710

Schöller, E., Weichmann, F., Treiber, T., Ringle, S., Treiber, N., Flatley, A., et al. (2018). Interactions, localization and phosphorylation of the m6A generating METTL3-METTL14-WTAP complex. *RNA* 24, 499–512. doi: 10.1261/rna.064063.117

Schorn, A. J., Gutbrod, M. J., LeBlanc, C., and Martienssen, R. (2017). LTR-retrotransposon control by tRNA-derived small RNAs. *Cell* 170, 61.e11–71.e11. doi: 10.1016/j.cell.2017.06.013

Schosserer, M., Minois, N., Angerer, T. B., Amring, M., Dellago, H., Harreither, E., et al. (2015). Methylation of ribosomal RNA by NSUN5 is a conserved mechanism modulating organismal lifespan. *Nat. Commun.* 6, 6158. doi: 10.1038/ncomms7158

Schwartz, S., Bernstein, D. A., Mumbach, M. R., Jovanovic, M., Herbst, R. H., León-Ricardo, B. X., et al. (2014a). Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. *Cell* 159, 148–162. doi: 10.1016/j.cell.2014.08.028

Schwartz, S., Mumbach, M. R., Jovanovic, M., Wang, T., Maciag, K., Bushkin, G. G., et al. (2014b). Perturbation of m6A writers reveals two distinct classes of mRNA methylation at internal and 5' sites. *Cell Rep.* 8, 284–296. doi: 10.1016/j.celrep.2014.05.048

Shaheen, R., Han, L., Faqeih, E., Ewida, N., Alobeid, E., Phizicky, E. M., et al. (2016). A homozygous truncating mutation in PUS3 expands the role of tRNA modification in normal cognition. *Hum. Genet.* 135, 707–713. doi: 10.1007/s00439-016-1665-7

Shao, Y., and Chen, Y. (2016). Roles of circular RNAs in neurologic disease. *Front. Mol. Neurosci.* 9:25. doi: 10.3389/fnmol.2016.00025

Sharma, S., Yang, J., Watzinger, P., Kötter, P., and Entian, K. D. (2013). Yeast Nop2 and Rcm1 methylate C2870 and C2278 of the 25S rRNA, respectively. *Nucleic Acids Res.* 41, 9062–9076. doi: 10.1093/nar/gkt679

Sibley, C. R., Emmett, W., Blazquez, L., Faro, A., Haberman, N., Briese, M., et al. (2015). Recursive splicing in long vertebrate genes. *Nature* 521, 371–375. doi: 10.1038/nature14466

Sledz, P., and Jinek, M. (2016). Structural insights into the molecular mechanism of the m(6)A writer complex. *Elife* 5:e18434. doi: 10.7554/eLife.18434

Sloan, K. E., Warda, A. S., Sharma, S., Entian, K. D., Lafontaine, D. L. J., and Bohnsack, M. T. (2017). Tuning the ribosome: the influence of rRNA modification on eukaryotic ribosome biogenesis and function. *RNA Biol.* 14, 1138–1152. doi: 10.1080/15476286.2016.1259781

Somme, J., Van Laer, B., Roovers, M., Steyaert, J., Versées, W., and Droogmans, L. (2014). Characterization of two homologous 2'-O-methyltransferases

showing different specificities for their tRNA substrates. *RNA* 20, 1257–1271. doi: 10.1261/rna.044054.114

Sridhar, P., Gan, H. H., and Schlick, T. (2008). A computational screen for C/D box snoRNAs in the human genomic region associated with Prader-Willi and Angelman syndromes. *J. Biomed. Sci.* 15, 697–705. doi: 10.1007/s11373-008-9271-x

Takano, K., Nakagawa, E., Inoue, K., Kamada, F., Kure, S., Goto, Y., et al. (2008). A loss-of-function mutation in the FTSJ1 gene causes nonsyndromic X-linked mental retardation in a Japanese family. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* 147B, 479–484. doi: 10.1002/ajmg.b.30638

Terrazas, M., and Kool, E. T. (2009). RNA major groove modifications improve siRNA stability and biological activity. *Nucleic Acids Res.* 37, 346–353. doi: 10.1093/nar/gkn958

Theler, D., Dominguez, C., Blatter, M., Boudet, J., and Allain, F. H. (2014). Solution structure of the YTH domain in complex with N6-methyladenosine RNA: a reader of methylated RNA. *Nucleic Acids Res.* 42, 13911–13919. doi: 10.1093/nar/gku1116

Torres, A. G., Batlle, E., and Ribas de Pouplana, L. (2014). Role of tRNA modifications in human diseases. *Trends Mol. Med.* 20, 306–314. doi: 10.1016/j.molmed.2014.01.008

Tuorto, F., Liebers, R., Musch, T., Schaefer, M., Hofmann, S., Kellner, S., et al. (2012). RNA cytosine methylation by Dnmt2 and NSun2 promotes tRNA stability and protein synthesis. *Nat. Struct. Mol. Biol.* 19, 900–905. doi: 10.1038/nsmb.2357

Uttara, B., Singh, A. V., Zamboni, P., and Mahajan, R. T. (2009). Oxidative stress and neurodegenerative diseases: a review of upstream and downstream antioxidant therapeutic options. *Curr. Neuropharmacol.* 7, 65–74. doi: 10.2174/157015909787602823

van Rossum, D., Verheijen, B. M., and Pasterkamp, R. J. (2016). Circular RNAs: novel regulators of neuronal development. *Front. Mol. Neurosci.* 9:74. doi: 10.3389/fnmol.2016.00074

Vitali, P., Basyuk, E., Le Meur, E., Bertrand, E., Muscatelli, F., Cavaillé, J., et al. (2005). ADAR2-mediated editing of RNA substrates in the nucleolus is inhibited by C/D small nucleolar RNAs. *J. Cell Biol.* 169, 745–753. doi: 10.1083/jcb.200411129

Wang, C. C., Lo, J. C., Chien, C. T., and Huang, M. L. (2011). Spatially controlled expression of the Drosophila pseudouridine synthase RluA-1. *Int. J. Dev. Biol.* 55, 223–227. doi: 10.1387/ijdb.103112cw

Wang, P., Doxtader, K. A., and Nam, Y. (2016a). Structural basis for cooperative function of Mettl3 and Mettl14 methyltransferases. *Mol. Cell* 63, 306–317. doi: 10.1016/j.molcel.2016.05.041

Wang, T., Bray, S. M., and Warren, S. T. (2012). New perspectives on the biology of fragile X syndrome. *Curr. Opin. Genet. Dev.* 22, 256–263. doi: 10.1016/j.gde.2012.02.002

Wang, X., Feng, J., Xue, Y., Guan, Z., Zhang, D., Liu, Z., et al. (2016b). Structural basis of N-6-adenosine methylation by the METTL3-METTL14 complex. *Nature* 534, 575–578. doi: 10.1038/nature18298

Wang, Y., Li, Y., Toth, J. I., Petroski, M. D., Zhang, Z. L., and Zhao, J. C. (2014). N-6-methyladenosine modification destabilizes developmental regulators in embryonic stem cells. *Nat. Cell Biol.* 16, 191–198. doi: 10.1038/ncb2902

Wang, Y., Li, Y., Yue, M., Wang, J., Kumar, S., Wechsler-Reya, R. J., et al. (2018). N(6)-methyladenosine RNA modification regulates embryonic neural stem cell self-renewal through histone modifications. *Nat. Neurosci.* 21, 195–206. doi: 10.1038/s41593-017-0057-1

Wen, J., Lv, R., Ma, H., Shen, H., He, C., Wang, J., et al. (2018). Zc3h13 regulates nuclear RNA m(6)A methylation and mouse embryonic stem cell self-renewal. *Mol. Cell* 69, 1028.e6–1038.e6. doi: 10.1016/j.molcel.2018.02.015

Weng, Y. L., Wang, X., An, R., Cassin, J., Vissers, C., Liu, Y., et al. (2018). Epitranscriptomic m(6)A regulation of axon regeneration in the adult mammalian nervous system. *Neuron* 97, 313.e6–325.e6. doi: 10.1016/j.neuron.2017.12.036

Widagdo, J., Zhao, Q. Y., Kempen, M. J., Tan, M. C., Ratnu, V. S., Wei, W., et al. (2016). Experience-dependent accumulation of N6-methyladenosine in the prefrontal cortex is associated with memory processes in mice. *J. Neurosci. Off. J. Soc. Neurosci.* 36, 6771–6777. doi: 10.1523/JNEUROSCI.4053-15.2016

Wilkinson, C. R., Bartlett, R., Nurse, P., and Bird, A. P. (1995). The fission yeast gene pmt1+ encodes a DNA methyltransferase homologue. *Nucleic Acids Res.* 23, 203–210. doi: 10.1093/nar/23.2.203

Willems, P., Vits, L., Buntinx, I., Raeymaekers, P., Van Broeckhoven, C., and Ceulemans, B. (1993). Localization of a gene responsible for nonspecific mental retardation (MRX9) to the pericentromeric region of the X chromosome. *Genomics* 18, 290–294. doi: 10.1006/geno.1993.1468

Wilusz, J. E. (2015). Controlling translation via modulation of tRNA levels. *Wiley Interdiscip. Rev. RNA* 6, 453–470. doi: 10.1002/wrna.1287

Xu, C., Wang, X., Liu, K., Roundtree, I. A., Tempel, W., Li, Y., et al. (2014). Structural basis for selective binding of m6A RNA by the YTHDC1 YTH domain. *Nat. Chem. Biol.* 10, 927–929. doi: 10.1038/nchembio.1654

Yang, Y., Fan, X., Mao, M., Song, X., Wu, P., Zhang, Y., et al. (2017). Extensive translation of circular RNAs driven by N(6)-methyladenosine. *Cell Res.* 27, 626–641. doi: 10.1038/cr.2017.31

Yao, P., and Fox, P. L. (2013). Aminoacyl-tRNA synthetases in medicine and disease. *EMBO Mol. Med.* 5, 332–343. doi: 10.1002/emmm.201100626

Yildirim, I., Kierzek, E., Kierzek, R., and Schatz, G. C. (2014). Interplay of LNA and 2'-O-methyl RNA in the structure and thermodynamics of RNA hybrid systems: a molecular dynamics study using the revised AMBER force field and comparison with experimental results. *J. Phys. Chem. B* 118, 14177–14187. doi: 10.1021/jp506703g

Yoon, K. J., Ringeling, F. R., Vissers, C., Jacob, F., Pokrass, M., Jimenez-Cyrus, D., et al. (2017). Temporal control of mammalian cortical neurogenesis by m(6)A methylation. *Cell* 171, 877–899. doi: 10.1016/j.cell.2017.09.003

Yu, A. T., Ge, J., and Yu, Y. T. (2011). Pseudouridines in spliceosomal snRNAs. *Protein Cell* 2, 712–725. doi: 10.1007/s13238-011-1087-1

Yu, B., Yang, Z., Li, J., Minakhina, S., Yang, M., Padgett, R. W., et al. (2005). Methylation as a crucial step in plant microRNA biogenesis. *Science* 307, 932–935. doi: 10.1126/science.1107130

Yu, J., Chen, M., Huang, H., Zhu, J., Song, H., Zhu, J., et al. (2018). Dynamic m6A modification regulates local translation of mRNA in axons. *Nucleic Acids Res.* 46, 1412–1423. doi: 10.1093/nar/gkx1182

Yue, Y., Liu, J., Cui, X., Cao, J., Luo, G., Zhang, Z., et al. (2018). VIRMA mediates preferential m(6)A mRNA methylation in 3'UTR and near stop codon and associates with alternative polyadenylation. *Cell Discov.* 4:10. doi: 10.1038/s41421-018-0019-0

Zahn, L. M. (2017). ERVs affect brain gene expression. *Science* 355, 491–492. doi: 10.1126/science.355.6324.491-d

Zalfa, F., Giorgi, M., Primerano, B., Moro, A., Di Penta, A., Reis, S., et al. (2003). The fragile X syndrome protein FMRP associates with BC1 RNA and regulates the translation of specific mRNAs at synapses. *Cell* 112, 317–327. doi: 10.1016/S0092-8674(03)00079-5

Zhang, Y., Shi, J., and Chen, Q. (2017). tsRNAs: new players in mammalian retrotransposon control. *Cell Res.* 27, 1307–1308. doi: 10.1038/cr.2017.109

Zhao, X., Patton, J. R., Davis, S. L., Florence, B., Ames, S. J., and Spanjaard, R. A. (2004). Regulation of nuclear receptor activity by a pseudouridine synthase through posttranscriptional modification of steroid receptor RNA activator. *Mol. Cell* 15, 549–558. doi: 10.1016/j.molcel.2004.06.044

Zhao, Y., Mo, B., and Chen, X. (2012). Mechanisms that impact microRNA stability in plants. *RNA Biol.* 9, 1218–1223. doi: 10.4161/rna.22034

Zheng, G., Dahl, J. A., Niu, Y., Fedorcsak, P., Huang, C. M., Li, C. J., et al. (2013). ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Mol. Cell* 49, 18–29. doi: 10.1016/j.molcel.2012.10.015

Zhou, C., Molinie, B., Daneshvar, K., Pondick, J. V., Wang, J., Van Wittenberghe, N., et al. (2017). Genome-wide maps of m6A circRNAs Identify Widespread and cell-type-specific methylation patterns that are distinct from mRNAs. *Cell Rep.* 20, 2262–2276. doi: 10.1016/j.celrep.2017.08.027

Link Between m6A Modification and Cancers

Zhen-Xian Liu[1†], Li-Man Li[1†], Hui-Lung Sun[2]* and Song-Mei Liu[1]*

[1] Center for Gene Diagnosis, Zhongnan Hospital of Wuhan University, Wuhan, China, [2] Department of Chemistry and Institute for Biophysical Dynamics, Howard Hughes Medical Institute, University of Chicago, Chicago, IL, United States

N6-methyladenosine (m6A) epitranscriptional modification has recently gained much attention. Through the development of m6A sequencing, the molecular mechanism and importance of m6A have been revealed. m6A is the most abundant internal modification in higher eukaryotic mRNAs, which plays crucial roles in mRNA metabolism and multiple biological processes. In this review, we introduce the characteristics of m6A regulators, including "writers" that create m6A mark, "erasers" that show demethylation activity and "readers" that decode m6A modification to govern the fate of modified transcripts. Moreover, we highlight the roles of m6A modification in several common cancers, including solid and non-solid tumors. The regulators of m6A exert enormous functions in cancer development, such as proliferation, migration and invasion. Especially, with the underlying mechanisms being uncovered, m6A and its regulators are expected to be the targets for the diagnosis and treatment of cancers.

Keywords: m6A, mRNA, cancers, function, structures

INTRODUCTION

More than 100 kinds of chemical modifications of RNA have been identified in living organisms (Boccaletto et al., 2018). Studies have widely reported certain types of RNA modifications in eukaryotic mRNA, including N1-methyladenosine (m1A), N6-methyladenosine (m6A) and 5-methylcytosine (m5C), among which m6A was first discovered in the 1970s. m6A is the most abundant internal modification of mRNA and long noncoding RNA (lncRNA) in the majority of eukaryotes. Besides, m6A significantly clusters around the stop codon and 3′ untranslated region (3′UTR) (Dominissini et al., 2012; Meyer et al., 2012; Bodi et al., 2015). m6A modification mostly occurs at RRACH motif (R denotes A or G, H denotes A, C, or U) (Narayan and Rottman, 1988; Csepany et al., 1990; Narayan et al., 1994).

As shown in **Figure 1**, the formation of m6A is a reversible process (Jia et al., 2013), m6A "writers" with methyltransferase activity are consisted of three individual proteins: methyltransferase-like 3 (METTL3), methyltransferase-like 14 (METTL14), and Wilms' tumor 1-associating protein (WTAP). Obesity-associated protein (FTO) and alkB homolog 5 (ALKBH5) are m6A demethylase (Jia et al., 2011; Zheng et al., 2013). Another protein family is m6A "readers," which can recognize m6A modification to modulate mRNA fate (Li et al., 2017a).

m6A modification regulates mRNA at different levels, including structure, maturation, stability, splicing, export, translation and decay (Liu and Zhang, 2018). Moreover, m6A is also involved in cell fate decision, cell cycle regulation, cell differentiation and circadian rhythm maintenance (Wu et al., 2018). Furthermore, multiple RNA binding proteins that are affected by m6A.
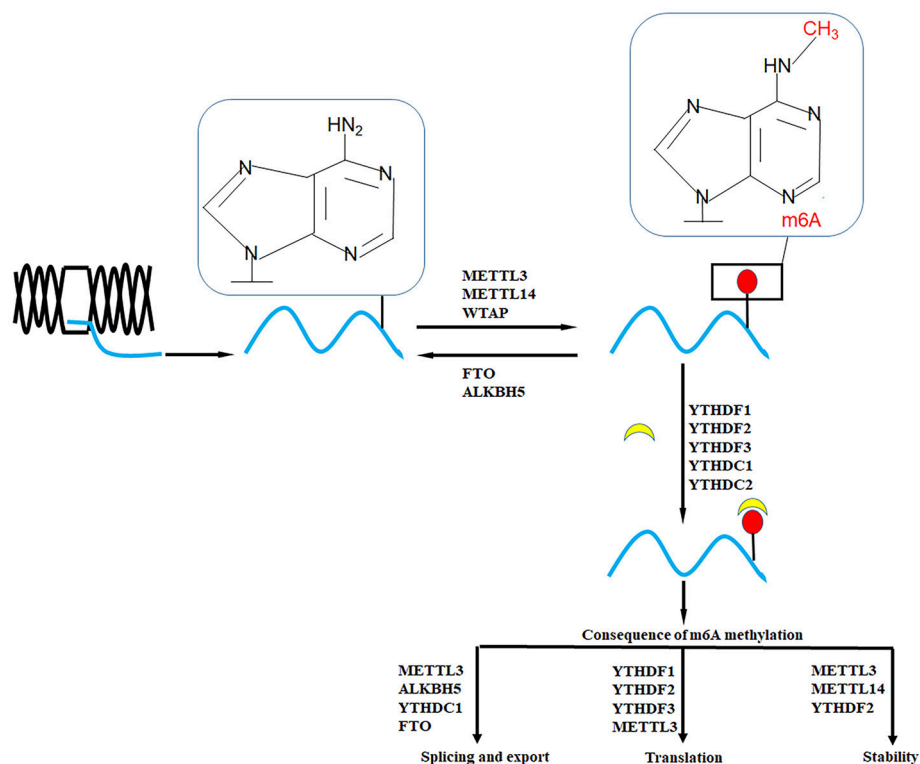
FIGURE 1 | The establishment and function of m6A RNA methylation. The m6A of RNA methylation is dynamically regulated by "writers" (METTL3, METTL14, WTAP, and others) and "erasers" (FTO and ALKBH5). "readers" (YTHDF1, YTHDF2, YTHFDF3, YTHDC1, and YTHDC2) are binding proteins of m6A. Different regulators of m6A have different functions on transcriptional process, including splicing, export, translation, and stability.

For example, heterogeneous nuclear ribonucleoprotein G (HNRNPG), a new m6A reader protein, utilizes a low-complexity region to recognize a motif exposed by m6A modification (Liu et al., 2017). HNRNPC and HNRNPA2B1 are two abundant nuclear RNA-binding proteins (Dai et al., 2018). Currently, increasing evidence has shown the roles of m6A in human diseases. Here, we summarized the functions and roles of m6A regulators in diverse cancers (Wang et al., 2017a), such as acute myeloid leukemia (AML), glioblastoma (GBM), lung cancer, liver cancer, hoping to elucidate the contributions of m6A in cancer process.

PROTEINS INVOLVED IN THE M6A METHYLATION

m6A "Writers"

METTL3, METTL14, and WTAP form m6A methyltransferase complex (Bokar et al., 1997; Liu et al., 2014; Ping et al., 2014). METTL3, a 70-kDa key protein, is firstly identified as m6A "writer" (Bokar et al., 1997). A recent study has indicated that knockdown of METTL3, METTL14, and WTAP could decrease m6A level in polyadenylated RNA. The gel filtration experiment has revealed that METTL3 and METTL14 form a stable METTL3-14 complex with a 1:1 stoichiometric ratio, then WTAP binds to the METTL3-14 complex (Liu et al., 2014). Further

crystallization and structure determination have demonstrated that METTL3-14 heterodimer is asymmetric. In this crystallized METTL3-14 complex, both METTL3 and METTL14 contain MTA-70 methyltransferase domain, two CCCH-type zinc finger motifs exist in the N-terminal region of METTL3 and an N-terminal extension presents in the N-terminal region of METTL14 (Sledz and Jinek, 2016). Additionally, METTL3 acts as the catalytic core, transferring methyl group from S-adenosylmethionine (SAM) to acceptor adenine moiety. METTL14 serves as the RNA-binding platform, promoting the binding of RNA substrate and enhancing the complex integrity (Wang et al., 2016a,b). METTL3-14 complex dimer induces m6A deposition on nuclear RNA. WTAP does not possess methyltransferase activity (Ping et al., 2014), however, it interacts with METTL3-14 complex to affect m6A methyltransferase activity *in vivo* and localization in nuclear speckles (Liu et al., 2014).

m6A "Erasers"

FTO and ALKBH5 have been reported to exhibit demethylation activity (Jia et al., 2011; Zheng et al., 2013). The level of m6A in mRNA increases after FTO knockdown, while m6A in mRNA notably decreases after overexpression of the wild-type FTO (Jia et al., 2011). However, Meyer C et al. reported that, compared with m6A, FTO showed higher affinity with the modification

N6, 2'-O-dimethyladenosine (m6A$_m$), a reversible modification influencing cellular mRNA fate. FTO preferentially demethylated m6A$_m$ and reduced the stability of m6A$_m$ mRNA (Mauer et al., 2017). The difference across the studies may be caused by the location of FTO in different cell lines. ALKBH5 localizes in the nucleus, the level of m6A in mRNA significantly decreases in ALKBH5 overexpressed cells (Zheng et al., 2013). ALKBH5 strictly selects substrate in catalysis progression, and the loss of ALKBH5 impairs RNA metabolism, mRNA export and assembly (Zheng et al., 2013).

m6A "Readers"

Remarkably, m6A regulates gene expression through m6A "readers," a group of various proteins which can recognize m6A modification. The highly conserved YT521-B homology YTH domain family proteins include YTHDF1, YTHDF2, and YTHDF3 in the cytoplasm, and YTH domain containing 1 (YTHDC1) in the nucleus (Wang et al., 2014a, 2015; Xu et al., 2014; Xiao et al., 2016; Shi et al., 2017). YTHDF1 promotes the translation of m6A-methylated mRNA, YTHDF2 accelerates the decay of m6A-methylated mRNA, and YTHDF3, together with YTHDF1 and YTHDF2, noticeably enhances the metabolism of m6A-methylated mRNA in the cytoplasm (Shi et al., 2017). In nuclear speckle localization, YTHDC1 influences mRNA splicing by facilitating SRSF3 but inhibiting SRSF10 (Roundtree et al., 2017). Although YTHDC1 knockdown does not significantly alter the distribution of non-target transcripts, it can affect mature mRNA transport from the nucleus to the cytoplasm (Xiao et al., 2016). YTH domain containing 2 (YTHDC2) can preferentially bind to m6A-containing transcripts resulting in the decrease of mRNA abundance and the enhancement of translation efficiency via the interaction with translation initiation and decay machineries (Hsu et al., 2017).

CONSEQUENCE OF M6A METHYLATION

Impact of m6A on Splicing and Export

mRNA is spliced into mature transcript and exported from nucleus to the cytoplasm, then they can be translated into proteins. Therefore, mRNA nuclear export is a vital step which connects transcription with translation. Simultaneously, mRNA export also regulates gene expression (Wickramasinghe and Laskey, 2015). Depletion of METTL3 delays the export of mature mRNA. In addition, circadian period shows elongation by prolonging nuclear retention of mature mRNA of the clock genes Per2 and Arntl (Fustin et al., 2013). ALKBH5 plays vital roles in mRNA export as well as roles in RNA metabolism and the association of the nuclear speckle proteins (Zheng et al., 2013). YTHDC1 can interact with SRSF3, a nuclear export adaptor protein and splicing factor, to modulate the binding of RNA with SRSF3 and NXF1, then SRSF3-NXF1 and YTHDC1-SRSF3 protein complexes lead m6A-modified mRNA into export pathway (Roundtree et al., 2017). FTO mediates nuclear pre-mRNA alternative splicing. For instance, FTO controls alternative splicing of RUNX1T1, an adipogenesis-related transcription factor, affecting adipogenesis (Zhao et al., 2014). Furthermore, FTO preferentially binds to pre-mRNA in intronic regions, and FTO knockdown leads the substantial changes in pre-mRNA splicing with exon skipping events (Bartosovic et al., 2017).

Impact of m6A on Translation

Translation is regulated by m6A modification through several mechanisms. YTHDF1 is known to promote translation efficiency by binding with m6A. The translation efficiency decreases notably after YTHDF1 knockdown. YTHDF1 ensures the efficient protein production from m6A modification transcript (Wang et al., 2015). In addition, YTHDF1 recruits eukaryotic initiation factor 3 (eIF3) to directly bind a single m6A modification in the 5′UTR, which facilitates ribosome loading and recruits the 43S complex to promote translation (Meyer et al., 2015).

According to previous studies, m6A sites are mainly enriched in the stop codon and 3'UTR (Dominissini et al., 2012; Meyer et al., 2012). Interestingly, heat shock stress induces an elevated m6A peak in the 5'UTR. Upon heat shock stress, the nuclear YTHDF2 limits the demethylation of m6A "eraser" FTO to preserve 5'UTR methylation of stress-induced transcripts. The increased m6A-marked 5'UTR methylation promotes cap-independent translation initiation, providing a mechanism for selective mRNA translation under heat shock stress (Zhou et al., 2015). In translational process, YTHDF3 interacts with ribosomal 40S/60S subunits and significantly enhances translation efficiency of YTHDF1 and YTHDF3 sharing targeting m6A-methylated mRNA (Li et al., 2017a).

METTL3 also directly promotes translation of certain mRNA in human cancer cells by recruiting eIF3. Reader proteins YTHDF1, YTHDF2 and binding-partners METTL14, WTAP are independent of translational control of METTL3 (Lin et al., 2016).

Impact of m6A on RNA Stability

The stability of RNA is closely associated with m6A-dependent degradation process. Knockout of METTL3 and METTL14 in embryonic stem cells can reduce mRNA decay, leading to increase of target mRNA expression, resistance of differentiation, enhancement of self-renewal and maintenance of the pluripotency state of embryonic stem cell (ESC) (Liu et al., 2014; Wang et al., 2014b). A RNA-binding protein, human antigen R (HuR), can increase RNA stability, however, m6A modification inhibits the binding ability by interfering with the HuR (Srikantan et al., 2012).

YTHDF2 directly modulates mRNA decay pathway in an m6A-dependent way (Wang et al., 2014a). Compared to the unmethylated mRNA, YTHDF2 has about 16-fold higher binding affinity to m6A-marked mRNA, showing significant increase of decay rate and shorter half-life (Batista et al., 2014; Fu et al., 2014). The carboxy-terminal domain of YTHDF2 selectively binds to methylated mRNA and leads to YTHDF2-bound transcripts to decay sites (Wang et al., 2014a). Furthermore, YTHDF2 N-terminal region interacts with the SH domain of the CNOT1 subunit recruiting the CCR4-NOT complex to accelerate the deadenylation of YTHDF2-bound m6A-containing mRNA (Du et al., 2016).

Roles of m6A Methylation in Cancers

Currently, increasing studies focus on the potential links between m6A and cancers. In different tumors, the effect of m6A modification could be different. The changes of m6A also affect the progression of tumors, including proliferation, growth, invasion and metastasis (**Figures 2–4**).

Acute Myeloid Leukemia (AML)

The common types of leukemia are acute lymphoblastic leukemia (ALL), chronic lymphocytic leukemia (CLL), acute myeloid leukemia (AML) and chronic myeloid leukemia (CML). As regard the incidence and mortality, AML ranks the first in hematopoietic malignancies. Moreover, different cytogenetic abnormalities [t(8;21)(q22;q22), inv(16)(p13q22)/t(16;16)(p13;q22), t(15;17)(q22;q11~21), and abn(11q23)] and molecular abnormalities (FLT3-ITD, MLL PTD and NPM1 mutations) show different pathogenesis (Frohling et al., 2005). Although new therapies for AML, such as epigenetic targeted drugs and immunotherapies (Yang and Wang, 2018), have been developed, the survival rates have not improved significantly.

In recent years, the molecular mechanism of m6A "writers" and "erasers" in leukemia has been explored (**Figure 2**). METTL3 expression was increased in AML patients and plays an oncogenic role. METTL3 inhibits cell differentiation and apoptosis, and promotes cell proliferation through increasing c-MYC, BCL-2, PTEN translation. MELLT3 also activate PI3K/AKT pathway to control cell differentiation and self-renewal (Vu et al., 2017). Interestingly, METTL3 works on a chromatin-based pathway independently of METTL14 by localizing to the transcriptional start sites of active genes through CAATT-box binding protein (CEBPZ), which resulting in an increase of translation of the corresponding mRNA (Barbieri et al., 2017). In addition, almost all members of "writers" are overexpressed in AML, including METTL3, METTL14, WTAP, and KIAA1429. METTL14 plays an inhibitory role in normal myelopoiesis, whereas suppresses cell differentiation in AML. In SPI1-METTL14-MYB/MYC axis, METTL14, is downregulated by SPI1, exerts an oncogenic role in enhancing leukemia stem/initiating cells self-renewal and repressing myeloid differentiation by regulating MYB and MYC via m6A modification (Weng et al., 2017). WTAP, another member of m6A writers, has been shown an association with AML. Elevated WTAP promotes cell proliferation and inhibits cell differentiation of AML. WTAP also directly links with Hsp90 (Bansal et al., 2014), a molecular chaperone maintains the stability of many tumor-promoting oncoproteins (Whitesell and Lindquist, 2005). More importantly, the knockdown of WTAP with etoposide treatment significantly promotes apoptosis. However, without etoposide treatment, this phenomenon doesn't occur (Bansal et al., 2014). These outcomes suggest WTAP is a potential target for the treatment of AML.

Although both FTO and ALKBH5 have no effect on AML patients' survival rates (Vu et al., 2017), in certain subtypes of AML, such as MLL-rearranged AML, acute promyelocytic leukemia (APL), t(11q23) and t(15;17) AMLs, FTO is highly expressed. FTO reduces the m6A levels of ASB2 and RARA by targeting their UTRs and affects all-trans-retinoic acid (ATRA)

treatment efficiency (Li et al., 2017c). R-2-hydroxyglutarate(R-2HG), which accumulates in isocitrate dehydrogenase 1/2 (IDH1/2) mutant cancers, can increase global m6A via inhibiting FTO, resulting in the decrease of MYC/CEBPA mRNA stability and inhibition of leukemia proliferation (Su et al., 2018).

Genetic alterations of m6A regulatory genes including "writers," "readers," and "erasers" have indicated an inferior cytogenetic risk in AML by analyzing datasets from the Cancer Genome Atlas Research Network (TCGA) (Kwok et al., 2017). METTL3, METTL14, WTAP FTO and R-2HG, interacting with FTO, are potential therapeutic targets for AML. However, study on the "readers" of m6A has not yet been reported.

Glioblastoma (GBM)

Glioblastoma [GBM; WHO grade IV (Xi et al., 2016)] is an invasive malignant primary brain tumor with a median survival of 10–11 months and a poor quality of life(Rick et al., 2018). With the development of epigenetics, the correlation of m6A with GBM has been uncovered in recent years (**Figure 3**).

In GBM, the level of m6A on mRNA is reduced, however, when glioblastoma stem-like cells (GSCs) are induced into differentiation, the m6A level is increased. The effect of knockdown METTL3 or METTL14 on promoting GSCs growth and self-renewal have been further confirmed by overexpression experiments. Additionally, depletion of METTL3 or METTL14 enhances tumor tumorigenicity while FTO inhibitor treatment prevents tumor deterioration. What's more, knockdown of METTL3 or METTL14 alters gene expression leading to elevation of oncogenes, including ADAM19, EPHA3, and KLF4. FTO inhibitor MA2 can prolong the life span of GSCs transplanted animals, suggesting m6A is expected to be a target for the treatment of GBM (Cui et al., 2017). m6A modification related genes and pathways could serve as promising molecular targets for GBM treatment.

A study has proved that METTL3 increases, METTL14 and ALKBH5 decrease, whereas FTO has no significant change in GSCs. METTL3 mediates GSCs maintenance and dedifferentiation by regulating the stability of the SOX2 mRNA though installing m^{6}A on the SOX2-3′UTR. The complete structures of METTL3 and HuR are vital to this procedure. Furthermore, suppressed METTL3 can result in the inhibition of GSC growth and the neurosphere formation, the reduction of stem cell-specific marker (SSEA1) and glioma reprogramming factors (including POU3F2, OLIG2, SALL2, and SOX2) expression. In particular, SOX2 mRNA has a high affinity to METTL3. Crucially, depletion of METTL3 leads to an increase of radiation sensitivity and a decrease of DNA repair, providing a direction for overcoming radiation tolerance (Visvanathan et al., 2018).

WTAP is overexpressed in GBM. WTAP enhances cell proliferation, migration, invasion and tumorigenicity of glioblastoma cells in xenograft via mediating phosphorylation of epidermal growth factor receptor (EGFR) and AKT. Besides, WTAP regulates the expression of certain genes related to motility of cancer cells, such as chemokine ligand 2 (CCL2), chemokine ligand 3 (CCL3), matrix metallopeptidase 3 (MMP3), lysyl oxidase-like 1 (LOXL1), hyaluronan synthase 1 (HAS1),
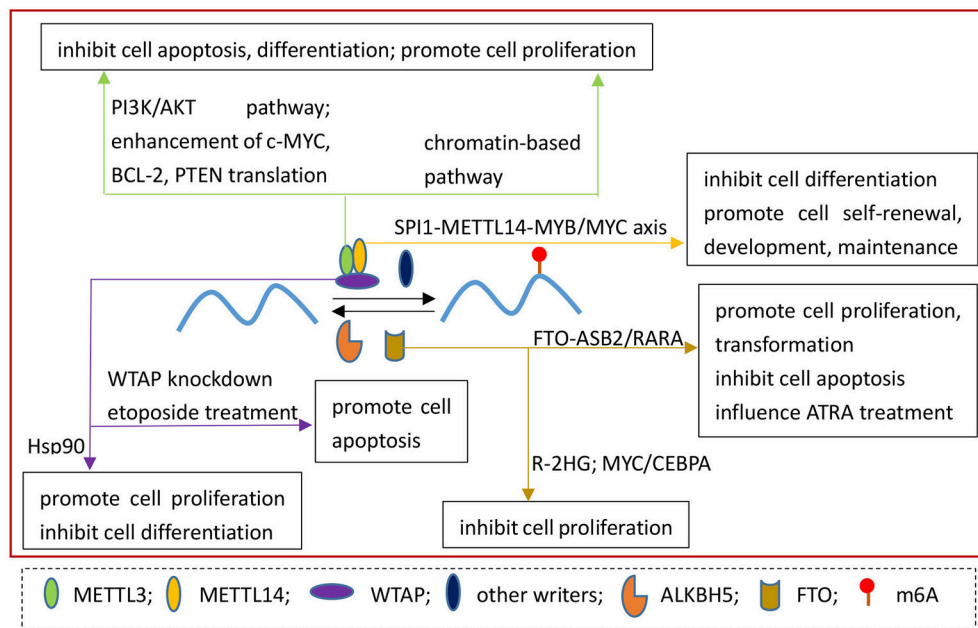
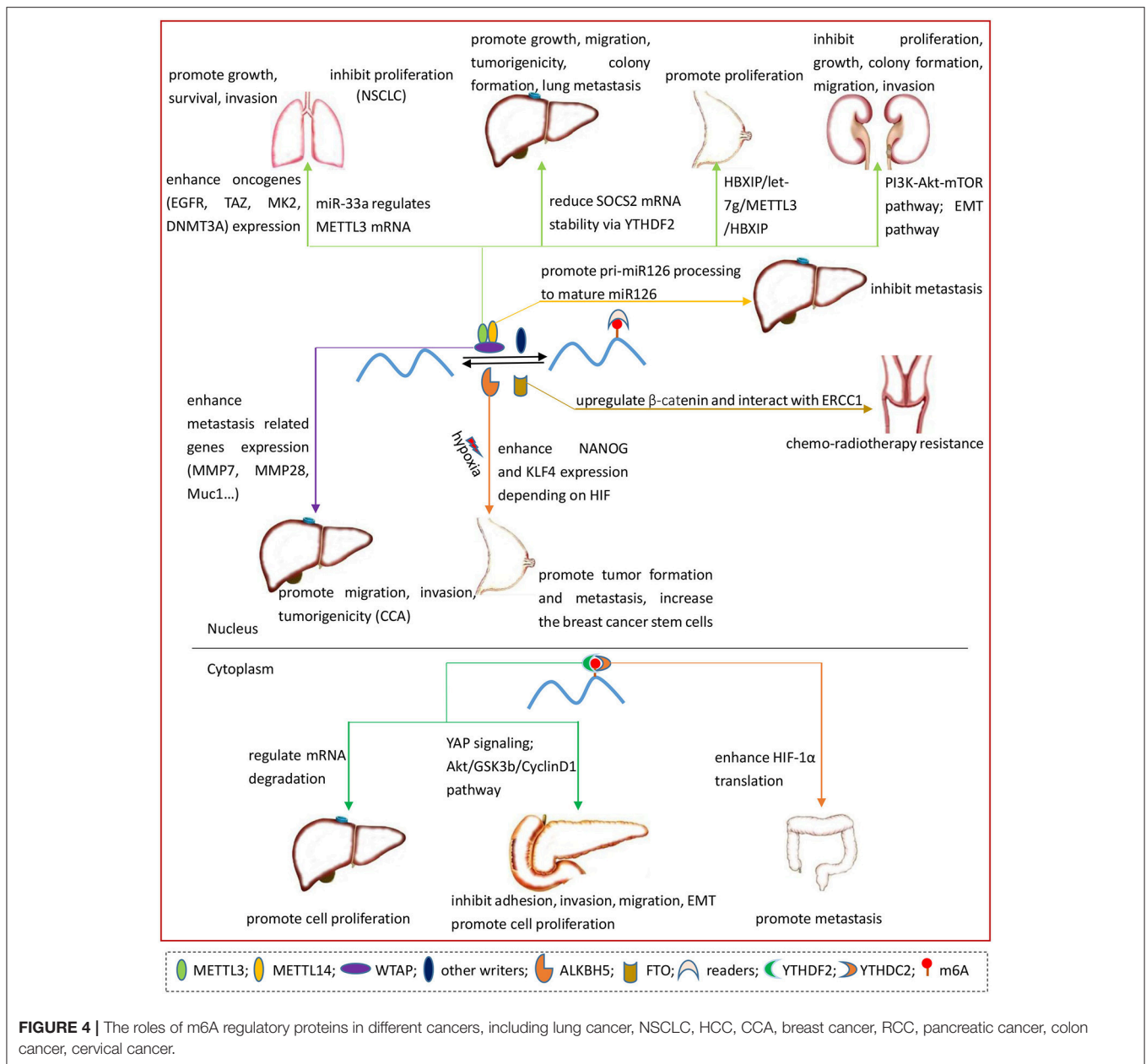**FIGURE 2 |** The roles of m6A regulatory proteins in AML.
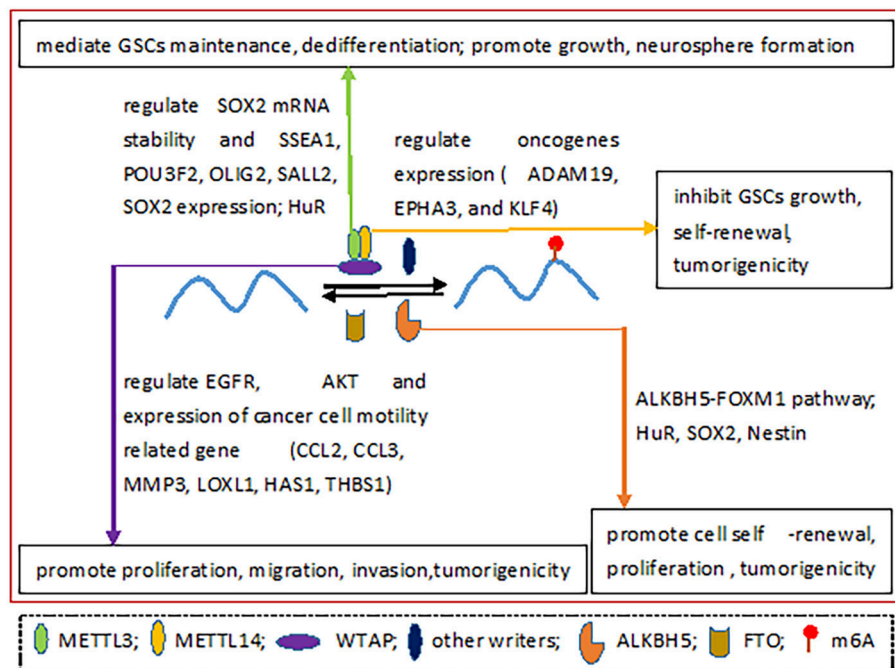
FIGURE 3 | The roles of m6A regulatory proteins in GBM.

and thrombospondin 1 (THBS1) (Jin et al., 2012). The high expression of WTAP is an independent negative prognostic factor that is associated with age and WHO grade, predicting poor overall survival for GBM patients (Xi et al., 2016). Therefore, WTAP may be a prognostic marker for GBM.

Contrary to the text mentioned above (Visvanathan et al., 2018), ALKBH5 is elevated in GSCs (Zhang et al., 2017), enhancing cell self-renewal, proliferation and tumorigenicity. ALKBH5 serves as a poor prognostic indicator in patients with glioma. ALKBH5 removes m^{6}A from FOXM1 (which is a

FIGURE 4 | The roles of m6A regulatory proteins in different cancers, including lung cancer, NSCLC, HCC, CCA, breast cancer, RCC, pancreatic cancer, colon cancer, cervical cancer.

transcription factor and is highly expressed in GBM patients) nascent transcripts by binding to the 3′UTR, enhancing FOXM1 expression. This process can be strengthened by a long non-coding RNA antisense to FOXM1 (FOXM1-AS). Depletion of ALKBH5 and FOXM1-AS inhibits GSCs tumorigenesis via the FOXM1 axis. In addition, certain evidence has confirmed that HuR, SOX2 and Nestin also play a crucial role in this process (Zhang et al., 2017).

Lung Cancer

Lung cancer includes small cell lung carcinoma (SCLC) and non-small-cell lung carcinoma (NSCLC), and NSCLC accounts for approximately 85% of all cases. Although the incidence and death

rate have declined, the 5-year survival rates remain poor (Molina et al., 2008).

As seen in **Figure 4**, METLL3 enhances the translation of certain oncogenes such as EGFR, TAZ, MAPKAPK2 (MK2), and DNMT3A by recruiting eIF3 to the translation initiation complex which is independent of METTL3 catalytic activity and m6A readers. Furthermore, METTL3 plays a driving role in cancer cell growth, survival and invasion (Lin et al., 2016). Another study shows that miR-33a inhibits the proliferation of NSCLC cells by binding to the 3′UTR of METTL3 mRNA (Du et al., 2017). The links between microRNA, METTL3, and NSCLC suggest METTL3 may be a novel target for NSCLC therapy.

Liver Cancer: Hepatocellular Carcinoma (HCC) and Cholangiocarcinoma (CCA)

Hepatocellular Carcinoma (HCC)

The majority of liver cancer is HCC. The incidence and mortality of HCC increase per year owing to the lack of precision diagnosis at an early-stage, and prediction of tumor metastasis and postsurgical recurrence. The 5-year survival rate is only 18%, thus the mechanism and pathogenesis of HCC are urgent to be addressed (Ma et al., 2017). Increasing evidence has shown that m6A and regulators are critical for the development of liver cancer (**Figure 4**).

