Nascent epitranscriptomics - Insights into RNA processing

Inaugural-Dissertation
To obtain the academic degree
Doctor rerum naturalium (Dr.rer.nat.)
Submitted to the Department of Biology, Chemistry and Pharmacy
of Freie Universität Berlin

by
Annita Louloupi
From Cyprus

Conducted at the Otto Warburg Laboratory,
Max Planck Institute for Molecular Genetics
Berlin, December 2018

Time period: November 2014 - December 2018
1st Reviewer: Dr. Ulf Andersson Vang Ørom
2nd Reviewer: Prof. Markus C. Wahl

Date of defense: 13/5/19

Science provides space for my imagination to be transformed into reality. In this space questions are created which I am eager to answer. The answers expand this space leaving more area for my imagination to take place. And this pattern is an infinite fractal! Enjoy your reading.

Acknowledgements

This thesis would have never been possible without the help, feedback, encouragement, inspiration and collaboration of many people. First and foremost, I would like to express my gratitude and appreciation to Ulf Andersson Vang Ørom for having me in his lab and being a good supervisor. Thank you for letting me choose my project, trusting in me and encouraging my work until the end. I offer my most sincere thanks to Evgenia Ntini, she was a great scientific partner, supervisor and friend. I learned a lot from her and I really appreciate every conversation and lab moment that we spent together. Thomas Conrad, for sharing knowledge and guiding many of my projects. His enthusiasm about science encouraged me to move on. Julia Liz, for being a wonderful bench partner and sharing her experience in many of the troubleshooting that I was facing. Dubravka Vucicevic for guiding me through my first days in the lab and giving support throughout the time I spent there. I truly appreciate your support and help until now, you are a great friend! The sequencing facility team and Nobert Mages for their patience and professional attitude towards all the alternative requests I had for the library preparations. I would also like to extend my thanks to Martin Vingron for his support until the end of the PhD. I appreciate his trust for letting me run a lab on my own in my final year. I wish to express my sincere thanks to Marcus C. Wahl for providing rigorous feedback on my PhD and Florian Heyd for welcoming me in his lab meetings where I had the chance to participate in stimulating discussions. Additionally, I would like to thank two very good friends of mine, Sara and Tomas for translating the summary of my thesis and for their support throughout my PhD years, especially in the last months. A very special thanks to my partner in crime Ivan for being next to me, encouraging me and believing in me all the time. My wonderful “black hole flat mates” for helping me thinking outside of the box, expanding the limits of my imagination. For one more time, I would like to thank Ulf, Julia and Evgenia for spending time reading and correcting my thesis. MPI has been my second home for a long time, I am very grateful to all my colleagues that were next to me until the end.

I would like to dedicate this work to my entire family and most importantly to my parents that have been and always shall be my rock.

Table of Contents

List of Figures

Author Contribution

I established, calibrated and performed: BrU-Chase Seq time points 0, 15, 30, 60 minutes, and qPCR experiments for pri-miRNA processing analysis. I developed and performed TNT-Seq, qTNT-chase Seq. In addition, I conducted qPCR experiments to calculate splicing kinetics. Furthermore, I performed functional analysis with METTL3 knock downs followed by BrU-Chase Seq and qPCR experiments to calculate splicing kinetics. I interpreted data, conceived the experiments and wrote the manuscripts[1–4]. Dr. Evgenia Ntini performed the bioinformatic and statistical analysis, interpreted data and wrote the manuscripts[2,3]. Dr. Thomas Conrad conceived experiments, interpreted data and wrote the the manuscript[2]. Dr. Ulf Ørom Anderson conceived experiments, interpreted data, supervised research, and wrote the manuscripts[2–4].

1. Introduction

The cellular RNA levels are determined by RNA synthesis[5], or transcription, RNA processing[6] and RNA decay[7,8]. RNA synthesis rates regulate how much of the RNA is transcribed in a given system whereas, the rates of RNA decay control for how long the transcripts will be present in the system. All these steps are highly regulated and some are coupled to each other, kinetically or mechanistically[9–11], providing substantial flexibility for gene expression adaptation to alterations in the environmental conditions.

The intermediate step, RNA processing, occurs co-transcriptionally[12] and involves several biochemical steps necessary for the maturation of the primary transcripts. First, the 5' methyl guanosine cap is added on the 5´ end of the precursor messenger (pre-mRNA)[13]. Second, most of the pre-mRNAs contain introns − sequences with no protein coding potential − that must be excised during mRNA maturation through a process called splicing[14,15] (Box 1). Finally, once transcription through the gene is complete, the pre-mRNA is cleaved and polyadenylated[16] to generate its mature 3´end. The biogenesis of microRNA (miRNAs) (Box 2) is also part of a distinct RNA processing pathway. The sequence of the primary microRNA (pri-miRNA) promotes the formation of the secondary hairpin loop that is further processed to a mature miRNA[17,18]. The contribution of RNA processing in gene expression as well as the importance of the processing kinetics has only recently received attention due to the lack of proper tools for sufficient analysis.

Next-generation sequencing of total RNA (RNA-seq) allows to study the outcome of RNA processing at a given time point on a whole-transcriptome level[19]. Whole cell total RNA represents mostly the mature processed transcripts that are more stable and highly abundant, while the transcripts with high RNA decay rates are not detected. Therefore, detecting gene expression changes caused by treatment, virus infection or alterations in the environmental conditions using total RNA quantification can be misleading[20]. Furthermore, the majority of the sequencing reads in RNA-seq are associated with exons derived from highly expressed and stable transcripts, whereas only a small percentage of the sequencing reads cover intronic sites that are mostly associated with long and slowly processed introns[21]. Thus, steady-state RNA-seq alone provides incomplete information regarding the kinetics of RNA splicing and processing.

To surmount the limitations of whole cell total RNA-seq, a diverse range of methods have been used to measure the kinetics of RNA processing. In eukaryotic cells the RNA synthesis and processing are confined in distinct subcellular compartments. Thus, enriching for RNA molecules that are found in the cytoplasmic, nucleoplasmic and chromatin-associated compartments prior to RNA-seq, enables the analysis of distinct steps in the lifetime of individual RNA molecules[22,23]. Previous studies using cell fractionation methods showed that pre-mRNAs and pri-miRNAs are co-purified with salt, detergent and urea-resistant pellet, while

Box 1 | RNA splicing

The beginning and the end of an intron is marked by 5′ and 3′ splice junctions (SJs) that contain consensus sequences[100]. In metazoas the SJs consensus sequences are poorly conserved with 5′ SJ GURAGU and a 3′ SJ YAG where R is a purine and Y is a pyrimidine[14]. During splicing these points undergo "cutting and sewing" in a form of two transesterification reactions that are executed by a large ribonucleoprotein complex called spliceosome[101].The SJs together with the branch point sequence (BPS; YNYURAY, N is any nucleotide) located 18-40 nucleotides upstream of the 3′ SJ are recognized in the premRNA by small nuclear ribonucleoproteins (snRNPs). U1, U2, U4/U6 and U5 snRNPs are forming the major spliceosome[102]. *In vitro* studies on pre-mRNA splicing identified a two-step transesterification reaction. The first transesterification step, known as branching, is held by a nucleophilic attack by the 2′OH group of the adenosine found in the BPS on the guanine nucleotide at 5′SJ forming a branched RNA intermediate known as intron lariat. This results in a free 5′ exon and an intron lariat-3′ exon intermediate. In the second transesterification reaction the 3′OH-group of the 3′nucleotide end of the 5′ exon attacks the phosphate group of the 3′exon producing the ligated spliced mRNA and the excised intron lariat. During splicing the spliceosome is responsible for the folding of introns to favor the splicing reactions, the correct recognition and pairing of the splicing sites. Each snRNP consist of a snRNA (two in the case of U4/U6) and a variable number of specific proteins. Additionally, the U1, U2, U4 and U5 contain seven Sm proteins that together form an extremely stable Sm core of each snRNP[103]. Splicing is comprised by three reoccurring steps. First, the reactive groups on premRNA are recognized multiple times by trans-factors RNA or proteins to ensure the precision of the splicing reaction. Second, the RNA-RNA interactions and many functionally important interactions occurring within the spliceosome are mainly weak yet, are stabilized by a combination of multiple interactions that allows flexibility and dynamic regulation of splicing. Third, during the spliceosome assembly and later catalytic activation many RNP binding partners follow a stepwise mechanism of rearrangements. The first spliceosome assembly called complex E, starts from the U1 snRNP interaction with the 5′ intron through base pairing interactions. Serine-arginine-rich (SR) proteins together with proteins of U1snRNP stabilize this step. In addition, SF1/BBP protein binds to the BPS while U2 auxiliary factor (U2AF) comprised by two subunits of 65 kDa (U2AF65) and 35 kDa (U2AF35) to the polypyrimidine track residing downstream of BPS. U2AF65 is in contact with SF1/BBP while U2AF35 binds the AG consensus of the 3′SJ. The next step is the formation of complex A with U2snRNP binding to BPS through base pairing interactions. SF1/BBP binding with BPS and U2AF65 is hand over to the U2-associated protein p14 and SF3b155 respectively[104]. After A complex formation, U4/U6 and U5 snRNPs are recruited as preassembled tri-snRNP in a reaction catalyzed by DExD/H helicase Prp28 forming complex B. The spliceosome is still catalytically inactive and to become eligible to facilitate the first transesterification reaction U1 and U4 need to be destabilized or released from the complex. U6 snRNA is free to engage U2snRNA and the premRNA[105,106]. Several U4, U6 and U5 protein partners are also released through the activation process whereas the spliceosome complex B* undergoes the first catalytic step generating C complex that contains free exon 1 and the intron–exon 2 lariat intermediate. On this stage the 5′SJ is positioned by U6snRNA through base pairing interactions with the conserved ACAGAG region of U6snRNA. The BSP is paired with the conserved sequence GUAGUA of U2snRNA. In addition, U5snRNA interacts with both of the exons and contributes to the positioning for the catalysis. Complex C undergoes additional ATP-dependent rearrangements before the second catalytic step[107]. The U2-specific SF3b14a/p14 protein and U5-specific Prp8 protein are directly at or near the spliceosome's active site and are important for the progression of the catalysis[108]. After the second catalytic step the spliceosome dissociates releasing U2, U5 and U6snRNPs to be recycled for other rounds of splicing and the mRNA is released. The release of the splice product from the spliceosome is catalyzed by the DExD/H helicase Prp22[109]

the soluble nuclear extract contains spliced and processed transcripts[24–26]. A study from our lab showed that endogenous Microprocessor activity can be determined using RNA sequencing[18]. Cellular fractionation protocol[23] was used to isolate the nuclei and chromatin associated RNA enriching for the pri-miRNA transcripts. The Microprocessor cleavage signature was identified and the MicroProcessing Index (MPI) was defined as a measure for processing efficiency. The MPI is considering the expression levels of pri-miRNA that are reflected by the sequencing reads adjacent to the pre-miRNA and the read density of the precursor region. Significant depletion of read density at the pre-miRNA region generates negative MPI and corresponds to

Box 2 | miRNA biogenesis

MicroRNAs (miRNAs) are small non-coding RNAs in 22nt length long that act in the cytoplasm to direct post-transcriptional repression. RNA pol II transcribes the primary miRNA (pri-mRNA); long transcripts that are also capped and polyadenylated. The sequence of the pri-miRNA promotes the formation of the secondary hairpin loop structure that is recognized by Microprocessor, a protein complex containing an RNaseIII enzyme Drosha and its cofactor DGCR8/pasha. Drosha recognizes the double-strand RNA-single-strand RNA junction formed at the hairpin base and serves as a ruler by measuring 11 bp from basal ssRNA-dsRNA junction[110]. Two DGR8 proteins bind the stem and apical elements to ensure efficient and accurate processing[110]. Alternative cleavage of Drosha leads to the production of isomirs that differ in length and sequence[111]. The pre-miRNA product of the first-step cleavage is a hairpin of ~70nt nucleotides. The 2-nucleotide overhang of the hairpin at the 3´ end is recognized by Exportin 5 that transports the pre-miRNA hairpin to the cytoplasm. There, Dicer binds the pre-miRNA hairpin through the 5´ phosphate, 3´ overhang and loop structure[112]. Dicer also acts as a molecular ruler that cleaves pre-miRNAs yielding a mature- miRNA duplex with another typical 2 nt 3´ overhang. Alternative cleavage by Dicer may also leads to isomirs[113]. Finally, one strand of the mature miRNA, the "guide" strand RNA is loaded into AGO whereas the passenger strand is discarded. The strand RNA with the less stable 5´ end will become the guide RNA[114].miRNAs recognize their mRNA targets via base-pair complementarity. The position 2-7nt of the mature miRNA is called the seed and is the most essential for target recognition[115]. The miRNA target sites are residing within the 3´ UTR of mRNAs that possess strong complementarity to the seed region. The AGO-miRNA binding to the target mRNA 3´UTR leads to gene silencing through translation repression and mRNA decay[116]

efficiently processed pri-miRNAs, whereas positive MPI values correspond to inefficiently processed pri-miRNAs. This study showed that the processing efficiency is highly variable among canonical pri-miRNAs and a major determinant for the expression levels of individual mature miRNAs. However, the processing efficiency of pri-miRNAs in steady-state chromatin-associated RNA is not dependent on the expression level of the pri-miRNA. Several features, such as sequence-motifs around the precursor pre-miRNA hairpin in the pri-miRNA transcript and in the hairpin loop have been shown to be involved in processing efficiency in mammals. These sequence features include the UG at the position -14 and -13 at the basal stem, the GC(-13) 13nt upstream of the pre-miRNA[18], the apical loop GUG motifs and CNNC at the positions 16 -18[17]. The variations in pri-miRNA sequence composition are responsible for fine tuning miRNA expression levels.

RNA-seq analysis from chromatin bound-transcripts showed that many of them are accumulated on chromatin and are already processed. In addition, many ncRNAs such as snoRNAs and lncRNAs can be found in chromatin due to their role in transcription regulation and processing[27]. Hence, cell fractionation is a method to enrich chromatin-associated transcripts regardless of their age. While steady state chromatin-associated RNA sequencing approach provides insights into the average processing efficiency of steady-state pri-miRNA associated to chromatin, it cannot address the dynamics of this process.

RNA labelling is another innovative biochemical method used to capture, isolate and follow the processing of the newly transcribed RNA *in vivo*. Nucleoside uridine analogs such as 4-thiouridine (4sU)[28,29], 5-ethylnyluridine (EU) and 5'bromo-uridine (BrU)[30,31] are able to incorporate in to the newly transcribed RNA in living cells. These analogues can be added to the cell media enter the cell membrane and get incorporated into the pool of nucleotide phosphates

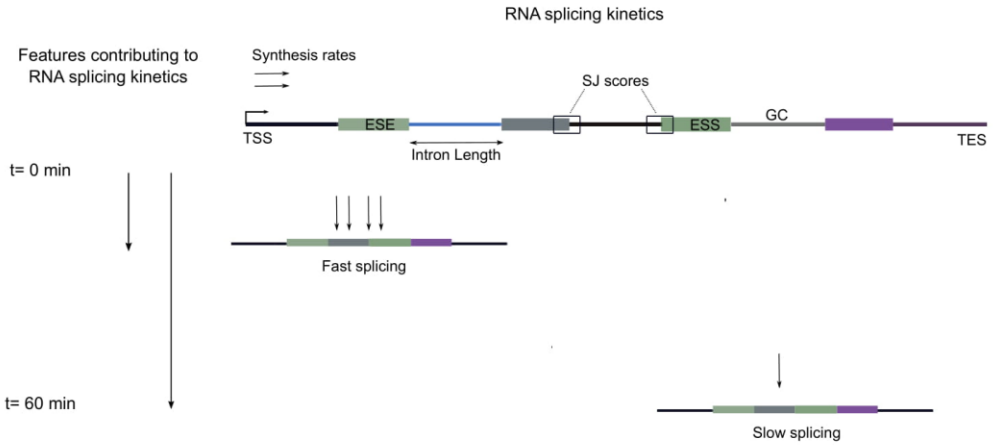


Figure 1: Schematic representation depicting features contributing to RNA splicing kinetics. Distance from TSS and TES, Intron length, SJ score, GC content, ESEs ESSs, and synthesis rates.

by means of the ribonucleoside salvage pathway[32], and further get diffused into the nucleus (without going through additional transfection techniques such as electroporation or lipofection). After short pulse incubation the cells can be lysed, total RNA is extracted and depending on the uridine analog the labelled RNA can be isolated.

4sU is rapidly taken up by cells and has minimal adverse effects on gene expression, RNA decay, protein stability and cell viability. The newly transcribed RNA (4sU-labelled RNA) is thiol-specifically biotinylated generating a disulphide bond with biotin and 4sU. The total RNA can be separated into labelled and unlabelled RNA with high purity using streptavidin-coated magnetic beads. A reducing reagent is then added to the beads to cleave the disulphide bond releasing the newly transcribed RNA. 4sU is a labelling analogue used efficiently in many studies however, it has been shown that prolonged incubation with 4sU reduced cell growth[33]. EU is similarly incorporated into the nascent RNA and the EU-labelled RNA is separated from total RNA by biotinylating of EU in copper-catalysed cycloaddition reaction (often referred to as click chemistry) followed by purification on streptavidin magnetic beads. However, prolonged incubation of cells with 4sU or EU causes inhibition of cell growth. BrU on the other hand, does not cause harmful effects compared to the rest uridine analogues[34]. BrU is incorporated into the cells, converted to BrUTP; similar to 4sUTP, and the nascent RNA is labeled by BrU. The BrU-labelled RNA can be isolated using anti-Bromodeoxyuridine (BrdU) antibody via simple immunoprecipitation. The BrU-RNAs do not cause misincorporation by reverse transcriptase and can be used as templates for further quantitative analysis and deep sequencing[35]. Additionally, in order to study the RNA splicing after the BrU-pulse the BrU can be removed with several washes and replaced with excess of uridine containing media. Then cells can be lysed at different time points and RNA splicing can be followed by analyzing the sequencing reads across the borders of exons and introns[36]. The analysis can provide information regarding RNA transcription and any RNA processing that took place during the

Box 3 | Alternative splicing

Studies using RNA-seq discovered numerous alternative isoforms of mammalian transcripts indicating that most multi-exon genes are alternatively spliced [117]. Alternative splicing allows the production of more than one unique mRNA from a single gene[117]. mRNAs generated from alternative splicing could differ in their UTRs and/or coding sequence. Events such as, exon skipping (removal of specific exons), exon inclusion (a choice between mutually exclusive exons), the choice of alternative splice sites and intron retention contribute to the final outcome of transcript diversity as well as protein complexity. Nearly 95% of human genes have been estimated to undergo alternative splicing[19] and approximately 37% [118] of the ~ 20,000 human protein-coding genes produce multiple protein isoforms. The spliceosome is capable to catalyze both constitutive and alternative splicing reactions and the decision is controlled mainly from the spliceosome assembly on the pre-mRNA. Human introns are on average ~5kb in length and contain many sequences that resemble the authentic consensus splice sites. However, the pseudo-exons are rarely spliced suggesting that the exon- intron definition is not only defined by the main consensus splicing sequences but also from other additional splicing regulatory elements (SREs)[119]. According to the location and function the SREs are classified as exonic splicing enhancers (ESEs), intronic splicing enhancers (ISEs), exonic splicing silencers (ESSs) or intronic splicing silencers (ISSs). *Trans*-acting protein splicing factors are recruited to the SREs and promote or hinder different steps during splicing reaction such as early and intermediate steps of the spliceosome assembly[120]. Constitutive spliced exons are enriched in ESEs that recruit members of the as Ser/Arg-rich (SR) protein family each of which facilitate the spliceosome assembly. More specifically, SR proteins interact with ESEs to form a barrier and to ensure the correct recognition of 5' and 3' order preventing prevents exon skipping[121]. On the other hand, various heterogeneous nuclear ribonucleoproteins (hnRNPs) typically recognize ESSs and inhibit splicing. In both cases U2 and U1 snRNPs are affected from SRs or hnRNPs presence. The activity of SRs and hnRNPs is context specific since SRs could also inhibit splicing when are bound to ISS and similarly hnRNPs can promote splicing when are bound to ISEs[122]. In addition, these two protein families exhibit antagonistic effects on splice site recognition with a well characterized example SR protein ASF/SF2 and hnRNP A1[123]

labelling period. Moreover, incorporating the chasing time points in to the analysis, allows to study the fate of the nascent RNA over time and RNA processing[30].

The combination of ultra-short and progressive 4sU labelling time points starting from 5 minutes to 60 minutes allowed to study the kinetics of RNA processing, in particular RNA splicing and alternative splicing (Box 3)[36,37]. In particular, this method enabled the identification of distinct intron classes with different splicing kinetics. Interestingly, each intron class of distinct splicing kinetics was characterised by specific intron length, gene length, splice site strength, distance from transcription start site, distance from transcription end site, presence of ESEs and/or ESSs, synthesis rates, and GC content (**Figure 1**). The fast-spliced introns were mostly characterised by low GC content, strong splice sites score, higher distance from TSS and TES, high levels of ESEs and low levels of ESSs in their upstream and downstream exons[36]. The slow processed introns show higher synthesis rates, have longer sequence and have a higher probability to be alternative spliced[36].

The dynamic interactions of the splicing regulatory factors with the spliceosome ultimately regulate the decision of constitutive or alternative splicing[38,39]. However, since splicing of most introns occurs co-transcriptionally[9–11,40], factors that regulate transcription also affect constitutive and alternative splicing[41]. The regulation of splicing and the final splicing product derives from a combination of RNA Binding Proteins (RBPs) presence, recruitment to their underlying sequence and the interactive environment. Direct competition or cooperative recruitments of RBPs as well as modulations of the RNA local secondary structure are means of interactions that govern the splicing regulation. Altogether, the information in the pre-mRNA sequence and how it is interpreted by RBPs is ruled by the so-called "splicing code". Deciphering the splicing code

is a necessary step to tackle the biological complexity of the splicing regulation and the splicing-caused diseases. This PhD thesis conducted extended research to investigate the role of N^6-Methyladenosine (m6A), an RNA modification involved in many RNA processes[42–48], in splicing kinetics.

Insights into m6A RNA methylation

All major biological macromolecules including DNA, RNA, proteins and lipids are being subjected to enzyme-catalyzed covalent modifications, following their synthesis, that are important for their structure, function and stability[49]. Modified nucleotides in RNA have been studied since the 1960s with N^6-Methyladenosine (m^6A)[1], 5-methylcytidine (m^5C)[2], and pseudouridine (Ψ)[3] to be the first identified in rRNAs, tRNAs and snRNAs. To date, more than 100 distinct RNA modifications have been identified within each of the four RNA nucleosides[50]. The multitude of RNA modifications constitute the 'Epitranscriptome'[45] and our understanding of this additional regulatory layer of biology resting between DNA and proteins, is still in its infancy.

m6A is a methylation on the nitrogen 6 position of adenosine and is the most abundant internal mRNA modification[43,45,51]. It is deposited by methyltransferase complex with methyltransferase-like 3 (METTL3)[52] as the main catalytical domain and can be removed by demethylases such as the Fat mass and obesity-associated (FTO)[53] and AlkB homolog 5 (ALKBH5)[54] (Box 4). Since the early 70s m6A has been estimated to be one per 700-800 nucleotides and is deposited preferentially within G (m^6A) C or A (m^6A) C sequence motifs[55]. However, only a portion of these motifs in the transcriptome bear detectable methylation[55,56]. The development of N^6-methyladenosine–sequencing (see m6A-Seq) encouraged transcriptome wide m6A-mapping studies on the mRNA and/or total RNA level extracted from different organisms and tissues. These studies located m6A in at least 8,000 transcripts with a consistent m6A peak distribution on coding sequences in long exons, in the 5´UTRs and near stop codons[43,51] .

m6A can alter the RNA structure modulating the accessibility of RNA binding factors to their RNA sequence[57]. Although m6A does not change the hydrogen-bonding donors and acceptors on the base, the energetics of A•U pair are affected. N^6-Methyladenosine in solution exists in two isomeric forms, in syn-orientation and in anti-orientation with a preference in syn orientation (**Figure 2**). In the syn-orientation the methyl group is located on the N1 side of the base and disrupts Watson-Crick base pairing. In the anti-orientation form the methyl group is on the N7 side of the base and the Watson-Crick base pairing is unhindered. However, the anti-conformation elicits an energetic penalty due to the steric clash between the methyl group and N7 that leads to destabilization of m^6A•U pairs in comparison to A•U pairs (**Figure 2**).

Thermodynamic measurements revealed that m6A found in duplexes causes a destabilization of 0.5–1.7 kcal/mol. However, m^6A in unpaired positions stacks more strongly than A stabilizing the single stranded locations[58]. In agreement, cellular RNAs show decreased base pairing around m6A sites as well as structural transition from paired to unpaired in the proximity of m^6A modifications

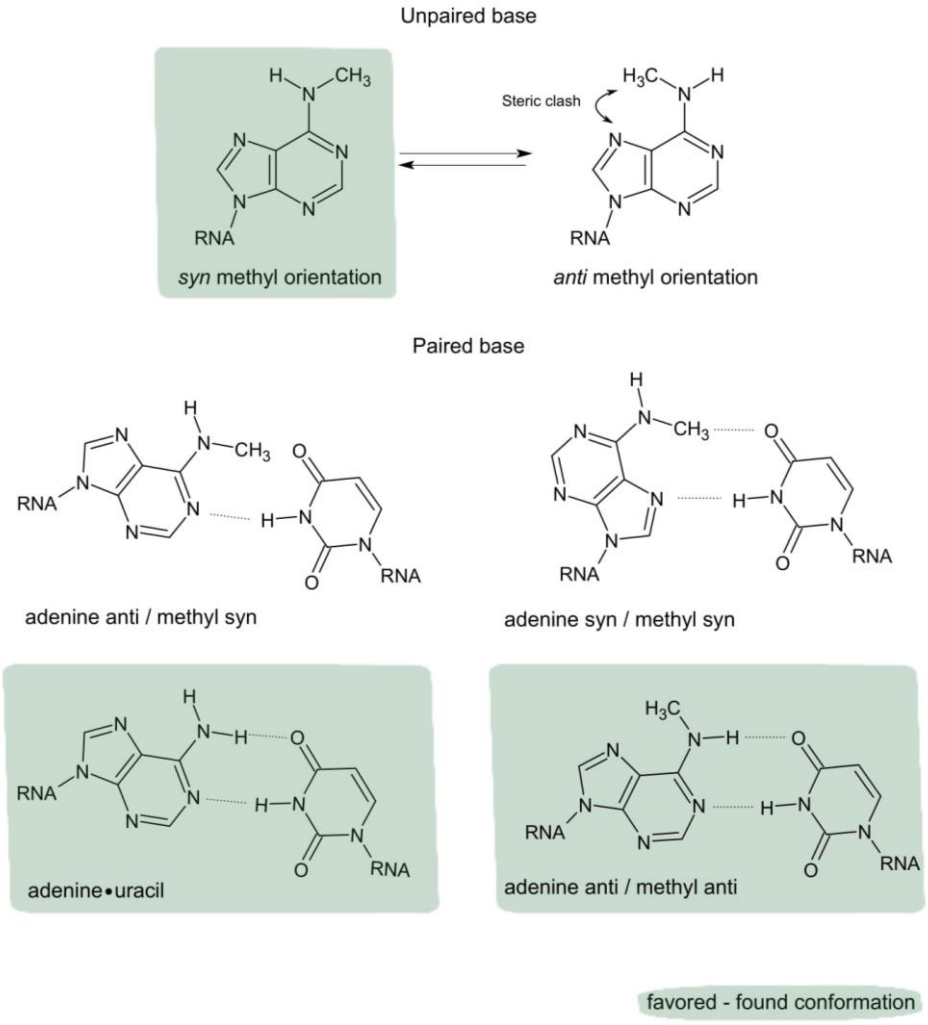
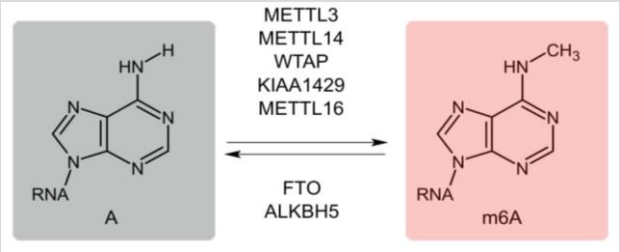


Figure 2: Conformation and structure of m6A in RNA.

(A) *Syn* methyl orientation is favored over *anti* when the base is unpaired due to the unfavoured steric clash between the methyl group and N7. (B) Multiple pairing configurations are possible for m^6A paired opposite U. The anti/anti structure has been found in an RNA duplex; the methyl group is spring-loaded into the high-energy *anti* conformation, trapped there by pairing with U and surrounding duplex structure. The green highlighted structure is the favoured and found structure [58]

Box 4 | m6A writers and erasers

Early studies discovered a large protein complex responsible for m6A methylation that is comprised by three components of 30, 200, 875 kDa[124]. Further analysis revealed that the 200-kDa component contains the S-adenosylmethionine-binding site on a 70-kDa subunit (The catalytic component was detected by its ability to crosslink to [^{3}H]-SAM). S-adenosynlmethionine (SAM), the enzymatic cofactor that participates in most of the methylation reactions in the cell. The 70-kDa subunit called MT-A70 or METTL3 is able to catalyze the transfer of a methyl group from the donor substrate S-adenosyl methionine (SAM) to the adenine nucleobases in acceptor RNA substrates[124]. Phylogenetic analysis of METTL3 family identified that METTL14 shares 43% identity with METTL3 and is highly conserved in mammals[52]. Crystal structures of METTL3/METTL14 as the ability to methylate RNA substrates[125]. The interaction of METTL14 with METTL3 complex revealed that METTL14 mainly functions as a structural scaffold and only METTL3 results in substantial higher methyltransferase activity than METTL3 alone. Proteome wide interaction analysis showed METTL14 and METTL3 form a complex that is regulated by the association of a WTAP, a pre-mRNA splicing regulator[52]. WTAP does not have a methyltransferase activity however, its presence is necessary for the localization of METTL3 and METTL14 in the nuclear speckles. Depletion of WTAP leads to substantial loss of m6A formation on mRNA[51]. Immunoprecipitation-mass spectrometry (IP-MS) was used to identify KIAA1429 (also known as vir-like m6A methyltransferase associated or VIRMA) an additional interaction partner of methyltransferase complex that is important to mediate the full activity of the complex[51]. Depletion of KIAA1429 causes loss of m6A in human cell lines as well as Drosophila S2R+ cells[51,71].RBM15 and its paralog RBM15B are two additional components of the methyltransferase complex. Both of the proteins interact with METTL3 in a WTAP dependent manner and knockdown experiments showed a significant reduction of m6A on mRNA level. These findings were also supported by studies in *Drosophila* where a homolog of RBM15, Spenito (Nito) is shown to be important for the m6A deposition in flies[71,72]. METTL16 is the only methyltransferase identified to methylate mRNAs encoding the SAM-synthetase MAT2A at 3´UTR. When the SAM cellular levels are depleted MAT2A mRNA is stabilized. METTL16 has been suggested to control mRNA stability of MAT2A and the SAM depended MAT2A expression[96]. The vertebrate conserved hairpin (hp1) loop structure found in the 3´UTR of MAT2A mRNA is the substrate for the METTL16. Under high concentrations of SAM METTL16 is able to methylate MAT2A mRNA leading to its degradation. Fine tuning MAT2A expression through m6A modification may contribute to the flexibility and precision of SAM-depended processes. According to the human protein atlas all the five proteins are expressed in all the human tissues[126]. An important finding was the identification of m6A demethylases that could be endogenously expressed. Fat-mass and obesity-associated protein (FTO) belongs to the non-heme Fe(II)- and α-KG-dependent dioxygenase AlkB family proteins and can catalyse the Fe(II)- and 2OG-dependent demethylation of 3-methylthymine in single-stranded DNA, with concomitant production of succinate, formaldehyde, and carbon dioxide[127]. Subsequent studies showed that FTO could also demethylate 3-methyluracil (3mU) in single stranded RNAs that can be found on rRNAs[128]. Further *in vitro* studies identified FTO demethylation activity can also be applied on m6A residues in RNA. In fact, FTO oxidize N^6-methyladenosine to generate N^6-hydroxymethyladenosine as an intermediate modification, and N^6-formyladenosine as a further oxidized product oxidize. These two chemical products have ~3h half-life times in aqueous solution under physiological conditions and are found in human cells and mouse tissues[129]. Overexpression of FTO in Hela cells reduced m6A levels on polyA enriched RNA to ~18% whereas FTO knockdown increased m6A levels to ~23%[53]. FTO is localized in the nucleus and cytoplasm and according to FTO-iCLIP data FTO targets are mostly intronic regions with no specificity for the m6A sequence motif [42]. FTO is highly expressed in the brain and widely expressed in all adult and fetal tissues[130] . Soon after FTO discovery a second mammalian demethylase was identified from the same protein family, called α‑ketoglutarate‑dependent dioxygenase alkB homologue 5 (ALKBH5). Both *in vitro and in vivo* ALKBH5 has the ability to catalyze the removal of the m6A modification on nuclear RNA[29]. ALKBH5 demethylates preferentially single stranded substrates DNA or RNA and similar to FTO, is able to recognize m6A modification on non-consensus sites [30]. ALKBH5 is enriched in the nucleus, is predominately expressed in testis and has relatively low expression levels in other tissues[54]. Knock down experiments increased m6A levels on mRNA by ~9% whereas ~50-fold overexpression of ALKBH5 lead to ~29% of m6A level reduction. The subtle changes on m6A level, similarly to FTO, indicate that both demethylases target specific m6A residues on mRNA level. m6A has been characterized as a "conformational marker" that regulates the conformational changes of the modified RNA that serves as the substrate for the demethylases. Thus m6A its self in different sequence environment profoundly impacts the interaction with m6A –recognizing proteins such as ALKBH5 and FTO[131].



m6A readers mediating m6A function in RNA processing

The protein factors that mediate the outcomes of m6A on RNAs are important to understand the biological role of m6A. m6A modification can be "read" in different forms such as by direct recognition from a binding pocket of YTH-domain proteins [59] or indirect recognition through a structural change that is caused by the modification[57]. In addition, hydrophobic modifications such as m6A induce solvation penalty in water and their interaction to hydrophobic protein side-chain residues can reduce the solvation penalty (**Figure 3**). SFSF and HNRNP proteins have no modification-specific binding domain however; they selectively bind adjacent to m6A residues[60].

The YTH domain is comprised ~ 145 amino acids that fold in to a distinct module with an aromatic cage of three tryptophan residues engaging the methyl group of m6A. The aromatic rings of the two of the three Trps that bind two m6A are almost parallel to each other where m6A adenine moiety is sandwiched by them. Mutations of the aromatic cage residues reduce significantly the binding of YTH-domain proteins to m6A RNA probes *in vitro*[60].

Initial *in-vitro* pulldown experiments with m6A modified RNA baits identified the YTHDF2 and YTHDF3 as m6A binding proteins[61]. YTH domain proteins can be divided based on their sequence to three major classes DC1; DC2 and the DF family comprising nearly three identical paralogs DF1, DF2 and DF3. The three DF proteins and DC2 are primarily cytoplasmic while DC1 is located in the nucleus[62,63]. DF proteins have similar structure and contain two domains; a C-terminally located YTH domain and a large low-complexity domain containing Glutamine (Q), Asparagine (N) and Proline (P) residues. Low–complexity domains are regions of protein sequences with biased amino acid composition, flexible enough to bind several different targets[64]. On the other hand, DC1 contains an YTH domain and multiple

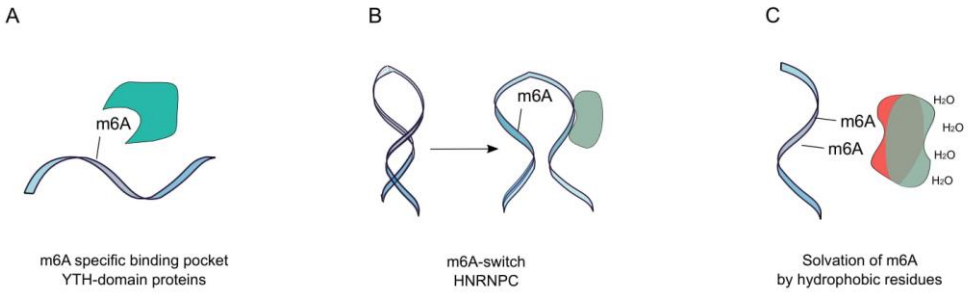


Figure 3: m6A interaction with m6A readers

(A) m6A readers such as YTH-domain proteins bind to the m6A directly and selectively lead the transcripts to distinct RNA pathways. (B) m6A alters the secondary structure of mRNA, exposing or masking potential RNA-binding motifs. (C) m6A introduces a hydrophobic chemical group. The hydrophobic amino acid chains or low complexity regions of proteins associate with m6A and reduce the solvation penalty.

nuclear localization elements and an SH2 domain. YTHDC2 is an RNA-induced ATPase with a 3′ to 5′ RNA helicase activity and it is found in the nucleus and in the cytoplasm[65].

