The role of m^{6}A RNA modification in post-transcriptional regulation of gene expression

Dissertation
zur Erlangung des Grades
Doktor der Naturwissenschaften

am Fachbereich Biologie
der Johannes Gutenberg-Universität Mainz, Deutschland

Nadine Körtel,

geb. am 03.07.1993 in Aschaffenburg

Mainz, März 2023

Dekan: Prof. Dr. Eckhard Thines

1. Berichterstatter:

2. Berichterstatter:

Tag der mündlichen Prüfung: 14th of April, 2023

Abstract

The epitranscriptome is increasingly recognized as a crucial player in post-transcriptional gene expression regulation. As the most abundant internal mRNA modification, m^{6}A is an essential regulator in almost all aspects of mRNA metabolism. Global mapping has only recently been enabled by developing RIP-seq and iCLIP-seq based methods. iCLIP is a state-of-the-art method to map RNA interactions with RNA-binding proteins in an transcriptome-wide manner. The m^{6}A-antibody based method miCLIP allowed to map m^{6}A in single-nucleotide resolution. This paved the way to further study the biological consequences of m^{6}A. Despite the important insights we gained from using these methods, they inherit several limitations and subsequent computational analysis remained challenging. Broad antibody reactivity results in high-background signal and hindered computational analysis. The high required input material made global m^{6}A-mapping exclusionary for rare samples e.g. for clinical samples or *in vivo* tissues, which are restricted in material.

In this study, we overcome these limitations and provide an improved m^{6}A-detection method combined with enhanced extensive computational analysis. First, we improved the iCLIP procedure for efficient library preparation and developed iCLIP2. We combined miCLIP with our recently improved iCLIP2 protocol, which we subsequently termed miCLIP2. Our improved protocol results in high-complexity libraries and allows to perform global mapping of m^{6}A in low-input samples. We combined the protocol with a robust computational pipeline and a machine learning classifier. Calibrating our machine learner with mESC WT and *Mettl3* KO miCLIP2 data helped to understand the characteristics of m^{6}A in miCLIP2 data. The now so-called m6Aboost predictor allows the identification of genuine m^{6}A sites without the need for a *Mettl3* KO or DRACH filtering steps. Importantly, we were able to identify m^{6}A sites outside of DRACH motifs and can apply m6Aboost for miCLIP2 data across species. Using miCLIP2 in combination with m6Aboost enables further dissection of the biological role of m^{6}A. Therefore, we found that m^{6}A accumulated towards the 5' splice site and introns are retained upon m^{6}A methylation.

The m^{6}A modification was found to act as an important player in gene expression regulation by promoting mRNA degradation. Overall, balancing gene expression is crucial and disturbance by aneuploidy can have detrimental consequences. Due to sex-chromosome evolution from an ancestral pair of autosomes, mammalian females possess two X chromosomes while males have one X and one Y chromosome. The genetic imbalance between both sexes is balanced by X-chromosome inactivation. The resulting imbalance between the single remaining X chromosome and the two active copies of autosomes is balanced by dosage compensation mechanisms. However, the existence and the mechanism behind dosage compensation is still actively debated.

This work unravelled a novel role of m^{6}A in dosage compensation. We find that m^{6}A modifications are less abundant on X-chromosomal transcripts. Due to its degrading nature, more m^{6}A results in higher degradation of autosomal transcript compared to their X-chromosomal counterparts. Hence, X-chromosomal transcripts are more stable, which is mediated by differential m^{6}A methylation numbers. Consequently, X-chromosomal transcripts are differentially affected by acute m^{6}A depletion. Importantly, we find that lower numbers of m^{6}A is internally hardcoded by reduced GGACH motifs on X-chromosomal transcripts. Taken together, we find that mammalian dosage compensation is accomplished by higher RNA stabilities of X-chromosomal transcripts, which is mediated by an epitranscriptomic RNA regulatory mechanism.

Zusammenfassung

Das Epitranskriptom (eng. epitranscriptome) wird zunehmend als entscheidender Akteur bei der posttranskriptionellen Regulierung der Genexpression anerkannt. Als häufigste interne mRNA-Modifikation ist m^6A ein wesentlicher Regulator für fast alle Aspekte des mRNA-Stoffwechsels. Eine globale Kartierung wurde erst kürzlich durch die Entwicklung von RIP-seq- und iCLIP-seq-Methoden ermöglicht. iCLIP ist eine hochmoderne Methode zur Kartierung von RNA-Interaktionsstellen eine RNA-bindenden Proteins des gesamten Transkriptoms. Die auf m^6A-Antikörpern basierende miCLIP-Methode ermöglichte die Kartierung von m^6A in Einzel-Nukleotid-Auflösung. Dies ebnete den Weg zur weiteren Untersuchung der biologischen Folgen von m^6A. Trotz der wichtigen Erkenntnisse, die wir mit diesen Methoden gewonnen haben, weisen sie mehrere Einschränkungen auf, und die anschließende computergestützte Analyse blieb weiterhin eine Herausforderung. Die breite Antikörperreaktivität führt zu einem hohen Hintergrundsignal und behindert die bioinformatische Analyse. Der hohe Bedarf an Ausgangsmaterial machte ein globales m^6A-Mapping für Proben mit niedrigem Ausgangsmaterial unmöglich, wie z. B. für klinische Proben oder *in-vivo* Proben, bei denen das Material begrenzt ist.

In dieser Studie haben wir diese Einschränkungen überwunden und eine verbesserte Methode zur m^6A-Detektion in Verbindung mit einer verbesserten, umfassenden bioinformatischen Analyse entwickelt. Wir haben ein verbessertes iCLIP2-protokoll entwickelt und miCLIP anschließend mit diesem kombiniert. Unser verbessertes Protokoll miCLIP2 führt zu hochkomplexen Bibliotheken und ermöglicht ein globales Mapping von m^6A in Proben mit geringem Ausgangsmaterial. Wir kombinierten das Protokoll mit einer robusten bioinformatischen Pipeline und einem maschinellen Lernklassifikator. Die Kalibrierung unseres maschinellen Lerners mit mESC WT- und *Mettl3* KO miCLIP2-Daten half uns, die Eigenschaften von m^6A in miCLIP2-Daten zu verstehen. Der jetzt so genannte m6Aboost-Prädiktor ermöglicht die Identifizierung echter m^6A-Stellen, ohne dass ein *Mettl3* KO- oder DRACH-Filterschritt erforderlich ist. Wichtig ist, dass wir in der Lage waren, m^6A-Stellen außerhalb von DRACH-Motiven zu identifizieren und m6Aboost für miCLIP2-Daten artübergreifend anwenden können. Die Verwendung von miCLIP2 in Kombination mit m6Aboost ermöglicht eine weitere Aufschlüsselung der biologischen Rolle von m^6A. So fanden wir heraus, dass sich m^6A in Richtung der 5'-Spleißstelle anreichert und Introns bei m^6A-Methylierung erhalten bleiben.

Es wurde festgestellt, dass die m^6A-Modifikation eine wichtige Rolle bei der Regulierung der Genexpression spielt, indem sie den mRNA-Abbau fördert. Eine ausgewogene Genexpression ist von entscheidender Bedeutung, und eine Störung durch Aneuploidie kann schwerwiegende Folgen haben. Aufgrund der Entwicklung der Geschlechtschromosomen aus einem ursprünglichen Paar Autosomen besitzen weibliche Säugetiere zwei X-Chromosomen, während männliche Tiere ein X- und ein Y-Chromosom besitzen. Das genetische Ungleichgewicht zwischen beiden Geschlechtern wird durch die Inaktivierung eines X-Chromosoms in weiblichen Säugetieren ausgeglichen. Das daraus resultierende Ungleichgewicht zwischen dem einzigen verbleibenden X-Chromosom und den beiden aktiven Kopien der Autosomen wird durch Dosierungsausgleichsmechanismen ausgeglichen. Die Existenz und der Mechanismus der Dosiskompensation werden jedoch noch immer kontrovers diskutiert.

Diese Arbeit hat eine neue Rolle von m^6A beim Dosiskompensationsmechanismus aufgedeckt. Wir haben festgestellt, dass m^6A-Modifikationen auf X-chromosomalen Transkripten weniger

häufig vorkommen. Aufgrund seiner degradierenden Natur führt mehr m^6A zu einem höheren Abbau von autosomalen Transkripten im Vergleich zu ihren X-chromosomalen Gegenstücken. Daher sind die X-chromosomalen Transkripte stabiler, was durch die unterschiedliche Anzahl von m^6A-Methylierungen vermittelt wird. Folglich sind X-chromosomale Transkripte von einer akuten m^6A-Depletion unterschiedlich betroffen. Wichtig ist, dass eine geringere Anzahl von m^6A intern durch reduzierte GGACH-Motive auf X-chromosomalen Transkripten fest kodiert ist. Zusammenfassend stellen wir fest, dass der Dosierungsausgleich bei Säugetieren durch eine höhere RNA-Stabilität von X-chromosomalen Transkripten erreicht wird, die durch einen epitranskriptomischen RNA-Regulationsmechanismus vermittelt wird.

Table of content

1. Introduction

1.1 The central dogma of molecular biology

The genetic information of a given organism is stored in deoxyribonucleic acid (DNA) and passed on to next generations. The four main components of DNA are the four chemical bases adenine (A), cytosine (C), guanine (G), thymine (T), which are attached to a sugar-phosphate backbone. The double stranded DNA is wrapped around histones and compacted to chromatin in the nucleus. The central dogma of biology describes the flow of our genetic information and is the guiding principle of gene expression (Crick *et al.,* 1958). Here, the DNA is converted to proteins through an intermediate messenger molecule, the mRNA. The central dogma entails three key stages – replication, transcription and translation. Usually, the flow of genetic information is depicted as straight-forward. However, the unidirectional interpretation of the central dogma and the assumption that every gene is finally translated into a functional protein is oversimplified (Crick *et al.,* 1970). For instance, during reverse transcription RNA is transcribed back into DNA (Temin and Mizutani *et al.,* 1970; Baltimore, 1970). Furthermore, not all genes translate to proteins. Non-coding RNAs (e.g. long non-coding RNA, small RNA, micro RNA) are not translated to proteins and can have regulatory functions themselves.

1.2 Epigenetic regulation of gene expression

Every cell in an organism contains the same set of DNA. However, multicellular organisms are defined by different cell types with distinct functions. The key to generating different cell types, despite the absence of variation in the primary set of DNA, is gene expression regulation. The cell is a dynamic environment and cellular fate is determined by selective activation or repression of genes. The selective up- or downregulation of genes defines the set of expressed genes and their degree of expression. Gene expression regulation begins with the synthesis of numerous transcription factors, which bind DNA, that can either activate or repress transcription, and ends with the regulation of correct protein translation and assembly. Hence, gene expression regulation is dynamically regulated by a plethora of mechanisms that shape cellular status and affect differentiation and development.
Essential mechanisms of gene expression regulation are epigenetic modifications (epi, on top). Without the ability to alter the primary DNA sequence, epigenetic mechanisms add an additional layer to the genetic code. A wide variety of chemical modifications on DNA, chromatin and proteins control gene expression. Thus, several dozen chemical modifications on DNA have been identified (https://dnamod.hoffmanlab.org/) (Korlach and Turner, 2012). The most widely studied epigenetic modification is the methylation of the 5^{th} carbon of cytosine (5mC) (Hotchkiss, 1948). The covalent addition of a methyl group (CH_3) is carried out

by a set of DNA methyltransferases (DNMTs) and is typically found in CpG dinucleotides (Meissner *et al.,* 2008). CpG islands are composed of CpG dinucleotide stretches and are found in roughly 70% of mammalian promoters (Saxonov and Brutlag, 2006). 5mC is a repressive epigenetic mark and was shown to affect DNA accessibility, X-chromosome inactivation, silencing of retroviral elements, tissue specific gene expression regulation and genomic imprinting (Mohandas *et al.,* 1981; Stuhlmann *et al.,* 1982; Bird, 2002). Similarly, post-transcriptional histone modification can alter the accessibility of the chromatin to the transcriptional machinery. For instance, while acetylation of histones generally make the chromatin less compact and more accessible to the transcriptional machinery, different types of histone methylations can either have repressive or activating consequences (Alhamwe *et al.,* 2018; Zentner *et al.,* 2013; Allfrey *et al.,* 1964; Pogo *et al.,* 1966; Greer and Shi, 2012). Like post-translational histone modifications, modifications of proteins (collectively described as the epiproteome) change the properties and diversify the fate and downstream function of a given protein (Dai and Rasmussen, 2007; Ramazi and Zahiri, 2021). Therefore, from the genome to the proteome, chemical modifications regulate gene expression and shape cellular fate. Resting between DNA and proteins, the intermediate messenger molecule RNA represents another layer of gene expression regulation. Finally, the field of epitranscriptomics emerged with the identification of chemical RNA modifications.

1.3 The epitranscriptome

Already in the early beginnings of RNA research, pseudouridine (Ψ), the first modification on RNA was identified (Cohn and Volkin, 1951). The high abundancy in which the modification was found, led to the misinterpretation as the fifth nucleotide (Cohn and Volkin, 1951). Further research unravelled that rather an isomer of uridine was found (Yu and Allen, 1959). During the earliest stages of RNA modification studies, research was limited to only highly expressed RNAs like ribosomal RNA (rRNA), transfer RNA (tRNA), followed by small nuclear RNA (snRNA) and small nucleolar RNA (snoRNA) (Grosjean, 2005; Grosjean, 2015; Greenberg and Penman, 1966). Thus far it remains true that the most highly modified classes of RNA are tRNAs and rRNA. While the human rRNA harbours 14 distinct types of chemical modifications at 228 nucleotide sites, $tRNA^{Try}$ harbours 17 modified sites including 12 different types of modifications (Taoka *et al*., 2018; Pan, 2018; Sprinzl and Vassilenko, 2005; Cantara *et al.,* 2011; Machnicka *et al.,* 2013). While mRNA modifications were already found in the 1970s, only recent technological advances allowed the identification of modified nucleotides (Desrosiers *et al.*, 1974; Wei (A) *et al.,* 1975; Dubin and Stollar, 1975;).