Currently, only two researches have revealed the association of m6A "writers" with HCC, focusing on METTL3 and METTL14, respectively. METTL3, increasing in HCC, can facilitate HCC cells growth, migration and colony formation *in vitro* and enhance HCC tumorigenicity, growth and lung metastasis *in vivo*. The stability of suppressor of cytokine signaling 2 (SOCS2) mRNA can be downregulated via YTHDF2. And SOCS2 mRNA is a downstream target of METTL3 and a tumor inhibitor in HCC. Particularly, as a clinical manifestation of HCC, the higher level of METTL3 might predict a poor prognosis. RBM15B, KIAA1429 and m6A level on mRNA also increase in HCC while METTL14 has no significant change. However, the disturbance of METTL14 obviously alters the proliferation, migration and colony formation of Huh-7 cell (Chen et al., 2017b). Interestingly, in another study, METTL14 and FTO decreased in HCC while METTL3, WTAP, KIAA1429, and ALKBH5 have no remarkable change. The downregulation of METTL14 acts as a poor prognostic indicator for survival without recurrence in HCC and has a close association with tumor metastasis. Specifically, METTL14 can promote pri-miR126 processing to mature miR126, a tumor suppressor in HCC metastasis, by mediating the recognition and binding of the microprocessor protein DGCR8 to pri-miRNA (Ma et al., 2017).

YTHDF2 is closely associated with the malignance of HCC. YTHDF2 regulates mRNA degradation by recognizing mRNA m6A sites, leading to the enhancement of proliferation of HCC cells. miR-145, which is down-regulated in HCC patients, can suppress the expression of YTHDF2 by directly targeting the 3′UTR of YTHDF2 mRNA(Yang et al., 2017). Thereby, miR-145 could be a candidate target for the treatment of liver cancer.

Cholangiocarcinoma (CCA)

Being diagnosed at advanced stage, CCA has a poor prognosis and limited curative options, and the overall survival rate is rather low. Invasion and migration cancer cells make CCA worse (Jo et al., 2013; Goldaracena et al., 2017; Kennedy et al., 2017).

As a nuclear protein, WTAP upregulates in CCA and has a positive correlation with TNM stage, lymph node metastasis and vascular invasion. WTAP siRNA inhibits CCA cells migration, invasion and tumorigenicity rather than proliferation. The results of cDNA microarray and real time PCR have revealed that WTAP could promote metastasis-related genes expression, such as MMP7, MMP28, and Muc1 (Jo et al., 2013). However, it is not clear whether WTAP function is related to m6A

methyltransferase in CCA. Further investigations remain to be done.

Breast Cancer

In breast cancer, METTL3, HBXIP (mammalian hepatitis B X-interacting protein) and let-7g miRNA form a positive feedback loop (HBXIP/let-7g/METTL3/HBXIP) to promote cell proliferation (**Figure 4**). HBXIP enhances the expression of METTL3 by suppressing the tumor suppressor let-7g. The increase of METTL3, in turn, enhances the level of HBXIP by facilitating m6A modification in mRNA (Cai et al., 2018).

Hypoxia is a vital feature of the tumor microenvironment that drives cancer progression. With hypoxia induction, hypoxia inducible factor (HIF) dependent ALKBH5 strengthens pluripotency factor NANOG mRNA stabilization and accumulation by demethylation (**Figure 4**). Moreover, elevation of ALKBH5 increases the expression of NANOG in breast cancer stem cells (Zhang et al., 2016a). Furthermore, depletion of ALKBH5 or ZNF217, an m6A methyltransferase inhibitor, suppresses NANOG and Kruppel-like factor 4 (KLF4) expression by increasing m6A-methylated RNA (Zhang et al., 2016b). Of note, depletion of ALKBH5 inhibits tumor formation, decreases the breast cancer stem cells and suppresses metastasis from breast to lung in immunodeficient mice (Zhang et al., 2016a,b). The hypoxic induction of pluripotency factor, ALKBH5 and ZNF217 expression depends on HIF, meaning that HIF is vital for breast cancer therapy.

Renal Cell Carcinoma (RCC)

METTL3 plays an inhibitory role in RCC and the higher level indicates a better prognosis. The expression levels of METTL3 mRNA and protein decline in RCC. Depletion of METTL3 promotes cell proliferation, growth, and colony formation through PI3K-Akt-mTOR pathway activation, and enhances cell migration and invasion by epithelial-mesenchymal transition (EMT) pathway. METTL3 knockdown also decreases cell cycle arrest in G1 phase and P21 expression, which acts as a suppressor in tumor (Li et al., 2017b). In a word, METTL3 exerts an essential role in the progression of RCC (**Figure 4**).

Pancreatic Cancer

High expression of YTHDF2 mRNA and protein in pancreatic cancer have a positive correlation with its progression (**Figure 4**). In pancreatic cancer cells, YTHDF2 serves as a suppressor in adhesion, invasion, migration and EMT through YAP signaling, but YTHDF2 acts as a promoter in proliferation via Akt/GSK3b/CyclinD1 pathway (Chen et al., 2017a). As a potential diagnostic and prognostic marker, whether YTHDF2 could be a new therapeutic target remains to be elucidated.

Colon Cancer

RNA helicase YTHDC2 acts as a promoter in colon cancer metastasis by enhancing the translation of hypoxia-inducible factor-1alpha (HIF-1α) gene, which could promote EMT via the transcription factor Twist1 (**Figure 4**). Further, the 5′UTR of both HIF-1α and Twist1 are the targets of YTHDC2 under hypoxia. YTHDC2 upregulates in colon cancer and has an

TABLE 1 | Clinical significance of the regulators of m6A in different cancers.

Cancer	m6A Regulators	Change	Sample Source	Clinical Significance	References
AML	METTL3, METTL14, WTAP, FTO	Up	Patients, nude mice, AML cells	Therapy	[a]
GBM	METTL3	Up	Tissue, nude mice, GSCs	Therapy	[b]
	METTL14/METTL3	Down	NSG mice, GSCs	Therapy	
	WTAP	Up	U87MG/ GBM05 cell, nude mice, tissue	Prognosis	
	ALKBH5	Up	Tissue, Foxn1$^{nu/nu}$ nude mice, GSCs	Prognosis	
Lung Cancer	METTL3	Up	A549/H1299/H1792/HEK cell	NA	Lin et al., 2016; Du et al., 2017
		Down	Tissue, A549/NCI-H460 cell	Therapy	
HCC	METTL3	Up	Tissue, HepG2/Huh-7 cell, BABL/cAnN-nude mice	Prognosis	Chen et al., 2017b; Ma et al., 2017; Yang et al., 2017
	METTL14	Down	Tissue, HepG2 cell, BALB/c nude mice	Prognosis	
	YTHDF2	Up	Tissue,HepG2 cell	Prognosis	
CCA	WTAP	Up	Tissue, HuCCT1/SUN196 cell, nude mice	NA	(Jo et al., 2013)
Breast Cancer	METTL3	Up	Tissue, MCF-7 cell	NA	[c]
	ALKBH5	Up	MCF-7/MDA-MB-435/T4TD cell, NSG mice	Therapy	
RCC	METTL3	Down	Tissue, CAKI-1/CAKI-2/ACHN cell, BALB/c nude mice	Diagnosis, prognosis	Li et al., 2017b
Pancreatic cancer	YTHDF2	Up	Tissue, SW1990/PaTu8988/BxPC3 cell	Diagnosis, prognosis	Chen et al., 2017a
Colon Tumor	YTHDC2	Up	HT29/HCT116/COS cell, BALB/c $^{nu/nu}$ mice	Diagnosis, therapy	Tanabe et al., 2016
Cervical Cancer	FTO	Up	Tissue, SiHa/c-33a cell, nude mice	Therapy, prognosis	Zhou et al., 2018

NA, Not Available.
[a]*Vu et al. (2017), Barbieri et al. (2017), Weng et al. (2017), Bansal et al. (2014), Li et al. (2017c), and Su et al. (2018).*
[b]*Xi et al. (2016), Cui et al. (2017), Visvanathan et al. (2018), Jin et al. (2012), and Zhang et al. (2017).*
[c]*Cai et al. (2018), Zhang et al. (2016a), and Zhang et al. (2016b).*

obviously positive correlation with tumor stage, indicating YTHDC2 may be a diagnostic marker for colon cancer patients (Tanabe et al., 2016). YTHDC2 serves as a RNA helicase rather than the reader of m6A in this study (Tanabe et al., 2016), although other studies have certified that YTHDC2 is a N6-methyladenosine binding protein (Hsu et al., 2017). Whether the two kinds of YTHDC2 belong to the same substance and the exact mechanism still need to be explored.

Cervical Cancer

In cervical cancer, decreased m6A is closely related to tumor size, FIGO stage, differentiation, lymph node infiltration and tumor recurrence (**Figure 4**). The reduction of m6A plays a positive role in cell proliferation and tumor development (Wang et al., 2017b). In particular, FTO might upregulate both mRNA and protein expression of β-catenin, interacting with excision repair cross-complementation group 1 (ERCC1). Ultimately, this process leads to resistance of chemo-radiotherapy on cervical squamous cell carcinoma, the major type of cervical cancer. More importantly, although FTO expression is not associated with the 5-year overall survival, the patients with high FTO and β-catenin expression has a poor prognosis (Zhou et al., 2018). The findings of FTO function in response to the treatment of cervical cancer expands our understanding of m6A role in cancer development.

CONCLUSIONS AND OUTLOOKS

RNA epigenetics has become a hot topic in recent years. Among more than 100 kinds of different chemical modifications, m6A is the most abundant modification with the feature of dynamic and reversibility. It is installed by "writers" and removed by

"erasers." "Readers" are the m6A recognition proteins. m6A is extremly important for mRNA metabolism at different stage, from processing in the nucleus to translation and decay in the cytoplasm. Besides, m6A regulates circadian rhythm, cell cycle, cell differentiation, reprogramming, state transitions and stress responses (Zhao et al., 2017). Except for m6A, other chemical modifications also are irreplaceable, such as m1A, m5C, 2′-O-methylation (2'OMe), pseudouridine (ψ) (Esteller and Pandolfi, 2017; Zhao et al., 2017). For example, m5C plays a significant role in translation efficency, mRNA structure, genetic recoding of coding gene. It also regulates vault ncRNA process into small RNA and tRNA cleavage (Esteller and Pandolfi, 2017).

Diverse cancers are influenced by the structures and functions of m6A modification. In this review, we summarized the mechanisms and functions of m6A-modified RNA in 10 cancers, including AML, GBM, lung cancer, HCC, CCA, breast cancer, RCC, pancreatic cancer, colon tumor and cervical cancer (**Table 1**, **Figures 2–4**). Compared to m6A "writers" and "erasers," only few articles studied the relationship between m6A "readers" and cancers. YTHDF2 in GBM, HCC, intestinal-type gastric adenocarcinoma and lung adenocarcinoma is upregulated. Genetic alterations of m6A readers indicates a poor survival in AML (Kwok et al., 2017). However, the underlying mechanism of m6A "readers" in cancers still remains a mystery. Meanwhile, other modifications also play a vital role in cancers. For instance, ALKBH3, a m1A RNA demethylase, can protect the genome against alkylation damage. TET1, a RNA m5C demethylase, acts as a tumor suppressor in leukemia (Esteller and Pandolfi, 2017). Whether there is joint effect or subtractive effect between m6A and other modifications remains to be further studied.

Among the regulators of m6A, METTL3, the crucial methyltransferase of m6A, plays a promotor in most of cancers. In colorectal, prostate and bile duct cancers, METTL3 has been reported to be significantly upregulated based on bioinformatic analysis (Chen et al., 2017b). In addition to methyltransferase activity, METTL3 also influences cancers indenpendently of its catalytic subunit. This typical function is to enhance the translation of oncogenes (Lin et al., 2016).

Insterestingly, the methyltransferases and demethylases have distinct impacts on various cancer cells. In AML cell, both of them promote cell proliferation and suppress cell differenation (Bansal et al., 2014; Li et al., 2017c; Vu et al., 2017; Weng et al., 2017). In GSCs, METTL3 and METTL14 suppress differenation while WTAP and ALKBH5 promote cell proliferation, self-renewal and tumorigenicity (Jin et al., 2012; Cui et al., 2017; Zhang et al., 2017; Visvanathan et al., 2018). The roles of METTL3 in lung cancer are inconsistent across different studies (Lin et al., 2016; Du et al., 2017). Elevated METTL3 palys an oncogenic role in lung cancer (Lin et al., 2016), while METTL3 is downregulated in NSCLC (Du et al., 2017). It is worth noting that alteration of m6A levels in HCC development is discordant across different studies, METTL3 and METTL14 show a completely contrary effects on migration of HCC cells (Chen et al., 2017b; Ma et al., 2017). These results may suggest that some functions of METTL3 are independent of m6A modification and the underlying mechanism remains to be explored.

Despite the roles of m6A in cancers have dramatically advanced in recent years, a large number of challenges still exist. First, the mechanisms of m6A regulators in some cancers are largely unkown. Second, if m6A level and its regulators could be potential biomarkers for diagnosis and prognosis of some caners, and the specificity and sensitivity of these biomarkers need to be explored. Third, a number of studies have suggested that m6A regulators and related pathways could be used as therapeutic targets, but lack of the specific applications in clinical practice with a large sample size, and the corresponding side effects are largely unkown. All of these issues should be clearly addressed.

AUTHOR CONTRIBUTIONS

Z-XL and L-ML are the co-first author who writer the article. S-ML and H-LS are corresponding author.

ACKNOWLEDGMENTS

REFERENCES

Bansal, H., Yihua, Q., Iyer, S. P., Ganapathy, S., Proia, D. A., Penalva, L. O., et al. (2014). WTAP is a novel oncogenic protein in acute myeloid leukemia. *Leukemia* 28, 1171–1174. doi: 10.1038/leu.2014.16

Barbieri, I., Tzelepis, K., Pandolfini, L., Shi, J., Millan-Zambrano, G., Robson, S. C., et al. (2017). Promoter-bound METTL3 maintains myeloid leukaemia by m(6)A-dependent translation control. *Nature* 552, 126–131. doi: 10.1038/nature24678

Bartosovic, M., Molares, H. C., Gregorova, P., Hrossova, D., Kudla, G., and Vanacova, S. (2017). N6-methyladenosine demethylase FTO targets pre-mRNAs and regulates alternative splicing and 3'-end processing. *Nucleic Acids Res.* 45, 11356–11370. doi: 10.1093/nar/gkx778

Batista, P. J., Molinie, B., Wang, J., Qu, K., Zhang, J., Li, L., et al. (2014). m(6)A RNA modification controls cell fate transition in mammalian embryonic stem cells. *Cell Stem Cell* 15, 707–719. doi: 10.1016/j.stem.2014.09.019

Boccaletto, P., Machnicka, M. A., Purta, E., Piatkowski, P., Baginski, B., Wirecki, T. K., et al. (2018). MODOMICS: a database of RNA modification pathways. 2017 update. *Nucleic Acids Res* 46, D303–D307. doi: 10.1093/nar/gkx1030

Bodi, Z., Bottley, A., Archer, N., May, S. T., and Fray, R. G. (2015). Yeast m6A Methylated mRNAs are enriched on translating ribosomes during meiosis, and under rapamycin treatment. *PLoS ONE* 10:e0132090. doi: 10.1371/journal.pone.0132090

Bokar, J. A., Shambaugh, M. E., Polayes, D., Matera, A. G., and Rottman, F. M. (1997). Purification and cDNA cloning of the AdoMet-binding subunit of the human mRNA (N6-adenosine)-methyltransferase. *RNA* 3, 1233–1247

Cai, X., Wang, X., Cao, C., Gao, Y., Zhang, S., Yang, Z., et al. (2018). HBXIP-elevated methyltransferase METTL3 promotes the progression of breast cancer via inhibiting tumor suppressor let-7g. *Cancer Lett.* 415, 11–19. doi: 10.1016/j.canlet.2017.11.018

Chen, J., Sun, Y., Xu, X., Wang, D., He, J., Zhou, H., et al. (2017a). YTH domain family 2 orchestrates epithelial-mesenchymal transition/proliferation dichotomy in pancreatic cancer cells. *Cell Cycle* 16, 2259–2271. doi: 10.1080/15384101.2017.1380125

Chen, M., Wei, L., Law, C. T., Tsang, F. H., Shen, J., Cheng, C. L., et al. (2017b). RNA N6-methyladenosine methyltransferase METTL3 promotes liver cancer progression through YTHDF2 dependent post-transcriptional silencing of SOCS2. *Hepatology.* 67, 2254–2270. doi: 10.1002/hep.29683

Csepany, T., Lin, A., Baldick, C. J., and Beemon, K. (1990). Sequence specificity of mRNA N6-adenosine methyltransferase. *J. Biol. Chem.* 265, 20117–20122

Cui, Q., Shi, H., Ye, P., Li, L., Qu, Q., Sun, G., et al. (2017). m(6)A RNA Methylation Regulates the Self-Renewal and Tumorigenesis of Glioblastoma Stem Cells. *Cell Rep.* 18, 2622–2634. doi: 10.1016/j.celrep.2017.02.059

Dai, D., Wang, H., Zhu, L., Jin, H., and Wang, X. (2018). N6-methyladenosine links RNA metabolism to cancer progression. *Cell Death Dis* 9:124. doi: 10.1038/s41419-017-0129-x

Dominissini, D., Moshitch-Moshkovitz, S., Schwartz, S., Salmon-Divon, M., Ungar, L., Osenberg, S., et al. (2012). Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* 485, 201–206. doi: 10.1038/nature11112

Du, H., Zhao, Y., He, J., Zhang, Y., Xi, H., Liu, M., et al. (2016). YTHDF2 destabilizes m(6)A-containing RNA through direct recruitment of the CCR4-NOT deadenylase complex. *Nat. Commun.* 7:12626. doi: 10.1038/ncomms12626

Du, M., Zhang, Y., Mao, Y., Mou, J., Zhao, J., Xue, Q., et al. (2017). MiR-33a suppresses proliferation of NSCLC cells via targeting METTL3 mRNA. *Biochem. Biophys. Res. Commun.* 482, 582–589. doi: 10.1016/j.bbrc.2016.11.077

Esteller, M., and Pandolfi, P. P. (2017). The epitranscriptome of noncoding RNAs in cancer. *Cancer Discov.* 7, 359–368. doi: 10.1158/2159-8290.CD-16-1292

Fröhling, S., Scholl, C., Gilliland, D. G., and Levine, R. L. (2005). Genetics of myeloid malignancies: pathogenetic and clinical implications. *J. Clin. Oncol.* 23, 6285–6295. doi: 10.1200/JCO.2005.05.010

Fu, Y., Dominissini, D., Rechavi, G., and He, C. (2014). Gene expression regulation mediated through reversible m(6)A RNA methylation. *Nat. Rev. Genet.* 15, 293–306. doi: 10.1038/nrg3724

Fustin, J. M., Doi, M., Yamaguchi, Y., Hida, H., Nishimura, S., Yoshida, M., et al. (2013). RNA-methylation-dependent RNA processing controls the speed of the circadian clock. *Cell* 155, 793–806. doi: 10.1016/j.cell.2013.10.026

Goldaracena, N., Gorgen, A., and Sapisochin, G. (2017). Current status of liver transplantation for cholangiocarcinoma. *Liver Transpl.* 24, 294–303. doi: 10.1002/lt.24955

Hsu, P. J., Zhu, Y., Ma, H., Guo, Y., Shi, X., Liu, Y., et al. (2017). Ythdc2 is an N(6)-methyladenosine binding protein that regulates mammalian spermatogenesis. *Cell Res.* 27, 1115–1127. doi: 10.1038/cr.2017.99

Jia, G., Fu, Y., and He, C. (2013). Reversible RNA adenosine methylation in biological regulation. *Trends Genet.* 29, 108–115. doi: 10.1016/j.tig.2012.11.003

Jia, G., Fu, Y., Zhao, X., Dai, Q., Zheng, G., Yang, Y., et al. (2011). N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat. Chem. Biol.* 7, 885–887. doi: 10.1038/nchembio.687

Jin, D. I., Lee, S. W., Han, M. E., Kim, H. J., Seo, S. A., Hur, G. Y., et al. (2012). Expression and roles of Wilms' tumor 1-associating protein in glioblastoma. *Cancer Sci.* 103, 2102–2109. doi: 10.1111/cas.12022

Jo, H. J., Shim, H. E., Han, M. E., Kim, H. J., Kim, K. S., Baek, S., et al. (2013). WTAP regulates migration and invasion of cholangiocarcinoma cells. *J. Gastroenterol.* 48, 1271–1282. doi: 10.1007/s00535-013-0748-7

Kennedy, L., Hargrove, L., Demieville, J., Francis, N., Seils, R., Villamaria, S., et al. (2017). Recent advances in understanding cholangiocarcinoma. *F1000Res* 6:1818. doi: 10.12688/f1000research.12118.1

Kwok, C. T., Marshall, A. D., Rasko, J. E., and Wong, J. J. (2017). Genetic alterations of m(6)A regulators predict poorer survival in acute myeloid leukemia. *J. Hematol. Oncol.* 10:39. doi: 10.1186/s13045-017-0410-6

Li, A., Chen, Y. S., Ping, X. L., Yang, X., Xiao, W., Yang, Y., et al. (2017a). Cytoplasmic m(6)A reader YTHDF3 promotes mRNA translation. *Cell Res.* 27, 444–447. doi: 10.1038/cr.2017.10

Li, X., Tang, J., Huang, W., Wang, F., Li, P., Qin, C., et al. (2017b). The M6A methyltransferase METTL3: acting as a tumor suppressor in renal cell carcinoma. *Oncotarget* 8, 96103–96116. doi: 10.18632/oncotarget.21726

Li, Z., Weng, H., Su, R., Weng, X., Zuo, Z., Li, C., et al. (2017c). FTO Plays an Oncogenic Role in Acute Myeloid Leukemia as a N(6)-Methyladenosine RNA Demethylase. *Cancer Cell* 31, 127–141. doi: 10.1016/j.ccell.2016.11.017

Lin, S., Choe, J., Du, P., Triboulet, R., and Gregory, R. I. (2016). The m(6)A Methyltransferase METTL3 promotes translation in human cancer cells. *Mol. Cell* 62, 335–345. doi: 10.1016/j.molcel.2016.03.021

Liu, J., Yue, Y., Han, D., Wang, X., Fu, Y., Zhang, L., et al. (2014). A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. *Nat. Chem. Biol.* 10, 93–95. doi: 10.1038/nchembio.1432

Liu, N., Zhou, K. I., Parisien, M., Dai, Q., Diatchenko, L., and Pan, T. (2017). N6-methyladenosine alters RNA structure to regulate binding of a low-complexity protein. *Nucleic Acids Res.* 45, 6051–6063. doi: 10.1093/nar/gkx141

Liu, Z., and Zhang, J. (2018). Human C-to-U coding RNA editing is largely nonadaptive. *Mol. Biol. Evol.* 35, 963–969. doi: 10.1093/molbev/msy011

Ma, J. Z., Yang, F., Zhou, C. C., Liu, F., Yuan, J. H., Wang, F., et al. (2017). METTL14 suppresses the metastatic potential of hepatocellular carcinoma by modulating N(6) -methyladenosine-dependent primary MicroRNA processing. *Hepatology* 65, 529–543. doi: 10.1002/hep.28885

Mauer, J., Luo, X., Blanjoie, A., Jiao, X., Grozhik, A. V., Patil, D. P., et al. (2017). Reversible methylation of m(6)Am in the 5' cap controls mRNA stability. *Nature* 541, 371–375. doi: 10.1038/nature21022

Meyer, K. D., Patil, D. P., Zhou, J., Zinoviev, A., Skabkin, M. A., Elemento, O., et al. (2015). 5' UTR m(6)A promotes cap-independent translation. *Cell* 163, 999–1010. doi: 10.1016/j.cell.2015.10.012

Meyer, K. D., Saletore, Y., Zumbo, P., Elemento, O., Mason, C. E., and Jaffrey, S. R. (2012). Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell* 149, 1635–1646. doi: 10.1016/j.cell.2012.05.003

Molina, J. R., Yang, P., Cassivi, S. D., Schild, S. E., and Adjei, A. A. (2008). Non-small cell lung cancer: epidemiology, risk factors, treatment, and survivorship. *Mayo Clin. Proc.* 83, 584–594. doi: 10.4065/83.5.584

Narayan, P., and Rottman, F. M. (1988). An *in vitro* system for accurate methylation of internal adenosine residues in messenger RNA. *Science* 242, 1159–1162

Narayan, P., Ludwiczak, R. L., Goodwin, E. C., and Rottman, F. M. (1994). Context effects on N6-adenosine methylation sites in prolactin mRNA. *Nucleic Acids Res.* 22, 419–426

Ping, X. L., Sun, B. F., Wang, L., Xiao, W., Yang, X., Wang, W. J., et al. (2014). Mammalian WTAP is a regulatory subunit of the RNA N6-methyladenosine methyltransferase. *Cell Res.* 24, 177–189. doi: 10.1038/cr.2014.3

Rick, J., Chandra, A., and Aghi, M. K. (2018). Tumor treating fields: a new approach to glioblastoma therapy. *J. Neurooncol.* 137, 447–453. doi: 10.1007/s11060-018-2768-x

Roundtree, I. A., Luo, G. Z., Zhang, Z., Wang, X., Zhou, T., Cui, Y., et al. (2017). YTHDC1 mediates nuclear export of N(6)-methyladenosine methylated mRNAs. *Elife* 6:pii: e31311. doi: 10.7554/eLife.31311

Shi, H., Wang, X., Lu, Z., Zhao, B. S., Ma, H., Hsu, P. J., et al. (2017). YTHDF3 facilitates translation and decay of N(6)-methyladenosine-modified RNA. *Cell Res.* 27, 315–328. doi: 10.1038/cr.2017.15

Sledz, P., and Jinek, M. (2016). Structural insights into the molecular mechanism of the m(6)A writer complex. *Elife* 5:e18434. doi: 10.7554/eLife.18434

Srikantan, S., Tominaga, K., and Gorospe, M. (2012). Functional interplay between RNA-binding protein HuR and microRNAs. *Curr. Protein Pept. Sci.* 13, 372–379.

Su, R., Dong, L., Li, C., Nachtergaele, S., Wunderlich, M., Qing, Y., et al. (2018). R-2HG Exhibits Anti-tumor Activity by Targeting FTO/m(6)A/MYC/CEBPA Signaling. *Cell* 172, 90–105.e123. doi: 10.1016/j.cell.2017.11.031

Tanabe, A., Tanikawa, K., Tsunetomi, M., Takai, K., Ikeda, H., Konno, J., et al. (2016). RNA helicase YTHDC2 promotes cancer metastasis via the enhancement of the efficiency by which HIF-1alpha mRNA is translated. *Cancer Lett.* 376, 34–42. doi: 10.1016/j.canlet.2016.02.022

Visvanathan, A., Patil, V., Arora, A., Hegde, A. S., Arivazhagan, A., Santosh, V., et al. (2018). Essential role of METTL3-mediated m6A modification in glioma stem-like cells maintenance and radioresistance. *Oncogene* 37, 522–533. doi: 10.1038/onc.2017.351

Vu, L. P., Pickering, B. F., Cheng, Y., Zaccara, S., Nguyen, D., Minuesa, G., et al. (2017). The N(6)-methyladenosine (m6A)-forming enzyme METTL3 controls myeloid differentiation of normal hematopoietic and leukemia cells. *Nat. Med.* 23, 1369–1376. doi: 10.1038/nm.4416

Wang, P., Doxtader, K. A., and Nam, Y. (2016a). Structural Basis for Cooperative Function of Mettl3 and Mettl14 Methyltransferases. *Mol. Cell* 63, 306–317. doi: 10.1016/j.molcel.2016.05.041

Wang, S., Sun, C., Li, J., Zhang, E., Ma, Z., Xu, W., et al. (2017a). Roles of RNA methylation by means of N(6)-methyladenosine (m6A) in human cancers. *Cancer Lett.* 408, 112–120. doi: 10.1016/j.canlet.2017.08.030

Wang, X., Feng, J., Xue, Y., Guan, Z., Zhang, D., Liu, Z., et al. (2016b). Structural basis of N(6)-adenosine methylation by the METTL3-METTL14 complex. *Nature* 534, 575–578. doi: 10.1038/nature18298

Wang, X., Li, Z., Kong, B., Song, C., Cong, J., Hou, J., et al. (2017b). Reduced m(6)A mRNA methylation is correlated with the progression of human cervical cancer. *Oncotarget* 8, 98918–98930. doi: 10.18632/oncotarget.22041

Wang, X., Lu, Z., Gomez, A., Hon, G. C., Yue, Y., Han, D., et al. (2014a). N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* 505, 117–120. doi: 10.1038/nature12730

Wang, X., Zhao, B. S., Roundtree, I. A., Lu, Z., Han, D., Ma, H., et al. (2015). N(6)-methyladenosine modulates messenger RNA translation efficiency. *Cell* 161, 1388–1399. doi: 10.1016/j.cell.2015.05.014

Wang, Y., Li, Y., Toth, J. I., Petroski, M. D., Zhang, Z., and Zhao, J. C. (2014b). N6-methyladenosine modification destabilizes developmental regulators in embryonic stem cells. *Nat. Cell Biol.* 16, 191–198. doi: 10.1038/ncb2902

Weng, H., Huang, H., Wu, H., Qin, X., Zhao, B. S., Dong, L., et al. (2017). METTL14 inhibits hematopoietic stem/progenitor differentiation and promotes leukemogenesis via mRNA m6A modification. *Cell Stem Cell.* 22, 191–205. doi: 10.1016/j.stem.2017.11.016

Whitesell, L., and Lindquist, S. L. (2005). HSP90 and the chaperoning of cancer. *Nat. Rev. Cancer* 5, 761–772. doi: 10.1038/nrc1716

Wickramasinghe, V. O., and Laskey, R. A. (2015). Control of mammalian gene expression by selective mRNA export. *Nat. Rev. Mol. Cell Biol.* 16, 431–442. doi: 10.1038/nrm4010

Wu, Y., Zhou, C., and Yuan, Q. (2018). Role of DNA and RNA N6-Adenine methylation in regulating stem cell fate. *Curr. Stem Cell Res. Ther.* 13, 31–38. doi: 10.2174/1574888X12666170621125457

Xi, Z., Xue, Y., Zheng, J., Liu, X., Ma, J., and Liu, Y. (2016). WTAP expression predicts poor prognosis in malignant glioma patients. *J. Mol. Neurosci.* 60, 131–136. doi: 10.1007/s12031-016-0788-6

Xiao, W., Adhikari, S., Dahal, U., Chen, Y. S., Hao, Y. J., Sun, B. F., et al. (2016). Nuclear m(6)A reader YTHDC1 regulates mRNA splicing. *Mol. Cell* 61, 507–519. doi: 10.1016/j.molcel.2016.01.012

Xu, C., Wang, X., Liu, K., Roundtree, I. A., Tempel, W., Li, Y., et al. (2014). Structural basis for selective binding of m6A RNA by the YTHDC1 YTH domain. *Nat. Chem. Biol.* 10, 927–929. doi: 10.1038/nchembio.1654

Yang, X., and Wang, J. (2018). Precision therapy for acute myeloid leukemia. *J. Hematol. Oncol.* 11:3. doi: 10.1186/s13045-017-0543-7

Yang, Z., Li, J., Feng, G., Gao, S., Wang, Y., Zhang, S., et al. (2017). MicroRNA-145 modulates N(6)-methyladenosine levels by targeting the 3'-Untranslated mRNA Region of the N(6)-Methyladenosine binding YTH Domain family 2 Protein. *J. Biol. Chem.* 292, 3614–3623. doi: 10.1074/jbc.M116.749689

Zhang, C., Samanta, D., Lu, H., Bullen, J. W., Zhang, H., Chen, I., et al. (2016a). Hypoxia induces the breast cancer stem cell phenotype by HIF-dependent and ALKBH5-mediated m(6)A-demethylation of NANOG mRNA. *Proc. Natl. Acad. Sci. U.S.A.* 113, E2047–E2056. doi: 10.1073/pnas.1602883113

Zhang, C., Zhi, W. I., Lu, H., Samanta, D., Chen, I., Gabrielson, E., et al. (2016b). Hypoxia-inducible factors regulate pluripotency factor expression by ZNF217- and ALKBH5-mediated modulation of RNA methylation in breast cancer cells. *Oncotarget* 7, 64527–64542. doi: 10.18632/oncotarget.11743

Zhang, S., Zhao, B. S., Zhou, A., Lin, K., Zheng, S., Lu, Z., et al. (2017). m(6)A Demethylase ALKBH5 maintains tumorigenicity of glioblastoma stem-like cells by sustaining FOXM1 expression and cell proliferation program. *Cancer Cell* 31, 591–606.e596. doi: 10.1016/j.ccell.2017.02.013

Zhao, B. S., Roundtree, I. A., and He, C. (2017). Post-transcriptional gene regulation by mRNA modifications. *Nat. Rev. Mol. Cell Biol.* 18, 31–42. doi: 10.1038/nrm.2016.132

Zhao, X., Yang, Y., Sun, B. F., Shi, Y., Yang, X., Xiao, W., et al. (2014). FTO-dependent demethylation of N6-methyladenosine regulates mRNA splicing and is required for adipogenesis. *Cell Res.* 24, 1403–1419. doi: 10.1038/cr.2014.151

Zheng, G., Dahl, J. A., Niu, Y., Fedorcsak, P., Huang, C. M., Li, C. J., et al. (2013). ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Mol. Cell* 49, 18–29. doi: 10.1016/j.molcel.2012. 10.015

Zhou, J., Wan, J., Gao, X., Zhang, X., Jaffrey, S. R., and Qian, S. B. (2015). Dynamic m(6)A mRNA methylation directs translational control of heat shock response. *Nature* 526, 591–594. doi: 10.1038/nature 15377

Zhou, S., Bai, Z. L., Xia, D., Zhao, Z. J., Zhao, R., Wang, Y. Y., et al. (2018). FTO regulates the chemo-radiotherapy resistance of cervical squamous cell carcinoma (CSCC) by targeting beta-catenin through mRNA demethylation. *Mol. Carcinog.* 57, 590–597. doi: 10.1002/mc.22782

frontiers
in Plant Science

Unveiling Chloroplast RNA Editing Events Using Next Generation Small RNA Sequencing Data

Nureyev F. Rodrigues[1], Ana P. Christoff[1], Guilherme C. da Fonseca[2], Franceli R. Kulcheski[3] and Rogerio Margis[1, 2, 4]*

[1] Programa de Posgraduação em Genética e Biologia Molecular, Departamento de Genética, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil, [2] Programa de Posgraduação em Biologia Celular e Molecular, Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil, [3] Programa de Pósgraduação em Biologia Celular e do Desenvolvimento, Departamento de Biologia Celular, Genética e Embriologia, Universidade Federal de Santa Catarina, Florianópolis, Brazil, [4] Departamento de Biofísica, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil

Organellar RNA editing involves the modification of nucleotide sequences to maintain conserved protein functions, mainly by reverting non-neutral codon mutations. The loss of plastid editing events, resulting from mutations in RNA editing factors or through stress interference, leads to developmental, physiological and photosynthetic alterations. Recently, next generation sequencing technology has generated the massive discovery of sRNA sequences and expanded the number of sRNA data. Here, we present a method to screen chloroplast RNA editing using public sRNA libraries from Arabidopsis, soybean and rice. We mapped the sRNAs against the nuclear, mitochondrial and plastid genomes to confirm predicted cytosine to uracil (C-to-U) editing events and identify new editing sites in plastids. Among the predicted editing sites, 40.57, 34.78, and 25.31% were confirmed using sRNAs from Arabidopsis, soybean and rice, respectively. SNP analysis revealed 58.2, 43.9, and 37.5% new C-to-U changes in the respective species and identified known and new putative adenosine to inosine (A-to-I) RNA editing in tRNAs. The present method and data reveal the potential of sRNA as a reliable source to identify new and confirm known editing sites.

Keywords: small RNA, chloroplast, RNA editing, NGS, SNP genotyping

INTRODUCTION

Chloroplasts are notable examples of successful endosymbiosis in the early origin of modern life forms. These organelles possess their own gene expression machinery, with complex posttranscriptional processes and fine nucleus-cytosol crosstalk. In plants, these organelles undergo a posttranscriptional process called RNA editing, corresponding to nucleotide changes from cytosine to uracil (C-to-U) and less frequently from uracil to cytosine (U-to-C), in some sites of coding sequences (Tillich et al., 2006; Chateigner-Boutin and Small, 2010). These nucleotide changes correct the codons to encode appropriate amino acids, maintaining the functional amino acid sequence of the evolutionarily conserved protein (Takenaka et al., 2013). Another well-known mechanism of RNA editing is the adenine to inosine (A-to-I) editing, as observed in the chloroplast tRNA^{Arg} (ACG). This type of editing enables hydrogen bond formation with more than one base in the corresponding codon position (Su and Randau, 2011). The A-to-I editing in position 34 of the tRNA^{Arg} (ACG) produces the wobble nucleotide described as essential for efficient chloroplast

translation (Delannoy et al., 2009). In *Arabidopsis thaliana*, arginine tRNA adenosine deaminase (TAD or ADAT) performs this deamination (Elias and Huang, 2005; Delannoy et al., 2009).

RNA editing in coding sequences increases the conservation levels among proteins across several plants species. Evolutionarily, codons generated by RNA editing are more conserved than codons encoded by genomic DNA (Guo et al., 2015). Editing sites located within coding sequences have been well studied, despite the existence of editing sites in non-coding regions, such as introns and tRNAs. There are several cases of different editing efficiencies from plant to plant, and even among different plant tissues (Peeters and Hanson, 2002; Chateigner-Boutin and Hanson, 2003; Tseng et al., 2013), suggesting that several different RNA editing sites remain to be elucidated.

The identification of all components from the RNA editing machinery has not yet been achieved, although several proteins have been identified as important for the maintenance of editing processes. The pentatricopeptide repeat proteins (PPR) are a highly diverse protein family. In the plant evolutionary landscape of PPR proteins, 109 genomes/proteomes were analyzed, resulting in a total of 49,204 PPR genes and 616,206 motifs (Cheng et al., 2016). Some of these PPRs harbor a DYW motif, similar to the deaminase motifs observed in other proteins, which could explain the C-to-U nucleotide conversion (Salone et al., 2007; Schallenberg-Rüdinger et al., 2013; Hayes et al., 2015). In addition, several studies have reported PPRs associated with specific RNA editing events, demonstrating that these molecules bind to specific *cis*-elements located upstream of the RNA editing site (Okuda et al., 2006; Barkan and Small, 2014). Moreover, the PPR alone is not sufficient to promote RNA editing but requires other proteins, such as RNA editing-interacting (RIP/MORF), OMMR and OZ proteins, to achieve a successful editing event (Bentolila et al., 2013; Sun et al., 2016).

The most frequent plastid RNA editing type in flowering plants is the C-to-U change, with approximately 40 sites detected thus far in Arabidopsis (Takenaka et al., 2013). To facilitate RNA editing site prediction in organelles, software, such as PREP suite has been developed (Mower, 2009). These programs enable RNA editing site prediction in genes from organelles by considering homology and conservation among protein sequences compared to genomic databases. Currently, thousands of partial and complete plastid genomes are available in NCBI, which can be used to extensively search for RNA editing events.

Different experimental techniques have identified chloroplast RNA editing sites. A widely used method is the reverse transcription PCR (RT-PCR) of plastid messenger RNAs in which several chloroplast cDNA fragments are cloned into vectors and further sequenced (Rüdinger et al., 2009). Additionally, if a chloroplast candidate gene sequence is previously known, then specific primers can be designed to direct the gene amplification from cDNA samples, with subsequent sequencing (Wolf et al., 2004). RNA editing events can also be detected through the Poisoned Primer Extension method or High Resolution Melting (HRM) analysis (Chateigner-Boutin and Small, 2007), using chloroplast cDNA as a template for amplification. Another method to measure RNA editing is multiplex RT-PCR mass spectrometry, described as a robust and convenient method

(Germain et al., 2015). Although robust, these methods are dependent on specific primers and are restricted to RNA editing studies only.

RNA sequencing has facilitated RNA editing analyses by comparing reads from RNA-seq data with organelle genome references. Currently, RNA-seq is primarily adapted to study polyadenylated transcripts. Thus, as their cyanobacterial ancestor, several plastid polyadenylated RNA transcripts are associated with the RNA decay pathway *via* degradation by 3′– 5′ exoribonucleases (Komine et al., 2002; Zimmer et al., 2009). Therefore, this approach generates RNA-seq libraries with smaller amounts of plastid reads than libraries generated from organelle-enriched RNA samples, with posterior reduction of ribosomal RNA (Guo et al., 2015). Furthermore, these approaches restrict the analysis to only transcripts located in chloroplasts, preventing a comparative analysis between nuclear and plastid transcripts.

In recent years, studies of small RNAs (sRNA) have considerably increased, particularly associated with the deep sequencing of microRNAs (miRNAs) and other small non-coding RNAs (ncRNAs) from nuclear origin, producing a large amount of new sequence data. These studies have focused on the roles of sRNAs in genome maintenance, development and plant responses to environmental stresses (Simon et al., 2009; Long et al., 2015; Xu et al., 2015). However, plastid-derived sRNA sequences have also been identified in these total sRNA libraries (Ruwe and Schmitz-Linneweber, 2012; Zhelyazkova et al., 2012; Ruwe et al., 2016). Therefore, considerable amounts of sRNA data are available in public databases and can be employed for RNA editing studies. In the present study, we propose that sRNA sequencing data could represent an additional resource to identify chloroplast RNA editing events, in addition to other approaches, such as strand-specific RNA sequencing and Single Nucleotide Polymorphism (SNP). Here, we describe a method for identifying a set of new editing sites in chloroplast transcripts using sRNA data. Analyses of sRNA libraries can provide a strong qualitative and reliable quantitative measure of plastid RNA editing events.

MATERIALS AND METHODS

sRNA Libraries and Chloroplast Genomes

Public RNA libraries deposited in NCBI GEO (www.ncbi.nlm.nih.gov/geo/) with accession numbers GSE85070 (Wu et al., 2016) (*Arabidopsis thaliana*, mRNA-seq and sRNA-seq), GSE69571 (da Fonseca et al., 2016) (*Glycine max*, soybean, mRNA-seq and sRNA-seq) and GSE77046 (Neto et al., 2015) (*Oryza sativa* japonica group, rice, sRNA-seq; mRNA-seq data unpublished) were used as input data to evaluate the proposed method. These libraries were produced from samples with no qualitative influence on RNA editing and did not use any method to enrich the isolation of plastid RNAs. The Arabidopsis mutant data present in the libraries were not used. For sRNA analyses, only reads with 18–24 nucleotides were selected from the libraries. Complete chloroplast genome, coding sequences and tRNAs from Arabidopsis (NC_000932), soybean (NC_007942), and rice (NC_001320) were obtained separately at the Index

of Genomes from The CpBase: Chloroplast Genome Database (http://chloroplast.ocean.washington.edu/).

Prediction of Conserved Editing Sites

The Predictive RNA Editor for Plants suite (PREP-Cp) (http://prep.unl.edu/) (Mower, 2009) was used to predict conserved plastid editing sites. These sites were used to evaluate read coverage and editing percentage using the sRNA data. Fasta files corresponding to plastid coding sequence data were manually formatted to be usedfor use as an input batch file in the PREP-Cp tool. To predict editing sites for each species, a less stringent cutoff value of 0.5 was used, despite the 0.8 default value. This lower cutoff value was used to evaluate the effective occurrence of the predicted editing sites and their efficacious detection from sRNA data.

RNA Mapping and Confirmation of Predicted Sites

The sRNA/mRNA libraries were primarily mapped using Bowtie (Langmead et al., 2009) with 0 mismatch and no reverse complement against the chloroplast genome, coding sequences and tRNAs. Mapped reads resulted in a new file (m0). Unmapped reads were submitted to a second round of mapping with no mismatches against nuclear and mitochondrial genomes. This step eliminates all reads with perfect matches against these genomes. Unmapped reads were further mapped with two mismatches and no reverse complement against chloroplast genome and coding sequences. This second group of mapped reads produced another file containing reads with editing events (m2). Both m0 and m2 fastq files were concatenated in an m0 + m2 file. The C-to-U editing sites predicted by PREP-Cp in the cpDNA coding sequence were subjected to m0 + m2 mapping and further manual inspection using Tablet software (Milne et al., 2013). The predicted editing sites were confirmed based on a C-to-T mapping change. The steps described above are summarized in **Figure 1**.