In the cytosol, YTHDF1 and YTHDF3 act cooperatively to promote ribosome loading of their m6A methylated mRNA targets[66,67]. More specifically YTHDF1 interacts with ribosomal subunits and translation initiation factors complex 3 (elF3) promoting translations of YTHDF1 bound transcripts[66]. On the other hand, YTHDF2 decreases the mRNA stability of its m6A mRNA targets and promotes degradation by recruiting CCR4-NOT deadenylase complex[68]. Over all the YTHDFs proteins have more than 50% common m6A mRNA targets and can interact with each other with an RNA independent manner. YTHDF3 seems to interact first with m6A mRNA targets and functions as a hub for partitioning its common targets with YTHDF1 and YTHDF2. YTHDF3 might contribute to RNA specificity while YTHDF1 and YTHDF2 contribute to RNA binding affinity. In this way the targeted methylated transcripts have enhanced translation efficiency and a fast degradation rate in a highly regulated environment controlled by the YTHDF proteins[67].

The nuclear m6A reader YTHDC1 promotes exon inclusion in targeted mRNAs through the recruitment of the splicing factor SRSF3. In the meantime, SRSF10, another splicing factor that promotes the opposite effect, exon exclusion, is competed away by the YTHDC1/SRSF3 complex. Additionally, YTHDC1 promotes SRSF3 localization to the nuclear speckles while repelling SRSF10. Thus, m6A and the stoichiometry of YTHDC1/SRSF3 and SRSF10 regulate the final outcome of their target mRNAs[69]. Independently from its role in splicing, YTHDC1 mediates export of methylated mRNA from nucleus to the cytoplasm through SRSF3 and the nuclear mRNA export receptor (NXF1). SRSF3 has a lack of biochemical selectivity for m6A in vitro yet, its interaction with YTHDC1 and NXF1 couple's m6A selectivity to in nuclear export[70]. In Drosophila, YT521-B the homologue of YTHDC1, regulates many splicing events of targeted methylated transcripts. More specifically, YT521-B binds to m6A residues in the Sex lethal mRNA and regulates sex determination by repressing the inclusion of the male-specific exon of the transcript[71]. YTHDC1 also, promotes the epigenetic silencing effects of XIST, a non-coding RNA important for the silencing of genes on one X chromosome in female cells[72]. XIST m6A residues are essential for the gene silencing effect on X chromosome through YTHDC1 that interacts with multiple epigenetic regulators[72]. YTHDC2 m6A reader is essential for male and female fertility in mice for maintaining transcripts essential for early meiotic progression[37]. Furthermore, YTHDC2 enhances the translation efficiency and mRNA degradation of its targeted methylated transcripts [42]. In addition, YTHDC2 was found to recruit the 5′ to 3′ exoribonuclease XRN1 suggesting a mechanism to destabilize its mRNA targets[73].

Heterogeneous nuclear ribonucleoprotein A2/B1 (HNRNPA2B1) is a nuclear reader of m6A and binds to RGm6AC sites on nuclear RNAs *in vivo* and *in vitro*[74]. HNRNPA2B1 regulates alternative splicing of exons in a set of transcripts in a similar manner as METTL3[74].

In addition, HNRNPA2B1 is responsible for the nuclear processing of a subset of miRNAs whose maturation is dependent on METTL3 activity[74]. More specifically, depletion of HNRNPA2B1 or METTL3 leads to accumulation of pri-miRNA transcripts in the nucleus. HNRNPA2B1 recognizes and binds methylated pri-miRNA transcripts interacts with DGCR8 by protein-protein interaction and facilitates their processing[74].

Heterogeneous nuclear ribonucleoprotein C (HNRNPC) protein is a member of a large ubiquitously expressed family that bind nascent RNA transcripts affecting pre-mRNA stability, splicing, nuclear export and translation. HNRNPC preferentially binds single stranded U-tracks (five or more contiguous uridines). M6A modification found within 50 residues away of HNRNPC binding site can alter the local RNA structure enhancing the accessibility of HNRNPC binding. The mechanism where m6A regulates the binding of RNA binding proteins via structural changes is called m6A-switch[57]. Most of the m6A switches for HNRNPC protein are found within intron regions of coding and non-coding RNAs. In coding RNAs the m6A-switches are located within long exons near stop codons and in the 3'UTR, following m6A known topology of mRNA. The gene expression of more than five thousand genes is co-regulated by the methyltransferase complex and HNRNPC. Furthermore, the splicing of more than 200 genes with multiple m6A-switched sites is co regulated from the methyltransferase complex and HNRNPC[57].

Heterogeneous nuclear ribonucleoprotein G (HNRNPG) is another m6A reader that uses a low-complexity region to recognize an RNA binding motif exposed by m6A modification[75]. The RNA binding motif of HNRNPG is a purine rich region that overlaps with the m6A consensus sequence. More than thirteen thousand m6A sites are bound by HNRNPG with more than fifty per cent residing in introns. Knock down experiments of HNRNPG and METTL3 or METTL14 led to similar gene expression changes as well as alternative splicing events[75].

m^{6}A detection and mapping methods

The first method used to detect base modifications was a combination of chemical or enzymatic digestion, radiolabeling and thin-layer chromatography[76]. Later, this method evolved into liquid chromatography coupled to mass spectrometry (LC-MS/MS), that nowadays is used to identify RNA modifications and to determine their global abundance[77]. LC-MS/MS is an extremely sensitive and accurate tool able to determine the amount of RNA modifications in the range of femtomole. To obtain such high sensitivity it is necessary to relate the sample signals with absolute amount of the modification such as spike-in measurements, or internal standards. Dynamic multiple reaction monitoring (DMRM) and neutral loss scan (NLS) are the two mainly modes used in LC-MS/MS[77]. DMRM method can be performed by a triple quadrupole mass analyzer where multiple precursor ions are chosen by adjusted mass-to-charge ratios and

fragmented in the collision cell. The ion products are then detected and the corresponding modified nucleosides are rendered in a mass spectrum. DMRM has reduced chemical background and it allows the quantification of different nucleosides at the same time. NLS allows the analysis of novel modified nucleosides however, the sensitivity is lower than DMRM method. The main disadvantage of LC-MS/MS is the inability to provide sequence-level information since the RNA is digested into single nucleotides or nucleosides which are then subjected for further analysis.

To gain more information regarding the location of m6A sites, m^6A-Seq is applied using RNA purified from various tissues or cell lines[43,45]. Total RNA or mRNA is chemically fragmented into ~100nt length and used as input for immunoprecipitation using an anti-m6A affinity purified antibody. The methylated RNA fragments are enriched over the randomly fragmented transcriptome and subjected to high-throughput sequencing to determine the identity of the methylated fragments. The randomly fragmented input is also sequenced and the m6A sites are identified using a peak-detection algorithm that finds the relative enrichment of methylated fragments over input. m^6A-Seq identifies 200nt regions in the transcriptome were m6A site could be found. The peak summit is assumed to be in close proximity with the actual methylated site[43]. The limiting factor of m^6A-Seq is the resolution since it does not provide single nucleotide information. In order to increase the resolution level and at the same time validate the m6A sites *de novo* motif search is applied. m6A motif should have a positional enrichment near the m6A peak summits over negative peak regions which are randomly generated genomic intervals. Nevertheless, single nucleotide resolution methods can be used for higher resolution on the m6A sites.

Site-specific cleavage and radioactive-labeling followed by ligation-assisted extraction and thin-layer chromatography (SCARLET) determines the exact location of the m6A residue and its modification fraction in single nucleotide resolution for a specific mRNA or long ncRNA[78]. The method starts from total RNA or mRNA isolation and RNase H cleavage guided by a complementary 2′-OMe/2′-H chimeric oligonucleotide leading to site specific cleavage 5′ to the candidate site. The point of digestion is radiolabeled with $_{32}P$ and the $_{32}P$-labeled RNA fragment is splint-ligated to a 116-nucleotide single-stranded DNA oligonucleotide using DNA ligase[78]. RNAseTI/A treatment is used to completely digest all the RNA, whereas the $_{32}P$-labeled candidate fragment remains intact. The DNA-$_{32}P$-labeled candidate is analysed on a denaturing gel, excised and digested with nuclease P1 to produce mononucleotides containing 5′ phosphate. Finally, thin-layer chromatography is applied to determine the m6A modification status.

m6A individual-nucleotide-resolution crosslinking and immunoprecipitation (miCLIP) is a method that enables single nucleotide resolution mapping of m6A on a transcriptome scale[79]. miCLIP is based on the property of certain m6A antibodies to crosslinked to m6A containing RNA fragments and cause a reverse transcriptase-induced mutation or truncation. miCLIP

follows the same steps as m6A Seq with an intermediated step of UV crosslink after the addition of the m6A antibody. After the elution a proteinase K treatment is followed and the m6A RNA fragments are subjected for sequencing. Finally, the mutation sites or C to T transitions at position +1 relative to A detected by sequencing are used to find where m6A residues are located in the transcriptome after bioinformatic analysis[79].

All the above-mentioned methods to identify m6A lack stoichiometry information and are insensitive to the proportion of methylated transcript or sites. m6A-level and isoform-characterization sequencing (m6A-LAIC-seq[80]) is a quantitative approach for assessing the methylation status on a whole-transcriptome scale. Total RNA or mRNA is subjected to immunoprecipitation without prior fragmentation. The eluted methylated transcripts together with the supernatant non-methylated transcripts are separated. External RNA Controls Consortium (ERCC) spike ins are added to each fraction before sequencing. The m6A levels of a gene can be quantified using the ERCC-normalizes RNA abundances in different pools.

Aim 1 | Studying pri-miRNA processing kinetics

To address the dynamics of the pri-miRNA processing we established a pulse-chase approach that allow us to capture an earlier RNA processing stage than chromatin-associated RNA and followed its processing during a 1-h chase. With this we were able to study the pri-miRNAs processing kinetics in high resolution through time. Partial results of the presented work have been published in[3,4]

Aim 2 | Nascent m6A role in splicing kinetics

In order to expand "the splicing code" and study the role of m6A in splicing kinetics we needed to use a technic that could identify m6A deposition on nascent RNA. All the existing technics mentioned above are not able to capture nascent RNA or to identify the nascent m6A deposition. Thus, two novel techniques have been developed to capture the nascent m6A deposition and at the same time follow the RNA processing kinetics. Partial results of the presented work have been published in[2].

2. Materials and Methods

Lab Equipment

Cell culture dish 100 x 20mm, 150 x 25mm
Laminar flow hood
Cell culture incubator
Refrigerated microcentrifuge
1.5 ml Protein LoBind tubes
1.5 ml DNA Lobind tubes
5ml Eppendorf tubes
Refrigerated Eppendorf 1.5 ml shaker
Agilent 2100 Bioanalyzer
Qybit ® Fluorometer
Dynal magnetic separation rack
Tube Rotator
Illumina HiSeq 2500

Chemicals	**Supplier, catalog number**
Opti-MEM™ I Reduced Serum Medium	Thermo Fisher Scientific, 31985062
(-)-5-Bromouridine	Santa Cruz Biotechnology, CAS 957-75-5
Uridine	Sigma-Aldrich, U3750-25G
anti-BrU purified	BD Pharmingen, 555627
TRIzol Reagent	Thermo Fisher Scientific, 5596-01
GlycoBlue™	Thermo Fisher Scientific, AM9515
Tris-HCl, pH 7.5	Alfa Aesal, N25B905
Igepal CA-6300	Sigma-Aldrich, I8896-100ML
BSA, acetylated (20 mg/mL)	Thermo Fisher Scientific, AM2614
Ethanol	Merck, 1.009.832.500
3M sodium acetate (pH 5.2)	Thermo Fisher Scientific, AM9740
SUPERase• In™ RNase Inhibitor	Thermo Fisher Scientific, AM2696
PBS, pH 7.4	Thermo Fisher Scientific, 10010056
anti-m6A polyclonal antibody	Synaptic Systems, 202 003
N6-methyl-ATP	TriLink, N-1013
$C_{11}H_{15}N_5O_7PNa$	Sigma-Aldrich, M2780
Chroroform	Merck, 1.02445.2500

Isoamyl alcohol	Serva, 39557.02
Isopropanol	Merck, 1.09634.2500
DMEM,	Thermo Fisher Scientific, 11965092
Fetal Bovine Serum	Thermo Fisher Scientific, 10500064
Protein G/A Dynabeads	Thermo Fisher Scientific, 10004D/ 10002D
Distilled water	Thermo Fisher Scientific, 10977049
EDTA pH 8,0 (0,5 M)	AppliChem GmbH, A3145,1000
NaCl	Sigma-Aldrich, S3014-1KG
$ZnCl_2$	Sigma-Aldrich, 96468
anti-METTL3 Polyclonal antibody	Proteintech, 15073-1-AP
ERCC RNA Spike-In Mix	Thermo Fisher Scientific, 4456740

Kits	**Supplier, catalog number**
GoScript Reverse Transcriptase	Promega, A5003
GoTaq qPCR Master mix	Promega, A6001
Qubit® RNA HS Assay Kit	Thermo Fisher Scientific, Q32852
Agilent RNA 6000 Pico kit	Agilent, 5067-1513
HiPerFect Transfection Reagent	Qiagen, 301704
T7 RNA Polymerase	Thermo Fisher Scientific, 18033019
TruSeq Stranded mRNA Library Preparation	Illumina, 20020594
SuperSignal West DURA Extended Duration	Thermo Fisher Scientific, 10445345

SiRNAs	**Integrated Device Technology, Inc. (IDT)**
NAME siRNA Target	**Sequence 5'→3'**
Mettl3-1	5'-ACUGCUCUUUCCUUAAUA
	5'-AAACAUGUAUUAAGGAAA
Mettl3-2	5'-CCAACAGUCCACUAAGGA
	5'-CUGUUGUUCCUUAGUGGA
Mettl3-3	5'-AGGCAAGGAACAAUCCAU
	5'-UUCAACAAUGGAUUGUUC
Mettl3-4	5'-AGCCAAGGAACAAUCCAU
	5'-UUCAACAAUGGAUUGUUC
Control	NCI IDT controls

Methods:

"Cell culture and BrU-chase Seq.

HEK293 cells were cultured in DMEM growth-medium supplemented with 10% Fetal Bovine Serum (FBS) under normal growth conditions (37°C and 5% CO_2). The day before bromouridine (BrU) labelling ~2.0 x 10^6 cells were seeded in 100 mm plates with 10ml media, one plate for each time point. Cells were 70-80% confluent before the addition bromouridine (BrU). BrU (-5-Bromouridine cat.no. CAS 957-75-5 Santa Cruz Biotechnology) was added to a final concentration of 2 mM to the media and cells were incubated at normal growth conditions for 15 minutes (pulse). Cells were washed thrice in PBS and either collected directly (0 minutes chase time point) or chased in conditional media supplemented with 20 mM uridine (Sigma cat.no U3750-25G) for 15, 30 and 60 minutes. RNA was purified using TRIzol following manufacturer's instructions. In this step we followed the protocol of (Paulsen et al., 2013) with some modifications. 35ul of anti of anti-mouse IgG magnetic Dynabeads (Invitrogen) were transferred to a 1.5ml microfuge Protein Low binding tube and washed 3 times with BrU-IP 1X buffer (0.1% BSA in RNAse free PBS). After the final wash, the beads were resuspended with BrU-IP 1X buffer supplemented with SUPERase• In™ RNase Inhibitor 1:2000 together with BrdU antibody (5μg of antibody per 100 μg RNA). Antibody-beads mixture was incubated for 1hour at room temperature with gentle rotation following 3 washes with 1X BrU-IP. 150 μg RNA was used for each BrU-IP and heated up for 4 minutes at 65°C prior to IP. The same amount of unlabeled total RNA was used as a negative control. 5X BrU-IP (0.5% BSA 5X PBS supplemented with SUPERase• In™ RNase Inhibitor 1:2000) was added to the RNA to have a final concertation of 1X. RNA-antibody-beads mixture was incubated for 90 minutes at room temperature with gentle rotation in a final volume of 800 μl. The beads were washed thrice with 800 μl 1X BrU-IP at room temperature. For all wash steps, with the exception of the elution step, the beads were washed for 5 min rotating then placed on a magnetic rack and the wash buffers were discarded. At the last wash the Protein low binding tubes were replaced with DNA LoBind tubes. For elution 200 μl of Elution buffer (0.1% BSA and 25 mM bromouridine in PBS) were added directly on the beads and the tubes were incubated for 60 minutes with continuous shaking (1100 rpm) at 4 °C. The supernatant (eluate w/o beads) was transferred to a new tube and RNA was precipitated by adding 1/10 volumes of 3M sodium acetate (pH 5.2) and 3-4 volumes of 100% ethanol. RNA was allowed to precipitate at −80 °C overnight. RNA pellet was washed twice with 75% ethanol and resuspended in RNase-free water. RNA quality was analyzed using Agilent 2100 Bioanalyzer with an Agilent RNA 6000 Pico kit according to the manufacturer's instructions.

Filtering and annotation of miRNAs

microRNAs used in the analysis were filtered to include only high-confidence microRNAs showing absence of other non-coding RNA species in the region; folding of the pre-miRNA into a hairpin; and homogenous reads in small-RNA sequencing data for both the 5' and 3' mature miRNA. We required conservation of the hairpin structure in orthologous members of the gene family for conserved microRNAs (as defined in mirBase) including mouse or other mammals and conservation of the seed in more than 50 per cent of the orthologous genes. The miRNAs used is from (4) and includes 229 miRNAs; 138 classified as broadly conserved; 52 classified as weakly conserved; and 39 as non-conserved. We determined the exact Microprocessor cleavage sites using the annotation of the 5p and 3p miRNA strands from miRBase and mapped them onto the sequence of the pre-miRNA.

TNT-seq

For one TNT-seq sample ~ 25 150mm plates were used for BrU labelling. RNA was metabolically labelled with BrU for 15 minutes and RNA was isolated as described above. RNA concentration was adjusted to 2μg/μl with nuclease free water. 18 μl of RNA was added to thin-walled 200μl PCR tube following addition of 2 μl of 10X fragmentation mixture (containing 800 μl of RNase-free water, 100 μl of 1M Tris-HCl pH 7.4 and 100 μl 1M of $ZnCl_2$). After vortex and quick spinning, the tubes were incubated in 94 °C for 3.5 minutes in a preheated thermal cycler block with the heated lid closed. Tubes were quickly removed from the thermocycler and placed on ice following addition of 2 μl of 0.5 M EDTA. After vortex and quick spin the RNA was collected in a tube to continue with for RNA precipitation using 1/10 volumes of 3 M sodium acetate (pH 5.2), 3-4 volumes of 100% ethanol. RNA was allowed to precipitate at −80 °C overnight. The following day tubes were centrifuged at full speed for 30 minutes at 4 °C. RNA pellet was washed twice with 75% ethanol and resuspended in 400-500 μl of RNase-free water. Validation of post fragmentation size (~100 nt) distribution was analyzed using Agilent 2,100 Bioanalyzer with an Agilent RNA 6,000 Pico kit according to the manufacturer's instructions. 400 μg-600 μg fragmented BrU labeled total RNA was used for each BrU-IP. BrU-RNA isolation was performed as described above. The BrU-IP recovery was approximately 0.09-0.16% of input. 4.5 μg of BrU fragmented RNA was used as input for the m6A immunoprecipitation. 35 μl of Dynabeads® Protein A (Invitrogen) per sample was transferred to a 1.5 ml microfuge Protein LoBind tube and washed 3 times with 1X m6A-IP (500 mM NaCl, 0.1% NP-40, 10 mM Tris-HCl, pH 7.5). After final wash the beads were resuspend in 800 μl 1X m6A-IP buffer supplemented with SUPERase• In™ RNase Inhibitor 1:1000. 1μg of affinity purified anti-m6A polyclonal antibody (Synaptic Systems) per 2.5 μg BrU-RNA was added to the beads and incubated for 60 minutes at room temperature with gentle rotation. As a negative control, we used Dynabeads® Protein A magnetic beads bound to an irrelevant IgG. Beads were washed 3 times with m6A-IP 1X buffer for 5 min

on the rotator. 5 μg of BrU Fragmented RNA was used as input. RNA was heated up for 4 minutes at 65°C. 5X m6A-IP buffer (50 mM Tris-HCl, 750 mM NaCl and 0.5% (vol/vol) Igepal CA-6300 supplemented with SUPERase• In™ RNase Inhibitor) was added to have the RNA in 1X m6A-IP buffer. RNA-antibody-beads mixture was incubated for 2h at 4°C with gentle rotation in a final volume of 0.8ml in Protein low binding tubes. Three washing steps followed using m6A-IP 1X buffer (1st and 2nd wash) and high salt m6A-IP buffer (500 mM NaCl, 0.1% Igepal CA-6,300, 10 mM Tris-HCl, pH 7.5) (3rd wash). For all wash steps, with the exception of the elution step, the beads were washed for 5 min then placed on a magnet and the wash buffers were discarded. At the last wash the Protein low binding tubes were replaced with DNA LoBind tubes. For elution 80 μl of Elution buffer (1X m6A-IP buffer + 6.7 mM m6A nucleotides) were added directly on the beads and the tubes were incubated for 1hour with continuous shaking (1100rpm) at 4 °C. The beads were spin down and the supernatant was transferred to a clean tube. After the second round of elution the eluted RNA was precipitated using ethanol precipitation as described above. RNA pellet was resuspended in 15 μl RNase-free water and using Qubit® RNA HS Assay Kit we measured the RNA concentration following manufacturer's instructions.

qTNTchase-seq, qPCR, RT-PCR.

RNA was metabolically labelled with BrU for 15 minutes and chased for 30 minutes as described above. RNA was purified using TRIzol following manufacturer's instructions. 200 ug total BrU labeled RNA was used as Input for the BrU-RNA isolation. After the elution step (200 μl of 0.1% BSA and 25mM bromouridine in PBS) we added 50ul of 5X m6A-IP buffer. 4 μg (1μg ab per 500ng RNA) m6A ab were coupled to 40ul Dynabeads® Protein A as described above, resuspended in 550 μl m6A-IP 1X buffer and added to the RNA mixture. RNA-antibody-beads mixture was incubated for 60 minutes at room temperature with gentle rotation. The supernatant was kept and RNA was isolated with TRIzol. The beads were washed 3 times for 5 minutes at RT (twice with low salt m6A-IP 1X buffer and last wash high salt m6A-IP 1X buffer). We eluted the RNA captured by m6A ab by competition as described in TNT-Seq section. cDNA synthesis was performed using the same amount of RNA (10-20 ng) from all fractions (Input BrU-RNA 0 min, Input BrU-RNA 30 minutes chase, Supernatant m6A-neg 0h, Supernatant m6A-neg 30 min chase, IP m6A-positive 0 min, IP m6A-positive 30 min chase). RT-PCR was performed using Q5 Hot Start High-Fidelity DNA Polymerase New England Biolabs with initial denaturation 98 °C 30s, then 32 cycles of 98 °C 10 s, 58 °C 20 s and 72 °C 55 s and final extension 72 °C 2 minutes. PCR products were resolved on agarose gel. Spike-in controls were in vitro transcribed using T7 RNA Polymerase Invitrogen following manufactures instructions. For the methylated transcripts N6-methyl-ATP (tri-link) was used in a ratio 4:1 to ATP in the *in vitro* transcription reaction. GFP and Luciferase sequences were used as template for the RNA transcription. For each qTNTchase-seq sample before m6A IP, *in vitro*–transcribed transcripts with and without m^6A modification

were mixed into the samples as spike-in controls at the indicated percentage of m6A-modified to m6A-unmodified transcript (Molinie et al., 2016). For all samples after BrU-IP but before m6A-IP we added $2.5x10^7$ copies from each spike included: 0% GFP, and 20% luciferase. For the sequencing; Post- qTNTchase seq 1 µl of 1:2000 dilution of the universal ERCC spike-in control A (Invitrogen) was added to each fraction.

SiRNA transfection

HEK293 cells were transfected with four different siRNAs targeting METTL3 transcript (see Supplemetary table 1) using HiPerFect Transfection Reagent from QIAGEN. In brief, reverse transfection was performed using 1 x 10^6 cells for a single 100mm plate. Cells were seeded in a final 4ml final volume of media without antibiotics. 12ul of transfection reagent together with siRNAs (25nM final concertation) were incubated at room temperature in 1ml Opti-MEM™ I Reduced Serum Media after mixing for 20 minutes. The transfection complexes were added dropwise into the plate. 16 hours after transfection 5 ml of cell culture media were added to each plate. 24 hours after the transfection we performed a second round of transfection using the same amount of transfection reagent and siRNAs as the first round. 40 hours after the first transfection 5 ml of cell culture media were added to each plate. We analyzed knock down efficiency with western blot (anti-METTL3 Polyclonal antibody, protein tech Catalog.number: 15073-1-AP) and continued with BrU-Chase Seq 72 hours after the first round of transfection. The experiment was performed in duplicates.

Transcript m6A level and splicing index

The m6A level per transcript from the qTNTchase-seq experiment were calculated as described in (Molinie et al., 2016). The ratio of the RNA abundance for each transcript between the eluate and the supernatant was represented by the ratio of the overlapping strand-specific RNA read counts normalized to the ratio of the reads of the ERCC RNAs. We used the log2-transformed read counts of ERCC RNAs to fit a linear regression model, computing the eluate ERCC reads as a function of the supernatant ERCC reads with a coefficient of 1(not shown). The log2 ratio between ERCC eluate counts and supernatant counts was indicated by the intercept of the regression formula. Only the ERCC RNAs with at least 100 read counts were used in this pipeline. M6A level = E/(E+S*2^*intercept*). Eluate read counts (*E*), supernatant read counts (*S*), and the intercept of ERCC regression (*intercept*). We assessed the splicing efficiency per transcript as the ratio of the overlapping strand-specific split reads (extracted by using bedtools coverage –s –F 1.0) to all (split + non-split) reads covering the transcript.

Quantitative real-time PCR

RNA was reverse transcribed using the Goscript reverse transcription Promega A500. cDNA was quantified on an 7900HT Fast real time PCR system (Applied Biosystems) using the Go Taq qPCR Master Mix Promega (A6001). The PCR was carried out using a standard protocol with melting curve. Primers for unspliced RNA transcripts were design to span exon – intron 5' splice junction and exon – exon boundaries for spliced RNA transcripts. Splicing efficiency (SE) was determined by the ration of $2^{-CT_{spliced}} / (2^{-CT_{spliced}} + 2^{-CT_{unspliced}})$ for each timepoint. SED was determined by the ration of SED = 1/ ((1- SE0 min) * (1- SE60 min))

RNA sequencing and data analysis

For the BrU-Chase Seq, the library preparation was performed using the TrueSeq Stranded Total RNA Kit (Illumina). Sequencing was performed on an Illumina HiSeq 2500 instrument to obtain around 200M reads per sample. For the TNT-Seq, 100 ng of Input BrU-labeled fragmented RNA and 100 ng of TNT-IP eluate RNA were subjected to library preparation following the TruSeq Stranded mRNA Library Preparation Kit instructions with some modifications. The protocol started from the first strand synthesis step and 3X Clean-NA-Beads beads volume was used for the buffer exchange to include shorter RNA fragments. Mapping of strand-specific reads to GRC37 genome assembly (hg19) was done using STAR (Dobin et al., 2013) and only uniquely mapped reads were kept for further downstream analyses. To extract read coverage per nucleotide position across the genome the strand-specific bed files were sorted by chromosome and start coordinate and converted into wig files with bedtools genomecov using –scale to normalize for library size. To assess the genome-wide correlation of the m6A signal from replicates, the ratio of normalized read counts per nucleotide position of IP to Eluate, rendering the m6A signal, was converted to bigWig using wigToBigWig (UCSC) and then bigWigCorrelate (UCSC) was used. To extract the m6A signal per nucleotide position in given intervals, the depth at each nucleotide position of the examined intervals (e.g. within +/- 500 bp windows around anchor points) was extracted using bedtools coverage –d –s from the m6A Input and the respective m6A IP, and then the ratio m6A IP/Input multiplied by (total number of mapped reads in the Input/ total number of mapped reads in the IP) was calculated. Then the average m6A signal was extracted at each nucleotide position from all examined entries.

m6A peak calling

We called m6A peaks based on a previously published pipeline (Ke et al., 2015; Ke et al., 2017). We first divided the genome into 20 bp non-overlapping bins with bedtools windowMaker and extracted the strand-specific read coverage from m6A Input and IP for all bins using bedtools coverageBed –s. Fisher's exact test p-value was extracted from the matrix (bin Input read counts, bin IP read counts, total number of mapped reads in the Input, total number of mapped reads in

the IP) and adjusted by the Benjamini and Hochberg method to determine the false discovery rate (FDR). Only windows with a p-adjusted < 0.05 in all three replicates and fold enrichment (score) minimum four in at least two out of the three replicates were kept as significant. Adjacent significant bins were merged using bedtools mergeBed into broader peaks (finally 95 % of the peaks were in the range 20-100 nt long). In the case of broad peaks, the peak summit is the midpoint of the 20 nt window with the maximum score, or the midpoint of the interval of merged adjacent bins sharing same maximum score within the same peak. In a few cases, a broad peak was assigned more than one summits if it contained non-adjacent windows sharing the same maximum score, finally yielding 58102 m6A peaks and 58311 peak summits. Custom scripts were written in awk programming language.

***De novo* motif search**

De novo motif search was run using HOMER (Heinz et al., 2010) within +/-150 nt intervals around the peak summit of 5651 best scoring exonic m6A peaks (minimum fold enrichment 20) and the same number of top best intronic peaks. Control sequences were generated from the respective input sequences with the scrambleFasta.pl script. Then, *de novo* motif search was run with 'findsMotifs.pl input_sequences.fa fasta –basic –rna –len 6,7,8 –fasta scrambled_sequences'. The results were inspected in terms of enrichment, significance and the presence of common consensus sequences, with the four motifs displayed in Figure S1B being the most represented. Those were used to scan the input sequences for the presence of match occurrences using the 'dna-pattern' search tool from the RSAT suite (Medina-Rivera et al., 2015) with parameters 'search given strand only, prevent overlapping matches, origin-start, return flanking nucleotide positions 2'. Motif search was also performed in the same number of random genomic intervals as a control, generated with bedtools (–length 300 –number 5651). The matches were aligned and the logo was generated with WebLogo3 (Crooks et al., 2004) .

Splicing kinetics and predictive models

To assess splicing efficiency we extracted the splicing index value Θ ψ as in (Mukherjee et al., 2017) . Θ equals to the ratio of the split reads mapping to the 5' and 3' SJ of an intron divided to the sum of split plus non-split reads (schematic representation in **Figure 7A**). The Θ value (representing Splicing Efficiency, SE) was extracted from all pulse-chase time points, for 13,532 introns with at least five reads coverage in both 5' and 3' SJ, and used in k-means clustering with k = 3 to call three groups of distinct splicing efficiency (fast, medium and slow) (**Figure 7E**). The Splicing Efficiency Dynamics metric was calculated as SED = 1/ ((1.001- Θ 0 min) * (1.001- Θ 60 min)) (plotted in the *log* scale for the three groups in **Figure 7D**). To assess constitutive versus alternative splicing we extracted the ψ value as in (Mukherjee et al., 2017) . ψ is the ratio of constitutive split reads assigned to a given intron's 5' and 3' SJ to all split reads (i.e. split reads

from the given intron 5' SJ to any downstream 3'SJ and from the intron's 3' SJ to any upstream 5' SJ, as depicted in **Figure 7A**). Therefore, ψ is in the range 0 to 1 with 1 meaning 100 % constitutive splicing. We then used the ψ value extracted from the pulse-chase time point 60 min (closer to steady-state) to perform k-means clustering with k = 2 and define two clusters of introns, constitutive (n = 11836, minimum ψ 0.5294) and alternative (n = 1696, maximum ψ 0.5278). In the case of introns classified as alternative spliced (ψ < 0.5278) upstream or downstream exon skipping takes place. The following features were used in logistic and linear regression models to predict splicing efficiency kinetics and alternative versus constitutive splicing: The 5' and 3' splice site underlying sequence scores extracted using MaxEntScan (http://genes.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq.html); distance of the 5' SJ to the annotated transcript first start site (TSS) and of the 3' SJ to the last end site (TES); expression calculated as coverage (reads per kb) from the m6A Input RNA-seq (15 min BrU pulse) for the whole transcript interval where the intron belongs to; intron length; intron overall m6A signal extracted as the strand-specific m6A IP read coverage divided to m6A Input read coverage, normalized by (total number of mapped m6A Input reads * total number of mapped m6A IP reads); m6A signal calculated the same way at the 5' SJ 100 nt exonic boundary, 5' SJ 100 nt intronic boundary, 3' SJ 100 nt exonic boundary and 3' SJ 100 nt intronic boundary.

To predict fast versus slow or alternative versus constitutive splicing, logistic regression was performed with R function glm (family = binomial) (all parameters apart from the sequence scores were first log scale transformed and all were then standardized). To evaluate the fitting of the model and assess discrimination, the Receiver Operating Characteristic Curve (ROC) and the area under the curve (AUC) were calculated with the R package ROCR (Sing et al., 2005). Linear regression to predict splicing efficiency using the continuous value ψ (in the range 0 to 1) was performed with R function lm().

CLIP data analysis

We used CLIP data for SRF3 and SRSF10 from[69](Xiao et al., 2016)(GEO GSE71096). To calculate the relative SRSF10/SRSF3 binding per nucleotide position, we used the ModeScore column from the GEO submitted PARalyzer output file, which is the score of the highest signal divided to the sum value (signal+backround) and ranges from 0.5 to 1. We first extracted the coverage for each SRSF per nucleotide position in the +/500 nt window around 5' or 3' SJ, or per bin for the length-binned introns (introns with length 1000-10000 nt, binned into 1000 non-overlapping windows), by using bedtools coverage –s –d. Nucleotide positions with overlapping CLIP binding sites were assigned the cluster's score (ModeScore column) whereas nucleotide positions with no CLIP data overlap were assigned a pseudo-score 0.1. We then computed the ratio SRFS10/SRSF3 per nucleotide position or per bin of all analyzed loci and the metagene

analysis extracting the average ratio SRFS10/SRSF3 per nucleotide position or per bin was run separately for each of the subgroups fast/medium/slow or constitutive/alternative.

Primer Sequences	**Integrated Device Technology, Inc. (IDT)**
NAME	**Sequence**
CDKN1B unspliced Forward	AATAAGGAAGCGACCTGCAA
CDKN1B unspliced Reverse	atacgccgaaaagcaagcta
CDKN1B spliced Forward	AATAAGGAAGCGACCTGCAA
CDKN1B spliced Reverse	GGGGAACCGTCTGAAACAT
LMAN2 unspliced Forward	GTGACTGCGGATATAACTGACG
LMAN2 unspliced Reverse	ctcgccctcactcttcactc
LMAN2 spliced Forward	GTGACTGCGGATATAACTGACG
LMAN2 spliced Reverse	ATAGTGCTGCCCTGGAAGTC
NASP unspliced Forward	CATGGAGTCCACAGCCACT
NASP unspliced Reverse	tgccttaagctttccacagtc
NASP spliced Forward	CATGGAGTCCACAGCCACT
NASP spliced Reverse	GCAGATGTAGAAGGAGCAGGA
ARF4 unspliced Forward	CCTCCCTCTTCTCCCGACT
ARF4 unspliced Reverse	attgtggagaccctgccttt
ARF4 spliced Forward	CCTCCCTCTTCTCCCGACT
ARF4 spliced Reverse	TTGTCTTGCCAGCAGCATC
C8orf33 Forward	TAAGAAGAAAACGCGGAACAGG
C8orf33 Reverse	GGTGGGTTTCTGCCTCTTGA
MSN unspliced Forward	TCAAGAAGCTGAAGAGGCCA
MSN unspliced Reverse	agttcccataatcccagccc
MSN spliced Reverse	CTGTCAGCTCTGCCATTTCC
SPTBN1 unspliced Forward	CTGGATGAGCGAGCAGGAG
SPTBN1 unspliced Reverse	aagtgtgcccagggtttgaa
SPTBN1 spliced Revers	GCATAGTCCTCCACAGCTTGT
NOL7 unspliced Forward	TCCTGAAGGAGAAGAGGAAGC
NOL7 unspliced Forward	aattctccctgagccgagtt
NOL7 spliced Forward	AACGCTCCTGAAGGAGAAGA
NOL7 spliced Reverse	TCCAAAATAGTGTCTGGAAGGA
Pri-let7a1/d/f1 Forward	GCATTTGTTTATGGCCTGGA
Pri-let7a1/d/f1 Forward	CACCCCCATCCAGTGTACTT
Pri-let7a1 unprocessed F	ACACCCACCACTGGGAGATA

Pri-let7a1 unprocessed R	GCCTGGATGCAGACTTTTCT
Pri-miR221/2 Forward	AGCAAAGAGAACACCAATCCTGT
Pri--miR221/2 Reverse	GTTCCAAGCTTTCCTCCCATGAT
Pri-mir221 unprocessed F	ACTTGCAAGCTGAACATCCA
Pri-mir221 unprocessed R	TGCCTAACGAACACAGAAATCT

” published in ref.[2–4].