To date, over 170 different modifications have been identified. The term epitranscriptome collectively describes all modifications occurring on RNA (He, 2010; (http://modomics.genesilico.pl/sequences/, http://mods.rna.albany.edu); Cantara *et al.,* 2011; Machnicka *et al.,* 2013). Every position of the base as well as the ribose can be chemically modified by a broad spectrum of modifications. However, the vast majority of

modifications comprise methylations (Motorin and Helm, 2021; Czerwoniec *et al.,* 2009; Cantara *et al.,* 2011). The methyl group (CH_3) can be attached to either the base [e.g. N^1-methyladenosine (m^1A), 5-methylcytidine (m^5C), 7-methylguanosine (m^7G), N^6-methyladenosine (m^6A)], the ribose (e.g. 2'-O-methyladenosine) or both [e.g. 2'-O-dimethyladenosine (m^6Am)] (Kumar and Mohapatra, 2021; Dominissini *et al.,* 2016; Dunn, 1961; Wyatt, 1950; Desrosiers *et al.,* 1974; Perry and Kelley, 1974; Wei (B) *et al.,* 1975; Furuichi, 2015).

These chemical modifications are able to expand the genetic code. Moreover, the epitranscriptome has emerged as a crucial and complex mechanism for gene expression regulation. The downstream processing and interpretation of a given RNA molecule can be altered or expanded. For instance, pseudouridine increases rigidity and has effects in stabilizing conformations, thermodynamic stability and structural dynamics (Charette and Gray, 2000; Kierzek *et al.,* 2014; Davis, 1995; Meroueh *et al.,* 2000). Pseudouridylation can alter RNA-protein interactions while several cellular processes such as RNA processing, localization and stability may be affected (Wu *et al.,* 2016; Desaulniers *et al.,* 2008).

Two decades after the discovery of pseudouridine, different RNA modifications in mRNA were observed (Desrosiers *et al.,* 1974; Wei (A) *et al.,* 1975; Adams and Cory, 1975; Dubin and Stollar, 1975). However, only technological advances in recent years enabled identification and transcriptome-wide mapping of individual modified nucleotides for a few modifications. This has re-established the interest in mRNA modifications and their role in gene expression regulation (Dominissini *et al.,* 2012; Schwartz *et al.,* 2014; Carlile *et al.,* 2014; Squires *et al.,* 2012; Dominissini *et al.,* 2016; Li *et al.,* 2016; Edelheit *et al.,* 2013). To date, 12 different modifications have been observed in mRNA [m^6A, m^5C, m^1A, Ψ, m^7G, m^6Am, Inosine (I), N^4-acetylcytidine (ac^4C), 2'O-methylations (Nm), 5-hydroxymethylcytosine (hm^5C), 3-methylcytosine (m^3C) and 8-oxoguanine (o^8G)] (Desrosiers *et al.,* 1974; Dubin and Taylo *et al.,* 1975; Dominissini *et al.,* 2016; Carlile *et al.,* 2014; Lovejoy *et al.,* 2014; Schwartz *et al.,* 2014; Rottman *et al.,* 1974; Wei (B) *et al.,* 1975; Bass and Weintraub, 1988; Arango *et al.,* 2018; Delatte *et al.,* 2016; Xu *et al.,* 2017; Shan *et al.,* 2003).

Destined enzymes catalyse the modification of bases. The modifications are deposited (writing), removed (erasing), bound and interpreted (reading) by RNA-modifying proteins (RMPs). The identification and characterization of RMPs is essential to further advance our knowledge about the cellular functions and biological consequences of epitranscriptomic marks. The writing, reading and erasing of RNA modifications are dynamic and tightly regulated processes. The significance of a precise regulation of RNA modifications is underlined by the importance in health and disease. Dysregulation of RMPs lead to hypo- or hypermodification or their misinterpretation, causing detrimental consequences during differentiation and development. Around half of the so far known RMPs have been linked to human diseases like neurological disorders, cancer or cardiovascular diseases as reviewed in Jonkhout *et al.,* 2017. For instance, defects in A-to-I editing enzyme ADAR2 have been linked to the motor neuron disease, amyotrophic lateral sclerosis (ALS) (Hideyama and Kwak, 2011; Hideyama *et al.,* 2012). Furthermore, the m^5C writer enzyme NSUN2 was demonstrated to be upregulated in breast cancer (Yi *et al.,* 2017).

Nevertheless, despite the impressive diversity of RNA modifications and the important insights we have gained over the last decades, the exact role and biological function remains enigmatic for the majority of modifications.

1.4 The N^6-methyladenosine (m^{6}A) modification

Among the identified mRNA modifications one stands out most. m^{6}A was found to be conserved in all three kingdoms of life and beyond, including the identification in several viruses (Deng *et al.,* 2015; Chen (A) *et al.,* 2015; Wan *et al.,* 2015; Canaani *et al.,* 1979). The m^{6}A RNA modification, was found to modify almost all types of RNA classes, including mRNA, tRNA, rRNA, circRNA and miRNA (Shi *et al.,* 2019; Liu and Pan 2016). Interestingly, ranging from one to thirteen m^{6}A sites per transcript, it was found to be the most prevalent internal mRNA modification in polyadenylated mRNA and long non-coding RNAs in higher eukaryotes (Desrosiers et al., 1974; Zaccara *et al.,* 2019; Tegowski *et al.,* 2022; Uzonyi *et al.,* 2023). Initial m^{6}A mapping studies proposed that around 7000 transcripts harbour m^{6}A sites. However, recent studies propose, it is likely that all RNAs eventually will be methylated, but potentially in low stoichiometry, which makes identification and mapping challenging (Uzonyi *et al.,* 2023; Tegowski *et al.,* 2022).

1.4.1 Deposition of m^{6}A

The deposition of m^{6}A in mRNA is carried out primarily by a highly conserved multicomponent methyltransferase writer complex. To date, multiple subunits have been identified, as reviewed in Shi *et al.,* 2019 and Zaccara *et al.,* 2019. Building the core subunit, methyltransferase-like 3 (METTL3) forms a stable heterodimer with methyltransferase-like 14 (METTL14) in a 1:1 ratio (Bokar *et al.,* 1997; Liu *et al.,* 2014; Wang (A) *et al.,* 2016). METTL3 is the only catalytic active subunit and facilitates the deposition of a methyl group from the donor molecule S-adenosylmethionine (SAM) to the acceptor adenine. Binding to SAM is facilitated by a SAM binding domain and the transfer of the methyl group is catalysed by the conserved DPPW (Asp-Pro-Pro-Trp) sequence (Bujnicki *et al.,* 2002; Martin and McMillan, 2002). METTL14 contains an EPPL (Glu-Pro-Pro-Leu) sequence, which is less well conserved and not catalytically active (Bujnicki *et al.,* 2002). METTL14 stabilizes METTL3 conformation to increase its catalytic activity and thus, acts as an allosteric activator (Wang (A) *et al.,* 2016; Wang (B) *et al.,* 2016). Furthermore, METTL14 plays a role in substrate recognition and facilitates RNA binding (Wang (A) *et al.,* 2016; Wang (B) *et al.,* 2016). Wilms' tumor 1-associating protein (WTAP) acts as another crucial core component and interacts with the METTL3/14 heterodimer, localizing the complex to nuclear speckles and increasing the RNA-binding capacity (Liu *et al.,* 2014; Ping *et al.,* 2014). Thus, WTAP indirectly enhances the methylation activity. Subsequent studies have identified four other subunits of the m^{6}A methyltransferase complex, including Vir like m^{6}A methylation associated (VIRMA), zinc finger CCCH-type containing 13 (ZC3H13), RNA binding motif protein 15/15B (RBM15/15B) and the

E3 ubiquitin ligase Hakai (HAKAI) (Yue *et al.,* 2018; Wen *et al.,* 2018; Patil *et al.,* 2016; Horiuchi *et al.,* 2013; Bawankar *et al.,* 2021). VIRMA was shown to be critical for mediating m^{6}A methylation specifically in the 3'UTR and near stop codons (Yue *et al.,* 2018). ZC3H13 localizes and retains the m^{6}A methyltransferase complex to the nucleus (Wen *et al.,* 2018). RBM15/15B facilitates binding to specific mRNAs and the lncRNA *XIST* (Patil *et al.,* 2016). HAKAI was identified as one of the strongest interactors of WTAP and it was shown to stabilize the core components of the m^{6}A methyltransferase complex (Horiuchi *et al.,* 2013; Bawankar *et al.,* 2021).

The deposition of m^{6}A by the methyltransferase complex occurs co-transcriptionally in the nucleus in a consensus sequence-dependent manner (Slobodin *et al.,* 2017). Early studies indicated the occurrence of m^{6}A in a Rm^{6}AC [R=G/A] motif, which was further expanded to the now well-established DRm^{6}ACH [D=G/A/T, R=G/A, H=T/C/A] consensus sequence motif (Wei *et al.,* 1976; Dominissini *et al.,* 2012; Meyer *et al.,* 2012; Linder *et al.,* 2015; Ke *et al.,* 2015). DRACH motifs are very frequent in the transcriptome, occurring every ~ 57 nucleotides (Zaccara *et al.,* 2019). However, not every transcript potentially gets methylated. Moreover, in transcripts that are methylated, only very few adenines in DRACH motifs eventually were found to carry m^{6}A modifications. RNA substrate and sequence specificity could be provided by the individual subunits of the methyltransferase complex. Furthermore, recent studies reported that the histone H3 trimethylation at Lysin36 (H3K36me3) marks guide m^{6}A deposition (Huang *et al.,* 2019). It was shown that METTL14 specifically binds the epigenetic H3K36me3 mark. Therefore, the crosstalk between epigenetic marks and the m^{6}A methylation machinery could dictate which DRACH motifs will eventually be methylated. Furthermore, the co-localization of METTL3 and RNA polymerase II (RNA PolII) was observed, indicating that co-transcriptional m^{6}A methylation may be directed by interaction with the transcriptional machinery (Slobodin *et al.,* 2017).

In addition to the main mRNA methyltransferase heterodimer METTL3/METTL14, other methyltransferases were shown to induce m^{6}A methylations. In complex with TRMT112 (TRNA Methyltransferase Activator Subunit 11-2), METTL5 induces m^{6}A methylation at a single site in rRNA of the 18S subunit (m^{6}A$_{1832}$) (van Tran *et al.,* 2019). ZCCHC4 was identified as the methyltransferase that induces m^{6}A methylation in rRNA of the 28S subunit (m^{6}A$_{4220)}$ (Ma *et al.,* 2019; van Tran *et al.,* 2019). METTL16 was shown to induce m^{6}A methylation in the U6 snRNA and methylation in a small number of mRNAs (Pendleton *et al.,* 2017; Shima *et al.,* 2017; Warda *et al.,* 2017). However, the sequence motif of METTL16 (UACm^{6}AGARAA) dramatically differs from the DRACH motif.

1.4.2 Demethylation of m^{6}A

The fat mass and obesity-associated protein (FTO) was discovered as the first m^{6}A demethylase (eraser) (Jia *et al.,* 2017). Subsequently, Alkb homolog 5 (ALKBH5) was identified as the second m^{6}A eraser protein (Zheng *et al.,* 2013). Both proteins belong to the non-heme

Fe(ii)- and α-ketoglutarate-dependent dioxygenase AlkB family, which demethylate DNA and RNA nucleotides (Ougland *et al.,* 2004). The m^6A eraser proteins allow removal of the methyl group, converting m^6A back into adenine and thus, indicating that m^6A is a dynamic reversible modification. FTO was shown to demethylate m^6A on mRNA and snRNA targets (Jia *et al.,* 2017). However, it was also demonstrated that FTO-mediated m^6A demethylation might be a nonspecific reaction and the physiological relevance appears to remain questionable, as reviewed in Zaccara *et al.,* 2019 (Jia *et al.,* 2017; Mauer *et al.,* 2017). FTO cross-linking and immunoprecipitation followed by sequencing (CLIP-seq) data showed no RRACH motif enrichment and depletion of FTO in different mouse cell types and tissues did not show an increase in m^6A levels (Bartosovic *et al.,* 2017; Hess *et al.,* 2013; Mauer *et al.,* 2017; Garcia-Campos *et al.,* 2019). Interestingly, depletion of FTO in acute myeloid leukaemia (AML) revealed a 20% increase of m^6A modifications (Li (A) *et al.,* 2017). Several studies further propose that m^6Am rather than m^6A is a substrate for FTO, since higher catalytic activity is shown towards m^6Am (Mauer *et al.,* 2017; Mauer *et al.,* 2019). Moreover, m^1A modifications on tRNAs were found to be additional substrates for FTO-mediated demethylation (Wei *et al.,* 2018). FTO was shown to have diverse cellular localization. While FTO was initially reported to localize in the nucleus and partially in nuclear speckles, further studies revealed that cellular localization is cell type specific (Sanchez-Pulido *et al.,* 2007; Jia *et al.,* 2017; Wei *et al.,* 2018; Aas *et al.,* 2017; Gulati *et al.,* 2014). FTO was found to be primarily expressed in adipose and cerebral tissues and thus, taken together, FTO-mediated m^6A demethylation appears to be highly context and cell type dependent (Zheng *et al.,* 2013).
As the second identified m^6A eraser protein, ALKBH5 is located in nuclear speckles (Zheng *et al.,* 2013). In contrast to FTO, depletion of ALKBH5 led to an increase in m^6A levels without increasing m^6Am levels in human cell lines (Mauer *et al.,* 2017). Thus, ALKBH5 showed no catalytic activity towards m^6Am and m^6A is the only known substrate, to date (Mauer et al., 2017). However, also ALKBH5 seems to act cell type specific, since it was found to be enriched in testis and tissues of the female reproduction system (Zheng *et al.,* 2013). Collectively, both demethylases have cell type specific expression patterns and thus, m^6A demethylation appears to be highly tissue and context-dependent.