Single Nucleotide Polymorphism Analysis

The m0 + m2 fastq files from sRNA libraries were mapped against the whole chloroplast genome, coding sequences and tRNAs using Geneious-R8 (Kearse et al., 2012), with the Bowtie algorithm and the same parameters of the previous mapping (**Figure 1**). The Geneious find variation/SNPs tool was used to search for A-to-G and C-to-T changes in putative new editing sites that were not predicted by PREP. The following parameters were used: Minimum Coverage of 5, Maximum Variant P-value of 10^{-2}, option to find polymorphism Inside and Outside coding sequence and P-value calculation method as approximate. In the manual inspection of mapping, reads with putative editing events in the $5'$ and $3'$ end were discarded to improve prediction and selection for validation using RT-qPCR assay.

Validation and Analysis of the RNA Editing Sites Using RT-qPCR

To validate predicted and new C-to-U RNA editing sites from the sRNA data in soybean chloroplast transcripts [*Glycine max*
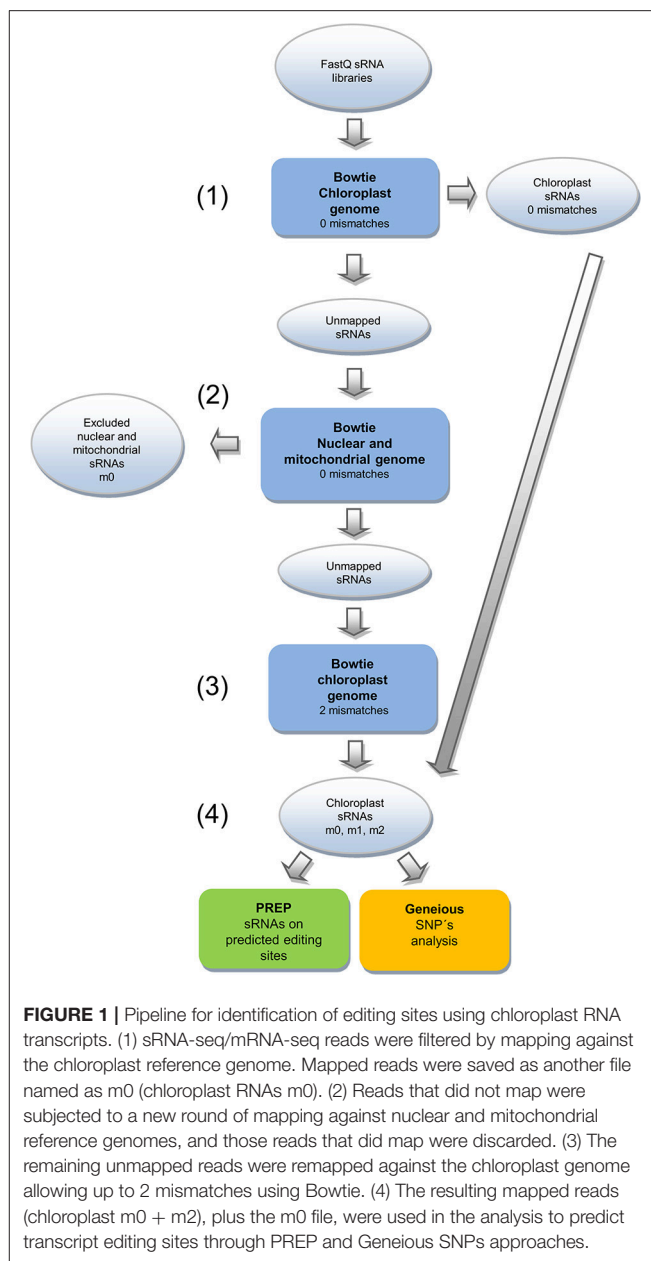
FIGURE 1 | Pipeline for identification of editing sites using chloroplast RNA transcripts. (1) sRNA-seq/mRNA-seq reads were filtered by mapping against the chloroplast reference genome. Mapped reads were saved as another file named as m0 (chloroplast RNAs m0). (2) Reads that did not map were subjected to a new round of mapping against nuclear and mitochondrial reference genomes, and those reads that did map were discarded. (3) The remaining unmapped reads were remapped against the chloroplast genome allowing up to 2 mismatches using Bowtie. (4) The resulting mapped reads (chloroplast m0 + m2), plus the m0 file, were used in the analysis to predict transcript editing sites through PREP and Geneious SNPs approaches.

(L.) Merrill], we collected the roots, leaves and petals from the soybean cultivar Conquista. These tissues were collected as biological triplicates. All samples were immediately frozen in liquid nitrogen, and total RNA was extracted using Trizol (Invitrogen, CA, USA). The RNA quality was evaluated through electrophoresis on a 1% agarose gel, and the RNA amount was verified using a Qubit fluorometer and Quant-iT RNA assay kit according to the manufacturer's instructions (Invitrogen, CA, USA).

Reverse transcription quantitative polymerase chain reaction (RT-qPCR) was performed to validate the C-to-U RNA editing rates for some predicted editing sites in soybean chloroplast genes across three different tissues (roots, leaves and petals). To validate and quantify new RNA editing sites, only leaf

samples were used. The cDNA synthesis was performed with approximately 1 µg of total RNA. Each reaction was primed with 1 µM dT25V oligonucleotide (Invitrogen, Carlsbad, CA, USA). Prior to transcription, RNA and the oligo(dT)25V primer oligo were mixed with RNase-free water to a total volume of 10 µL and incubated at 70°C for 5 min, followed by cooling on ice. The reactions were reverse transcribed with 1X M-MLV RT buffer, 0.5 mM dNTPs (Ludwig, Porto Alegre, RS, Brazil) and 200 U of M-MLV RT Enzyme (Promega, Madison, WI, USA) in a final volume of 30 µL. The synthesis was performed at 40°C for 60 min. All cDNA samples were diluted 100-fold with RNase-free water and subsequently used as templates in RT-qPCR analysis. The subsequent PCR amplification was performed using a set of primers designed according to Chen et al. (2008), with modifications. A set of primers, comprising two specific editing primers and one unique universal primer, were designed for each editing site. Specific editing primers were characterized by a unique difference in the last nucleotide at the 3′ end that recognizes and differentiates edited and unedited sites. All primers employed in the reaction are listed in Table S1.

All RT-qPCR reactions were performed on a Bio-Rad CFX384 real-time PCR detection system (Bio-Rad, Hercules, CA, USA) using SYBR Green I (Invitrogen, Carlsbad, CA, USA) to detect double-stranded cDNA synthesis. The reactions were conducted in a 10 µL volume containing 5 µL of diluted cDNA (1:100), 0.2X SYBR Green I, 0.1 mM dNTP, 1X PCR buffer, 3 mM MgCl$_2$, 0.25 U Platinum Taq DNA Polymerase (Invitrogen, Carlsbad, CA, USA) and 200 nM of each forward and reverse primer. The samples were analyzed as biological triplicates and technical quadruplicates in a 384-well plate. A non-template control was also included. The PCR reactions were run under the following conditions: an initial polymerase hot start at 94°C for 5 min, followed by 40 cycles at 94°C for 15 s, 60°C for 15 s and 72°C for 10 s. A melting curve analysis was programmed at the end of the PCR run over the range of 65 to 99°C, and the temperature increased stepwise by 0.5°C. The threshold and baseline were manually determined using Bio-Rad CFX manager software.

To calculate the RNA editing rates, we used the threshold cycle (Ct) generated during the qPCR amplifications. To calculate the percentage of editing, an equation that considered the difference between the Ct-values of each editing variant was used:

$$\% \; RNA \; editing \; = \frac{2^{(Ct \; mean \; of \; T \; variant \; - \; Ct \; mean \; of \; C \; variant)}}{2^{(Ct \; mean \; of \; T \; variant \; - \; Ct \; mean \; of \; C \; variant)} + 1} \times 100$$

RESULTS

sRNA Reads Mapped to Chloroplast Genomes

The sRNA libraries sequenced without plastid RNA isolation were mapped to Arabidopsis, soybean and rice chloroplast genomes using an in-house pipeline (**Figure 1**). Approximately 3.2, 1.6, and 0.9 million reads did not map to nuclear and mitochondrial genomes but mapped to Arabidopsis, soybean and rice chloroplast genomes, respectively. These chloroplast (cp)-mapped reads represented approximately 22.9% (Arabidopsis), 4.79% (soybean), and 3.62% (rice) of the total reads in these libraries (**Table 1**). The editing informative m2 reads corresponded to 455,904 (Arabidopsis), 208,417 (soybean), and 144,609 (rice). The histograms representing the percentage length distribution of each individual class are shown in **Figure S1**. The mean coverage was 838.6 in Arabidopsis, 358.6 in soybean and 222 in rice. The maximum coverage values were 872,674 in Arabidopsis, 380,116 in soybean and 166,534 in rice. Some chloroplast regions were not covered by the sRNA library reads, with minimal coverage of zero. The number of plastid genome positions with no coverage was 47,057 in Arabidopsis, 24,505 in soybean and 3,039 in rice, representing approximately 30.46, 16.09, and 2.25% of each chloroplast genome, respectively. The genome fraction coverage for Arabidopsis, soybean and rice is represented in **Figure S2**.

sRNA Polymorphisms Confirm PREP Editing Site Prediction in Coding-Sequence Genes

The conserved chloroplast C-to-U RNA editing sites were predicted using the Predictive RNA Editor for Plants (PREP-Cp) (http://prep.unl.edu/) (Mower, 2009). The PREP suite predicted 69 potential editing sites in Arabidopsis, 92 sites in soybean and 79 sites in rice chloroplast genes. These predicted editing sites

TABLE 1 | Distribution of sRNA sequences among nuclear, mitochondrial and plastid genomes.

Organism	Total	Nuclear	mtDNA	cpDNA (m0)	cpDNA (m2)	cpDNA total	Not aligned
Arabidopsis	14,113,280	6,369,985	18,393	2,778,067	454,904	3,232,971	4,491,931
	100%	45.13%	0.13%	19.68%	3.22%	22.9%	31.82%
Soybean	34,313,559	28,219,467	46,399	1,438,193	208,417	1,646,610	4,401,083
	100%	82.23%	0.13%	4.19%	0.60%	4.79%	12.82%
Rice	25,247,958	21,479,400	12,003	768,437	144,609	913,046	2,843,509
	100%	85.07%	0.05%	3.04%	0.57%	3.62%	11.27%

m0, reads with no mismatches.
m2, reads with until 2 mismatches.

were distributed in 21 different coding sequences in Arabidopsis and rice and 23 coding sequences in soybean. The mapped chloroplast sRNA reads were analyzed using Tablet software to evaluate the presence/absence of C-to-U editing events in the predicted sites. Different numbers of confirmed editing sites were observed among the three species: 28 sites in Arabidopsis, 32 sites in soybean and 20 sites in rice, corresponding to 40.57, 34.78, and 25.31% of the total sites, respectively. The PREP score (values between 0 and 1) indicates editing site prediction confidence to control the relative proportion of false positive and false negative predictions. When a more stringent score value (≥0.8) was considered, the predicted editing site numbers decreased to 45, 59, and 29 for Arabidopsis, soybean and rice, respectively. Analyses of chloroplast sRNA alignment confirmed the 23 predicted editing sites in Arabidopsis, 28 sites in soybean,

and 14 sites in rice, corresponding to 51.1, 47.45, and 48.27% of the total predicted editing sites, respectively (**Figure 2A**). Even with a higher score value, some predicted sites were not confirmed, reflecting the absence of reads corresponding to editing or not enough coverage (Table S2). Four editing sites were conservatively predicted and confirmed among the three species. These sites corresponded to three sites inside the *ndhB* transcript and one site in the *rps14* transcript. Soybean and Arabidopsis shared 11 common editing sites in the *atpF*, *clpP*, *ndhB*, *ndhD*, *psbE*, *psbF*, *rpoB*, *rpoC1*, and *rps14* transcripts. Concerning the rice *atpF*, *clpP*, *ndhB*, *psbE*, and *psbF* genes, a thymine was already present in these editing sites. Rice shared a single editing site with Arabidopsis in the *ndhB* transcript at position 467, which in soybean corresponds to a thymine. The numbers of unique confirmed editing sites for each species were 12, 16, and 14
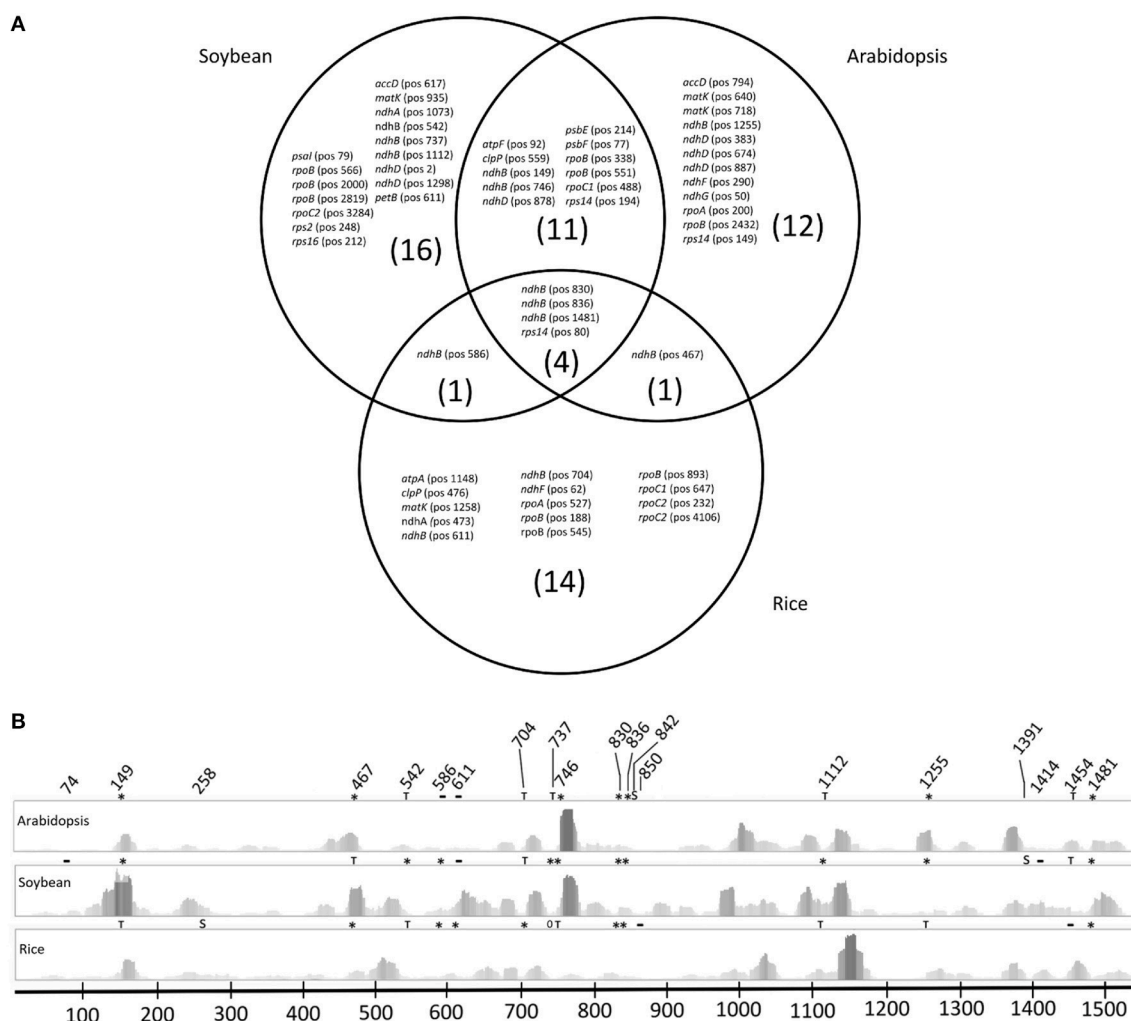
FIGURE 2 | PREP predicted editing sites and graphical read distribution and editing in the *ndhB* transcript. **(A)** Venn diagram with confirmed RNA editing sites predicted by PREP in Arabidopsis, soybean and rice. Gene names followed by the position numbering of the editing site in the coding sequence are indicated. **(B)** Graphical representation of sRNA coverage and predicted editing sites in the *ndhB* gene; (S) editing sites identified by SNP analysis, (T) predicted editing site in another species that already has a thymine in the species, (*) editing site predicted by PREP and confirmed by read mapping and coverage, (−) predicted sites with reads but not confirmed by editing and (0) predicted editing sites without read coverage.

for Arabidopsis, soybean and rice, respectively (**Figure 2A**). The complete distribution of PREP predicted editing sites according to species is described in Table S2.

## mRNA-Seq and sRNA-Seq Differences in RNA Editing Analysis

To provide information concerning sRNA data reliability, the C-to-U RNA editing profiles were compared to the PREP predicted editing sites between the sRNA and mRNA (messenger RNA) libraries in Arabidopsis, soybean and rice. The mRNA-Seq data confirmed 27 predicted editing sites in Arabidopsis, 37 sites in soybean and 20 sites in rice, corresponding to 39.13, 40.21, and 25.31% of the predicted sites, respectively (Table S3). One predicted editing site was exclusively confirmed using mRNA-Seq libraries in Arabidopsis, and 11 predicted editing sites were confirmed in soybean and rice. However, analyses using sRNA-Seq libraries detected two exclusively confirmed editing sites in Arabidopsis, six sites in soybean and eight sites in rice. The confirmed predicted editing sites shared between mRNA and sRNA data corresponded to 37.68, 28.26, and 15.19% of the total predicted editing sites in Arabidopsis, soybean and rice, respectively (**Figure 3**).

Confirmation of PREP Predicted Editing Sites and New Editing Site Prediction through SNP Analysis in Coding-Sequences Using sRNA Data

In addition to the confirmation of the predicted editing sites, new candidates for editing sites were searched. A SNP analysis was used with a minimum P-value of $\leq 10^{-10}$ to identify sites with C-to-T changes. This parameter enabled the identification of 59 potential editing sites in Arabidopsis, 43 sites in soybean, and 19 sites in rice. Among these editing sites, 58, 37, and 15

sites encode amino acid changes in Arabidopsis, soybean and rice, respectively (Table S4). These editing sites were distributed in 27 genes in Arabidopsis, 24 genes in soybean and 11 genes in rice. Comparison of these editing sites against the editing sites predicted using PREP revealed that 20, 18, and 7 sites were previously predicted in Arabidopsis, soybean and rice, respectively (Table S5). Among these sites, 18, 18, and 6 sites were predicted with a higher score value in Arabidopsis, soybean and rice, respectively.

When the edited transcript distribution was evaluated in all species (**Figure 4A**), a higher editing frequency was associated with a core of genes (*clpP*, *ndhB*, *ndhF*, *rpoA*, *rpoB*, *rpoC1*, *rpoC2*, and *rps14*) and confirmed with at least one method used for all species evaluated. Considering exclusive edited genes, Arabidopsis showed 14 editing sites distributed among nine genes identified using SNP analysis. The editing in the rice *atpA* gene, detected through SNP analysis, was predicted by PREP. Soybean presented four exclusive editing sites confirmed by sRNA reads and predicted by PREP. They sites were distributed among the *petB*, *rps2*, and *rps14* genes. C-to-U changes promote a serine to leucine amino acid change in *petB* and *rps14* and
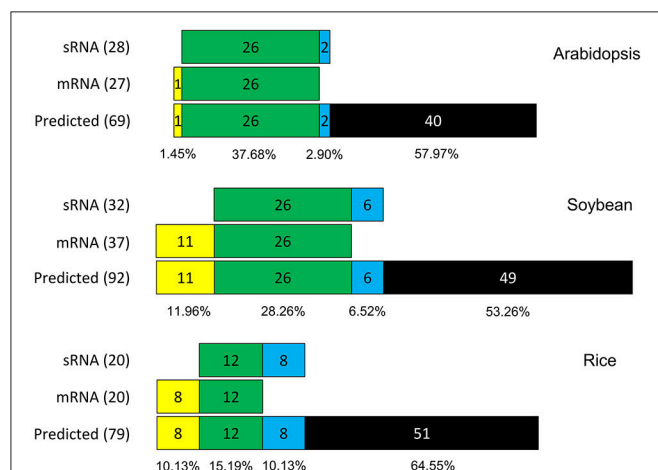
FIGURE 3 | Comparison of predicted editing site confirmation between sRNA and mRNA data. On the left, values of total confirmed predicted editing sites by data type (mRNA or sRNA). Green boxes represent editing sites confirmed in both data; yellow boxes represent editing sites confirmed only in mRNA data; blue boxes represent editing sites confirmed only in sRNA data; and black boxes represent unconfirmed predicted editing sites.
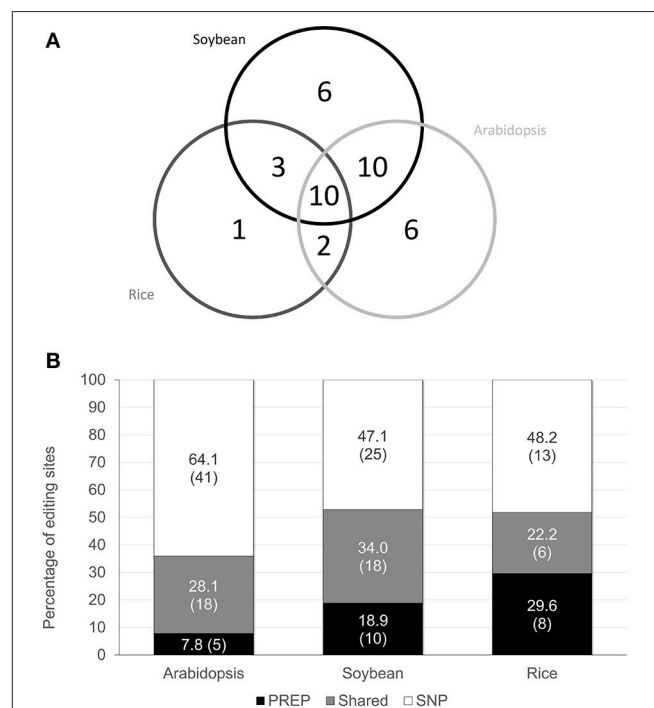
FIGURE 4 | Number of genes with C-to-U editing sites in the studied species. **(A)** Venn diagram with the total number of genes with editing sites in Arabidopsis, soybean and rice, when using both PREP (only confirmed) and SNP analysis. Not all genes share common editing sites among species. The gene identities are described in Table S6. **(B)** Percentages of total RNA editing sites identified by distinct approaches, as observed in Arabidopsis, soybean and rice. The absolute number of editing sites for each method is in parentheses. Black bars correspond to the percentage of total sites confirmed only by PREP prediction (>0.8 in prediction score); white bars indicate the percentage of total sites confirmed by the SNP approach; and gray bars show the percentage of total sites confirmed using both approaches.

a histidine to tyrosine amino acid change in *rps2*. Arabidopsis, soybean and rice SNP analysis revealed 19, 15, and 7 C-to-T changes distributed among 11, 10 and five exclusive genes, respectively. All genes and their respective editing sites are listed in Table S6. The comparative C-to-T analysis using different identification methods demonstrated that the SNP method could identify reliable C-to-U editing events, including events previously predicted using PREP at a lower PREP score (> 0.5) (**Figure S3**) or a more stringent cutoff (PREP score >0.8) (**Figure 4B**).

C-to-U RNA Editing in the ndhB Gene

The well-studied *ndhB* gene was the most frequently edited gene detected through PREP prediction in all plants. The number of editing sites predicted by PREP in this gene varied between species: 9 sites in Arabidopsis, 13 in soybean and 10 in rice. The number of editing sites confirmed by sRNA alignment was 7 sites in Arabidopsis, 9 sites in soybean and 7 sites in rice, representing 77.7, 69.23, and 70% of the predicted editing sites, respectively. Other editing sites could not be confirmed, reflecting insufficient read coverage (**Table 2**). In contrast, despite high predicted editing site numbers, 7 sites in Arabidopsis, 9 sites in soybean and 5 sites in rice, the *matK* gene had only two confirmed predicted editing sites in Arabidopsis and one confirmed predicted editing site in soybean and rice (Table S2).

In the *ndhB* gene, SNP analysis detected potential new editing sites in all three species (**Table 2**). However, this gene was not the most edited gene according to SNP analysis in rice. In this species, *ndhB* had three new potential editing sites, while *rpoC2* gene had four new sites. In Arabidopsis, *ndhD* had 8 new potential editing sites according to SNP analysis. In soybean, the *ndhB* gene remained as the most edited gene (Table S6). Comparative analyses showed a different read distribution of the predicted sites in *ndhB* among species (**Figure 2B**). Some regions showed higher coverage, not only in the editing site, but also in neighboring sites. For example, PREP predicted 467 editing sites (C-to-U), with varied coverage between species, but reads confirming the editing event were observed in both Arabidopsis and rice. Although soybean had a higher amount of reads in this site, a T was present in this genomic position. Notably, several sites showed more than 10 reads of coverage but did not confirm editing events. Some putative editing sites predicted using SNP analysis showed higher coverage than the predicted sites confirmed using PREP (**Table 2**).

A-to-I Editing Events Predicted Using SNP Analysis in Chloroplast tRNA Genes

Chloroplast sRNAs can also be useful in adenosine to inosine (A-to-I) RNA editing screening. tRNA genes were used to evaluate editing events, by searching for a guanosine (G) SNP in sRNA mapping since inosine is read as G by cellular machineries (Kim, 2004).

tRNA genes showed at least one position with an A-to-G change in at least two species (Table S7), totaling 11, 4, and 12 putative A-to-I editing events in Arabidopsis, soybean and rice, respectively. These A-to-G changes were distributed in 8, 4, and 10 tRNAs in Arabidopsis, soybean and rice, respectively. Among

these sites, two sites were conserved between species: position 58 of tRNA-Trp (CCA) between soybean and rice and position 35 of tRNA-Arg (ACG) among all species evaluated. In tRNA-Arg (ACG), nucleotide 35 presented 40, 58.8, and 67.8% of the edited reads in Arabidopsis, soybean and rice, respectively (**Table 3**). The tRNAs most frequently edited were tRNA-Ser (UGA), with 3 A-to-G changes in Arabidopsis, and tRNA-Leu (UAG) and tRNA-Trp (CCA) with two A-to-G changes in Arabidopsis and rice, respectively.

Validation of C-to-U RNA Editing in Soybean Plastid Genes

To validate some predicted editing sites and demonstrate sRNA data reliability as a resourceful tool for the identification of RNA editing sites, four PREP predicted editing sites were selected for C-to-U RNA editing analysis using RT-qPCR. The *ndhA* (position 1073), *ndhB* (position 149), *rps14* (position 80), and *rps16* (position 212) editing sites were comparatively quantified in different soybean tissues (**Figures 5A–D**). Five new putative editing sites, identified by SNP analysis, were also confirmed and quantified in leaf samples: *accD* (position 617), *ndhE* (position 233), *petB* (position 611), *rps2* (position 248), and *rps3* (position 383) (**Figure 5E**). RT-qPCR showed that the percentage of *ndhA* editing was higher in leaves (76.75%) than in petals (20.11%) or roots (30.23%) (**Figure 5A**). The same editing pattern was observed for *ndhB* and *rps14*. In *ndhB,* the percentage editing was 72.41, 30.54, and 16.55% (**Figure 5B**), while values of 74, 17.86, and 8.15% were obtained in *rps14* editing in the leaves, petals and roots, respectively (**Figure 5C**). The *rps16* editing profile was different, with an editing percentage that was higher than 60% in all tissues (**Figure 5D**). With respect to putative new C-to-U editing sites identified using SNP analysis, RT-qPCR confirmed C-to-U editing events and demonstrated different editing rates among genes: *accD* (60.2%), *ndhE* (39.85%,) *petB* (54.3%), *rps2* (71.52%), and *rps3* (20.02%) (**Figure 5E**).

DISCUSSION

In the present study, we propose an additional resource and new method to identify conserved and new RNA editing sites in plastid RNA sequences. Currently, an increasing number of high-throughput sequencing data have become available. Among these datasets, there are substantial data corresponding to sRNA sequencing libraries. After analyzing some of these libraries, we observed that even without previous isolation of chloroplasts for further RNA extraction and sequencing, millions of chloroplast-derived sRNA reads could be recovered, reflecting mapping against the chloroplast genome. An important constraint of the presented method refers to the library quality and the read coverage of reference genomes.

In the present study, Arabidopsis libraries had the highest mean coverage using sRNA reads, which likely facilitated the recovery of the largest number of confirmed editing sites. The coverage percentage across genomes was different between species, with lower values detected in Arabidopsis. This result

TABLE 2 | *NdhB* C-to-U editing events by PREP and SNP approach using reads derived from sRNA-seq.

Organism	Codon change	Nucleotide position	AA change	AA position	Total coverage	Edited coverage	% Editing	SNP *P*-value	PREP score
Arabidopsis	TCA–TTA	149	S–L	50	40	32	80	4.8E-108	1
(1,539: 870)*	CCA–CTA	467	P–L	156	40	28	75	8.5E-109	1
	CAT–TAT	586	H–Y	196	1	0	no editing	-	1
	TCA–TTA	611	S–L	204	5	0	no editing	-	0.8
	TCT–TTT	746	S–F	249	12	5	41.7	5.3E-109	1
	TCA–TTA	830	S–L	277	20	9	45	8.4E-29	1
	TCA–TTA	836	S–L	279	21	10	47.6	1.1E-23	1
	GCC–GTC	842	T–I	281	19	2	10.5	1.3E-8	nd
	CAT–TAT	1,255	H–Y	419	47	47	100	nd	1
	CCA–CTA	1,481	P–L	494	34	14	41.2	5.5E-40	1
Soybean	CCT–CTT	74	P–L	25	4	0	no editing	nd	1
(1,533: 543)*	TCA–TTA	149	S–L	50	35	10	28	3.3E-11	1
	ACG–ATG	542	T–M	181	1	1	100	nd	1
	CAT–TAT	586	H–Y	196	11	2	18.2	0.0000038	1
	TCA–TTA	611	S–L	204	14	0	no editing	nd	0.8
	CCA–CTA	737	P–L	246	2	2	100	nd	1
	TCT–TTT	746	S–F	249	12	4	33.3	2.0E-14	1
	TCA–TTA	830	S–L	277	12	5	41.7	3E-17	1
	TCA–TTA	836	S–L	279	11	5	45.5	2.6E-15	1
	TCA–TTA	1,112	S–L	371	22	5	22.7	4.3E-17	1
	CAT–TAT	1,255	H–Y	419	1	0	no editing	nd	1
	CCT–CTT	1,391	P–L	464	9	2	22.7	0.0000036	1
	CCC–TCC	1,414	P–S	472	10	0	no editing	nd	1
	CCA–CTA	1,481	P–L	494	13	8	64.3	1.3E-31	1
Rice	AGC–AGT	258	S–S	86	8	2	25	1.6E-8	nd
(1,533: 619)*	CCA–CTA	467	P–L	156	14	9	64.3	1.30E-31	1
	CAT–TAT	586	H–Y	196	5	3	60	4.00E-12	1
	TCA–TTA	611	S–L	204	2	1	50	nd	0.8
	TCC–TTC	704	S–F	235	16	3	18.8	7.10E-08	1
	CCA–CTA	737	P–L	246	0	0	nd	nd	1
	TCA–TTA	830	S–L	277	3	1	33	nd	1
	TCA–TTA	836	S–L	279	4	1	25	nd	1
	CTC–TTC	850	L–F	284	2	0	no editing	nd	0.6
	ACT–ATT	1,454	T–I	485	30	0	no editing	nd	0.6
	CCA–CTA	1,481	P–L	494	6	5	83	8.0E-17	1

*Coding sequence length and coverage values.

"Nucleotide position": position in base pair is from the A of the initiator codon.

"Total Coverage": total mapped reads in respective nucleotide position.

"Edited Coverage": number of reads shown T, instead C.

"% Editing": percentage of RNA editing using the edited reads divided by total mapped reads.

"PREP score": confidence value of prediction according PREP.

"nd": no defined.

demonstrated that the use of sRNA libraries for mapping editing events is not directly related to a significant coverage across the entire plastid genome. Although this method has the capacity to confirm and discover editing sites in chloroplasts, a smaller number of mitochondrial reads would likely affect RNA editing analysis in this organelle. In the present study, the approach for the identification of editing sites was compared to the PREP and SNP strategies. The editing sites and percentage editing may vary between species because some species may already possess a thymine in the genome. In these cases, C-to-U editing will not occur. The same situation can occur with some A-to-I editing

sites, which could affect the general percentage of editing among species. The use of a different PREP score, resulting in distinct cut-off values, may also affect these percentages. In addition, editing factors and their editing sites may evolve differently among species.

The elementary step employed in the pipeline used in the present study was the initial sRNA library mapping against the chloroplast genome, considering 0 mismatches. Plastid DNA insertions in nuclear genomes have been demonstrated for partial, intact or even truncated coding sequences in several species (Chen et al., 2015). Thus, an initial filtration step against

Organism	Nucleotide position	Nucleotide change	Total coverage	Edited coverage	% Editing	Variant *P*-value
Arabidopsis (74: 3,015)*	35	A–G	80	32	40	3.8E-655
Soybean (74: 65,787)*	35	A–G	80	47	58.5	2.3E-144
Rice (74: 1,673)*	35	A–G	214	145	67.8	1.5E-465

*tRNA sequence length and coverage values.
"Total Coverage": total mapped reads in respective nucleotide position.
"Edited Coverage": number of reads shown G, instead A.
"% Editing": percentage of RNA editing using the edited reads divided by total mapped reads.

the chloroplast genome prevents the loss of unedited reads to those loci present in nuclear insertions. Unedited reads are necessary, particularly in quantitative editing analysis, where the editing percentage is measured and cannot be ruled out.

Some C-to-U editing studies have previously used mRNA-Seq to demonstrate and quantify editing events in plant mitochondria (Bentolila et al., 2013) and chloroplasts (Guo et al., 2015). Comparison of sRNAs and mRNA data sequences demonstrated that most of the confirmed editing sites can be recovered using both datasets. However, there are differences between these data, demonstrating that sRNAs can identify editing sites that were not detected using mRNA data and vice versa (**Figure 3**). The use of sRNA data to complement RNA editing analysis can improve the identification and measurement of RNA editing in various aspects.

In the present study, a new set of plastid editing sites was identified in soybean. The C-to-U editing events have previously been demonstrated in other species, and we recovered several edited transcripts, including *ndhB*, *ndhD*, *ndhG*, *rpoB*, and *rpoC1* (Corneille et al., 2000; Okuda et al., 2009; Zhou et al., 2009; Chateigner-Boutin et al., 2011; Boussardon et al., 2012; Tseng et al., 2013), in the present analysis. For most known C-to-U editing sites predicted through PREP and confirmed by sRNA reads in the present study, 21 sites have previously been demonstrated in Arabidopsis (Tsudzuki et al., 2001; Tillich et al., 2005) and 19 sites have previously been demonstrated in rice (Corneille et al., 2000; Tsudzuki et al., 2001), representing 30.43 and 24% of the total predicted editing sites, respectively (Table S2). Moreover, we showed editing events in soybean plastid genes, including *ndhA*, *psaI*, and *petB*, which had not previously been demonstrated for rice or Arabidopsis. In the SNP analysis, we identified new C-to-U editing sites. For example, in the Arabidopsis *ndhF* gene, a putative C-to-U editing site was identified at position 884, leading to a serine to phenylalanine change. In the soybean *ndhE* gene, a putative C-to-U editing site at position 233 was observed in 73.7% of the reads. This editing led to a proline to leucine change in the encoded protein. Despite this information, the impact of amino acid modifications on respective protein structures remains unclear. Both *ndh* genes encode thylakoid Ndh complex components involved in photosynthesis optimization under different stress conditions conditions (Casano, 2001; Martin et al., 2004; Rumeau

et al., 2007). *NdhB* mutants under lower air humidity conditions or following exposure to ABA present a reduction in the photosynthetic level, likely mediated through stomatal closure triggered under these conditions (Horvath, 2000). Therefore, a protein structure modification, resulting from a loss or decrease in RNA editing events could affect adaptations to stress conditions or cause other unknown changes.

The coding sequence of protein D2, encoded by the *psbD* gene, a photosystem II (PSII) core protein, showed a putative new editing event in rice at positions 1006 and 1007. However, reflecting low coverage, these new editing sites still require further experimental confirmation. Maintenance of the D2 protein structure is important not only for proton transport (Pokhrel et al., 2013) but also for the phosphorylation dynamics of this protein (Tikkanen and Aro, 2012) and its interaction with the proteins responsible for PSII maintenance (Liu and Last, 2015). If this editing site is confirmed, then alterations in editing site patterns resulting from factors, such as abiotic stress could be associated with photo-oxidative damage susceptibility. Previous studies have demonstrated that abiotic stress influences the editing process and consequently plastid physiology (Nakajima and Mulligan, 2001; Karcher and Bock, 2002).

Five putative C-to-U editing sites predicted using SNP analysis were validated through RT-qPCR. This result demonstrates the reliability and accuracy of sRNA data resources and the method presented herein to confirm predicted sites *in silico* and identify new RNA editing sites. Position 1073 in the *ndhA* gene is an editing site identified only in the soybean chloroplast editome. RT-qPCR revealed that the editing percentage varies among different soybean tissues. The *ndhB* (position 149) gene was previously evaluated in the non-photosynthetic tissues of Arabidopsis. An RNA editing pattern previously demonstrated in Arabidopsis (Tseng et al., 2013), with a higher percentage in leaves (>75% edited), followed by flowers (25–75% edited) and roots (unedited), was similarly observed in the present study. An exception was observed for the root tissue, which showed a low editing percentage (16.5%) in soybean instead of an unedited rate, as observed in Arabidopsis. The editing site at position 80 in *rps14* also was evaluated across different tissues in Arabidopsis. A high editing percentage was demonstrated in Arabidopsis leaves (Tseng et al., 2013), a pattern also demonstrated in soybean using RT-qPCR. The RNA editing percentages observed in roots
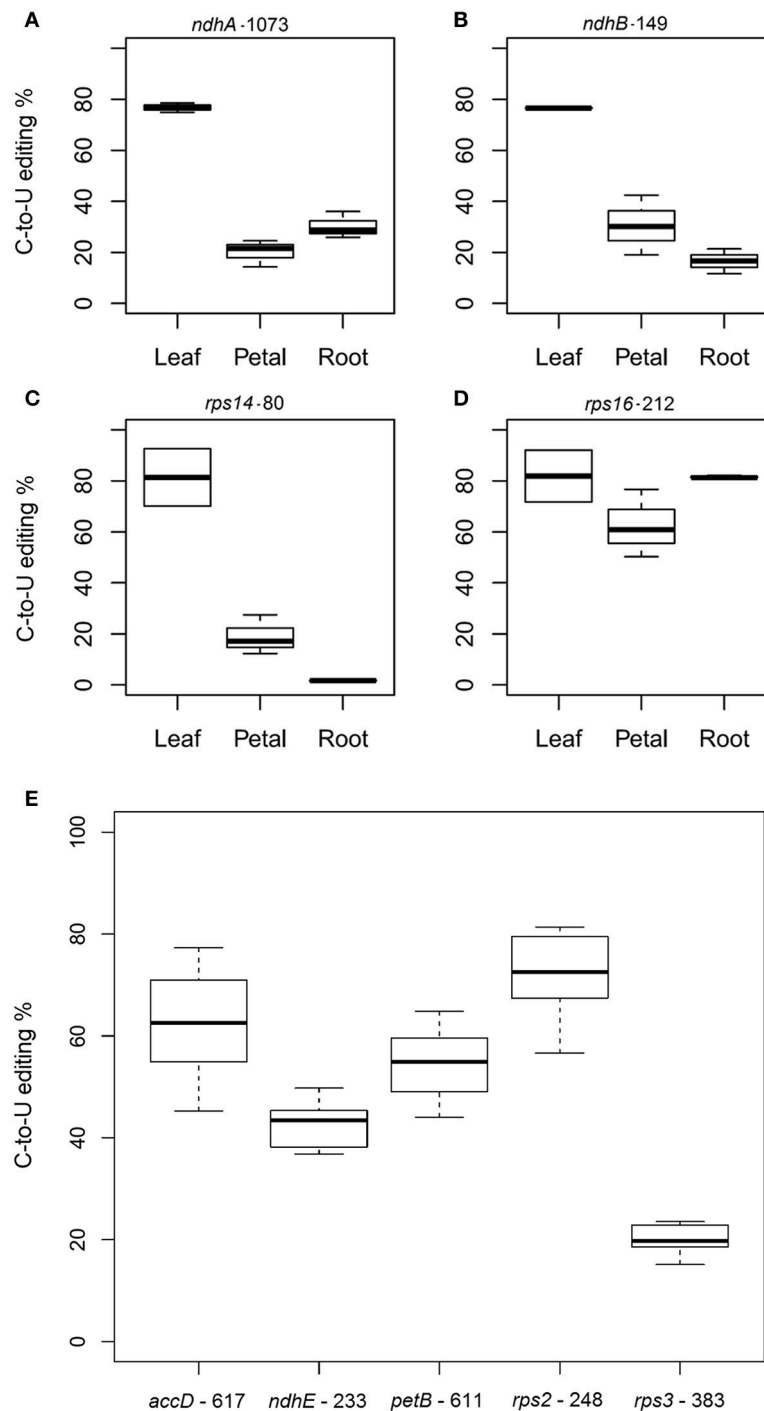
FIGURE 5 | Confirmation and quantitation of soybean editing sites predicted by PREP. **(A)** *ndhA-1053*, **(B)** *ndhB*-149, **(C)** *rps14*-80, and **(D)** *rps16*-212 were analyzed in leaves, petals and roots. Box area represents the lower and upper percentiles; **(E)** confirmation and quantitation of soybean editing sites identified by SNP analysis. Transcripts from soybean leaves were analyzed for C-to-U editing in specific nucleotide positions: *accD*-617, *ndhE*-233, *petB*-611, *rps2*-248, and *rps3*-383. Box area represents the lower and upper percentiles. The upper whisker of the boxplot indicates the highest editing value observed; the lower whisker, the lowest editing value; and the middle line, the median.

and petals showed different patterns between Arabidopsis and soybean, although a decrease in these values was observed in the root tissue of both species. The editing of *rps16* at position 212

was predicted and confirmed only in soybean and did not show differences in the editing percentage between leaf and root tissues. These results indicate that sRNA sequence mapping can not only

be used to confirm the predicted editing sites, but also to quantify the editing percentage.

The plastid acetyl-CoA carboxylase, necessary for *de novo* fatty acid synthesis, comprises two components, *accA* and *accD* proteins; *accD* encodes the β-carboxyl transferase subunit and is required in tobacco plants for a functional enzyme (Kode et al., 2005). The vanilla cream1 (vac1) albino mutant, reflecting a PPR-DYW protein required for editing in *accD* and *ndhF* in Arabidopsis, exhibits albino to pale yellow phenotype and an RNA editing reduction in those transcripts (Tseng et al., 2010). The requirement of plastid *accD* editing for functional protein has previously been demonstrated (Sasaki et al., 2001), and this new editing site, which promotes a serine to leucine change, could also be important for the maintenance of protein structure and functionality. The *ndhE* gene encodes a subunit of a membrane subcomplex of the NAD(P)H dehydrogenase complex (Peng et al., 2011). NdhE protein interacts with the membrane subcomplex proteins, NdhC and NdhG, and with subcomplex proteins, NhdH and NdhK (Efremov et al., 2010; Peng et al., 2011). The new editing site described here promotes a proline to leucine change, which could modify the interaction between these proteins and lead to changes in electron transfer to quinone. The *petB* gene encodes the cytochrome b_6 protein, a cytochrome b_6f complex component responsible for mediating electron transfer between photosystem I (PSI) and plastocyanin (Baniulis et al., 2008); mutants of *petB* in tobacco showed reduced levels of PSI, PSII and light-harvesting complex proteins (Monde et al., 2000), indicating a requirement of cytochrome b_6 to correct photosynthetic apparatus assembly. The new editing site involving a serine to leucine change in *petB* at position 611, identified in the present study, could be required for the maintenance of cytochrome b_6f complex structure and stability. Proteins S2 and S3 are located on the solvent side of ribosome small subunit (Manuell et al., 2004), and RNA editing events can modify their interactions among other ribosomal proteins and likely with mRNA, with potential effects on the regulatory aspects of plastid translation in response to stress or other homeostasis processes.