3. Results

3.1. Set-up of nascent RNA pulse-chase sequencing[3,4]

RNA-seq provides an average view of RNA in the cell or in the respective purified subcellular compartment, reflecting a mixture of RNA of different age compared to the time of

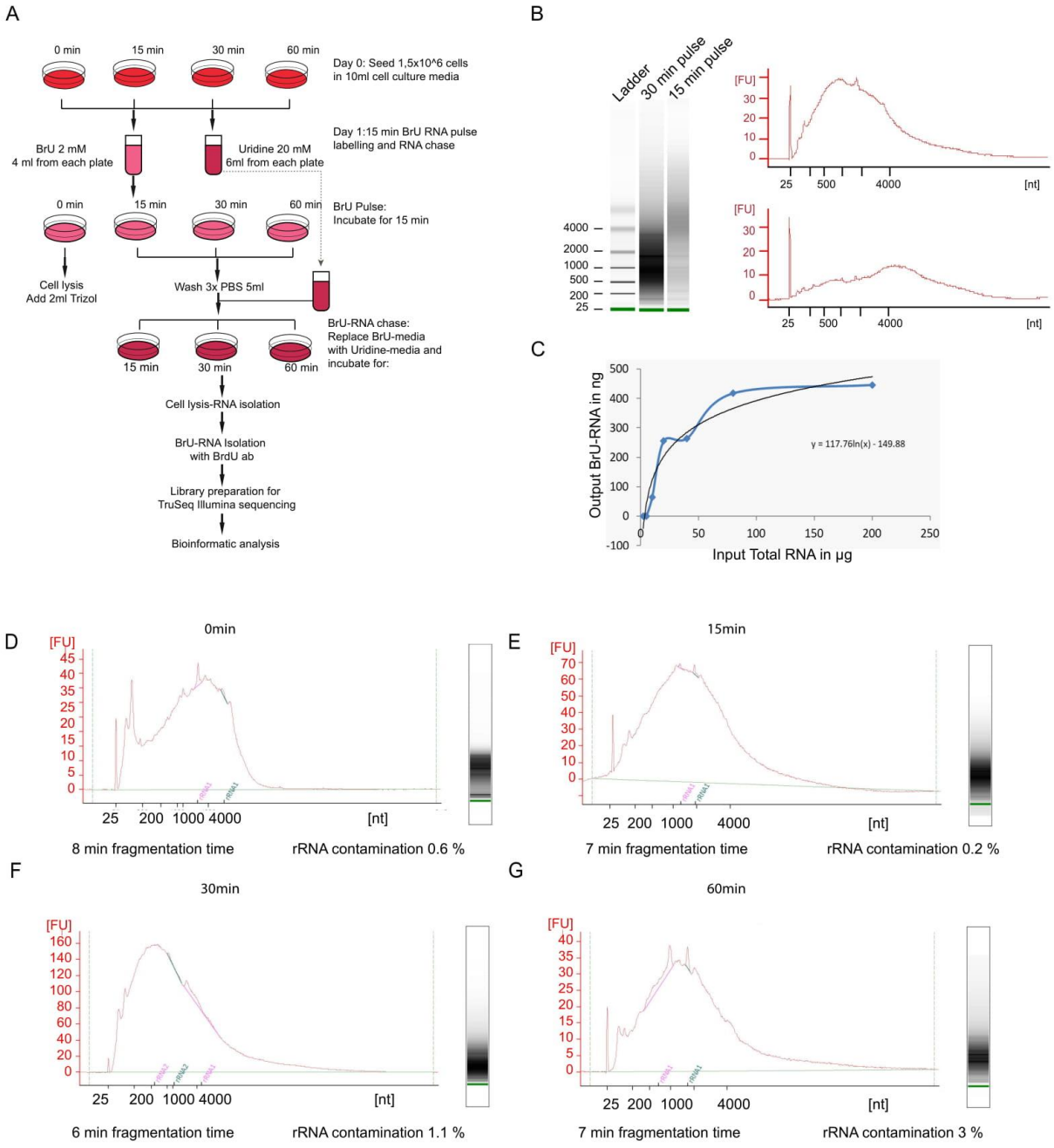


Figure 4: Set-up of nascent RNA pulse-chase sequencing

(A) Workflow for RNA pulse-labeling with BrU and chase to follow nascent RNA. (B) Calibration curve for 5ug BrdU antibody. (C) Bioanalyser Results from BrU-RNA labelled for 30 minutes and 15 minutes. (D-G) Bioanalyzer results for BrU-Chase Seq time points 0, 15,30, 60 minutes. The indicated RNA fragmentation time is according to TruSeq Stranded mRNA library prep kit protocol

transcription. To follow RNA from transcription through processing, nascent RNA can be obtained by labelling actively transcribed RNA with a pulse of a modified nucleotide analogue that allows for subsequent purification.

For this study we chose BrU analog to label nascent RNA not only because it does not affect cell growth, but also because it is a simple and a cost-effective technic; ideally suited for *in vivo* studies[30]. We established BrU-Chase Seq (**Figure 4A**) from ref.[81] with some modifications that improved the quality of the BrU labelled RNA. First, we reduced the BrU pulse time to 15 minutes instead of 30 minutes. The 15 minutes labelling time was the fittest and earliest time point from a series of calibrating timepoints that gave enough nascent RNA as output for further experiments. We compared the Bioanalyser results from isolated BrU labelled RNA of 30 minutes versus 15 minutes pulse and observed that the length distribution of the two timepoints differ. More specifically, the majority of the nascent transcripts isolated from the 15 minutes pulse are longer than 4000 nt whereas the majority of transcripts from 30 minutes pulse are approximately 2000 nt (**Figure 4B**). This implicates that within 30 minutes pulse, the isolated RNA is enriched in many transcripts that undergo processing. Second, we calibrated the amount of Total labelled RNA (Input) that is needed to saturate 5 μg of BrdU antibody and found that ~50 μg Total labelled RNA should be used per 5 μg BrdU antibody (**Figure 4C**). Third, we eluted the BrU labelled RNA bound to the antibody via BrU competition rather than heating for 10 minutes in 80 °C. In this way, we isolate only the RNA molecules that are bound to the BrdU-coupled beads and at the same time, we avoid any degradation of the RNA caused by high temperatures. Fourth, the eluted BrU labelled RNA from all time points was subjected to Bioanalyser analysis for quality check. In addition, according to the Bioanalyzer results, RNA was fragmented during different incubation times at 94 °C (**Figure 4D-G**) prior to the library preparation. Finally, we subjected BrU labelled nascent RNA from all chased time points obtained from HEK293 cells to next-generation sequencing using an Illumina Hi-Seq 2500 to obtain around 200 M reads per sample. In our data we did not observe rRNA enrichment in the eluted BrU-labelled RNA (**Figure 4D-G**). Given that the rRNA synthesis and ribosome biogenesis are regulated to meet cells growth rate and proliferation and that the average doubling time of HEK293 is 24h, it is unlikely that within 15 minutes BrU pulse we could capture rRNA transcription[82]. (See Appendix: Metabolic Pulse-Chase RNA Labeling for pri-miRNA Processing Dynamics Chapter published in[3] for further details regarding the protocol).

3.2. *In vivo* profiles of pri-miRNA processing dynamics from whole cells[3]

To further extend our previous findings[18] on steady-state pri-miRNA processing efficiency we used nascent RNA obtained after a short (15 min) BrU pulse and subsequent chase for 0, 15, 30 and 60 minutes (Samples 15, 30, 45 and 75 min after BrU, respectively) to follow the processing kinetics. We have previously reported a specific profile for steady-state chromatin-

associated RNA around the site of pre-miRNA processing within the pri-miRNA transcript[18] (**Figure 5A**). Following the nascent RNA during the chase we could record the time-course of processing of pri-miRNAs in HEK293 cells. Interestingly, for the 38 pri-miRNAs (see appendix) where we observed a pronounced profile, we noticed different processing kinetics across pri-miRNA transcripts, and within polycistronic pri-miRNAs. The profiles for miR-221, let-7a-1 are depicted in **Figure 5B** to represent the intermediate and fast processing kinetics. The profile of miR-21 was chosen as a representative for not a pronounced processing profile. The processing efficiencies were analysed also using quantitative PCR (qPCR) of individual pri-miRNAs, as shown for miR-221, let-7a-1 and miR-21 in **Figure 5C**. As described in ref.[18], we used primers spanning the processing site and primers amplifying the total of pri-miRNA transcript (processed and unprocessed) to determine the relative amounts of unprocessed pri-miRNAs.

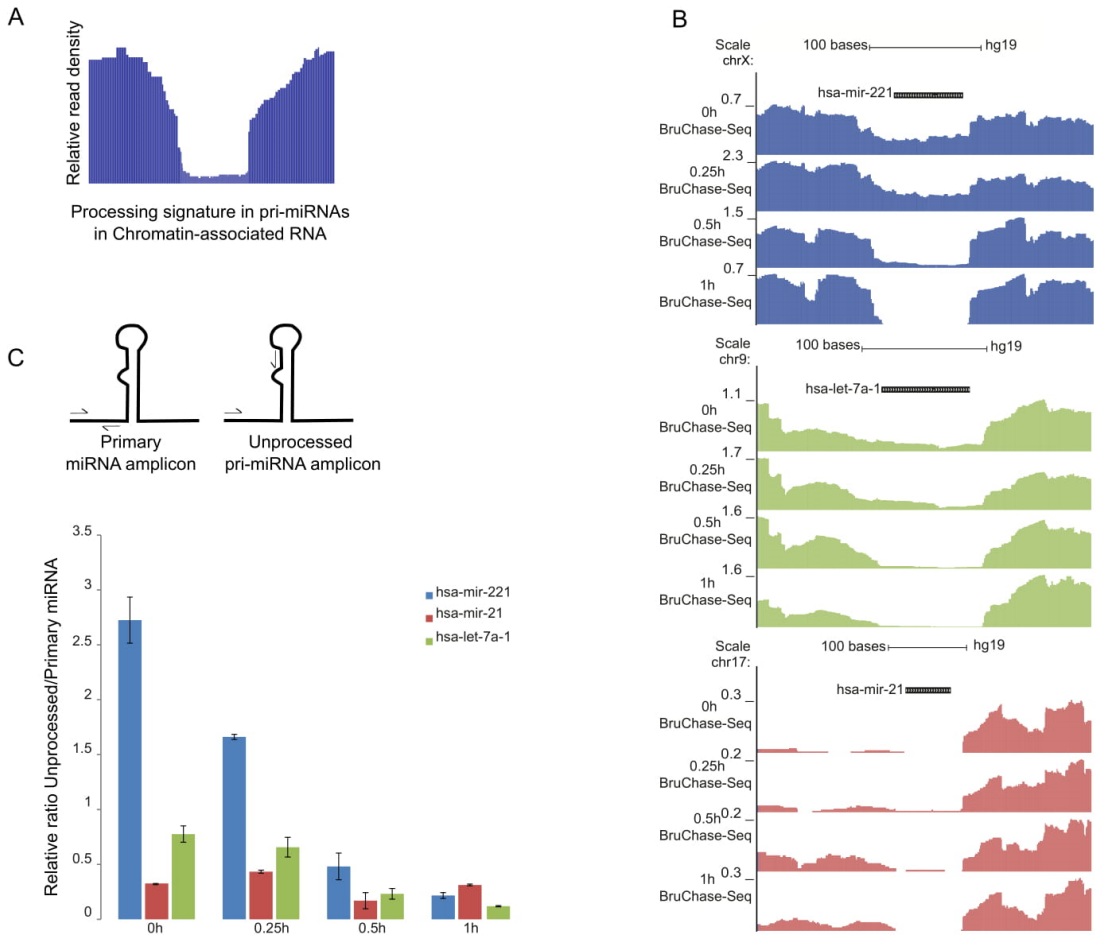


Figure 5: In vivo profiles of pri-miRNAs processing dynamics from whole cells

(A) Concept of processing signature in pri-miRNAs. Processing extent is calculated as the read-density in the pre-miRNA region compared to the flanking regions. Processing efficiency is calculated as (1 – processing extent). (B) Processing signatures in RNA-sequencing data from nascent RNA in pulse-chase experiment for pri-miR-221, pri-let-7a-1 and pri-miR-21. (C) Quantification by PCR of unprocessed/primary pri-miRNA for examples shown in c from two independent experiments.

3.3. Differential processing within polycistronic pri-miRNAs[3]

Many miRNAs are expressed from polycistronic pri-miRNAs and these miRNAs often belong to the same families and thus predicted to target the same mRNAs for translation regulation and target RNA degradation [83,84]. Let-7a/f and miR-221/222 are prominent polycistronic pri-miRNAs with crucial roles in the development of cancer and cell cycle[85]. We found differential processing kinetics within both these polycistronic pri-miRNAs. The miR-221/222 pri-miRNA is a 25kb long transcript (**Figure 6A**) encoding miR-221 and miR-222. While the two miRNAs are adjacent to each other, they exhibit very different processing kinetics **(Figure 6B-C**), demonstrating that processing kinetics, for the miR-221/222 cluster, are not defined by the primary transcript or its association to chromatin, as has recently been suggested[86]. The processing efficiency of miR-221 and miR-222 over time was quantified by qPCR shown in (**Figure 6C**).

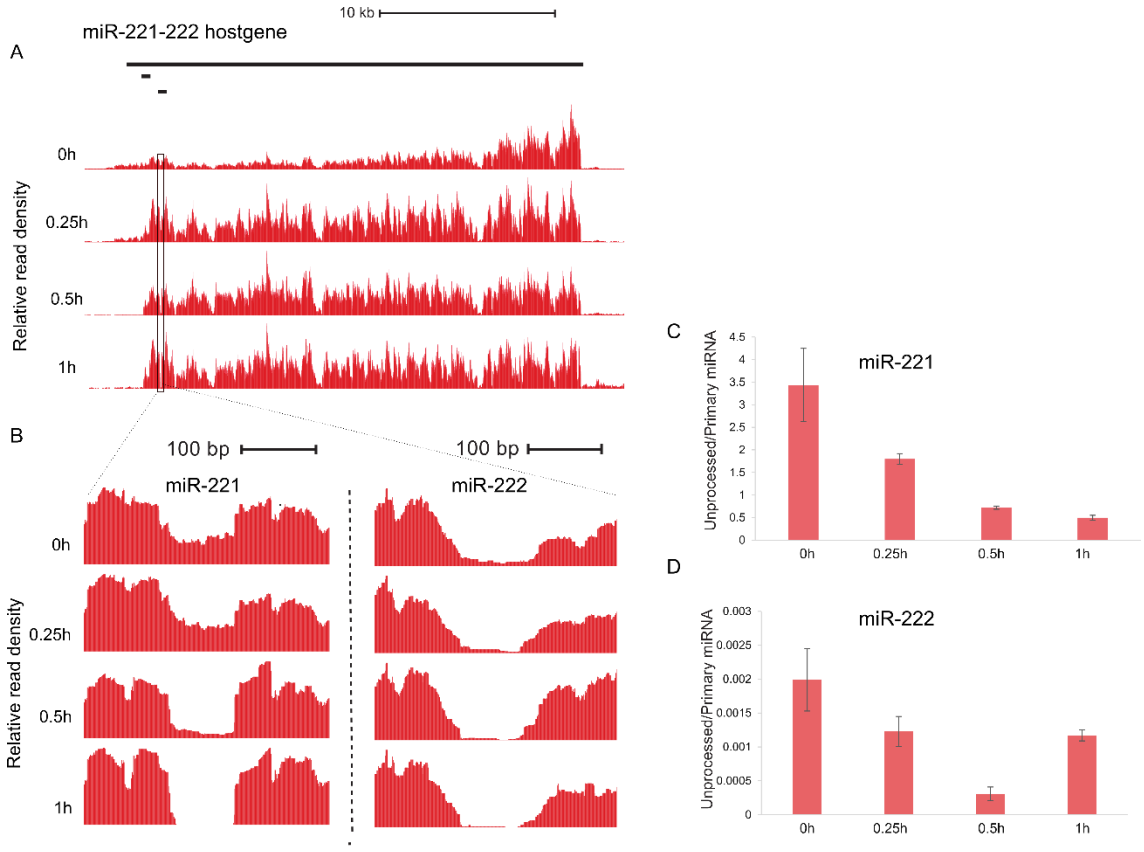


Figure 6: Differential processing within polycistronic pri-miRNAs

(A) Overview of the genomic region and full pri-miRNA transcript. (B) Enlarged read-densities around pre-miRNAs for miR-221 and miR-222. (C-D) Quantification by PCR of unprocessed/primary pri-miRNA for (C) miR-221 and miR-222.

3.4. *In vivo* profiles of pre-mRNA processing dynamics from whole cells[2]

To study the splicing kinetics of nascent RNA, we calculated the splicing index value Θ[36]**(Figure 7A**) and we determined the splicing efficiency across all time points for introns that had at least 5 reads coverage on both 5' and 3' SJ for all RNA sequencing libraries (four time points of BrU Chase-seq and the three Input samples for TNT-seq). Accordingly, we extracted the Θ value from 13,532 introns, ranging from 0 (unspliced) to 1 (fully spliced). We calculated the cumulative distribution of the splicing index for all four time points and steady-state chromatin-associated RNA[18] and as expected, the BrU-Chase Seq 0 min, captured more unspliced pre-mRNAs than CA-RNA and rest of the BrU-Chase Seq time points (**Figure 7B**).

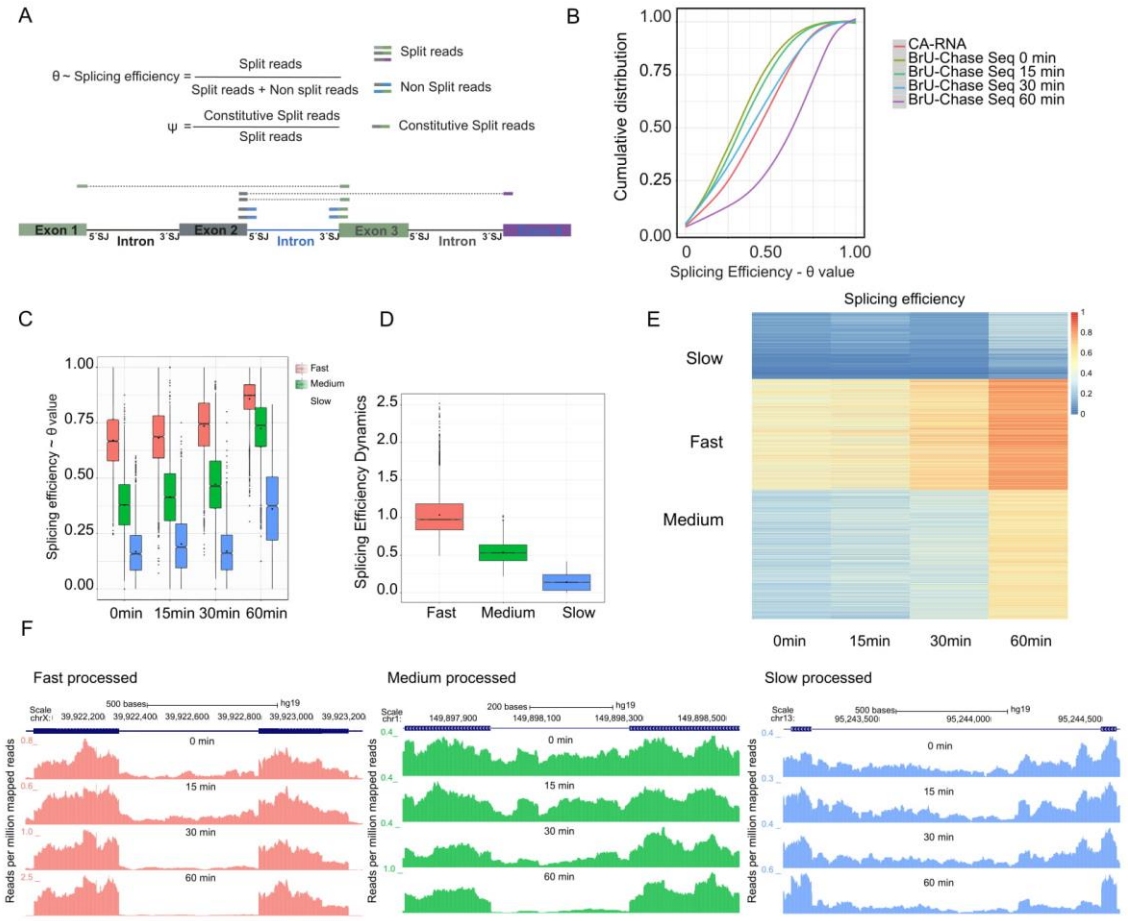


Figure 7: *In vivo* profiles of pre-mRNA processing dynamics from whole cells.

(A) Definition of Θ and ψ value. (B) Cumulative distribution of the SE index from chromatin-associated RNA-seq (Conrad et al., 2014), BrU-Chase Seq 0 min, 15 min, 30 min and 60 min. (C) Box plot representing the density of the SE index (θ value) distribution for introns grouped on the basis of differential splicing kinetics. (D) Boxplot showing distribution of the Splicing Efficiency Dynamics (SED) for the Fast, Medium and Slow processed intron groups. SED=1/((1.001-SE 0 min)*(1.001-SE 60 min)) (E) Heatmap showing the k-means clustering results (with k = 3) of the splicing SE index (θ value) of the 13,532 filtered introns measured for the BrU-Chase time points. Introns are clustered into fast-, medium- and slow-processed. (F) UCSC genome-browser views of representative cases of introns from each of the three clustering groups.

Using k-means clustering with k = 3 we obtained three clusters of distinct splicing efficiency dynamics (SED) representing 4,882 fast, 5,702 medium, and 2,948 slowly processed introns (**Figure 7C-E).** The value of SED is calculated by SED=1/ ((1.001-SE 0 min) * (1.001-SE 60 min)) and represents the splicing dynamics for each intron including the initial splicing efficiency (SE 0 min). Snapshots from the UCSC genome browser for three representative cases are shown in **Figure 7F**.

3.5. Transient N-6-methyladensosine Transcriptome sequencing[2]

We developed TNT-seq, a technique to detect m6A on nascent RNA, enabling us to study the deposition of m6A on short-lived RNA processing intermediates. Concisely, we applied MeRIP-Seq[43,45] on metabolically labeled transcripts that are produced within a 15 minutes window of active transcription (**Figure 8A**). Directly after a 15 minutes BrU-pulse, cells were collected and the isolated RNA was heat-fragmented to ~100 nt length. After calibration experiments, we chose 94°C for 3 minutes to succeed the desired fragment length (**Figure 1B**). The length of the RNA fragments was also verified with Agilent 2100 Bioanalyser shown in (**Figure 8C**). BrU-labeled RNA was subsequently eluted via BrU competition, to reduce background from contaminating unlabeled RNA, and the eluate was then subjected to immunoprecipitation with an m6A-specific antibody to enrich for methylated RNA fragments. The BrU-labeled nascent RNA (BrU-RNA Input) and the m6A enriched RNA fragments (BrU-m6A-RNA IP eluate) were then subjected to deep sequencing to identify positions of m6A on nascent RNA (**Figure 8A**). We detected localized enrichment of m6A deposition at start and stop codons as well as at 5' and 3' SJs as a reproducible profile from independent replicates (**Figure 8D**), suggesting a robust experimental pipeline (genome-wide m6A signal correlation = 0.58).

3.6. TNT-seq reveals m6A deposition on newly transcribed RNA[2]

m6A peaks were called using a published pipeline[44]. We show in Figure 9A that the majority of early m6A peaks (57 %) reside within intronic sequences, 22 % in coding sequences (CDS), 5 % in 5' UTRs and 9 % in 3' UTRs. To compare m6A peak distribution in newly transcribed RNA with steady-state mRNA we reexamined MeRIP-Seq data from ref.[3] and called m6A peaks using the same pipeline. The majority of steady-state mRNA m6A peaks reside in the CDS (52 %), 3' UTR (28 %) and 5' UTR (12 %), while only a minor fraction (4 %) is intronic (**Figure 9B**). Nearly half of the CDS-associated nascent m6A peaks reside within 100 nt upstream of the 5' SJ and about one fifth are within 100 nt downstream of the 3'SJ (**Figure 9A**). For steady-state mRNA only 17 and 11 % of the CDS peaks are within the respective intervals, which could

suggest a transient functional role of early m6A deposition (**Figure 9B**). We normalized the number of m6A peaks to the length of the analyzed intervals and the respective input read coverage and find that the early m6A deposition is enriched within 100 nt of the 5' SJ exonic boundary (**Figure 9C-D**). To validate the authenticity of m6A sites on nascent RNA we evaluated the presence of the DRACH m6A consensus motif by performing a *de novo* motif search with HOMER in the regions +/-150 nt around the peak summit of best scoring peaks (score > 20, n= 5651) or in randomly generated 300 nt genomic intervals (See Methods under '*De novo* Motif Search'). This analysis showed a positional enrichment of a DGACH motif, around the m6A peak summits in particular for exonic peaks (**Figure 9E**). Furthermore, by *de novo* motif search we identified three additional motifs, sharing a SAG core, with a strong positional enrichment around the peak summit, especially for intronic peaks (**Figure 9E)**. Then, we analyzed the positional distribution of m6A peak summits around 5' SJs, 3' SJs, and start- and stop-codon anchor points for both newly transcribed and steady-state mRNA (**Figure 9F**). Early m6A peaks at and in close

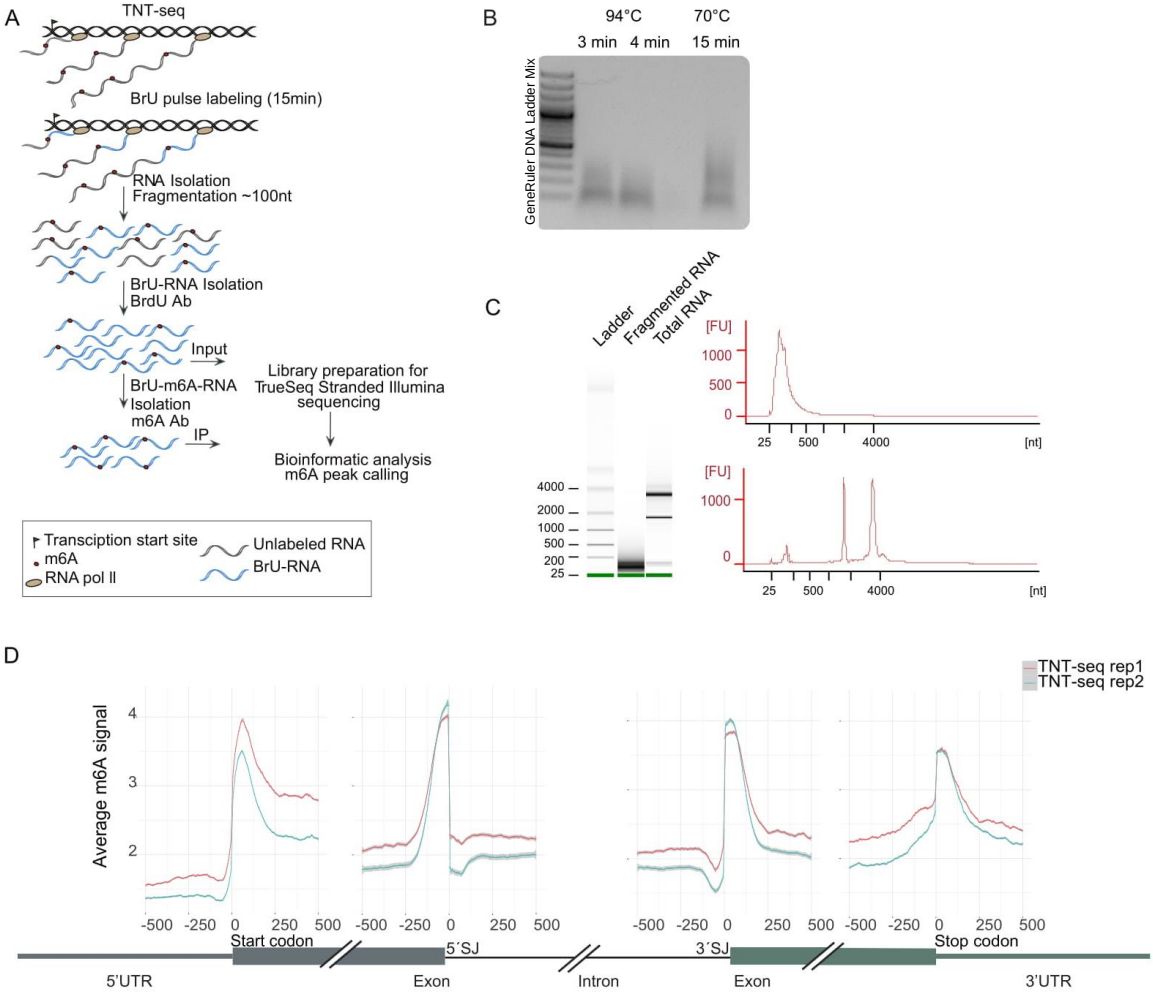


Figure 8: Transient N-6-methyladensosine Transcriptome sequencing

(A) Schematic representation of the TNT-seq protocol. (B) Agarose gel analysis of RNA fragments with different heat fragmentation conditions. (C) Bioanalyser Results of RNA fragmented for 3 minutes in 94 °C and before fragmentation-total RNA. (D) Average m6A signal per nucleotide position around start and stop codons, 5´ and 3´ SJs for the two TNT-Seq replicates

proximity to splice junctions are relatively more compared to steady-state mRNA m6A peaks (**Figure 9G-H**), whereas around start- and stop-codons the picture is inversed (**Figure 9F, I**). This result led us to examine whether early m6A deposition in close proximity to SJs has a role in splicing of RNA.

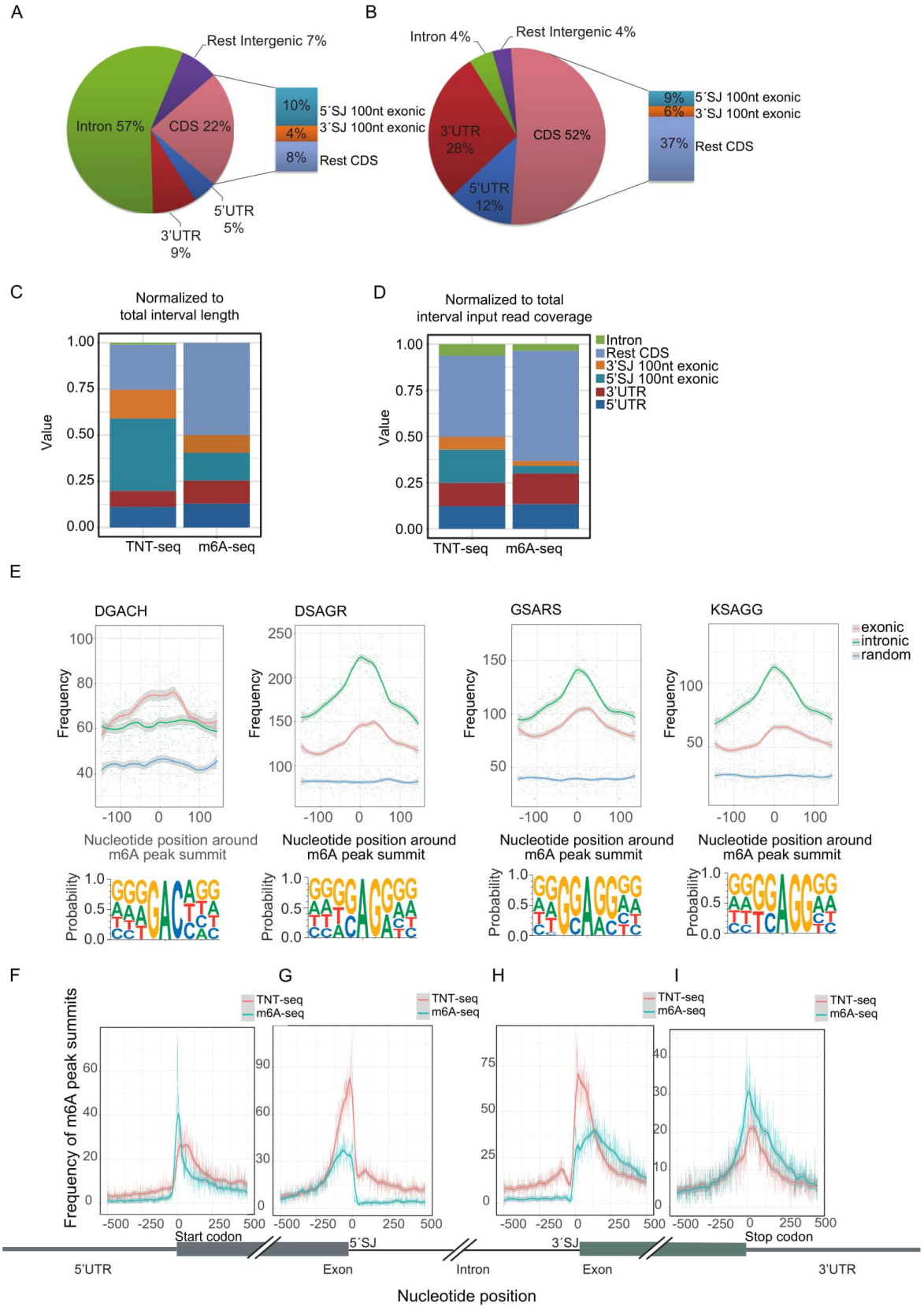


Figure 9: TNT-seq reveals m6A deposition on newly transcribed RNA

m6A peak distribution in (A) newly transcribed RNA and (B) mRNA from[51] . (C) Distribution of the normalized number of m6A peaks to the length of the analyzed intervals and (D) the respective input read coverage for TNT-seq and mRNA m6A-seq data[51]. (E) Number of motif occurrences (sum) at nucleotide positions around the m6A peak summit of the top scoring 5,651 exonic peaks, intronic peaks or random intervals. The line represents loess curve fitting (local polynomial regression) with the 95% confidence interval shaded grey. (F) Distribution (frequency) of the distance of m6A peak summits to the closest given anchor point, (E) Start codon; (G) 5´SJ, (H) 3´SJ and (I) Stop codon for nascent RNA (TNT-Seq) and mRNA (m6A-Seq[51]).

3.7. m6A signatures separate distinct intron classes[2]

To study how early m6A deposition varies with different processing efficiencies, we plotted the average m6A signal per nucleotide position around 5' and 3' SJ (**Figure 10A, B, D, E**) and within length-binned introns for the three groups (**Figure 10C**). To avoid the overlap of +/- 500nt from 5' and 3' SJ we analysed 6742 with length 1000-10000 nt and their adjacent 5´ and 3 ´SJ but also the average m6A signal of 5' and 3' SJ for all 13532 introns. Notably, we found that fast processed introns show greater m6A deposition at SJs with an overall positive relationship between m6A deposited at 5' and 3' SJ exonic boundaries and processing efficiency (**Figure 10A, B, D, E**). We reached the same conclusion also when we plotted the average frequency of m6A peak summits per nucleotide position (instead of the average m6A signal) for the three subgroups (**Figure 10G-I**). Contrary, slowly processed introns are associated with increased m6A deposition within the intron (**Figure 10B, H**). To address whether the position of an intron affects m6A signal and splicing efficiency we plotted the average m6A signal per nucleotide position around the 5' and 3' SJs of only the first and last introns (of transcripts with at least four exons). This analysis show that the observed effect is independent of the position of the intron (**Figure 10K-M**).