1.4.3 Reading of m^6A

The question how m^6A exactly exerts its function remained to be unveiled. Biophysical studies revealed that m^6A can have an effect on secondary structure of RNA duplexes and m^6A-to-U Watson-Crick base pairing (Liu and Pan 2016; Roost *et al.,* 2015; Kierzek and Kierzek, 2003). It was shown that upon methylation, RNA duplexes are destabilized (Kierzek and Kierzek, 2003). However, m^6A in an unpaired hairpin loop structure contributes to the stabilization of the secondary structure (Roost *et al.,* 2015; Kierzek and Kierzek, 2003). Hence, the effect is strongly context-dependent. Despite the biophysical properties of m^6A, the function is exerted by the recruitment of destined reader proteins that specifically recognise and bind the

modified nucleotides. The identification and characterization of m^{6}A reader proteins has provided significant insights into the biological and functional roles. The binding of reader proteins can alter the fate of methylated RNAs as elaborated below.

1.4.3.1 Direct readers of m^{6}A

The first m^{6}A readers were identified by RNA-pulldown experiments using methylated RNA probes as a bait (Dominissini *et al.,* 2012). Proteins containing an YT521-B homology (YTH) domain were identified and following studies revealed that YTH domains contain a specific m^{6}A binding pocket and are highly conserved (Zhu *et al.,* 2014; Li *et al.,* 2014). The class of direct m^{6}A readers include the three paralogs: YTH domain family 1 to 3 (YTHDF1/2/3) as well as the two YTH domain containing 1 and 2 (YTHDC1/2) proteins.

YTHDF (hereafter denoted as DF) proteins share high amino acid homology, containing a C-terminal YTH domain and a low-complexity region with several prion-like domains (Patil *et al.,* 2017). The DF proteins are exclusively found in the cytoplasm (Wang *et al.,* 2015). The low-complexity region of DF proteins enables liquid-liquid phase separation (LLPS) (Ries *et al.,* 2019). Therefore, the DF proteins are found within cytoplasmic phase-separated compartments such as processing bodies (P bodies), stress granules (SGs) and neuronal RNA granules (Ries *et al.,* 2019). In unstressed cells, the DF proteins are primarily found in P bodies. Upon stress induction, DF proteins relocalize to SGs (Ries *et al.,* 2019). Despite the high sequence homology and identical subcellular localization, several studies demonstrated individual functions for the three DF proteins. Initially, YTHDF2 was found to bind m^{6}A-modified RNAs and recruit the RNA to P bodies to promote RNA degradation (Zaccara *et al.,* 2020; Du *et al.,* 2016). YTHDF1 was found to promote translation by increasing translational efficiency of m^{6}A-modified transcripts (Wang *et al.,* 2015). Lastly, YTHDF3 was found to facilitate both, promoting degradation and translation (Li (B) *et al.,* 2017; Shi *et al.,* 2017). However, due to the high protein homology of the YTHDF1-3 proteins, it remains elusive how selectivity towards different m^{6}A methylated RNA transcripts is achieved and how different functions could be facilitated. Further studies provided conflicting evidence, which revealed similar roles in mRNA degradation and the association with the deadenylation complex CCR4/NOT for all DF proteins (Zaccara *et al.,* 2020; Du *et al.,* 2016). The CCR4/NOT complex is localized to P bodies in the cytoplasm and facilitates mRNA degradation by deadenylation (Teixeira *et al.,* 2007; Collart *et al.,* 2016). Consequently, a recent study proposed a unified function in promoting RNA degradation for the three DF proteins, indicating their redundancy (Zaccara *et al.,* 2020). Depletion of individual DF proteins led to no effect on mRNA stability. However, simultaneous depletion of all DFs led to stabilization, demonstrating a compensatory effect for the DF paralogs (Zaccara *et al.,* 2020). Moreover, for all DF paralogs, similar binding proportions for each m^{6}A site was demonstrated (Zaccara *et al.,* 2020). Nevertheless, in order to fully resolve this discrepancy, future research is required.

The YTH domain containing 1 and 2 (hereafter referred to as DC1 and DC2) proteins share no great homology in amino acid sequence beside the YTH domain. Therefore, they are not

classified as paralogs or as a protein family (Patil *et al.,* 2017). While DC1 predominantly resides in nuclear speckles in the nucleus, DC2 localizes to the cytoplasm and the nucleus (Hartmann *et al.,* 1999; Wojtas *et al.,* 2017). Similarly to the three DF paralogs, DC1 contains a low-complexity domain, indicating a potential function in phase separation (Patil *et al.,* 2016; Zaccara *et al.,* 2019). DC1 was found to mediate the subcellular localization of m^{6}A methylated transcripts by promoting nuclear export (Roundtree *et al.,* 2017). It was shown that DC1 competitively interacts with Serine and Arginine rich Splicing factor 3 (SRSF3) to mediate delivery of methylated RNA to the nuclear export receptor Nuclear RNA export factor 1 (NXF1) (Roundtree *et al.,* 2017). The interaction of DC1 with SRSF3 is in competition with Serine and Arginine rich Splicing factor 10 (SRSF10) (Xiao *et al.,* 2016). Upon interaction of DC1 with SRSF3, exon inclusion is promoted (Xiao *et al.,* 2016). The competitive interaction of DC1 with SRSF10 conversely facilitates exon skipping (Xiao *et al.,* 2016).

In contrast to the DF proteins, DC1 appears to preferentially bind non-coding RNAs (Patil *et al.,* 2016). For instance, DC1 binds the heavily modified lncRNA *XIST* and thereby mediates X-chromosome inactivation in an m^{6}A-dependent manner, which will be further discussed in chapter 1.4.5 (Patil *et al.,* 2016).

YTHDC2 resides in the nucleus and in the cytoplasm and is tissue-specific expressed in testes (Wojtas *et al.,* 2017; Hsu *et al.,* 2017). Different from the other YTH proteins, DC2 only shows weak binding affinity to m^{6}A and low overlaps in binding to m^{6}A sites throughout the transcriptome (Zaccara *et al.,* 2019; Xu *et al.,* 2015; Wojtas *et al.,* 2017; Patil *et al.,* 2016). DC2 is suggested to play a role in regulating nuclear mRNA stability by recruitment of 5'-3' Exoribonuclease 1 (Xrn1) (Kretschmer *et al.,* 2018). Moreover, interaction of DC2 and the small ribosomal subunit in proximity to the mRNA entry site is suggesting a role in m^{6}A methylated mRNA translation (Kretschmer *et al.,* 2018).

In further RNA-pull down experiments, several other proteins have been identified as potential direct m^{6}A readers. The insulin like growth factor proteins (IGF2BP) 1-3 as well as the fragile X mental retardation protein (FMRP) were proposed to directly bind m^{6}A-modified RNA (Edupuganti *et al.,* 2017, Huang *et al.,* 2018). It was shown that the RNA binding K homology (KH) domains of IGF2BP1-3 directly interact with m^{6}A, while CLIP-seq data revealed a binding consensus motif resembling the DRACH m^{6}A consensus motif. Furthermore, stability measurements demonstrated an m^{6}A-dependent effect on stabilizing transcripts upon IGF2BP binding (Huang *et al.,* 2018). However, conflicting studies show no DRACH motif enrichment for IGF2BPs (Hafner *et al.,* 2010). Similarly, m^{6}A binding via the KH domains of FMR1 was demonstrated and CLIP-seq data revealed a binding motif resembling the m^{6}A consensus motif (Edupuganti *et al.,* 2017). FMR1 was shown to mediate m^{6}A-dependent translation, mRNA stability and mRNA export (Edupuganti *et al.,* 2017; Edens *et al.,* 2019). Recent studies propose that rather than directly binding to m^{6}A, m^{6}A-dependency might be observed due to interactions between the YTH domain containing proteins and the IGF2BPs and FMRP or by m^{6}A-structural switches (m^{6}A-switches are discussed in Chapter 1.4.3.3) (Zaccara *et al.,* 2019; Zhang *et al.,* 2018; Youn *et al.,* 2018; Sun *et al.,* 2019).

1.4.3.2 Indirect readers and anti-readers of m^{6}A

Several other proteins have been identified, which might indirectly interact with m^{6}A-methylated RNA. Due to the biophysical properties of m^{6}A in altering the secondary structure of RNA, methylation could provide RNA-structure-dependent accessibility for RNA-binding proteins (RBP), which were inaccessible prior to m^{6}A modification. It was shown that local structure of ncRNAs and mRNAs were altered, which allowed access of the heterogeneous nuclear ribonucleoprotein C (HNRNPC) (Liu *et al.,* 2015). Several m^{6}A sites were found in hairpin loops, which destabilized the secondary structure and allowed binding of HNRNPC. Subsequently, the mechanism of m^{6}A-dependent RNA structural remodelling was termed m^{6}A-switch (Liu *et al.,* 2015). It was shown that m^{6}A-switch-regulated binding of HNRNPC affects the abundance and alternative splicing of target RNAs (Liu *et al.,* 2015). The concept of m^{6}A-switches potentially allows many RBPs greater access to their binding sites throughout the transcriptome. Subsequent studies identified several other proteins, where the m^{6}A-structural switch allowed increased access to their binding sites. For instance, binding of HNRNPG as well as HNRNPA2B1 were shown to rely on m^{6}A-switches. Here, splicing, alternative splicing and primary miRNA processing was facilitated in an m^{6}A-dependent manner (Liu *et al.,* 2017; Wu *et al.,* 2018; Alarcón *et al.,* 2015).
In contrast to m^{6}A-switches, where the modification allows binding of certain RBPs upon structural changes, m^{6}A sites can also repel certain proteins. Proteins, where binding is inhibited upon m^{6}A, are termed m^{6}A-anti-readers. For instance, the concept of anti-readers was demonstrated for the proteins stress granule assembly factor (G3BP) 1 and 3, EWS RNA binding protein 1 (EWSR1) and human single-stranded RNA-binding protein Pumilio 2 (hPUM2) (Edupuganti *et al.,* 2017; Vaidyanathan *et al.,* 2017). It was demonstrated that upon m^{6}A modification, the binding of these proteins was inhibited (Edupuganti *et al.,* 2017; Vaidyanathan *et al.,* 2017).

1.4.4. CLIP-seq techniques and improvements for characterization of RNA-modifying proteins (RMPs)

Writer, reader and eraser proteins are RNA-binding proteins (RBPs) and are collectively described as RNA-modifying proteins (RMPs). In order to further characterize the function of these RNA-binding proteins, it is essential to identify their target RNAs. Therefore, CLIP-seq techniques have been an important tool and are commonly used to broaden our knowledge about RMP-RNA interactions. The original CLIP-seq protocol combined UV-crosslinking with high-throughput sequencing and provided a tool to identify transcriptome-wide RNA-protein maps (Ule *et al.,* 2003). Since then, several methods have been developed based on CLIP-seq, which significantly improved transcriptome-wide mapping and enabled mapping in single-nucleotide resolution (Hafner *et al.,* 2010; König *et al.,* 2010).

In iCLIP, after UV-crosslinking of the RNA-RBP contacts only the direct contacts are preserved. Subsequently, cells are harvested and lysis is performed. RNases are used to induce RNA fragmentation, which ensures purification of only the desired RBP, since several different RBPs could bind the same RNA transcripts (Lee and Ule, 2018). Additionally, RNA fragmentation also provides a suitable size for sequencing (Lee and Ule, 2018). This is followed by bead-based immunoprecipitation of the RBP-RNA complexes accompanied by stringent washing steps to ensure no co-purification of other RBPs. The ligation of adapters provide sequences necessary for RT and subsequent polymerase chain reaction (PCR) steps. Here, the iCLIP protocol made use of a circularization step, which enabled amplification of prematurely truncated cDNA (Lee and Ule, 2018; König *et al.,* 2010). Truncations are later used for individual-nucleotide resolution mapping. The RNA-RBP complexes are visualized using radioactive labelling of the 5'ends followed by SDS-PAGE and membrane transfer. Proteinase K treatment digests the RBP and leaves a small peptide due to the covalent bond formed by UV-crosslinking. Upon RT, the reverse transcriptase truncates at the crosslinking site and later allows the mapping in individual-nucleotide resolution. During iCLIP, RT primers were introduced containing additional sequences. Unique molecular identifier (UMIs) and barcodes were introduced, which allowed the multiplexing followed by demultiplexing of several replicates (König *et al.,* 2010). Moreover, during computational analysis, PCR artefacts and PCR duplicates can be distinguished and subsequently will be removed. The cDNA is recovered by gel-based clean-up systems and excess adaptors and RT oligos are removed. Subsequently, the cDNA is amplified before subjecting the libraries to high-throughput sequencing.

Photoactivatable-ribonucleoside-enhanced crosslinking and immunoprecipitation (PAR-CLIP) is based on the incorporation of photoactivatable nucleotide analogs to the RNA and UV-crosslinking at 365 nm (Hafner *et al.,* 2010). The nucleotide analogs contain an exocyclic thione group, 4-thiouridine (4SU) or 6-thioguanosine (6SG), which increases the photoreactivity. Upon reverse transcription, the nucleotide analog 4SU induces characteristic T-to-C transitions and the enrichment upon identification after high-throughput sequencing indicate RNA-RBP interaction sites (Hafner *et al.,* 2010; Hafner *et al.,* 2021). However, the pre-incubation of cells with 4SU or 6SG could cause cellular toxicity and induce stress (Huppertz *et al.,* 2014).

Based on the iCLIP protocol, infrared CLIP (irCLIP) an enhanced CLIP (eCLIP) used the truncation-read information but introduced further alternative steps to improve library preparation steps (Zarnegar *et al.,* 2016; Van Nostrand *et al.,* 2016). Instead of using radioactive labelling for visualization, irCLIP uses an infrared dye attached to the adaptor. Moreover, a bead-based clean-up system was introduced to reduce material loss during the gel clean-up step. A pre-PCR amplification step enables the reduction of PCR cycles and duplication rates. eCLIP omits the circularization step and introduces two separately ligated adapters in order to save time and to reduce costs. Similar to irCLIP, a bead-based clean-up strategy is used.