The SNP analysis facilitated the evaluation of not only C-to-U editing but also A-to-I editing events in chloroplast tRNAs. The tRNA-Arg (ACG) A-to-I editing event was also observed in all three species in the present study. This change corresponds to an inosine in the wobble position, which encodes three arginine codons CGU, CGC, and CGA that play a critical role in plastid protein synthesis (Rogalski et al., 2008). The enzyme involved in this mechanism in Arabidopsis, At1g68720, encodes a tRNA adenosine deaminase (TADA), which is targeted to plastids. RNAi lines of this gene show markedly reduced A-to-I editing efficiency, displaying phenotype consequences, such as growth and development delays (Elias and Huang, 2005; Delannoy et al., 2009; Karcher and Bock, 2009). Editing events in others tRNAs have been shown in some species and have been well studied in animals (Su and Randau, 2011) and previously demonstrated in moss *Takakia lepidozioides* (Miyata et al., 2008). The method described here can help to identify and measure other tRNA editing events not yet described in plants.

In addition to the high amount of data currently available in public databases that can readily be assessed, there are some

plastid sRNAs biological features that can reveal important mechanisms of RNA editing. The precise plastid sRNA biogenesis remains unknown because there is no evidence of any RNAi machinery in organelles that could originate small RNAs thus far. Notably, there is evidence of a relaxed plastid genome transcription mechanism, resulting in full plastid genome transcription (Hotto et al., 2012). It has been suggested that plastid sRNAs originated from RNA sequence regions protected against degradation by forming secondary structures or from associations with RNA-binding proteins regions (Pfalz et al., 2009). The results of the present study demonstrated that sRNAs are not necessarily over-represented in regions of editing sites but are also evident in coding sequences with smaller lengths, where these sRNAs can still be observed. These biological features enable the use of sRNA datasets to confirm the results of different RNA editing prediction tools and enable the analysis of editing events not only in a qualitative but also a quantitative manner, depending on the library quality and read coverage.

The identification of editing sites and measurement of editing levels have demonstrated differences among tissues (Tseng et al., 2013) and developmental stages (Miyata and Sugita, 2004). These findings can be used to evaluate the impact of different stresses on these mechanisms (Nakajima and Mulligan, 2001; Van Den Bekerom et al., 2013). Thus, the use of sRNA data to confirm predicted editing sites in association with SNP searches can provide a powerful and reliable plastid editome characterization and measurement, and the results can be applied to compare editing levels in different tissues, developmental stages and physiological conditions.

CONCLUSION

Analysis of sRNA libraries can be used to identify and quantify RNA editing events. Using this source of sequence data and pipeline of analyses, we obtained, for the first time, a consistent set of non-conserved and new editing sites in soybean. We propose the use of plastid sRNA libraries as a novel source and approach to study RNA editing events. Until recently, no other studies have taken advantage of such data to screen for RNA editing sites. Thus, the results from the present study should encourage researchers to use small RNA libraries to compare RNA editing in different plants under different conditions to improve knowledge on the editing role of plastid RNA in plant biology.

AUTHOR CONTRIBUTIONS

RM, NR, and AC conceived and designed the study. NR conducted *in silico* analysis. NR and FK conducted the RT-qPCR experiments. NR and GdF analyzed the data. NR and AC drafted the manuscript. All authors have read and approved the manuscript.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fpls.2017.01686/full#supplementary-material

REFERENCES

Baniulis, D., Yamashita, E., Zhang, H., Hasan, S. S., and Cramer, W. A. (2008). Structure-function of the cytochrome b 6 f complex. *Photochem. Photobiol.* 84, 1349–1358. doi: 10.1111/j.1751-1097.2008.00444.x

Barkan, A., and Small, I. (2014). Pentatricopeptide repeat proteins in plants. *Annu. Rev. Plant Biol.* 65, 415–442. doi: 10.1146/annurev-arplant-050213-040159

Manuell, A., Beligni, M. V., Yamaguchi, K., and Mayfield, S. P. (2004). Regulation of chloroplast translation: interactions of RNA elements, RNA-binding proteins and the plastid ribosome. *Biochem. Soc. Trans.* 32, 601–605. doi: 10.1042/BST0320601

Bentolila, S., Oh, J., Hanson, M. R., and Bukowski, R. (2013). Comprehensive high-resolution analysis of the role of an Arabidopsis gene family in RNA editing. *PLoS Genet.* 9:e1003584. doi: 10.1371/journal.pgen.1003584

Boussardon, C., Salone, V., Avon, A., Berthome, R., Hammani, K., Okuda, K., et al. (2012). Two interacting proteins are necessary for the editing of the NdhD-1 Site in Arabidopsis plastids. *Plant Cell* 24, 3684–3694. doi: 10.1105/tpc.112.099507

Casano, L. M. (2001). Hydrogen peroxide mediates the induction of chloroplastic Ndh complex under photooxidative stress in barley. *Plant Physiol.* 125, 1450–1458. doi: 10.1104/pp.125.3.1450

Chateigner-Boutin, A.-L., des Francs-Small, C. C., Delannoy, E., Kahlau, S., Tanz, S. K., de Longevialle, A. F., et al. (2011). OTP70 is a pentatricopeptide repeat protein of the E subgroup involved in splicing of the plastid transcript rpoC1. *Plant J.* 65, 532–542. doi: 10.1111/j.1365-313X.2010.04441.x

Chateigner-Boutin, A. L., and Hanson, M. R. (2003). Developmental co-variation of RNA editing extent of plastid editing sites exhibiting similar cis-elements. *Nucleic Acids Res.* 31, 2586–2594. doi: 10.1093/nar/gkg354

Chateigner-Boutin, A.-L., and Small, I. (2010). Plant RNA editing. *RNA Biol.* 7, 213–219. doi: 10.4161/rna.7.2.11343

Chen, H., Yu, Y., Chen, X., Zhang, Z., Gong, C., Li, J., et al. (2015). Plastid DNA insertions in plant nuclear genomes: the sites, abundance and ages, and a predicted promoter analysis. *Funct. Integr. Genomics* 15, 131–139. doi: 10.1007/s10142-014-0422-z

Chen, Y., Kao, S., Chou, H., Lin, W., Wong, F.-H., and Chow, W.-Y. (2008). A real-time PCR method for the quantitative analysis of RNA editing at specific sites. *Anal. Biochem.* 375, 46–52. doi: 10.1016/j.ab.2007.12.037

Cheng, S., Gutmann, B., Zhong, X., Ye, Y., Fisher, M. F., Bai, F., et al. (2016). Redefining the structural motifs that determine RNA binding and RNA editing by pentatricopeptide repeat proteins in land plants. *Plant J.* 85, 532–547. doi: 10.1111/tpj.13121

Corneille, S., Lutz, K., and Maliga, P. (2000). Conservation of RNA editing between rice and maize plastids: are most editing events dispensable? *Mol. Gen. Genet.* 264, 419–424. doi: 10.1007/s004380000295

da Fonseca, G. C., de Oliveira, L. F. V., de Morais, G. L., Abdelnor, R. V., Nepomuceno, A. L., Waterhouse, P. M., et al. (2016). Unusual RNA plant virus integration in the soybean genome leads to the production of small RNAs. *Plant Sci.* 246, 62–69. doi: 10.1016/j.plantsci.2016.01.011

Delannoy, E., Le Ret, M., Faivre-Nitschke, E., Estavillo, G. M., Bergdoll, M., Taylor, N. L., et al. (2009). Arabidopsis tRNA adenosine deaminase arginine edits the wobble nucleotide of chloroplast tRNAArg(ACG) and Is essential for efficient chloroplast translation. *Plant Cell* 21, 2058–2071. doi: 10.1105/tpc.109.066654

Efremov, R. G., Baradaran, R., and Sazanov, L. A (2010). The architecture of respiratory complex I. *Nature* 465, 441–445. doi: 10.1038/nature09066

Elias, Y., and Huang, R. H. (2005). Biochemical and Structural Studies of A-to-I Editing by tRNA:A34 Deaminases at the Wobble Position of Transfer RNA. *Biochemistry* 44, 12057–12065. doi: 10.1021/bi050499f

Germain, A., Hanson, M. R., and Bentolila, S. (2015). High-throughput quantification of chloroplast RNA editing extent using multiplex RT-PCR mass spectrometry. *Plant J.* 83, 546–554. doi: 10.1111/tpj.12892

Guo, W., Grewe, F., and Mower, J. P. (2015). Variable frequency of plastid RNA editing among ferns and repeated loss of uridine-to-cytidine editing from vascular plants. *PLoS ONE* 10:e0117075. doi: 10.1371/journal.pone.0117075

Hayes, M. L., Dang, K. N., Diaz, M. F., and Mulligan, R. M. (2015). A conserved glutamate residue in the C-terminal deaminase domain of pentatricopeptide repeat proteins is required for RNA editing activity. *J. Biol. Chem.* 290, 10136–10142. doi: 10.1074/jbc.M114.631630

Horvath, E. M. (2000). Targeted inactivation of the plastid ndhB gene in tobacco results in an enhanced sensitivity of photosynthesis to moderate stomatal closure. *Plant Physiol.* 123, 1337–1350. doi: 10.1104/pp.123.4.1337

Hotto, A. M., Germain, A., and Stern, D. B. (2012). Plastid non-coding RNAs: emerging candidates for gene regulation. *Trends Plant Sci.* 17, 737–744. doi: 10.1016/j.tplants.2012.08.002

Karcher, D., and Bock, R. (2002). Temperature sensitivity of RNA editing and intron splicing reactions in the plastid ndhB transcript. *Curr. Genet.* 41, 48–52. doi: 10.1007/s00294-002-0278-y

Karcher, D., and Bock, R. (2009). Identification of the chloroplast adenosine-to-inosine tRNA editing enzyme. *RNA* 15, 1251–1257. doi: 10.1261/rna.1600609

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., et al. (2012). Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649. doi: 10.1093/bioinformatics/bts199

Kim, D. D. Y. (2004). Widespread RNA editing of embedded alu elements in the human transcriptome. *Genome Res.* 14, 1719–1725. doi: 10.1101/gr.2855504

Kode, V., Mudd, E. A., Iamtham, S., and Day, A. (2005). The tobacco plastid accD gene is essential and is required for leaf development. *Plant J.* 44, 237–244. doi: 10.1111/j.1365-313X.2005.02533.x

Komine, Y., Kikis, E., Schuster, G., and Stern, D. (2002). Evidence for *in vivo* modulation of chloroplast RNA stability by 3'-UTR homopolymeric tails in *Chlamydomonas reinhardtii*. *Proc. Natl. Acad. Sci. U.S.A.* 99, 4085–4090. doi: 10.1073/pnas.052327599

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10:R25. doi: 10.1186/gb-2009-10-3-r25

Liu, J., and Last, R. L. (2015). A land plant-specific thylakoid membrane protein contributes to photosystem II maintenance in *Arabidopsis thaliana*. *Plant J.* 82, 731–743. doi: 10.1111/tpj.12845

Long, R. C., Li, M. N., Kang, J. M., Zhang, T. J., Sun, Y., and Yang, Q. C. (2015). Small RNA deep sequencing identifies novel and salt-stress-regulated microRNAs from roots of *Medicago sativa* and *Medicago truncatula*. *Physiol. Plant.* 154, 13–27. doi: 10.1111/ppl.12266

Chateigner-Boutin, A. L., and Small, I. (2007). A rapid high-throughput method for the detection and quantification of RNA editing based on high-resolution melting of amplicons. *Nucleic Acids Res.* 35:e114. doi: 10.1093/nar/gkm640

Martin, M., Casano, L. M., Zapata, J. M., Guera, A., del Campo, E. M., Schmitz-Linneweber, C., et al. (2004). Role of thylakoid Ndh complex and peroxidase in the protection against photo-oxidative stress: fluorescence and enzyme

activities in wild-type and ndhF-deficient tobacco. *Physiol. Plant.* 122, 443–452. doi: 10.1111/j.1399-3054.2004.00417.x

Milne, I., Stephen, G., Bayer, M., Cock, P. J. A., Pritchard, L., Cardle, L., et al. (2013). Using tablet for visual exploration of second-generation sequencing data. *Brief. Bioinformatics* 14, 193–202. doi: 10.1093/bib/bbs012

Miyata, Y., Sugita, C., Maruyama, K., and Sugita, M. (2008). RNA editing in the anticodon of tRNA Leu (CAA) occurs before group I intron splicing in plastids of a moss *Takakia lepidozioides* S. Hatt. & Inoue. *Plant Biol.* 10, 250–255. doi: 10.1111/j.1438-8677.2007.00027.x

Miyata, Y., and Sugita, M. (2004). Tissue- and stage-specific RNA editing of rps14 transcripts in moss (*Physcomitrella patens*) chloroplasts. *J. Plant Physiol.* 161, 113–115. doi: 10.1078/0176-1617-01220

Monde, R.-A., Zito, F., Olive, J., Wollman, F.-A., and Stern, D. B. (2000). Post-transcriptional defects in tobacco chloroplast mutants lacking the cytochrome b6/f complex. *Plant J.* 21, 61–72. doi: 10.1046/j.1365-313x.2000.00653.x

Mower, J. P. (2009). The PREP suite: predictive RNA editors for plant mitochondrial genes, chloroplast genes and user-defined alignments. *Nucleic Acids Res.* 37, W253–W259. doi: 10.1093/nar/gkp337

Nakajima, Y., and Mulligan, M. (2001). Heat stress results in incomplete C-to-U editing of maize chloroplast mRNAs and correlates with changes in chloroplast transcription rate. *Curr. Genet.* 40, 209–213. doi: 10.1007/s002940100249

Neto, L. B., Arenhart, R. A., de Oliveira, L. F. V., de Lima, J. C., Bodanese-Zanettini, M. H., Margis, R., et al. (2015). ASR5 is involved in the regulation of miRNA expression in rice. *Plant Cell Rep.* 34, 1899–1907. doi: 10.1007/s00299-015-1836-3

Okuda, K., Hammani, K., Tanz, S. K., Peng, L., Fukao, Y., Myouga, F., et al. (2009). The pentatricopeptide repeat protein OTP82 is required for RNA editing of plastid ndhB and ndhG transcripts. *Plant J.* 61, 339–349. doi: 10.1111/j.1365-313X.2009.04059.x

Okuda, K., Nakamura, T., Sugita, M., Shimizu, T., and Shikanai, T. (2006). A Pentatricopeptide repeat protein is a site recognition factor in chloroplast RNA editing. *J. Biol. Chem.* 281, 37661–37667. doi: 10.1074/jbc.M608184200

Peeters, N. M., and Hanson, M. R. (2002). Transcript abundance supercedes editing efficiency as a factor in developmental variation of chloroplast gene expression. *RNA* 8, 497–511. doi: 10.1017/S1355838202029424

Peng, L., Yamamoto, H., and Shikanai, T. (2011). Structure and biogenesis of the chloroplast NAD(P)H dehydrogenase complex. *Biochim. Biophys. Acta Bioenerg.* 1807, 945–953. doi: 10.1016/j.bbabio.2010.10.015

Pfalz, J., Bayraktar, O. A., Prikryl, J., and Barkan, A. (2009). Site-specific binding of a PPR protein defines and stabilizes 5′ and 3′ mRNA termini in chloroplasts. *EMBO J.* 28, 2042–2052. doi: 10.1038/emboj.2009.121

Pokhrel, R., Service, R. J., Debus, R. J., and Brudvig, G. W. (2013). Mutation of lysine 317 in the D2 subunit of photosystem II alters chloride binding and proton transport. *Biochemistry* 52, 4758–4773. doi: 10.1021/bi301700u

Rogalski, M., Karcher, D., and Bock, R. (2008). Superwobbling facilitates translation with reduced tRNA sets. *Nat. Struct. Mol. Biol.* 15, 192–198. doi: 10.1038/nsmb.1370

Rüdinger, M., Funk, H. T., Rensing, S. A., Maier, U. G., and Knoop, V. (2009). RNA editing: only eleven sites are present in the *Physcomitrella patens* mitochondrial transcriptome and a universal nomenclature proposal. *Mol. Genet. Genomics* 281, 473–481. doi: 10.1007/s00438-009-0424-z

Rumeau, D., Peltier, G., and Cournac, L. (2007). Chlororespiration and cyclic electron flow around PSI during photosynthesis and plant stress response. *Plant Cell Environ.* 30, 1041–1051. doi: 10.1111/j.1365-3040.2007.01675.x

Ruwe, H., and Schmitz-Linneweber, C. (2012). Short non-coding RNA fragments accumulating in chloroplasts: footprints of RNA binding proteins? *Nucleic Acids Res.* 40, 3106–3116. doi: 10.1093/nar/gkr1138

Ruwe, H., Wang, G., Gusewski, S., and Schmitz-Linneweber, C. (2016). Systematic analysis of plant mitochondrial and chloroplast small RNAs suggests organelle-specific mRNA stabilization mechanisms. *Nucleic Acids Res.* 44, 7406–7417. doi: 10.1093/nar/gkw466

Salone, V., Rüdinger, M., Polsakiewicz, M., Hoffmann, B., Groth-Malonek, M., Szurek, B., et al. (2007). A hypothesis on the identification of the editing enzyme in plant organelles. *FEBS Lett.* 581, 4132–4138. doi: 10.1016/j.febslet.2007.07.075

Sasaki, Y., Kozaki, A., Ohmori, A., Iguchi, H., and Nagano, Y. (2001). Chloroplast RNA editing required for functional acetyl-CoA Carboxylase in Plants. *J. Biol. Chem.* 276, 3937–3940. doi: 10.1074/jbc.M008166200

Schallenberg-Rüdinger, M., Lenz, H., Polsakiewicz, M., Gott, J. M., and Knoop, V. (2013). A survey of PPR proteins identifies DYW domains like those of land plant RNA editing factors in diverse eukaryotes. *RNA Biol.* 10, 1549–1556. doi: 10.4161/rna.25755

Simon, S. A., Zhai, J., Nandety, R. S., McCormick, K. P., Zeng, J., Mejia, D., et al. (2009). Short-read sequencing technologies for transcriptional analyses. *Annu. Rev. Plant Biol.* 60, 305–333. doi: 10.1146/annurev.arplant.043008.092032

Su, A. A. H., and Randau, L. (2011). A-to-I and C-to-U editing within transfer RNAs. *Biochemistry* 76, 932–937. doi: 10.1134/S0006297911080098

Sun, T., Bentolila, S., and Hanson, M. R. (2016). The unexpected diversity of plant organelle RNA editosomes. *Trends Plant Sci.* 21, 962–973. doi: 10.1016/j.tplants.2016.07.005

Takenaka, M., Zehrmann, A., Verbitskiy, D., Härtel, B., Brennicke, A., Barbara, H., et al. (2013). RNA editing in plants and its evolution. *Annu. Rev. Genet.* 47, 335–352. doi: 10.1146/annurev-genet-111212-133519

Tikkanen, M., and Aro, E. M. (2012). Thylakoid protein phosphorylation in dynamic regulation of photosystem II in higher plants. *Biochim. Biophys. Acta Bioenerg.* 1817, 232–238. doi: 10.1016/j.bbabio.2011.05.005

Tillich, M., Funk, H. T., Schmitz-Linneweber, C., Poltnigg, P., Sabater, B., Martin, M., et al. (2005). Editing of plastid RNA in *Arabidopsis thaliana* ecotypes. *Plant J.* 43, 708–715. doi: 10.1111/j.1365-313X.2005.02484.x

Tillich, M., Lehwark, P., Morton, B. R., and Maier, U. G. (2006). The evolution of chloroplast RNA editing. *Mol. Biol. Evol.* 23:1912. doi: 10.1093/molbev/msl054

Tseng, C. C., Lee, C. J., Chung, Y. T., Sung, T. Y., and Hsieh, M. H. (2013). Differential regulation of Arabidopsis plastid gene expression and RNA editing in non-photosynthetic tissues. *Plant Mol. Biol.* 82, 375–392. doi: 10.1007/s11103-013-0069-5

Tseng, C. C., Sung, T. Y., Li, Y. C., Hsu, S. J., Lin, C. L., and Hsieh, M. H. (2010). Editing of accD and ndhF chloroplast transcripts is partially affected in the *Arabidopsis vanilla* cream1 mutant. *Plant Mol. Biol.* 73, 309–323. doi: 10.1007/s11103-010-9616-5

Tsudzuki, T., Wakasugi, T., and Sugiura, M. (2001). Comparative analysis of RNA editing sites in higher plant chloroplasts. *J. Mol. Evol.* 53, 327–332. doi: 10.1007/s002390010222

Van Den Bekerom, R. J. M., Dix, P. J., Diekmann, K., and Barth, S. (2013). Variations in efficiency of plastidial RNA editing within ndh transcripts of perennial ryegrass (*Lolium perenne*) are not linked to differences in drought tolerance. *AoB Plants* 5:plt035. doi: 10.1093/aobpla/plt035

Wolf, P. G., Rowe, C. A., and Hasebe, M. (2004). High levels of RNA editing in a vascular plant chloroplast genome: analysis of transcripts from the fern *Adiantum capillus-veneris. Gene* 339, 89–97. doi: 10.1016/j.gene.2004.06.018

Wu, C., Li, X., Guo, S., and Wong, S.-M. (2016). Analyses of RNA-Seq and sRNA-Seq data reveal a complex network of anti-viral defense in TCV-infected *Arabidopsis thaliana. Sci. Rep.* 6:36007. doi: 10.1038/srep36007

Xu, T., Wang, Y., Liu, X., Lv, S., Feng, C., Qi, M., et al. (2015). Small, RNAs and degradome sequencing reveals microRNAs and their targets involved in tomato pedicel abscission. *Planta* 242, 963–984. doi: 10.1007/s00425-015-2318-0

Zhelyazkova, P., Hammani, K., Rojas, M., Voelker, R., Vargas-Suarez, M., Borner, T., et al. (2012). Protein-mediated protection as the predominant mechanism for defining processed mRNA termini in land plant chloroplasts. *Nucleic Acids Res.* 40, 3092–3105. doi: 10.1093/nar/gkr1137

Zhou, W., Cheng, Y., Yap, A., Chateigner-Boutin, A.-L., Delannoy, E., Hammani, K., et al. (2009). The Arabidopsis gene YS1 encoding a DYW protein is required for editing of rpoB transcripts and the rapid development of chloroplasts during early growth. *Plant J.* 58, 82–96. doi: 10.1111/j.1365-313X.2008.03766.x

Zimmer, S. L., Schein, A., Zipor, G., Stern, D. B., and Schuster, G. (2009). Polyadenylation in Arabidopsis and *Chlamydomonas organelles*: the input of nucleotidyltransferases, poly(A) polymerases and polynucleotide phosphorylase. *Plant J.* 59, 88–99. doi: 10.1111/j.1365-313X.2009.03853.x

REDIdb 3.0: A Comprehensive Collection of RNA Editing Events in Plant Organellar Genomes

Claudio Lo Giudice[1], Graziano Pesole[1,2] and Ernesto Picardi[1,2]*

[1] Institute of Biomembranes, Bioenergetics and Molecular Biotechnologies, Consiglio Nazionale delle Ricerche, Bari, Italy,
[2] Department of Biosciences, Biotechnology and Biopharmaceutics, University of Bari A. Moro, Bari, Italy

RNA editing is an important epigenetic mechanism by which genome-encoded transcripts are modified by substitutions, insertions and/or deletions. It was first discovered in kinetoplastid protozoa followed by its reporting in a wide range of organisms. In plants, RNA editing occurs mostly by cytidine (C) to uridine (U) conversion in translated regions of organelle mRNAs and tends to modify affected codons restoring evolutionary conserved aminoacid residues. RNA editing has also been described in non-protein coding regions such as group II introns and structural RNAs. Despite its impact on organellar transcriptome and proteome complexity, current primary databases still do not provide a specific field for RNA editing events. To overcome these limitations, we developed REDIdb a specialized database for RNA editing modifications in plant organelles. Hereafter we describe its third release containing more than 26,000 events in a completely novel web interface to accommodate RNA editing in its genomics, biological and evolutionary context through whole genome maps and multiple sequence alignments. REDIdb is freely available at http://srv00.recas.ba.infn.it/redidb/index.html

Keywords: organellar genomes, RNA editing, plant database, mitochondria, chloroplasts

INTRODUCTION

RNA editing is an essential co/post transcriptional process able to expand transcriptome and proteome diversity in addition to alternative splicing. The term RNA editing was first introduced in 1986 to describe the addition and deletion of uridine nucleotides to and from mRNAs in trypanosome mitochondria (Benne et al., 1986). Since then, RNA editing events have been found in a wide range of organisms and can occur in the nucleus and cytoplasm as well as in organelles (Bowe and depamphilis, 1996). Modifications due to RNA editing comprise nucleotide substitutions and insertions or deletions that can affect both protein coding and Non-protein coding RNAs (Maier et al., 1996; Steinhauser et al., 1999).

In humans, the most prevalent type of RNA editing event is the deamination of adenosine (A) in inosine (I) in double RNA strands (dsRNAs) through the catalytic activity of the adenosine deaminase (ADAR) family of enzymes. To date, more than 4 million events have been collected and annotated in dedicated resources such as DARNED, RADAR, and REDIportal (Kiran et al., 2013; Ramaswami and Li, 2014; Picardi et al., 2017).

In plants, RNA editing occurs mostly in organelles in the form of cytidine (C) to uridine (U) conversion particularly in translated regions of mRNAs, albeit the opposite event (U-to-C substitutions) has been observed in some taxa, especially in chloroplasts RNAs (Takenaka et al., 2013). Plant RNA editing sites are recognized by specific pentatricopeptide repeat (PPR) proteins that are encoded in the nuclear genome. In flowering plants, the editosome machinery requires several additional Non-PPR protein factors, even though its molecular assembly has yet to be clarified (Sun et al., 2016).

Most of the C-to-U changes in the protein coding regions tends to modify affected codons restoring evolutionary conserved aminoacid residues (Gray, 2003). Therefore, plant RNA editing is believed to act as an additional proofreading mechanism to generate fully functional proteins. Occasionally, C-to-U modifications occur in untranslated regions, structural RNAs and intervening sequencing, affecting splicing and translation efficiency. Indeed, RNA editing changes in the domain V of plant group II introns is mandatory for the splicing process (Castandet et al., 2010).

With the advent of high-throughput sequencing technologies, many complete plant organellar genomes have been released and numerous novel RNA editing events uncovered. Nevertheless, RNA editing changes are not always correctly or completely annotated in primary databases (GenBank, ENA and DDBJ) and an appropriate field to unambiguously describe them is not provided. RNA editing modifications are often reported as misc_feature or even as simple exception notes. With the aim to overcome these limitations and create a cured catalog of plant RNA editing events, we developed the specialized REDIdb database. Its first release stored 9,964 modifications distributed over 706 different nucleotide sequences, increased to 11,897 in the following update.

After 10 years of massively parallel sequencing, we present here REDIdb 3.0, an upgraded release that annotates 26,618 RNA editing events distributed among 281 organisms and 85 complete organellar genomes.

All changes have been recovered from Genbank and literature using a semi-automated bioinformatics procedure in which each annotation has been manually checked to avoid redundancy or inconsistencies due to errors in flatfiles.

The web-interface was totally restyled and developed using the latest computational technologies in the field of database querying and managing.

Furthermore, many computational facilities have been integrated to improve the user experience and ensure continuous and future updates of the database. Indeed, REDIdb 3.0 accommodates RNA editing in its genomics, biological and evolutionary context through whole genome maps and multiple sequence alignments.

Although a variety of RNA editing databases have been released such as DARNED (Kiran et al., 2013), RADAR (Ramaswami and Li, 2014), and REDIportal (Picardi et al., 2017), REDIdb is the only one devoted to editing changes in plant organelles. Indeed, similar resources such as dbRES (He et al., 2007), RESOPS (Yura et al., 2009), ChloroplastDB (Cui et al.,

2006), or GOBASE (O'Brien et al., 2009) have been dismissed or not updated.

MATERIALS AND METHODS

All editing events stored in REDIdb derive from GenBank flatfiles through a semi-automated parsing algorithm implemented in custom python (2.7.13) scripts. Each flatfile is screened for RNA editing features using the SeqIO parser included in the Biopython (1.68) module (Cock et al., 2009).

All annotations have been manually checked to identify and correct potential errors, taking into account other related flatfile fields or literature. REDIdb database is organized in MySQL tables and queries are in python employing the MySQL-python (1.2.5) module, a data access library to MySQL engine. The web interface, instead, is built in BootStrap (3.3.7), while data presentation is based on DataTables, an *ad hoc* Javascript library (1.10.13) to efficiently show large tables in html documents. Genome rendering, available for complete organellar genomes, has been developed in pure python, mimicking OGDraw graphics (Lohse et al., 2013).

Query results are dynamically generated using the CGI (common gateway interface) technology. Multiple sequence alignments of edited cDNAs and proteins have been generated by ClustalOmega (Sievers et al., 2011) and displayed in html pages through the MSAViewer (Yachdav et al., 2016), a JavaScript component of the BioJS collection (https://biojs.net/).

The distribution of RNA editing events along functional domains and predicted protein secondary structures are shown by the feature-viewer JavaScript library (https://github.com/calipho-sib/feature-viewer) based on the powerful D3 JavaScript library for visualizing data using web standards (https://d3js.org/). Functional domains have been detected using InterPro engine (Jones et al., 2014), while protein secondary structures have been predicted using the stand-alone version of Spider2 program (Yang et al., 2017).

All the scripts to parse multiple alignments, InterPro html files and Spider2 outputs have been created in Python. Scripts used to extract RNA editing positions from Genbank flatfiles are freely available at the REDIdb help page. Additional details and supplementary scripts are available upon request.

RESULTS

Database Content

Previous REDIdb release contained 11,897 editing events distributed over 198 organisms and 929 different nucleotide sequences. This upgraded version, instead, collects more than 26,000 editing events from 281 organisms, 85 complete organellar genomes and 3,467 sequences. REDIdb 3.0 includes 26,545 events in protein coding sequences and 73 in untranslated regions, structural RNAs and introns. The vast majority of editing changes occur in the mitochondrion, accounting for a total of 23,553 events over 2,300 sequences.

The most recurrent RNA editing modification is the C-to-U substitution, that accounts for more than 92% of all

TABLE 1 | Number of RNA Editing events in complete genomes stored in REDIdb.

Location	Protein_coding_events	No_protein_coding_events	Genbank_id	Organism_name	Cultivar	Strain
mito	6		NC_021931	*Anomodon attenuatus*		
mito	6		NC_016121	*Anomodon rugelii*		
mito	493		NC_001284	*Arabidopsis thaliana*		
mito	2		NC_024520	*Atrichum angustatum*		
mito	1		NC_024519	*Bartramia pomiformis*		
mito	344		NC_015994	*Beta macrocarpa*		
mito	340		NC_015099	*Beta vulgaris* subsp. maritima		
mito	4		NC_031212	*Brachythecium rivulare*		
mito	417		AP006444	*Brassica napus*	Westar	
mito	467		KJ820683	*Brassica oleracea* var. botrytis		
mito		5	KJ820683	*Brassica oleracea* var. botrytis		
mito	2		NC_024518	*Buxbaumia aphylla*		
mito	2		NC_012116	*Carica papaya*	SunUp	
mito	547		NC_014043	*Citrullus lanatus*	Florida giant	
mito	3		NC_024515	*Climacium americanum*		
mito	475		NC_014050	*Cucurbita pepo*	Dark green zucchini	
mito	719		NC_027976	*Ginkgo biloba*		
mito		8	AY182006	*Harpochytrium* sp. JEL105		JEL105
mito		5	AY182005	*Harpochytrium* sp. JEL94		JEL94
mito	1		AP014526	*Hevea brasiliensis*	BPM 24	
mito	1		AP017300	*Hordeum vulgare* subsp. spontaneum		H602
mito	1		AP017301	*Hordeum vulgare* subsp. vulgare		HarunaNijo
mito	3		NC_026515	*Hyoscyamus niger*		
mito	5		NC_024516	*Hypnum imponens*		
mito	888		KC821969	*Liriodendron tulipifera*		
mito	488		NC_016743	*Lotus japonicus*		
mito		1	NC_016743	*Lotus japonicus*		
mito	488		JN872551	*Lotus japonicus*		MG-20
mito		1	JN872551	*Lotus japonicus*		MG-20
mito	485		NC_016742	*Millettia pinnata*		
mito		1	NC_016742	*Millettia pinnata*		
mito		10	AY182007	*Monoblepharella* sp. JEL15		JEL15
mito	847		NC_030753	*Nelumbo nucifera*		
mito	3		NC_029805	*Nicotiana sylvestris*	TW 137	
mito	1		NC_006581	*Nicotiana tabacum*	Bright Yellow 4	
mito	60		NC_012651	*Nothoceros aenigmaticus*		
mito		1	NC_012651	*Nothoceros aenigmaticus*		
mito	1		NC_029356	*Orthotrichum diaphanum*		
mito	1		NC_029355	*Orthotrichum macrocephalum*		
mito	2		NC_024522	*Orthotrichum stellatum*		
mito	3		AP017386	*Oryza sativa* Indica Group		BT-CMS
mito	97		NC_013765	*Phaeoceros laevis*		
mito	16		NC_017755	*Phlegmariurus squarrosus*		
mito	8		NC_013444	*Pleurozia purpurea*		
mito	8		NC_024514	*Ptychomnion cygnisetum*		
mito	60		JQ083668	*Raphanus sativus*		

(Continued)

TABLE 1 | Continued

Location	Protein_ coding_events	No_protein_ coding_events	Genbank_id	Organism_name	Cultivar	Strain
mito	7		NC_027974	*Sanionia uncinata*		
mito	326		NC_014487	*Silene latifolia*		
mito	3		NC_024521	*Sphagnum palustre*		
mito	1		NC_017840	*Spirodela polyrhiza*		7498
mito	1		NC_027515	*Syntrichia filaris*		
mito	5		KC784953	*Tetraphis pellucida*		
mito	2		NC_028191	*Tetraplodon fuegianus*		
mito	1		NC_016122	*Treubia lacunosa*		
mito	2		NC_024517	*Ulota hutchinsiae*		
mito	7		NC_015121	*Vigna radiata*		
mito	109		AY506529	*Zea mays*		NB
chloro	79		KU764518	*Actinostachys pennula*		
chloro	343		AY178864	*Adiantum capillus-veneris*		
chloro		1	AY178864	*Adiantum capillus-veneris*		
chloro	564		NC_004543	*Anthoceros formosae*		
chloro	56		NC_019628	*Apopellia endiviifolia*		
chloro	25		LC154068	*Arabidopsis lyrata* subsp. lyrata		MN47
chloro	37		NC_000932	*Arabidopsis thaliana*		
chloro	36		NC_004561	*Atropa belladonna*		Ab5p(kan)
chloro	1		NC_031894	*Citrus depressa*		
chloro	1		NC_009618	*Cycas taitungensis*		
chloro	65		NC_028542	*Cyrtomium devexiscapulae*		
chloro	55		NC_028705	*Cyrtomium falcatum*		
chloro	2		NC_031159	*Ipomoea nil*	Tokyo-kokei standard	
chloro	1		NC_016058	*Larix decidua*		
chloro	3		NC_017006	*Mankyua chejuensis*		
chloro	38		NC_016708	*Millettia pinnata*		
chloro		1	NC_016708	*Millettia pinnata*		
chloro	58		NC_014592	*Myriopteris lindheimeri*		
chloro	40		NC_001879	*Nicotiana tabacum*	Bright Yellow 4	
chloro	5		NC_007602	*Nicotiana tomentosiformis*		
chloro	44		AY916449	*Phalaenopsis aphrodite* subsp. formosana	Taisugar TS-97	
chloro	1		NC_005087	*Physcomitrella patens* subsp. patens		
chloro	23		NC_001631	*Pinus thunbergii*		
chloro	66		NC_014348	*Pteridium aquilinum* subsp. aquilinum		
chloro	16		NC_005878	Saccharum hybrid cultivar SP80-3280		
chloro	1		NC_020098	*Tectona grandis*		
chloro	72		NC_028543	*Woodwardia unigemmata*		
chloro	32		NC_001666	*Zea mays*		

Events are divided by sequence (coding/Non-coding) and according to their intracellular location. In presence of multiple accession numbers for the same organism, only the RefSeq record (if present in Genbank) has been considered.

annotated events and, when located in protein coding regions, tends to modify the aminoacid coded by the edited codon. Indeed, the majority of RNA editing events affects the first and second codon position leading to aminoacid changes resulting the most conserved in the comparison with related orthologs.

FIGURE 1 | (A) REDIdb query form. Searches can be performed by Organism, Location, Gene or a combination of them. Additional filters (RefSeq, Exons, Full Orfs) are also available in order to refine the results. **(B)** Once a query has been submitted, the corresponding results are displayed in a sortable and exportable table report.
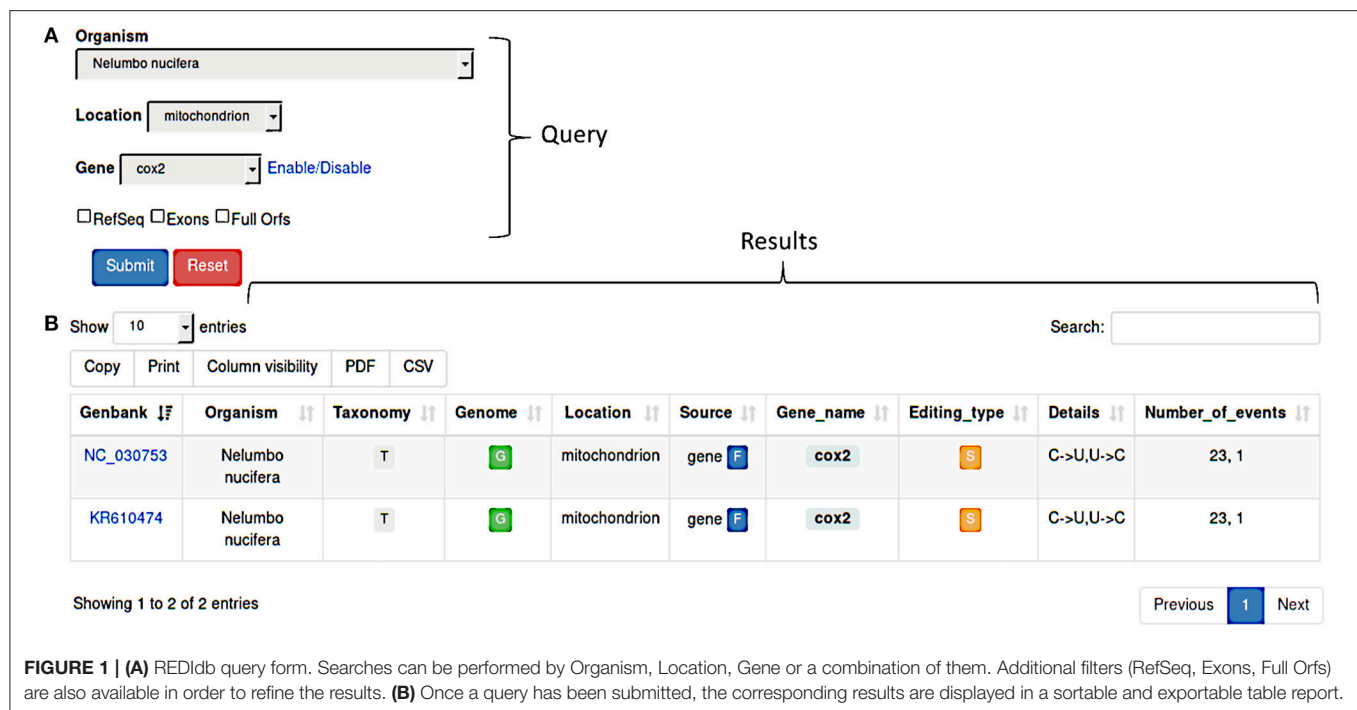
Differently from the previous releases, the novel REDIdb database annotates 85 complete organellar genomes. Of these 57 are mitochondrial genomes and include 7791 events. As reported in **Table 1**, the most edited mitochondrial genomes are those from *Liriodendron tulipifera*, *Nelumbo nucifera* and *Ginkgo biloba* with 888, 847, and 717 events, respectively. Of 27 annotated chloroplast genomes, instead, the one from *Anthoceros formosae* comprising 564 modifications results the richest in editing events.

All REDIdb sequences including RNA editing events are identified by unique accession numbers (e.g., EDI0000.). To preserve the full compatibility with previous database versions, accession numbers linked to old entries have been maintained unchanged.

Query Form and Output Tables

REDIdb implements a modular query form (**Figure 1A**) allowing users to make flexible searches by selecting the organism or the intracellular location or the gene name. Regarding nucleotide sequences, users can retrieve the original sequence submitted to the primary database or the RefSeq version or both. In addition, the search can be limited to full open reading frames and include individual exons in case of interrupted genes.

Query results are shown in a sortable and exportable summary table (**Figure 1B**) comprising several info such as the GenBank accession number, the organism and the link to the related taxonomy, the organelle type and the link to the complete genome (if available), the gene name and a flag indicating its partial or full nature, the editing types and details and the

total number of events. Column can be selectively included in the final table and results are downloadable in pdf or csv format. The "Taxonomy" column includes a link to an interactive taxonomy chart, while the "Genome" column contains a link to the complete genome (if available in primary databases) chart in which RNA editing events are displayed in their genomics context.

Using the link in the "Gene_name" column, users can browse individual RNA editing events organized in flatfiles.

Entry Organization

RNA editing events stored in REDIdb are organized in specific flat-files comprising four main sections. The first section (**Figure 2A**) contains a general description of the entry including the organism name, the taxonomy (according with the NCBI Taxonomy database), the GenBank and PubMed accession numbers, the intracellular location (mitochondrion or chloroplast) and the official gene name.