3.8. m6A deposition at nascent RNA predicts splicing efficiency dynamics[2]

To further investigate the role of early m6A sites in shaping the splicing efficiency dynamics we used several features in a logistic regression model fit to predict fast versus slowly processed introns (**Figure 11A, B**). We show that inclusion of the m6A signal at SJs as an additional parameter improves the predictive power of the model (**Figure 11A**), with the m6A contribution in predicting fast processing being comparable to other previously shown features[36], such as the 5' and 3' SJ sequence scores and distance to TSS/TES (**Figure 11B**). In contrast, the overall intronic internal m6A signal and intron length are significantly associated with slow processing (**Figure 11B**).

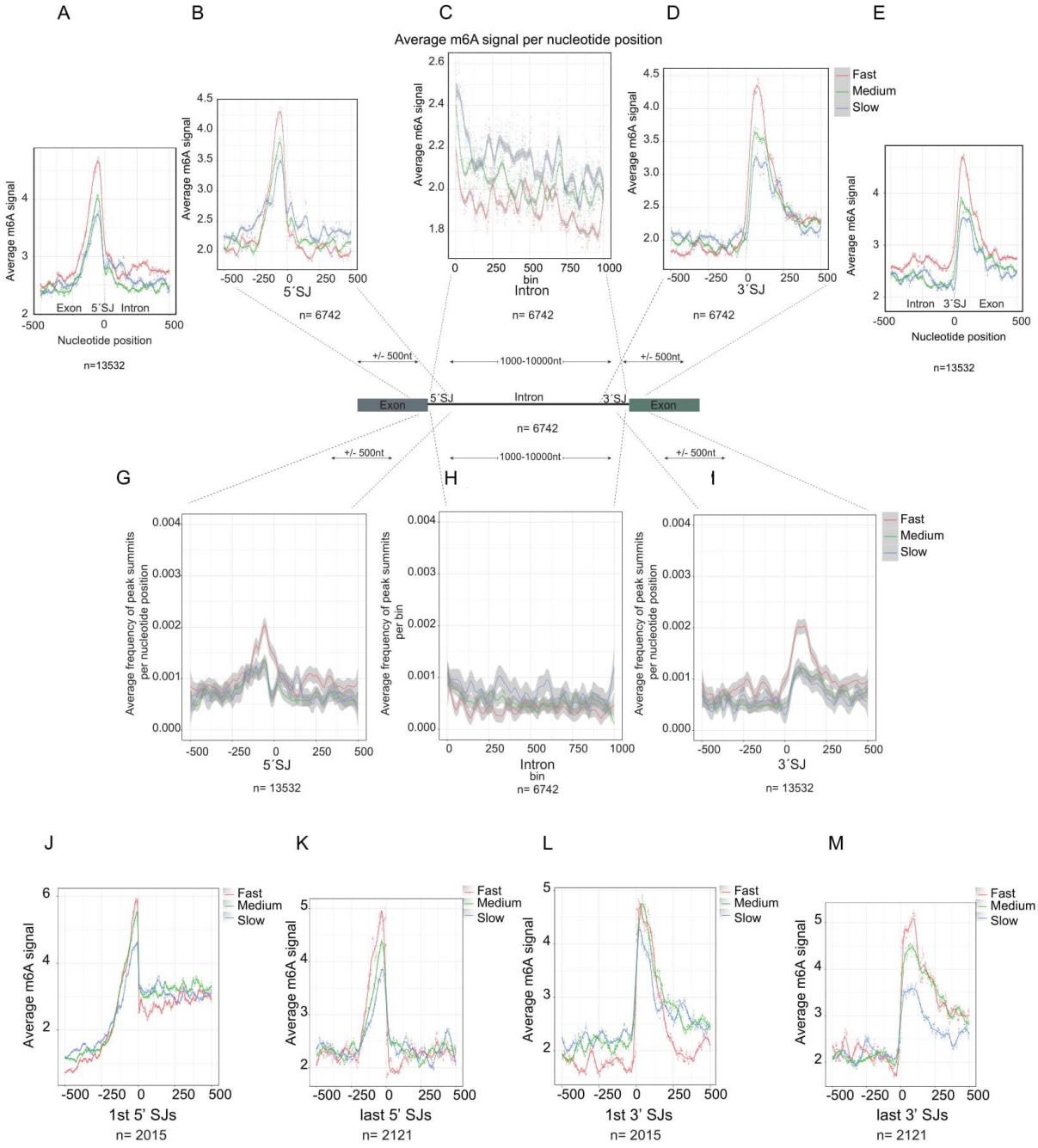


Figure 10: m6A signatures separate distinct intron classes.

Average m6A signal per nucleotide position in a +/- 500 nt window around (A, B) 5´SJ and (D, E) 3´SJ of the 13532 filtered introns and for the 6742 introns with length 1000-10000 nt, for fast, medium and slow processed introns. (C) Average m6A signal per nucleotide position internally per bin for 6742 introns (with length 1000-10000 nt), for fast, medium and slow processed introns. (F-H) Average frequency of m6A peak summits per nucleotide position in the window +/- 500 nt around (F) 5' SJ, (H) 3' SJ of all 13,532 filtered introns, and (G) per bin of 6722 introns 1000-10000 nt long, extracted separately for fast, medium, slow subgroups. The lines represent loess curve fitting (local polynomial regression) with the 95% confidence interval grey shaded. (K-M) Average m6A signal per nucleotide position around the 5' and 3' SJs of only the first and last introns. n = number of introns

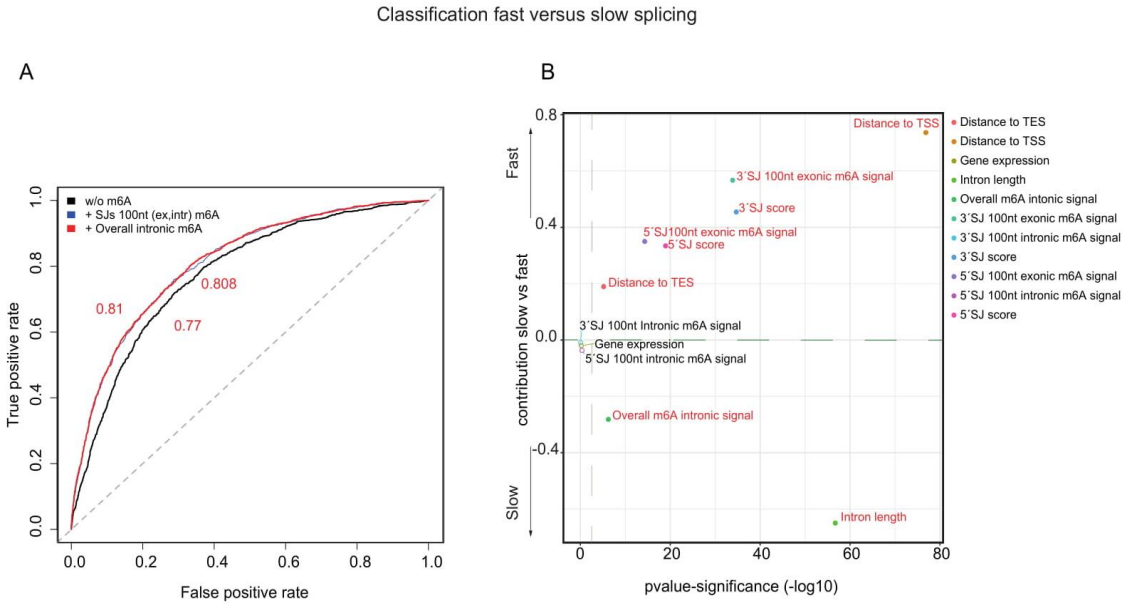


Figure 11: m6A deposition at nascent RNA predicts splicing efficiency dynamics

(A) Average receiver operating characteristics (ROC) curve for discrimination of fast versus slow introns including all characteristics and excluding m6A. The respective Area Under the Curve (AUC number) is indicated. (B) Contribution of each feature to the model fit of fast versus slow processing calculated as the coefficients from the binary logistic regression with the associated estimated significance (-log10 p-value). The features with p-value <0.001 are colored red

3.9. Internal intronic m6A deposition associates with alternative splicing[2].

Slow pre-mRNA processing has been linked with the occurrence of alternative splicing, *i.e.* exon-skipping[36]. We examined alternative versus constitutive splicing by extracting the intron-centric ψ value as in ref.[36] (**Figure 12A**). Our analysis further supports that alternative splicing events are significantly enriched in slowly processed introns (odds ratio 3.84, Fisher's exact test p-value < 2.2e-16) (**Figure 12A**). We next asked whether intronic m6A deposition could affect alternative splicing. We show that intronic m6A peaks correlate with upstream or downstream exon-skipping about two times more often than expected by random chance (odds ratio 1.7, Fisher's exact test p-value < 2.2e-16), indicating that internal intronic m6A deposition is significantly enriched in alternative splicing events. In agreement, we found that the average m6A signal is greater along alternative versus constitutively spliced introns (**Figure 12C**) and that the average m6A signal is greater at constitutive versus alternatively spliced SJ exonic boundaries (**Figure 12B, 12D**). In the prediction of alternative versus constitutive splicing shown in Figure 12E, the overall intronic m6A, along with the physical characteristic of intron length, are significant contributors in determining alternative splicing. On the other hand, m6A at SJ exonic boundaries and strong splice site consensus sequences (SJ score) ensure constitutive splicing (**Figure 12E**). The inclusion of m6A signal once more improves the predictive power of the model

fit of constitutive versus alternative splicing (**Figure 12F**), highlighting the impact of m6A deposition on nascent RNA in shaping splicing efficiency.

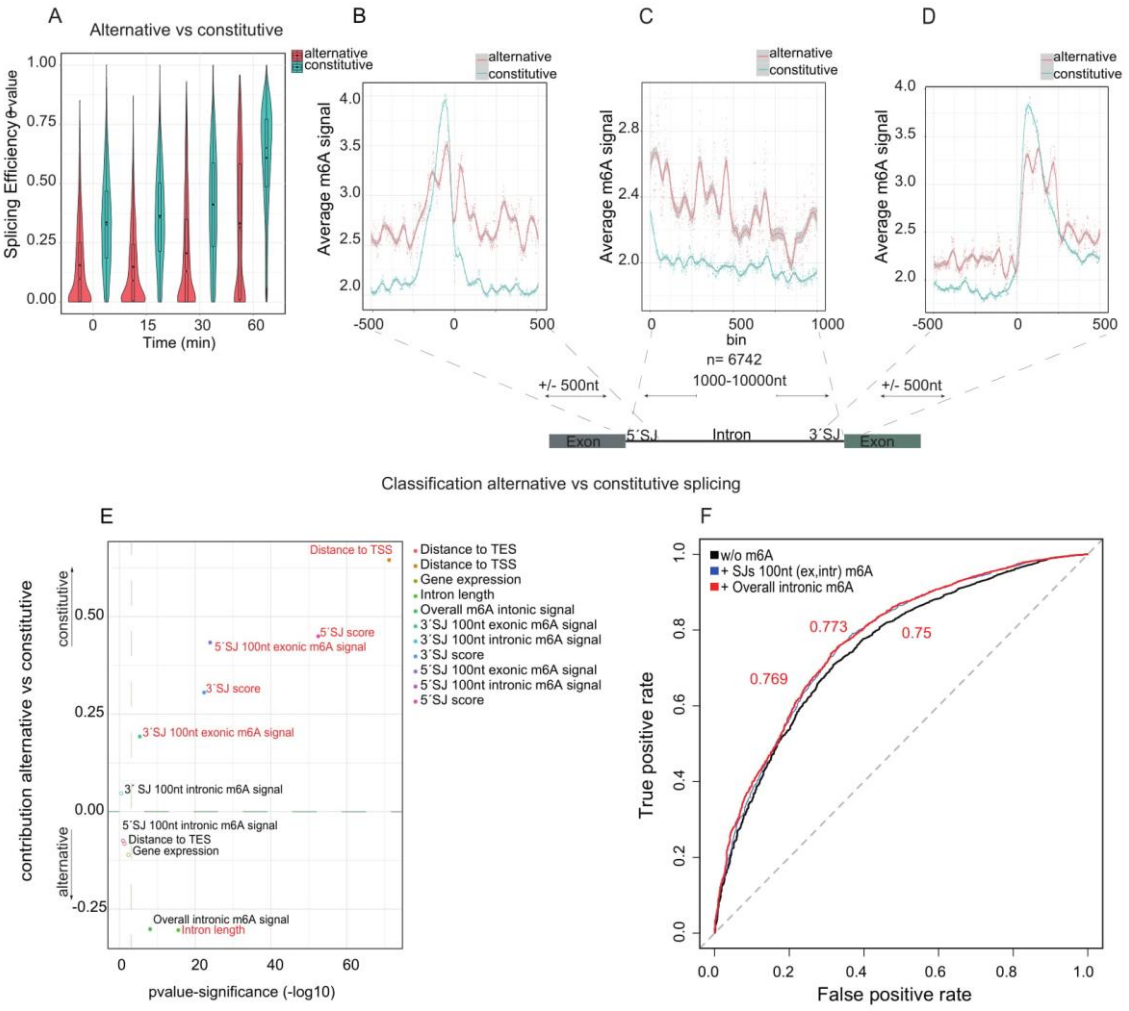


Figure 12: Intronic m6A deposition associates with alternative splicing

(A) Violin plots showing density of the distribution (with embedded box-and-whiskers plots) of θ value for introns classified as either constitutive or alternative spliced extracted from all pulse-chase time points. (B-D) Average m6A signal per nucleotide position in a +/- 500 nt window around (B) the 5' SJ and (D) 3' SJ, and per bin (C) of 6,742 introns with length 1,000-10,000 nt. The average m6A signal is extracted separately for the two subgroups, constitutive and alternative. The lines represent loess curve fitting (local polynomial regression) with the 95% confidence interval shaded grey. (E) The contribution of each feature to alternative versus constitutive splicing, calculated as the coefficients of the binary logistic regression fit with associated estimated significance (-log10 p-value). Features with $p<0.001$ are colored red. (F) Average ROC for the logistic regression prediction of the alternative versus constitutive splicing using all features, with and without m6A data. The respective AUC number is indicated.

3.10. qTNTchase-seq identifies m6A-marked fast-track RNAs[2]

We examined the direct impact of m6A modifications at the individual transcript level by developing and applying qTNTchase-seq (quantitative TNT pulse-chase sequencing). This method enabled us to clearly separate directly m6A-mediated from sequence specific effects on RNA processing. Here, labeled RNA was isolated at 0 and 30 min chase after a short BrU pulse and then, without prior fragmentation to maintain transcript level information and the m6A methylated transcripts were immunoprecipitated with an m6A-specific antibody (**Figure 13A**). Importantly, we kept the supernatant representing the m6A negative transcripts, and both the supernatant (m6A negative transcripts) and the eluate (m6A positive transcripts) from each time-point were sequenced to obtain quantitative information. To validate that qTNTchase-seq can quantify m6A levels or stoichiometry, we mixed non-mammalian m6A-modified RNAs and unmodified RNAs generated *in vitro*, at ratios ranging from 0% for GFP and 20% for Luciferase (modified to unmodified). qPCR measurements prior to sequencing showed a quantitative

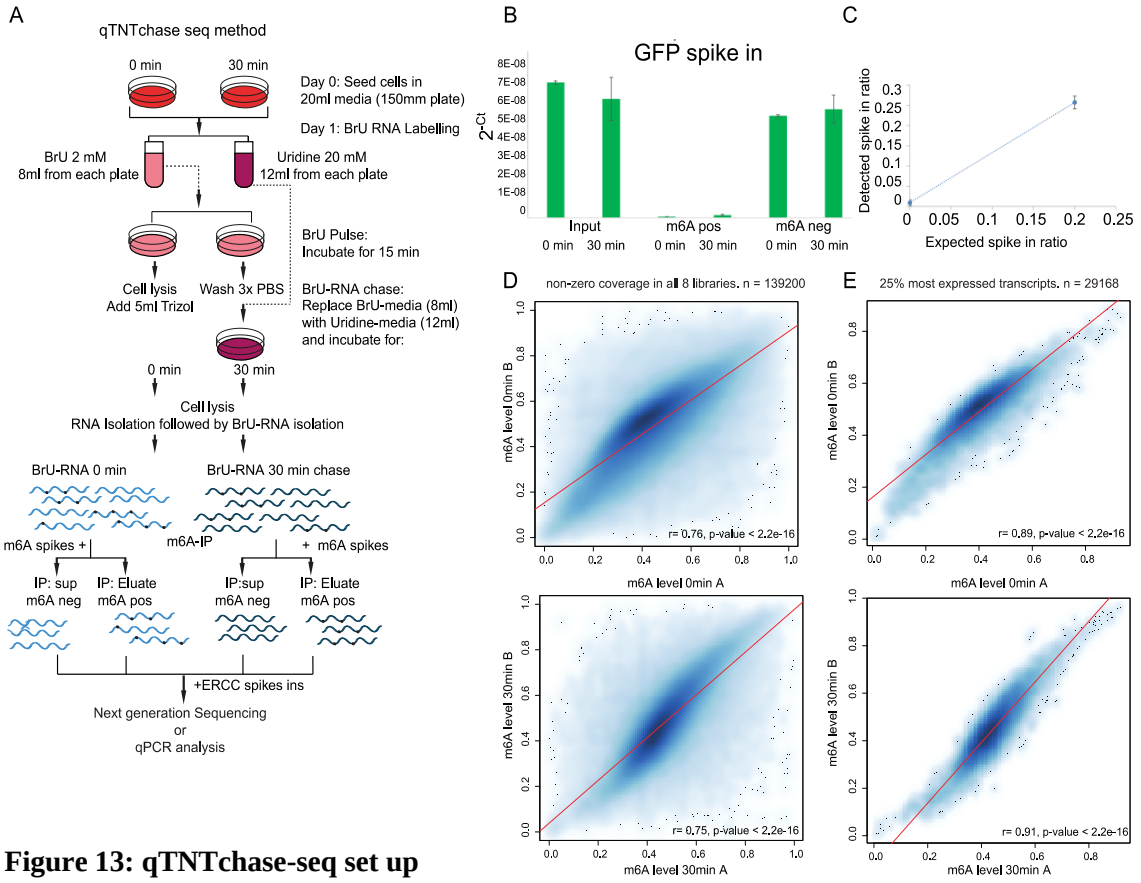


Figure 13: qTNTchase-seq set up

(A) Schematic description of the qTNTchase-seq method. (B) qPCR measurments of spike in GFP enrichment in the Inputs 0 min, 30min, Eluates (m6A pos 0 min, 30min,) and Supernatant (m6A neg 0 min, 30min). (C) qTNTchase-seq quantifies m6A levels. Scatterplot comparing expected versus observed spike in % ratio of 0% GFP and 20% Luciferase. The spike ins were generated in vitro, with each transcript harboring m6A modifications mixed with unmodified counterparts at indicated ratios as spike-ins before anti-m6A RIP. The error bars represent standard error (n=2).

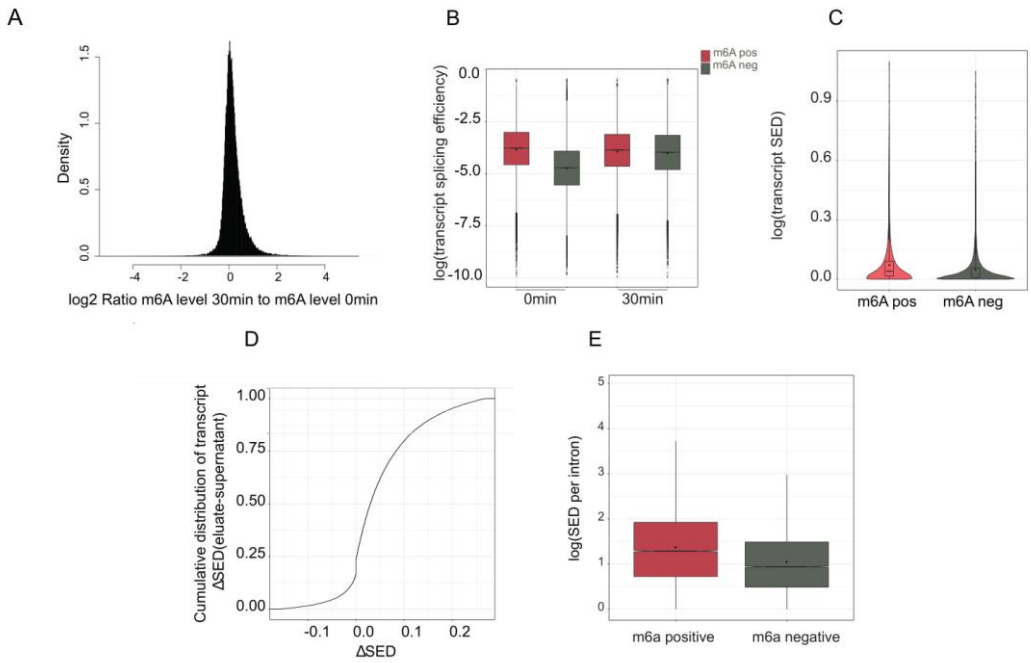


Figure 14: qTNTchase-seq identifies m6A-marked fast-track RNAs

(A) Density plot of the log2 ratio of the m6A levels per transcript measured at 30 minutes divided to 0 minutes (all transcripts with non-zero coverage. (B) Box plot representing the overall SE of methylated (m6A positive) versus non-methylated (m6A negative) transcripts at time points 0 min and 30 min. (C) Violin plots showing distribution of the transcript SED in m6A positive and m6A negative fractions (two-tailed Student's t-test p-value < 2.2e-16). (D) Cumulative distribution of transcript SED differences between the methylated and unmethylated state (ΔSED = SED m6A-positive – SED m6A-negative). (E) Box plot displaying SED per intron in m6A positive and m6A negative transcripts (two tailed paired t-test p-value < 2.2e-16).

agreement with the known stoichiometry (**Figure 13B-C**). We performed two biological replicates of qTNTchase-seq and calculated the m6A levels per transcript according to ref.[80]. On a transcriptome-wide scale we observe a strong concordance of m6A levels between the two biological replicates, irrespective if only the top 25% expressed transcripts or all transcripts with non-zero coverage are included in the analysis (for 0 minutes Pearson r = 0.89 p value < 2.2e-16 and for 30 min Pearson r = 0.91 p value < 2.2e-16) (**Figure 13D-E**). When comparing m6A levels between 0 min and 30 min chase we did not observe any significant differences indicating that overall m6A modification levels on transcripts remain the same for at least ~45 minutes after transcription (**Figure 14A**). Then, we analysed splicing efficiency on the transcript level by extracting the transcript splicing index and compared this for methylated versus non-methylated transcripts at 0 min and 30 min separately. Within the pulse (0 min), corresponding to a 15-minute window of transcription, methylated transcripts show significantly higher splicing efficiency than non-methylated transcripts (**Figure 14B**), further supporting the role of the early m6A deposition in enhancing processing efficiency. In addition, by measuring the splicing efficiency dynamics (SED) at the transcript level from 0 to 30 minutes chase, we show that methylated transcripts

show on average significantly greater processing than unmethylated transcripts (two tailed paired *t-test* p-value < 2.2e-16) (**Figure 14C**). Notably, processing is significantly enhanced for the same individual transcripts in the methylated compared to the unmethylated state; ~76% of the transcripts show gain of SED in the methylated versus unmethylated state revealing a direct and sequence independent role of m6A on processing kinetics (**Figure 14D**). Furthermore, we examined locally the splicing efficiency for the dataset of the 13,532 filtered introns. We found, that ~14% have significantly higher splicing efficiency in the m6A positive than in the m6A negative transcripts and show a 1.26 fold enrichment over random chance to possess an m6A peak in the 5' SJ 250 nt exonic boundary (odds ratio 1.265, Fisher's exact test p-value 0.0006745). In addition, individual intron loci show on average significantly higher SED in methylated versus non-methylated transcripts (two tailed paired *t- test* p-value < 2.2e-16) (**Figure 14E**). We then, analyzed the splicing kinetics of four candidate splice junctions that have at least one m6A peak (+/-250nt) by qPCR on qTNTchase-seq RNA. We calculated splicing efficiency as the ratio of the spliced signal over total (spliced + unspliced) signal. Notably, at time point 0 min, methylated transcripts show higher splicing efficiency compared to their unmethylated counterparts that share the same nucleotide sequence (**Figure 15A-B**). This result was recapitulated with RT-PCR analysis where the fragments corresponding to spliced and unspliced transcripts were analyzed on an agarose gel, showing the positive effect of m6A on RNA splicing (**Figure 15C**)

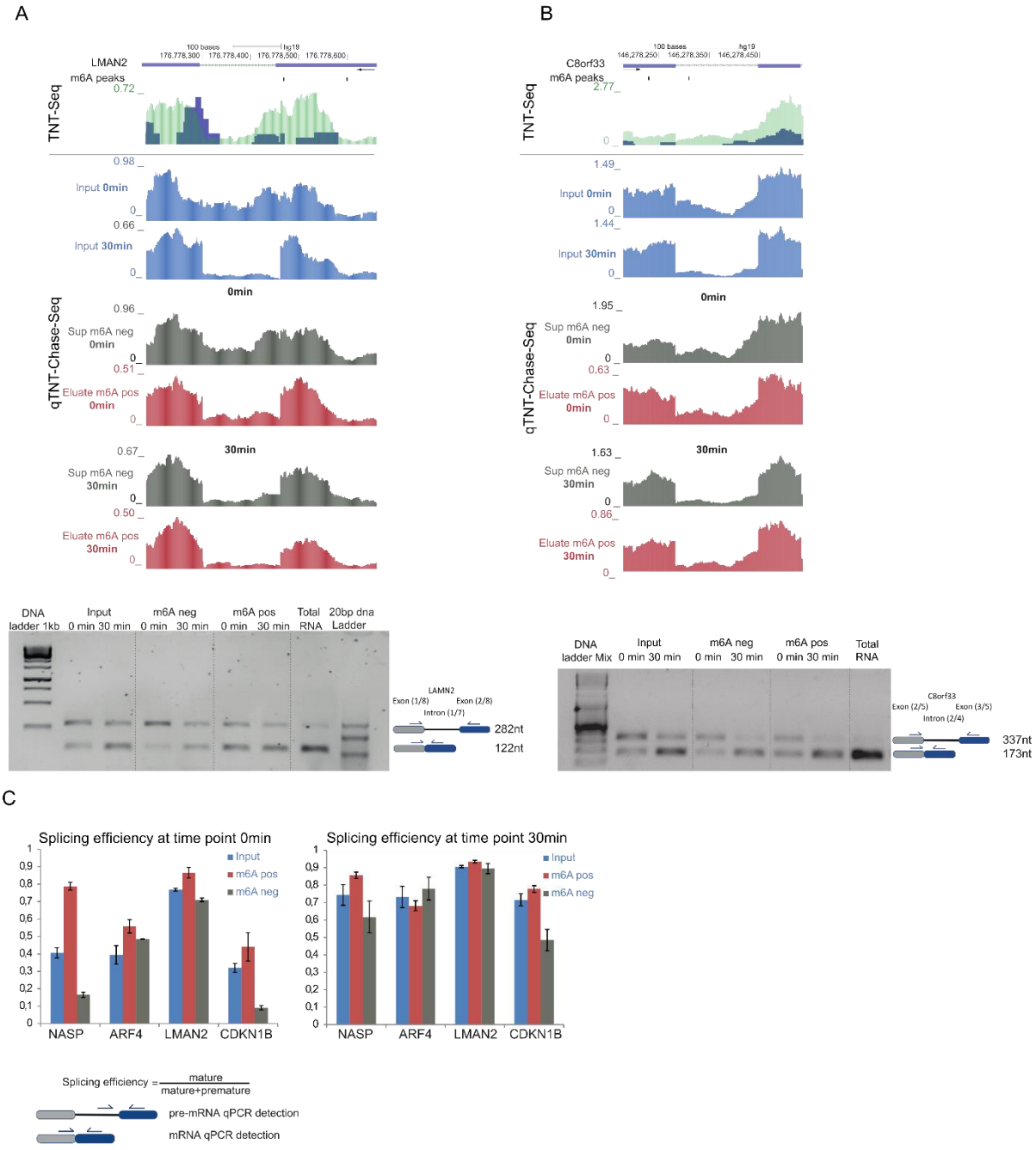


Figure 15: m6A-mediated effects on RNA processing is independent from the underlining sequence.

(A) UCSC genome browser tracks of qTNTchase-seq data representing the transcript regions used for the RT-qPCR analysis. Normalized read coverage (reads per million of total number of mapped reads) tracks for Input (blue), Supernatant m6A negative (grey), Eluate m6A positive (pink). The upper overlay track represents the TNT-seq with purple for Input and green for IP; black rectangles above represent the called m6A peaks. Below tracks for each sample are agarose gels depicting semi-quantitative PCR of Input, m6A positive and m6A negative samples for 0 min and 30 min. (B) qPCR analysis of the local intronic SE of methylated versus non-methylated transcripts for 0 minutes and 30 min

3.11. Splicing factors coincide with m6A deposition[2]

We sought to investigate the functionality of early m6A positive role in regulating RNA processing. For this we analyzed available CLIP-data for SRSF factors with a known role in splicing. We found that both SRSF3 and SRSF10 show a high probability to have a m6A peak summit in close proximity (< 250 nt) (**Figure 16A-B**), with SRSF10 showing relatively greater affinity (**Figure 16C**). Furthermore, the SAG core that we identified by *de novo* motif search in early m6A peaks (**Figure 9**) is similar to the SRSF binding site motifs[87]. In addition, both SRSF3 and SFRF10 have been shown to bind near m6A. More specifically, while SRSF3 binding can be synergistically augmented through interaction with YTHDC1, SRSF10 can independently bind to m6A modified regions[69]. In agreement, we found that the ratio of SRSF10/SRSF3 binding is greater at the SJ exonic boundaries for fast processed introns, and internally along within slowly processed introns (**Figure 16D-F**), in agreement with the respective relative enrichment of early m6A deposition (**Figure 10**). Furthermore, the average ratio of SRSF10/SRSF3 binding clearly separates alternative and constitutive spliced introns (**Figure 16 G-I**), most prominently along length-binned introns (**Figure 16H**). This result is in agreement with a previous study suggesting that alternative splicing activity can be antagonistically regulated by SRSF10 versus SRSF3 binding[10]. The above data support that early m6A deposition could play an early role in shaping the final outcome of alternative splicing activity via resolving the relative recruitment of various splicing factors with varying m6A affinities.

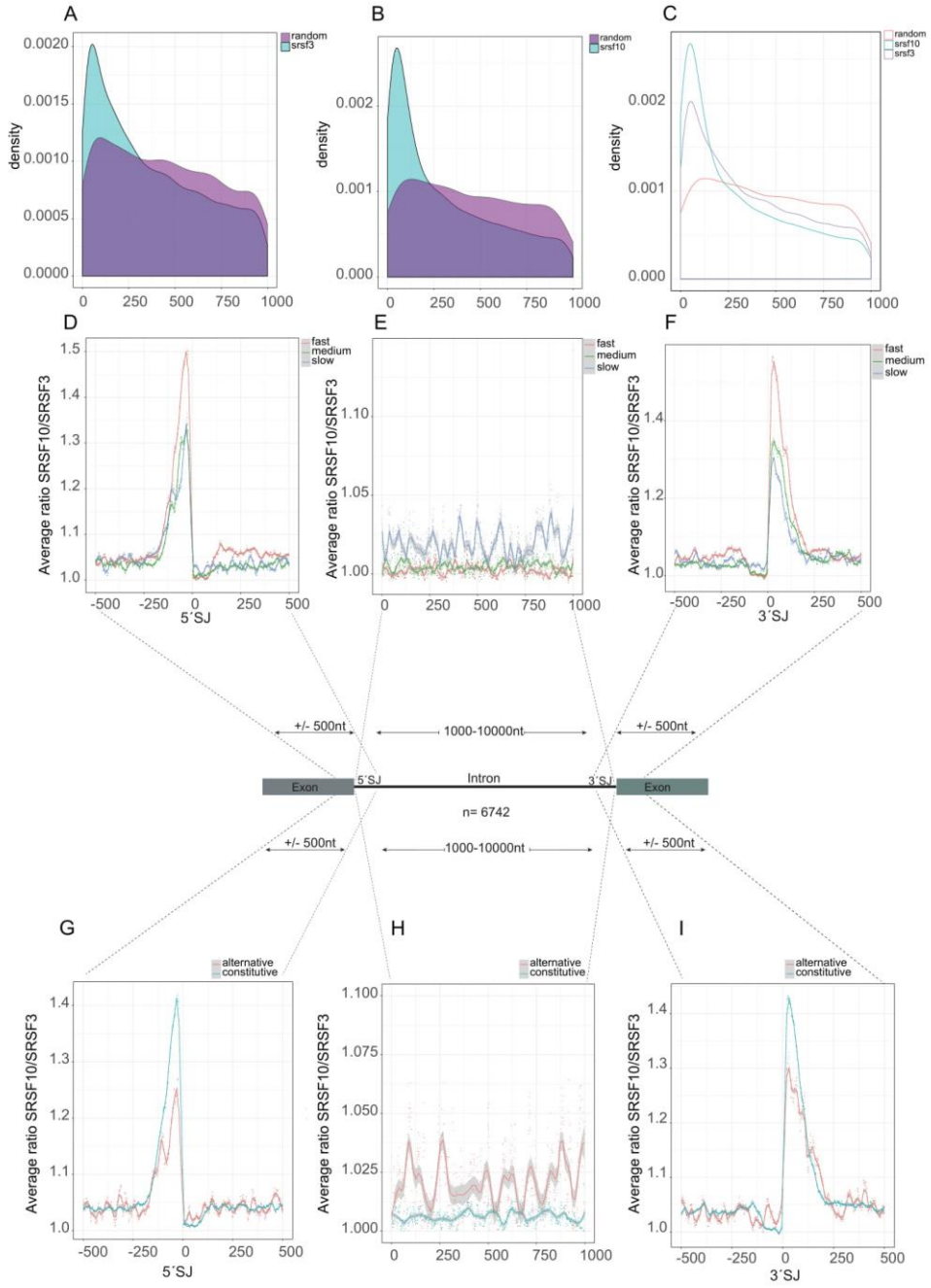


Figure 16: Splicing factors coincide with m6A deposition

(A-C) Distribution of the interdistances of factor binding sites to closest m6A peak summit for (A) SRSF3 (B) SRSF10 and (C) overlap. As a control, distance from the midpoint of the respectively same number of randomly generated genomic intervals is also plotted. (D-F) Distribution of the average ratio SRSF10/SRSF3 binding, extracted separately for the three subgroups fast/medium/slow per nucleotide position in the window +/-500 nt around the 5'SJ (D) and 3'SJ (F), or per bin (E) for 6,742 length-binned introns (with a length 1,000-10,000 nt). (G-I) Same analysis as in (D-F) but comparing the average SRSF10/SRSF3 ratio for the two subgroups constitutive versus alternative.