1.4.5 m^{6}A detection methods

1.4.5.1 Transcriptome-wide mapping methods

Next to the characterization of RMPs, the identification of individual modified nucleotides is essential to understand the biological relevance and function of m^{6}A. After the discovery of m^{6}A, only a few sites have been mapped, which were all found within the consensus sequence RRm^{6}ACH (Horowitz *et al.,* 1984; Kane and Beemon, 1985; Harper *et al.,* 1990). Since then, significant technological advances have been made to map m^{6}A in a transcriptome-wide manner. In the following section, commonly used techniques will be explained in detail.

1.4.5.1.1 Antibody-based m^{6}A sequencing methods

MeRIP-Seq and m^{6}A seq

In 2012, two independent studies established the RIP-seq based methods m^{6}A-seq and MeRIP-seq (m^{6}A-specific methylated RNA immunoprecipitation with next generation sequencing) (Dominissini *et al.,* 2012; Meyer *et al.,* 2012). Both methods rely on an anti-m^{6}A antibody to capture and enrich for m^{6}A-modified transcripts. The antibody is incubated with fragmented RNA followed by RNA immunoprecipitation. Subsequent high-throughput next-generation sequencing allowed the transcriptome-wide mapping of m^{6}A. Thus, the first human and mouse methylomes were generated. However, these methods only allowed mapping of m^{6}A in a ~50-200 nucleotide (nt) window (Dominissini *et al.,* 2012; Meyer *et al.,* 2012). These windows could contain several DRACH motifs and thus, multiple m^{6}A sites within a peak could not be distinguished. Only the approximate location of m^{6}A could be given and individual-nucleotide resolution could not be provided. Nevertheless, important insights into the distribution of m^{6}A were gained. m^{6}A was found in ~7000 transcripts (Dominissini *et al.,* 2012; Meyer *et al.,* 2012). The identified m^{6}A sites were found to cluster predominantly around the stop codon, in the 3' UTR and within long internal exons (Dominissini *et al.,* 2012; Meyer *et al.,* 2012). However, limited resolution required further improvements to precisely locate m^{6}A.

miCLIP and m^{6}A-CLIP

Significant improvements by two independent studies in 2015 led to greater resolution in m^{6}A detection (Ke *et al.,* 2015; Linder *et al.,* 2015). Here, the principles of UV-crosslinking employed by iCLIP was implemented and miCLIP and m^{6}A-CLIP (m^{6}A-individual nucleotide resolution UV-crosslinking and immunoprecipitation) were developed (König *et al.,* 2010; Ke *et al.,* 2015; Linder *et al.,* 2015). In brief, purified RNA is incubated with an m^{6}A-specific antibody followed by UV-crosslinking. Upon UV-crosslinking, a covalent bond forms between the m^{6}A and the antibody. After subsequent digestion of the antibody using proteinase K, a peptide adduct

stays attached to the RNA, which will induce truncations or transitions after reverse transcription (RT). High-throughput sequencing allows the computational identification of the transitions and truncations. Therefore, these methods allowed mapping of m^{6}A in individual-nucleotide resolution. However, also these methods suffer from certain unavoidable limitations. Different commercially available antibodies differ in m^{6}A affinity and induce different patterns of truncations and different types of mutations (Linder *et al.,* 2015). The very similar RNA modification m^{6}Am is additionally recognized (Schwartz *et al.,* 2013; Linder *et al.,* 2015). Furthermore, considering that the primary epitope can expand from the methylgroup to the modified nucleobase (here adenine), all adenines potentially bind the antibody (Helm *et al.,* 2019). Therefore, non-m^{6}A methylated bases compete with m^{6}A sequences, although with lower affinities (Slama *et al.,* 2019). This was demonstrated using enrichment factor measurements, reporting only a 4-10 fold enrichment for m^{6}A (Slama *et al.,* 2019). This is demonstrating the limited antibody-specificity that commonly results in high background signal (Helm *et al.,* 2019). Thus, background noise in miCLIP data is globally observed in the field. Another important drawback of miCLIP is the high-required amount of input material, making miCLIP experiments exclusionary for rare and limited input samples. Nevertheless, miCLIP/m^{6}A-CLIP experiments paved the way to expand our knowledge about the transcriptomic distribution and biological function of m^{6}A.

1.4.5.1.2 Enzyme-based m^{6}A-detection methods

MAZTER-seq/m^{6}A-REF-seq

Acknowledging the need for an antibody-free method, two groups developed m^{6}A-mapping methods, which employ the endoribonuclease MazF (Zhang *et al.,* 2019; Garcia-Campos *et al.,* 2019). MazF is an m^{6}A-sensitive enzyme that cuts RNA in an ACA sequence motif, but not in m^{6}ACA sequences (Imanishi *et al.,* 2017). Thus, all unmodified ACA sites within a transcript are cut by MazF. Both approaches treat the mRNA with MazF, leaving ACA sites at the 5' end and the 3' end of the fragments. All other ACA sites within the fragment indicate uncut, and therefore, m^{6}A sites. High-throughput sequencing followed by computational analysis allows the identification of cut versus uncut ratios. This allows quantification of m^{6}A sites and thus, stoichiometry information about m^{6}A sites can be provided. Comparison of digested mRNA in control condition versus m^{6}A-depleted condition allows precise mapping in a high-confidence manner. The depletion of m^{6}A is achieved either by FTO-guided demethylation reactions or by depletion of the main methyltransferase METTL3. Despite this great advantage of MAZTER-seq and m^{6}A-REF-seq, also these methods suffer from certain limitations. The need of an m^{6}A-depleted condition by *METTL3* knock out (KO) limits usage in several the cell types, since *METTL3* knock out generation can lead to lethality in cells and organisms. Furthermore, the authors report the requirement of a >10% decrease of m^{6}A modification upon FTO demethylation reaction (Capitanchik *et al.,* 2020; Zhang *et al.,* 2019). Most importantly, since MazF cuts in an ACA context, only ~25% of modified sites in yeast and ~16% in mammals are detectable and quantifiable (Zhang *et al.,* 2019, Garcia-Campos *et al.,* 2019). Thus, full-

transcriptome mapping of m^{6}A cannot be provided by MAZTER-seq or m^{6}A-REF-seq. Further studies recently revealed that over ~50% of m^{6}A sites reported by MAZTER-seq were found to be false-positive sites introduced by RNA secondary structure and the sequence bias of MazF (Zhang (A) *et al.,* 2021).

1.4.5.1.2 Metabolic-labelling m^{6}A-detection methods

m^{6}A-label-seq

m^{6}A is formed by attaching a methyl group to an unmodified adenosine. SAM is the cofactor and acts as the donor to provide the methyl group. Therefore, starting from the m^{6}A biogenesis process, the metabolic labelling method m^{6}A-label-seq was developed. Here, cells are fed with the S-adenosyl methionine analogue Se-allyl-l-selenohomocystein, which substitutes the methyl group from the SAM donor with the allyl group (Shu *et al.,* 2020). Upon methylation reaction, sites that are supposed to get m^{6}A-modified are now N^6-allyladenosines (a^{6}A) modified sites. After iodine-induced cyclization, the reverse transcriptase induces misincorporations upon reverse transcription. Subsequent high-throughput sequencing and computational analysis allows transcriptome-wide identification of m^{6}A. However, the authors report moderate cellular stress induced upon Se-allyl-l-selenohomocystein feeding (Shu *et al.,* 2020). Moreover, the labelling yield is low and the labelling time remains to be improved (Shu *et al.,* 2020).

1.4.5.1.3 Chemical-assisted m^{6}A detection methods

m^{6}A-SEAL-seq

Upon demethylation reaction of m^{6}A-mediated by FTO, FTO oxidizes m^{6}A twice. The first oxidation leads to N^6-hydroxymethyladenosine (hm^{6}A) and further oxidation generates N^6-formyladenosine (f^{6}A). The demethylation products can be further processed by biotinylation and subsequent purification can be performed. Therefore, m^{6}A-SEAL-seq was developed, a FTO-assisted m^{6}A-selective chemical labelling method (Wang (A) *et al.,* 2020). Dithiothreitol (DTT) is used to modify the unstable hm^{6}A to a more stable N^6-dithiolsitolmethyladenosine (dm^{6}A). First, FTO is used to generate the unstable intermediate hm^{6}A followed by DTT-mediated thiol-addition to generate the more stable dm^{6}A. Biotin was added using the thiol-reactive biotin reagent MTSEA-biotin. After subsequent streptavidin pulldown and DTT cleavage, high-throughput sequencing is used to detect m^{6}A sites (Wang (A) *et al.,* 2020). However, the resolution of m^{6}A-SEAL-seq is limited to ~200nt, which is comparable to MeRIP-seq approaches and thus, is unable to provide single-nucleotide resolution.

GLORIseq

Glyoxal and nitrite-mediated deamination of unmethylated adenosine (GLORI-seq) was recently developed (Liu *et al.,* 2022). GLORI-seq allows absolute quantification of m^{6}A at single-nucleotide resolution in a transcriptome-wide manner. Nitrous acid was reported to

cause deamination, leading to G-to-xanthosine (X), A-to-I and C-to-U conversions (Shapiro and Pohl, 1968; Schuster and Wilhelm, 1963). Subsequently, glyoxal was used to protect G from deamination. The authors report that they achieve A-to-I conversions of ~99% of unmethylated adenosines while m^6A sites stay intact. The reaction is followed by high-throughput sequencing, where A converted to inosine is read as G during reverse transcription. False-positives could derive from endogenously present inosines. However, the high copy numbers of RNA could potentially buffer these effects. Nevertheless, false-positives could still be present for lowly expressed genes and m^6A sites below a 10% methylation level can not be reported (Liu *et al.,* 2022). Other adenosine modifications such as m^6Am and m^1A can not be distinguished by gylaoxal and nitrite-mediated deamination. While m^6Am can be removed due to the distinct occurrence in transcription start sites, m^1A can be distinguished due to m^1A-induced stop signals during reverse transcription and therefore is not leading to false-positives (Liu *et al.,* 2022). GLORI-seq identified over 176,000 m^6A sites in HEK293T cells with a median methylation level of ~40%. In accordance with other recently developed methods, GLORI-seq confirms that the number of m^6A sites have been greatly underestimated in the field (Liu *et al.,* 2022; Uzonyi *et al.,* 2023).

1.4.5.1.4 Fusion-domain m^6A-detection methods

DART-seq

In order to provide an *in vivo* method not relying on m^6A-specific antibodies, the deamination adjacent to RNA modification targets (DART-seq) method was developed. DART-seq utilizes the cytidine-deaminase APOBEC1 (Apolipoprotein B mRNA editing enzyme catalytic subunit 1), which induces C-to-U editing (Navaratnam *et al.,* 1993; Meyer, 2019). Fusing APOBEC1 to the m^6A-binding YTH domain in cells allows C-to-U editing of m^6A-adjacent cytidines, which can be detected by high-throughput sequencing (Meyer, 2019). DART-seq allowed mapping of m^6A in very low input material (10 ng of total RNA) and can be utilized for single-cell m^6A detection (Tegowski *et al.,* 2022). However, this method could possibly induce sequence preferences. APOBEC1 alone induces C-to-U deamination preferentially in the 3' UTR. Thus, comparing APOBEC1 control with APOBEC1-YTH could lead to high false-negative rates (Capitanchik *et al.,* 2020). High false-positive rates may also derive from off-targets of APOBEC1 or the fused YTH-domain.

1.4.5.1.3 Direct RNA sequencing m^6A-detection methods

Nanopore sequencing

A new technology was recently developed by Oxford Nanopore Technologies for direct RNA sequencing. Direct RNA sequencing methods make use of a nanoscale pore embedded into a membrane. Here, single-stranded RNA or DNA can pass through the pore in a nucleobase-by-

nucleobase-based manner (Deamer *et al.,* 2016). In contrast to next-generation sequencing methods, direct RNA sequencing omits the RT step to produce cDNA, which is required by the Illumina and similar sequencing platforms.

Pore-based sequencing methods utilize the ionic current that differ for each individual nucleotide passing through the pore. It was demonstrated that modified nucleotides differ from unmodified nucleotides in their ionic current and pore dwell time, enabling their identification (Workman *et al.,* 2019; Smith *et al.,* 2019). The first transcriptome-wide mapping of m^{6}A by direct RNA sequencing using the Oxford Nanopore technology was performed in 2019 (Liu *et al.,* 2019). In order to effectively distinguish m^{6}A sites from unmodified adenosines or different modifications, the authors trained a machine learning model, termed EpiNano (Liu *et al.,* 2019). Following studies trained other machine learning models (MINES [m^{6}A Identification using Nanopore Sequencing] and NanoCompore) and validated identified sites using METTL3-depleted conditions (Lorenz *et al.,* 2020; Leger *et al.,* 2021). Direct RNA sequencing methods allow insights into the stoichiometry of m^{6}A modifications and could potentially allow mapping of multiple different modifications at once. However, certain drawbacks are still to overcome. The accuracy of m^{6}A sites for transcripts with low sequencing depth remains limited as reviewed in Capitanchik *et al.,* 2020.

1.4.5.2 Individual m^{6}A site detection

Next to transcriptome-wide m^{6}A detection methods, methods for m^{6}A detection of individual nucleotides have been developed. These methods can provide stoichiometric information about individual m^{6}A nucleotides and allow reliable validation opportunities for m^{6}A sites that have been mapped by transcriptome-wide mapping methods.

SCARLET

Site-specific cleavage and radioactive labelling followed by ligation-assisted extraction and thin-layer chromatography (SCARLET) is generally depicted as the gold standard method for detecting individual m^{6}A sites (Liu and Pan, 2015). A complementary 2'-OMe/2'-H chimeric oligonucleotide is used to guide RNase H in order to cleave the 5'-site of the desired site. This is followed by radioactive labelling and splint-ligation of the single-stranded DNA (ssDNA) (Liu and Pan, 2015). Subsequent RNase T1/A digestion followed by thin-layer chromatography allows to distinguish adenosine and m^{6}A (Liu and Pan, 2015). SCARLET provided the first method to precisely detect m^{6}A and the fraction of methylated sites within specific positions. Despite this great advantage, SCARLET is very time-consuming, laborious and can not be subjected to high-throughput sequencing or used for extensive applications (Wei *et al.,* 2017; Zhang (B) *et al.,* 2021).