The second section (**Figure 2B**) is devoted to Gene Ontologies (GO), obtained by matching each protein sequence contained in REDIdb against the InterPro database (Finn et al., 2017). In the case of protein coding genes, it contains information regarding the molecular functions, the biological processes and the cellular localization of the protein product. The third section (**Figure 2C**) shows all the editing features that characterize the record. Here, for each editing event the position on the transcript is reported and, if the complete reference genome is available, also the genomic location. In case of editing within protein coding genes, the genomic codon, edited codon and aminoacidic change are determined and reported. Finally, the fourth

GENERAL INFORMATIONS

ORGANISM Nelumbo nucifera

DIVISION Eukaryota, Viridiplantae, Streptophyta, Embryophyta, Tracheophyta, Spermatophyta, Magnoliophyta, eudicotyledons, Proteales, Nelumbonaceae, Nelumbo

LOCATION mitochondrion

STATUS complete genome

SEQUENCE Full

SOURCE gene

NAME cox2

GENBANK NC_030753

PUBMED 27444405

A

GO Gene Ontology

MOLECULAR FUNCTION GO:0004129 cytochrome-c oxidase activity

BIOLOGICAL PROCESS GO:0022900 electron transport chain

CELLULAR COMPONENT GO:0016020 membrane

B

EDITING FEATURES

Genomic position	cDNA position	Genomic codon	Edited codon	Genomic AA	Edited AA
105715	161	tCa	tTa	S	L
105713	163	Cgg	Tgg	R	W
105849	27	ctC	ctT	L	L
105598	278	cCg	cTg	P	L
105843	33	atC	atT	I	I
105522	354	atC	atT	I	I
105497	379	Cgg	Tgg	R	W
105838	38	cCt	cTt	P	L
103699	443	aCg	aTg	T	M
103682	460	Cca/CCa	Tca/TTa	P	S/L
103681	461	cCa/CCa	cTa/TTa	P	L/L
103666	476	tCa	tTa	S	L
103598	544	Cct	Tct	P	S
103585	557	cCt	cTt	P	L
103561	581	tCa	tTa	S	L
103528	614	tCa	tTa	S	L
103510	632	tCg	tTg	S	L
103466	676	Cgt	Tgt	R	C
103447	695	tCt	tTt	S	F
103444	698	aCg	aTg	T	M
100871	700	Tct	Cct	S	P
105805	71	tCt	tTt	S	F
100850	721	Cct	Tct	P	S
100829	742	Cgg	Tgg	R	W

C

SEQUENCES

GENOMIC SEQUENCE 783 bp FASTA

BASE COUNT 214a 169c 159g 241t

```
    1 ATGATTGTTC TAGAATGGCT ATTCCTCACA ATCGCTCCTT GTGATGCTGC GGAACCATGG CAATTAGGAT CTCAAGACGC AGCAACACCT ATGATGCAAG GAATAATGGA CTTACATCAC
  121 GATATCTTTT TCTTCCTCAT TCTGATTTTG GTTTTCGTAT CACGGATGTT GGTTCGCGCT TTATGGCATT TCCACTATCA AACGAACCCA ATCCCGCAAA GGATTGTTCA TGGAACTACT
  241 ATCGAGATTC TTTGGACCAT ATTTCCTAGT ATCATCCCGA TGTTCATTGC TATACCATCA TTTGCTCTGT TATACTCAAT GGACGAGGTA GTAGTAGATC CAGCCATTAC TATCAAAGCT
  361 ATTGGACATC AATGGTATCG GACTTATGAG TATTCGGACT ATAACAGTTC CGATGAACAG TCACTCACTT TTGACAGTTA TACGATTCCA GAAGTAGATC CAGAATTGGG TCAATCACGT
  481 TTATTAGAAG TGGACAATAG AGTGGTTGTA CCAGCCAAAA CTAATCTACG TATTATTGTA ACACCTGCTG ATGTACCTCA TAGTTGGGCT GTACCTTCCT CAGGTGTCAA ATGTGATGCT
  601 GTACCTGGTC GTTCAAATCA GACCTCTATT TCGGTACAAC GAGAAGGAGT TTACTATGGT CAGTGCAGTG AGATTCGTGG AACTAATCAT GCCTCTACGT CTATCGTCGT AGAAGCTGTT
  721 CCTAGGAAAG ATTATGGTTC TCGGGTATCC AATCAATTAA TCCCCCAAAC CGGGGAAGCT TAA
```

cDNA SEQUENCE 783 bp FASTA MSA

BASE COUNT 214a 147c 159g 263t

```
    1 ATGATTGTTC TAGAATGGCT ATTCCTTACA ATTGCTCTTT GTGATGCTGC GGAACCATGG CAATTAGGAT TTCAAGACGC AGCAACACCT ATGATGCAAG GAATAATGGA CTTACATCAC
  121 GATATCTTTT TCTTCCTCAT TCTGATTTTG GTTTTCGTAT TATGGATGTT GGTTCGCGCT TTATGGCATT TCCACTATCA AACGAACCCA ATCCCGCAAA GGATTGTTCA TGGAACTACT
  241 ATCGAGATTC TTTGGACCAT ATTTCCTAGT ATCATCCCGA TGTTCATTGC TATACCATCA TTTGCTCTGT TATACTCAAT GGACGAGGTA GTAGTAGATC CAGCCATTAC TATTAAAGCT
  361 ATTGGACATC AATGGTATTG GACTTATGAG TATTCGGACT ATAACAGTTC CGATGAACAG TCACTCACTT TTGACAGTTA TAGTATTCCA GAAGTAGATT TAGAATTGGG TCAATTACGT
  481 TTATTAGAAG TGGACAATAG AGTGGTTGTA CCAGCCAAAA CTAATCTACG TATTATTGTA ACATCTGCTG ATGTACTTCA TAGTTGGGCT GTACCTTCCT TAGGTGTCAA ATGTGATGCT
  601 GTACCTGGTC GTTTAAATCA GACCTCTATT TTGGTACAAC GAGAAGGAGT TTACTATGGT CAGTGCAGTG AGATTTGTGG AACTAATCAT GCCTTTATGC CTATCGTCGT AGAAGCTGTT
  721 TCTAGGAAAG ATTATGGTTC TTGGGTATCC AATCAATTAA TCCCCCAAAC CGGGGAAGCT TAA
```

PROTEIN SEQUENCE 261 AA FASTA MSA DOMAINS

```
    1 mivlewlflt ialcdaaepw qlgfqdaatp mmqgimdlhh diffflilil vfvlwmlvra lwhfhyqtnp ipqrivhgtt ieilwtifps iilmfiaips fallysmdev vvdpaitika
  121 ighqwywtye ysdynssdeq sltfdsymip eddlelgqlr llevdnrvvv paktnlriiv tsadvlhswa vpslgvkcda vpgrlnqtsi lvqregvyyg qcseicgtnh afmpivveav
  241 srkdygswvs nqlipqtgea x
```

D

FIGURE 2 | Editing informations stored in REDIdb are organized in specific flat-files in which it is possible to distinguish a header **(A)** containing the main features of the record (organism, Genbank accession, intracellular location, gene name, PubMed references, ecc.), a gene ontology box **(B)** describing the gene product properties, a feature table **(C)** with all the editing events and a sequence zone **(D)** with both the genomic sequence and the corresponding edited transcript/protein.

section (**Figure 2D**) contains the genomic sequence and the corresponding edited transcript. In coding protein genes, also the edited protein is displayed. Genomic sequences as well as edited transcripts and proteins can be retrieved in Fasta format.

Graphical Visualization

Edited cDNA and protein sequences can be explored in their evolutionary context through multiple alignments of available orthologs sequences. Since plant RNA editing tends to increase the sequence conservation along the evolution, annotated RNA editing changes are marked and visualized in the multiple alignment by the MSAViewer, to give rise to conservation levels and provide valuable comparative genomics information (**Figure 3A**).

In addition, RNA editing events are displayed along the edited sequence showing known functional domains and predicted secondary protein structures in order to better interpret the biological role of specific C-to-U or U-to-C changes (**Figure 3B**).

In case of complete organellar genomes, each genome is graphically rendered and edited genes can be selectively highlighted. Genome graphs are generated in SVG and include links to edited genes by mousing over. Further statistics such as

the coding potential of the genome as well as the fraction of edited genes are also reported (**Figure 4**).

CONCLUSIONS AND PERSPECTIVES

As already mentioned, RNA editing plays an important role in transcriptome and proteome diversity. Since its first discovery in 1986 (Benne et al., 1986), a large number of events have been found in a wide range of eukaryotic organisms (Ichinose and Sugita, 2016). Only in humans more than 4 million events have been reported and dedicated resources such as DARNED, RADAR, and REDIportal have been developed to contain them into suitable specialized databases (Kiran et al., 2013; Ramaswami and Li, 2014; Picardi et al., 2017).

In the plant kingdom, RNA editing was first identified as C-to-U substitutions in mitochondrial transcripts (Hiesel et al., 1989), followed by its identification also in chloroplasts (Höch et al., 1991). In order to maintain a cured catalog of such events, we developed the specialized REDIdb database. Its third release, described here, contains three times more entries than the first version and two times more entries than the second version. To date, REDIdb is the unique bioinformatics resource collecting plant organellar RNA editing events. Indeed, similar databases such as dbRES

FIGURE 3 | (A) Multiple alignments of cox2 orthologous sequences. Editing conservation across species can be easily obtained considering both the sequence logo and the bar chart relative to each position. **(B)** Protein domains and structure of *Nelumbo nucifera*'s cox2 protein. Protein domains are obtained by querying multiple databases (CDD, PROSITE, Pfam ecc.).

FIGURE 4 | In REDIdb 3.0 complete genomes are graphically rendered, allowing users to visualize the reciprocal order of the edited genes. Further statistics such as the coding potential of the genome as well as the fraction of edited genes are reported in the same page.

(He et al., 2007) or RESOPS (Yura et al., 2009) have been dismissed or are no more updated. Plant RNA editing events are also annotated in CloroplastDB (Cui et al., 2006), devoted to chloroplast genomes, and GOBASE (O'Brien et al., 2009), the organelle genome database. However, such resources are not specialized for RNA editing and include

potential not fixed errors due to the lack of manual curation (Picardi et al., 2011).

REDIdb 3.0 has been completely redrawn keeping in mind the simplicity as its working principle. RNA editing events are always shown in their biological context and novel graphical facilities have been added. Edited genes are now depicted in complete

genome maps and RNA editing conservation can be investigated in pre-calculated multiple alignments of orthologous sequences. REDIdb 3.0 allows also the visualization of aminoacid changes induced by RNA editing in protein domains or secondary structures, providing insights into the potential functional consequences.

Next generation sequencing technologies, now arrived at their third generation, are expected to greatly increase the number of RNA editing candidates in the next future. Therefore, it will be indispensable to collect and annotate them in their biological context taking into account also the RNA editing levels.

Due to the unicity in its field, REDIdb is planned to be maintained and updated over time (as new editing sites or complete genomes are released), taking into account, as much as possible, eventual feedbacks from the users.

AUTHOR CONTRIBUTIONS

CL conducted the bioinformatics analyses and wrote the first manuscript draft; EP and GP conceived the study and contributed to writing and revising the manuscript.

FUNDING

ACKNOWLEDGMENTS

REFERENCES

Benne, R., Van Den Burg, J., Brakenhoff, J. P. J., Sloof, P., Van Boom, J. H., and Tromp, M. C. (1986). Major transcript of the frameshifted coxII gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. *Cell* 46, 819–826. doi: 10.1016/0092-8674(86)90063-2

Bowe, L. M., and depamphilis, C. W. (1996). Effects of RNA editing and gene processing on phylogenetic reconstruction. *Mol. Biol. Evol.* 13, 1159–1166. doi: 10.1093/oxfordjournals.molbev.a025680

Castandet, B., Choury, D., Bégu, D., Jordana, X., and Araya, A. (2010). Intron RNA editing is essential for splicing in plant mitochondria. *Nucleic Acids Res.* 38, 7112–7121. doi: 10.1093/nar/gkq591

Cock, P. J., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., et al. (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* 25, 1422–1423. doi: 10.1093/bioinformatics/btp163

Cui, L., Veeraraghavan, N., Richter, A., Wall, K., Jansen, R. K., Leebens-Mack, J., et al. (2006). ChloroplastDB: the chloroplast genome database. *Nucleic Acids Res.* 34, D692–D696. doi: 10.1093/nar/gkj055

Finn, R. D., Attwood, T. K., Babbitt, P. C., Bateman, A., Bork, P., Bridge, A. J., et al. (2017). InterPro in 2017–beyond protein family and domain annotations. *Nucleic. Acids. Res.* 45, D190–D199. doi: 10.1093/nar/gkw1107

Gray, M. W. (2003). Diversity and evolution of mitochondrial RNA editing systems. *IUBMB Life* 55, 227–233. doi: 10.1080/1521654031000119425

He, T., Du, P., and Li, Y. (2007). dbRES: a web-oriented database for annotated RNA editing sites. *Nucleic Acids Res.* 35, D141–D144. doi: 10.1093/nar/gkl815

Hiesel, R., Wissinger, B., Schuster, W., and Brennicke, A. (1989). RNA editing in plant mitochondria. *Science* 246, 1632–1634. doi: 10.1126/science.2480644

Höch, B., Maier, R. M., Appel, K., Igloi, G. L., and Kossel, H. (1991). Editing of a chloroplast mRNA by creation of an initiation codon. *Nature* 353, 178–180. doi: 10.1038/353178a0

Ichinose, M., and Sugita, M. (2016). RNA editing and its molecular mechanism in plant organelles. *Genes* 8:5. doi: 10.3390/genes8010005

Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., Mcanulla, C., et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30, 1236–1240. doi: 10.1093/bioinformatics/btu031

Kiran, A. M., O'Mahony, J. J., Sanjeev, K., and Baranov, P. V. (2013). Darned in 2013: inclusion of model organisms and linking with Wikipedia. *Nucleic Acids Res.* 41, D258–D261. doi: 10.1093/nar/gks961

Lohse, M., Drechsel, O., Kahlau, S., and Bock, R. (2013). OrganellarGenomeDRAW–a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic Acids Res.* 41, W575–W581. doi: 10.1093/nar/gkt289

Maier, R. M., Zeltz, P., Kössel, H., Bonnard, G., Gualberto, J. M., and Grienenberger, J. M. (1996). RNA editing in plant mitochondria and chloroplasts. *Plant Mol. Biol.* 32, 343–365. doi: 10.1007/BF00039390

O'Brien, E. A., Zhang, Y., Wang, E., Marie, V., Badejoko, W., Lang, B. F., et al. (2009). GOBASE: an organelle genome database. *Nucleic Acids Res.* 37, D946–D950. doi: 10.1093/nar/gkn819

Picardi, E., D'erchia, A. M., Lo Giudice, C., and Pesole, G. (2017). REDIportal: a comprehensive database of A-to-I RNA editing events in humans. *Nucleic Acids Res.* 45, D750–D757. doi: 10.1093/nar/gkw767

Picardi, E., Regina, T. M., Verbitskiy, D., Brennicke, A., and Quagliariello, C. (2011). REDIdb: an upgraded bioinformatics resource for organellar RNA editing sites. *Mitochondrion* 11, 360–365. doi: 10.1016/j.mito.2010.10.005

Ramaswami, G., and Li, J. B. (2014). RADAR: a rigorously annotated database of A-to-I RNA editing. *Nucleic Acids Res.* 42, D109–D113. doi: 10.1093/nar/gkt996

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7:539. doi: 10.1038/msb.2011.75

Steinhauser, S., Beckert, S., Capesius, I., Malek, O., and Knoop, V. (1999). Plant mitochondrial RNA editing. *J. Mol. Evol.* 48, 303–312. doi: 10.1007/PL00006473

Sun, T., Bentolila, S., and Hanson, M. R. (2016). The Unexpected Diversity of Plant Organelle RNA Editosomes. *Trends Plant Sci.* 21, 962–973. doi: 10.1016/j.tplants.2016.07.005

Takenaka, M., Zehrmann, A., Verbitskiy, D., Härtel, B., and Brennicke, A. (2013). RNA editing in plants and its evolution. *Annu. Rev. Genet.* 47, 335–352. doi: 10.1146/annurev-genet-111212-133519

Yachdav, G., Wilzbach, S., Rauscher, B., Sheridan, R., Sillitoe, I., Procter, J., et al. (2016). MSAViewer: interactive JavaScript visualization of multiple sequence alignments. *Bioinformatics* 32, 3501–3503. doi: 10.1093/bioinformatics/btw474

Yang, Y., Heffernan, R., Paliwal, K., Lyons, J., Dehzangi, A., Sharma, A., et al. (2017). SPIDER2: a package to predict secondary structure, accessible surface area, and main-chain torsional angles by deep neural networks. *Methods Mol. Biol.* 1484, 55–63. doi: 10.1007/978-1-4939-6406-2_6

Yura, K., Sulaiman, S., Hatta, Y., Shionyu, M., and Go, M. (2009). RESOPS: a database for analyzing the correspondence of RNA editing sites to protein three-dimensional structures. *Plant Cell Physiol.* 50, 1865–1873. doi: 10.1093/pcp/pcp132

Transcriptome-Wide Annotation of m⁵C RNA Modifications Using Machine Learning

Jie Song [1,2†], Jingjing Zhai [1†], Enze Bian [3†], Yujia Song [3], Jiantao Yu [3] and Chuang Ma [1,2*]

[1] State Key Laboratory of Crop Stress Biology for Arid Areas, Center of Bioinformatics, College of Life Sciences, Northwest A&F University, Shaanxi, China, [2] Key Laboratory of Biology and Genetics Improvement of Maize in Arid Area of Northwest Region, Ministry of Agriculture, Northwest A&F University, Shaanxi, China, [3] College of Information Engineering, Northwest A&F University, Shaanxi, China

The emergence of epitranscriptome opened a new chapter in gene regulation. 5-methylcytosine (m⁵C), as an important post-transcriptional modification, has been identified to be involved in a variety of biological processes such as subcellular localization and translational fidelity. Though high-throughput experimental technologies have been developed and applied to profile m⁵C modifications under certain conditions, transcriptome-wide studies of m⁵C modifications are still hindered by the dynamic nature of m⁵C and the lack of computational prediction methods. In this study, we introduced PEA-m5C, a machine learning-based m⁵C predictor trained with features extracted from the flanking sequence of m⁵C modifications. PEA-m5C yielded an average AUC (area under the receiver operating characteristic) of 0.939 in 10-fold cross-validation experiments based on known *Arabidopsis* m⁵C modifications. A rigorous independent testing showed that PEA-m5C (Accuracy [Acc] = 0.835, Matthews correlation coefficient [MCC] = 0.688) is remarkably superior to the recently developed m⁵C predictor iRNAm5C-PseDNC (Acc = 0.665, MCC = 0.332). PEA-m5C has been applied to predict candidate m⁵C modifications in annotated *Arabidopsis* transcripts. Further analysis of these m⁵C candidates showed that 4nt downstream of the translational start site is the most frequently methylated position. PEA-m5C is freely available to academic users at: https://github.com/cma2015/PEA-m5C.

Keywords: AUC, Epitranscriptome, machine learning, RNA modification, RNA 5-methylcytosine

INTRODUCTION

The epitranscriptome, also known as chemical modifications of RNA (CMRs), is a newly discovered layer of gene expression (Meyer and Jaffrey, 2014). With advances in mass spectrometry and high-throughput sequencing technologies, the field of epitranscriptome is rapidly expanding and attracting a comparable degree of research interests to DNA and histone modifications in the field of epigenetics (Helm and Motorin, 2017). Among more than 150 types of CMRs identified, most of them have been found in transfer RNAs (tRNAs) and ribosomal RNAs (rRNAs) (Hussain et al., 2013), but some can occur in mRNAs and noncoding RNAs (Machnicka et al., 2013; Pan, 2013; Carlile et al., 2014; Dominissini et al., 2016; David et al., 2017). A growing line of evidences indicated that CMRs located in both coding and noncoding regions can play essential roles in a variety of biological processes. For instance, N⁶-methyladenosine (m⁶A) sites in 5′-untranslated

region (UTR) can promote cap-independent translation under heat stress (Meyer et al., 2015; Zhou et al., 2015); while m⁶A sites in coding regions can affect translation dynamics by inducing steric constraints and destabilizing pairing between codons and tRNA anticodons (Choi et al., 2016; Zhao et al., 2017). Thus, the transcriptome-wide annotation of RNA modifications is essential for fully understanding the biological functions of CMRs.

Compared with those well characterized modifications such as m⁶A and N^1-methyladenosine (m¹A), the transcriptome-wide annotation of 5-methylcytosine (m⁵C) modifications is more challenging. First, bisulfite sequencing technologies are difficult to implement for profiling m⁵C modifications because of the instability of mRNA molecules treated with bisulfite (Amort et al., 2013; Li et al., 2016). In addition, other existing high-throughput sequencing technologies, such as m⁵C-RIP (Edelheit et al., 2013), can localize m⁵C residues to transcript regions of 100–200 nucleotide (nt) long, but fail to accurately identify m⁵C modifications at single-nucleotide solution. Second, because of the dynamic nature of m⁵C (Wang and He, 2014), existing high-throughput sequencing technologies can only capture a snapshot of RNA modifications under certain experimental conditions, and cover just a small fraction of the whole transcriptome of a given sample (Zhou et al., 2016), resulting in the generation of significant numbers of false negatives (non-detected true m⁵C modifications). Third, the base preferences around the m⁵C sites are not strong enough, increasing the difficulties in computational predictions with traditional statistical approaches. Machine learning (ML) is a branch of artificial intelligence technology that has been widely used in engineering, computer science, informatics and biology (Ma et al., 2014a, 2017; Cui et al., 2015; Libbrecht and Noble, 2015; Zhai et al., 2016). The biggest advantage of ML systems is that they can automatically learn interesting patterns from existing datasets and bring about self-improvement of system performance for accurately predicting novel knowledge from a new data set (Ma et al., 2014a,b). Therefore, computational methods coupled with machine learning technologies may provide an option to accurately annotate RNA modifications like m⁵C in the transcriptome-wide manner.

Until now, iRNAm5C-PseDNC is the exclusive m⁵C predictor, which was built using random forest (RF) algorithm based on sequence-based features, and has been reported to have a good predictive performance for mammalian m⁵C prediction (Qiu et al., 2017). However, because of the lineage-specific sequence and structural properties differences between plant and mammalian species, tools developed for mammal species can't always retain their original performance when applied to other organisms (Leclercq et al., 2013; Zhai et al., 2017). This particular issue underscores the need for accurate transcriptome-wide m⁵C prediction tools in plants, which may lay a foundation for elucidating the mechanisms of formation and the cellular functions of m⁵C modifications.

In this study, we developed PEA-m5C, an accurate transcriptome-wide m⁵C predictor under a ML framework with an ensemble of 10 RF-based prediction models. PEA-m5C was trained with features extracted from the flanking sequence of m⁵C modifications, and showed promising performance when applied to predict m⁵C modifications in *Arabidopsis thaliana*. We further applied PEA-m5C to predict candidate m⁵C modifications in annotated *Arabidopsis* transcripts, and found that candidate m⁵C modifications are enriched in the coding region of mRNAs. In addition, 4-nt downstream of the translational start site is the most frequently methylated position. All candidate m⁵C modifications have been deposited in a public database named Ara-m5C for follow-up functional studies. In order to facilitate the application of PEA-m5C, we have implemented the proposed model into a cross-platform, user-friendly and interactive interface with R and JAVA programming languages.

MATERIALS AND METHODS

Dataset Generation

In this study, we constructed four m⁵C datasets: DatasetCV (cross-validation dataset), DatasetHT (hold-out test dataset), DatasetIT1 (independent test dataset for samples from the *Arabidopsis* silique tissue) and DatasetIT2 (independent test dataset for samples from the *Arabidopsis* shoot tissue).

DatasetCV and DatasetHT were constructed based on m⁵C modifications in transcripts expressed in the *Arabidopsis* root tissue at single-nucleotide resolution using RNA bisulfite sequencing technology (David et al., 2017). During bisulfite conversion, unmethylated cytosines were converted into uracils, while methylated cytosines were not converted. Bisulfite-treated RNA samples were sequenced to generated 100-nt paired-end reads using the Illumina HiSeq 2500. Low-quality reads were processed using Trimmomatic (Bolger et al., 2014), and the left clean reads were globally mapped to *in silico* bisulfite-converted *Arabidopsis* reference genome sequences using the RNA mode of B-Solana (Kreck et al., 2012). For each cytosine site in the *Arabidopsis* reference genome, the methylation level was calculated using a proportion statistic: $P = (C+\Psi)/(T+C)$, where C and T represent the number of cytosines and thymines in aligned reads at the cytosine site under analysis, respectively. Ψ specifies the added pseudo counts (1/8 counts). The false discovery rate (FDR) was calculated using the R package qvalue (Storey, 2002). Cytosines were regarded as positive samples (m⁵C modifications) if they satisfied the following criteria: methylation level $\geq 1\%$ and FDR ≤ 0.3. After the removal of sequence redundancy, we finally obtained 1,296 m⁵C modifications in 885 transcripts (**Table S1**). In these 885 transcripts, cytosines were regarded as negative samples (non-m⁵C modifications) if they were not annotated as m⁵C modifications. In order to avoid over-fitting and GC bias in training process, we limited the number of negative samples to be 10 times of positive samples. Thus, for each positive sample, 10 samples were selected in the 200-nt region around the positive sample, among which GC content difference is not more than 5%. This allows a similar distribution of positive and GC-matched negative samples, which is markedly different from the background distribution of all cytosines in these 885 transcripts (**Figure S1**). Note

that some of the negative samples may in fact be true m⁵C modifications not yet discovered. We randomly divided these 1,296 positive samples and 12,960 negative samples into two parts for constructing DatasetCV and DatasetHT, respectively. The DatasetCV comprises 1,196 positive samples and 11,960 negative samples, while the DatasetHT has a balanced number (100) of positive and negative samples (**Table S1**).

Using the same criteria mentioned above, another two datasets (DatasetIT1: 79 positive and negative samples; DatasetIT2: 73 positive and negative samples) were also constructed for *Arabidopsis* silique and shoot tissues, respectively (**Table S1**). Of note, positive and negative samples in DatasetIT1 and DatasetIT2 were not overlapped with those in DatasetCV and DatasetHT.

Each sample in these four datasets was represented by a sequence window of 43 nucleotides centered around the respective cytosine site. For samples near the borders of the available RNA sequence, the positions missing from the 43-nt window were filled with "N," the symbol for unknown. The *Arabidopsis* reference genome sequences (TAIR10) and annotated transcripts used in this study were downloaded from the Araport 11 database (https://www.araport.org/data/araport11).

Feature Encoding

In order to be recognized by ML-based systems, each sample of *L*-nt window size, was represented as a numeric vector (length: $4*L + 106$) using the binary, k-mer and PseDNC encoding schemes. The details of these three encoding schemes are described in the following.

Binary Encoding

This encoding strategy generates a vector of $4*L$ features by characterizing "A," "C,", "G," "U," and "N" with (1, 0, 0, 0), (0, 1, 0, 0), (0, 0, 1, 0), (0, 0, 0, 1), and (0, 0, 0, 0) for each sample, respectively.

K-mer Encoding

In this scheme, the composition of short sequence with different lengths was considered to explore its potential effect on the identification of m⁵C. In order to avoid the curse of dimensionality, we set $k = 1$, 2, and 3 to generate 84 features for calculating the frequency of mononucleotide occurrence ($k = 1$; four features), dinucleotide occurrence ($k = 2$; 16 features) and trinucleotide occurrence ($k = 3$; 64 features).

PseDNC Encoding

The pseudo dinucleotide composition (PseDNC) is a widely used encoding strategy that considers sequential information as well as physicochemical properties of dinucleotides in the RNA sequence (Chen et al., 2015, 2017). For each sample, it generates $16+\lambda$ numeric features, the first 16 of which are features extracted from adjacent dinucleotide pairs, and the other λ are features extracted from distant dinucleotide pairs (λ denotes the maximal distance between two dinucleotides). The detailed definition of PseDNC is presented in **Supplementary Data 1**.

Development of ML-Based m⁵C Predictor

Figure 1 illustrates the workflow of PEA-m5C, which consists of three phases, namely, (A) model construction, (B) model optimization, and (C) model prediction. Model construction and optimization were performed on the DatasetCV.
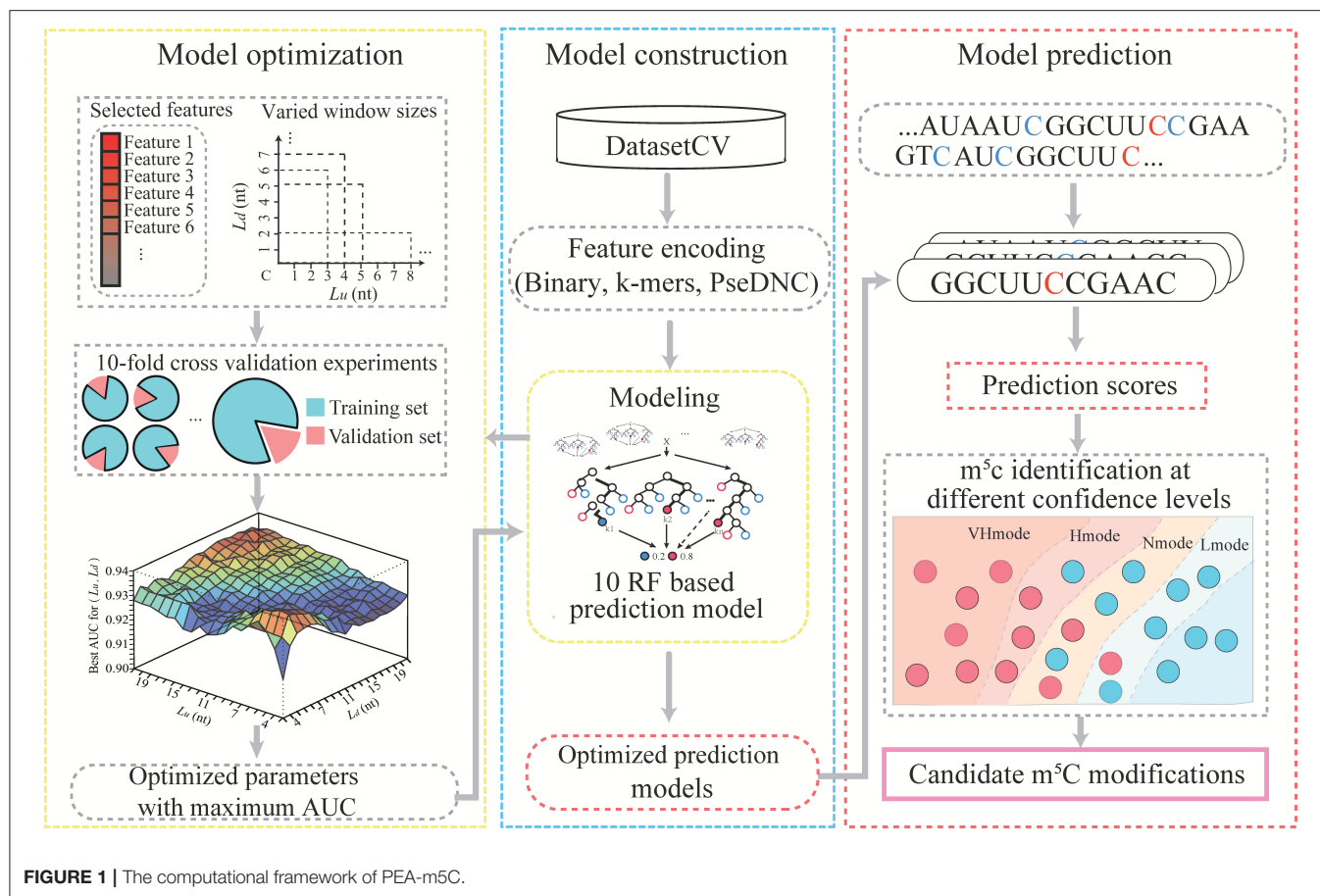
Model Construction

To construct an m⁵C prediction model, PEA-m5C required an input of a set of positive and negative samples. These samples were transformed into a feature matrix using three different encoding schemes (binary, k-mer, and PseDNC). The feature matrix was input into the RF algorithm to construct an m⁵C prediction model, which consisted of 100 classification trees. Each of the classification trees was built using a set of bootstrapped samples and features. The output of the RF-based m⁵C prediction model was determined by a majority vote of the classification trees. The RF algorithm was implemented using the R package "Rweka" (Hornik et al., 2009), which provides an R environment to invoke the ML package "weka" (v3.9.1; https://www.cs.waikato.ac.nz/ml/weka).

Model Optimization

Ten-fold cross-validation experiments were performed to optimize m⁵C prediction models in PEA-m5C by iteratively varying window size and feature number. Cross-validation is a standard method for estimating the generalization accuracy of ML systems. In a ten-fold cross-validation, the DatasetCV was randomly divided into 10 equal subsets and each subset was iteratively selected as a testing set for evaluating the model trained with other nine subsets. In each fold of cross-validation, considering the high unbalance between positive and negative samples (1:10), the negative samples were randomly divided into 10 parts, each of which coupled with the set of positive samples were used for training an RF-based m⁵C prediction model. Therefore, ten RF-based m⁵C prediction models were constructed in the training process. In the testing process, each sample was scored using these ten RF-based m⁵C prediction models. The corresponding ten prediction scores were averaged as the final prediction score of the sample under analysis. Once the testing process was completed, the prediction accuracy of PEA-m5C (an ensemble of ten RF-based m⁵C prediction models) was evaluated using the receiver operating characteristic (ROC) analysis, which plots a curve of false positive rate (FPR) varying at different true positive rate (TPR). The value under the ROC curve (AUC) was used to quantitatively score the prediction performance of PEA-m5C. AUC is ranged from 0 to 1, the higher the better prediction performance. After 10 subsets have been successively used as the testing set, the corresponding 10 AUC values were averaged as the overall prediction performance of PEA-m5C.

The PEA-m5C was optimized to maximize the AUC by iteratively varying window size *L* from 5- to 43-nt and feature number from 2 to $4*L+106$. The feature subset was selected according to the feature importance estimated using the information gain approach implemented in R package "FSelector" (Cheng et al., 2012). The detailed process of model optimization is given in **Figure 2**. We initialize AUC matrix

FIGURE 1 | The computational framework of PEA-m5C.

FIGURE 2 | The pseudo-code for model optimization.

("AUCMatrix") and feature matrix ("FMatrix") as two empty sets (**Lines 1-2**). Then for a given window size L (5-nt $\leq L \leq$ 43-nt) (**Line 3**), we varied the upstream sequence length (L_u) from 1-nt to (L-2)-nt and the number of feature subset from 2 to 4*L+106 (**Lines 4-7**). Subsequently, for each feature subset, we performed a 10-fold cross-validation experiment and stored the corresponding AUC value into a vector ("AUCVector") (**Lines 8-9**). After all possible feature subsets have been examined using 10-fold cross-validation experiments, the maximum AUC in "AUCVector" will be stored in the "AUCMatrix" (**Lines 11-12**), and the corresponding feature subset with maximum AUC will be stored in "FMatrix" (**Lines 13-15**). Finally, after all possible window sizes have been performed, the optimized L_u and L_d can be obtained by searching the maximum value in "AUCMatrix" (**Lines 18-19**), and the optimized feature subset can be obtained by searching "FMatrix" with L_u and L_d (**Lines 20-21**).

Model Prediction

PEA-m5C predicted all candidate m^{5}C modifications in given RNA sequences in FASTA format. For each cytosine site, PEA-m5C firstly extracted the flanking sequence with the optimized window size. Then, three feature encoding schemes were performed to transform the flanking sequence to a numeric vector. Subsequently, the optimized feature subset was input into the ten RF-based m5C prediction models. Finally, PEA-m5C generated a prediction score to reflect the possibility of this cytosine to be a real m^{5}C modification. Of note, four thresholds have been also included in the PEA-m5C, which were automatically determined in the 10-fold cross-validation at the specificity level of 99, 95, 90, and 85%, respectively. These four thresholds corresponded to four different confidence modes of PEA-m5C: VHmode (very high confidence mode), HMode (high confidence mode), NMode (normal confidence mode) and LMode (low confidence mode), respectively. Cytosine sites with a prediction score higher than the threshold were predicted as positive samples; otherwise, they were predicted as negative samples.

Model Comparisons

The iRNAm5C-PseDNC is only available m^{5}C predictor that aims to accurately predict m^{5}C modifications in mammalian genomes. It was constructed using the RF algorithm with only PseDNC features, and was trained with mammalian m^{5}C modifications (window size: 41-nt) (Sun et al., 2016). In order to fairly compare prediction performance between iRNAm5C-PseDNC and our proposed model PEA-m5C, we also re-trained iRNAm5C-PseDNC with positive and negative samples of 41-nt in the DatasetCV, and this re-trained predicted model was named as iRNAm5C-PseDNC*. Prediction performance of iRNAm5C-PseDNC, iRNAm5C-PseDNC* and PEA-m5C was estimated on DatasetHT, DatasetIT1 and DatasetIT2 using six widely used measures: sensitivity (Sn, also known as recall), specificity (Sp), precision (Pr), accuracy (Acc), F$_1$-score (F$_1$), and Matthews correlation coefficient (MCC). These measures were defined as

follows:

$$Sn = \frac{TP}{TP + FN},$$

$$Sp = \frac{TN}{TN + FP},$$

$$Pr = \frac{TP}{TP + FP},$$

$$Acc = \frac{TP + TN}{TP + TN + FP + FN},$$

$$F_1 = \frac{2*Pr \times Sn}{Pr + Sn} = \frac{2*TP}{2*TP + FP + FN},$$

$$MCC = \frac{TP*TN - FP*FN}{\sqrt{(TP + FP)*(TP + FN)*(TN + FP)*(TN + FN)}},$$

where TP, TN, FP, and FN represent the number of true positives, true negatives, false positives and false negatives, respectively. F$_1$ is the harmonic mean of Pr and Sn. Compared with Sn, Sp, Pr, and F$_1$, Acc and MCC are two more information measures which combine all of the predictions (TP, TN, FP, and FN) into a single score. Acc, which ranges from 0 to 1, measures the proportion of correct predictions. MCC, also known as the phi coefficient, measures the correlation between the observations and predictions. It is generally regarded as a balanced measure, which can be used even if the two classes are of very different size. The value of MCC ranges from -1 to 1, where 1 represents a perfect prediction, 0 indicates no better than random prediction and -1 means total disagreement between observations and predictions.

Transcriptome-Wide m^{5}C Annotation and Analysis

Candidate m^{5}C sites in the annotated *Arabidopsis* transcripts were predicted using the PEA-m5C. The spatial distribution of candidate m^{5}C modifications was statistically analyzed in three aspects: (i) feature enrichment (e.g., 5′-UTR, coding region [CDS] and 3′-UTR) analysis of candidate m^{5}C modifications in coding RNAs; (ii) the most frequently methylated position relative to the translational start site; (iii) functional enrichment analysis of genes containing candidate m^{5}C modifications.

The base preference around candidate m^{5}C modification sites was also explored, including: (i) the proportion of m^{5}C modifications in different sequence contexts: CG, CHG and CHH (H: A, T or C); (ii) sequence motifs of candidate m^{5}C modifications.

RESULTS

Characterization of m^{5}C Modifications Using Sequence-Based Features

To investigate whether m^{5}C modifications can be identified using sequence-based features, we first examined the positional frequencies of four bases in positive and negative samples in the DatasetCV (**Figures 3A,B**). We observed that the positional base frequency appears to be stable in negative samples. In
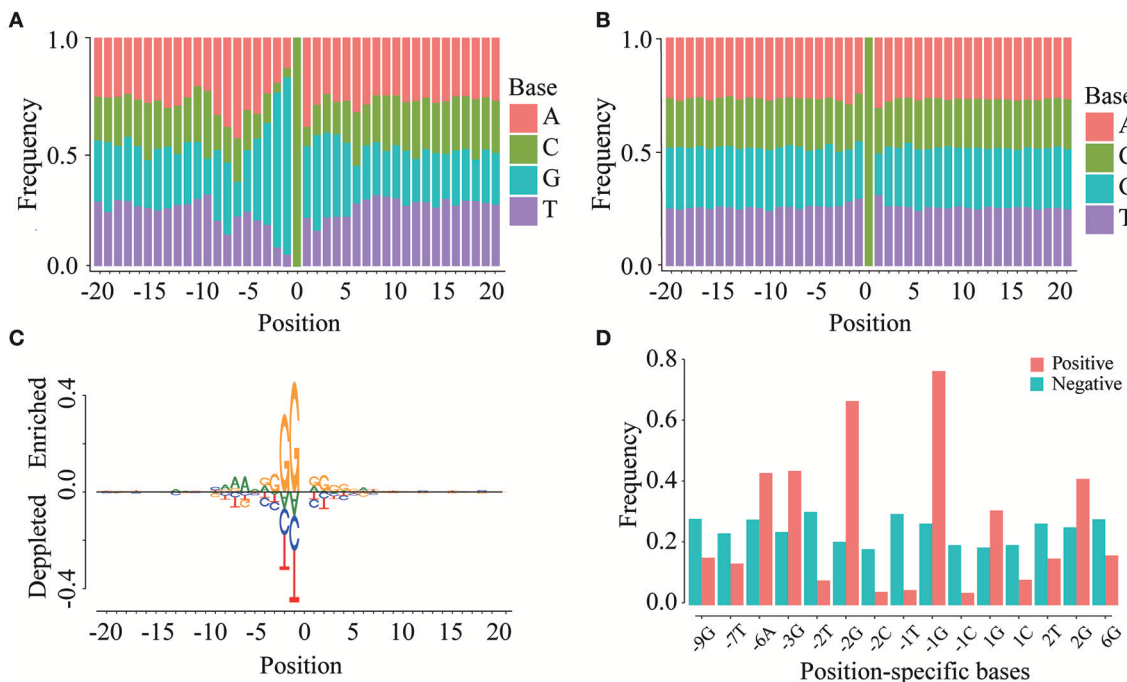
FIGURE 3 | Position-specific base usage in positive and negative samples. The region around (−20, 20) around the cytosine sites is used to perform position-specific base usage analysis, where the cytosines are defined as 0. **(A)** Frequencies of 41 × 4 position-specific bases in positive samples. **(B)** Frequencies of 41 × 4 position-specific bases in negative samples. **(C)** Two sample logo of the differences between m5C and non-m5C modifications. It shows nucleotides which are enriched or depleted in the surrounding regions of m5C modifications. **(D)** With a significance level of 1.0E-10, the usage of 15 position-specific bases is significantly different between m5C and non-m5C modifications.

contrast, the positional base frequency was biased to guanine (G) in the region near m5C sites in positive samples. We then detected position-specific base usages by using rank sum test. Setting significant level (p-value) to be 1.0E-10, we found that 15 position-specific base usages are significantly different between positive and non-m5C modifications. They are −9G, −7T, −6A, −3G, −2T, −2G, −2C, −1T, −1C, 1G, 1C, 2T, 2G, 6G. The difference can be visualized by comparing the frequencies of these position-specific bases in m5C and non-m5C modifications (**Figure 3C**). Furthermore, through two sample logo analysis using R package "DiffLogo" (Nettling et al., 2015), we discovered the similar trend of some specific nucleotide usage preferences around m5C modifications (**Figure 3D**). These results indicate that base frequency differences exist between m5C and non-m5C modifications.
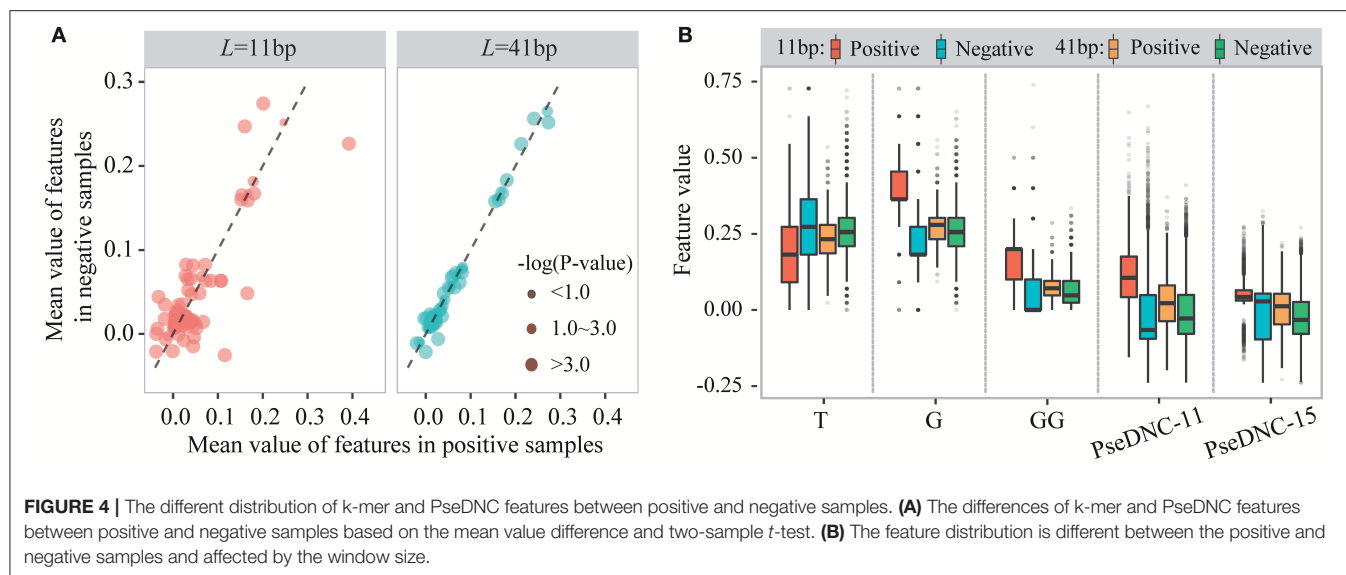
We then examined sequence-based features generated from k-mer and PseDNC encoding schemes. **Figure 4A** displays the mean values of these features for positive and negative samples. When the window size is 11-nt ($L_u = L_d = 5$), we detected 70 k-mer-based features and 19 PseDNC-based features significantly different between positive and negative samples (two-sample t-test; $p \leq 1.0E-4$). The top-five ranked features are the frequency of T, G, GG and PseDNC-11, PseDNC-15 (**Figure 4B**). When the window size was extended from 11-nt to 41-nt ($L_u = L_d = 20$), we also detected 32 k-mer-based features and 12 PseDNC-based features at the significance level of 1.0E-4. The top-five ranked features are

the frequency of G, GG and GGC, PseDNC-11 and PseDNC-15 (**Figure 4B**).

Taken together, these results indicate that the three encoding schemes, binary, k-mer and PseDNC, can generate discriminative features for m5C prediction. However, the importance of different features is affected by the window size used.

A Machine Learning-Based m5C Predictor With Optimized Window Size and Features

To obtain the optimized window size and feature subset, we iteratively performed ten-fold cross-validation experiments on the DatasetCV by varying window size L from 5-nt to 43-nt and the feature number F from 2 to 106+4*L (**Figure 5A**). For a given window size of L (e.g., upstream region: $L_u = 10$ and downstream region: $L_d = 5$) and feature number of F (e.g., $F = 50$), we performed a 10-fold cross-validation experiment to calculate an AUC value for evaluating the prediction performance of PEA-m5C. Then, at the given window size L, the best AUC value achieved by PEA-m5C can be found according to the curve depicted in **Figure 5B**, where x axis represents the number of selected features and y axis represents the AUC yielded by PEA-m5C. After examining all possible combinations of window sizes and feature numbers, we observed that PEA-m5C achieved the highest AUC value of 0.939 (**Figure 5A**), when the window size was set as 11-nt ($L_u = L_d = 5$) and 50 top ranked features were used (**Figure 5C**, **Table S2**).