3.12. m6A effects are METTL3 dependent[2]

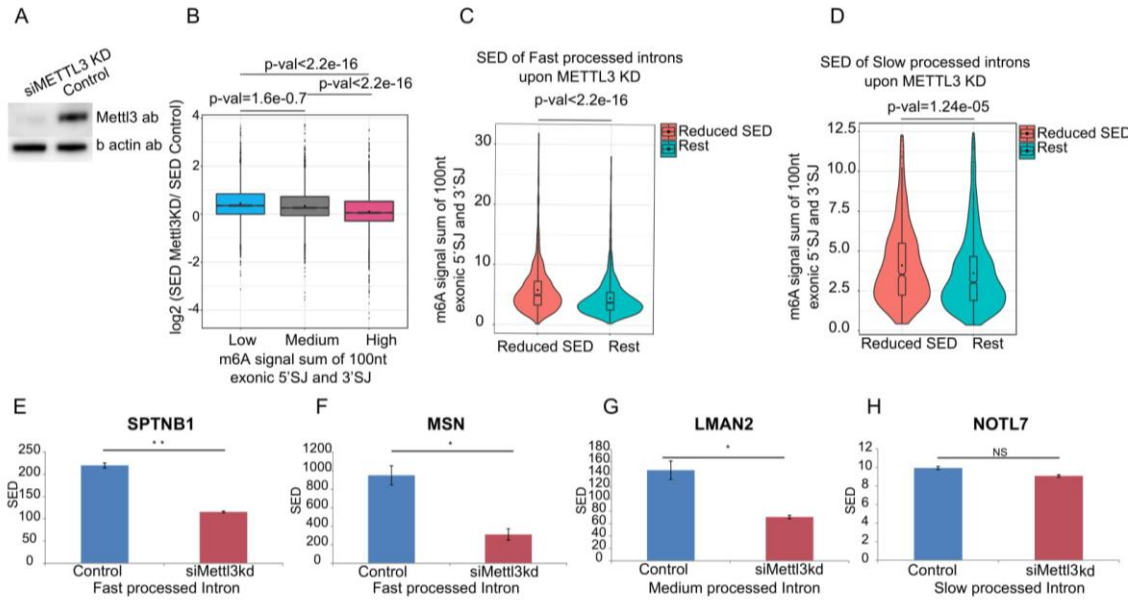


Figure 17: Nascent m6A effects are METTL3-dependent

(**A**) Western blot for METTL3 KD. **(B**) Log2 ratio of SED in METTL3 KD to Control for introns with low, medium and high m6A signal at both 5' and 3'SJ (100 nt exonic area). (**C**) m6A signal at both 5'SJ and 3'SJ (100 nt exonic area) for the Fast-processed introns that show reduced SED in the METTL3 KD condition versus the rest (two-tailed Student's t-test p-value < 2.2e-16). (**D**) m6A signal at both 5'SJ and 3'SJ (100 nt exonic area) for the Slow processed introns that show reduced SED in METTL3 KD condition versus the rest (two-tailed Student's t-test p-value < 2.2e-16). (**E-H**) qPCR analysis of SED for (**E-F**) Fast (**G**) Medium and (**H**) Slow processed introns (n = 2 biological replicates, *p<0.05 and **p<0.01, two-tailed Student's t-test)

To provide a direct link between RNA splicing kinetics and early m6A deposition at SJs we examined the splicing kinetics after METTL3 knock-down (60 min chase) (**Figure 17A**). We divided the intron dataset into three equal-size quantiles based on the m6A signal at 5' and 3'SJ (5' and 3'SJ 100 nt exonic intervals) and calculate the SED respectively. We plotted the log2 ratio of SED for METTL3 KD to Control for introns with low, medium and high m6A signal (**Figure 17B**). For introns with high m6A signal on both 5' and 3'SJ we notice a decreased SED upon METTL3 KD, for about half of the entries (log2 SED ratio METTL3 KD/Control <0) (**Figure 17B**). For introns with medium and low m6A signal (log2 SED ratio METTL3 KD/Control > 0) we observe an increased SED (**Figure 17B**). The difference in the SED ratio (log2 METTL3 KD/Control) of high m6A signal compared to low or medium is significant (t-test p-value < 2.2e-16). Then, we focused on Fast processed introns and plotted the m6A signal (sum of 5'SJ and 3'SJ 100 nt exonic area) for those that show reduced SED upon METTL3 KD versus the rest (**Figure 17C**). We found that the METTL3-affected introns have significantly higher m6A at the 5' and 3'SJ exonic boundaries. This confirms that the 5' and 3' SJ exonic methylation promotes

fast splicing kinetics, as also shown by the logistic regression model fit (**Figure 12**). We observed the same but less pronounced tendency for the Slow processed introns (**Figure 17D**). Figures 17E-H show qPCR analysis of SED for four candidates that are in agreement with the transcriptome-wide data.

4. Discussion of the Thesis.

RNA is found everywhere in the cell; in different shapes, sizes, ages and most importantly RNA participates in a vast array of cellular events. At any given time, RNA transcripts can be found in different processing stages. As a consequence, studying the entire population of cellular RNA is extraordinarily complex. A way to tackle this complexity is to label, enrich and follow a population of RNA through time. Another way is to isolate a specific cellular compartment where most likely the nascent RNA will be found. In our study, we used a metabolic labelling pulse-chase approach to enrich for the nascent RNA transcribed within 15 minutes of transcription. We show that the transcripts captured from this method are more nascent than cellular fractionation protocols[23] (**Figure 7b**). BrU-Chase Seq is a method used by Paulsen et al.,[30] to study the inflammation response in human cells. The inflammatory response in human fibroblast involves rapid and dramatic gene expression changes that could be captured by the BrU-Chase Seq method. Our questions revolved around RNA processing of pri-miRNA and pre-mRNA and to answer them we established an improved version of BrU-Chase Seq[4].

BrU-Chase Seq revealed the complexity of pri-miRNA processing with distinct processing dynamics. Given that the miRNA expression is directly related to the pri-miRNA processing rather than pri-miRNA expression[18], it is important to understand which is the molecular mechanism that drives the different processing kinetics. Several protein co-factors may facilitate or inhibit the miRNA biogenesis either through the recruitment of Microprocessor to the pri-miRNA transcripts[86] or by inhibiting Microprocessors binding. Our method[4] allowed us to detect the differential processing within polycistronic pri-miRNAs. More specifically, miR-221 and miR-222 are derived from the same pri-miRNA gene however, miR-222 has faster processing than miR-221. The miR-221/222 cluster is characterised as oncogene and shares many targets such as p27 tumor suppressor[85]. On the other hand, miR-221 and miR-222 have also non-overlapping targets suggesting that processing kinetics could define the biological regulation that could modulate the levels of miRNAs against a specific target through-out the cell cycle. It would be interesting to study at a single RNA molecule level whether these miRNAs are processed from the same pri-miRNA transcript copy. Single RNA molecule analysis coupled to FRET technology (smFRET)[88,89] should help answering this question.

The second part of the thesis provides the first high-resolution view of the transient, nascent N-6-methyladenosine transcriptome[2]. We identify an enrichment of m6A deposition

near the 5' SJs of nascent RNA transcripts and we find that early m6A deposition at SJs and within introns is associated with distinct RNA processing kinetics. A recent study from Ke et al. that was conducted in HeLa cells, reported that m6A is deposited mostly on chromatin-associated pre-mRNAs (CA-RNAs) while observing an enrichment of m6A peaks in exonic regions[44]. In agreement with our results they show an enrichment of m6A signal near exon-intron junction implicating that the m6A deposition occurs before splicing. However, they report higher m6A peak density in the last exons relative to the rest in all three cell fractions; CA-RNAs, nucleoplasmic RNA and cytoplasmic RNA fraction[44]. These results show that the m6A is deposited in most pre-mRNAs while being attached to chromatin, before the splicing is completed and that the m6A distribution remains the same during the mRNAs release from chromatin (nucleoplasmic RNAs) and after their nuclear export (cytoplasmic RNAs). The establishment of m6A-CLIP[90] allowed more precise identification of m6A signal already within ~50nt away from the SJs that rises almost double at ~100nt. TNT-seq provides a direct assessment of m6A on nascent RNA and shows that the majority of m6A peaks is found within intronic sequences consistent with METTL3-METTL14 PAR-CLIP data showing 29 %-34 % intronic binding sites[52]. However, after normalizing the number of m6A peaks to the length of the analysed intervals and the respective input read coverage, we found in agreement with Ke et al., that the early m6A deposition is enriched within 100 nt of the 5' SJ exonic boundary. On the other hand, the earliest time point BrU-Chase Seq time point 0 min (15 minutes BrU labeling without chase) captured more nascent RNA transcripts than CA-RNA[18]. The protocol for cell fractionation includes multiple biochemical steps to purify chromatin together with the associated RNA. The protocol lasts more than 15 minutes and is carried out in 4° that does not exclude the possibility that the transcription and processing is on during the phase separation[23]. Thus, the intronic sites that TNT-seq detected could be potentially removed at a later stage or co-transcriptionally while transcripts are still associated with chromatin. One hypothesis could be that RNA transcripts associated with chromatin are more prone to be bound by demethylases such as FTO. FTO loss of function experiments coupled with CA-RNA m6A-CLIPs are needed to exclude the above hypothesis. A recent study contacted an extended research on FTO localization and its function in different cell lines identifying that in HeLa cells FTO is predominantly located in the cell nucleus while in HEK293 cells FTO is located both in the cytoplasm and cell nucleus[91]. In addition, FTO demethylation activity is more profound in HeLa cells in comparison with HEK 293 and 3T3-L1 cell lines. This indicates that differences in m6A distribution among different cell types are expected due to the different FTO localization and function [91].

TNT-seq in conjunction with BrU pulse-chase reveals that the signature of early m6A deposition at splice junctions and within introns is associated with distinct RNA processing kinetics[2]. Interestingly, m6A position is a feature within many that improves the prediction power

of a logistic regression model fit contributing to fast or slow splicing (**Figure 11**). More specifically, the 5´and 3´ SJ 100nt m6A signal contributes to fast splicing in very similar way as the 5´ and 3´ SJs score. This means that the lack of strong sequence consensus at SJs of many introns may be compensated by the presence of m6A that could eventually attract splicing factors to exert their function. In other words, m6A could potentially act as a splicing enhancer of splicing inhibitor either from via attracting proteins or repelling them[92]. In addition, m6A presence changes the structure of RNA locally that could either reveal or hide protein binding sites[93]. Long introns together with intronic methylation contribute to slow processing and alternative splicing. More than 50% of the nascent m6A peaks are found in introns. This is not surprising since introns are much longer than exons and have relatively increased possibility to be bound by methyltransferases. Bartosovic et al.,[42] showed that FTO, an m6A demethylase, binds mostly to introns, mediating m6A removal. FTO knockout causes alternative splicing events with a preference for exon-skipping, suggesting that demethylation of mRNA transcripts promotes exon-inclusion under normal conditions[42]. Taken together, these findings suggest that intronic m6A marks that are not targeted or not yet removed by FTO mediate exon skipping while introns involved in constitutive splicing show no enrichment in the m6A signal and most probably are targets of FTO[42]. In mRNAs, m6A is enriched in the consensus DRACH motif; however not all DRACH motifs are methylated, indicating that the presence of the sequence motif alone is not enough to drive m6A deposition. FTO CLIP data show no significant enrichment of the DRACH motif[42] leading us to hypothesize that early m6A intronic deposition is mostly in non-DRACH sequences where FTO can detect and eventually remove the m6A marks.

Using *de novo* motif analysis, we identified three motifs sharing a SAG core with a higher positional enrichment compared to the consensus DRACH motif. Interestingly, the three novel *de novo* found motifs that are enriched in our m6A peaks resemble binding sites of SRSF splicing factors known as splicing enhancers or splicing inhibitors[87]. m6A position alone, or the altered RNA structure after m6A deposition could favor the binding of an SRSF splicing inhibitor and/or could repel the binding of an SRSF splicing enhancer and vice versa. On the other hand, if m6A favors the binding of both splicing inhibitors and splicing enhancers then the stoichiometry and binding affinity of each protein could play an important role for the final outcome. Recently, Xiao et al. demonstrated that the m6A reader YTHDC1 recruits SRSF3 while competing away SRSF10 and binds to m6A sites promoting exon inclusion[69]. In the absence of YTHDC1 and SRSF3, SRSF10 has the availability to bind to free m6A sites independently, promoting exon skipping. This is also supported by a previous study from Ajiro et al., 2016 showing that SRSF3 knockdown in U2OS cells causes exon skipping events[87]. When we calculated the average SRSF10/SRSF3 ratio per nucleotide position for the three subgroups fast, medium, slow and constitutive versus alternatively spliced transcripts we observed a similar distribution and profile to the m6A signal, confirming the competitive binding of SRSF10 versus SRSF3 to m6A regions.

A key method in our study was qTNTchase-seq that enabled us to directly compare the processing of individual transcripts in the methylated versus unmethylated state. Our results show that on transcript level m6A directly controls splicing kinetics irrespectively of the underlying transcript sequence. A gene can be transcribed several times generating transcripts with identical sequence (transcript copies). We show that transcript copies can have different methylation levels. We cannot exclude the possibility that within the transcript copies m6A can be found in different positions. With TNT-seq we identified the positions that the transcripts could potentially be methylated, within methylated transcript copies though, some of the potential sites could be methylated or not. Furthermore, our analysis on the splicing kinetics for each individual transcript is the average of the splicing kinetics from all the transcripts produced from the same gene. Different RNA molecules with identical sequences could have different methylation status and if they are methylated then the position of the m6A can be also different. It still remains unknown what drives the methyltransferase complex to RNA substrate selection and further investigation is necessary to answer this key question. m6A could potentially facilitate as a labelling signal that could be recognized by m6A reader proteins to sort methylated transcripts into a fast-track processing. In a next step the position of m6A together with the presence of splicing inhibitors or enhancers direct the kinetics of the introns within a transcript.

qTNTchase-seq enabled us to have a genome-wide quantitative measurement of how many transcripts copies of a particular gene are m6A modified (m6A level) of nascent RNA (BrU-Chase seq 0 min) and also 30 minutes chased nascent RNA (BrU-Chase Seq 30 min). The m6A levels at 0 time point as well as 30 minutes chase follow almost the same normal distribution implicating that the overall m6A modification levels on transcripts remain the same for at least ~45 minutes after transcription. This finding does not exclude that demethylation occurs within the ~45 minutes since even if a transcript is demethylated but leaves only one m6A modification it would be pulled down and end up in the eluate contributing to the reads that increase the assessed transcript m6A level. qTNTchase-seq (as well as the previously published method[80] of Molinie et al., 2016) does not employ an initial RNA fragmentation step, hence this approach allows for quantitative measurement of the m6A-methylated transcripts per gene; however, it does not provide the precise positional information about where exactly the m6A modification is deposited along the transcript. Furthermore, we were not able to draw any conclusions regarding different m6A levels (i.e. high, medium and low m6A level, representing the differential relative fraction of methylated transcript copies versus non-methylated transcript copies within each cluster) and splicing efficiency. The splicing efficiency is measured either locally per splicing junction, or at transcript level. High m6A level at transcript level can inevitably also include transcripts that are methylated within the introns, which in turn correlates with slow processing. qTNTchase-seq does not distinguish whether the transcripts are

methylated within the introns or at SJs. Thus, within the methylated transcripts one could find all three cases of processing.

Our analysis proposes a role of m6A in splicing that contradicts the finding of Ke et al., suggesting no role for m6A in splicing[44]. Ke et al., examined the RNA splicing profiles in steady state poly-A isolated RNA in control and METTL3 knock-out ESC mouse cells and found that all m6A-containing constitutive exons were spliced quantitatively the same in both conditions. In our functional analysis in HEK293 cells we identified a significant difference of the METTL3 knock down effect between the highly methylated SJs from compare to low and medium. Even though we observed an overall increase of SED upon METTL3 knock down, when focusing on the fast-processed introns we identified that only highly methylated ones showed significantly decreased SED. The slow-processed introns did not show any effect upon METTL3 knock down suggesting that another methyltrasferase is responsible for the intronic methylation of slow processed genes, or that m6A has relatively low impact on slow processing kinetics. METTL16 is a promising candidate to test since its m6A marks do not occur within the DGACH sequence motif and they are rather found in introns[94].

The lack of a method to deplete m6A alone and not the methyltransferase make it difficult to conclude whether the effects upon METTL3 depletion are m6A driven. METTL3 physically interacts with RNA pol II and it has been proposed that transcription rates influence the engagement of the methyltransferase complex with slow transcription or RNA pol II pausing to increase the probability of METTL3 binding[46]. The crosstalk between slow or paused RNA pol II and METTL3 engagement could explain the overall splicing efficiency increase independent from m6A presence. Further investigation on the role of METTL3 and m6A on transcription elongation rates and pausing are needed to draw legitimate conclusions.

Our findings are in the same line with a previous study that describes m6A methylation as a mark for selective nuclear processing providing evidence for an m6A dependent mRNA metabolism[59]. Roundtree et al, showed that the knock-down of YTHDC1, results in defective nuclear export of target mRNAs, and does so in an m6A-dependent, splicing-independent manner[59]. This agrees with the tight biochemical coupling of mRNA splicing and export, that is reflected by the dual roles observed for similar adaptor proteins, particularly those of the SRSFs family. One of the pioneered studies in 2012 from Dominissini et al., had already observed a relationship with alternative splicing events and m6A deposition[43]. Recently, two studies[71,95] in Drosophila melanogaster revealed methylation-dependent changes in splicing modulating sex determination. Mettl16 was found to methylate the 3´UTR of MAT2A leading to the splicing of retained introns and production of more MAT2A[96]. In spermatogenic cells in murine testes *Alkbh5* knock-out enhanced splicing events[97]. Our findings moved a step forward our understanding of the RNA splicing code, coupling RNA methylation with the direction and kinetics of splicing.

5. Conclusions-Outlook

We are still far from fully understanding the "epitranscriptome code". Many other RNA modifications could contribute to the regulation of RNA processing, thus different combinations of RNA modifications could drive the final outcome. Our study shows that the crucial role of m6A on splicing efficiency dynamics as well as on alternative splicing is positional dependent. m6A deposited in intronic regions sorts transcripts to a slow-track processing pathway and is associated with alternative splicing while, m6A found in splice-junction exonic boundaries, sorts transcripts to a fast-track processing pathway and constitutive splicing

The technological advances have been beneficial in biological studies aiming to elucidate the complexity of cellular RNA. However, future experiments are needed to further our understanding in RNA modifications and RNA processing. It has been shown that METTL3 methylates pri-miRNAs, facilitating the recognition and processing by DGCR8. METTL3 depletion resulted in global reduction of mature miRNAs and a concomitant accumulation of unprocessed pri-miRNAs[98]. However, TNT-Seq did not detect any m6A peaks on the 38 pri-miRNAs that we studied. It would be interesting to perform an additional analysis to correlate the pri-miRNA processing efficiency in the methylated and unmethylated state transcriptome-wide using qTNT-chase Seq data.

Furthermore, it is necessary to identify the underlying mechanism driving the fast track RNA splicing. Many m6A readers could be the proteins factors behind this mechanism. YTH-protein family, HNRNPG, HNRNPC and splicing factors SRSFs are potential candidates that could be functionally tested via siRNAs or auxin-inducible degradation system following qTNT-chase Seq. Recently, a method was developed called RNA interactome using click chemistry (RICK) that can be used to capture proteins bound to nascent RNA[99]. Combining our nascent RNA m6A data set together with RICK on m6A depleted conditions we could identify new m6A readers that bind specifically on nascent RNA with high affinity on m6A.

We are still far from understanding the m6A role on single molecule RNAs. With the establishment of TNT-Seq we identified possible m6A locations on nascent transcripts. However, is not clear whether all the transcript copies acquire m6A on the identified locations. Single molecule m6A detection techniques are needed to be developed in order to dig deeper into m6A role on single RNA molecules.

The field is still in the early stage of discoveries. It remains a mystery how the selectivity, both at the transcript level and at specific modification sites, is achieved, and how this is coupled with transcriptional events. In addition, we still do not know how effector proteins, m6A readers, are regulated. Importantly, beyond simple cell lines, we have to dive into complex *in vivo* biological systems and further reveal functional relevance, as well as the potential implications

for human diseases. Our knowledge of this field is a continuously evolving process and new discoveries, mechanistic investigations, and in-depth discussions will further broaden our views.

6. Summary in English and German

The current thesis studies the RNA processing dynamics in a high-resolution time scale manner. The method used, called BrU Chase Seq (described in detail in ref.[4]) , determined the processing kinetics of pri-miRNAs within intact cells over time using a pulse-chase approach to obtain nascent RNA within a 1-hour window after transcription. Further analysis showed that pri-miRNAs exhibit different processing kinetics ranging from fast over intermediate to slow processing. In addition, polycistronic pri-miRNAs show differential processing. The first part of the thesis is part a study published in[3].

The second part of the thesis describes for the first time the role of m6A RNA modification impact on RNA splicing kinetics. Two techniques have been developed namely TNT-Seq and qTNT-Chase Seq. These techniques provide the first time-resolved high-resolution assessment of m6A on nascent RNA transcripts and unveil its importance for the control of RNA splicing kinetics. More specifically, the early co-transcriptional m6A deposition near splice junctions promotes fast splicing, while m6A modifications in introns are associated with long, slowly processed introns and alternative splicing events. In conclusion, the early m6A deposition specifies the fate of transcripts regarding splicing kinetics and alternative splicing. The second part of the thesis was published in[2].

Die vorliegende Arbeit untersucht die RNA-Prozessierungsdynamik auf einer hochaufgelösten Zeitskala. Durch die verwendete Methode, der sogenannten BrU Chase Seq (detailliert in ref.[4] beschrieben), konnte die Prozessierungskinetik von pri-miRNAs in intakten Zellen in ihrem Zeitverlauf unter Verwendung eines Puls-Chase-Ansatzes, um innerhalb eines 1-stündigen Zeitraumes nach der Transkription nascest RNA zu erhalten bestimmt werden. Weitere Analysen zeigten, dass pri-miRNAs unterschiedliche Prozesskinetiken aufweisen, von schnell über intermediär bis langsam. Darüber hinaus zeigen polycistronische Pri-miRNAs eine differentielle Prozessierung. Der erste Teil dieser Arbeit ist Teil einer in [3] veröffentlichten Studie.

Der zweite Teil dieser Arbeit beschreibt zum ersten Mal die Rolle der m6A-RNA-Modifikation für die Kinetik der RNA-Splicing Es wurden zwei Techniken entwickelt, die TNT-Seq und die qTNT-Chase Seq. Diese Techniken bieten die erste zeitaufgelöste Bewertung von m6A an aufkommenden RNA-Transkripten mit hoher Auflösung und zeigen deren Bedeutung für die Kontrolle der RNA-Splicing-Kinetik auf. Genauer gesagt, die frühe co-transkriptionelle m6A-Anlagerung in der Nähe von Spleißverbindungen fördert ein schnelles Spleißen, während m6A-Modifikationen in Intronsmit langen, langsam verarbeiteten Introns und alternativen Spleißereignissen verbunden sind. Zusammenfassend lässt sich sagen, dass die frühe Anlagerung

von m6A das Schicksal von Transkripten in Bezug auf die Kinetik von Spleißen und alternatives Spleißen bestimmt. Der zweite Teil der Dissertation wurde in[2] veröffentlicht.

7. Bibliography

1. Louloupi, A. & Ørom, U. A. V. Inhibiting Pri-miRNA Processing with Target Site Blockers. in 63–68 (Springer, New York, NY, 2018). doi:10.1007/978-1-4939-8624-8_6
2. Louloupi, A., Ntini, E., Conrad, T. & Ørom, U. A. V. Transient N-6-Methyladenosine Transcriptome Sequencing Reveals a Regulatory Role of m6A in Splicing Efficiency. *Cell Rep.* **23,** (2018).
3. Louloupi, A., Ntini, E., Liz, J. & Ørom, U. A. Microprocessor dynamics shows co- and post-transcriptional processing of pri-miRNAs. *RNA* **23,** 892–898 (2017).
4. Louloupi, A. & Ørom, U. A. V. Metabolic Pulse-Chase RNA Labeling for pri-miRNA Processing Dynamics. in *Methods in molecular biology (Clifton, N.J.)* **1823,** 33–41 (2018).
5. Selth, L. A., Sigurdsson, S. & Svejstrup, J. Q. Transcript Elongation by RNA Polymerase II. *Annu. Rev. Biochem.* **79,** 271–293 (2010).
6. Herzel, L., Ottoz, D. S. M. M., Alpert, T. & Neugebauer, K. M. Splicing and transcription touch base: Co-transcriptional spliceosome assembly and function. *Nat. Rev. Mol. Cell Biol.* **18,** 637–650 (2017).
7. Schmid, M. & Jensen, T. H. Controlling nuclear RNA levels. *Nat. Rev. Genet.* **19,** 518–529 (2018).
8. Schwanhäusser, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473,** 337–342 (2011).
9. Nojima, T. *et al.* RNA Polymerase II Phosphorylated on CTD Serine 5 Interacts with the Spliceosome during Co-transcriptional Splicing. *Mol. Cell* **72,** 369–379.e4 (2018).
10. Zaborowska, J., Egloff, S. & Murphy, S. The pol II CTD: new twists in the tail. *Nat. Struct. Mol. Biol.* **23,** 771–777 (2016).
11. Heidemann, M., Hintermair, C., Voß, K. & Eick, D. Dynamic phosphorylation patterns of RNA polymerase II CTD during transcription. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1829,** 55–62 (2013).
12. Herzel, L. & Neugebauer, K. M. Quantification of co-transcriptional splicing from RNA-Seq data. *Methods* **85,** 36–43 (2015).
13. Cho, E. J., Takagi, T., Moore, C. R. & Buratowski, S. mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. *Genes Dev.* **11,** 3319–26 (1997).
14. Wahl, M. C., Will, C. L. & Lührmann, R. The Spliceosome: Design Principles of a

Dynamic RNP Machine. *Cell* **136,** 701–718 (2009).

15. Moore, M. J. & Proudfoot, N. J. Pre-mRNA processing reaches back to transcription and ahead to translation. *Cell* **136,** 688–700 (2009).
16. Kuehner, J. N., Pearson, E. L. & Moore, C. Unravelling the means to an end: RNA polymerase II transcription termination. *Nature Reviews Molecular Cell Biology* **12,** 283–294 (2011).
17. Auyeung, V. C., Ulitsky, I., McGeary, S. E. & Bartel, D. P. Beyond Secondary Structure: Primary-Sequence Determinants License Pri-miRNA Hairpins for Processing. *Cell* **152,** 844–858 (2013).
18. Conrad, T., Marsico, A., Gehre, M. & Ørom, U. A. Microprocessor Activity Controls Differential miRNA Biogenesis In Vivo. *Cell Rep.* **9,** 542–554 (2014).
19. Pan, Q., Shai, O., Lee, L. J., Frey, B. J. & Blencowe, B. J. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* **40,** 1413–1415 (2008).
20. Marcinowski, L. *et al.* Real-time Transcriptional Profiling of Cellular and Viral Gene Expression during Lytic Cytomegalovirus Infection. *PLoS Pathog.* **8,** e1002908 (2012).
21. Ameur, A. *et al.* Total RNA sequencing reveals nascent transcription and widespread co-transcriptional splicing in the human brain. *Nat. Struct. Mol. Biol.* **18,** 1435–1440 (2011).
22. Khodor, Y. L. *et al.* Nascent-seq indicates widespread cotranscriptional pre-mRNA splicing in Drosophila. *Genes Dev.* **25,** 2502–12 (2011).
23. Conrad, T. & Ørom, U. A. Cellular Fractionation and Isolation of Chromatin-Associated RNA. in 1–9 (Humana Press, New York, NY, 2017). doi:10.1007/978-1-4939-4035-6_1
24. Bhatt, D. M. *et al.* Transcript Dynamics of Proinflammatory Genes Revealed by Sequence Analysis of Subcellular RNA Fractions. *Cell* **150,** 279–290 (2012).
25. Wuarin, J. & Schibler, U. Physical isolation of nascent RNA chains transcribed by RNA polymerase II: evidence for cotranscriptional splicing. *Mol. Cell. Biol.* **14,** 7219–25 (1994).
26. Pandya-Jones, A. & Black, D. L. Co-transcriptional splicing of constitutive and alternative exons. *RNA* **15,** 1896–1908 (2009).
27. Werner, M. S. & Ruthenburg, A. J. Nuclear Fractionation Reveals Thousands of Chromatin-Tethered Noncoding RNAs Adjacent to Active Genes. *Cell Rep.* **12,** 1089–1098 (2015).
28. Friedel, C. C., Dölken, L., Ruzsics, Z., Koszinowski, U. H. & Zimmer, R. Conserved principles of mammalian transcriptional regulation revealed by RNA half-life. *Nucleic Acids Res.* **37,** e115–e115 (2009).
29. Melvin, W. T., Milne, H. B., Slater, A. A., Allen, H. J. & Keir, H. M. Incorporation of 6-thioguanosine and 4-thiouridine into RNA. Application to isolation of newly synthesised

RNA by affinity chromatography. *Eur. J. Biochem.* **92,** 373–9 (1978).

30. Paulsen, M. T. *et al.* Coordinated regulation of synthesis and stability of RNA during the acute TNF-induced proinflammatory response. *Proc. Natl. Acad. Sci.* **110,** 2240–2245 (2013).
31. Tani, H. *et al.* Genome-wide determination of RNA stability reveals hundreds of short-lived noncoding transcripts in mammals. *Genome Res.* **22,** 947–956 (2012).
32. Jao, C. Y. & Salic, A. Exploring RNA transcription and turnover in vivo by using click chemistry. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 15779–84 (2008).
33. Tani, H. & Akimitsu, N. Genome-wide technology for determining RNA stability in mammalian cells: Historical perspective and recent advantages based on modified nucleotide labeling. *RNA Biol.* **9,** 1233–1238 (2012).
34. Choy, M.-K. *et al.* Genome-wide conserved consensus transcription factor binding motifs are hyper-methylated. *BMC Genomics* **11,** 519 (2010).
35. Core, L. J., Waterfall, J. J. & Lis, J. T. Nascent RNA Sequencing Reveals Widespread Pausing and Divergent Initiation at Human Promoters. *Science (80-.).* **322,** 1845–1848 (2008).
36. Mukherjee, N. *et al.* Integrative classification of human coding and noncoding genes through RNA metabolism profiles. *Nat. Struct. Mol. Biol.* **24,** 86–96 (2017).
37. Windhager, L. *et al.* Ultrashort and progressive 4sU-tagging reveals key characteristics of RNA processing at nucleotide resolution. *Genome Res.* **22,** 2031–2042 (2012).
38. Ast, G. How did alternative splicing evolve? *Nat. Rev. Genet.* **5,** 773–782 (2004).
39. Chen, M. & Manley, J. L. Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nat. Rev. Mol. Cell Biol.* **10,** 741–54 (2009).
40. Tilgner, H. *et al.* Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. *Genome Res.* **22,** 1616–25 (2012).
41. Ip, J., Schmidt, D. & Pan, Q. Global impact of RNA polymerase II elongation inhibition on alternative splicing regulation. *Genome …* 390–401 (2011). doi:10.1101/gr.111070.110.390
42. Bartosovic, M. *et al.* N6-methyladenosine demethylase FTO targets pre-mRNAs and regulates alternative splicing and 3′-end processing. *Nucleic Acids Res.* **45,** 11356–11370 (2017).
43. Dominissini, D. *et al.* Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* **485,** 201–6 (2012).
44. Ke, S. *et al.* m ^{6} A mRNA modifications are deposited in nascent pre-mRNA and are not required for splicing but do specify cytoplasmic turnover. *Genes Dev.* **31,** 990–1006 (2017).

45. Meyer, K. D. *et al.* Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell* **149,** 1635–46 (2012).
46. Slobodin, B. *et al.* Transcription Impacts the Efficiency of mRNA Translation via Co-transcriptional N6-adenosine Methylation Article Transcription Impacts the Efficiency of mRNA Translation via Co-transcriptional N6-adenosine Methylation. *Cell* **169,** 326–337.e12 (2017).
47. Wang, X. *et al.* N^6-methyladenosine modulates messenger RNA translation efficiency. *Cell* **161,** 1388–1399 (2015).
48. Roundtree, I. A. *et al.* YTHDC1 mediates nuclear export of N6-methyladenosine methylated mRNAs. *Elife* **6,** (2017).
49. Boriack-Sjodin, P. A., Ribich, S. & Copeland, R. A. RNA-modifying proteins as anticancer drug targets. *Nat. Rev. Drug Discov.* **17,** 435–453 (2018).
50. Machnicka, M. A. *et al.* MODOMICS: a database of RNA modification pathways—2013 update. *Nucleic Acids Res.* **41,** D262–D267 (2012).
51. Schwartz, S. *et al.* Perturbation of m6A Writers Reveals Two Distinct Classes of mRNA Methylation at Internal and 5' Sites. *Cell Rep.* **8,** 284–96 (2014).
52. Liu, J. *et al.* A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. *Nat. Chem. Biol.* **10,** 93–5 (2014).
53. Jia, G. *et al.* N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat. Chem. Biol.* **7,** 885–7 (2011).
54. Zheng, G. *et al.* ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Mol. Cell* **49,** 18–29 (2013).
55. Lavi, U., Fernandez-Mufioz, R. & Darnell, J. E. Content of N-6 methyl adenylic acid in heterogeneous nuclear and messenger RNA of HeLa cells. *Nucleic Acids Res.* **4,** 63–69 (1977).
56. Roundtree, I. A., Evans, M. E., Pan, T. & He, C. Dynamic RNA Modifications in Gene Expression Regulation. *Cell* **169,** 1187–1200 (2017).
57. Liu, N. *et al.* N6-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions. *Nature* **518,** 560–564 (2015).
58. Roost, C. *et al.* Structure and thermodynamics of N6-methyladenosine in RNA: A spring-loaded base modification. *J. Am. Chem. Soc.* **137,** 2107–2115 (2015).
59. Roundtree, I. A. & He, C. Nuclear m6A Reader YTHDC1 Regulates mRNA Splicing. *Trends Genet.* **32,** 320–321 (2016).
60. Li, F., Zhao, D., Wu, J. & Shi, Y. Structure of the YTH domain of human YTHDF2 in complex with an m6A mononucleotide reveals an aromatic cage for m6A recognition. *Cell Res.* **24,** 1490–1492 (2014).
61. Zhao, B. S., Roundtree, I. A. & He, C. Post-transcriptional gene regulation by mRNA

modifications. *Nat. Rev. Mol. Cell Biol.* **18,** 31–42 (2017).

62. Wang, X. *et al.* N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* **505,** 117–20 (2014).
63. Berglund, L. *et al.* A Genecentric Human Protein Atlas for Expression Profiles Based on Antibodies. *Mol. Cell. Proteomics* **7,** 2019–2027 (2008).
64. Coletta, A. *et al.* Low-complexity regions within protein sequences have position-dependent roles. *BMC Syst. Biol.* **4,** 43 (2010).
65. Wojtas, M. N. *et al.* Regulation of m6A Transcripts by the 3′→5′ RNA Helicase YTHDC2 Is Essential for a Successful Meiotic Program in the Mammalian Germline. *Mol. Cell* **68,** 374–387.e12 (2017).
66. Wang, X. *et al.* N6-methyladenosine modulates messenger RNA translation efficiency. *Cell* **161,** 1388–1399 (2015).
67. Shi, H. *et al.* YTHDF3 facilitates translation and decay of N 6-methyladenosine-modified RNA. *Cell Res.* **27,** 315–328 (2017).
68. Du, H. *et al.* YTHDF2 destabilizes m 6 A-containing RNA through direct recruitment of the CCR4-NOT deadenylase complex. *Nat. Commun.* **7,** (2016).
69. Xiao, W. *et al.* Nuclear m 6 A Reader YTHDC1 Regulates mRNA Splicing. *Mol. Cell* **61,** 507–519 (2016).
70. Roundtree, I. A. *et al.* YTHDC1 mediates nuclear export of N 6 - methyladenosine methylated mRNAs. 1–28 (2017). doi:10.7554/eLife.31311
71. Lence, T. *et al.* m6A modulates neuronal functions and sex determination in Drosophila. *Nature* **540,** 242–247 (2016).
72. Patil, D. P. *et al.* m6A RNA methylation promotes XIST-mediated transcriptional repression. *Nature* **537,** 369–373 (2016).
73. Hsu, P. J. *et al.* Ythdc2 is an N6-methyladenosine binding protein that regulates mammalian spermatogenesis. *Cell Res.* **27,** 1115–1127 (2017).
74. Alarcón, C. R. R. *et al.* HNRNPA2B1 Is a Mediator of m6A-Dependent Nuclear RNA Processing Events. *Cell* **162,** 1299–1308 (2015).
75. Liu, N. *et al.* N6-methyladenosine alters RNA structure to regulate binding of a low-complexity protein. *Nucleic Acids Res.* **45,** 6051–6063 (2017).
76. Desrosiers, R., Friderici, K. & Rottman, F. Identification of methylated nucleosides in messenger RNA from Novikoff hepatoma cells. *Proc. Natl. Acad. Sci. U. S. A.* **71,** 3971–5 (1974).
77. Thüring, K., Schmid, K., Keller, P. & Helm, M. Analysis of RNA modifications by liquid chromatography–tandem mass spectrometry. *Methods* **107,** 48–56 (2016).
78. Liu, N. *et al.* Probing N6-methyladenosine RNA modification status at single nucleotide resolution in mRNA and long noncoding RNA. *RNA* **19,** 1848–56 (2013).