SELECT

The single-base elongation- and ligation-based qPCR method SELECT makes use of two m^{6}A-sensitive enzymes followed by qPCR (Xiao *et al.,* 2018). First, two probes are designed, which are annealed to the candidate transcript, leaving a gap at the site that is either unmodified or

m^{6}A-modified (Xiao *et al.,* 2018). SELECT utilizes the feature of m^{6}A to inhibit the *Bst1* polymerase and *SplintR* nick-ligase, resulting in less efficient production of complementary DNA. Thus, upon comparing m^{6}A versus m^{6}A-depleted conditions, the difference in qPCR cycle number reveals pre- or absence of m^{6}A modification. The m^{6A}-depletion is achieved either by FTO demethylation reaction or METTL3 depletion. Furthermore, SELECT allows quantification of the m^{6}A-modified fraction (Xiao *et al.,* 2018).

1.4.6 The biological role of m^{6}A

As the most abundant internal mRNA modification, m^{6}A has been found in most types of RNA and was found to be involved in all aspects of mRNA metabolism (Desrosiers et al., 1974; Zaccara *et al.,* 2019; Tegowski *et al.,* 2022; Uzonyi *et al.,* 2023; Shi *et al.,* 2019; Liu and Pan 2016). As elaborated in the previous chapters, recent technological advances allowed transcriptome-wide mapping of the modification, which paved the way to broaden the knowledge about the positional information of m^{6}A. Accompanied by characterization of m^{6}A specific RMPs, the biological functions have been extensively studied in the last decade. For instance, m^{6}A was found to play a role in cancer, Diabetes mellitus Type 2 and the circadian rhythm, with a few more examples of biological roles of m^{6}A given in detail below (Chen (B) *et al.,* 2019; Fustin *et al.,* 2013; Fustin *et al.,* 2018; Shen *et al.,* 2015).

1.4.6.1 The role of m^{6}A in development

From human to plant, the role of m^{6}A in development seems to be conserved. Several studies suggest that correct m^{6}A biogenesis enables to sustain pluripotent stem cell state in mouse embryonic stem cells (mESCs) and plays a role in the reprogramming of mouse embryonic fibroblasts (MEFs) (Wang *et al.,* 2014; Chen (B) *et al.,* 2015). Depletion of METTL3/METTL14 led to the loss of self-renewal capability in mESC, which were unable to maintain their pluripotent ground state (Wang *et al.,* 2014; Chen (B) *et al.,* 2015). Further studies revealed that m^{6}A regulates the termination of naïve pluripotency (Geula *et al.,* 2015). The finely timed m^{6}A-mediated downregulation of pluripotency factors was shown to guarantee proper lineage priming and differentiation (Geula *et al.,* 2015). While loss of *Mettl3* in mESC is viable, precocious differentiation led to embryonic lethality in mice (Geula *et al.,* 2015). Similarly, loss of *METTL3* paralog *MTA* in *Arabidopsis Thaliana* leads to embryonic lethality due to failed progression from the embryonic to the globular stage, demonstrating a conserved role from plants to mammals (Luo *et al.,* 2014; Zhong *et al.,* 2008).

Moreover, from *Drosophila* to mammals, m^{6}A plays a crucial role during neuronal development. During neurogenesis, neuronal progenitor cells (NPCs) differentiate into various glial and neural cell types. In the developing brain, m^{6}A was shown to be highly enriched during neurogenesis (Yoon *et al.,* 2017). The conditional knock out of *Mettl14* and *Mettl3* were shown

to prolong the cell cycle of radial glia cells, extend the cortical neurogenesis and to impair NPC differentiation (Yoon *et al.,* 2017). Similarly, human induced pluripotent stem cells (iPSCs)-derived forebrain brain organoids showed that m^{6}A-signalling also regulates human cortical neurogenesis (Yoon *et al.,* 2017). Additionally, in adult brains, m^{6}A was found to play crucial roles. For instance, m^{6}A was found to be highly enriched in the cerebellum (Chang *et al.,* 2017). It was found that m^{6}A regulates synaptic function and dendritic development of cortical neurons (Chang *et al.,* 2017; Merkurjev *et al.,* 2018). Furthermore, synaptic organization, maturation and transmission modulation in hippocampal neurons are m^{6}A-regulated (Merkurjev *et al.,* 2018). Dysregulation of m^{6}A biogenesis have been linked to neurological disorders. For instance, Alzheimer disease (AD) mice models showed increased m^{6}A levels in hippocampus and cortex due to decreased *Fto* and increased *Mettl3* expression (Han *et al.,* 2020; Shafik *et al.,* 2021). Moreover, *Fto* loss-of-function mutations were shown to result in developmental retardation and delay as well as growth defects, underlining the importance of m^{6}A regulation in human neuronal development (Daoud *et al.,* 2016).

Several studies have provided insights into the role of m^{6}A in oocyte maturation and spermatogenesis. For instance, while depletion of DF2 in mice led to infertility in females, knock out of *DC1* led to defects in spermatogenesis (Ivanova *et al.,* 2017; Wojtas *et al.,* 2017; Hsu *et al.,* 2017). DF2 was demonstrated to be required for oocyte competence and sustaining early zygotic development (Ivanova *et al.,* 2017). DC1 is required for spermatogonial development and for oocyte growth (Kasowitz *et al.,* 2018). Moreover, *DC2* KO mice were reported to have smaller testes and ovaries (Hsu *et al.,* 2017). The germ cells are unable to develop, leading to infertility (Hsu *et al.,* 2017). Depletion of the m^{6}A writer Mettl3 reduced fertility in *Danio rerio*, *Drosophila melanogaster* and mice, indicating a conserved function (Xia *et al.,* 2018; Hongay *et al.,* 2011; Mu *et al.,* 2021). Depletion of the m^{6}A eraser Alkbh5 in mice led to increased levels of m^{6}A and resulted in impaired fertility due to abnormal spermatogenesis (Zheng *et al.,* 2013). Thus, precise maintenance of m^{6}A biogenesis is critical for reproductive system development and fertility.

1.4.6.2 The role of m^{6}A during immune response

The m^{6}A modification has been recognized as a crucial regulator during immune response and viral infections. Since the discovery of m^{6}A, viral RNAs were known to contain m^{6}A methylations (Lavi and Shatkin, 1975). During viral infections, the viral as well as host m^{6}A-methylated RNA can alter virus/host interactions. For instance, hepatitis C virus (HCV), human immunodeficiency virus type 1 (HIV-1) and Zika virus (ZIKV) RNA are known to harbour m^{6}A. The m^{6}A reading proteins were shown to promote or inhibit the regulation of virus particle production. Methylations can alter metabolism of viral RNA by alterations in splicing, nuclear export or processing. For instance, m^{6}A decorating HIV-1 RNA enhances virus mRNA expression and could help to avoid recognition by the host immune system (Wei *et al.,* 2017; Kennedy *et al.,* 2016; Lichinchi *et al.,* 2016; Tirumuru *et al.,* 2016).

1.4.6.3 The role of m^{6}A in X chromosome inactivation

Moreover, m^{6}A was found to play a role in X-chromosome inactivation. In mammals, males possess one X and one Y chromosome, while females have two X chromosomes. The process of X-chromosome inactivation provides transcriptional silencing of one X chromosome (X_i) in females, which equalizes dosage of X-linked genes between males and females. The long non-coding RNA *XIST* plays a crucial role by mediating the silencing of transcription on the inactivated X chromosome (Penny *et al.,* 1996). Once the inactivation has been induced, *XIST* is expressed exclusively from the inactive X chromosome, which spreads from their location of synthesis and coats the whole X chromosome (Panning, 2008). Here, m^{6}A was shown to play a crucial role. In proteomic screens, the methyltransferase complex subunits WTAP, SPEN and RBM15 have been revealed as *XIST*-binding proteins (Chu *et al.,* 2015; McHugh *et al.,* 2015; Moindrot *et al.,* 2015). With 78 putative m^{6}A sites, *XIST* was shown to be highly methylated (Patil *et al.,* 2016). The methyltransferase complex subunit RBM15/15B is suggested to facilitate direct binding of *XIST* and recruitment of the methyltransferase complex (Patil *et al.,* 2016). Moreover, the authors demonstrate recruitment of DC1 to m^{6}A-methylated *XIST*, which mediates transcriptional repression (Patil *et al.,* 2016).

1.5 Dosage compensation

m^{6}A acts on almost every type of RNA and is involved in almost all aspects of RNA metabolism. As elaborated above, m^{6}A acts as a crucial player in gene expression regulation and has a broad spectrum of biological consequences. Generally, balancing gene expression is crucial and its disturbance by aneuploidy leads to reduced fitness and lethality (Torres *et al.,* 2008; Tang *et al.,* 2013; Epstein, 1990). However, despite the role of m^{6}A in regulating RNA stability and other known functions, it remains elusive whether and how m^{6}A could influence the global X-to-autosome gene dosage balancing mechanisms.

1.5.1 Sex chromosomes

Sex-determining chromosomes evolved from a pair of ancestral autosomes (Lahn and Page, 1999; Livernois and Waters, 2012). This has led to the existence of a homogametic and a heterogametic sex. Females possess two X chromosomes (homogametic sex) and males possess one X chromosomes and one Y chromosome (heterogametic sex) in mammals. Very few genes cause detrimental consequences for the cell when present in only one copy (haploinsufficient genes) (Morrill and Amon, 2019; Zug, 2022). While the variation in copy numbers of a given locus in most cases are not causing deleterious consequences, loss of larger chromosomal fractions or whole chromosomes often is incompatible with life (Torres *et al.,* 2008; Tang *et al.,* 2013; Epstein, 1990). The male Y chromosome accumulated many sex-

linked mutations and lost many genes by progressive suppression of recombination of the X and the Y chromosome (Charlesworth, 1996; Charlesworth *et al.,* 2005; Rice *et al.,* 1996; Heard and Disteche 2006). XY recombination is limited only to the so-called pseudoautosomal region (PAR), which are small regions of homology (Burgoyne, 1982). Thus, due to this non-recombinant nature, the Y chromosome is highly degenerated and contains only ~100 genes compared to ~1100 genes on the X chromosome in human (Heard and Disteche 2006). This divergence results in an imbalance of gene dosage between mammalian males and females. Therefore, the haploinsufficiency of X-linked genes need compensatory mechanisms.

1.5.2 X-chromosome inactivation

As elaborated in chapter 1.4.5.3, one X chromosome is inactivated in mammalian males and females early in development to provide equalised expression (Lyon, 1961). During X-chromosome inactivation, the chromosome undergoes several epigenetic modifications, providing transcriptional silencing of the inactive X chromosome (X_i).
Initiation of X-chromosome inactivation is controlled by the genetic locus called X-inactivation center, which contains the X chromosome *cis*-regulators such as the activator lncRNA *Xist* as well as the negative regulators *Tsix* and *Xite* (Nora *et al.,* 2012; Rastan, 1983; Rastan and Robertson, 1985). In mice, two subsequent waves of X chromosome inactivation are required early in development. The first wave leads to imprinted inactivation of the paternal X chromosome (Kay *et al.,* 1993; Takagi *et al.,* 1978; Pinheiro and Heard, 2017). During the blastocyst stage, the parental X chromosome is reactivated and random X-chromosome inactivation is initiated, that inactivates either the parental or the maternal X chromosome. Upon further cell divisions, the chosen silenced X_i chromosome stays silenced.
As described above, *Xist* is exclusively expressed from the X chromosome, which is subsequently silenced (Brown *et al.,* 1991; Clemson *et al.,* 1996). In undifferentiated female cells, low levels of *Xist* and high levels of *Tsix* were observed (Debrand *et al.,* 1999; Lee and Lu, 1999). Depletion of *Tsix* resulted in elevated *Xist* expression levels, indicating the repressive effect of *Tsix* on *Xist* (Lee and Lu, 1999; Luikenhuis and Jaenisch, 2001). The transcriptionfactors Rex1 and Rnf12 were additionally highlighted to play important roles in regulation of *Xist* and *Tsix* (Navarro and Anver, 2010; Navarro *et al.,* 2010). Rex1 suppresses levels of the *Xist* activator Rnf12, while it promotes the expression of *Tsix* (Navarro *et al.,* 2010). Upon differentiation, transcriptionfactor levels decline and the accumulation of *Xist* leads to the inhibition of *Tsix* expression (Lee and Lu, 1999; Luikenhuis and Jaenisch, 2001).
Xist coats the X chromosome in *cis*, starting from the locus of expression, leading to exclusion of RNA Pol II, removal of active histone marks and enrichment of repressive histone marks (Brown *et al.,* 1991; Jeppesen and Turner, 1993). The accumulation of *Xist* after initiation of X-chromosome inactivation is followed by several chromosomal changes. The histone modifications H3K9 acetylation and H3K4 methylation, which are associated with active chromatin, are subsequently lost (Heard *et al.,* 2001; Goto *et al.,* 2002). Subsequent

enrichment of other histone modifications, which are associated with repression such as H3K27me3 (H3 trimethylation of lysine 27), H2K119ub1 (H2A lysine 119-monoubiquititinated), H3K9me3 (H3 trimethylation of lysine 9) and H4K20me1 (H3 monomethylation of lysine 20) was demonstrated (Brockdorff and Turner, 2015). H3K9me3 is generally found within constitutive pericentric heterochromatin, while the H4K20me1was associated with chromosome condensation (Probst *et al.,* 2007; Santos *et al.,* 2005; Kohlmaier *et al.,* 2004).