FIGURE 4 | The different distribution of k-mer and PseDNC features between positive and negative samples. **(A)** The differences of k-mer and PseDNC features between positive and negative samples based on the mean value difference and two-sample *t*-test. **(B)** The feature distribution is different between the positive and negative samples and affected by the window size.

Prediction Evaluation and Comparison Using Hold-Out and Independent Testing Sets

After training PEA-m5C using the DatasetCV with the optimized window size and feature subset, we next evaluated the performance of PEA-m5C on a hold-out test set (DatasetHT). As shown in **Figure 6A**, the prediction score of positive samples (mean ± standard deviation [sd]: 0.775 ± 0.223) was significantly higher than that of negative samples (mean ± sd: 0.194± 0.225). This result indicates that PEA-m5C could provide a competitive performance in discriminating positive and negative samples. Indeed, PEA-m5C gave an area under ROC (AUC) and an area under the precision-recall curve (auPRC) of 0.939 and 0.945, respectively (**Figures 6B,C**). To assess the performance more comprehensively, six measures (Sn, Sp, Pr, Acc, MCC, and F$_1$) were examined at four thresholds, corresponding to the specificity level of 99% (very high confidence mode; VHmode), 95% (high confidence mode; HMode), 90% (normal confidence mode; NMode) and 85% (Low confidence mode; LMode) in the 10-fold cross-validation experiment, respectively (**Table 1**). In line with the intuitive observations of ROC curve (**Figure 6B**) and precision-recall curve (**Figure 6C**), PEA-m5C performed markedly better than random selection (AUC = 0.5, auPRC = 0.5, and MCC = 0) in predicting m^{5}C modifications at four different specificity levels (**Table 1**).
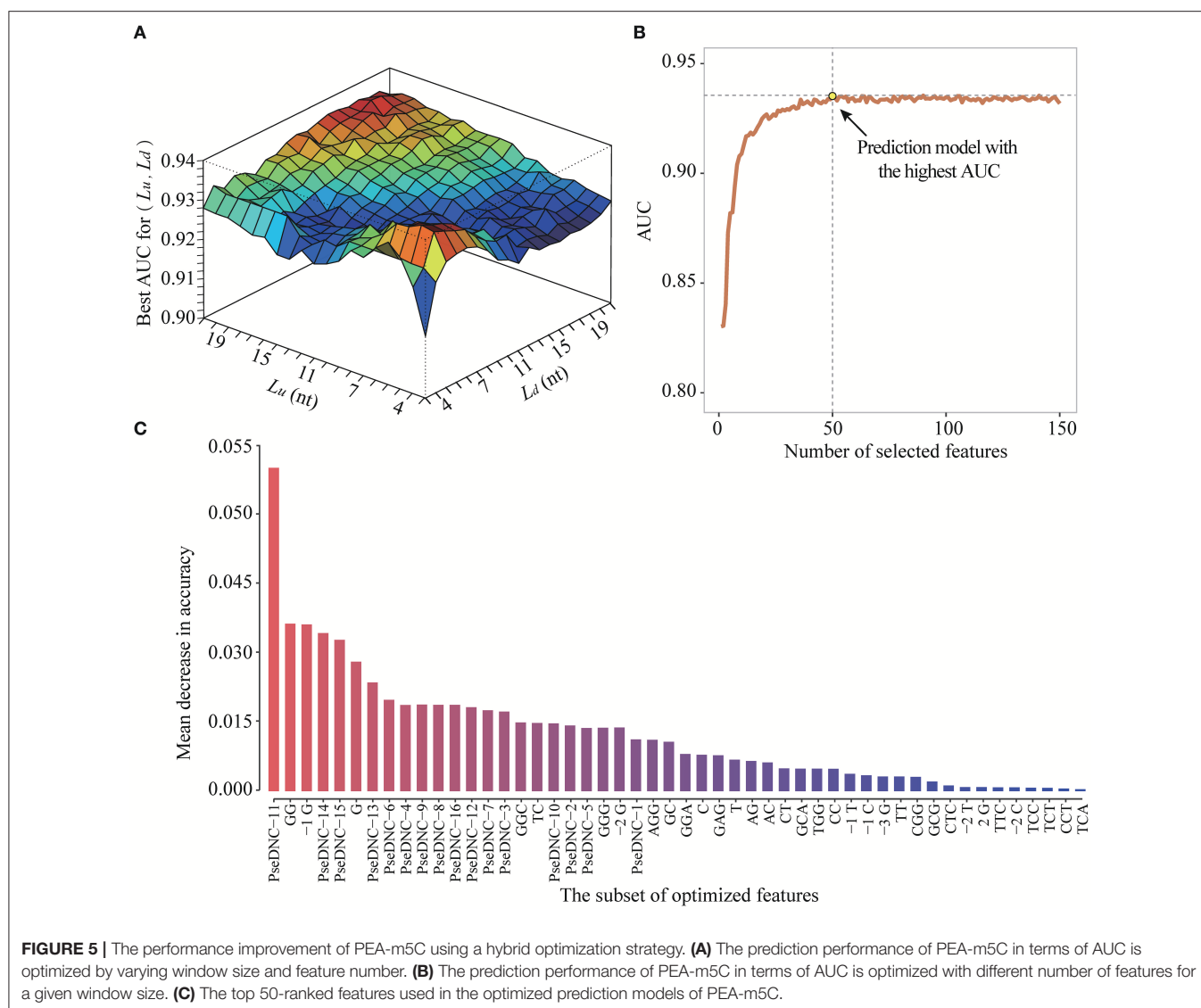
Currently, iRNAm5C-PseDNC is the only software available for m^{5}C prediction; however, it was built based on mammalian m^{5}C modifications. This provides us an opportunity to evaluate whether iRNAm5C-PseDNC could retain prediction accuracy on *Arabidopsis* m^{5}C modifications. We observed that iRNAm5C-PseDNC yielded a high specificity of 0.980, but an extremely low sensitivity of 0.010. The main reason is that there are significant differences between mammalian and *Arabidopsis* m^{5}C modifications (**Figure S2**). To examine the effectiveness of ML algorithms in iRNAm5C-PseDNC, we generated a new prediction model (named as iRNAm5C-PseDNC*) by re-training

iRNAm5C-PseDNC using positive and negative samples from the DatasetCV and evaluated its performance using the DatasetHT. Compared with iRNAm5C-PseDNC, iRNAm5C-PseDNC* yielded higher prediction accuracy at the level of Sn, Sp, Pr, Acc, MCC, and F$_1$. However, PEA-m5C still achieved higher prediction accuracy than iRNAm5C-PseDNC and iRNAm5C-PseDNC* (**Table 1**). The prediction performance of PEA-m5C was also better than iRNAm5C-PseDNC and iRNAm5C-PseDNC* on DatasetIT1 and Dataset2, which consist of samples from *Arabidopsis* silique and shoot tissues, respectively (**Table S3**).

Taken together, these results indicate that the construction of *Arabidopsis thaliana*-specific predictor is necessary and crucial. In addition, PEA-m5C is a useful tool for the prediction of m^{5}C sites in *Arabidopsis* transcripts.

Transcriptome-Wide Annotation and Analysis of Candidate m^{5}C Modifications

The encouraging performance of PEA-m5C in the cross-validation and validation testing experiments provide us an opportunity to accurately predict m^{5}C sites in the annotated *Arabidopsis* transcripts. At the threshold of 0.891 (VHMode), PEA-m5C predicted 303,421 candidate m^{5}C modifications (**Table 2**), covering 4.56% cytosines (303,421/6,650,570) in all annotated transcripts in Araport 11 database (https://www.araport.org/data/araport11). During the writing of our manuscript, Cui and colleagues identified 4,439 m^{5}C peaks in 3,534 expressed genes (**Table S4**) in young seedlings of *Arabidopsis* (Cui et al., 2017), by applying m^{5}C RNA immunoprecipitation followed by a deep-sequencing approach. We validated the m^{5}C predictions using these 4,439 m^{5}C peaks. Among the 3,534 expressed genes, PEA-m5C identified 5,463 candidate m^{5}C modifications, covering 2,724 of 4,439 reported peak regions. We note that the proportion of covered m^{5}C peaks increased from 61.4% (2,724/4,439) to 89.4% (3,968/4,439), when the HMode was used.
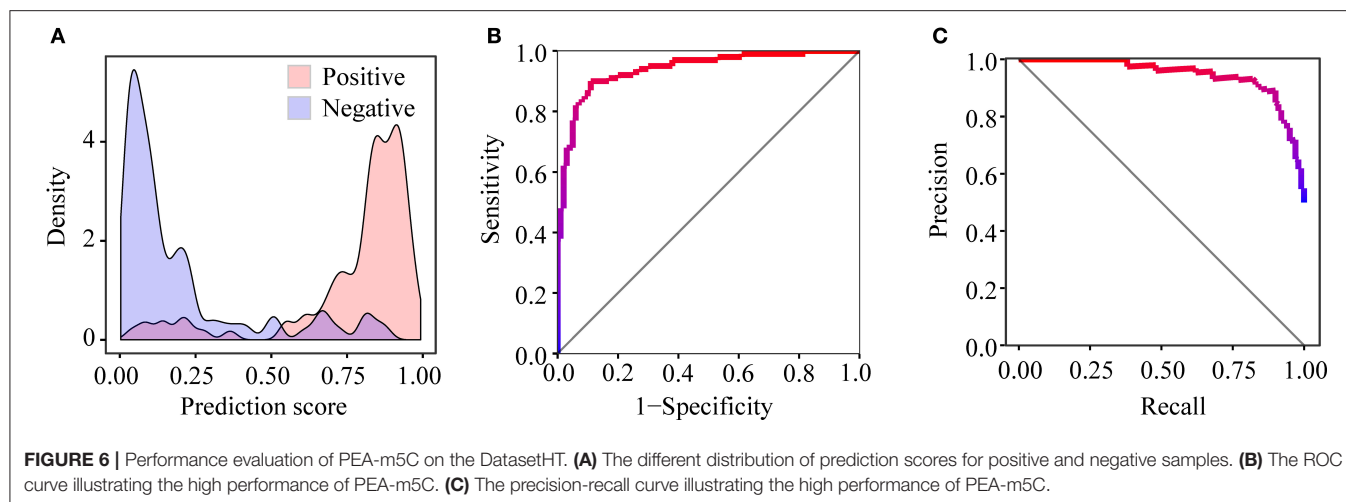
FIGURE 5 | The performance improvement of PEA-m5C using a hybrid optimization strategy. **(A)** The prediction performance of PEA-m5C in terms of AUC is optimized by varying window size and feature number. **(B)** The prediction performance of PEA-m5C in terms of AUC is optimized with different number of features for a given window size. **(C)** The top 50-ranked features used in the optimized prediction models of PEA-m5C.

As is known to us all, cytosines in DNA sequences can be methylated in three sequence context, namely CG, CHG, and CHH (H = A, C, or T) (Smith and Meissner, 2013). In this study, we explored the levels of cytosine methylation in RNA sequences. We observed that 24.7, 27.8, and 47.5% of the candidate m^{5}C modifications are methylated in the CG, CHG, and CHH sequence context, respectively. These proportions are markedly different from those of cytosines in background sequences (CG: 15.1%, CHG: 17.9%, CHH: 67.0%) (**Figure 7A**). Statistical analysis of base preference showed that there are very strong "G" signal around candidate m^{5}C modifications (**Figure 7B**). These results indicate that candidate m^{5}C modifications predicted by PEA-m5C may have potential biological functions.

Toward a better understanding of these candidate m^{5}C modifications, we further analyzed the enrichment of m^{5}C within three different regions of mRNAs: 5′-UTR, CDS and 3′-UTR. It can be seen from **Figure 7C** that the majority

of m^{5}C modifications are located in CDS regions. Recent studies have indicated that the m^{5}C modification prefers to occur at the downstream of translational start sites in mammal mRNAs (Amort et al., 2017; Yang et al., 2017). We calculated the distance between candidate m^{5}C modifications and translational start sites, and found that the most frequently m^{5}C modification position is the 4nt downstream of the translational start site (AUG*<u>**C**</u>; methylated cytosines are in bold and underlined) (**Figure 7D**). In order to further investigate the potential function of those 1,063 genes with m^{5}C modifications located at 4-nt downstream of the translational start site, we performed a GO (gene ontology) enrichment analysis using agriGO 2.0 (Tian et al., 2017) and found that in the BP (Biological Progress) sub-category, 166 genes (**Table 3**) are enriched in the term "response to stimulus" with FDR of 2.40E-4; For the MF (Molecular Function) sub-category, 350 genes are significantly enriched in "catalytic activity" with FDR of 9.80E-07 (**Table 3**). We also performed

FIGURE 6 | Performance evaluation of PEA-m5C on the DatasetHT. **(A)** The different distribution of prediction scores for positive and negative samples. **(B)** The ROC curve illustrating the high performance of PEA-m5C. **(C)** The precision-recall curve illustrating the high performance of PEA-m5C.

TABLE 1 | Prediction performance of m^{5}C predictors on DatasetHT.

m^{5}C predictor	Mode	Threshold	Sn	Sp	Pr	Acc	MCC	F$_1$
PEA-m5C	VHmode	0.891	0.330	1.000	1.000	0.665	0.445	0.496
	HMode	0.765	0.720	0.950	0.935	0.835	0.688	0.814
	NMode	0.622	0.860	0.900	0.896	0.880	0.761	0.878
	LMode	0.484	0.900	0.860	0.865	0.880	0.761	0.882
iRNAm5C-PseDNC (web server)	–	–	0.010	0.980	0.333	0.495	−0.041	0.019
iRNAm5C-PseDNC (Method)*	–	0.500	0.610	0.720	0.685	0.665	0.332	0.646

*An m^{5}C prediction model generated by re-training iRNAm5C-PseDNC using positive and negative samples from the DatasetCV.

TABLE 2 | Candidate m^{5}C modifications in different types of RNAs. Num: number, Prop: proportion, trans: transcripts.

RNA	Num of trans	Num of cytosines	Num (Prop) of cytosines are methylated	Num (Prop) of trans containing m^{5}C
Long noncoding RNA	2,455	161,608	1,480 (0.92%)	967(39.39%)
miRNA	387	1,082	15 (1.29%)	12(3.10%)
Primary miRNA transcript	325	9,717	100 (1.03%)	74(22.77%)
mRNA	48,353	16,727,847	225,348 (1.35%)	44,350(91.72%)
rRNA	15	4,041	147 (3.64%)	12(80.00%)
snoRNA	287	5,049	70 (1.39%)	55(19.16%)
snRNA	82	2,906	62 (2.13%)	35(42.68%)
tRNA	689	10,227	272 (2.66%)	232(33.67%)

pathway enrichment analysis on these 1063 genes using the hypergeometric distribution test. Pathway information was obtained from KEGG (http://www.genome.jp/kegg) and AraCyc (http://www.plantcyc.org) databases. At the level of $p \leq 1.0E-2$, we identified four significantly enriched pathways, including L-lysine biosysthesis VI pathway, glutathione metabolism, N-Glycan biosynthesis, and phosphatidylinositol signaling system (**Table S5**).
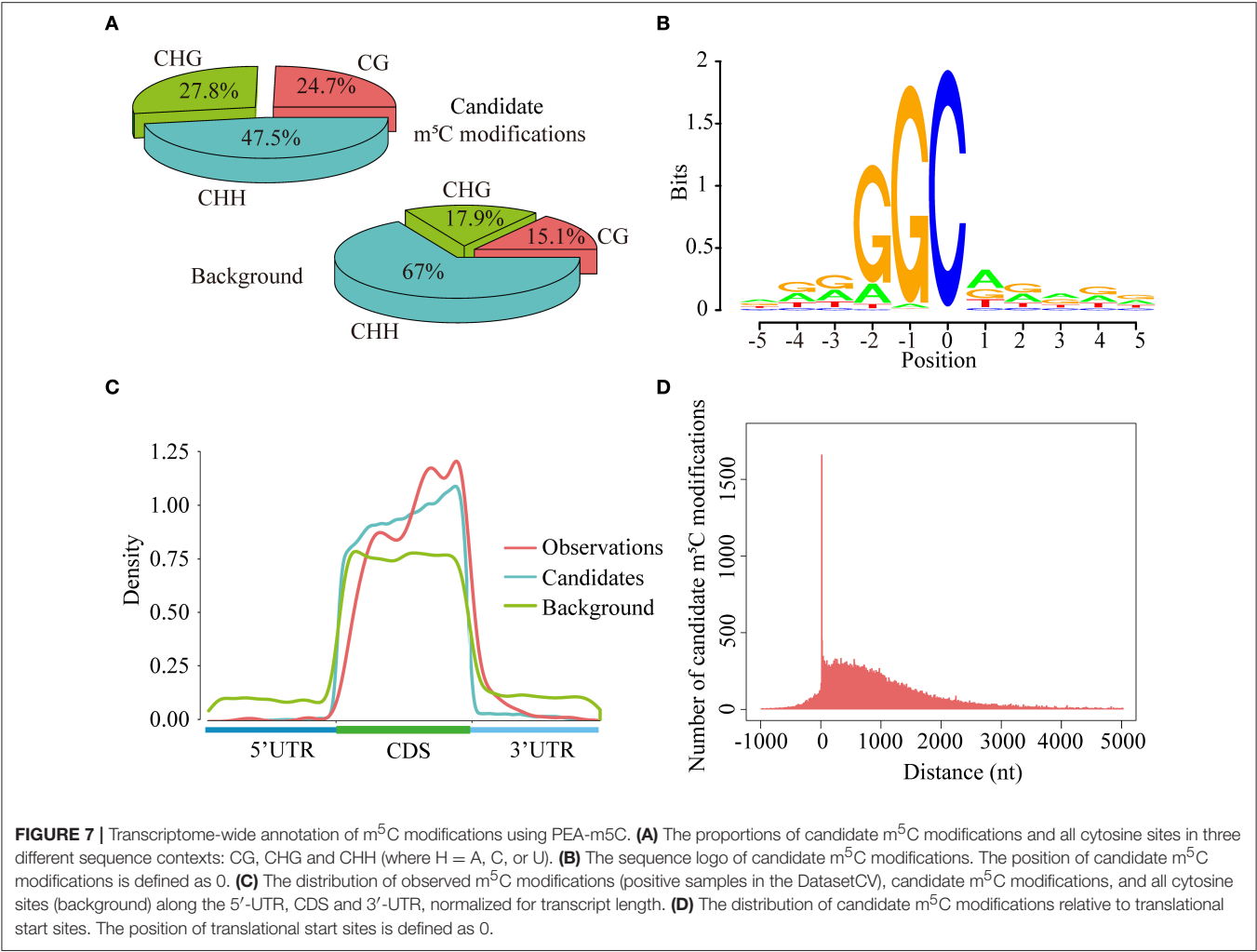
Implementation of PEA-m5C

To facilitate the practicability, we implemented PEA-m5C into an R package named "PEA-m5C". We also provided a cross-platform, user-friendly and interactive interface for PEA-m5C with JAVA programming language (**Figure 8**). This allows the user to easily implement PEA-m5C without the requirement of any programming skills or knowledge. To expand the application of PEA-m5C to other species, users can also retrain prediction models through the pre-specified dataset using the "Self-Defined Mode" option in PEA-m5C, with the input of positive and negative samples in FASTA format. PEA-m5C is freely available to academic users at: https://github.com/cma2015/PEA-m5C.

DISCUSSION

In this study, we developed PEA-m5C, a computationally framework for accurate identification of m^{5}C modifications in *Arabidopsis*. PEA-m5C predictor was constructed using RF algorithm with optimized window size and sequence-based features, achieving a considerable promising performance no matter from 10-fold cross-validation experiment or hold-out test experiment. The PEA-m5C is superior to the newly developed and only available m^{5}C predictor iRNAm5C-PseDNC in several aspects.

First, besides the PseDNC encoding scheme used in iRNAm5C-PseDNC, PEA-m5C additionally integrates

**FIGURE 7** | Transcriptome-wide annotation of m⁵C modifications using PEA-m5C. **(A)** The proportions of candidate m⁵C modifications and all cytosine sites in three different sequence contexts: CG, CHG and CHH (where H = A, C, or U). **(B)** The sequence logo of candidate m⁵C modifications. The position of candidate m⁵C modifications is defined as 0. **(C)** The distribution of observed m⁵C modifications (positive samples in the DatasetCV), candidate m⁵C modifications, and all cytosine sites (background) along the 5′-UTR, CDS and 3′-UTR, normalized for transcript length. **(D)** The distribution of candidate m⁵C modifications relative to translational start sites. The position of translational start sites is defined as 0.

TABLE 3 | Top five significant GO terms in the sub-category of biological progress (BP), molecular function (MF), and cellular component (CC).

GO Term	Enriched gene number	FDR	Category
Lipid localization	10	2.60E-05	BP
Response to stimulus	166	0.00024	BP
Macromolecule localization	34	0.00032	BP
Localization	91	0.00032	BP
Toxin metabolic process	11	0.00032	BP
Transporter activity	91	3.70E-09	MF
Transmembrane transporter activity	69	9.80E-07	MF
Catalytic activity	350	9.80E-07	MF
Substrate-specific transporter activity	63	6.90E-06	MF
Substrate-specific transmembrane transporter activity	56	1.10E-05	MF
Cytoplasm	299	2.60E-14	CC
Cell part	548	1.00E-13	CC
Cell	548	1.00E-13	CC
Cytoplasmic part	276	1.40E-13	CC
Membrane	200	1.40E-13	CC

another two encoding schemes (binary and k-mer) to make more use of sequence-based features. Both 10-fold cross-validation and independent testing experiments have demonstrated that higher prediction accuracy can be achieved by PEA-m5C when more feature encoding schemes were used (**Figure S3**; **Table S6**). For instance, in the 10-fold cross-validation, PEA-m5C yielded an AUC of 0.904, 0.914 and 0.939 when PseDNC, PseDNC + k-mer, PseDNC + k-mer + binary encoding schemes were used, respectively.

Second, PEA-m5C uses a hybrid optimization strategy to produce better prediction accuracy (**Table S6**), while iRNAm5C-PseDNC didn't perform the model optimization process. This is understandable as the model optimization is a rather timing-consuming process (**Figure 2**). However, the results shown in **Figure 5** illustrated the importance of model optimization in developing accurate m⁵C predictors. We also would like to note that the process of model optimization requires to be finely tuned, such as the choice of appropriate feature selection approaches. To select informative features for m⁵C prediction, we preferred to use the information
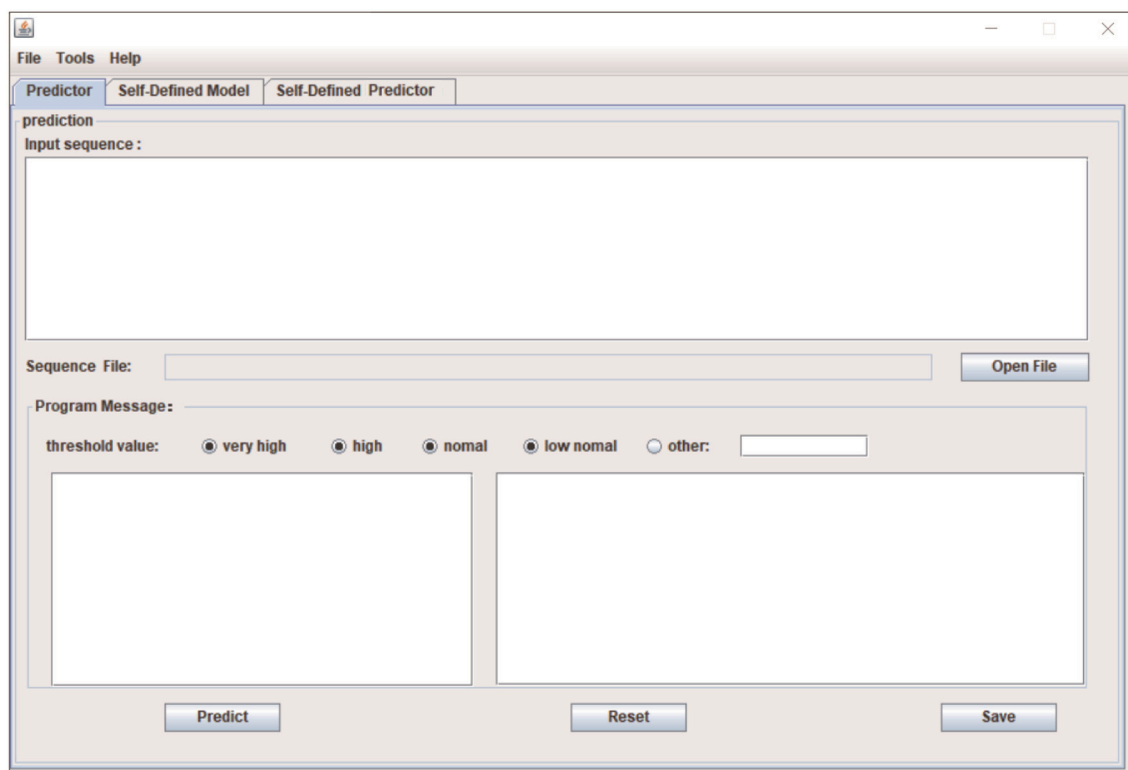
FIGURE 8 | The Java interface implementation of PEA-m5C.

gain approach rather than statistical analysis approaches (e.g., chi-square test for binary features, student's *t*-test for k-mer- and PseDNC-based features). While testing on the DatasetHT, PEA-m5C using the information gain approach yielded a slightly higher maximum MCC (0.790) than that using the chi-square test and the student's *t*-test (0.770).

Finally, PEA-m5C has been implemented into a user-friendly interface with JAVA programming language and an R package to maximize its practicality. It also includes a self-training module that provides an option to automatically build m^{5}C predictors for specific species, tissues, or conditions. This is very important as m^{5}C modifications exhibit different sequence patterns in different issues (**Figure S4**).

In the future, we will endeavor to incorporate more features (e.g., structure-based features) to further improve the performance of PEA-m5C. If possible, specie-specific or tissue-specific predictors will be developed to facilitate the functional investigation of m^{5}C modifications in plants.

AUTHOR CONTRIBUTIONS

CM: Designed the experiments; JS, JZ, and EB: Performed the experiments; JS, JZ, EB, CM, JY, and YS: Analyzed the data; CM, JZ, and JS: Wrote the paper. All authors read and approved the final manuscript.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2018.00519/full#supplementary-material

Supplementary Data 1 | The description of PseDNC encoding.

Figure S1 | Location distribution of positive, negative and background samples along the 5′-UTR, CDS and 3′-UTR, normalized for transcript length (relative distance).

Figure S2 | Two sample logos of *Arabidopsis* and mammalian m^{5}C modifications. It shows nucleotides which are enriched or depleted in the surrounding region of m^{5}C modifications.

Figure S3 | The ROC curve of 10-fold cross-validation illustrating the performance of PEA-m5C with different feature encoding schemes.

Figure S4 | Different sequence patterns of m5C modifications in DatasetHT, DatasetIT1, and DatasetIT2. **(A)** Frequencies of 41 * 4 position-specific bases in

DatasetHT (Root tissue) and DatasetIT1 (Silique tissue). **(B)** Two sample logos DatasetHT (Root tissue) and DatasetIT1 (Silique tissue) m^{5}C modifications. **(C)** Frequencies of 41 * 4 position-specific bases in DatasetHT (Root tissue) and DatasetIT2 (Shoot tissue). **(D)** Two sample logos DatasetHT (Root tissue) and DatasetIT2 (Shoot tissue) m^{5}C modifications.

Table S1 | Four benchmark datasets constructed for the prediction of m^{5}C modifications in this study.

Table S2 | The feature importance measured using the information gain approach at the window size of 11-nt ($L_u = L_d = 5$).

Table S3 | Prediction performance of m^{5}C predictors on DatasetIT1 and DatasetIT2.

Table S4 | Peak regions used for validating transcriptome-wide candidate m^{5}C modifications predicted by PEA-m^{5}C.

Table S5 | Enriched pathways of genes containing m^{5}C modifications at 4-nt downstream of the translational start site.

Table S6 | The performance of m^{5}C predictors on DatasetHT using different encoding schemes.

REFERENCES

Amort, T., Rieder, D., Wille, A., Khokhlova-Cubberley, D., Riml, C., Trixl, L., et al. (2017). Distinct 5-methylcytosine profiles in poly(A) RNA from mouse embryonic stem cells and brain. *Genome Biol.* 18:1. doi: 10.1186/s13059-016-1139-1

Amort, T., Souliere, M. F., Wille, A., Jia, X. Y., Fiegl, H., Wörle, H., et al. (2013). Long non-coding RNAs as targets for cytosine methylation. *RNA Biol.* 10, 1003–1008. doi: 10.4161/rna.24454

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Carlile, T. M., Rojas-Duran, M. F., Zinshteyn, B., Shin, H., Bartoli, K. M., and Gilbert, W. V. (2014). Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature* 515, 143–146. doi: 10.1038/nature13802

Chen, W., Feng, P., Ding, H., Lin, H., and Chou, K. C. (2015). iRNA-Methyl: identifying N(6)-methyladenosine sites using pseudo nucleotide composition. *Anal. Biochem.* 490, 26–33. doi: 10.1016/j.ab.2015.08.021

Chen, W., Xing, P., and Zou, Q. (2017). Detecting N(6)-methyladenosine sites from RNA transcriptomes using ensemble Support Vector Machines. *Sci. Rep.* 7:40242. doi: 10.1038/srep40242

Cheng, T., Wang, Y., and Bryant, S. H. (2012). FSelector: a Ruby gem for feature selection. *Bioinformatics* 28, 2851–2852. doi: 10.1093/bioinformatics/bts528

Choi, J., Ieong, K. W., Demirci, H., Chen, J., Petrov, A., Prabhakar, A., et al. (2016). N(6)-methyladenosine in mRNA disrupts tRNA selection and translation-elongation dynamics. *Nat. Struct. Mol. Biol.* 23, 110–115. doi: 10.1038/nsmb.3148

Cui, H., Zhai, J., and Ma, C. (2015). miRLocator: machine learning-based prediction of mature microRNAs within plant pre-miRNA sequences. *PLoS ONE* 10:e0142753. doi: 10.1371/journal.pone.0142753

Cui, X., Liang, Z., Shen, L., Zhang, Q., Bao, S., Geng, Y., et al. (2017). 5-Methylcytosine RNA methylation in *Arabidopsis thaliana*. *Mol. Plant* 10, 1387–1399. doi: 10.1016/j.molp.2017.09.013

David, R., Burgess, A., Parker, B., Li, J., Pulsford, K., Sibbritt, T., et al. (2017). Transcriptome-wide mapping of RNA 5-Methylcytosine in *Arabidopsis* mRNAs and noncoding RNAs. *Plant Cell* 29, 445–460. doi: 10.1105/tpc.16.00751

Dominissini, D., Nachtergaele, S., Moshitch-Moshkovitz, S., Peer, E., Kol, N., Ben-Haim, M. S., et al. (2016). The dynamic N(1)-methyladenosine methylome in eukaryotic messenger RNA. *Nature* 530, 441–446. doi: 10.1038/nature16998

Edelheit, S., Schwartz, S., Mumbach, M. R., Wurtzel, O., and Sorek, R. (2013). Transcriptome-wide mapping of 5-methylcytidine RNA modifications in bacteria, archaea, and yeast reveals m^{5}C within archaeal mRNAs. *PLoS Genet.* 9:e1003602. doi: 10.1371/journal.pgen.1003602

Helm, M., and Motorin, Y. (2017). Detecting RNA modifications in the epitranscriptome: predict and validate. *Nat. Rev. Genet.* 18, 275–291. doi: 10.1038/nrg.2016.169

Hornik, K., Buchta, C., and Zeileis, A. (2009). Open-source machine learning: r meets Weka. *Comput. Stat.* 24, 225–232. doi: 10.1007/s00180-008-0119-7

Hussain, S., Aleksic, J., Blanco, S., Dietmann, S., and Frye, M. (2013). Characterizing 5-methylcytosine in the mammalian epitranscriptome. *Genome Biol.* 14:215. doi: 10.1186/gb4143

Kreck, B., Marnellos, G., Richter, J., Krueger, F., Siebert, R., and Franke, A. (2012). B-SOLANA: an approach for the analysis of two-base encoding bisulfite sequencing data. *Bioinformatics* 28, 428–429. doi: 10.1093/bioinformatics/btr660

Leclercq, M., Diallo, A. B., and Blanchette, M. (2013). Computational prediction of the localization of microRNAs within their pre-miRNA. *Nucleic Acids Res.* 41, 7200–7211. doi: 10.1093/nar/gkt466

Li, X., Xiong, X., and Yi, C. (2016). Epitranscriptome sequencing technologies: decoding RNA modifications. *Nat. Methods* 14, 23–31. doi: 10.1038/nmeth.4110

Libbrecht, M. W., and Noble, W. S. (2015). Machine learning applications in genetics and genomics. *Nat. Rev. Genet.* 16, 321–332. doi: 10.1038/nrg3920

Ma, C., Xin, M., Feldmann, K. A., and Wang, X. (2014a). Machine learning-based differential network analysis: a study of stress-responsive transcriptomes in Arabidopsis. *Plant Cell* 26, 520–537. doi: 10.1105/tpc.113.121913

Ma, C., Zhang, H. H., and Wang, X. (2014b). Machine learning for Big Data analytics in plants. *Trends Plant Sci.* 19, 798–808. doi: 10.1016/j.tplants.2014.08.004

Ma, W., Qiu, Z., Song, J., Cheng, Q., and Ma, C. (2017). DeepGS: predicting phenotypes from genotypes using deep learning. *bioRxiv*. doi: 10.1101/241414. [Epub ahead of print].

Machnicka, M. A., Milanowska, K., Osman Oglou, O., Purta, E., Kurkowska, M., Olchowik, A., et al. (2013). MODOMICS: a database of RNA modification pathways−2013 update. *Nucleic Acids Res.* 41, D262–D267. doi: 10.1093/nar/gks1007

Meyer, K. D., and Jaffrey, S. R. (2014). The dynamic epitranscriptome: N6-methyladenosine and gene expression control. *Nat. Rev. Mol. Cell Biol.* 15, 313–326. doi: 10.1038/nrm3785

Meyer, K. D., Patil, D. P., Zhou, J., Zinoviev, A., Skabkin, M. A., Elemento, O., et al. (2015). 5′ UTR m(6)A promotes cap-independent translation. *Cell* 163, 999–1010. doi: 10.1016/j.cell.2015.10.012

Nettling, M., Treutler, H., Grau, J., Keilwagen, J., Posch, S., and Grosse, I. (2015). DiffLogo: a comparative visualization of sequence motifs. *BMC Bioinformatics* 16:387. doi: 10.1186/s12859-015-0767-x

Pan, T. (2013). N6-methyl-adenosine modification in messenger and long non-coding RNA. *Trends Biochem. Sci.* 38, 204–209. doi: 10.1016/j.tibs.2012.12.006

Qiu, W. R., Jiang, S. Y., Xu, Z. C., Xiao, X., and Chou, K. C. (2017). iRNAm5C-PseDNC: identifying RNA 5-methylcytosine sites by incorporating physical-chemical properties into pseudo dinucleotide composition. *Oncotarget* 8, 41178–41188. doi: 10.18632/oncotarget.17104

Smith, Z. D., and Meissner, A. (2013). DNA methylation: roles in mammalian development. *Nat. Rev. Genet.* 14, 204–220. doi: 10.1038/nrg3354

Storey, J. D. (2002). A direct approach to false discovery rates. *J. R. Stat. Soc. B Stat. Methodol.* 64, 479–498. doi: 10.1111/1467-9868.00346

Sun, W. J., Li, J. H., Liu, S., Wu, J., Zhou, H., Qu, L. H., et al. (2016). RMBase: a resource for decoding the landscape of RNA modifications from high-throughput sequencing data. *Nucleic Acids Res.* 44, D259–D265. doi: 10.1093/nar/gkv1036

Tian, T., Liu, Y., Yan, H., You, Q., Yi, X., Du, Z., et al. (2017). agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Res.* 45, W122–W129. doi: 10.1093/nar/gkx382

Wang, X., and He, C. (2014). Dynamic RNA modifications in posttranscriptional regulation. *Mol. Cell* 56, 5–12. doi: 10.1016/j.molcel.2014.09.001

Yang, X., Yang, Y., Sun, B. F., Chen, Y. S., Xu, J. W., Lai, W. Y., et al. (2017). 5-methylcytosine promotes mRNA export - NSUN2 as the methyltransferase and ALYREF as an m(5)C reader. *Cell Res.* 27, 606–625. doi: 10.1038/cr.2017.55

Zhai, J., Song, J., Cheng, Q., Tang, Y., and Ma, C. (2017). PEA: an integrated R toolkit for plant epitranscriptome analysis. *bioRxiv*. doi: 10.1101/240887. [Epub ahead of print].

Zhai, J., Tang, Y., Yuan, H., Wang, L., Shang, H., and Ma, C. (2016). A meta-analysis based method for prioritizing candidate genes involved in

a pre-specific function. *Front. Plant Sci.* 7:1914. doi: 10.3389/fpls.2016.
01914

Zhao, B. S., Roundtree, I. A., and He, C. (2017). Post-transcriptional gene
regulation by mRNA modifications. *Nat. Rev. Mol. Cell Biol.* 18, 31–42.
doi: 10.1038/nrm.2016.132

Zhou, J., Wan, J., Gao, X., Zhang, X., Jaffrey, S. R., and Qian, S. B. (2015). Dynamic
m(6)A mRNA methylation directs translational control of heat shock response.
Nature 526, 591–594. doi: 10.1038/nature15377

Zhou, Y., Zeng, P., Li, Y. H., Zhang, Z., and Cui, Q. (2016). SRAMP: prediction
of mammalian N6-methyladenosine (m6A) sites based on sequence-derived
features. *Nucleic Acids Res.* 44:e91. doi: 10.1093/nar/gkw104

Corrigendum: Transcriptome-Wide Annotation of m^{5}C RNA Modifications Using Machine Learning

Jie Song [1,2†], Jingjing Zhai [1†], Enze Bian [3†], Yujia Song [3], Jiantao Yu [3] and Chuang Ma [1,2]*

[1] State Key Laboratory of Crop Stress Biology for Arid Areas, Center of Bioinformatics, College of Life Sciences, Northwest A&F University, Shaanxi, China, [2] Key Laboratory of Biology and Genetics Improvement of Maize in Arid Area of Northwest Region, Ministry of Agriculture, Northwest A&F University, Shaanxi, China, [3] College of Information Engineering, Northwest A&F University, Shaanxi, China

A corrigendum on

Transcriptome-Wide Annotation of m^{5}C RNA Modifications Using Machine Learning
by Song, J., Zhai, J., Bian, E., Song, Y., Yu, J., and Ma, C. (2018). Front. Plant Sci. 9:519. doi: 10.3389/fpls.2018.00519

In the original article, there was an error, the word "reversible" is misleading. A correction has been made to the Abstract and the Introduction, paragraph 2.

Though high-throughput experimental technologies have been developed and applied to profile m^{5}C modifications under certain conditions, transcriptome-wide studies of m^{5}C modifications are still hindered by the dynamic nature of m^{5}C and the lack of computational prediction methods.

Second, because of the dynamic nature of m^{5}C (Wang and He, 2014), existing high-throughput sequencing technologies can only capture a snapshot of RNA modifications under certain experimental conditions, and cover just a small fraction of the whole transcriptome of a given sample (Zhou et al., 2016), resulting in the generation of significant numbers of false negatives (non-detected true m^{5}C modifications).

The authors apologize for the mistake. This error does not change the scientific conclusions of the article in any way. The original article has been updated.

REFERENCES

Wang, X., and He, C. (2014). Dynamic RNA modifications in posttranscriptional regulation. *Mol. Cell* 56, 5–12. doi: 10.1016/j.molcel.2014.09.001

Zhou, Y., Zeng, P., Li, Y. H., Zhang, Z., and Cui, Q. (2016). SRAMP: prediction of mammalian N6-methyladenosine (m6A) sites based on sequence-derived features. *Nucleic Acids Res.* 44:e91. doi: 10.1093/nar/gkw104

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Cadmium Stress Leads to Rapid Increase in RNA Oxidative Modifications in Soybean Seedlings

Jagna Chmielowska-Bąk[1]*, Karolina Izbiańska[1], Anna Ekner-Grzyb[1], Melike Bayar[2] and Joanna Deckert[1]

[1] Department of Plant Ecophysiology, Faculty of Biology, Institute of Experimental Biology, Adam Mickiewicz University in Poznań, Poznań, Poland, [2] Department of Molecular Biology and Genetics, Faculty of Science, Istanbul University, Istanbul, Turkey

Increase in the level of reactive oxygen species (ROS) is a common response to stress factors, including exposure to metals. ROS over-production is associated with oxidation of lipids, proteins, and nucleic acids. It is suggested that the products of oxidation are not solely the markers of oxidative stress but also signaling elements. For instance, it has been shown in animal models that mRNA oxidation is a selective process engaged in post-transcriptional regulation of genes expression and that it is associated with the development of symptoms of several neurodegenerative disorders. In the present study, we examined the impact of short-term cadmium (Cd) stress on the level of two RNA oxidation markers: 8-hydroxyguanosine (8-OHG) and apurinic/apyrimidinic sites (AP-sites, abasic sites). In the case of 8-OHG, a significant increase was observed after 3 h of exposure to moderate Cd concentration (10 mg/l). In turn, high level of AP-sites, accompanied by strong ROS accumulation and lipid peroxidation, was noted only after 24 h of treatment with higher Cd concentration (25 mg/l). This is the first report showing induction of RNA oxidations in plants response to stress factors. The possible signaling and gene regulatory role of oxidatively modified transcripts is discussed.

Keywords: cadmium, soybean, RNA oxidation, 8-hydroxyguanosine, AP-sites, abasic sites, epitranscriptomics, oxidative stress

INTRODUCTION

Elevated levels of cadmium (Cd) were found in the soil in several regions of the world (Khan et al., 2011; Wang et al., 2015; Dumitrel et al., 2017). Contamination of the environment with this metal possesses serious threat to plants growth as Cd is highly mobile and toxic. Importantly, it can be absorbed by crop plants and in this mode enter human organisms leading to serious disorders. Cd is considered as class I carcinogen. The main targets of its toxicity are kidneys, lungs, and skeleton system (Kah et al., 2012).

In recent years, significant effort has been made to elucidate Cd impact on plants. The research has been focused on the toxicity mechanisms, Cd sensing, and the development of metal tolerance. The most universal responses to this toxic element, not only limited to plants but also found in bacteria and animals, are over-production of ROS (reviewed in

---

Chmielowska-Bąk and Deckert, 2012). ROS are "double-faced" molecules, which on one hand can lead to significant damage of cellular compounds but on the other hand are crucial components of signaling network indispensable for the activation of the defense (Cuypers et al., 2016; Sewelam et al., 2016). Recently it has been proposed that ROS signal is transmitted by the products of the oxidation of biological molecules including lipids, proteins, and nucleic acids (reviewed in Chmielowska-Bąk et al., 2015).

Indeed, various studies carried out on human, animal, and plant system demonstrated correlation between oxidation of certain species of mRNA and decrease in the level of encoded proteins, indicating that the process constitutes a post-transcriptional gene regulatory mechanism (Shan et al., 2003, 2007; Tanaka et al., 2007; Chang et al., 2008; Bazin et al., 2011; Gao et al., 2013). The decrease in the amount of proteins encoded by oxidized transcripts is most probably dependent on recently described ribosome stalling. *In vitro* studies carried out on reconstituted bacterial system demonstrated that occurrence of 8-OHG, the most common oxidative modification of RNA, causes slowing down of translation process by 2–4 magnitudes. The effect has been observed regardless of the position of oxidized bases in the codon, even in the wobble position. In eukaryotic extracts, translation was nearly completely inhibited by the presence of 8-OHG. It was suggested that occurrence of 8-OHG in transcripts leads to alterations in RNA–RNA interactions and prevents adaptation of active conformation in the decoding center. At the same time, it has been shown that oxidized transcripts are subjected to ribosome-based quality control and are predestined for degradation through No-Go decay pathway (NGD) (Simms et al., 2014). In concordance, yeast mutants with alerted decapping system leading to less efficient mRNA degradation showed elevated level of 8-OHG in the transcripts accompanied by premature death of the cells. The yeast mutants were also characterized by higher frequency of reversion from Trp⁻ (tryptophan minus) phenotype (Stirpe et al., 2016).