79. Linder, B. *et al.* Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. *Nat. Methods* **12,** 767–772 (2015).
80. Molinie, B. *et al.* m6A-LAIC-seq reveals the census and complexity of the m6A epitranscriptome. *Nat. Methods* **13,** 692–698 (2016).
81. Paulsen, M. T. *et al.* Use of Bru-Seq and BruChase-Seq for genome-wide assessment of the synthesis and stability of RNA. *Methods* **67,** 45–54 (2014).
82. Moss, T., Langlois, F., Gagnon-Kugler, T. & Stefanovsky, V. A housekeeper with power of attorney: the rRNA genes in ribosome biogenesis. *Cell. Mol. Life Sci.* **64,** 29–49 (2007).
83. Du, P., Wang, L., Sliz, P. & Gregory, R. I. A Biogenesis Step Upstream of Microprocessor Controls miR-17∼92 Expression. *Cell* **162,** 885–899 (2015).
84. Eichhorn, S. W. *et al.* mRNA Destabilization Is the Dominant Effect of Mammalian MicroRNAs by the Time Substantial Repression Ensues. *Mol. Cell* **56,** 104–115 (2014).
85. le Sage, C. *et al.* Regulation of the p27Kip1 tumor suppressor by miR-221 and miR-222 promotes cancer cell proliferation. *EMBO J.* **26,** 3699–3708 (2007).
86. Liu, H. *et al.* HP1BP3, a Chromatin Retention Factor for Co-transcriptional MicroRNA Processing. *Mol. Cell* **63,** 420–432 (2016).
87. Ajiro, M., Jia, R., Yang, Y., Zhu, J. & Zheng, Z.-M. A genome landscape of SRSF3-regulated splicing events and gene expression in human osteosarcoma U2OS cells. *Nucleic Acids Res.* **44,** 1854–1870 (2016).
88. Ha, T. Single-Molecule Fluorescence Resonance Energy Transfer. *Methods* **25,** 78–86 (2001).
89. Sasmal, D. K., Pulido, L. E., Kasal, S. & Huang, J. Single-molecule fluorescence resonance energy transfer in molecular biology. *Nanoscale* **8,** 19928–19944 (2016).
90. Ke, S. *et al.* A majority of m ^{6} A residues are in the last exons, allowing the potential for 3′ UTR regulation. *Genes Dev.* **29,** 2037–2053 (2015).
91. Wei, J. *et al.* Differential m6A, m6Am, and m1A Demethylation Mediated by FTO in the Cell Nucleus and Cytoplasm. *Mol. Cell* **71,** 973–985.e5 (2018).
92. Edupuganti, R. R. *et al.* N 6-methyladenosine (m 6 A) recruits and repels proteins to regulate mRNA homeostasis. (2017). doi:10.1038/nsmb.3462
93. Liu, N. *et al.* N6-methyladenosine-dependent RNA structural switches regulate RNA–protein interactions. *Nature* **518,** 560–564 (2015).
94. Warda, A. S. *et al.* Human METTL16 is a N6-methyladenosine (m6A) methyltransferase that targets pre-mRNAs and various non-coding RNAs. *EMBO Rep.* **18,** 2004–2014 (2017).
95. Haussmann, I. U. *et al.* m6A potentiates Sxl alternative pre-mRNA splicing for robust Drosophila sex determination. *Nature* **540,** 301–304 (2016).

96. Shima, H. *et al.* S-Adenosylmethionine Synthesis Is Regulated by Selective N6-Adenosine Methylation and mRNA Degradation Involving METTL16 and YTHDC1. *Cell Rep.* **21,** 3354–3363 (2017).
97. Tang, C. *et al.* ALKBH5-dependent m6A demethylation controls splicing and stability of long 3'-UTR mRNAs in male germ cells. *Proc. Natl. Acad. Sci. U. S. A.* **115,** E325–E333 (2018).
98. Alarcón, C. R., Lee, H., Goodarzi, H., Halberg, N. & Tavazoie, S. F. N6-methyladenosine marks primary microRNAs for processing. *Nature* **519,** 482–5 (2015).
99. Bao, X. *et al.* Capturing the interactome of newly transcribed RNA. *Nat. Methods* **15,** 213–220 (2018).
100. Mount, S. M. A catalogue of splice junction sequences. *Nucleic Acids Res.* **10,** 459–72 (1982).
101. Grabowski, P. J., Seiler, S. R. & Sharp, P. A. A multicomponent complex is involved in the splicing of messenger RNA precursors. *Cell* **42,** 345–53 (1985).
102. Mercer, T. R. *et al.* Genome-wide discovery of human splicing branchpoints. *Genome Res.* **25,** 290–303 (2015).
103. Séraphin, B. Sm and Sm-like proteins belong to a large family: identification of proteins of the U6 as well as the U1, U2, U4 and U5 snRNPs. *EMBO J.* **14,** 2089–98 (1995).
104. Valcárcel, J., Gaur, R. K., Singh, R. & Green, M. R. Interaction of U2AF65 RS region with pre-mRNA branch point and promotion of base pairing with U2 snRNA [corrected]. *Science* **273,** 1706–9 (1996).
105. Sun, J. S. & Manley, J. L. A novel U2-U6 snRNA structure is necessary for mammalian mRNA splicing. *Genes Dev.* **9,** 843–54 (1995).
106. Raghunathan, P. L. & Guthrie, C. RNA unwinding in U4/U6 snRNPs requires ATP hydrolysis and the DEIH-box splicing factor Brr2. *Curr. Biol.* **8,** 847–55 (1998).
107. Konarska, M. M., Vilardell, J. & Query, C. C. Repositioning of the Reaction Intermediate within the Catalytic Center of the Spliceosome. *Mol. Cell* **21,** 543–553 (2006).
108. Pena, V., Rozov, A., Fabrizio, P., Lührmann, R. & Wahl, M. C. Structure and function of an RNase H domain at the heart of the spliceosome. *EMBO J.* **27,** 2929–40 (2008).
109. Ilagan, J. O., Chalkley, R. J., Burlingame, A. L. & Jurica, M. S. Rearrangements within human spliceosomes captured after exon ligation. *RNA* **19,** 400–412 (2013).
110. Nguyen, T. A. *et al.* Functional Anatomy of the Human Microprocessor. *Cell* **161,** 1374–1387 (2015).
111. Kim, B., Jeong, K. & Kim, V. N. Genome-wide Mapping of DROSHA Cleavage Sites on Primary MicroRNAs and Noncanonical Substrates. *Mol. Cell* **66,** 258–269.e5 (2017).
112. MacRae, I. J., Zhou, K. & Doudna, J. A. Structural determinants of RNA recognition and

cleavage by Dicer. *Nat. Struct. Mol. Biol.* **14,** 934–940 (2007).

113. Neilsen, C. T., Goodall, G. J. & Bracken, C. P. IsomiRs – the overlooked repertoire in the dynamic microRNAome. *Trends Genet.* **28,** 544–549 (2012).
114. Khvorova, A., Reynolds, A. & Jayasena, S. D. Functional siRNAs and miRNAs exhibit strand bias. *Cell* **115,** 209–16 (2003).
115. Lewis, B. P., Burge, C. B. & Bartel, D. P. Conserved Seed Pairing, Often Flanked by Adenosines, Indicates that Thousands of Human Genes are MicroRNA Targets. *Cell* **120,** 15–20 (2005).
116. Guo, H., Ingolia, N. T., Weissman, J. S. & Bartel, D. P. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* **466,** 835–840 (2010).
117. Wang, E. T. *et al.* Alternative isoform regulation in human tissue transcriptomes. *Nature* **456,** 470–476 (2008).
118. Kim, M.-S. *et al.* A draft map of the human proteome. *Nature* **509,** 575–581 (2014).
119. Sun, H. & Chasin, L. A. Multiple splicing defects in an intronic false exon. *Mol. Cell. Biol.* **20,** 6414–25 (2000).
120. Wang, Z. & Burge, C. B. Splicing regulation : From a parts list of regulatory elements to an integrated splicing code Splicing regulation : From a parts list of regulatory elements to an integrated splicing code. *Rna* **14,** 802–813 (2008).
121. Corbo, C., Orrù, S. & Salvatore, F. SRp20: An overview of its role in human diseases. *Biochem. Biophys. Res. Commun.* **436,** 1–5 (2013).
122. Levy, J. B., Dorai, T., Wang, L. H. & Brugge, J. S. The structurally distinct form of pp60c-src detected in neuronal cells is encoded by a unique c-src mRNA. *Mol. Cell. Biol.* **7,** 4142–5 (1987).
123. Expert-Bezançon, A. *et al.* hnRNP A1 and the SR Proteins ASF/SF2 and SC35 Have Antagonistic Functions in Splicing of β-Tropomyosin Exon 6B. *J. Biol. Chem.* **279,** 38249–38259 (2004).
124. Bokar, J. A., Rath-shambaugh, M. E., Ludwiczak, R., Narayan, P. & Rottman, F. Characterization and Partial Purification of mRNA NG-Adenosine Methyltransferase from HeLa Cell Nuclei. *J. Biol. Chem.* **269,** 17697–17704 (1994).
125. Śledź, P. & Jinek, M. Structural insights into the molecular mechanism of the m6A writer complex. *Elife* **5,** 1–16 (2016).
126. Pendleton, K. E. *et al.* Article The U6 snRNA m 6 A Methyltransferase METTL16 Regulates SAM Synthetase Intron Retention. *Cell* **169,** 824–829.e14 (2017).
127. Gerken, T. *et al.* The Obesity-Associated FTO Gene Encodes a 2-Oxoglutarate-Dependent Nucleic Acid Demethylase. *Science (80-.).* **318,** 1469–1472 (2007).
128. Jia, G. *et al.* Oxidative demethylation of 3-methylthymine and 3-methyluracil in single-stranded DNA and RNA by mouse and human FTO. *FEBS Lett.* **582,** 3313–3319 (2008).

129. Fu, Y. *et al.* FTO-mediated formation of N6-hydroxymethyladenosine and N6-formyladenosine in mammalian RNA. *Nat. Commun.* **4,** 1798 (2013).
130. Frayling, T. M. *et al.* A Common Variant in the FTO Gene Is Associated with Body Mass Index and Predisposes to Childhood and Adult Obesity. *Science (80-.).* **316,** 889–894 (2007).
131. Zou, S. *et al.* N 6 -Methyladenosine: A conformational marker that regulates the substrate specificity of human demethylases FTO and ALKBH5. *Sci. Rep.* **6,** 1–12 (2016).

8. Appendix

Published Manuscripts

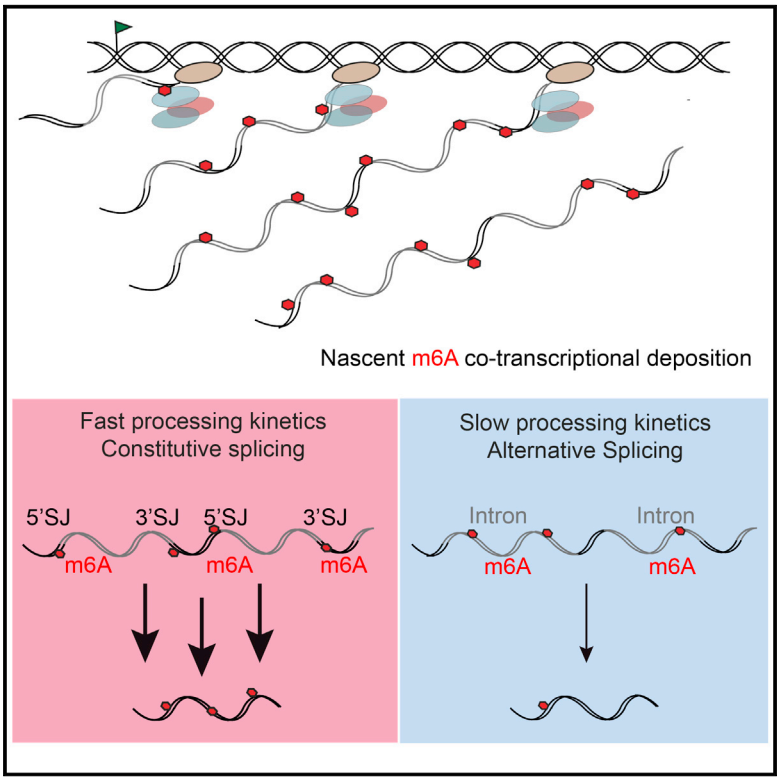
1. Transient N-6-methyladensosine Transcriptome sequencing reveals a regulatory role of m6A in splicing efficiency. Cell Rep. 2018 Jun 19;23(12):3429-3437. doi: 10.1016/j.celrep.2018.05.077.
2. Microprocessor dynamics shows co- and post-transcriptional processing of pri-miRNAs RNA. 2017 Jun;23(6):892-898. doi: 10.1261/rna.060715.117. Epub 2017 Mar 1.
3. Metabolic pulse-chase RNA labelling for pri-miRNA processing dynamics. Methods Mol Biol. 2018;1823:33-41. doi: 10.1007/978-1-4939-8624-8_3.
4. Inhibiting pri-miRNA processing with Target Site Blockers. Methods Mol Biol. 2018;1823:63-68. doi: 10.1007/978-1-4939-8624-8_6.

Report

Cell Reports

Transient N-6-Methyladenosine Transcriptome Sequencing Reveals a Regulatory Role of m6A in Splicing Efficiency

Graphical Abstract



Authors

Annita Louloupi, Evgenia Ntini, Thomas Conrad, Ulf Andersson Vang Ørom

Correspondence

ulf.orom@mbg.au.dk

In Brief

Louloupi et al. describe an approach to detect m6A RNA methylation on nascent RNA. They find that nascent transcripts are often marked by m6A at splice junctions and in introns. The authors show that m6A at splice junctions contributes to faster splicing, while m6A in introns is associated with alternative splicing.

Highlights

- A time-resolved high-resolution picture of m6A on nascent RNA transcripts
- m6A is deposited at nascent RNA and in introns
- m6A deposition at splice-junctions increases splicing kinetics
- High m6A levels in introns is associated with slow and alternative splicing

Data and Software Availability

GSE92565
GSE83561

Louloupi et al., 2018, Cell Reports *23*, 3429–3437
June 19, 2018
https://doi.org/10.1016/j.celrep.2018.05.077

CellPress

Cell Reports
Report

Transient N-6-Methyladenosine Transcriptome Sequencing Reveals a Regulatory Role of m6A in Splicing Efficiency

Annita Louloupi,[1,2,5] Evgenia Ntini,[1,5] Thomas Conrad,[1,3] and Ulf Andersson Vang Ørom[1,4,6,*]
[1]Otto Warburg Laboratories, Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany
[2]Free University, Department of Biology, 14195 Berlin, Germany
[3]BIMSB Genomics, Max Delbrück Center, 13092 Berlin, Germany
[4]Institute for Molecular Biology and Genetics, Aarhus University, 8000 Aarhus, Denmark
[5]These authors contributed equally
[6]Lead Contact
*Correspondence: ulf.orom@mbg.au.dk
https://doi.org/10.1016/j.celrep.2018.05.077

SUMMARY

Splicing efficiency varies among transcripts, and tight control of splicing kinetics is crucial for coordinated gene expression. N-6-methyladenosine (m6A) is the most abundant RNA modification and is involved in regulation of RNA biogenesis and function. The impact of m6A on regulation of RNA splicing kinetics is unknown. Here, we provide a time-resolved high-resolution assessment of m6A on nascent RNA transcripts and unveil its importance for the control of RNA splicing kinetics. We find that early co-transcriptional m6A deposition near splice junctions promotes fast splicing, while m6A modifications in introns are associated with long, slowly processed introns and alternative splicing events. In conclusion, we show that early m6A deposition specifies the fate of transcripts regarding splicing kinetics and alternative splicing.

INTRODUCTION

The RNA nucleotide code is supplemented by more than a hundred chemical modifications, greatly extending the functionality and information content of RNA (Fu et al., 2014; Harcourt et al., 2017). N-6-methyladenosine (m6A) is deposited by a protein complex consisting of the methyltransferase-like 3 and 14 (METTL3 and METTL14), Wilms' tumor 1-associating protein (WTAP), and the Virilizer homolog (KIAA1429) (Liu et al., 2014; Ping et al., 2014; Schwartz et al., 2014). Early studies have demonstrated that adenosine methylation frequently occurs within a subset of RRA*CH consensus sites (R, purine; A*, methylatable A; H, non-guanine base) (Narayan and Rottman, 1988). Fat mass and obesity associated (FTO) and AlkB homolog 5 (ALKBH5) are m6A demethylases, adding dynamics to the function of m6A in RNA biogenesis (Jia et al., 2011; Zheng et al., 2013). m6A is involved in a number of RNA processes, including splicing, RNA degradation, and translation (Bartosovic et al., 2017; Dominissini et al., 2012; Ke et al., 2017; Meyer et al., 2015; Slobodin et al., 2017; Wang et al., 2014; Xiao et al., 2016). These pathways are mediated in part by members of the YTH-domain protein family called m6A readers, which recognize and bind specifically to sequences marked with m6A (Xiao et al., 2016; Xu et al., 2014). The presence of m6A can affect the RNA structure and increase the accessibility of the adjacent RNA sequence for the heterogeneous nuclear ribonucleoproteins HNRNPG and HNRNPC, with an effect on splicing (Liu et al., 2015, 2017). Because of the challenging nature of addressing the impact of m6A on splicing at the mature RNA level, the direct role of m6A on splicing dynamics has not been investigated so far. Here, using TNT-seq (transient N-6-methyladenosine transcriptome sequencing) and qTNTchase-seq (quantitative TNT pulse-chase sequencing), we show that m6A modifications deposited early and co-transcriptionally near splice junctions (SJs) positively affect RNA splicing kinetics. Furthermore, we show that intronic m6A deposition is connected with slow processing kinetics and alternative splicing events. Our results strongly support a scenario where nascent m6A deposition is functionally involved in regulating splicing efficiency (SE) and alternative splicing.

RESULTS

TNT-Seq Reveals m6A Deposition on Newly Transcribed RNA

We developed TNT-seq to identify and study m6A on nascent RNA. In brief, bromouridine (BrU)-labeled RNA was isolated, fragmented, and purified with a BrU-specific antibody. Subsequently, m6A methylated fragments were isolated using an m6A-specific antibody. The labeled RNA (BrU-RNA input) and the m6A-enriched RNA fragments (BrU-m6A-RNA IP eluate) were sequenced to identify positions of m6A on nascent RNA (Figure S1A). We find enrichment of m6A around start and stop codons as well as at 5′ and 3′ SJs reproducibly across independent replicates (Figure S1B), demonstrating a robust experimental pipeline (genome-wide m6A signal correlation = 0.58). The majority (57%) of early m6A peaks (Experimental Procedures) reside within introns, whereas 22% reside in coding sequences (CDSs), 5% are in 5′ UTRs, and 9% are in 3′ UTRs (Figure S1C). To compare m6A peak distribution in newly transcribed RNA with steady-state mRNA, we reanalyzed published m6A-sequencing (m6A-seq)

Cell Reports 23, 3429–3437, June 19, 2018

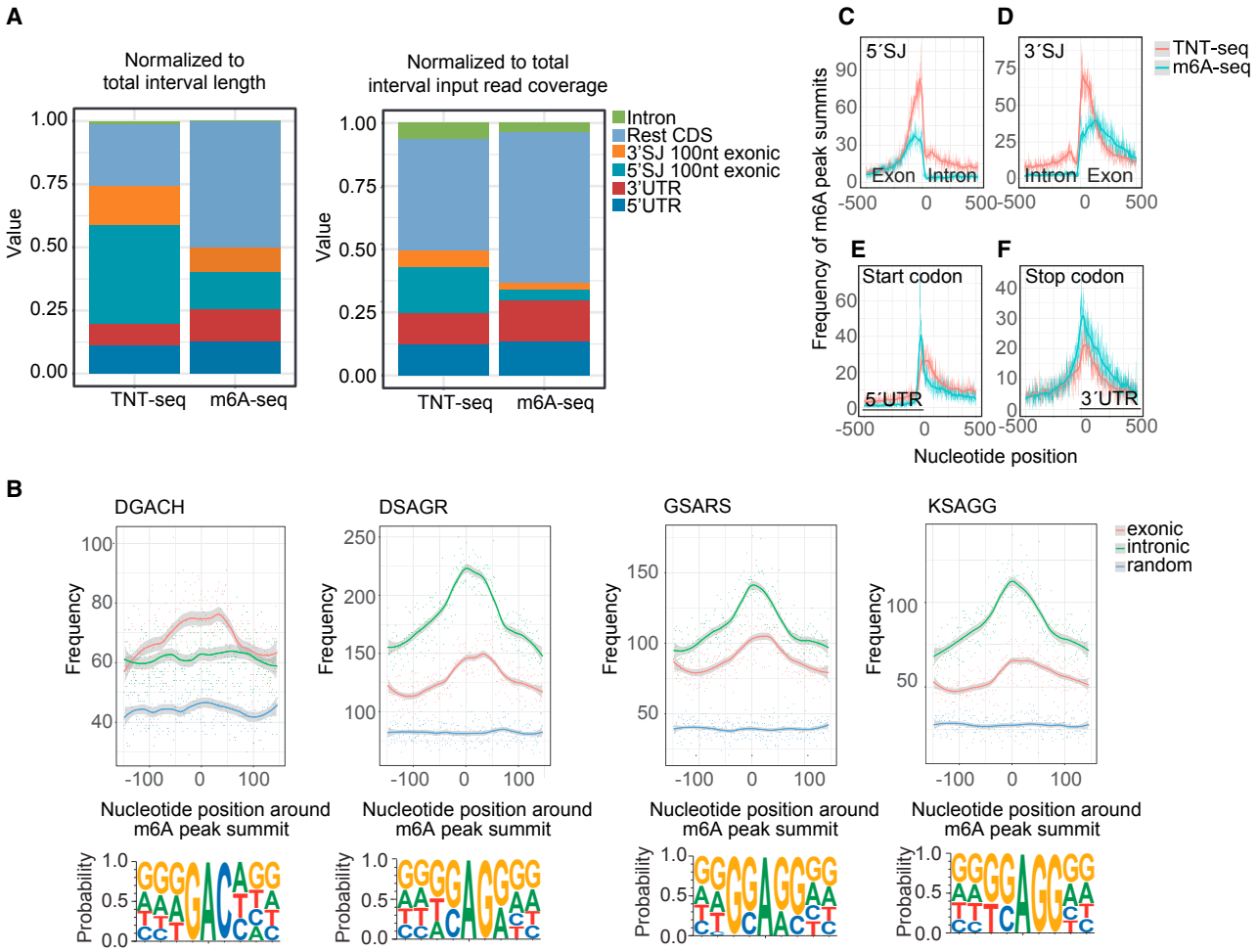


Figure 1. TNT-Seq Reveals m6A Deposition on Newly Transcribed RNA

(A) Distribution of the normalized number of m6A peaks to the length of the analyzed intervals and the respective input read coverage for TNT-seq and mRNA m6A-seq data.

(B) Number of motif occurrences (sum) at nucleotide positions around the m6A peak summit of the top scoring 5,651 exonic peaks, intronic peaks, or random intervals. The line represents loess curve fitting (local polynomial regression), with the 95% confidence interval shaded gray.

(C–F) Distribution (frequency) of the distance of m6A peak summits to the closest given anchor point 5′ SJ (C), 3′ SJ (D), start codon (E), and stop codon (F) for nascent RNA (TNT-seq) and mRNA (m6A-seq; Schwartz et al., 2014).

See also Figure S1.

data (Schwartz et al., 2014) and called m6A peaks using the same pipeline. The majority of steady-state mRNA m6A peaks reside in the CDS (52%), 3′ UTR (28%), and 5′ UTR (12%), while only a minor fraction (4%) is intronic (Figure S1D). Almost half of the CDS-associated nascent m6A peaks reside within 100 nt upstream of the 5′ SJ, and approximately one-fifth are within 100 nt downstream of the 3′ SJ (Figure S1C). For steady-state mRNA, only 17% and 11% of the CDS peaks are within the respective intervals, suggesting a transient functional role of early m6A deposition (Figure S1D). By normalizing the number of m6A peaks to the length of the analyzed intervals and the respective input read coverage, we find that the early m6A deposition is enriched within 100 nt of the 5′ SJ exonic boundary (Figure 1A). To validate the early m6A sites, we assessed the presence of the m6A consensus DRACH motif by performing *de novo* motif search with HOMER (Heinz et al., 2010) in the regions ±150 nt around the peak summit of the top scoring peaks (score >20, n = 5,651) or in randomly generated 300-nt genomic intervals. We find a DGACH motif with a positional enrichment around the peak summit, in particular for exonic peaks (Figure 1B). We also identify three additional motifs sharing an SAG core, with a strong positional enrichment around the peak summit, especially for intronic peaks (Figure 1B). Early m6A deposition is predominant at and in close proximity to SJs (Figures 1C and 1D). In contrast, the picture is inversed around start and stop codons, with a relatively greater number of peaks in steady-state mRNA (Figures 1E and 1F). This finding led us to examine whether early m6A deposition in close proximity to SJs has an impact on splicing of RNA.

m6A Signatures Separate Distinct Intron Classes

To determine the splicing kinetics of newly transcribed RNA, we used BrU-Chase Seq as described previously (Louloupi et al., 2017; Paulsen et al., 2013). Cells were labeled with a 15-min BrU pulse and chased for 0, 15, 30, and 60 min, followed by RNA purification. To determine SE across all time points, we calculated the splicing index value θ (Mukherjee et al., 2017) (Figure 2A), yielding 13,532 introns with an extracted θ value ranging from 0 (unspliced) to 1 (fully spliced). The degree of splicing at 0 min, representing nascent RNA, is lower compared to steady-state chromatin-associated RNA (Conrad et al., 2014), indicating that nascent pre-mRNA is more efficiently captured by our approach than by chromatin fractionation (Figure 2B). Using k-means clustering, we called three clusters of distinct SE dynamics (SED; Experimental Procedures) representing 4,882 fast-, 5,702 medium-, and 2,948 slow-processed introns (Figures 2C–2F). Three representative cases are depicted in Figure 2E. We plotted the average m6A signal per nucleotide position around 5′ and 3′ SJs (Figures 2G–2H) and within length-binned introns for the three groups (Figure S2B). Strikingly, we find that fast-processed introns show greater m6A deposition at SJs, with an overall positive relationship between m6A deposited at 5′ and 3′ SJ exonic boundaries and processing efficiency (Figures 2G–2J and Figures S2A–S2C). By plotting the average frequency of m6A peak summits per nucleotide position (instead of the average m6A signal) for the three subgroups, we reach the same conclusion (Figures S2D–S2F). In contrast, slowly processed introns are associated with increased m6A deposition within the intron (Figures S2B and S2E). To address whether the position of an intron affects m6A signal and SE, we looked at the average m6A signal per nucleotide position around the 5′ and 3′ SJs of only the first and last introns (of transcripts with at least four exons), showing that the effect is independent of the position of the intron (Figures S2G–S2J).

m6A Deposition at Nascent RNA Predicts SED

To further investigate the impact of m6A deposition on nascent RNA in shaping the SED, we used a logistic regression model fit to predict fast- versus slow-processed introns (Figures 2I and 2J). We find that inclusion of the m6A at SJs as an additional parameter improves the predictive power of the model (Figure 2I), with the m6A contribution in predicting fast processing being comparable to other previously shown features, such as the 5′ and 3′ SJ sequence scores and distance to transcription start site (TSS) and transcription end site (TES) (Figure 2J) (Mukherjee et al., 2017). Intron length and internal m6A signal are significantly associated with slow processing (Figure 2J). To complement this analysis, we applied linear regression to predict SED as a continuous value (Figure S3). Again, introducing the m6A at SJs improves the correlation between predicted and measured SED (Figures S3A–S3C), further confirming the impact of early m6A deposition on RNA processing.

Intronic m6A Deposition Associates with Alternative Splicing

We assessed alternative versus constitutive splicing (by extracting the c value), as slow pre-mRNA processing has been shown to favor the occurrence of alternative splicing (Mukherjee et al., 2017) (Figure 2A). Alternative splicing events are significantly enriched in slow processed introns (odds ratio, 3.84; Fisher's exact test p value < 2.2e-16) (Figure 3A). Additionally, intronic m6A peaks are associated with upstream or downstream exon skipping approximately two times more often than expected by random chance (odds ratio, 1.7; Fisher's exact test p value < 2.2e-16), suggesting that intronic m6A deposition is involved in alternative splicing. In concurrence, the average m6A signal is greater along alternative versus constitutively spliced introns and the average m6A signal is greater at constitutive versus alternatively spliced SJ exonic boundaries (Figures 3B–3D). The overall intronic m6A, along with the intron length, are significant contributors in determining alternative splicing (Figure 3E). In contrast, m6A at SJ exonic boundaries and strong splice site consensus sequences (SJ score) ensure constitutive splicing (Figure 3E). Inclusion of m6A improves the predictive power of the model fit of constitutive versus alternative splicing (Figure 3F).

Splicing Factors Coincide with m6A Deposition

To investigate how m6A functionality in splicing is mediated, we analyzed available crosslinking immunoprecipitation sequencing (CLIP-seq) data for SRSF factors with an established role in splicing (Xiao et al., 2016). We find that both SRSF3 and SRSF10 show a high probability to have an m6A peak summit in close proximity (<250 nt) (Figures S4A and S4B), with SRSF10 showing relatively greater affinity (Figure S4C). The SAG motif core that we identify in early m6A peaks is reminiscent of the SRSF binding site motifs (Ajiro et al., 2016; Xiao et al., 2016). In addition, both SRSF3 and SFRF10 have been shown to bind near m6A, and while SRSF3 binding is augmented through interaction with YTHDC1, SRSF10 can bind independently to m6A modified regions (Xiao et al., 2016). In agreement with this observation, we find that the ratio of SRSF10/SRSF3 binding is greater at the SJ exonic boundaries for fast-processed introns and internally along within slow-processed introns (Figures S4D–S4F), in concordance with the respective relative enrichment of early m6A deposition (Figures 2G, 2H, and S2). The average ratio of SRSF10/SRSF3 binding clearly separates alternative and constitutive spliced introns (Figures S4G–S4I), most prominently along length-binned introns (Figure S4H). This result is in agreement with the observation that alternative splicing can be antagonistically regulated by SRSF10 versus SRSF3 binding (Xiao et al., 2016). These results suggest that m6A could play a role in shaping the final outcome of splicing through the recruitment of splicing factors with varying m6A affinities.

qTNTchase-Seq Identifies m6A-Marked Fast-Track RNAs

To separate direct m6A-mediated effects on RNA processing from sequence specific ones, we used qTNTchase-seq. Here, BrU-labeled RNA was isolated at 0 and 30 min chase and m6A transcripts were isolated with an m6A-specific antibody without fragmentation. Both supernatant (m6A negative transcripts) and eluate (m6A positive transcripts) were sequenced for each time point to obtain quantitative information, and we calculated the

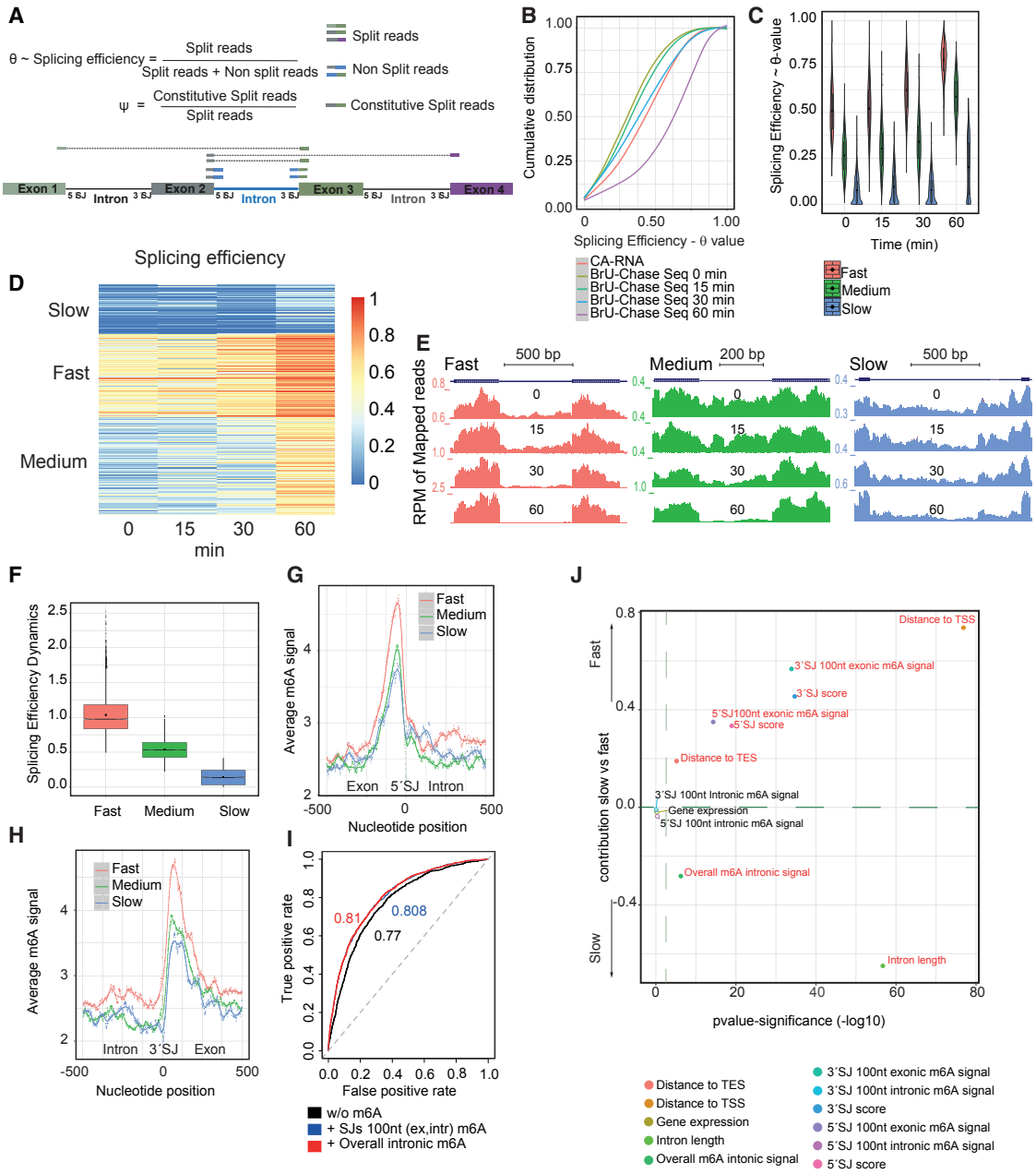


Figure 2. m6A Deposition at Nascent RNA Determines SED

(A) Definition of θ and c value.

(B) Cumulative distribution of the SE index from chromatin-associated RNA-seq (Conrad et al., 2014), BrU-Chase Seq (0, 15, 30, and 60 min).

(C) Violin plot representing the density of the SE index (θ value) distribution with embedded box and whisker plots for introns grouped on the basis of differential splicing kinetics.

(D) Heatmap showing the k-means clustering results (with k = 3) of the splicing SE index (θ value) of the 13,532 filtered introns measured for the BrU-Chase time points. Introns are clustered into fast, medium, and slow processed.

(E) UCSC genome browser views of representative cases of introns from each of the three clustering groups.

(F) Boxplot showing distribution of the SED for the fast-, medium-, and slow-processed intron groups.

(G and H) Average m6A signal per nucleotide position in a ±500-nt window around 5′ SJs (G) and 3′ SJs (H) of the filtered introns.

(I) Average receiver-operating characteristic (ROC) curve for discrimination of fast versus slow introns, including all characteristics and excluding m6A. The respective area under the curve (AUC number) is indicated.

(J) Contribution of each feature to the model fit of fast versus slow processing calculated as the coefficients from the binary logistic regression with the associated estimated significance (−log10 p value). The features with p value < 0.001 are colored red.

See also Figures S2–S4.