1.5.2.1 X-chromosome inactivation escapers

Despite that X-chromosome inactivation is depicted as a chromosome-wide effect, some loci on the X_i are able to escape the gene silencing. Around 15% of genes on the human X_i constitutively escape inactivation (Galupa and Heard, 2018; Carrel and Willard, 2005). Thus, these genes are expressed from both female X chromosomes. These loci are frequently located in the pseudoautosomal region (PAR). Since genes located in the PAR region are present in the Y and the X chromosome, silencing of one loci in females is not required and thus, X-chromosome inactivation depends on loss of Y-linked genes (Jegalian and Page, 1998; Lahn and Page, 1999; Heard and Disteche 2006). However, many other genes that are not located in the PAR can escape X inactivation (Carrel and Willard, 2005). Escapees are frequently found clustered along the chromosome, indicating the separation from neighbouring silenced loci (Payer and Lee, 2008). Many escapees were found on the short arm of the human X chromosome (Disteche, 1999). Therefore, incomplete silencing could be a barrier effect due to the centromeric heterochromatin, separating the X_i short arm (Disteche, 1999; Heard and Disteche 2006). Ctcf was demonstrated to prevent spreading of repressive CpG island DNA methylation marks, therefore shielding the escape regions from transcriptional silencing (Filippova *et al.,* 2005). The importance of escapees is reflected by ensuring a normal human phenotype. Individuals with only a single X chromosome (X0, Turner syndrome) exhibit high foetal lethality and reduced fertility due to haploinsufficiency of escapees (Ashworth *et al.,* 1991; Heard and Disteche 2006). Thus, indicating dosage-sensitivity for escapees in females.

1.5.3 X-to-autosome dosage compensation

X-chromosome inactivation equalizes gene dosage of X-linked genes between both sexes in mammals. However, the gene dosage of the single remaining active X chromosome compared to two active copies of autosomes is imbalanced. While monoallelic gene expression can be tolerated by cells rather well, the consequences of halving the gene dosage of a whole chromosome might be fatal. Therefore, Susumu Ohno hypothesised a two-fold upregulation of X-linked genes to restore gene dosage (Ohno, 1966). In order to provide evidence for Ohno's' hypothesis, several studies assessed overall expression levels for X-linked genes

compared to autosomal genes. An X-to-autosome expression ratio (X:AA) of 0.5 could indicate the absence of upregulation. A 1.0 ratio would suggest a complete upregulation of the X chromosome and validate Ohno's hypothesis. However, several studies demonstrate conflicting evidence. Using microarrays, it was shown that the X:AA ratio in mammalian cells is ~1 (Nguyen and Disteche, 2006; Lin et al., 2007). RNA sequencing experiments initially provided contradicting evidence and subsequently rejected Ohno's hypothesis (Xiong *et al.,* 2010). Analysing public RNA sequencing data, the authors calculated an X:AA ratio of 0.5 in human and 0.3 in mice. The interpretation of RNA sequencing data needs careful consideration since the X chromosome contains higher proportions of tissue specific expressed genes compared to autosomes. This set of genes contains reproductive-related genes and is not expressed in somatic tissues (Vicoso and Charlesworth, 2006; Deng *et al.,* 2011). Thus, subsequent studies took this into consideration and reanalysed published and new experimental data sets, which supported Ohno's hypothesis (Deng *et al.,* 2011; Kharchenko *et al.,* 2011; Lin *et al.,* 2011). Although the X:AA expression ratio was consistently found to be >0.5, it did not reach 1.0 which might indicate incomplete X-upregulation where only dosage-sensitive genes undergo upregulation (Deng *et al.,* 2011; Lin *et al.,* 2012; Pessia *et al.,* 2012; Brockdorff and Turner, 2015). However, how the upregulation of the X chromosome in mammals is achieved is still under active debate.

In mammals, several chromatin-mediated, transcriptional, post-transcriptional and translational mechanisms have been identified. Genome-wide investigation of chromatin signatures demonstrated an enrichment of active histone marks compared to autosomes (Yildirim *et al.,* 2011). Using chromatin immunoprecipitation with deep sequencing (ChIP-seq), the authors showed enrichment for H3K4me3 and H3Kme3, which are associated with active chromatin. Moreover, enhanced chromatin accessibility was observed (Talon *et al.,* 2021). Talon *et al.,* suggests that previously identified increased H4K16 acetylation levels on the active X chromosome (X_a) is consistent with enhanced chromatin accessibility (Talon *et al.,* 2021; Deng *et al.,* 2013). As previously suggested that not all X-linked genes require dosage compensation, region-specific enhanced chromatin accessibility was observed (Talon *et al.,* 2021). Moreover, several studies reported higher RNA pol II occupation on the X chromosome compared to autosomes (Deng *et al.,* 2011; Yildirim *et al.,* 2011). ChIP-seq experiments revealed that the RNA pol II, which is phosphorylated at Ser5 (PolII-S5p), showed higher enrichment in the 5' end of X-linked genes (Deng *et al.,* 2011). PolII-S5p is associated with active transcriptional activation and thus, provides increased transcriptional initiation compared to autosomes (Phatnanai and Greenleaf, 2006). A subsequent study confirmed the enrichment of PolII-Sp5 at the transcription start site and additionally found an enrichment of PolII-S2P. PolII-S2P is associated with transcription elongation (Yildirim *et al.,* 2011).

In addition to regulating transcription, several post-transcriptional mechanisms have been identified. The nonsense-mediated decay (NMD) pathway is a surveillance mechanism to deplete aberrant transcripts such as transcripts with prematurely termination codons (Hillman *et al.,* 2004; Yin *et al.,* 2009). However, NMD was also shown to have a gene expression regulatory function by targeting functional transcripts as well (Isken and Maquat, 2008). Upon recognition of NMD targets, mRNA decay pathway enzymes induce degradation (Lykke-

Andersen and Jensen, 2015; He and Jacobson, 2015). A previous study found that X-chromosomal transcripts have the least NMD targets compared to all other chromosomes, reflecting a skewed chromosomal distribution and induction of RNA decay (Yin *et al.,* 2009). Importantly, after NMD inhibition by KO of the key factor *UPF1* (Up-Frameshift Suppressor 1 Homolog), gene expression was differentially influenced. The X:AA ratio of ~1.0 was reduced by 10-15% (Yin *et al.,* 2009).

Higher RNA half-lives have been observed for X-linked transcripts (Deng *et al.,* 2013; Faucillion and Larsson, 2015). For both, male and female cell lines from mouse and human datasets, increased RNA stability was globally observed. This led to higher transcript levels observed in RNA sequencing data (Faucillion and Larsson, 2015; Deng *et al.,* 2013). Thus, these results indicate a conserved post-transcriptional dosage compensation mechanism in RNA stability control. As described above, the NMD pathways could partially contribute to observed higher RNA stabilities. However, how higher RNA stability of X-chromosomal transcripts is globally achieved remains elusive.

Further studies provided evidence that dosage compensation could be achieved by elevated translation (Faucillion and Larsson, 2015; Wang (B) *et al.,* 2020). Analysis of ribosome-profiling data revealed higher ribosome density on X-chromosomal transcripts, suggesting that the translational efficiency is elevated (Faucillion and Larsson, 2015). Another recent study provided consistent results by demonstrating significant higher translational efficiencies for X-linked genes (Wang (B) *et al.,* 2020). Elevated expression for X-linked genes was demonstrated from the transcriptome to the translatome level (Wang (B) *et al.,* 2020).

Using single cell RNA sequencing (scRNA-seq), another recent study confirmed elevated expression levels for X-chromosomal transcripts (Larrson *et al.,* 2019). Moreover, by breaking down the expression levels into transcriptional kinetic parameters, the authors demonstrated that increased transcriptional burst frequency of the X chromosome leads to dosage compensation (Larrson *et al.,* 2019).

Taken together, clear evidence for accepting Ohno's hypothesis have been provided over the last decade. Several transcriptional, post-transcriptional and translational dosage compensation mechanisms could play a role. However, whether these mechanisms act together and how the chromosomal origin of a transcript is recognized for these downstream processes in the cell remains unclear. Moreover, whether epitranscriptomic mechanisms could potentially play a role in balancing gene expression has not yet been addressed and remains elusive.

1.6 Aim of this work

The m^{6}A RNA modification is the most abundant internal mRNA modification and conserved in all kingdoms of life and beyond. Recent technological advances allowed global transcriptome-wide mapping in single-nucleotide resolution. Nevertheless, these methods suffer several limitations. Moreover, despite discovering m^{6}A as a crucial player in gene expression regulation, a role in X-to-autosome dosage compensation remains elusive.

The first aim of this PhD work is to significantly enhance the detection of m^{6}A modifications. To overcome several limitations introduced by using m^{6}A-specific antibodies during miCLIP experiments, we first developed the iCLIP2 protocol. Therefore, we introduce bead-based clean-up systems, two separately ligated linker, a pre-amplification step and an optimized size selection procedure. After that, we combined our iCLIP2 protocol with the miCLIP procedure. Moreover, due to challenges during the computational m^{6}A detection in miCLIP data, we developed a machine learning classifier for enhanced m^{6}A detection. In order to avoid the need of a Mettl3-depleted condition or DRACH filtering steps, we first calibrate our machine learning classifier by performing miCLIP2 in mESC WT and *Mettl3* KO conditions. We use differential methylation analysis to identify true m^{6}A sites that we can use to train and validate our machine learning model. Therefore, our classifier can learn about the characteristics of m^{6}A sites in miCLIP2 data. After careful validation of predicted m^{6}A sites, we aim to generate high-confidence transcriptome-wide m^{6}A maps in single nucleotide resolution that can be a rich resource for future research.

In the second part of this PhD work, I aim to uncover novel m^{6}A-mediated functions. Since the modification is increasingly recognized as a crucial player in post-transcriptional gene expression regulation and due to the prominent role in promoting RNA degradation, we hypothesize that m^{6}A may be a player in mammalian dosage compensation by controlling RNA stability. In order to test whether m^{6}A mediates dosage compensation, we use our transcriptome-wide m^{6}A maps generated by using miCLIP2 and m6Aboost and analyse the methylation numbers on transcripts deriving from different chromosomes. We further aim to identify differential effects on X-chromosomal and autosomal transcripts after m^{6}A depletion. To this end, we perform RNA-sequencing upon acute m^{6}A depletion. Finally, we analyse RNA half-lives of X-chromosomal and autosomal transcripts in pre- and absence of m^{6}A. To this end, we perform SLAM-seq (Thiol(SH)-linked alkylation for the metabolic sequencing of RNA) experiments upon acute m^{6}A depletion and analyse the differential changes on transcripts from different chromosomes. In order to uncover the mechanism behind m^{6}A-mediated dosage compensation, we analyse differences in DRACH motif content in transcripts from the X chromosome and from the autosomes. To further check whether a potential discrepancy in DRACH motif content evolved in a sex-chromosome specific manner, we analyse the DRACH content in the outgroup species chicken and compare DRACH content on mouse orthologs. To further provide evidence for a hardcoded mechanism, we perform miCLIP2 in combination

with m6Aboost in X0 and XX subclones of female mESC cells, where dosage compensation mechanisms are in principle not required. This will provide us insights into the global role of m^6A in X-to-autosome dosage compensation mechanism.

2. Publications

2.1 Improved library preparation with the new iCLIP2 protocol

2.1.1 Zusammenfassung

Um RNA-Interaktionsstellen von RNA-bindenden Proteinen (RBPs) transkriptomweit zu kartieren, wird die hochmoderne Technologie der UV-Kreuzvernetzungs- und Immunpräzipitation in Einzel-Nukleotid-Auflösung (iCLIP) in diesem Forschungsgebiet weithin eingesetzt. In dieser Studie haben wir das neue iCLIP2-Protokoll entwickelt. Damit können qualitativ hochwertige iCLIP2-Bibliotheken auf effizientere und schnellere Weise gewonnen werden. Die Neuerungen des Protokolls umfassen die separate Ligation von Adaptern, zwei aufeinander folgende cDNA-Amplifikationsschritte und eine optimierte Größenselektion auf der Basis von Beads. Das gesamte Protokoll kann in vier Tagen abgeschlossen werden. Die Komplexität der iCLIP2-Bibliotheken ist deutlich erhöht. Dies führt zu einer verbesserten Darstellung von RBP-RNA-Bindungsstellen. Insgesamt bietet iCLIP2 ein Werkzeug für eine effiziente und verbesserte Bibliotheksgenerierung.

2.1.2 Abstract

In order to map RNA interaction sites of RNA-binding proteins (RBPs) transcriptome-wide, the state-of-the-art technology individual-nucleotide resolution UV crosslinking and immunoprecipitation (iCLIP) is broadly used in the field. In this study, we develop the new iCLIP2 protocol. This allows obtaining of high-quality iCLIP2 libraries in a more efficient and fast way. The advances of the protocol comprise the separate ligation of adapters, two consecutive cDNA amplification steps and an optimized bead-based size selection. The entire protocol can be completed in four days. The iCLIP2 libraries are significantly increased in complexity. This results in an enhanced representation of RBP-RNA binding sites. Taken together, iCLIP2 provides a tool for efficient and enhanced library generation.

2.1.3 Statement of contribution

In order to optimise the iCLIP library preparation efficiency, we have taken the best features of other CLIP-based techniques and developed iCLIP2 (Buchbender *et al.,* 2020). I helped during experiment optimisation. I prepared the respective figures, wrote respective parts of the manuscript and reviewed the manuscript.

Supervisor confirmation: _____

Methods 178 (2020) 33–48

Contents lists available at ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

Improved library preparation with the new iCLIP2 protocol

Andreas Buchbender[a], Holger Mutter[a], F.X. Reymond Sutandy[a,b], Nadine Körtel[a], Heike Hänel[a], Anke Busch[a], Stefanie Ebersberger[a], Julian König[a,*]

[a] Institute of Molecular Biology (IMB), Ackermannweg 4, 55128 Mainz, Germany
[b] Goethe University Medical School, Institute of Biochemistry II, Theodor-Stern-Kai 7, 60590 Frankfurt, Germany

ARTICLE INFO

Keywords:
CLIP
iCLIP
Protein-RNA interaction
RNA-binding protein
UV crosslinking
High-throughput sequencing

ABSTRACT

Individual-nucleotide resolution UV crosslinking and immunoprecipitation (iCLIP) is a state-of-the-art technology to map the RNA interaction sites of an RNA-binding protein (RBP) across the transcriptome. Here, we present the new iCLIP2 protocol that allows to obtain high-quality iCLIP libraries in a fast and efficient manner. The new protocol comprises separate adapter ligations, two cDNA amplification steps and bead-based size selection. The full procedure can be completed within four days. Our advances significantly increase the complexity of the iCLIP2 libraries, resulting in a more comprehensive representation of RBP binding sites. Overall, the methodological advances in iCLIP2 allow efficient library generation and thereby promote the versatile and flexible application of this important technology.