Another major discovery is the fact that mRNA oxidation is a selective process. Study carried out on postmortem isolated brain tissues of patients suffering from Alzheimer's disease demonstrated that oxidation did not occur in the most abundant transcripts such as β-*actin*. On the other hand, the most prominent oxidation was always noted in specific mRNAs encoding proteins engaged in signal transduction, cellular transport, gene expression regulation, and response to alerted ROS metabolism (Shan et al., 2003). These findings have been confirmed by other reports carried out on human, animal, and plant models (Shan et al., 2007; Tanaka et al., 2007; Chang et al., 2008; Bazin et al., 2011; Gao et al., 2013). Despite the importance of transcript oxidation in genes regulation and cell functioning so far, only two studies were dedicated to elucidation of this phenomenon in plants. It was demonstrated on sunflower seeds that alleviation of seed dormancy during dry after-ripening was associated with increase in the 8-OHG in mRNA. The observed oxidation was limited to 24 definite transcripts encoding proteins associated with metabolism, response to stress factors, and transport (Bazin et al., 2011). Similarly, studies on wheat showed

selective oxidation of certain transcripts associated with changes in protein levels and release from seed dormancy (Gao et al., 2013).

Beside the 8-OHG, oxidation might lead to the formation of numerous other modified bases in nucleic acids (Barciszewski et al., 1999). Studies using *in vitro* translation system showed that the most common oxidative modifications of RNA, namely 8-OHG, 5-hydroxyuridine (5-OHU), 5-hydroxycytidine (5-OHC), 8-oxo-7,8-dihydroadenosine (8-OHA), 1,N6-ethenoadenosine (ε-A), 3,N4-ethenocytidine (ε-C), and abasic sites (AP), result in slowing down or complete inhibition of the translation process (Calabretta et al., 2015). In turn, oxidation of DNA is associated with decrease in its stability and enhanced mutation rate. High level of oxidatively modified bases has been noted in various types of cancer (Sedelnikova et al., 2010). The repair of DNA lesions induced by oxidation is carried out by the Base Excision Repair (BER) pathway. The initial step of BER is recognition and excision of the modified bases by DNA glycosylases leading to the formation of AP-sites (abasic sites), which are considered markers of oxidative stress (Čolak, 2008; Antoniali et al., 2017). Recently, a method of abasic sites detection has been successfully applied in the evaluation of RNA oxidation (Tanaka et al., 2011). However, so far the changes in the level AP-sites in transcripts of plants exposed to stress factors have not been described.

The aim of present study was examination of the influence of Cd in two concentrations (10 and 25 mgl⁻¹) on the intensity of oxidative stress and frequency of RNA oxidation-dependent modifications – 8-OHG and AP-sites. Our previous research showed that in the earliest period of Cd stress (3 h), ROS are engaged in the regulation of gene expression, while strong accumulation of O_2^- and H_2O_2 has been marked after 24 h (Chmielowska-Bąk et al., 2017). Therefore, these two time points, 3 and 24 h, were applied in the present study.

MATERIALS AND METHODS

Growth Conditions and Treatment Procedures

Soybean (*Glycine max* L cv. Naviko) seeds were kindly supplied by the Department of Genetics and Plant Breeding, University of Life Sciences in Poznań, Poland. The seeds were surface-sterilized for 5 min with 75% ethanol and for 10 min with 1% sodium hyperchlorite. Thereafter, the seed were washed under running water for 30 min and imbedded in distilled water for 2 h. The seeds were placed on Petri dishes (30 cm of diameter) lined with two layers of moistened lignin covered by one layer of blotting paper and transferred to growth chamber with stable temperature of 22°C for 48 h. Germinated seedlings, selected in respect of similar roots length, were transferred to new Petri dishes (10 cm of diameter), wherein the roots were placed between two layers of blotting paper in cutout wholes. Afterward, the seedlings were treated with 5 mL of distilled water (control) or CdCl₂ with Cd at the concentration 10 and 25 mg l⁻¹ (corresponding to 89 and 223 µM, respectively).

Estimation of the Amount of Dead Cells

Cell viability was estimated on the basis of Blue Evans uptake according to Lehotai et al. (2011). Approximately 200 mg of roots were cut off on ice, weighted (200 mg), and incubated for 20 min in 0.25% Evans blue (Sigma, E-2129). Then the roots were washed twice for 15 min in distilled water and homogenized in mortar with – destaining solution (50 ml of ethanol, 49 ml of distilled water, and 1 ml of 10% SDS). Samples were incubated in heating block for 15 min at 50°C and centrifuged (12,000 rpm, 20°C, 15 min). The Blue Evans uptake, indicating cells death, was measured spectrophotometrically at $\lambda = 600$ nm.

Total RNA and mRNA Isolation

For RNA isolation, approximately 100 mg of soybean roots were cut off on ice, immediately frozen in liquid nitrogen, and stored in −80°C. The RNA was isolated from frozen tissue with the use of TriReagent (BioShop Canada Inc., Canada, TRI118) according to the manufacturer's instructions. The mRNA has been purified from the total RNA using GenElute^{TM} mRNA Miniprep Kit (Sigma–Aldrich, MRN10-1KT). The amount and purity of the obtained RNA and mRNA has been measured on NanoCell (Thermo Scientific) at spectrophotometer Biomate^{TM} 3S (Thermo Scientific).

Measurement of 8-OHG Level

The level of 8-OHG has been quantified with OxiSelect^{TM} Oxidative RNA Damage ELISA-8OHG Quantification Kit (BioCells, STA-325). For the analysis, 10 μg of sample (total RNA or mRNA) was digested with 20U of Nuclease S1 (BioShop Canada Inc., Canada, NUC333.50) for 2 h at 37°C followed by digestion with 10U of alkaline phosphatase from bovine intestinal mucosa (Sigma–Aldrich, P6774-2KU) for 1 h at 37°C. Further procedures were carried out according to the manufacturer's instructions. The absorbance of the samples was measured on İMARK^{TM} Microplate Reader (Bio-Rad) and the 8-OHG concentrations were calculated using ELISA Analysis software with 4-parameter logistic regression algorithm.

ROS Detection

The general ROS were detected *in vivo* in the roots of soybean seedlings using fluorescent dye, CM-H$_2$DCFDA (Life Technologies, C6827) dissolved in dimethyl sulfoxide (DMSO; Sigma–Aldrich, 472301) and diluted in phosphate-buffered saline (PBS) buffer (BioShop Canada Inc., Canada, PBS404) to the total concentration of 10 μM. The roots of seedlings were incubated for 1 h in CM-H$_2$DCFDA, washed with distilled water, and treated for 3 or 24 h with Cd solutions (10 or 25 mg L^{-1}) or distilled water (experimental control). To exclude the possibility of autofluorescence, the negative control incubated for 1 h in PBS buffer instead of CM-H$_2$DCFDA has been applied. All procedures were carried out in the dark room. The level of general ROS was visualized by means of Zeiss Axiovert 200M confocal microscope with 450–490 nm excitation and 515 nm emission light wave length. The 5× magnified images were photographed using AxioCam MRC5 camera.

Measurements of Lipid Peroxidation

Lipid peroxidation was evaluated on the basis of the amount of TBARS according to Cuypers et al. (2011) with small modifications. Roots of soybean seedlings (200 mg) were cut off on ice and homogenized with 3 ml of 10% TCA (Sigma–Aldrich, TO699). The samples were centrifuged (12,000 rpm, 4°C, 10 min) and 1 ml of supernatant was transferred to glass tubes. Thereafter, the tubes were filled with 4 ml of 0.5% TBA, dissolved in 10% TCA, and incubated for 30 min in 95°C. Subsequently, the samples were cooled, mixed by inversion, and centrifuged (5,000 rpm, 4°C, 2 min). The absorbance of supernatant was measured at $\lambda = 532$ nm and corrected for unspecific absorbance at $\lambda = 600$ nm. Amount of TBARS was calculated on the basis of extinction factor (155 mM^{-1}cm^{-1}).

Estimation of the Level of AP-Sites in RNA

The level of abasic sites (AP-sites) has been evaluated using method based on reaction with ARP described by Tanaka et al. (2011). Approximately 5 μg of mRNA dissolved in 50 μl of DNAase- and RNAase-free water was incubated with 50 μl of 2mM N-(aminooxyacetyl)-N′-(D-Biotinoyl)hydrazine (Life Technologies, A-10550) in Tris–EDTA buffer (Sigma–Aldrich, T9285) for 1 h at 37°C. The reaction was stopped by addition of 50 μl of 50 mM formaldehyde (BioShop Canada Inc., Canada, FOR201.1). RNA precipitation has been carried out by addition of 15 μl of 3M sodium acetate (BioShop Canada Inc., Canada, SAA333.100) and 450 μl of pure ethanol (POCH Basic, BA 6480111). The precipitation proceeded for 24 h at −20°C. Thereafter, the samples were centrifuged by 12,000 rcf at 4°C, washed with 500 μl of 75% ethanol and dissolved in 10 μl of Tris–EDTA.

The amount and purity of the obtained mRNA has been measured on NanoCell (Thermo Scientific) at spectrophotometer Biomate^{TM} 3S (Thermo Scientific). A total of 1 μg of the mRNA has been spotted on membrane (Zeta-Probe® Blotting Membranes, Bio-Rad) previously soaked in Tris–EDTA and air-dried. The membrane with samples was irradiated for 15 min with UV light, incubated for 30 min in Casein Blocking Buffer (Sigma–Aldrich, B6429), followed by incubation for 1 h with Streptavidin-Horseradish Peroxidase (HRP) Conjugate (Sigma–Aldrich, GERPN1231) in Casein Blocking Buffer (1:20,000). Thereafter, the membrane was washed 6 times for 4 min with PBS (BioShop Canada Inc., Canada, PBS404.200) containing 0.05% Tween (Sigma–Aldrich, P1379) and developed with Clarity^{TM} Western ECL Substrate (Bio-Rad). The blocking, incubation with streptavidin-HRP, washing, and developing proceeded on rocking platform (Shaker-Rocker MR12, BioSan). Chemiluminescence has been captured on ChemiDoc^{TM} Touch Imaging System (Bio-Rad) with 15 min exposure time. The intensity of spots was measured with Multi Gauge Software (Fuji) as Q-B/pixel^2, where Q is quantity and B is background. The relative density has been expressed as percentage in relation to the control.

Statistical Analysis

The measurements of the 8-OHG level in total RNA and mRNA were carried out on 4 and 5–6 experimental repetitions, respectively. The measurements of growth, lipid peroxidation, and abasic sites level in mRNA were conducted on 3 experimental repetitions, while the estimation of the amount of dead cells in 2–3 experimental repetitions. For evaluation of statistically significant differences, obtained data were analyzed with the use of one-way ANOVA ($p = 0.05$). In the case of evaluation of the level of abasic sites in mRNA, due to the non-parametric distribution, the Mann–Whitney U-test has been applied ($p = 0.05$). Results which showed no statistically significant differences are marked with the same letter.

RESULTS

Exposure of seedlings to Cd for 3 h had no effect on their morphology, growth, or the amount of dead cells (**Figures 1A,C,D**). In turn, 24 h of Cd stress resulted in roots browning and inhibition of their growth (**Figures 1B,C**). Exposure for 24 h to higher Cd concentration led to significant increase in the amount of dead cells (**Figure 1D**).

The level of 8-OHG (oxidatively modified base) was approximately 5 times higher in mRNA in relation to the total RNA (**Figures 2A,B**, **3A,B**). In response to Cd, significant increase in the level of 8-OHG has been noted in the total RNA and mRNA after 3 h of treatment with lower concentration (10 mgl^{-1}) (**Figures 2A,B**). This effect was not observed after 24 h or in response to higher Cd concentration (25 mgl^{-1}) (**Figures 3A,B**).

The opposite tendency has been noted in the case of ROS accumulation. General ROS were detected with specific dye, CM-H$_2$DCFD, which in response to oxidation emits green fluorescent signal. The fluorescence signal was generally lower after 3 than after 24 h. No differences in fluorescence intensity

have been observed between the roots of control and Cd-stressed seedlings after 3 h (**Figure 2C**). However, a visible increase in the fluorescence signal has been noted in roots of the seedling treated with Cd for 24 h in relation to the control (**Figure 3C**).
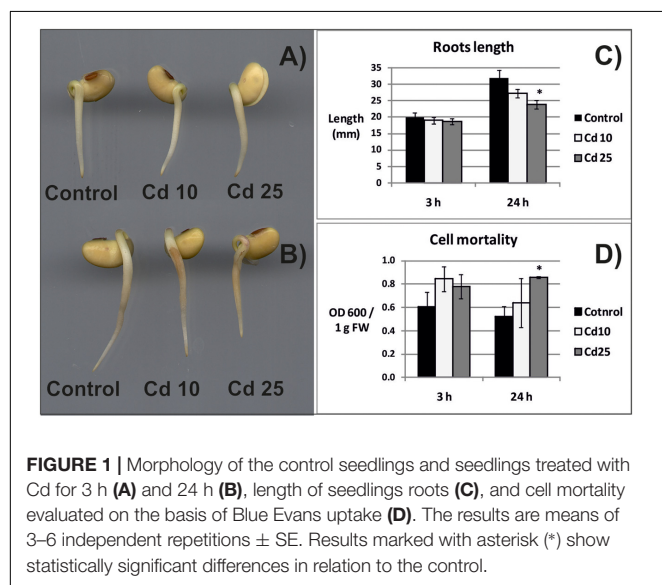
The ROS over-production was correlated in time with increased lipid peroxidation (**Figure 3D**) and AP-sites frequency in the mRNA (**Figure 3E** and **Supplementary Figure S1**). After 3 h, no significant changes in TBARS or AP-sites level were observed between Cd-stressed and control seedlings (**Figures 2D,E**). In turn, after 24 h, the level of TBARS and AP-sites was significantly higher in the roots of seedlings treated with Cd at higher concentration (25 mgl^{-1}) (**Figures 3D,E** and **Supplementary Figure S1**).
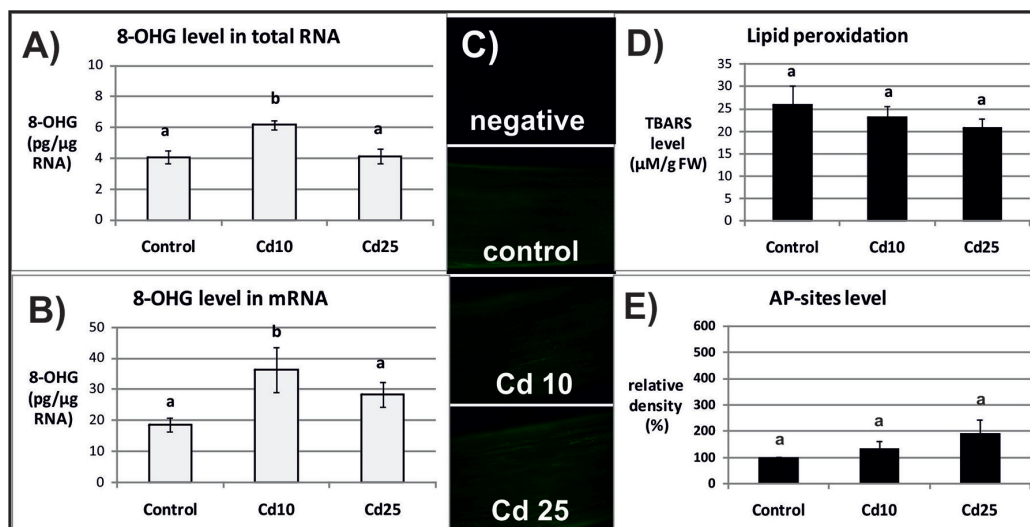
DISCUSSION

Exposure of plants to Cd leads to the development of various symptoms of toxicity (reviewed in Gallego et al., 2012). In the present study, 3 h long treatment with this metal did not affect the morphology and growth of soybean seedlings roots. At this time point, no statistically significant differences in cells viability were noted (**Figures 1A,C,D**). However, already after 24 h, browning and shortening of the roots in response to Cd treatment has been observed (**Figures 1A,C**). Additionally, 24 h long exposure to higher Cd concentration led to significant increase in the amount of dead cells (**Figure 1D**).

Cadmium (Cd) toxicity might be, at least partially, mediated by ROS. These molecules, which include hydrogen peroxide (H$_2$O$_2$), hydroxyl radical (HO), superoxide anion (O$_2^-$), and singlet oxygen (^{1}O$_2$), are highly reactive and mediate oxidation of various cellular compounds. Numerous reports showed that exposure to stresses leads to increase in the level of ROS accompanied by oxidation of proteins and membrane lipids (Gill and Tuteja, 2010; Cuypers et al., 2016; Sewelam et al., 2016). In fact, accumulation of the product of lipid peroxidation, malondialdehyde (MDA) and TBARS, is considered a typical symptom of oxidative stress. However, so far little attention has been given to the oxidation of nucleic acids in plant exposed to unfavorable condition. Moreover, the studies were limited only to changes in the level of DNA oxidation markers (Macovei et al., 2010; Yin et al., 2010), while there is no information concerning ROS impact on RNA. It is worth highlighting that studies on bacteria showed higher susceptibility of RNA to oxidation when compared to DNA. Under the same conditions, the level of oxidatively modified RNA exceeded level of oxidized DNA (Liu et al., 2012). This might be explained by the fact that DNA molecules are more protected due to localization in the nucleus, higher level of packing, and association with numerous proteins. Among the main RNA types, mRNA seems to be the most vulnerable to the oxidation (Bazin et al., 2011).

Indeed in the present study, the frequency of the most common oxidative modification, 8-OHG, was approximately 5 times higher in mRNA than in the total RNA (**Figures 2, 3**). Another modification associated with oxidation processes, abasic sites (AP-sites), were undetectable in the total RNA using same or even higher concentrations of sample as in the case of mRNA

FIGURE 1 | Morphology of the control seedlings and seedlings treated with Cd for 3 h **(A)** and 24 h **(B)**, length of seedlings roots **(C)**, and cell mortality evaluated on the basis of Blue Evans uptake **(D)**. The results are means of 3–6 independent repetitions ± SE. Results marked with asterisk (*) show statistically significant differences in relation to the control.

FIGURE 2 | The level of 8-OHG in the total RNA **(A)**, mRNA **(B)**, general ROS level **(C)**, lipid peroxidation expressed as changes in TBARS level **(D)**, and the level of abasic (apurinic/apyrimidinic) sites (AP-sites) in mRNA expressed as percentage of relative density **(E)** after 3 h of treatment with distilled water (control) or Cd at the concentration 10 mgl^{-1} (Cd 10) or 25 mgl^{-1} (Cd 25). The results are means of 3–5 independent repetitions ± SE. Results marked with the same letter show no statistically significant differences.
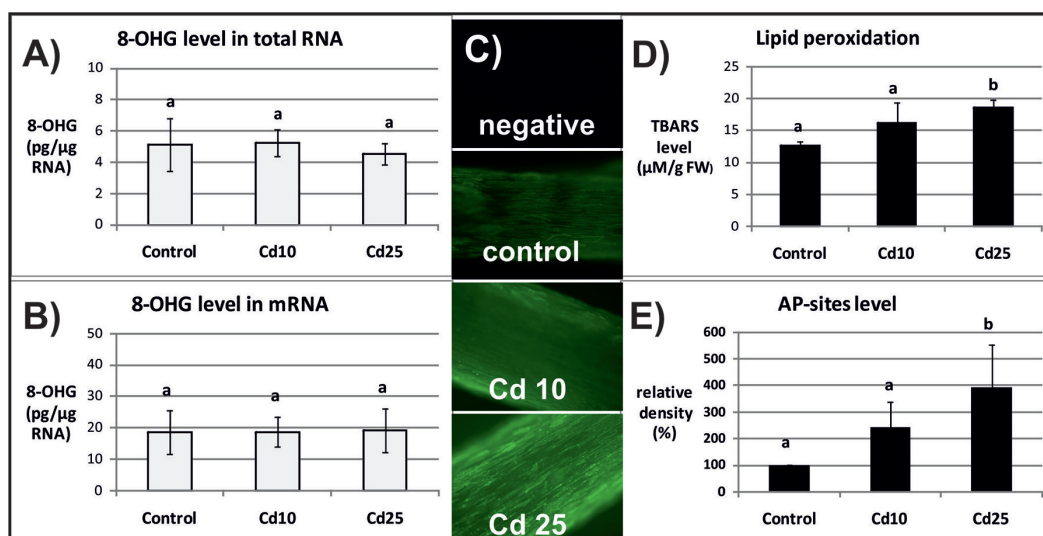
FIGURE 3 | The level of 8-OHG in the total RNA **(A)**, mRNA **(B)**, general ROS level **(C)**, lipid peroxidation expressed as changes in TBARS level **(D)**, and the level of abasic (apurinic/apyrimidinic) sites (AP-sites) in mRNA expressed as percentage of relative density **(E)** after 24 h of treatment with distilled water (control) or Cd at the concentration 10 mgl^{-1} (Cd 10) or 25 mgl^{-1} (Cd 25). The results are means of 3–5 independent repetitions ± SE. Results marked with the same letter show no statistically significant differences.

(data not shown). The results indicate that also this modification occurs more frequently in the transcripts than other RNA types. Interestingly, Cd-dependent induction of 8-OHG and AP-sites were separated in time. Significantly, higher levels of 8-OHG in response to Cd were noted only after 3 h of treatment with lower metal concentration (**Figures 2A,B**). In turn, increase in the level of AP-sites was observed after 24 h of exposure to the higher concentration (**Figure 3E**) and was accompanied by increase in the level of other markers of oxidative stress – strong ROS

over-production (**Figure 3C**) and enhanced lipid peroxidation (**Figure 3D**).

The observed variable in time effect of Cd-dependent ROS signal is in concordance with other studies. For example, several reports indicate that Cd stress leads to the generation of differing in the time of occurrence ROS waves (Garnier et al., 2006; Peréz-Chaca et al., 2014; Lv et al., 2017). In the case of the study on tobacco suspension cells, it has been shown that exposure to this metal leads to generation of three ROS waves, whereas

the earliest one occurred already within the first hour of metal treatment and was dependent on the activity of membrane bond enzyme, NADPH oxidase. Same research reported that Cd cytotoxicity was associated with the second ROS wave resulting from disturbances in mitochondria functioning (Garnier et al., 2006). Also our earlier research showed that ROS signal in soybean seedlings exposed to Cd differs in time. The activity of NADPH oxidase modulated expression of signaling associated genes after 3 h of metal treatment, while significant ROS accumulation and lipid peroxidation were noted only after 24 h (Chmielowska-Bąk et al., 2017). Apparently, the role of ROS in plant cells is temporal, species, and spatial specific (Miller et al., 2008; Cuypers et al., 2016). For example, it has been evidenced in *Arabidopsis thaliana* that ROS signal originating from peroxisomes and chloroplasts has distinct effect on the transcriptome (Sewelam et al., 2014). Another study using agents, which induce distinct ROS types, showed that some transcripts are modulated specifically by O_2^-, H_2O_2, or $^1O_2$ (Gadjev et al., 2006).

Reactive oxygen species might play various roles in plants exposed to stress conditions. On one hand, these molecules are responsible for oxidative damage of membrane lipids, proteins, and nucleic acids (Das and Roychoudhury, 2014; Rio, 2015). On the other hand, ROS are engaged in signaling network and defense mechanism (Cuypers et al., 2016). ROS-dependent signaling includes direct sensing, for example, through transcription factors or serine/threonine protein kinase oxidative signal-inducible 1 (OXI1) and indirect modulation of signal through changes in cellular redox status and/or interaction with other signaling molecules such as nitric oxide, calcium ions, mitogen-activated protein kinases, or plant hormones (Neill et al., 2002; Chi et al., 2013; Kopczewski and Kuźniak, 2013; Wrzaczek et al., 2013; Waszczak et al., 2015; Cuypers et al., 2016). Recently, it has been proposed that ROS signal might be also transmitted by products of oxidation such as oxylipins, peptides derived from protein oxidation, and oxidatively modifies nucleic acids (reviewed in Chmielowska-Bąk et al., 2015). One of the important future challenges in research concerning the role of ROS in plants response to stresses is elucidation of the exact role of specific ROS signals.

In the case of present research, the observed Cd-dependent induction of AP sites is most probably a symptom of oxidative stress (**Figure 3E** and **Supplementary Figure S1**). The assumption is based on the fact that AP sites induction was correlated in time with significant increase in ROS level and lipid peroxdiation (**Figures 3C,D**). However, at the present stage of research, it is difficult to explain the exact role of rapid induction of 8-OHG formation in RNA noted already after 3 h (**Figures 2A,B**), when the symptoms of oxidative stress were still not detectable (**Figures 2C,D**). Studies carried out on animal, human, and plant models showed that 8-OHG formation is not a random but highly selective process limited to defined transcripts, although the mechanism of its selectivity has not been yet discovered. High rate of 8-OHG in transcripts leads to ribosome stalling and in consequence to the decrease in the amount of encoded proteins (Shan et al., 2007; Tanaka et al., 2007; Chang et al., 2008; Bazin et al., 2011; Gao et al., 2013;

Simms et al., 2014). The selective nature of 8-OHG formation and its impact on protein biosynthesis indicate that this process constitutes a newly discovered mechanism of post-transcriptional gene regulation. Interestingly, it seems that the gene regulatory function of transcripts abundant in 8-OHG plays distinct role in animals and plants. In the case of animal and human models, the high 8-OHG level in mRNA has been shown to be associated with the development of neurodegenerative disorders (Shan et al., 2003; Chang et al., 2008; Kong and Lin, 2010). In turn in plants, this oxidative modification of transcripts is essential for regulation of the level of certain proteins and alleviation of seed dormancy – a natural process in plants' life cycle (Bazin et al., 2011; Gao et al., 2013).

In summary, this is the first report showing increased oxidation of total RNA and mRNA in plants exposed to stress. The observed increase in the level of AP-site was correlated in time with strong ROS accumulation and lipid peroxidation indicating that this mRNA modification constitutes a marker of oxidative challenge. However, the rapid induction of 8-OHG is puzzling and its exact role in plants response to unfavorable conditions needs further elucidation.

AUTHOR CONTRIBUTIONS

JC-B and JD designed the research. JC-B and AE-G carried out the cultivation of material, collection of samples, and RNA and mRNA isolation. JC-B and KI conducted the evaluation of ROS and AP-sites level. JC-B, MB, and AE-G performed the measurements of 8-OHG level including isolation of mRNA. JC-B carried out the examination of lipid peroxidation. All authors analyzed the obtained results. JC-B wrote the manuscript. All authors critically read, corrected, and approved the manuscript.

FUNDING

ACKNOWLEDGMENT

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2017.02219/full#supplementary-material

FIGURE S1 | Photograph of an exemplary membrane showing formation of abasic sites (AP-sites) in mRNA isolated from the roots of seedlings treated with distilled water (control) or Cd at the concentration 10 mgl^{-1} (Cd 10) or 25 mgl^{-1} (Cd 25) for 3 and 24 h.

REFERENCES

Antoniali, G., Malfatti, C., and Tell, G. (2017). Unveiling the non-repair face of the base excision repair pathway in RNA processing: a missing link between DNA repair and gene expression? *DNA Repair* 56, 65–74. doi: 10.1016/j.dnarep.2017.06.008

Barciszewski, J., Barciszewska, M. Z., Siboska, G., Rattan, S. I. S., and Clark, F. C. C. (1999). Some unusual nucleic acid bases are products of hydroxyl radical oxidation of DNA and RNA. *Mol. Biol. Rep.* 26, 231–238. doi: 10.1023/A:1007058602594

Bazin, J., Langlade, N., Vincourt, P., Arribat, S., Balzergue, S., El-Maarouf-Bouteau, H., et al. (2011). Targeted mRNA oxidation regulates sunflower seed dormancy alleviation during dry after-ripening. *Plant Cell* 23, 2196–2208. doi: 10.1105/tpc.111.086694

Calabretta, A., Küpfer, P. A., and Leumann, J. (2015). The effect of RNA base lesions on mRNA translation. *Nucleic Acids Res.* 19, 4713–4720. doi: 10.1093/nar/gkv377

Chang, Y., Kong, Q., Shan, X., Tian, G., Llieva, H., Cleveland, D. W., et al. (2008). Messenger RNA oxidation occurs early in disease pathogenesis and promotes motor neuron degeneration in ALS. *PLOS ONE* 3:e2849. doi: 10.1371/journal.pone.0002849

Chi, Y. H., Paeng, S. K., Kim, M. J., Hwang, G. Y., Melencion, S. M. B., Oh, H. T., et al. (2013). Redox-dependent functional switching of plant proteins accompanying with their structural changes. *Front. Plant Sci.* 41:277. doi: 10.3389/fpls.2013.00277

Chmielowska-Bąk, J., Arasimowicz-Jelonek, M., Izbiańska, K., Frontasyeva, M., Zinicovskaia, I., Guance-Varela, C., et al. (2017). NADPH oxidase is involved in regulation of gene expression and ROS overproduction in soybean (*Glycine max* L.) seedlings exposed to cadmium. *Acta Soc. Bot. Pol.* 86, 1–17. doi: 10.5586/asbp.3551

Chmielowska-Bąk, J., and Deckert, J. A. (2012). Common response to common danger? Comparison of animal and plant signaling pathways involved in cadmium sensing. *J. Cell Commun. Signal.* 6, 191–204. doi: 10.1007/s12079-012-0173-3

Chmielowska-Bąk, J., Izbiańska, K., and Deckert, J. (2015). Products of lipid, protein and RNA oxidation as signals and regulators of gene expression in plants. *Front. Plant Sci.* 6:405. doi: 10.3389/fpls.2015.00405

Cuypers, A., Hendrix, S., Amaral dos Reis, R., De Smet, S., Deckers, J., Gielen, H., et al. (2016). Hydrogen peroxide, signaling in disguise during metal phytotoxicity. *Front. Plant Sci.* 7:470. doi: 10.3389/fpls.2016.00470

Cuypers, A., Smeets, K., Ruytinx, J., Opdenakker, K., Keunen, E., Remens, T., et al. (2011). The cellular redox state as a modulator in cadmium, and copper responses in *Arabidopsis thaliana* seedlings. *J. Plant Physiol.* 168, 309–316. doi: 10.1016/j.jplph.2010.07.010

Čolak, E. (2008). New markers of oxidative damage to macromolecules. *J. Med. Biochem.* 27, 1–16. doi: 10.2478/v10011-007-0049-x

Das, K., and Roychoudhury, A. (2014). Reactive oxygen species (ROS) and response of antioxidants as ROS-scavengers during environmental stress in plants. *Front. Environ. Sci.* 2:53. doi: 10.3389/fenvs.2014.00053

Dumitrel, G.-A., Glevitzky, M., Popa, M., Chirila, D., and Palea, A. (2017). Monitoring of lead, copper and cadmium contamination level of soil from Zlatna Region – Romania. *J. Environ. Prot. Ecol.* 18, 55–62.

Gadjev, I., Vanderauwera, S., Gechev, T. S., Laloi, C., Minkov, I. N., Shulaev, V., et al. (2006). Transcriptomic footprints disclose specificity of reactive oxygen species signaling in Arabidopsis. *Plant Physiol.* 141, 436–445. doi: 10.1104/pp.106.078717

Gallego, S. M., Pena, L. B., Barcia, R. A., Azpilicueta, C. E., Iannone, M. F., Rosales, E. P., et al. (2012). Unravelling cadmium toxicity and tolerance in plants: insight into regulatory mechanisms. *Environ. Exp. Bot.* 83, 33–46. doi: 10.1016/j.envexpbot.2012.04.006

Gao, F., Rampitsch, C., Chitnis, V. R., Humphreys, G. D., Jordan, M. C., and Ayele, B. T. (2013). Integrated analysis of seed proteome and mRNA oxidation reveals distinct post-transcriptional features regulating dormancy in wheat (*Triticum aestivum* L.). *Plant Biotechnol. J.* 11, 921–932. doi: 10.1111/pbi.12083

Garnier, L., Simon-Plas, F., Thuleau, P., Agnel, J-P., Blein, J-P., Ranjeva, R., et al. (2006). Cadmium affects tobacco cells by a series of three waves of reactive oxygen species that contribute to cytotoxicity. *Plant Cell Environ.* 29, 1956–1969. doi: 10.1111/j.1365-3040.2006.01571.x

Gill, S. S., and Tuteja, N. (2010). Reactive oxygen species and antioxidant machinery in abiotic stress tolerance in crop plants. *Plant Physiol. Biochem.* 48, 909–930. doi: 10.1016/j.plaphy.2010.08.016

Kah, M., Levy, L., and Brown, C. (2012). Potential role for effects of land contamination on human health. 1 the case of cadmium. *J. Toxicol. Environ. Health* 15, 348–363. doi: 10.1080/10937404.2012.705107

Khan, S., Khan, M. A., and Rehman, S. (2011). Lead and cadmium contamination of different roadside soils and plants in Peshawar City, Pakistan. *Pedosphere* 21, 351–357. doi: 10.1016/S1002-0160(11)60135-5

Kong, Q., and Lin, C. G. (2010). Oxidative damage to RNA: mechanism, consequences, and diseases. *Cell. Mol. Life Sci.* 67, 1817–1829. doi: 10.1007/s00018-010-0277-y

Kopczewski, T., and Kuźniak, E. (2013). Redox signals as a language of inter-organellar communication in plant cells. *Cent. Eur. J. Biol.* 8, 1153–1163.

Lehotai, N., Petö, A., Bajkán, S., Erdei, L., Tari, I., and Kolbert, Z. (2011). In vivo and in situ visualization of early physiological events induces by heavy metals in pea root meristem. *Acta Physiol. Plant.* 33, 2199–2207. doi: 10.1007/s11738-011-0759-z

Liu, M., Gong, X., Alluri, R. K., Wu, J., Sablo, T., and Li, Z. (2012). Characterization of RNA damage under oxidative stress in *Escherichia coli*. *Biol. Chem.* 393, 123–132. doi: 10.1515/hsz-2011-0247

Lv, W., Yang, L., Xu, C., Shi, Z., Shao, J., Xian, M., et al. (2017). Cadmium disrupts the balance between hydrogen peroxide and superoxide radical by regulating endogenous hydrogen sulfide in the root tip of *Brassica rapa*. *Front. Plant Sci.* 8:232. doi: 10.3389/fpls.2017.00232

Macovei, A., Balestrazzi, A., Confalonieri, M., and Carbonera, D. (2010). The tyrosyl-DNA phosphodiesterae gene family in *Medicago truncatula* Gaertn.: bioinformatic investigation and expression profiles in response to copper- and PEG-mediated stress. *Planta* 232, 393–407. doi: 10.1007/s00425-010-1179-9

Miller, G., Shulaev, V., and Mittler, R. (2008). Reactive oxygen signaling and abiotic stress. *Physiol. Plant.* 133, 481–489. doi: 10.1111/j.1399-3054.2008.01090.x

Neill, S., Desikan, R., and Hancock, J. (2002). Hydrogen peroxide signaling. *Curr. Opin. Plant Biol.* 5, 388–395. doi: 10.1016/S1369-5266(02)00282-0

Peréz-Chaca, M. V., Rodrígues-Serrano, M., Molina, A. S., Pedranzani, H. E., Zirulnik, F., Sandalio, L. M., et al. (2014). Cadmium induces two waves of reactive oxygen species in *Glycine max* (L.) roots. *Plant Cell Environ.* 37, 1672–1687. doi: 10.1111/pce.12280

Rio, L. A. (2015). ROS and RNS in plant physiology: an overview. *J. Exp. Bot.* 66, 2827–2837. doi: 10.1093/jxb/erv099

Sedelnikova, O. A., Redon, C. E., Dickeym, J. S., Nakamura, A. J., Georgakilas, A. G., and Bonner, W. M. (2010). Role of oxidatively induced DNA lesions in human pathogenesis. *Mutat. Res.* 704, 152–159. doi: 10.1016/j.mrrev.2009.12.005

Sewelam, N., Jaspert, N., Van Der Kelen, K., Tognetti, V. B., Schmitz, J., Frerigmann, H., et al. (2014). Spatial H_2O_2 signaling specificity: H_2O_2 from chloroplasts and peroxisomes modulates the plant transcriptome differentially. *Mol. Plant* 7, 1191–1210. doi: 10.1093/mp/ssu070

Sewelam, N., Kazan, K., and Schenk, P. M. (2016). Global plant stress signaling: reactive oxygen species at the cross-road. *Front. Plant Sci.* 7:187. doi: 10.3389/fpls.2016.00187

Shan, X., Chang, Y., and Lin, G. (2007). Messenger RNA oxidation is an early event preceding cell death and causes reduced protein expression. *FASEB J.* 21, 2753–2764. doi: 10.1096/fj.07-8200com

Shan, X., Tashiro, H., and Lin, C. G. (2003). The identification and characterization of oxidized RNAs in Alzheimer's disease. *J. Neurosci.* 23, 4913–4921.

Simms, C. L., Hudson, B. H., Mosior, J. W., Rangwala, A. S., and Zaher, H. S. (2014). An active role for the ribosome in determining the fate of oxidized mRNA. *Cell Rep.* 9, 1256–1264. doi: 10.1016/j.celrep.2014.10.042

Stirpe, M., Palermo, V., Ferrari, M., Mroczek, S., Kufel, J., Falcone, C., et al. (2016). Increased levels of RNA oxidation enhances the reversion frequency in aging pro-apoptotic yeast mutants. *Apoptosis* 22, 200–206. doi: 10.1007/s10495-016-1319-1

Tanaka, M., Chock, P. B., and Stadtman, E. R. (2007). Oxidized Messenger RNA induces translation errors. *Proc. Natl. Acad. Sci. U.S.A.* 104, 66–71. doi: 10.1073/pnas.0609737104

Tanaka, M., Song, H., Küpfer, P. A., Leuman, C. J., and Sonntag, W. E. (2011). An assay for RNA oxidation induced abasic sites using the Aldehyde

Reactive Probe. *Free Radic. Res.* 45, 237–247. doi: 10.3109/10715762.2010.535529

Wang, L., Cui, X., Cheng, H., Chen, F., Wang, J., Zhao, X., et al. (2015). A review of soil cadmium contamination in China including a health risk assessment. *Environ. Sci. Pollut. Res.* 22, 16441–16452. doi: 10.1007/s11356-015-5273-1

Waszczak, C., Akter, S., Jacques, S., Huang, J., Messens, J., and Breusegem, F. (2015). Oxidative post-translational modifications of cysteine residues in plant signal transduction. *J. Exp. Bot.* 66, 2923–2934. doi: 10.1093/jxb/erv084

Wrzaczek, M., Brosché, M., and Kangasjärvi, J. (2013). ROS signaling Loops – production, perception and regulation. *Curr. Opin. Plant Biol.* 16, 575–582. doi: 10.1016/j.pbi.2013.07.002

Yin, L., Wang, S., Eltayeb, A. E., Uddin, M. I., Yamamoto, Y., Tsuji, W., et al. (2010). Overexpression of dehydroascorbate reductase, but not monodehydroascorbate reductase, confers tolerance to aluminum stress in transgenic tobacco. *Planta* 231, 609–621. doi: 10.1007/s00425-009-1075-3

The Effect of Centrifugal Force in Quantification of Colorectal Cancer-Related mRNA in Plasma Using Targeted Sequencing

Vivian Weiwen Xue[1], Simon Siu Man Ng[2], Wing Wa Leung[2], Brigette Buig Yue Ma[3], William Chi Shing Cho[4], Thomas Chi Chuen Au[3], Allen Chi Shing Yu[5], Hin Fung Andy Tsang[1] and Sze Chuen Cesar Wong[1]*

[1] Department of Health Technology and Informatics, Faculty of Health and Social Sciences, Hong Kong Polytechnic University, Kowloon, Hong Kong, [2] Department of Surgery, Faculty of Medicine, The Chinese University of Hong Kong, Shatin, Hong Kong, [3] State Key Laboratory in Oncology in South China, Sir YK Pao Centre for Cancer, Department of Clinical Oncology, Hong Kong Cancer Institute and Li Ka Shing Institute of Health Sciences, The Chinese University of Hong Kong, Shatin, Hong Kong, [4] Department of Clinical Oncology, Queen Elizabeth Hospital, Kowloon, Hong Kong, [5] Department of Computer Science, University of Oxford, Oxford, United Kingdom

In our previous study, we detected the effects of centrifugal forces on plasma RNA quantification by quantitative reverse transcription PCR. The aims of this study were to perform targeted mRNA sequencing and data analysis in healthy donors' plasma prepared by two centrifugation protocols and to investigate the effects of centrifugal forces on plasma mRNA quality and quantity. Targeted mRNA sequencing was performed using a custom panel with 108 colorectal cancer-related genes in 18 healthy donors' plasma that prepared by (1) 3,500 g for 10 min at 4°C and (2) 1,600 g for 10 min at 4°C followed by 16,000 g for 10 min at 4°C. Results showed that plasma ribosomal RNA was detected in 16/18 (88.9%) 3,500 g and 6/18 (33.3%) 1,600 g followed by 16,000 g centrifuged plasma. For targeted sequencing, 75/108 (69.4%) and 86/108 (79.6%) genes were detected in 3,500 and 1,600 g followed by 16,000 g, respectively, while 16/108 (14.8%) genes were not detected in both centrifugations. Detailed analysis showed that 2 of 108 (1.85%) genes showed lower expressions in 3,500 g than in 1,600 g followed by 16,000 g. The median expressions of genes in 3,500 g were positively correlated with the expressions in 1,600 g followed by 16,000 g ($R^2 = 0.9471$, $P < 0.0001$, Spearman rank correlation). Meanwhile, plasma samples were not distinctively clustered based on centrifugal forces according to hierarchical clustering. Targeted mRNA sequencing and subsequent data analysis were performed in this study to investigate the effects of two different centrifugal forces that are commonly used in plasma collection. Our targeted sequencing results help to understand the centrifugal force effects on plasma mRNA, and these findings show that the centrifugation protocol for plasma mRNA research using targeted sequencing can be standardized which facilitates multicenter studies for comparison and quality assurance in the future.

Keywords: centrifugal force, targeted sequencing, plasma mRNA, colorectal cancer, gene expression

INTRODUCTION

Colorectal cancer (CRC) is one of the most serious health issues worldwide (Brenner et al., 2014). The survival rate of primary CRC is significantly higher than the rate of advanced cancer, which means that a more effective cancer screening is helpful in CRC prevention (Levin et al., 2008). Currently recommended annual screening for people age ≥ 50 years is Guaiac-based fecal occult blood test (gFOBT) or fecal immunochemical test (FIT) (Smith et al., 2017). Although they are economical and non-invasive tests, gFOBT is not specific enough due to false-positive detection from hemorrhoids and ulcers, and both gFOBT and FIT are not sensitive in neoplasm detection (Brenner and Tao, 2013; Widlak et al., 2017). The insufficient sensitivity of current screening for non-invasive and early detection of CRC urges a more effective non-invasive detection based on liquid biopsy such as plasma. Plasma mRNA has been used as diagnostic and prognostic tumor markers in various cancers (García et al., 2007). The detection of plasma mRNA is non-invasive and flexible, which is beneficial to cancer patients' follow-up after surgeries or adjuvant therapies, and it has potential to monitor cancer recurrence as well (Wong et al., 2004b; Stein et al., 2011). However, mRNA has a low abundance, and it is fragmented in plasma (Savelyeva et al., 2017). Moreover, different blood processing protocols, such as filtering and different centrifugal forces used in plasma preparation, may result in different quantification of plasma mRNA (Ng et al., 2002; El-Hefnawy et al., 2004; Wong et al., 2007). In a previous study, an additional centrifugation in 1,300 g for 10 min after a routine centrifugation in 1,800 g for 10 min was found to decrease plasma RNA concentration over 20 times (El-Hefnawy et al., 2004). Moreover, our earlier work showed that mRNA quantity was significantly different in metastatic CRC patients' plasma samples prepared by two centrifugations, 800 g for 8 min at 4°C and 4,500 g for 8 min at 4°C, which detected *CTNNB1*, *SELP*, *KRT20,* and *GAPDH* mRNA using quantitative reverse transcription PCR (RT-qPCR) (Wong et al., 2007). Those results demonstrated that different centrifugal forces could lead to different quantities of mRNA in plasma samples from CRC patients. This artifact alerts us the importance to standardize the centrifugal protocol for plasma mRNA analysis. Otherwise, the data obtained cannot be interpreted and compared with other studies.