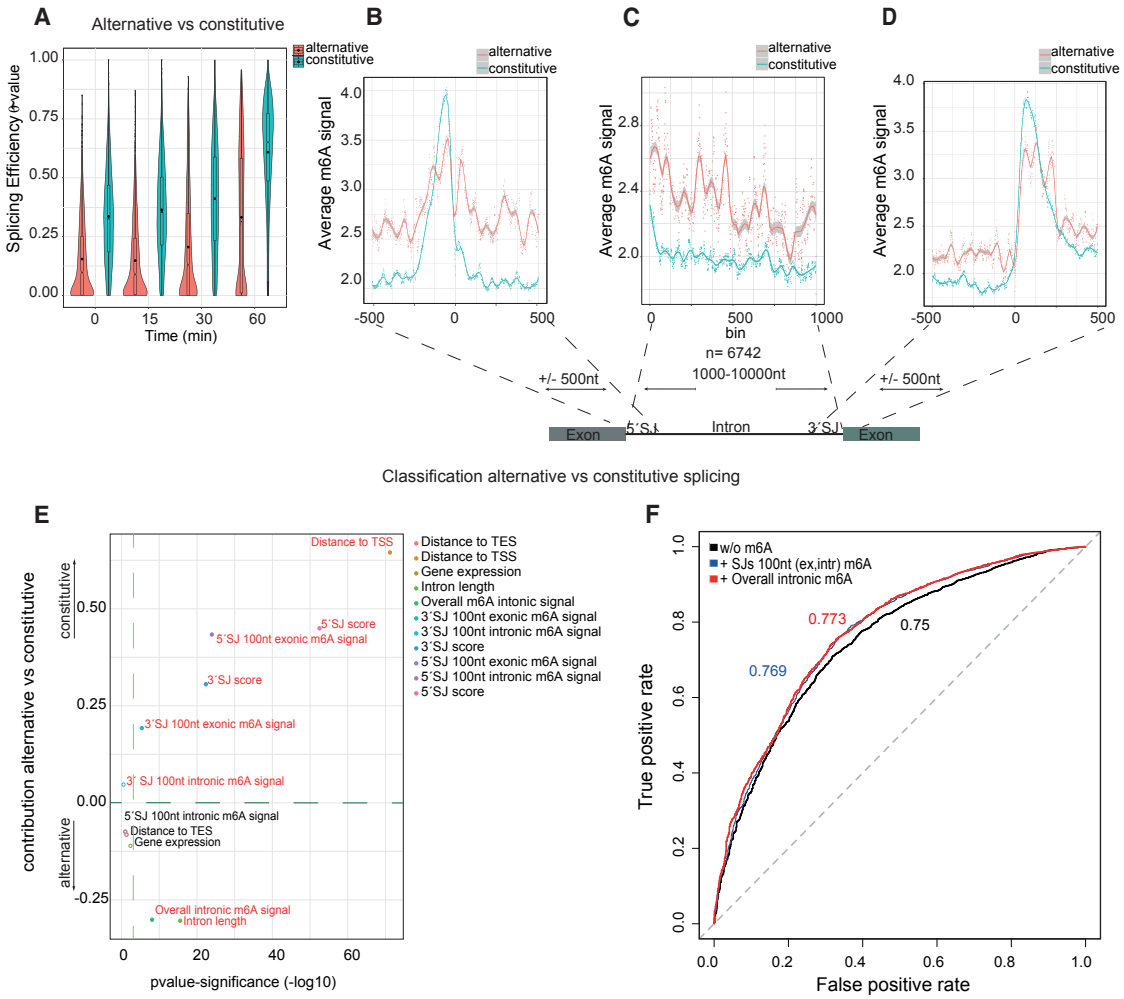


Figure 3. Intronic m6A Deposition Associates with Alternative Splicing

(A) Violin plots showing density of the distribution (with embedded box-and-whiskers plots) of θ value for introns classified as either constitutive or alternative spliced extracted from all pulse-chase time points.

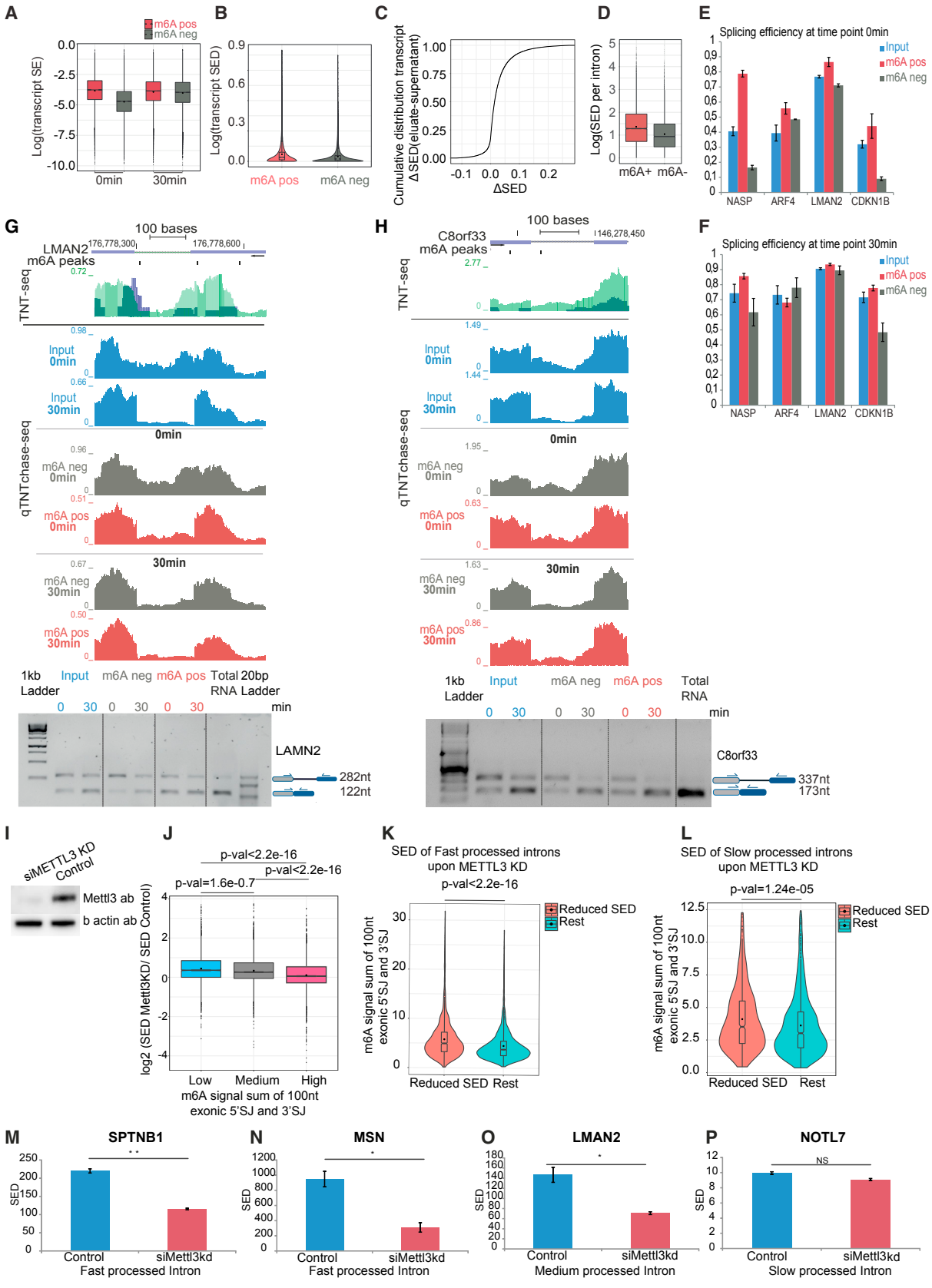
(B–D) Average m6A signal per nucleotide position in a ±500-nt window around the 5′ SJ (B) and 3′ SJ (D) and per bin (C) of 6,742 introns with length 1,000–10,000 nt. The average m6A signal is extracted separately for the two subgroups (constitutive and alternative). The lines represent LOESS curve fitting (local polynomial regression), with the 95% confidence interval shaded gray.

(E) The contribution of each feature to alternative versus constitutive splicing, calculated as the coefficients of the binary logistic regression fit with associated estimated significance (−log10 p value). Features with $p < 0.001$ are colored red.

(F) Average ROC for the logistic regression prediction of the alternative versus constitutive splicing using all features, with and without m6A data. The respective AUC number is indicated.

See also Figure S4.

m6A level per transcript (Molinie et al., 2016). On a transcriptome-wide scale, we observe a strong concordance of m6A levels between biological replicates, both for the top 25% expressed transcripts and for all transcripts with non-zero coverage (0 min: Pearson $r = 0.89$, p value < 2.2e-16; 30 min: Pearson $r = 0.91$, p value < 2.2e-16). The m6A levels do not significantly differ between 0 and 30 min chase, indicating that the overall m6A modification levels of transcripts remain the same for at least ~45 min after transcription (not shown). To follow SE, we extracted the transcript splicing index from m6A-positive and m6A-negative transcripts at 0 and 30 min chase. Within the pulse, corresponding to a 15-min window of transcription, m6A-positive transcripts show significantly higher SE than m6A-negative transcripts (Figure 4A). In addition, by measuring SED at the transcript level, we find that the m6A-positive transcripts show significantly greater processing than their m6A-negative counterparts (two tailed paired *t test* p value < 2.2e-16) (Figure 4B). Importantly, processing appears significantly enhanced for the m6A fraction of individual transcripts; ~76% show gain of SED in the m6A fraction, revealing a direct and sequence-independent impact of m6A on processing kinetics (Figure 4C). We further examined the SE locally for the 13,532



(legend on next page)

filtered introns. We find that ~14% have significantly higher SE in the m6A-positive transcripts and show a 1.26-fold enrichment over random chance to have an m6A peak in the 5′ SJ 250-nt exonic boundary (odds ratio, 1.265; Fisher's exact test *p* value = 0.0006745). In addition, individual intron loci show on average significantly higher SED in the m6A-positive versus m6A-negative transcripts (two-tailed paired t test p value < 2.2e-16) (Figure 4D).

We used qPCR to analyze the splicing kinetics of four candidate SJs that have at least one m6A peak (±250 nt). Strikingly, at time point 0, m6A-positive transcripts show higher SE than the m6A-negative transcripts (Figures 4E and 4F). We confirmed this result with semiquantitative PCR (Figures 4G and 4H).

Nascent m6A Effects Are METTL3 Dependent

To provide a direct link between RNA splicing kinetics and m6A deposition at SJs, we assessed the splicing kinetics after METTL3 knockdown (METTL3 KD) (60 min chase) (Figure 4I). The intron dataset was divided into three equal-size quantiles based on the m6A signal at 5′ and 3′ SJs (5′ and 3′ SJ 100-nt exonic intervals), and the SED was calculated. We plotted the log2 ratio of SED for METTL3 KD to control for introns with low, medium, and high m6A signal (Figure 4J). For introns with high m6A signal on both 5′ and 3′ SJs, we observe a decreased SED upon METTL3 KD for approximately half of the entries (log2 SED ratio METTL3 KD/control < 0) (Figure 4J). For introns with low and medium m6A signal (log2 SED ratio METTL3 KD/control > 0), we observe an increased SED (Figure 4J). The difference in the SED ratio (log2 METTL3 KD/control) of high m6A signal compared to low or medium is significant for both comparisons (t test p value < 2.2e-16). We then focused on fast-processed introns and plotted the m6A signal (sum of 5′ SJ and 3′ SJ 100-nt exonic area) for those that show reduced SED upon METTL3 KD versus the rest (Figure 4K). We find that the METTL3-affected introns have significantly higher m6A at the 5′ and 3′ SJ exonic boundaries. This verifies that the 5′ and 3′ SJ exonic methylation promotes fast splicing kinetics, as also shown by the logistic regression model fit (Figure 2J). We see the same but less pronounced tendency for the slow-processed introns (Figure 4L). qPCR analysis of SED for four candidates confirms the transcriptome-wide data (Figures 4M–4P).

DISCUSSION

We identify an enrichment of m6A deposition near the 5′ SJs of nascent RNA transcripts, and we show that early m6A deposition is associated with distinct RNA processing kinetics. Most importantly, we compare the processing of individual m6A-positive transcripts versus their m6A-negative counterparts, demonstrating that m6A directly controls splicing kinetics irrespectively of the underlying transcript sequence. Our findings suggest that m6A serves as a labeling signal that could be recognized by m6A reader proteins to destine methylated transcripts for specific splicing kinetics. This is in agreement with a study describing m6A methylation as a mark for selective nuclear processing, providing evidence for an m6A-dependent mRNA metabolism (Roundtree et al., 2017).

Our findings furthermore reveal that intronic m6A peaks are enriched in introns involved in alternative splicing. The m6A demethylase FTO binds mostly to introns and mediate removal of m6A. Knockout of FTO causes alternative splicing events with a preference for exon skipping, suggesting that demethylation of mRNA transcripts promotes exon inclusion under normal conditions (Bartosovic et al., 2017). Taken together, these findings suggest that intronic m6A marks that are not targeted or not yet removed by FTO mediate exon skipping, while introns involved in constitutive splicing show no enrichment in the m6A signal and most probably are targets of FTO (Bartosovic et al., 2017). In mRNAs, m6A is enriched in the consensus DRACH motif; however, not all DRACH motifs are methylated, indicating that the presence of the sequence motif alone is not enough to drive m6A deposition. FTO CLIP data show no significant enrichment of the DRACH motif (Bartosovic et al., 2017), leading us to hypothesize that early intronic m6A deposition is mostly in non-DRACH sequences where FTO can detect and eventually remove the m6A marks.

Recently, the m6A reader YTHDC1 was shown to recruit SRSF3 while competing away SRSF10. YTHDC1 binds m6A sites and promote exon inclusion (Xiao et al., 2016). In the absence of YTHDC1 and SRSF3, SRSF10 has the availability to bind to free m6A sites independently, promoting exon skipping. SRSF3 knockdown in U2OS cells has also been shown to cause exon-skipping events (Ajiro et al., 2016). Using *de*

Figure 4. qTNTchase-Seq Identifies m6A-Marked Fast-Track RNAs

(A) Boxplot representing the overall SE of methylated (m6A positive) versus non-methylated (m6A negative) transcripts at time points 0 and 30 min.

(B) Violin plots showing distribution of the transcript SED in m6A-positive and m6A-negative fractions (two-tailed Student's t test p value < 2.2e-16).

(C) Cumulative distribution of transcript SED differences between the methylated and unmethylated state (⊿SED = SED m6A positive − SED m6A negative).

(D) Boxplot displaying SED per intron in m6A-positive and m6A-negative transcripts (two-tailed paired t test p value < 2.2e-16).

(E and F) qPCR analysis of the local intronic SE of methylated versus non-methylated transcripts for 0 min (E) and 30 min (F).

(G and H) UCSC genome browser tracks of qTNTchase-seq data for LMAN2 (G) and C8orf33 (H) representing the transcript regions used for the qRT-PCR analysis. Normalized read coverage (reads per million of total number of mapped reads) tracks for input (blue), supernatant m6A negative (gray), and eluate m6A positive (pink). The upper overlay track represents the TNT-seq with purple for input and green for IP; black rectangles above represent the called m6A peaks. Below tracks for each sample are agarose gels depicting semiquantitative PCR of input, m6A-positive, and m6A-negative samples for 0 and 30 min.

(I) Western blot for METTL3 KD.

(J) Log2 ratio of SED in METTL3 KD to control for introns with low, medium, and high m6A signal at both 5′ and 3′ SJs (100-nt exonic area).

(K) m6A signal at both 5′ SJs and 3′ SJs (100 nt exonic area) for the fast-processed introns that show reduced SED in the METTL3 KD condition versus the rest (two-tailed Student's t test p value < 2.2e-16).

(L) m6A signal at both 5′ SJs and 3′ SJs (100 nt exonic area) for the slow-processed introns that show reduced SED in the METTL3 KD condition versus the rest (two-tailed Student's t test p value < 2.2e-16).

(M–P) qPCR analysis of SED for fast- (M and N), medium- (O), and slow-processed (P) introns (error bars show SD, n = 2 biological replicates (*p < 0.05 and **p < 0.01, two-tailed Student's t test).

novo motif analysis, we identify three additional motifs sharing a SAG core reminiscent of the SRSF binding site consensus, suggesting that m6A could be involved in recruiting splicing factors to control SE and alternative splicing.

The lack of strong consensus sequences at SJs of many introns may be compensated by the presence of m6A that could eventually attract splicing factors to exert their function. Our study shows that the crucial role of m6A on SED as well as on alternative splicing is position dependent. m6A deposited in intronic regions sort transcripts to a slow-track processing pathway and is associated with alternative splicing while m6A deposited at exonic boundaries of SJs sort transcripts to a fast-track processing pathway and constitutive splicing.

EXPERIMENTAL PROCEDURES

Cell Culture and BrU-Chase Sequencing

HEK293 cells were cultured in DMEM growth medium supplemented with 10% fetal bovine serum (FBS) under normal growth conditions (37°C and 5% CO_2). Cells were 70%–80% confluent before addition of BrU. BrU (−5-bromouridine, Santa Cruz Biotechnology catalog number CAS 957-75-5) was added to a final concentration of 2 mM to the medium and cells were incubated at normal growth conditions for 15 min. Cells were washed three times in PBS and either collected directly or chased in conditional medium supplemented with 20 mM uridine (Sigma catalog number U3750-25G) for 15, 30, and 60 min. RNA was purified using TRIzol following the manufacturer's instructions.

TNT-Seq

For one TNT-seq sample, ~25 150-mm plates were used for BrU labeling. RNA was labeled and isolated as described above. RNA concentration was adjusted to 2 μg/μL with nuclease-free water. 18 μL RNA was added to a thin-walled 200-μl PCR tube following the addition of 2 μL 10X fragmentation mixture (100 mM Tris-HCl [pH 7.4] and 100 mM $ZnCl_2$ in nuclease-free water). Distribution of post-fragmentation size (~100 nt) was analyzed using an Agilent 2100 Bioanalyzer with an Agilent RNA 6000 Pico kit according to the manufacturer's instructions. 400–600 μg fragmented BrU-labeled total RNA was used for each BrU immunoprecipitation (IP). BrU-RNA isolation was performed as described above. 5 μg BrU fragmented RNA was used as input for the m6A-IP buffer. An RNA-antibody-beads mixture was incubated for 2 hr at 4°C with gentle rotation in a final volume of 0.8 mL in protein low-binding tubes. Three washing steps were performed with 1X m6A-IPbuffer (1st and 2nd wash) and high-salt m6A-IP buffer (500 mM NaCl, 0.1% Igepal CA-6300, 10 mM Tris-HCl [pH 7.5]) (3rd wash). At the last wash, the protein low-binding tubes were replaced with DNA LoBind tubes. For elution, 80 μL elution buffer (1X m6A-IP buffer + 6.7 mM m6A nucleotides) was added directly on the beads, and the tubes were incubated for 1 hr with continuous shaking (1,100 rpm) at 4°C. After the second round of elution, RNA was ethanol precipitated and resuspended in 15 μL RNase-free water, and the RNA concentration was measured using the Qubit RNA HS Assay Kit as per the manufacturer's instructions.

siRNA Transfection

HEK293 cells were transfected with four different siRNAs targeting METTL3 transcript (Table S1) using HiPerFect Transfection Reagent from QIAGEN. In brief, reverse transfection was performed using 1×10^6 cells for a single 100-mm plate. Cells were seeded in a final 4 mL volume of medium without antibiotics. 12 μL transfection reagent together with siRNAs (25 nM final concentration) was incubated at room temperature (RT) in 1 mL Opti-MEM I Reduced Serum Medium after mixing for 20 min. The transfection complexes were added drop-wise into the plate. 16 hr after transfection, 5 mL cell culture medium was added to each plate. A second transfection was performed 24 hr after the first transfection. After 40 hr, 5 mL cell culture medium was added to each plate. Knockdown efficiency was analyzed with western blot (anti-METTL3 polyclonal antibody; Protein Tech catalog number 15073-1-AP). BrU-Chase Seq samples were prepared 72 hr after the first transfection. The experiments were performed in duplicate.

RNA Sequencing and Data Analysis

For the BrU- Chase Seq, the library preparation was done using the TrueSeq Stranded Total RNA Kit (Illumina). Sequencing was performed on an Illumina HiSeq 2500 instrument to obtain ~200 million reads per sample. For the TNT-Seq, 100 ng of Input BrU-labeled fragmented RNA and 100 ng of TNT-IP eluate RNA were subjected to library preparation following the TruSeq Stranded mRNA Library Preparation Kit instructions with some modifications.

DATA AND SOFTWARE AVAILABILITY

The accession numbers for the BrU-chase-seq, TNT-seq, qTNTchase-seq, and METTL3 KD BrU-chase-seq data reported in this paper are GEO: GSE92565 and GSE83561.

SUPPLEMENTAL INFORMATION

Supplemental Information includes four figures and one table and can be found with this article online at https://doi.org/10.1016/j.celrep.2018.05.077.

ACKNOWLEDGMENTS

E.N. has been funded by a postdoctoral stipend from the Alexander von Humboldt Foundation. Work in the author's laboratory is funded by the German Research Council (DFG), the Alexander von Humboldt Foundation (Sofja Kovalevskaja Award), and the Novo Nordisk Foundation (Hallas Møller Investigator) to U.A.V.Ø.

AUTHOR CONTRIBUTIONS

Conceptualization, A.L., E.N., T.C., and U.A.V.Ø.; Methodology, A.L., E.N., T.C., and U.A.V.Ø.; Investigation, A.L. and E.N.; Validation, A.L; Formal Analysis, Data Curation, Software, E.N.; Writing – Original Draft, Visualization, A.L, E.N., and U.A.V.Ø.; Writing – Review & Editing, A.L., E.N., T.C., and U.A.V.Ø.; Supervision, Funding Acquisition, U.A.V.Ø.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: January 10, 2018
Revised: April 30, 2018
Accepted: May 23, 2018
Published: June 19, 2018

REFERENCES

Ajiro, M., Jia, R., Yang, Y., Zhu, J., and Zheng, Z.-M. (2016). A genome landscape of SRSF3-regulated splicing events and gene expression in human osteosarcoma U2OS cells. Nucleic Acids Res. *44*, 1854–1870.

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.

Conrad, T., Marsico, A., Gehre, M., and Orom, U.A. (2014). Microprocessor activity controls differential miRNA biogenesis In Vivo. Cell Rep. *9*, 542–554.

Dominissini, D., Moshitch-Moshkovitz, S., Schwartz, S., Salmon-Divon, M., Ungar, L., Osenberg, S., Cesarkas, K., Jacob-Hirsch, J., Amariglio, N., Kupiec, M., et al. (2012). Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. Nature *485*, 201–206.

Fu, Y., Dominissini, D., Rechavi, G., and He, C. (2014). Gene expression regulation mediated through reversible m^6A RNA methylation. Nat. Rev. Genet. *15*, 293–306.

Harcourt, E.M., Kietrys, A.M., and Kool, E.T. (2017). Chemical and structural effects of base modifications in messenger RNA. Nature *541*, 339–346.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589.

Jia, G., Fu, Y., Zhao, X., Dai, Q., Zheng, G., Yang, Y., Yi, C., Lindahl, T., Pan, T., Yang, Y.-G., and He, C. (2011). N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. Nat. Chem. Biol. *7*, 885–887.

Ke, S., Pandya-Jones, A., Saito, Y., Fak, J.J., Vågbø, C.B., Geula, S., Hanna, J.H., Black, D.L., Darnell, J.E., Jr., and Darnell, R.B. (2017). m^6A mRNA modifications are deposited in nascent pre-mRNA and are not required for splicing but do specify cytoplasmic turnover. Genes Dev. *31*, 990–1006.

Liu, J., Yue, Y., Han, D., Wang, X., Fu, Y., Zhang, L., Jia, G., Yu, M., Lu, Z., Deng, X., et al. (2014). A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. Nat. Chem. Biol. *10*, 93–95.

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.

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.

Louloupi, A., Ntini, E., Liz, J., and Ørom, U.A. (2017). Microprocessor dynamics shows co- and post-transcriptional processing of pri-miRNAs. RNA *23*, 892–898.

Meyer, K.D., Patil, D.P., Zhou, J., Zinoviev, A., Skabkin, M.A., Elemento, O., Pestova, T.V., Qian, S.B., and Jaffrey, S.R. (2015). 5′ UTR m(6)A promotes cap-independent translation. Cell *163*, 999–1010.

Molinie, B., Wang, J., Lim, K.S., Hillebrand, R., Lu, Z.X., Van Wittenberghe, N., Howard, B.D., Daneshvar, K., Mullen, A.C., Dedon, P., et al. (2016). m(6)A-LAIC-seq reveals the census and complexity of the m(6)A epitranscriptome. Nat. Methods *13*, 692–698.

Mukherjee, N., Calviello, L., Hirsekorn, A., de Pretis, S., Pelizzola, M., and Ohler, U. (2017). Integrative classification of human coding and noncoding genes through RNA metabolism profiles. Nat. Struct. Mol. Biol. *24*, 86–96.

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.

Paulsen, M.T., Veloso, A., Prasad, J., Bedi, K., Ljungman, E.A., Tsan, Y.C., Chang, C.W., Tarrier, B., Washburn, J.G., Lyons, R., et al. (2013). Coordinated regulation of synthesis and stability of RNA during the acute TNF-induced proinflammatory response. Proc. Natl. Acad. Sci. USA *110*, 2240–2245.

Ping, X.-L., Sun, B.-F., Wang, L., Xiao, W., Yang, X., Wang, W.-J., Adhikari, S., Shi, Y., Lv, Y., Chen, Y.-S., et al. (2014). Mammalian WTAP is a regulatory subunit of the RNA N6-methyladenosine methyltransferase. Cell Res. *24*, 177–189.

Roundtree, I.A., Luo, G.-Z., Zhang, Z., Wang, X., Zhou, T., Cui, Y., Sha, J., Huang, X., Guerrero, L., Xie, P., et al. (2017). YTHDC1 mediates nuclear export of N^6-methyladenosine methylated mRNAs. eLife *6*, e31311.

Schwartz, S., Mumbach, M.R., Jovanovic, M., Wang, T., Maciag, K., Bushkin, G.G., Mertins, P., Ter-Ovanesyan, D., Habib, N., Cacchiarelli, D., et al. (2014). Perturbation of m6A writers reveals two distinct classes of mRNA methylation at internal and 5′ sites. Cell Rep. *8*, 284–296.

Slobodin, B., Han, R., Calderone, V., Vrielink, J.A.F.O., Loayza-Puch, F., Elkon, R., and Agami, R. (2017). Transcription impacts the efficiency of mRNA translation via co-transcriptional N6-adenosine methylation. Cell *169*, 326–337.e12.

Wang, X., Lu, Z., Gomez, A., Hon, G.C., Yue, Y., Han, D., Fu, Y., Parisien, M., Dai, Q., Jia, G., et al. (2014). N6-methyladenosine-dependent regulation of messenger RNA stability. Nature *505*, 117–120.

Xiao, W., Adhikari, S., Dahal, U., Chen, Y.-S., Hao, Y.-J., Sun, B.-F., Sun, H.-Y., Li, A., Ping, X.-L., Lai, W.-Y., et al. (2016). Nuclear m(6)A reader YTHDC1 regulates mRNA splicing. Mol. Cell *61*, 507–519.

Xu, C., Wang, X., Liu, K., Roundtree, I.A., Tempel, W., Li, Y., Lu, Z., He, C., and Min, J. (2014). Structural basis for selective binding of m6A RNA by the YTHDC1 YTH domain. Nat. Chem. Biol. *10*, 927–929.

Zheng, G., Dahl, J.A., Niu, Y., Fedorcsak, P., Huang, C.M., Li, C.J., Vågbø, C.B., Shi, Y., Wang, W.L., Song, S.H., et al. (2013). ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. Mol. Cell *49*, 18–29.

Cell Reports, Volume *23*

Supplemental Information

Transient N-6-Methyladenosine Transcriptome Sequencing Reveals a Regulatory Role of m6A in Splicing Efficiency

Annita Louloupi, Evgenia Ntini, Thomas Conrad, and Ulf Andersson Vang Ørom

Supplemental Experimental Procedures

Cell culture and BrU-chase Seq.

HEK293 cells were cultured in DMEM growth-medium supplemented with 10% Fetal Bovine Serum (FBS) under normal growth conditions (37°C and 5% CO_2). The day before bromouridine (BrU) labelling ~2.0 x 10^6 cells were seeded in 100 mm plates with 10ml media, one plate for each time point. Cells were 70-80% confluent before the addition bromouridine (BrU). BrU (-5-Bromouridine cat.no. CAS 957-75-5 Santa Cruz Biotechnology) was added to a final concentration of 2 mM to the media and cells were incubated at normal growth conditions for 15 min (pulse). Cells were washed three times in PBS and either collected directly (0 min chase time point) or chased in conditional media supplemented with 20 mM uridine (Sigma cat.no U3750-25G) for 15, 30 and 60 min. RNA was purified using TRIzol following manufacturer's instructions.

In the next step we followed the protocol of (Paulsen et al., 2013) with some modifications.

40ul of anti of anti-mouse IgG magnetic Dynabeads (Invitrogen) were transferred to a 1.5ml microfuge Protein Low binding tube and washed 3 times with BrU-IP 1X buffer (0.1% BSA in RNAse free PBS). After the final wash, the beads were resuspended in BrU-IP 1X buffer supplemented with SUPERase• In™ RNase Inhibitor 1:2000 together with BrdU antibody (5µg of antibody per 100 µg RNA). Antibody-beads mixture was incubated for 1 hour at room temperature with gentle rotation following 3 washes with 1X BrU-IP. 150 µg RNA was used for each BrU-IP and heated up for 4 min at 65°C prior to IP. The same amount of unlabeled total RNA was used as a negative control. 5X BrU-IP (0.5% BSA 5X PBS supplemented with SUPERase• In™ RNase Inhibitor 1:2000) was added to the RNA to a final concertation of 1X. RNA-antibody-beads mixture was incubated for 90 min at room temperature with gentle rotation in a final volume of 800 ul. The beads were washed three times with 800 ul 1X BrU-IP at room temperature. For all wash steps, with the exception of the elution step, the beads were washed for 5 min rotating then placed on a magnetic rack and the wash buffers were

discarded. At the last wash the Protein low binding tubes were replaced with DNA LoBind tubes. For elution 200 ul of Elution buffer (0.1% BSA and 25 mM bromouridine in PBS) were added directly on the beads and the tubes were incubated for 60 min with continuous shaking (1100 rpm) at 4 °C. The supernatant (eluate w/o beads) was transferred to a new tube and RNA was precipitated by adding 1/10 volumes of 3M sodium acetate (pH 5.2) and 3-4 volumes of 100% ethanol. RNA was allowed to precipitate at −80 °C overnight. RNA pellet was washed twice with 75% ethanol and resuspended in RNase-free water. RNA quality was analyzed using Agilent 2100 Bioanalyzer with an Agilent RNA 6000 Pico kit according to the manufacturer's instructions.

TNT-seq

For one TNT-seq sample ~25 150 mm plates were used for BrU labelling. RNA was metabolically labelled with BrU for 15 min and RNA was isolated as described above. RNA concentration was adjusted to 2 µg/ul with nuclease free water. 18 ul of RNA was added to thin-walled 200µl PCR tube following addition of 2 ul of 10X fragmentation mixture (containing 800 µl of RNase-free water, 100 µl of 1M Tris-HCl pH 7.4 and 100 µl 1M of $ZnCl_2$). After vortex and quick spinning the tubes were incubated in 94 °C for 3.5 min in a preheated thermal cycler block with the heated lid closed. Tubes were quickly removed from the thermocycler and placed on ice following addition of 2 µl of 0.5 M EDTA. After vortex and quick spin the RNA was collected in a tube to continue with for RNA precipitation using 1/10 volumes of 3 M sodium acetate (pH 5.2), 3-4 volumes of 100% ethanol. RNA was allowed to precipitate at −80 °C overnight. The following day tubes were centrifuged at full speed for 30 min at 4 °C. RNA pellet was washed twice with 75% ethanol and resuspended in 400-500 ul of RNase-free water. Validation of post fragmentation size (~100 nt) distribution was analyzed using Agilent 2100 Bioanalyzer with an Agilent RNA 6000 Pico kit according to the manufacturer's instructions. 400-600 ug fragmented BrU labeled total RNA was used

for each BrU-IP. BrU-RNA isolation was performed as described above. The BrU-IP recovery was approximately 0.09-0.16% of input. 4.5 ug of BrU fragmented RNA was used as input for the m6A immunoprecipitation. 40 ul of Dynabeads® Protein A (Invitrogen) per sample was transferred to a 1.5 ml microfuge Protein LoBind tube and washed 3 times with 1X m6A-IP (500 mM NaCl, 0.1% NP-40, 10 mM Tris-HCl, pH 7.5). After final wash the beads were resuspended in 800 ul 1X m6A-IP buffer supplemented with SUPERase• In™ RNase Inhibitor 1:1000. 1ug of affinity purified anti-m6A polyclonal antibody (Synaptic Systems) per 2.5 ug BrU-RNA was added to the beads and incubated for 60 min at room temperature with gentle rotation. As a negative control, we used Dynabeads® Protein A magnetic beads bound to an irrelevant IgG. Beads were washed 3 times with m6A-IP 1X buffer for 5 min on the rotator. 5 ug of BrU Fragmented RNA was used as input. RNA was heated up for 4 min at 65 °C. 5X m6A-IP buffer (50 mM Tris-HCl, 750 mM NaCl and 0.5% (vol/vol) Igepal CA-6300 supplemented with SUPERase• In™ RNase Inhibitor) was added to have the RNA in 1X m6A-IP buffer. RNA-antibody-beads mixture was incubated for 2 hours at 4°C with gentle rotation in a final volume of 0.8ml in Protein low binding tubes. Three washing steps followed using m6A-IP 1X buffer (1^{st} and 2^{nd} wash) and high salt m6A-IP buffer (500 mM NaCl, 0.1% Igepal CA-6300, 10 mM Tris-HCl, pH 7.5) (3^{rd} wash). For all wash steps, with the exception of the elution step, the beads were washed for 5 min then placed on a magnet and the wash buffers were discarded. At the last wash the Protein low binding tubes were replaced with DNA LoBind tubes. For elution 80 ul of Elution buffer (1X m6A-IP buffer + 6.7 mM m6A nucleotides) were added directly on the beads and the tubes were incubated for 1 hour with continuous shaking (1100 rpm) at 4 °C. The beads were spun down and the supernatant was transferred to a clean tube. After the second round of elution the eluted RNA was precipitated using ethanol precipitation as described above. RNA pellet was resuspended in 15 ul RNase-free water and using Qubit® RNA HS Assay Kit we measured the RNA concentration following manufacturer's instructions.

qTNTchase-seq, qPCR, RT-PCR.

RNA was metabolically labelled with BrU for 15 min and chased for 30 min as described above. RNA was purified using TRIzol following manufacturer's instructions. 200 ug total BrU labeled RNA was used as Input for the BrU-RNA isolation. After the elution step (200 ul of 0.1% BSA and 25mM bromouridine in PBS) we added 50 ul of 5X m6A-IP buffer. 4 ug (1 ug ab per 500 ng RNA) m6A ab were coupled to 40 ul Dynabeads® Protein A as described above, resuspended in 550 ul m6A-IP 1X buffer and added to the RNA mixture. RNA-antibody-beads mixture was incubated for 60 min at room temperature with gentle rotation. The supernatant was kept and RNA was isolated with TRIzol. The beads were washed 3 times for 5 min at RT (twice with low salt m6A-IP 1X buffer and last wash high salt m6A-IP 1X buffer). We eluted the RNA captured by m6A antibody by competition as described in TNT-Seq section. cDNA synthesis was performed using the same amount of RNA (10-20 ng) from all fractions (Input BrU-RNA 0 min, Input BrU-RNA 30 min chase, Supernatant m6A-negative 0 min, Supernatant m6A-negative 30 min chase, IP m6A-positive 0 min, IP m6A-positive 30 min chase). RT-PCR was performed using Q5 Hot Start High-Fidelity DNA Polymerase New England Biolabs with initial denaturation 98 °C 30s, then 32 cycles of 98 °C 10 s, 58 °C 20 s and 72 °C 55 s and final extension 72 °C 2 min. PCR products were resolved on agarose gel. Spike-in controls were *in vitro* transcribed using T7 RNA Polymerase Invitrogen following manufactures instructions. For the methylated transcripts N6-methyl-ATP (TriLink) was used in a ratio 4:1 to ATP in the *in vitro* transcription reaction. GFP and Luciferase sequences were used as template for the RNA transcription. For each qTNTchase-seq sample before m6A IP, *in vitro*–transcribed transcripts with and without m^6A modification were mixed into the samples as spike-in controls at the indicated percentage of m6A-modified to m6A-unmodified transcript (Molinie et al., 2016). For all samples after BrU-IP but before m6A-IP we added $2.5x10^7$ copies from each spike including: 0% GFP and

20% luciferase. Before the library preparation for sequencing, 1 ul of 1:2000 dilution of the universal ERCC spike-in control A (Invitrogen) was added to each fraction (100 ng).

siRNA transfection

HEK293 cells were transfected with four different siRNAs targeting METTL3 transcript (see Supplemetary Table 1) using HiPerFect Transfection Reagent from QIAGEN. In brief, reverse transfection was performed using 1 x 10^6 cells for a single 100 mm plate. Cells were seeded in a final 4 ml final volume of media without antibiotics. 12 ul of transfection reagent together with siRNAs (25 nM final concentration) were incubated at room temperature in 1 ml Opti-MEM^{TM} I Reduced Serum Media after mixing for 20 min. The transfection complexes were added dropwise into the plate. 16 hours after transfection 5 ml of cell culture media were added to each plate. 24 hours after the first transfection we performed a second hit using the same amount of transfection reagent and siRNAs as the first round. 40 hours after the first transfection 5 ml of cell culture media were added to each plate. We analyzed knock down efficiency with western blot (anti-METTL3 Polyclonal antibody, protein tech Catalog.number: 15073-1-AP) and continued with BrU-Chase Seq 72 hours after the first hit. The experiment was performed in duplicates.