1. Introduction

Posttranscriptional regulation of RNA controls gene expression in various facets of biology, from embryonic development to memory formation in the adult brain, but also in diseases, like cancer. Central regulators are RNA-binding proteins (RBPs) that recognise RNA sequence and structural elements and control processes such as RNA capping, splicing, polyadenylation, subcellular localisation, translation and decay [1]. Recent studies indicate that there are more than one thousand RBPs in mammalian cells, underlining the universal importance of RBP-mediated posttranscriptional processes [2–4].

In order to understand the function of RBPs, it is central to identify the full repertoire of their binding sites in the transcriptome. The *in vivo* approach 'UV crosslinking and immunoprecipitation' (CLIP) combined with high-throughput sequencing provides transcriptome-wide maps of protein-RNA interactions from cells or tissues [5,6]. To achieve this, RBPs are covalently crosslinked to their RNA targets by irradiating living cells with UV light. The crosslinked protein-RNA complexes are then stringently purified with an antibody against the RBP of interest followed by gel purification and membrane transfer. The crosslinked RNA is then isolated, reverse transcribed, PCR-amplified and subjected to high-throughput sequencing.

Based on the original CLIP approach, 'photoactivatable-ribonucleoside-enhanced crosslinking and immunoprecipitation' (PAR-CLIP) [7] and 'individual-nucleotide resolution CLIP' (iCLIP) [8] were developed. Both protocols significantly improved the CLIP technology by mapping the protein-RNA crosslink sites with nucleotide resolution. For iCLIP, this was achieved by capturing cDNAs that truncate at the crosslinked peptide during reverse transcription (see detailed description below). Computational analysis of the resulting data can provide information on binding sites, binding strength and the function of the investigated RBP (see Busch et al. in this issue [9]). Based on the truncation principle, further CLIP variants, such as 'enhanced CLIP' (eCLIP) [10] or 'infrared CLIP' (irCLIP) [11], introduced alternative routes for CLIP library preparation. For a detailed overview of CLIP variants, we refer to a recent review by Lee and Ule [12].

Even though the original iCLIP protocol provides datasets of high quality, the efficiency of the library preparation can become limiting. This is particularly the case for low-input samples, i.e. when working with lowly abundant proteins or small cell numbers. The restricted starting material can lead to over-amplified libraries with many PCR duplicates and low complexity. Here, we present iCLIP2, a new iCLIP library preparation workflow that overcomes these limitations and allows to produce high-complexity libraries at little effort and cost (see Table 1). The four most important innovations include a PCR pre-amplification step, two independent adapter ligations, as well as the optimised size selection and RNA clean-up strategies (see below). iCLIP2 thereby brings together the best features of iCLIP, eCLIP and irCLIP (Fig. 1). This makes iCLIP experiments more sensitive, robust, easier, cheaper and faster than the original iCLIP protocol. We now routinely obtain up to 50 million unique sequencing reads after PCR duplicate removal in iCLIP2 libraries from 2.5 million human cells for abundant

* Corresponding author.
E-mail address: j.koenig@imb-mainz.de (J. König).

https://doi.org/10.1016/j.ymeth.2019.10.003
Received 20 May 2019; Received in revised form 29 September 2019; Accepted 9 October 2019
Available online 11 October 2019

Table 1
Updates in iCLIP2 compared to the previous protocol published in 2014 [13].

iCLIP protocol	Reason for change	iCLIP2 protocol	Benefit of change
Gel-based cDNA size selection after reverse transcription	Gel-based approach is time consuming and technically challenging	Bead-based cDNA size selection	Time saving, cheap, easy-to-handle and standardisable
Second adapter ligation to the cDNA 3′ end by circularisation and relinearisation	Time-consuming and prone to sample loss	The second adapter is directly ligated to the 3′end	Time saving and minimised sample loss
4-nt barcode and 5-nt Unique Molecular Identifier (UMI; previously called random barcode)	5-nt long UMI becomes saturated for high-complexity datasets	6-nt barcode and 9-nt UMI (split in 4 nt and 5 nt on either side of the barcode; see Fig. 2)	Higher specificity of the barcode and no saturation of the UMI
Overnight precipitation after reverse transcription	Time-consuming	MyONE Silane clean-up of the reverse transcription products	Time saving; clean-up can be immediately followed by ligation of the second adapter
cDNA size selection followed by PCR amplification	Sample loss: cDNA molecules lost in the size selection are irretrievable	First PCR to pre-amplify cDNA before the first bead-based size selection	cDNA molecules that are lost in the size selection had already been pre-amplified (first PCR) and are therefore still part of the final library
Primer removal by AMPure XP bead-based size selection	Inconsistent size yield of AMPure XP beads	Primer removal by a second ProNex bead-based size selection after the second PCR	Accurate and reproducible size distribution

RBPs such as U2AF2 (see showcase example below).

In detail, the iCLIP2 protocol benefits from four major updates:

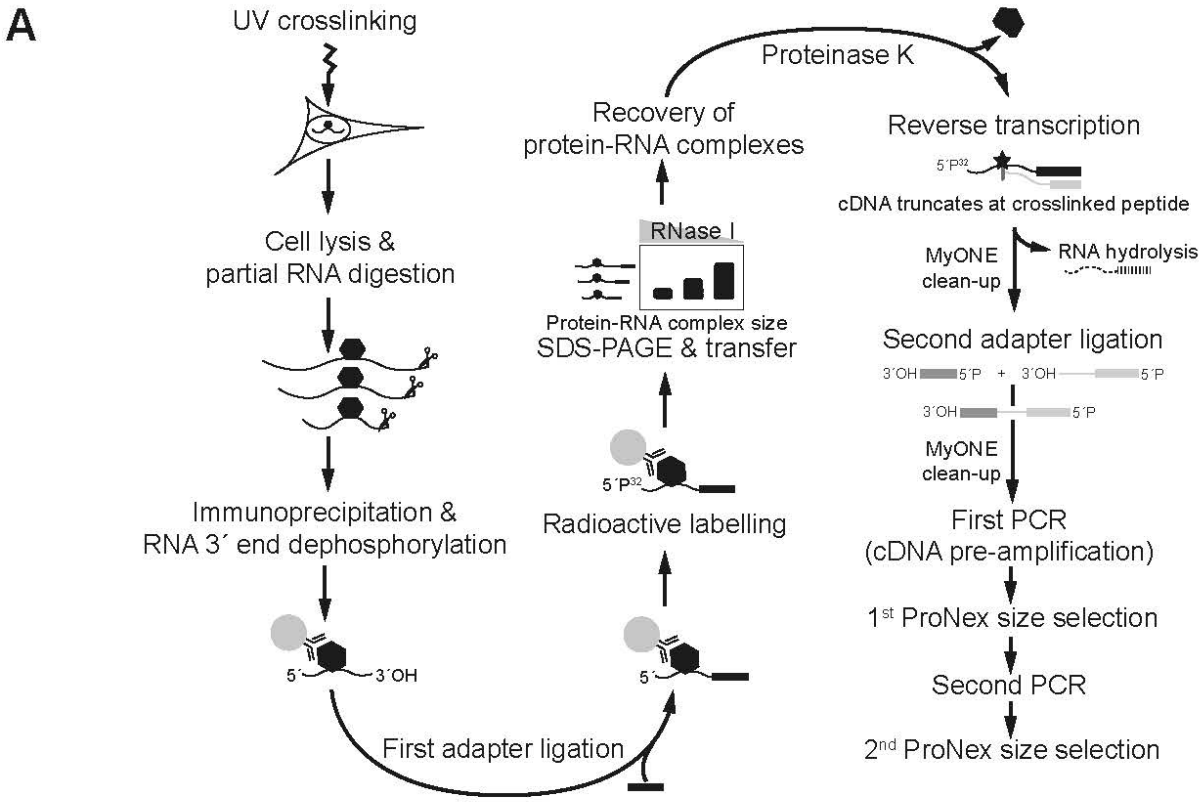
(1) Similar to irCLIP [11], we introduce a first PCR reaction for cDNA pre-amplification prior to size selection. In the original iCLIP protocol, the cDNA pool was size-selected with denaturing gel electrophoresis. Since cDNAs had not been amplified at this point, every lost molecule directly resulted in loss of information. Hence, iCLIP2 now introduces a first PCR reaction before size selection.
(2) We replaced cDNA circularisation by ligation of an independent adapter to the cDNA 3′ end to save time and money. This step was adopted from the eCLIP protocol [10].
(3) We now use bead-based size selection instead of denaturing gel electrophoresis. Compared to denaturing gel electrophoresis, bead-based size selection of cDNAs is more efficient, easier to handle and to standardise. In the new protocol we established a two-step bead-based size selection clean-up. While in the first clean-up, primer dimers and too short cDNAs are removed, the second clean-up discards PCR primers before sequencing.
(4) We revised the clean-up of the cDNAs during library preparation. Originally, all cDNA clean-up steps involved multiple ethanol precipitation and washing steps. The very small and fragile pellets did not attach well to the non-stick reaction tubes and were often lost during washing. We therefore now implement the purification by MyOne Silane beads that was introduced with the eCLIP protocol [10].

Overall, the iCLIP2 protocol harbours the following steps (Fig. 1): First, the protein-RNA complexes are covalently crosslinked by *in vivo* UV irradiation. After irradiation, the cells are lysed, and the lysate is treated with RNase I. The partial RNA digestion trims the covalently linked RNA molecules to a length spectrum suitable for sequencing. Immunoprecipitation with a specific antibody purifies the desired protein-RNA complexes. Hereafter, T4 polynucleotide kinase (PNK) is used to modify of the RNA: PNK dephosphorylates the 3′ end of the RNA, thereby generating a substrate for the first adapter ligation (L3-App) which is catalysed by T4 RNA ligase. Then, PNK is used to label the 5′ end with a radioactive ^{32}P isotope. The isotopic label allows to visualise the size and purity of the immunoprecipitated protein-RNA complexes after electrophoretic separation and transfer to a nitrocellulose membrane. The protein-RNA complexes of expected molecular weight are subsequently excised from the membrane, and the RNAs are released from the membrane-attached RBP by proteinase K treatment. From this point, the iCLIP2 protocol deviates in some aspects from the previous iCLIP protocol [13]. Using a short reverse transcription (RT) primer that is complementary to the L3-App adapter, the RNA is reverse transcribed into cDNA. After a clean-up using MyONE Silane beads, a second adapter is ligated to the 3′OH of the cDNAs. This second adapter contains a bipartite unique molecular identifier (UMI) and an experimental barcode, allowing for PCR duplicate removal and sample multiplexing, respectively (Fig. 2). After another MyONE Silane clean-up, the cDNA library is pre-amplified in a first PCR. This pre-amplification step compensates for the inevitable material loss during the subsequent ProNex size selection, which removes short cDNAs and primer dimers. The cDNA library is then amplified in a second PCR. A second ProNex size selection removes PCR primers and finally prepares the cDNA library for sequencing.

In the following sections, we give a detailed introduction into the new iCLIP2 protocol, including required materials, a cost estimation and a detailed description of all necessary steps to efficiently prepare iCLIP2 libraries. Since the initial immunoprecipitation steps remain unchanged from the original protocol, for further recommendations for these initial purification steps and important controls we refer to the manuscripts from Huppertz et al. [13] and Sutandy et al. [14]. In a showcase experiment, we apply iCLIP2 to study U2AF2 binding to RNA in human HeLa cells.

2. iCLIP2 materials

All buffers, reagents and oligonucleotides required to perform the iCLIP2 protocol are listed in the following Sections 2.1–2.4. Researchers, who are planning to introduce iCLIP2 in their laboratory, should anticipate the following approximate costs: If newly purchased, the listed consumables will amount to ca. 7300€. This initial investment will be sufficient for at least ten separate samples. Note that only few reagents, such as ProNex chemistry, RNase-free non-stick tubes and precast NuPAGE gels, are limiting, while most reagents will last for many more experiments. If all reagents are fully exploited, the price per sample will be ca. 130€ for iCLIP2, compared to ca. 160€ with the original iCLIP protocol. Hence, iCLIP2 lowers the sample price by approximately 30€ (18%).



B

	iCLIP Huppertz et al. (2014)	iCLIP2 Buchbender et al. (2019)	eCLIP Van Nostrand et al. (2016)	irCLIP Zarnegar et al. (2016)
Sample preparation				
UV crosslinking	254 nm	254 nm	254 nm	254 nm
Cell lysis	✓	✓	✓	✓
Partial RNA digestion	RNase I	RNase I	RNase I	RNase I+A or S1
Immunoprecipitation				
Dephosphorylation	PNK	PNK	FastAP & PNK	PNK
First adapter ligation	DNA adapter ligation	DNA adapter ligation	RNA adapter ligation	DNA adapter ligation
RNA labelling	Radioactive with PNK	Radioactive with PNK	No	With infrared dye
Purification of protein-RNA complexes				
PAGE & transfer	✓	✓	✓	✓
RNA visualisation	Autoradiograph	Autoradiograph	No	Infrared scan
Proteinase K treatment	✓	✓	✓	✓
RNA purification	Phenol & precipitation	Phenol & precipitation	Phenol & Zymo column	Phenol & precipitation
Library preparation				
Reverse transcription	SuperScript III	SuperScript III	Affinity Script	TIGRT-II
cDNA purification	Precipitation	MyONE	ExoSAP-IT & MyONE	Streptavidin clean-up
Second adapter ligation	Circularise & linearise	2nd adapter ligation	2nd adapter ligation	Circularisation
PCR amplification	Single PCR	Two-step PCR	Single PCR	Two-step PCR
Size selection	cDNA gel purification	ProNex beads	Gel purification	AMPure XP beads
PCR clean-up	AMPure XP beads	ProNex beads	AMPure XP beads	AMPure XP beads
Duration (days)	6	4	4	4

Fig. 1. Overview of the iCLIP2 method. A) Schematic representation of the different steps of the iCLIP2 workflow. B) Comparison of the iCLIP2 workflow to iCLIP [13], eCLIP [10] and irCLIP [11].