With the prevalence of next-generation sequencing, RNA deep sequencing has been used as an approach for transcriptome profiling in plasma samples (Wang et al., 2014; Shih et al., 2015). However, there is no standardized protocol of plasma preparation or explicit descriptions of centrifugal forces effects on plasma mRNA, which has crucial impacts on reproducibility of applications and multicenter researches. Some studies used plasma samples collected from centrifugal forces, such as 1,400 or 1,600 g, to study microRNAs (miRNAs) and mRNAs cancer markers using RNA deep sequencing, respectively (Wang et al., 2014; Shih et al., 2015). On the other hand, some researchers used a higher centrifugal forces to profile plasma extracellular RNAs using small RNA deep sequencing (Freedman et al., 2016; Danielson et al., 2017). In one study, plasma samples were prepared in 2,500 g for 22 min at 4°C with an additional centrifugation in 8,000 g for 5 min after plasma thawing prior to RNA extraction (Freedman et al., 2016). In another study, plasma samples were prepared in 1,000 g for 10 min at room temperature with an additional centrifugation in 2,000 g after plasma thawing prior to RNA extraction (Danielson et al., 2017). Up to now, no study has reported the effect of centrifugal force on plasma mRNA based on RNA sequencing data. Therefore, we aim to examine the effect of centrifugal force on plasma mRNA quantity and quality that may be important for cancer detection and monitoring.

In this study, we examined a panel of CRC-related mRNA in healthy donors' (HDs) plasma prepared by two commonly used centrifugations (1) 3,500 g for 10 min at 4°C and (2) 1,600 g for 10 min at 4°C followed by 16,000 g for 10 min at 4°C. CRC-related mRNA expression was detected using targeted deep sequencing. Subsequently, differential expression and correlation of expression in two centrifugal forces were analyzed. The information obtained from this study will be helpful for us to understand the centrifugal force effects on the expression level of CRC-related mRNAs in plasma samples, which could facilitate us to develop a standardized and effective protocol of CRC biomarker detection using targeted mRNA sequencing in plasma samples.

MATERIALS AND METHODS

Healthy Donors Recruitment and Plasma Collection

Eighteen HDs were recruited in this study. For each donor, 15 ml peripheral blood was collected in K3 EDTA tubes (Greiner Bio-one, Austria) and divided to two parts evenly. One portion was centrifuged for 3,500 g, 10 min at 4°C, and 3.2 ml plasma was collected and preserved by 9.6 ml Trizol LS Reagent (Thermo Fisher Scientific, USA) before storage at −80°C. Another portion was centrifuged for 1,600 g, 10 min at 4°C followed by 16,000 g for 10 min at 4°C, and 3.2 ml plasma was collected and preserved in the same way. Microfuge 22R Centrifuge and F301.5 rotor (Beckman Coulter) were used for centrifugation in plasma preparation. Blood processing was done within 4 h after blood draw. All donors were recruited with written informed consent. The study was approved by the Joint Chinese University of Hong Kong and New Territories Easter Cluster Clinical Research Ethics Committee (CREC-2014.224).

RNA Extraction and Purification

For each sample, 3.2 ml plasma was used for RNA extraction using our established protocol (Wong et al., 2004a,b). In brief, the aqueous layer with RNA was separated after adding chloroform (Sigma-Aldrich, USA) followed by centrifugation for 12,000 g, 15 min at 4°C. Then, 0.54 volume of absolute ethanol (Sigma-Aldrich, USA) was added to the aqueous layer to achieve appropriate binding conditions. The mixture was purified using RNeasy Mini Kit (Qiagen, Germany) (Wong et al., 2004a,b). Subsequently, DNase digestion using TURBO DNA-free Kit (Invitrogen, Lithuania) was performed, and the DNA-free RNA was concentrated using RNeasy MinElute Cleanup Kit (Qiagen, Germany) according to the manufacturer's instructions (Tsui

et al., 2014). Plasma total RNA was eluted in 14 μl RNase-free water, and it was stored at −80°C until use. The quality and quantity of RNA was detected using Agilent RNA 6000 Pico Kit (Agilent Technologies, Lithuania) on 2100 Bioanalyzer. RNA integrity number (RIN) and the percentage of RNA fragments > 200 nt (DV$_{200}$) were detected as the quality indicators. RIN is a standardized value to describe RNA quality, which has considered the 28S/18S ratio and other features from electrophoretic RNA separation results (Schroeder et al., 2006). DV$_{200}$ is a parameter to evaluate the length distribution of fragmented RNA (Landolt et al., 2016).

Sequencing Library Preparation and Data Analysis

Sequencing library was prepared using a custom designed TruSeq Targeted RNA Expression Kit (Illumina, USA), which was used to examine a panel of 108 CRC-related genes including 93 Wnt-signaling genes, existing CRC markers from literatures and a control gene (Supplementary Table 1). The cDNA libraries were synthesized using 5 μl extracted plasma RNA, which was equivalent to about 670 pg RNA per sample, and the preparation of sequencing libraries was according to the manufacturer's instructions with slight modifications, including (1) 2-fold diluted adapters to amplify libraries and (2) two times of clean-up for PCR products using AMPure XP beads (Beckman Coulter, USA) (Tsui et al., 2014). The quality and quantity of prepared cDNA libraries were checked by Agilent High Sensitivity DNA Kit (Agilent Technologies, Lithuania) and qPCR, respectively. FastStart Universal SYBR Green Master (Roche, Germany) was used in quantification. Primers with 5′-AATGATACGGCGACCACCGAGAT-3′ and 5′-CAAGCAGAAGACGGCATACGA-3′ matched sequences within adapters were used. Illumina format DNA standard (Qiagen, Germany) was prepared by a serial dilution to achieve the standard curve for absolute quantification. The pooled sequencing library with 5% PhiX control (Illumina) was sequenced for single-end 51 bp length on MiSeq System using MiSeq Reagent Kit v3 (Illumina, Singapore). The targeted RNA sequencing data were available in Sequence Read Archive (SRA) database (SRP125573).

Data analysis for targeted mRNA sequencing included two parts. The primary analysis was performed on MiSeq reporter. After base calling, FASTQ files of sequences with high sequencing quality were aligned to hg19 reference genome based on custom designed regions. Raw aligned replicate counts of each gene for each sample were output. Counts per million (CPM) of genes were calculated as normalized expression for correcting biases due to library sizes. Let the raw count of gene i in a sample j is C_{ij}, with $i = 1$ to n and $j = 1$ to m. The calculation of CPM_{ij} is below (Rau et al., 2013; Law et al., 2014):

$$CPM_{ij} = \frac{C_{ij}}{\sum_{i=1}^{n} C_{ij}} \times 10^6$$

The normalized expression of gene was shown using log$_2$ scale as $log_2 \ (CPM+1)$ to avoid log transformation for zero CPM (Law et al., 2014; Tsui et al., 2014). The secondary analysis was performed as below pipeline: a non-specific filter as "keep the gene if it has > 1 CPM in ≥ 18 plasma samples" was used to remove uninformative signals and increase detection power without dependency on centrifugal force labels, which generally excluded low-abundance genes and reduced dispersion (Bourgon et al., 2010; Robinson et al., 2010; Rau et al., 2013). After the filtering, DESeq2 was used to estimate dispersion and detect differential expression using paired sample test for the remaining genes (Love et al., 2014). A cutoff of fold change > 4 and adjusted P value < 0.05 was used to identify significant differences. Adjusted P value was calculated based on Benjamini-Hochberg correction (Benjamini and Hochberg, 1995).

TABLE 1 | The summary of detection of genes in two centrifugal forces.

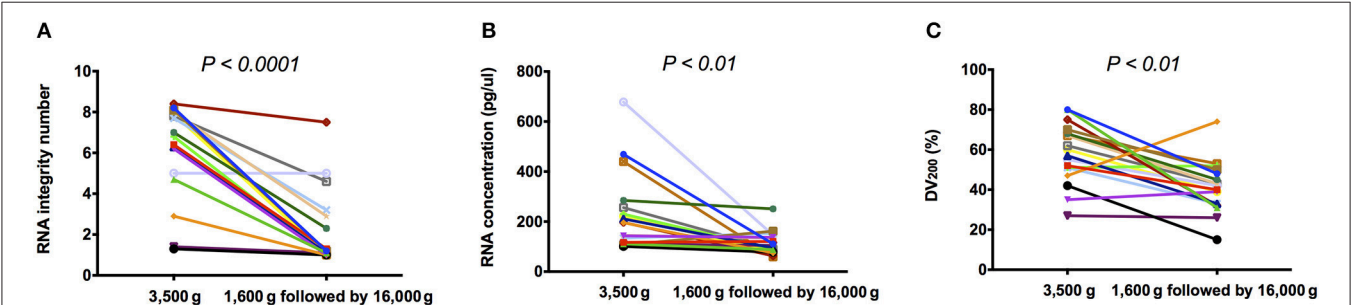|	3,500 g	1,600 g followed by 16,000 g
Genes (n = 108)		
Detectable	75 (69.4%)	86 (79.6%)
In both centrifugations	69 (63.8%)	69 (63.8%)
Only in this centrifugation	6 (5.6%)	17 (15.8%)
Undetectable	33 (30.6%)	22 (20.4%)
In both centrifugations	16 (14.8%)	16 (14.8%)
Only in this centrifugation	17 (15.8%)	6 (5.6%)

FIGURE 1 | The quality and quantity of plasma RNA. **(A)** RIN; **(B)** RNA concentration; and **(C)** DV$_{200}$ of extracted plasma total RNA in 3,500 g were significantly higher than those in 1,600 g followed by 16,000 g (P < 0.05, Wilcoxon matched-pairs signed rank test).

Statistical Analysis

Statistical analysis was performed by Wilcoxon matched-pairs signed rank test and Spearman correlation in Prism 5. $P < 0.05$ was regarded as significant difference and significant correlation, respectively.

RESULTS

Plasma RNA Quality and Quantity From Two Centrifugal Forces

Using Bioanalyzer, 18S and 28S rRNAs were detected in 16/18 (88.9%) 3,500 g centrifuged plasma and 6/18 (33.3%) 1,600 g followed by 16,000 g centrifuged plasma, respectively (Supplementary Figure 1). Besides, RIN was detected with median of 6.90 (range: 1.3–8.4) and 1.15 (range: 1–7.5) in 3,500 and 1,600 g followed by 16,000 g centrifuged plasma, respectively. RIN in 3,500 g was significantly higher than those in 1,600 g followed by 16,000 g (**Figure 1A**, $P < 0.0001$, Wilcoxon matched-pairs signed rank test). RNA concentration was detected with median concentration of 196.5 (range: 101–678) and 100.0 (range: 59–251) pg/μl in 3,500 and 1,600 g followed by 16,000 g centrifuged plasma, respectively. RNA concentration in 3,500 g was significantly higher than those in 1,600 g followed by 16,000 g (**Figure 1B**, $P < 0.01$, Wilcoxon matched-pairs signed rank test). DV_{200} was detected with median percentage of 58.5 (range: 27–80) and 41.0 (range: 15–74) in 3,500 and 1,600 g followed by 16,000 g centrifuged plasma, respectively. DV_{200} in 3,500 g was significantly higher than those in 1,600 g followed by 16,000 g (**Figure 1C**, $P < 0.01$, Wilcoxon matched-pairs signed rank test).

Summary of Plasma mRNA Targeted Sequencing

Overall, the number of total raw reads in targeted sequencing on MiSeq is 39.0 million with at least 93% $\geq$ Q30. Among them, about 32.9 million reads were high quality, and 20.9% of them were aligned to the targeted regions in human genome hg19.

Detection of the 108 CRC-related genes was summarized in **Table 1**. Sixteen (14.8%) of 108 genes were undetectable in both centrifugations. Besides, 6 genes could only be detected in 3,500 g but not in 1,600 g followed by 16,000 g centrifuged plasma, while 17 genes could only be detected in 1,600 g followed by 16,000 g but not in 3,500 g centrifuged plasma. Details of the genes detected only in one centrifugal force condition were listed in **Table 2**. These 23 genes were detected in $\leq$ 5 plasma samples, and the majority of them (73.9%) were low-abundance (raw counts $\leq$ 10 counts).

Expression Filtering and Gene Expression Levels in Two Centrifugal Forces

Detectable genes in 3,500 g (**Figure 2A**) and 1,600 g followed by 16,000 g (**Figure 2B**) were 75 genes and 86 genes, respectively. Based on the normalized expression levels (CPM), 25 of 108

TABLE 2 | The expression of genes detected only in one of two centrifugal forces.

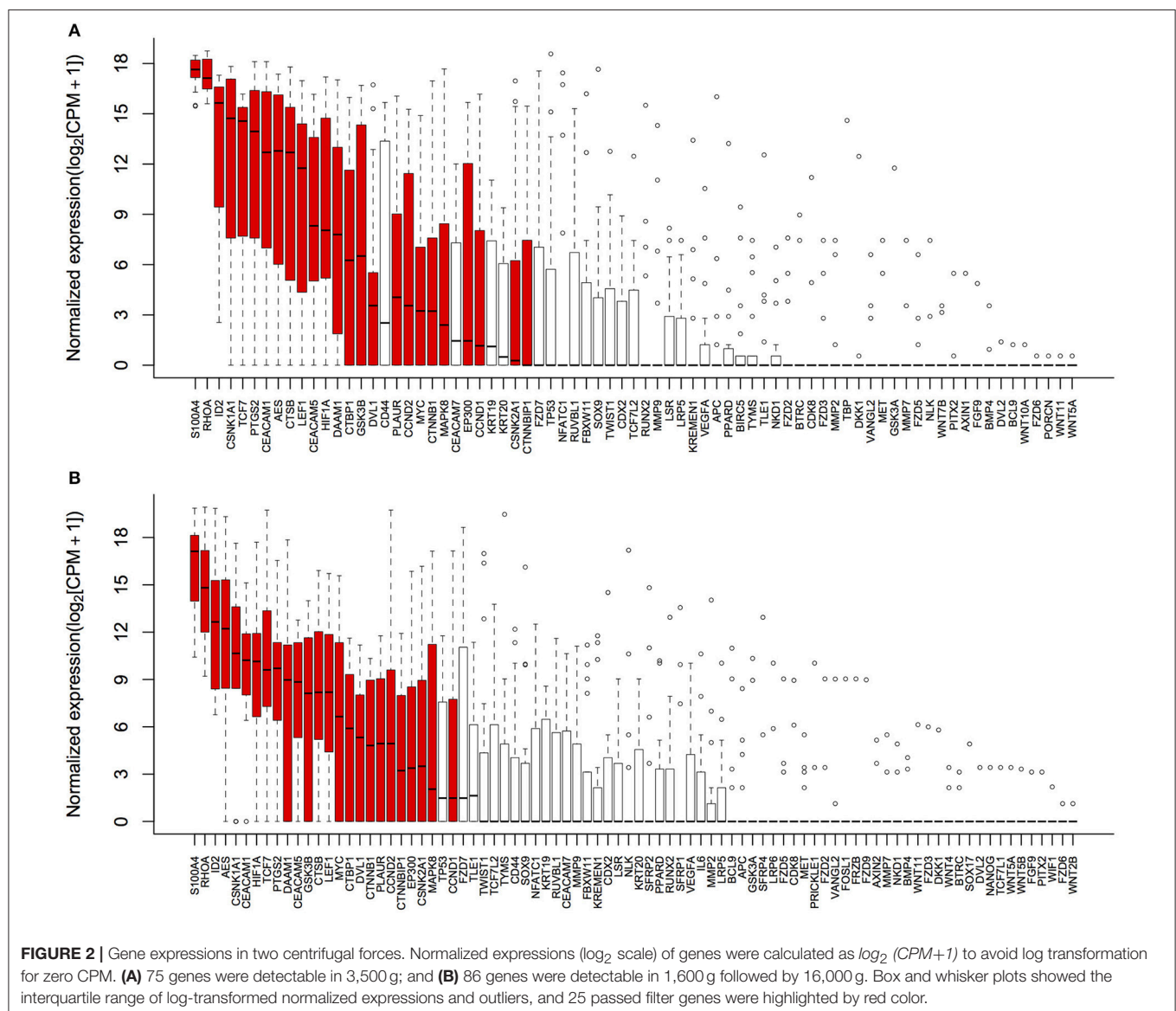| Centrifugation	Gene	Detectable	Raw counts in each detectable sample (count)	Normalized expression in each detected sample (CPM)	Range of normalized expression (CPM)
3,500 g	AXIN1	1/18	1	43.6	43.6
	BIRC5	5/18	2, 1, 2, 4, 1	2.7, 0.5, 191.6, 691.7, 10.7	0.5–691.7
	PORCN	1/18	1	0.5	0.5
	TBP	1/18	844	24825.0	24825.0
	WNT7B	2/18	8, 1	7.9, 10.7	7.9–10.7
	WNT10A	1/18	1	1.3	1.3
1,600 g followed by 16,000 g	AXIN2	2/18	1, 1	11.8, 34.6	11.8–34.6
	FOSL1	1/18	1	524.4	524.4
	FRZB	1/18	1	524.4	524.4
	FZD9	1/18	56	503.0	503.0
	IL6	5/18	1, 1, 1, 3, 25	7.8, 44.0, 9.0, 1573.2, 242.4	7.8–1573.2
	LRP6	2/18	2, 6	1048.8, 58.2	58.2–1048.8
	NANOG	1/18	1	9.7	9.7
	PRICKLE1	2/18	2, 1	1048.8, 9.7	9.7–1048.8
	SFRP1	3/18	4, 23, 18	984.0, 12060.8, 174.6	174.6–12060.8
	SFRP2	4/18	6, 1, 55, 10	2051.3, 11.8, 28841.1, 97.0	11.8–28841.1
	SFRP4	2/18	1, 15	44.0, 7865.8	44.0–7865.8
	SOX17	1/18	3	29.1	29.1
	TCF7L1	1/18	1	9.7	9.7
	WIF1	1/18	3	3.6	3.6
	WNT2B	1/18	1	1.2	1.2
	WNT4	2/18	1, 1	3.4, 9.7	3.4–9.7
	WNT5B	1/18	1	9.0	9.0

(23.1%) genes that had the higher expression compared to the majority of other genes in both centrifugations passed the filter (> 1 CPM in ≥ 18 plasma samples), and they were highlighted by red color in **Figure 2**. Only these 25 genes were included in the following differential expression analysis. The median expressions of those genes in 3,500 g were positively correlated with the expressions in 1,600 g followed by 16,000 g (**Figure 3**, $R^2 = 0.9471$, $P < 0.0001$, Spearman rank correlation). On the other hand, 83 genes were filtered out because of their low sequencing coverage, and they were not included in downstream differential expression analysis.

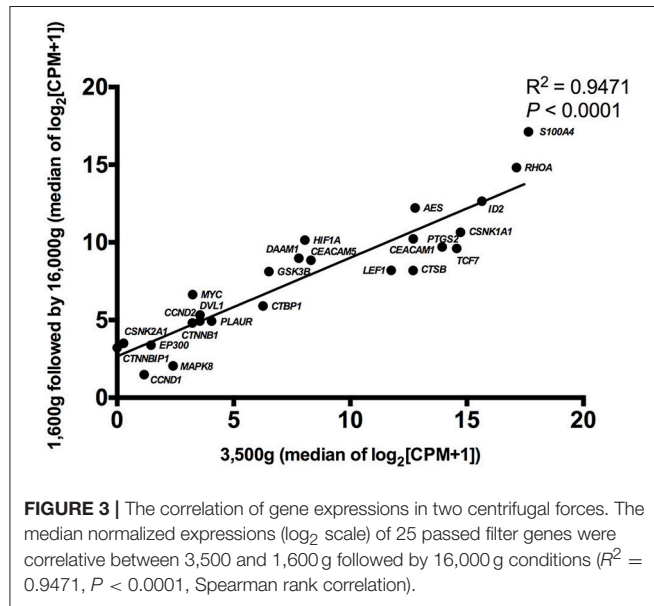## Differential Gene Expression in Two Centrifugal Forces

The differential expression analysis and the fold-change estimation of 25 passed filter genes were performed by DESeq2. Results were listed in **Table 3**, and genes with significant difference in expression were highlighted in bold. Among them, MYC proto-oncogene (*MYC*) and hypoxia inducible factor 1 alpha subunit (*HIF1A*) showed significantly lower expressions in 3,500 g compared with in 1,600 g followed by 16,000 g (16.67-fold with adjusted $P < 0.005$ and 5.56-fold with adjusted $P < 0.05$, respectively). In detail, *MYC* was detected in 11/18 (61.1%) plasma samples in both centrifugations with the median normalized expression of 9.2 (range: 0.0–30443.0) and 98.9 (range: 0.0–48931.0) CPM in 3,500 g and 1,600 g followed by 16,000 g, respectively. *HIF1A* was detected in 15/18 (83.3%) plasma samples in both centrifugations with the median normalized expression of 266.0 (range: 0.0–149880.3) and 1135.8 (range: 0.0–212939.9) CPM in 3,500 and 1,600 g followed by 16,000 g, respectively.

The hierarchical clustering for samples and gene expressions in different centrifugations was achieved by complete linkage of Euclidean distances (**Figure 4**). Plasma samples were not

FIGURE 2 | Gene expressions in two centrifugal forces. Normalized expressions ($\log_2$ scale) of genes were calculated as $\log_2 (CPM+1)$ to avoid log transformation for zero CPM. **(A)** 75 genes were detectable in 3,500 g; and **(B)** 86 genes were detectable in 1,600 g followed by 16,000 g. Box and whisker plots showed the interquartile range of log-transformed normalized expressions and outliers, and 25 passed filter genes were highlighted by red color.

distinctively clustered based on centrifugal forces, which indicated 3,500 and 1,600 g followed by 16,000 g centrifugations could not cause distinguished differential expression to the panel of CRC-related genes in HDs' plasma samples.

FIGURE 3 | The correlation of gene expressions in two centrifugal forces. The median normalized expressions ($\log_2$ scale) of 25 passed filter genes were correlative between 3,500 and 1,600 g followed by 16,000 g conditions ($R^2 = 0.9471$, $P < 0.0001$, Spearman rank correlation).

DISCUSSION

Plasma RNA sequencing has been used to investigate circulating cancer markers. However, majority of previous studies focused on profiling miRNA markers, because miRNAs were relative stable in human plasma (Mitchell et al., 2008; Wang et al., 2014). For plasma mRNA, we previously reported that its quantification could be affected by different centrifugal forces based on RT-qPCR results (Wong et al., 2007). Therefore, it is important to explicate centrifugal force effects based on RNA deep sequencing data before examining plasma mRNA using RNA deep sequencing technologies. Here, we provided important information on the centrifugal force effects using a custom panel of CRC-related mRNAs in plasma samples. Basically, we found that there were 2 of 108 CRC-related genes showed differential expression in plasma samples prepared by protocols with different centrifugal forces. Besides, the results of clustering and correlation for gene expression showed that two centrifugal forces used in this study were not cause distinguished differential expression to the panel of CRC-related genes in HDs' plasma samples. This is the first study to evaluate the effects of centrifugal force on plasma mRNA quantity and quality using targeted RNA sequencing.

Through comparing plasma mRNA extracted in two centrifugal forces, which were (1) 3,500 g and (2) 1,600 g followed by 16,000 g, three important findings were observed.

TABLE 3 | The differential gene expression in two centrifugal forces analyzed by DESeq2 (3,500 vs. 1,600 g followed by 16,000 g).

Gene ID	Chr	Start	Stop	Fold change	P Value	Adjusted P Value
MYC	**8q24.21**	**128748825**	**128750534**	**0.06**	**0.0001**	**0.0025**
HIF1A	**14q23.2**	**62187260**	**62188264**	**0.18**	**0.0037**	**0.0457**
CCND1	11q13.3	69456247	69457833	0.15	0.0077	0.0645
AES	19p13.3	3061188	3057697	0.26	0.0159	0.0993
CCND2	12p13.32	4385360	4387954	0.25	0.1028	0.3402
CEACAM5	19q13.2	42212680	42213625	0.39	0.1225	0.3402
CTNNB1	3p22.1	41241134	41265540	0.33	0.1154	0.3402
DAAM1	14q23.1	59730348	59757978	0.35	0.0961	0.3402
PLAUR	19q13.31	44171775	44169568	0.30	0.1138	0.3402
DVL1	1p36.33	1284308	1278111	0.38	0.1631	0.4078
CSNK2A1	20p13	485804	480536	0.44	0.2674	0.4489
CTNNBIP1	1p36.22	9937997	9932118	0.40	0.2694	0.4489
MAPK8	10q11.22	49609793	49612930	0.39	0.2324	0.4489
PTGS2	1q31.1	186648486	186648291	1.96	0.2334	0.4489
S100A4	1q21.3	153517158	153516365	0.60	0.2344	0.4489
CTSB	8p23.1	11710868	11710176	1.77	0.3313	0.5019
GSK3B	3q13.33	119812228	119721043	0.54	0.3413	0.5019
EP300	22q13.2	41489064	41513219	0.55	0.4247	0.5898
CTBP1	4p16.3	1232001	1222088	0.63	0.5014	0.6597
ID2	2p25.1	8822615	8823016	0.72	0.5326	0.6657
CSNK1A1	5q32	148930433	148929706	1.23	0.7224	0.8209
LEF1	4q25	109086283	109084819	0.79	0.7068	0.8209
CEACAM1	19q13.2	43031227	43026331	1.09	0.8773	0.9139
TCF7	5q31.1	133473825	133474683	0.89	0.8506	0.9139
RHOA	3p21.31	49412899	49405938	1.02	0.9613	0.9613
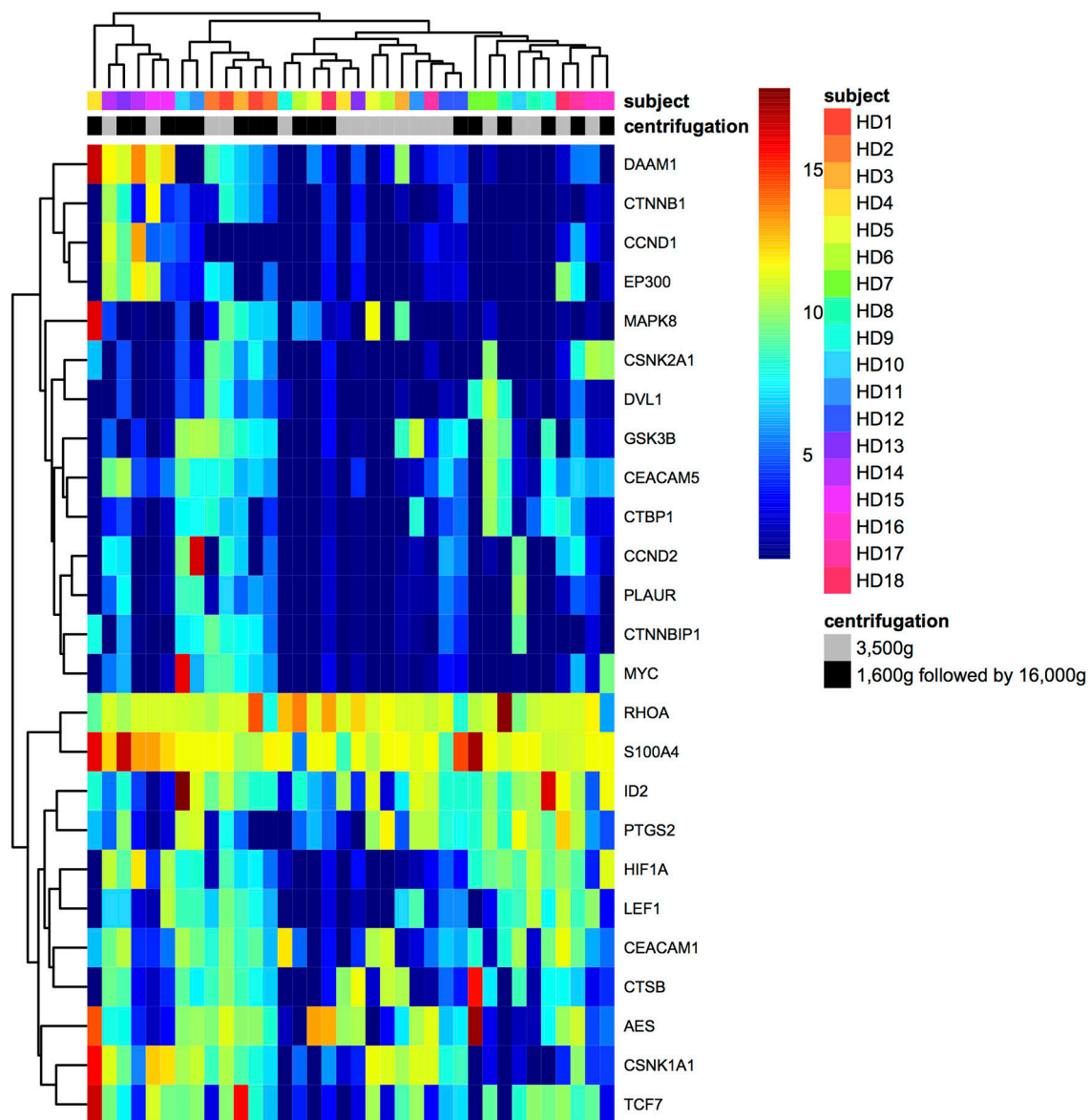
FIGURE 4 | The hierarchical clustering for samples and gene expressions.

First, RNA concentration, integrity and the percentage of longer fragments were significantly decreased in plasma samples prepared by the high centrifugal force. Compared with the cell-free plasma samples prepared by 1,600 g followed by 16,000 g (Chiu et al., 2001), plasma samples prepared by 3,500 g include more RNAs-associated particles (Ng et al., 2002). Those particles may include cell debris, extracellular vesicles, and other particles that can combine with mRNA molecules, which contribute to the increased amount of 18S and 28S rRNAs and the corresponding increase in RNA concentration. Meanwhile, the decrease of RIN and DV_{200} in plasma samples prepared by the high centrifugal force was probably due to depleted RNAs-associated particles, which account for fewer 18S and 28S rRNAs and

more fragmented RNA when plasma was subjected to the high centrifugal force, respectively. This phenomenon emphasized the effects from RNAs-associated particles to the quality and quantity of plasma total RNA in different centrifugal forces.

However, our attention is also focused on whether mRNAs could be affected by different centrifugal forces. Our second finding was that most of mRNAs in our CRC-related panel were detectable in plasma samples prepared by both centrifugations using targeted mRNA sequencing. This phenomenon is not surprising, because circulating mRNAs exist and are prevented from endogenous RNase digestion due to combination and protection from particles, for example apoptotic bodies and protein complexes (Wieczorek et al., 1985; Hasselmann et al.,

2001; Ng et al., 2002). However, a majority of detectable mRNAs had low and overdispersed expression, and the similar phenomenon was found in the recent study on plasma mRNA sequencing in pregnant women (Chim et al., 2017). Besides, we found several low-abundance mRNAs ($\leq$ 10 counts) that were only detectable in one of two centrifugations (**Table 2**). Most of these mRNAs were only detectable in one or two plasma samples, and they accounted for the situation that more genes were detected in 1,600 g followed by 16,000 g than in 3,500 g. It was difficult to determine which kind of particles protected these mRNAs from RNase in plasma samples. For transcripts only detected in 3,500 g-centrifuged plasma samples, their existence could be related to the presence of cell debris, extracellular vesicles and other particles, which cannot be removed effectively by the protocol with 3,500 g centrifugal force.

Filters are generally used in RNA sequencing data analysis to eliminate uninformative data points and increase detection power, and data filters were required to be chosen prudently to avoid losing type I error control in differential analysis (Bourgon et al., 2010; Rau et al., 2013). In this study, we used CPM filter for excluding mRNAs with the expression lower than filter criteria from subsequent differential expression analysis as stated in the methodology section, which was previously defined in edgeR (Robinson et al., 2010). Our third finding was that in differential expression analysis, 25 genes were detected for downstream analysis after filtering. Among them, *MYC* and *HIF1A* showed significantly lower expressions in 3,500 g than in 1,600 g followed by 16,000 g. This phenomenon implied that plasma mRNA of these two genes was hardly affected by centrifugal force effects and mainly preserved as cell-free format, which resulted in the increased relative expressions after normalization. *MYC* encodes the transcription factor c-Myc. It showed elevated expressions in different tumor cells, and it worked with the promoter regions of targeted genes (Lin et al., 2012). *HIF1A* encodes a transcription factor that responds to hypoxia through recruiting specific cyclin dependent kinase, stimulating RNA polymerase elongation and activating transcription of downstream genes (Galbraith et al., 2013). There was no previous study to describe how *MYC* and *HIF1A* mRNAs exist and are preserved as cell-free format in human plasma. The sequencing results of *MYC* and *HIF1A* expressions have been validated using RT-qPCR. Overall, those 25 genes had the high median of normalized expressions compared with other genes in both centrifugations, and their expressions in two different centrifugations were significantly correlated (**Figure 3**). This result demonstrated that detected gene expressions depended on the intrinsic expression levels of gene itself instead of effects from different centrifugal forces.

Moreover, plasma samples were not be clustered based on the centrifugal forces used in plasma preparation, which showed that the two centrifugal forces used in this study did not lead to distinctive difference in the concentration of CRC-related mRNAs in plasma samples (**Figure 4**).

To conclude, we achieved targeted mRNA sequencing using a custom panel of CRC-related mRNAs in plasma samples. Our sequencing results demonstrated these plasma mRNAs were not distinctly affected by two widely different centrifugal forces. However, considering mRNA from cell debris possibly interferes disease-derived plasma mRNA quantification and the efficiency of circulating markers selection for CRC in future studies, we suggest using the protocol with 1,600 g followed by 16,000 g centrifugal force in plasma preparation, which efficiently removes cell debris from plasma and is more likely to expose disease-derived mRNA information. These findings have laid down a solid foundation in plasma RNA properties upon centrifugation for downstream RNA deep sequencing. Moreover, it is helpful for researchers to standardize their protocol so that the results generated can be compared in multicenter studies with more precision and confidence. For future works, we may use transmission electron microscopy, ultra-centrifugation and other technologies to further study which components or extracellular vesicles result in differential plasma mRNA quantification caused by centrifugal force effects.

AUTHOR CONTRIBUTIONS

SW conceived and designed the experiments; VX performed the experiments; SW, VX, AY, and WC analyzed the data; SN, WL, BM, and WC gave invaluable comments on the subject recruitment and data interpretation; VX and SW wrote the paper; TA and HT processed the patient samples and technical work before library preparation.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2018.00165/full#supplementary-material

REFERENCES

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate - a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Methodol.* 57, 289–300.

Bourgon, R., Gentleman, R., and Huber, W. (2010). Independent filtering increases detection power for high-throughput experiments. *Proc. Natl. Acad. Sci. U.S.A.* 107, 9546–9551. doi: 10.1073/pnas.0914005107

Brenner, H., and Tao, S. (2013). Superior diagnostic performance of faecal immunochemical tests for haemoglobin in a head-to-head comparison with guaiac based faecal occult blood test among 2235 participants of screening colonoscopy. *Eur. J. Cancer* 49, 3049–3054. doi: 10.1016/j.ejca.2013.04.023

Brenner, H., Kloor, M., and Pox, C. P. (2014). Colorectal cancer. *Lancet* 383, 1490–1502. doi: 10.1016/S0140-6736(13)61649-9

Chim, S. S. C., Wong, K. K. W., Chung, C. Y. L., Lam, S. K. W., Kwok, J. S. L., Lai, C. Y., et al. (2017). Systematic selection of reference genes for the normalization

of circulating RNA transcripts in pregnant women based on RNA-seq data. *Int. J. Mol. Sci.* 18:1709. doi: 10.3390/ijms18081709

Chiu, R. W. K., Poon, L. L. M., Lau, T. K., Leung, T. N., Wong, E. M. C., and Lo, Y. M. D. (2001). Effects of blood-processing protocols on fetal and total DNA quantification in maternal plasma. *Clin. Chem.* 47, 1607–1613.

Danielson, K. M., Rubio, R., Abderazzaq, F., Das, S., and Wang, Y. Y. (2017). High throughput sequencing of extracellular RNA from human plasma. *PLoS ONE* 12:e0164644. doi: 10.1371/journal.pone.0164644

El-Hefnawy, T., Raja, S., Kelly, L., Bigbee, W. L., Kirkwood, J. M., Luketich, J. D., et al. (2004). Characterization of amplifiable, circulating RNA in plasma and its potential as a tool for cancer diagnostics. *Clin. Chem.* 50, 564–573. doi: 10.1373/clinchem.2003.028506

Freedman, J. E., Gerstein, M., Mick, E., Rozowsky, J., Levy, D., Kitchen, R., et al. (2016). Diverse human extracellular RNAs are widely detected in human plasma (2016). *Nat. Commun.* 7:11106. doi: 10.1038/ncomms11106

Galbraith, M. D., Allen, M. A., Bensard, C. L., Wang, X., Schwinn, M. K., Qin, B., et al. (2013). HIF1A employs CDK8-mediator to stimulate RNAPII elongation in response to hypoxia. *Cell* 153, 1327–1339. doi: 10.1016/j.cell.2013.04.048

García, J. M., Peña, C., Garcia, V., Domínguez, G., Muñoz, C., Silva, J., et al. (2007). Prognostic value of LISCH7 mRNA in plasma and tumor of colon cancer patients. *Clin. Cancer Res.* 13, 6351–6358. doi: 10.1158/1078-0432.CCR-07-0882

Hasselmann, D. O., Rappl, G., Tilgen, W., and Reinhold, U. (2001). Extracellular tyrosinase mRNA within apoptotic bodies is protected from degradation in human serum. *Clin. Chem.* 47, 1488–1489.

Landolt, L., Marti, H. P., Beisland, C., Flatberg, A., and Eikrem, O. S. (2016). RNA extraction for RNA sequencing of archival renal tissues. *Scandi. J. Clin. Lab. Investig.* 76, 426–434. doi: 10.1080/00365513.2016.1177060

Law, C. W., Chen, Y., Shi, W., and Smyth, G. K. (2014). voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* 15:R29. doi: 10.1186/gb-2014-15-2-r29

Levin, B., Lieberman, D. A., McFarland, B., Andrews, K. S., Brooks, D., Bond, J., et al. (2008). Screening and surveillance for the early detection of colorectal cancer and adenomatous polyps, 2008: a joint guideline from the American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer, and the American College of Radiology. *Gastroenterology* 134, 1570–1595. doi: 10.1053/j.gastro.2008.02.002

Lin, C. Y., Lovén, J., Rahl, P. B., Paranal, R. M., Burge, C. B., Bradner, J. E., et al. (2012). Transcriptional amplification in tumor cells with elevated c-Myc. *Cell* 151, 56–67. doi: 10.1016/j.cell.2012.08.026

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550. doi: 10.1186/s13059-014-0550-8

Mitchell, P. S., Parkin, R. K., Kroh, E. M., Fritz, B. R., Wyman, S. K., Pogosova-Agadjanyan, E. L., et al. (2008). Circulating microRNAs as stable blood-based markers for cancer detection. *Proc. Natl. Acad. Sci. U.S.A.* 105, 10513–10518. doi: 10.1073/pnas.0804549105

Ng, E. K. O., Tsui, N. B. Y., Lam, N. Y. L., Chiu, R. W. K., Yu, S. C. H., Wong, S. C. C., et al. (2002). Presence of filterable and nonfilterable mRNA in the plasma of cancer patients and healthy individuals. *Clin. Chem.* 48, 1212–1217.

Rau, A., Gallopin, M., Celeux, G., and Jaffrézic, F. (2013). Data-based filtering for replicated high-throughput transcriptome sequencing experiments. *Bioinformatics* 29, 2146–2152. doi: 10.1093/bioinformatics/btt350

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. doi: 10.1093/bioinformatics/btp616

Savelyeva, A. V., Kuligina, E. V., Bariakin, D. N., Kozlov, V. V., Ryabchikova, E. I., Richter, V. A., et al. (2017). Variety of RNAs in peripheral blood cells, plasma, and plasma fractions. *Biomed. Res. Int.* 2017:7404912. doi: 10.1155/2017/7404912

Schroeder, A., Mueller, O., Stocker, S., Salowsky, R., Leiber, M., Gassmann, M., et al. (2006). The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Mol. Biol.* 7:3. doi: 10.1186/1471-2199-7-3

Shih, C. L., Luo, J. D., Chang, J. W., Chen, T. L., Chien, Y. T., Yu, C. J., et al. (2015). Circulating messenger RNA profiling with microarray and next-generation sequencing: cross-platform comparison. *Cancer Genomics Proteomics* 12, 223–230.

Smith, R. A., Andrews, K. S., Brooks, D., Fedewa, S. A., Manassaram-Baptiste, D., Saslow, D., et al. (2017). Cancer screening in the United States, 2017: a review of current American Cancer Society guidelines and current issues in cancer screening. *CA Cancer J. Clin.* 67, 100–121. doi: 10.3322/caac.21392

Stein, U., Burock, S., Herrmann, P., Wendler, I., Niederstrasser, M., Wernecke, K. D., et al. (2011). Diagnostic and prognostic value of metastasis inducer S100A4 transcripts in plasma of colon, rectal, and gastric cancer patients. *J. Mol. Diagn.* 13, 189–198. doi: 10.1016/j.jmoldx.2010.10.002

Tsui, N. B., Jiang, P. Y., Wong, Y. F., Leung, T. Y., Chan, K. C., Chiu, R. W. K., et al. (2014). Maternal plasma RNA sequencing for genome-wide transcriptomic profiling and identification of pregnancy-associated transcripts. *Clin. Chem.* 60, 954–962. doi: 10.1373/clinchem.2014.221648

Wang, H. Y., Yan, L. X., Shao, Q., Fu, S., Zhang, Z. C., Ye, W., et al. (2014). Profiling plasma MicroRNA in nasopharyngeal carcinoma with deep sequencing. *Clin. Chem.* 60, 773–782. doi: 10.1373/clinchem.2013.214213

Widlak, M. M., Thomas, C. L., Thomas, M. G., Tomkins, C., Smith, S., O'Connell, N., et al. (2017). Diagnostic accuracy of faecal biomarkers in detecting colorectal cancer and adenoma in symptomatic patients. *Aliment. Pharmacol. Ther.* 45, 354–363. doi: 10.1111/apt.13865

Wieczorek, A. J., Rhyner, C., and Block, L. H. (1985). Isolation and characterization of an RNA proteolipid complex associated with the malignant state in humans. *Proc. Natl. Acad. Sci. U.S.A.* 82, 3455–3459.

Wong, S. C., Lo, E. S., and Cheung, M. T. (2004a). An optimised protocol for the extraction of non-viral mRNA from human plasma frozen for three years. *J. Clin. Pathol.* 57, 766–768. doi: 10.1136/jcp.2003.007880

Wong, S. C., Lo, S. F., Cheung, M. T., Ng, K. O., Tse, C. W., Lai, B. S., et al. (2004b). Quantification of plasma beta-catenin mRNA in colorectal cancer and adenoma patients. *Clin. Cancer Res.* 10, 1613–1617. doi: 10.1158/1078-0432.Ccr-1168-3

Wong, S. C., Ma, B. B., Lai, P. B., Ng, S. S., Lee, J. F., Hui, E. P., et al. (2007). The effect of centrifugation on circulating mRNA quantitation opens up a new scenario in expression profiling from patients with metastatic colorectal cancer. *Clin. Biochem.* 40, 1277–1284. doi: 10.1016/j.clinbiochem.2007.08.006

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read
for greatest visibility
and readership

FAST PUBLICATION
Around 90 days
from submission
to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative,
and constructive
peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers
acknowledged by name
on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: info@frontiersin.org | +41 21 510 17 00

REPRODUCIBILITY OF RESEARCH
Support open data
and methods to enhance
research reproducibility

DIGITAL PUBLISHING
Articles designed
for optimal readership
across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics
track visibility across
digital media

EXTENSIVE PROMOTION
Marketing
and promotion
of impactful research

LOOP RESEARCH NETWORK
Our network
increases your
article's readership