Transcript m6A-level and splicing index

The m6A level per transcript from the qTNTchase-seq experiment were calculated as described in (Molinie et al., 2016). The ratio of the RNA abundance for each transcript between the eluate and the supernatant was represented by the ratio of the overlapping strand-specific RNA read counts normalized to the ratio of the reads of the ERCC RNAs. We used the log2-transformed read counts of ERCC RNAs to fit a linear regression model, computing the eluate ERCC reads as a function of the supernatant ERCC reads with a coefficient of 1 (not shown). The log2 ratio between ERCC eluate counts and supernatant counts was

indicated by the intercept of the regression formula. Only the ERCC RNAs with at least 100 read counts were used in this pipeline.

M6A level = E/(E+S*2^*intercept*)

Eluate read counts (*E*), supernatant read counts (*S*), and the intercept of ERCC regression (*intercept*)

We assessed the splicing efficiency per transcript as the ratio of the overlapping strand-specific split reads (extracted by using bedtools coverage –s –F 1.0) to all (split + non-split) reads covering the transcript.

Quantitative real-time PCR

RNA was reverse transcribed using the GoScript reverse transcription Promega A500. cDNA was quantified on an 7900HT Fast real time PCR system (Applied Biosystems) using the Go Taq qPCR Master Mix Promega (A6001). The PCR was carried out using a standard protocol with melting curve. Primers for unspliced RNA transcripts were design to span exon – intron 5' splice junction and exon – exon boundaries for spliced RNA transcripts. Splicing efficiency (SE) was determined by the ration of $2^{-CT_{spliced}}$ / ($2^{-CT_{spliced}}+2^{-CT_{unspliced}}$) for each timepoint. SED was determined by the ration of SED = 1/ ((1- SE0 min) * (1- SE60 min))

RNA sequencing and data analysis

For the BrU-Chase Seq, the library preparation was performed using the TrueSeq Stranded Total RNA Kit (Illumina). Sequencing was performed on an Illumina HiSeq 2500 instrument to obtain around 200M reads per sample. For the TNT-Seq, 100 ng of Input BrU-labeled fragmented RNA and 100 ng of TNT-IP eluate RNA were subjected to library preparation following the TruSeq Stranded mRNA Library Preparation Kit instructions with some modifications. The protocol started from the first strand synthesis step and 3X Clean-NA-Beads beads volume was used for the buffer exchange to include shorter RNA fragments.

Mapping of strand-specific reads to GRC37 genome assembly (hg19) was done using STAR (Dobin et al., 2013) and only uniquely mapped reads were kept for further downstream analyses. To extract read coverage per nucleotide position across the genome the strand-specific bed files were sorted by chromosome and start coordinate and converted into wig files with bedtools genomecov using –scale to normalize for library size. To assess the genome-wide correlation of the m6A signal from replicates, the ratio of normalized read counts per nucleotide position of IP to Eluate, rendering the m6A signal, was converted to bigWig using wigToBigWig (UCSC) and then bigWigCorrelate (UCSC) was used. To extract the m6A signal per nucleotide position in given intervals, the depth at each nucleotide position of the examined intervals (e.g. within +/- 500 bp windows around anchor points) was extracted using bedtools coverage –d –s from the m6A Input and the respective m6A IP, and then the ratio m6A IP/Input multiplied by (total number of mapped reads in the Input/ total number of mapped reads in the IP) was calculated. Then the average m6A signal was extracted at each nucleotide position from all examined entries.

m6A peak calling

We called m6A peaks based on a previously published pipeline (Ke et al., 2015; Ke et al., 2017). We first divided the genome into 20 bp non-overlapping bins with bedtools windowMaker and extracted the strand-specific read coverage from m6A Input and IP for all bins using bedtools coverageBed –s. Fisher's exact test p-value was extracted from the matrix (bin Input read counts, bin IP read counts, total number of mapped reads in the Input, total number of mapped reads in the IP) and adjusted by the Benjamini and Hochberg method to determine the false discovery rate (FDR). Only windows with a p-adjusted < 0.05 in all three replicates and fold enrichment (score) minimum four in at least two out of the three replicates were kept as significant. Adjacent significant bins were merged using bedtools mergeBed into broader peaks (finally 95 % of the peaks were in the range 20-100 nt long). In the case of

broad peaks, the peak summit is the midpoint of the 20 nt window with the maximum score, or the midpoint of the interval of merged adjacent bins sharing same maximum score within the same peak. In a few cases, a broad peak was assigned more than one summits if it contained non-adjacent windows sharing the same maximum score, finally yielding 58102 m6A peaks and 58311 peak summits. Custom scripts were written in awk programming language.

De novo motif search

De novo motif search was run using HOMER (Heinz et al., 2010) within +/-150 nt intervals around the peak summit of 5651 best scoring exonic m6A peaks (minimum fold enrichment 20) and the same number of top best intronic peaks. Control sequences were generated from the respective input sequences with the scrambleFasta.pl script. Then, *de novo* motif search was run with 'findsMotifs.pl input_sequences.fa fasta –basic –rna –len 6,7,8 –fasta scrambled_sequences'. The results were inspected in terms of enrichment, significance and the presence of common consensus sequences, with the four motifs displayed in Figure 1B being the most represented. Those were used to scan the input sequences for the presence of match occurrences using the 'dna-pattern' search tool from the RSAT suite (Medina-Rivera et al., 2015) with parameters 'search given strand only, prevent overlapping matches, origin-start, return flanking nucleotide positions 2'. Motif search was also performed in the same number of random genomic intervals as a control, generated with bedtools (–length 300 –number 5651). The matches were aligned and the logo was generated with WebLogo3 (Crooks et al., 2004) .

Splicing kinetics and predictive models

To assess splicing efficiency we extracted the splicing index value θ as in (Mukherjee et al., 2017). θ equals to the ratio of the split reads mapping to the 5' and 3' SJ of an intron divided

to the sum of split plus non-split reads (schematic representation in Figure 2A). The θ value (representing Splicing Efficiency, SE) was extracted from all pulse-chase time points, for 13,532 introns with at least five reads coverage at both 5' and 3' SJ, and used in k-means clustering with k = 3 to call three groups of distinct splicing efficiency (fast, medium and slow) (Supplementary Figure 2B). The Splicing Efficiency Dynamics metric was calculated as SED = 1/ ((1.001- θ 0 min) * (1.001- θ 60 min)) (plotted in the *log* scale for the three groups in Figure 2F). To assess constitutive versus alternative splicing we extracted the ψ value as in (Mukherjee et al., 2017). ψ is the ratio of constitutive split reads assigned to a given intron's 5' and 3' SJ to all split reads (i.e. split reads from the given intron 5' SJ to any downstream 3'SJ and from the intron's 3' SJ to any upstream 5' SJ, as depicted in Figure 2A). Therefore ψ is in the range 0 to 1 with 1 meaning 100 % constitutive splicing. We then used the ψ value extracted from the pulse-chase time point 60 min (closer to steady-state) to perform k-means clustering with k = 2 and define two clusters of introns, constitutive (n = 11836, minimum ψ 0.5294) and alternative (n = 1696, maximum ψ 0.5278). In the case of introns classified as alternative spliced (ψ < 0.5278) upstream or downstream exon skipping takes place.

The following features were used in logistic and linear regression models to predict splicing efficiency kinetics and alternative versus constitutive splicing:

The 5' and 3' splice site underlying sequence scores extracted using MaxEntScan (http://genes.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq.html); distance of the 5' SJ to the annotated transcript first start site (TSS) and of the 3' SJ to the last end site (TES); expression calculated as coverage (reads per kb) from the m6A Input RNA-seq (15 min BrU pulse) for the whole transcript interval where the intron belongs to; intron length; intron overall m6A signal extracted as the strand-specific m6A IP read coverage divided to m6A Input read coverage, normalized by (total number of mapped m6A Input reads * total number

of mapped m6A IP reads); m6A signal calculated the same way at the 5' SJ 100 nt exonic boundary, 5' SJ 100 nt intronic boundary, 3' SJ 100 nt exonic boundary and 3' SJ 100 nt intronic boundary.

To predict fast versus slow or alternative versus constitutive splicing, logistic regression was performed with R function glm (family = binomial) (all parameters apart from the sequence scores were first log scale transformed and all were then standardized). To evaluate the fitting of the model and assess discrimination, the Receiver Operating Characteristic Curve (ROC) and the area under the curve (AUC) were calculated with the R package ROCR (Sing et al., 2005). Linear regression to predict splicing efficiency using the continuous value θ (in the range 0 to 1) was performed with R function lm().

CLIP data analysis

We used CLIP data for SRF3 and SRSF10 from (Xiao et al., 2016)(GEO GSE71096). To calculate the relative SRSF10/SRSF3 binding per nucleotide position, we used the ModeScore column from the GEO submitted PARalyzer output file, which is the score of the highest signal divided to the sum value (signal+backround) and ranges from 0.5 to 1. We first extracted the coverage for each SRSF per nucleotide position in the +/-500 nt window around 5' or 3' SJ, or per bin for the length-binned introns (introns with length 1000-10,000 nt, binned into 1000 non-overlapping windows), by using bedtools coverage –s –d. Nucleotide positions with overlapping CLIP binding sites were assigned the cluster's score (ModeScore column) whereas nucleotide positions with no CLIP data overlap were assigned a pseudo-score 0.1. We then computed the ratio SRFS10/SRSF3 per nucleotide position or per bin of all analyzed loci and the metagene analysis extracting the average ratio SRFS10/SRSF3 per nucleotide position or per bin was run separately for each of the subgroups fast/medium/slow or constitutive/alternative.

Table 1: Primer Sequences. All the primers and siRNAs where purchased from Integrated DNA Technology, Inc. (IDT)

RT-PCR and qPCR primers

NAME	**Sequence**
CDKN1B unspliced Forward	AATAAGGAAGCGACCTGCAA
CDKN1B unspliced Reverse	atacgccgaaaagcaagcta
CDKN1B spliced Forward	AATAAGGAAGCGACCTGCAA
CDKN1B spliced Reverse	GGGGAACCGTCTGAAACAT
LMAN2 unspliced Forward	GTGACTGCGGATATAACTGACG
LMAN2 unspliced Reverse	ctcgccctcactcttcactc
LMAN2 spliced Forward	GTGACTGCGGATATAACTGACG
LMAN2 spliced Reverse	ATAGTGCTGCCCTGGAAGTC
NASP unspliced Forward	CATGGAGTCCACAGCCACT
NASP unspliced Reverse	tgccttaagctttccacagtc
NASP spliced Forward	CATGGAGTCCACAGCCACT
NASP spliced Reverse	GCAGATGTAGAAGGAGCAGGA
ARF4 unspliced Forward	CCTCCCTCTTCTCCCGACT
ARF4 unspliced Reverse	attgtggagaccctgccttt
ARF4 spliced Forward	CCTCCCTCTTCTCCCGACT
ARF4 spliced Reverse	TTGTCTTGCCAGCAGCATC
C8orf33 Forward	TAAGAAGAAAACGCGGAACAGG
C8orf33 Reverse	GGTGGGTTTCTGCCTCTTGA
MSN unspliced Forward	TCAAGAAGCTGAAGAGGCCA
MSN unspliced Reverse	agttcccataatcccagccc
MSN spliced Reverse	CTGTCAGCTCTGCCATTTCC

SPTBN1 unspliced Forward	CTGGATGAGCGAGCAGGAG
SPTBN1 unspliced Reverse	aagtgtgcccagggtttgaa
SPTBN1 spliced Revers	GCATAGTCCTCCACAGCTTGT
NOL7 unspliced Forward	TCCTGAAGGAGAAGAGGAAGC
NOL7 unspliced Forward	aattctccctgagccgagtt
NOL7 spliced Forward	AACGCTCCTGAAGGAGAAGA
NOL7 spliced Reverse	TCCAAAATAGTGTCTGGAAGGA

NAME siRNA Target	**Sequence 5'→3'**
Mettl3-1	5'-ACUGCUCUUUCCUUAAUA 5'-AAACAUGUAUUAAGGAAA
Mettl3-2	5'-CCAACAGUCCACUAAGGA 5'-CUGUUGUUCCUUAGUGGA
Mettl3-3	5'-AGGCAAGGAACAAUCCAU 5'-UUCAACAAUGGAUUGUUC
Mettl3-4	5'-AGCCAAGGAACAAUCCAU 5'-UUCAACAAUGGAUUGUUC
Control	NCI IDT controls

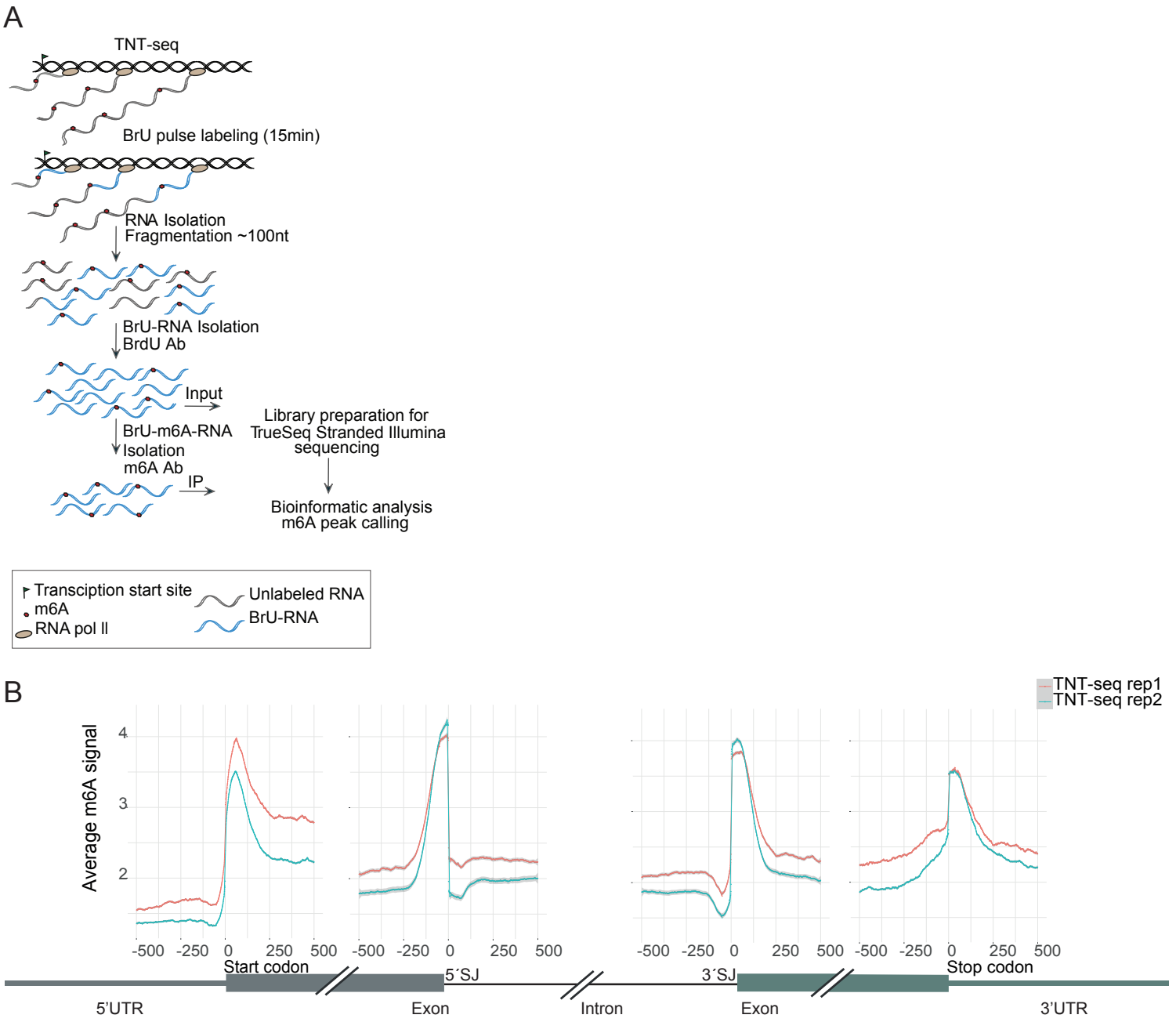


Figure S1;Related to Figure 1:m6A signal and motif analysis
(A) Schematic representation of the TNT-seq protocol. (B) The average m6A signal per nucleotide position around start and stop codons, 5' and 3' SJs is shown. (C-D) m6A peak distribution in (C) newly transcribed RNA and (D) mRNA(Schwartz et al., 2014)

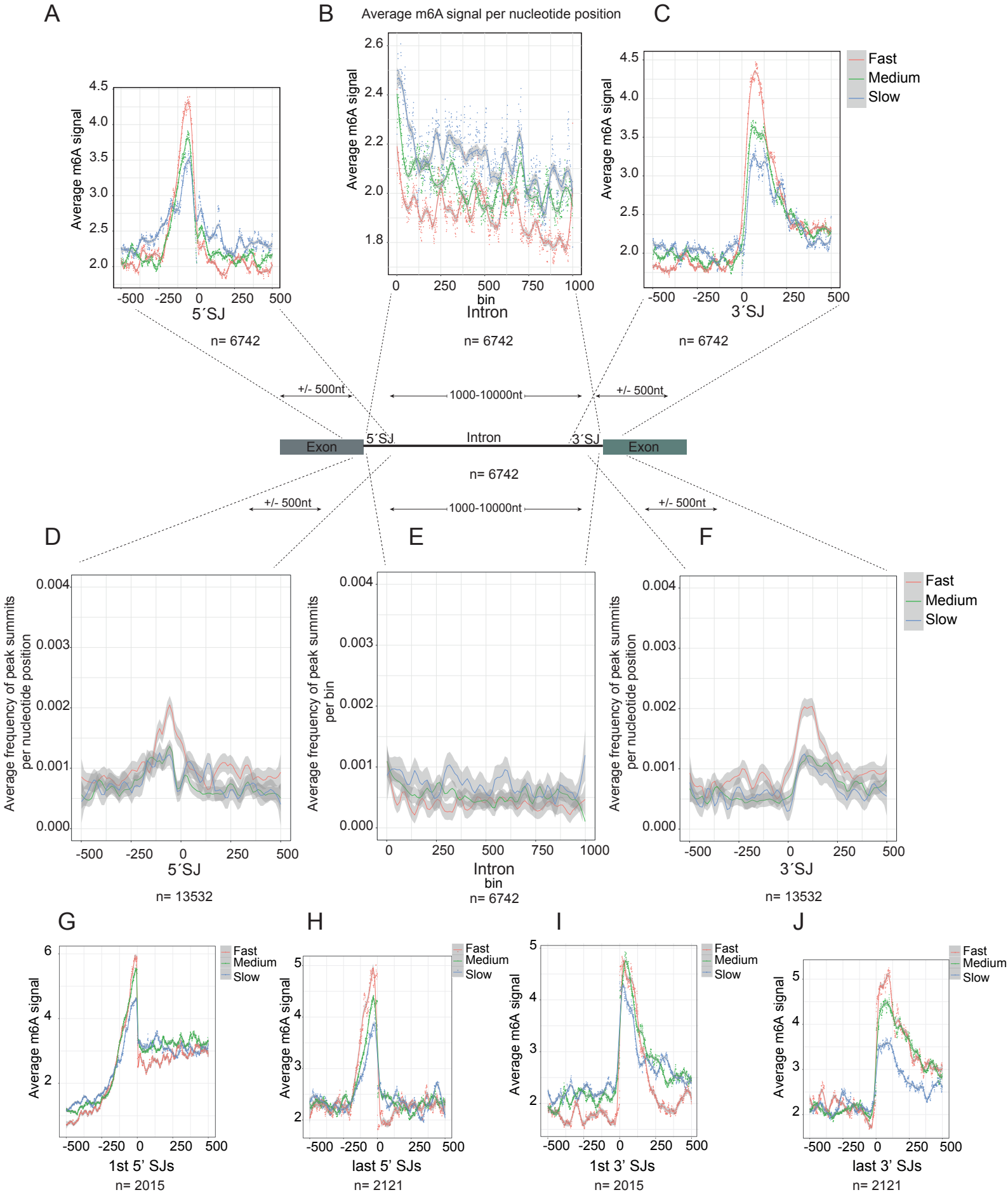


Figure S2; Related to Figure 2: M6A signal and splicing kinetics (A-C) Average m6A signal per nucleotide position in the window +/-500 nt around (A) 5´SJ and (C) 3´SJ and (B) internally per bin for 6742 introns (with length 1000-10000 nt), for fast, medium and slow processed introns. (D-F) Average frequency of m6A peak summits per nucleotide position in the window +/- 500 nt around (D) 5' SJ, (F) 3' SJ of all 13,532 filtered introns, and (E) per bin of 6722 introns 1000-10000 nt long, extracted separately for fast, medium, slow subgroups. The lines represent loess curve fitting (local polynomial regression) with the 95% confidence interval grey shaded.(G-J) Average m6A signal per nucleotide position around the 5' and 3' SJs of only the first and last introns.n = number of introns

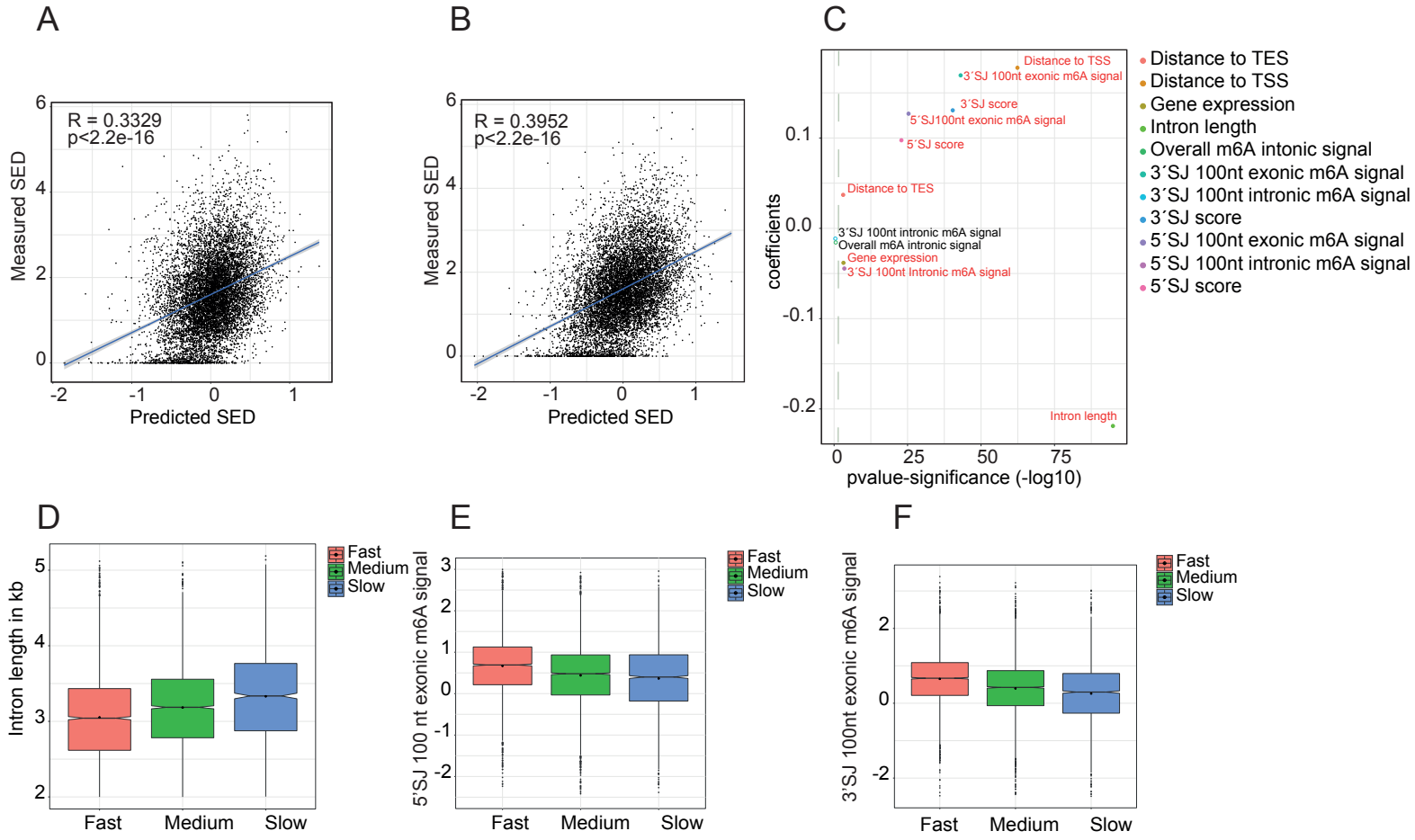


Figure S3; Related to Figure 2:Prediction of SED in a linear regression model
(A-C) Prediction of Splicing Efficiency Dynamics (SED) (as a continuous value) from (A) all features excluding m6A data (B) including m6A data using linear regression. The line indicates the linear regression fit between predicted and measured SED and the correlation value (R) with the associated significance is indicated.
(C) Plot depicting the contribution of each feature to the linear fit predicting SED. (D) Boxplots depicting the distribution of intron length for the three groups Fast (pink), Medium (Green) and Slow (blue) introns.
(E) Boxplots depicting the distribution of the m6A signal in the 100 nt exonic window boundary adjacent to the 5'SJ, for the three groups. (F) Boxplots depicting the distribution of the m6A signal in the 100 nt exonic window boundary adjacent to the 3'SJ, for the three groups.

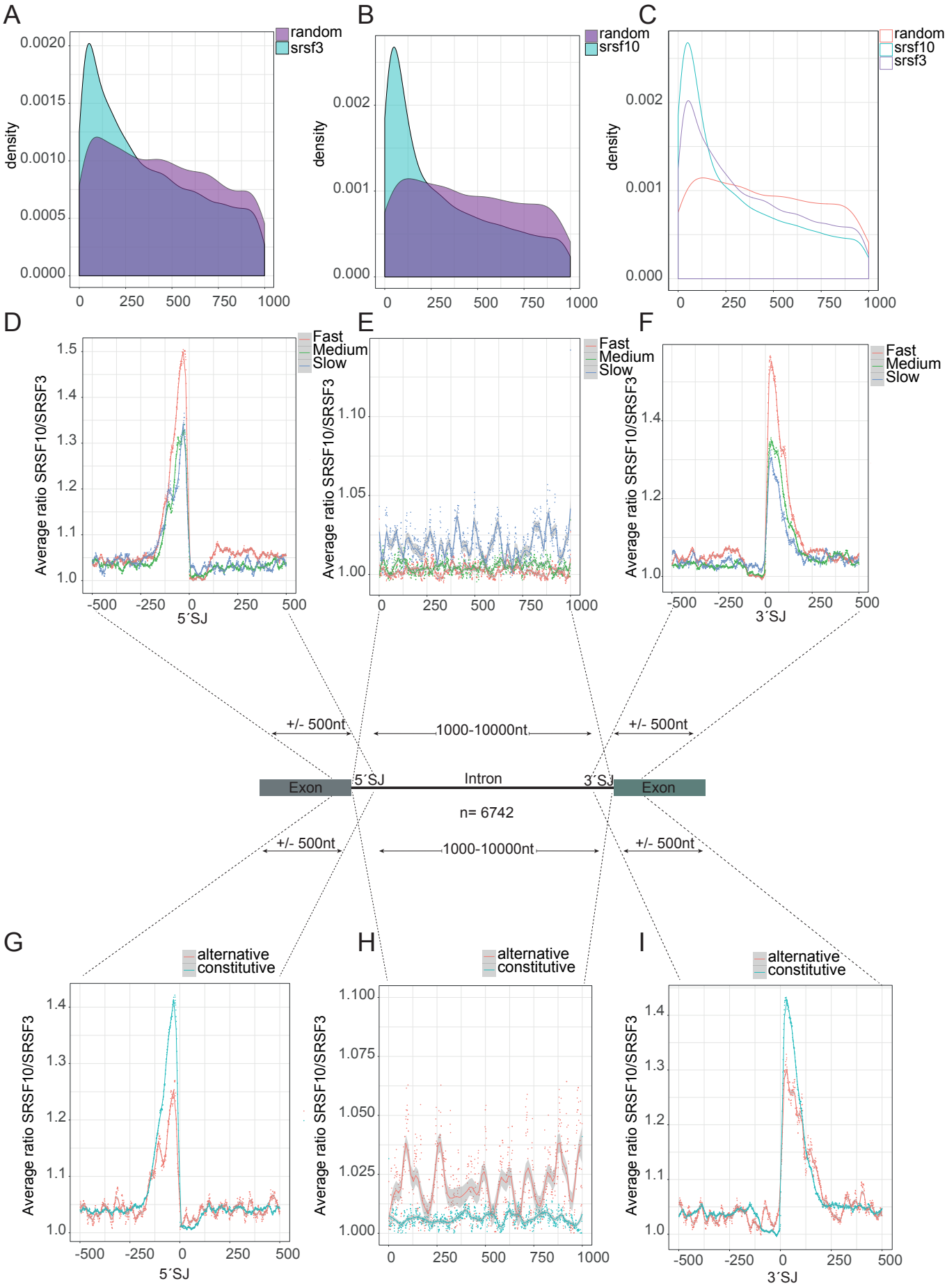


Figure S4; Related to Figure 2 and 3: Splicing factors coincide with m6A deposition
(A-C) Distribution of the interdistances of factor binding sites to closest m6A peak summit for (A) SRSF3 (B) SRSF10 and (C) overlap. As a control, distance from the midpoint of the respectively same number of randomly generated genomic intervals is also plotted. (D-F) Distribution of the average ratio SRSF10/SRSF3 binding, extracted separately for the three subgroups fast/medium/slow per nucleotide position in the window +/-500 nt around the 5'SJ (D) and 3'SJ (F), or per bin (E) for 6,742 length-binned introns (with a length 1,000-10,000 nt). (G-I) Same analysis as in (D-F) but comparing the average SRSF10/SRSF3 ratio for the two subgroups constitutive versus alternative.

References

Crooks, G.E., Hon, G., Chandonia, J.-M., and Brenner, S.E. (2004). WebLogo: A Sequence Logo Generator. Genome Research *14*, 1188-1190.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15-21.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. Molecular Cell *38*, 576-589.

Ke, S., Alemu, E.A., Mertens, C., Gantman, E.C., Fak, J.J., Mele, A., Haripal, B., Zucker-Scharff, I., Moore, M.J., Park, C.Y.*, et al.* (2015). A majority of m6A residues are in the last exons, allowing the potential for 3' UTR regulation. Genes Dev *29*, 2037-2053.

Ke, S., Pandya-Jones, A., Saito, Y., Fak, J.J., Vågbø, C.B., Geula, S., Hanna, J.H., Black, D.L., Darnell, J.E., and Darnell, R.B. (2017). m6A mRNA modifications are deposited in nascent pre-mRNA and are not required for splicing but do specify cytoplasmic turnover. Genes & Development *31*, 990-1006.

Medina-Rivera, A., Defrance, M., Sand, O., Herrmann, C., Castro-Mondragon, J.A., Delerce, J., Jaeger, S., Blanchet, C., Vincens, P., Caron, C.*, et al.* (2015). RSAT 2015: Regulatory Sequence Analysis Tools. Nucleic Acids Research *43*, W50-W56.

Molinie, B., Wang, J., Lim, K.S., Hillebrand, R., Lu, Z.-x., Van Wittenberghe, N., Howard, B.D., Daneshvar, K., Mullen, A.C., Dedon, P.*, et al.* (2016). m6A-LAIC-seq reveals the census and complexity of the m6A epitranscriptome. Nat Meth *13*, 692-698.

Mukherjee, N., Calviello, L., Hirsekorn, A., de Pretis, S., Pelizzola, M., and Ohler, U. (2017). Integrative classification of human coding and noncoding genes through RNA metabolism profiles. Nat Struct Mol Biol *24*, 86-96.

Paulsen, M.T., Veloso, A., Prasad, J., Bedi, K., Ljungman, E.A., Tsan, Y.C., Chang, C.W., Tarrier, B., Washburn, J.G., Lyons, R.*, et al.* (2013). Coordinated regulation of synthesis and stability of RNA during the acute TNF-induced proinflammatory response. P Natl Acad Sci USA *110*, 2240-2245.

Sing, T., Sander, O., Beerenwinkel, N., and Lengauer, T. (2005). ROCR: visualizing classifier performance in R. Bioinformatics *21*, 3940-3941.

Xiao, W., Adhikari, S., Dahal, U., Chen, Y.-S., Hao, Y.-J., Sun, B.-F., Sun, H.-Y., Li, A., Ping, X.-L., Lai, W.-Y.*, et al.* (2016). Nuclear m6A Reader YTHDC1 Regulates mRNA Splicing. Molecular Cell *61*, 507-519.

Microprocessor dynamics shows co- and post-transcriptional processing of pri-miRNAs RNA. 2017 Jun;23(6):892-898. doi: 10.1261/rna.060715.117. Epub 2017 Mar 1. https://doi.org/10.1261/rna.060715.117

Metabolic pulse-chase RNA labelling for pri-miRNA processing dynamics. Methods Mol Biol. 2018;1823:33-41. doi: 10.1007/978-1-4939-8624-8_3.a

https://doi.org/10.1007/978-1-4939-8624-8_3

Inhibiting pri-miRNA processing with Target Site Blockers.
Methods Mol Biol. 2018;1823:63-68. doi: 10.1007/978-1-4939-8624-8_6.

https://doi.org/10.1007/978-1-4939-8624-8_6

Selbstständigkeitserklärung
Statement of authorship

Ich erkläre ausdrücklich, dass es sich bei der von mir eingereichten Arbeit um eine von mir selbstständig und ohne fremde Hilfe verfasste Arbeit handelt.
I expressly declare that the work I have submitted was written independently and without external help.

Ich erkläre ausdrücklich, dass ich sämtliche in der oben genannten Arbeit verwendeten fremden Quellen, auch aus dem Internet (einschließlich Tabellen, Grafiken u. Ä.) als solche kenntlich gemacht habe. Insbesondere bestätige ich, dass ich ausnahmslos sowohl bei wörtlich übernommenen Aussagen bzw. unverändert übernommenen Tabellen, Grafiken o. Ä. (Zitaten) als auch bei in eigenen Worten wiedergegebenen Aussagen bzw. von mir abgewandelten Tabellen, Grafiken o.Ä. anderer Autorinnen und Autoren die Quelle angegeben habe.
I expressly declare that all sources used in the abovementioned work – including those from the Internet (including tables, graphic and suchlike) – have been marked as such. In particular, I declare that I have, without exception, stated the source for any statements quoted verbatim and/or unmodified tables, graphics etc. (i.e. quotations) of other authors.

Mir ist bewusst, dass Verstöße gegen die Grundsätze der Selbstständigkeit als Täuschung betrachtet und entsprechend geahndet werden.
I am aware that violations against the principles of academic independence are considered deception and are punished accordingly.

22-5-2019

Datum
date

Unterschrift Doktorand/in
signature of doctoral student