2.1. Buffers

Buffer	Ingredients	Manufacturer	Order No.
iCLIP lysis buffer			
	50 mM Tris-HCl, pH 7.4	Merck (Sigma)	T2194-1L
	100 mM NaCl	Merck (Sigma)	S5150-1L
	1% Igepal CA-630	Merck (Sigma)	I8896-50ML
	0.1% SDS	Bio-Rad	1610416
	0.5% Sodium deoxycholate	Merck (Sigma)	D6750-100G
Optional: add directly before usage	1:100 (v/v) Protease inhibitors	Merck (Sigma)	P8340-5ML
(add protease inhibitors only when explicitly mentioned in the protocol below)			
High-salt wash buffer			
	50 mM Tris-HCl, pH 7.4	Merck (Sigma)	T2194-1L
	1 M NaCl	Merck (Sigma)	S5150-1L
	1 mM EDTA, pH 8.0	Life Technologies	AM9261
	1% Igepal CA-630	Merck (Sigma)	I8896-50ML
	0.1% SDS	Bio-Rad	1610416
	0.5% Sodium deoxycholate	Merck (Sigma)	D6750-100G
PNK wash buffer			
	20 mM Tris-HCl, pH 7.4	Merck (Sigma)	T2194-1L
	10 mM $MgCl_2$	Life Technologies	AM9530G
	0.2% Tween-20	Merck (Sigma)	P9416-50ML
5× PNK buffer, pH 6.5			
	350 mM Tris-HCl, pH 6.5	Merck (Sigma)	T2194-1L
	50 mM $MgCl_2$	Life Technologies	AM9530G
	5 mM Dithiothreitol	Merck (Sigma)	D9779-10G
(freeze aliquots of the buffer, do not thaw and freeze again!)			
4× Ligation buffer			
	200 mM Tris-HCl, pH 7.8	Rockland	MB-004 1L
	40 mM $MgCl_2$	Life Technologies	AM9530G
	4 mM Dithiothreitol	Merck (Sigma)	D9779-10G
(freeze aliquots of the buffer, do not thaw and freeze again!)			
Proteinase K (PK) buffer			
	100 mM Tris-HCl, pH 7.4	Merck (Sigma)	T2194-1L
	50 mM NaCl	Merck (Sigma)	S5150-1L
	10 mM EDTA	Life Technologies	AM9261
PK + urea buffer			
	100 mM Tris-HCl, pH 7.4	Merck (Sigma)	T2194-1L
	50 mM NaCl	Merck (Sigma)	S5150-1L
	10 mM EDTA	Life Technologies	AM9261
	7 M Urea	Merck (Sigma)	U5378-1 KG

2.2. Reagents

Manufacturer	Company	Order number
Bead preparation		
Dynabeads for immunoprecipitation	Life Technologies	10004D (prot. G) 10002D (prot. A)
1.5 ml RNase-free non-stick tubes	Life Technologies	AM12450
Lysis		
Protease inhibitors	Merck (Sigma)	P8340-5ML
RNase I	Life Technologies	AM2295
Turbo DNase	Life Technologies	AM2238
Proteus clarification mini spin column	SERVA Electrophoresis GmbH	42225.01
De-phosphorylation		
T4 PNK (with 3′ phosphatase activity)	NEB	M0201S
RNasin Ribonuclease Inhibitors	Promega GmbH	N2615
First adapter ligation		
T4 RNA Ligase 1	NEB	M204L
PEG400	Merck (Sigma)	202398-500G
RNasin Ribonuclease Inhibitors	Promega GmbH	N2615
Radioactive labelling		
T4 PNK (with 3′ phosphatase activity)	NEB	M0201S
10× PNK buffer	NEB	B0201S
^{32}P-γ-ATP	Perkin Elmer	NEG502A250UC
SDS-PAGE		
4–12% NuPAGE gels	Life Technologies	NP0322BOX
20× Transfer buffer	Life Technologies	NP0006-1
20× MOPS-SDS running buffer	Life Technologies	NP0001

LDS 4× sample buffer	Life Technologies	NP0007
Prestained protein marker	NEB	P7712S
Nitrocellulose membrane protran BA85	VWR International GmbH	10600002
Methanol	Carl Roth	00.82.3
Whatman filter paper	GE Healthcare	3030917
Complex isolation, RNA extraction and precipitation		
Proteinase K	Merck (Sigma)	3115828001
Phenol:chloroform	Sigma	P3803
Phase lock heavy columns	VWR International GmbH	733–2478
3 M Sodium acetate pH 5.5	Life Technologies	AM9740
GlycoBlue	Life Technologies	AM9516
100% Ethanol	Merck (Sigma)	32205-2.5L-M
Razor blades	Life Technologies	YSJ-762-Q
30G syringe needles	BD Microlance	304000
Reverse transcription		
SuperScript III Reverse Transcriptase	Life Technologies	18080085
10 mM dNTPs	NEB	N0447L
RNasin Ribonuclease Inhibitors	Promega GmbH	N2615
HEPES	Merck (Sigma)	H0887-100ML
NaOH	Merck (Sigma)	S8045-500G
TE buffer	VWR International GmbH	A2575 / E112
0.2 ml RNase-free PCR tubes	Life Technologies	AM12230
MyONE clean-up		
MyONE Silane beads	Life Technologies	37002D
RLT buffer	Qiagen	79216
100% Ethanol	Merck (Sigma)	32205-2.5L-M
Second adapter ligation		
High conc. RNA ligase	NEB	M0437M
DMSO	NEB	from M0531L
PCR amplification		
2× Phusion High-Fidelity PCR MasterMix	NEB	M0531L
0.2 ml PCR tubes	Peqlab (VWR)	732–3206
High Sensitivity DNA screen tape	Agilent Technologies	5067–5584
High Sensitivity DNA reagent	Agilent Technologies	5067–5585
TapeStation tube strips	Agilent Technologies	401428
TapeStation tube caps	Agilent Technologies	401425
ProNex size selection		
GeneRuler Ultra Low Range DNA Ladder	Life Technologies	SM1211
ProNex Chemistry	Promega GmbH	NG2001
2× Phusion High-Fidelity PCR MasterMix	NEB	M0531L
High Sensitivity DNA Screen Tape	Agilent Technologies	5067–5584
High Sensitivity DNA reagent	Agilent Technologies	5067–5585
TapeStation tube strips	Agilent Technologies	401428
TapeStation tube caps	Agilent Technologies	401425

2.3. Devices

Manufacturer	Company	Order number	in Hotlab
Tissue/cell preparation			
UV-C Crossadapter Bio-Link 365	Vilber	n.a.	
Heraeus Fresco 21 Microcentrifuge	Life Technologies	n.a.	
Bead preparation			
Vakuum-Controller CVC 3000	Vacuubrand	n.a.	
Rotator RS-TR05	Carl Roth	XK30.1	
Immunoprecipitation			
DynaMag-2 Magnet	Invitrogen	12321D	yes
SDS-PAGE			
Xcell SureLock Mini-Cell and Xcell II Blot Module	Life Technologies	EI0002	yes
PowerPac Basic Power Supply	Bio-Rad	1645050	yes
Electrophoresis chamber	Life Technologies	EI0002	yes
Transfer apparatus	Life Technologies	EI0002	yes
Films	Fuji	4741019236	yes
Exposure cassette	GE Healthcare	29-1755-23	yes
Complex isolation, RNA extraction and isolation			
GE Typhoon FLA9500 Imaging System	GE Healthcare	n.a.	
GE FLA Image eraser	GE Healthcare	n.a.	
Reverse transcription			
TProfessional TRIO PCR Thermocycler	Biometra GmbH	n.a.	
PCR amplification			
TapeStation 2200	Agilent Technologies	n.a.	

Qubit Qubit 2.0 Fluorometer	Life Technologies	n.a.	
MyONE clean-up Heraeus Pico 21 Microcentrifuge	Life Technologies	n.a.	yes
Second adapter ligation Eppendorf ThermoMixer F1.5	Eppendorf	5384000012	yes

2.4. Oligonucleotide sequences

Note – The quality of the L#clip2.0 primers depends strongly on the individual synthesis. Therefore, it is important to compare all primers on the same input RNA when reordering primers or when ordering new primers for the first time.

Name	Sequence (IDT)
RToligo	GGATCCTGAACCGCT
L3-App	/rApp/AGATCGGAAGAGCGGTTCAG/ddC/
L01clip2.0	/5Phos/NNNNATCACGNNNNNAGATCGGAAGAGCGTCGTG/3ddC/
L02clip2.0	/5Phos/NNNNCGATGTNNNNNAGATCGGAAGAGCGTCGTG/3ddC/
L03clip2.0	/5Phos/NNNNTTAGGCNNNNNAGATCGGAAGAGCGTCGTG/3ddC/
L04clip2.0	/5Phos/NNNNTGACCANNNNNAGATCGGAAGAGCGTCGTG/3ddC/
L05clip2.0	/5Phos/NNNNACAGTGNNNNNAGATCGGAAGAGCGTCGTG/3ddC/
L06clip2.0	/5Phos/NNNNGCCAATNNNNNAGATCGGAAGAGCGTCGTG/3ddC/
L07clip2.0	/5Phos/NNNNCAGATCNNNNNAGATCGGAAGAGCGTCGTG/3ddC/
L08clip2.0	/5Phos/NNNNACTTGANNNNNAGATCGGAAGAGCGTCGTG/3ddC/
L09clip2.0	/5Phos/NNNNGATCAGNNNNNAGATCGGAAGAGCGTCGTG/3ddC/
L10clip2.0	/5Phos/NNNNTAGCTTNNNNNAGATCGGAAGAGCGTCGTG/3ddC/
L11clip2.0	/5Phos/NNNNATGAGCNNNNNAGATCGGAAGAGCGTCGTG/3ddC/
L12clip2.0	/5Phos/NNNNCTTGTANNNNNAGATCGGAAGAGCGTCGTG/3ddC/
L13clip2.0	/5Phos/NNNNAGTCAANNNNNAGATCGGAAGAGCGTCGTG/3ddC/
L14clip2.0	/5Phos/NNNNAGTTCCNNNNNAGATCGGAAGAGCGTCGTG/3ddC/
L15clip2.0	/5Phos/NNNNATGTCANNNNNAGATCGGAAGAGCGTCGTG/3ddC/
L16clip2.0	/5Phos/NNNNCCGTCCNNNNNAGATCGGAAGAGCGTCGTG/3ddC/
L17clip2.0	/5Phos/NNNNCAACTANNNNNAGATCGGAAGAGCGTCGTG/3ddC/
L18clip2.0	/5Phos/NNNNGTCCGCNNNNNAGATCGGAAGAGCGTCGTG/3ddC/
L19clip2.0	/5Phos/NNNNGTGAAANNNNNAGATCGGAAGAGCGTCGTG/3ddC/
L20clip2.0	/5Phos/NNNNCACGGNNNNNAGATCGGAAGAGCGTCGTG/3ddC/
L21clip2.0	/5Phos/NNNNGTTTCGNNNNNAGATCGGAAGAGCGTCGTG/3ddC/
L22clip2.0	/5Phos/NNNNCGTACGNNNNNAGATCGGAAGAGCGTCGTG/3ddC/
L23clip2.0	/5Phos/NNNNCACGATNNNNNAGATCGGAAGAGCGTCGTG/3ddC/
L24clip2.0	/5Phos/NNNNATTCCTNNNNNAGATCGGAAGAGCGTCGTG/3ddC/
L25clip2.0	/5Phos/NNNNACTGATNNNNNAGATCGGAAGAGCGTCGTG/3ddC/
P5Solexa_s	ACACGACGCTCTTCCGATCT
P3Solexa_s	CTGAACCGCTCTTCCGATCT
P5Solexa	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT
P3Solexa	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCT

3. iCLIP2 protocol

In the following section, we describe all steps needed to perform an iCLIP2 experiment. Besides notes addressing practical handling and theoretical background, this section also contains the test results for both ProNex size selection optimisations.

Note – Store all buffers (see 2.1) at 4 °C and perform the procedure on ice.

Note – If not stated differently, room temperature is defined as 23 °C.

Note – If not stated differently, all washing steps throughout the protocol are performed with a volume of 900 µl.

Note – All revolutions per minute of rotor (rpm) values are based on a rotor radius of 86 mm.

3.1. UV-C crosslinking

- Remove the medium, wash once with 5 ml PBS and add 6 ml ice-cold PBS to cells growing in a 10 cm plate (ca. 80% confluent; usually, enough for three immunoprecipitations). Place on ice-plate covered with a thin layer of water.

Note – Be careful that the cells do not freeze on the ice-plate.

Note – Fill a tray with 2 cm water and freeze in the −20 °C freezer to prepare the ice-plate. The tray should fit into the crosslinker.

- Remove the lid from cells and irradiate once with 150 mJ/cm^2 in a UV-C crosslinker (CL-1000 Ultraviolet Crosslinker, UVP) at 254 nm.

Note – Cells grown in a monolayer are equally exposed to the UV light and hence only require a single round of irradiation to crosslink equally.

Note – Make sure not to cover the energy detector of the crosslinker with the ice-plate tray.

- Harvest cells by scraping, using cell lifters.
- Add 2 ml cell suspension per 2 ml microtube, spin at 1000*g* (3200 rpm) for 1 min at 4 °C to pellet cells, then remove the supernatant.
- Snap freeze pellets on dry ice and store at −80 °C until use.

3.2. Immunoprecipitation

Note – Prepare the bead-antibody suspension in advance and then store it on ice (if preparation of the cell extract takes more than 60 min).

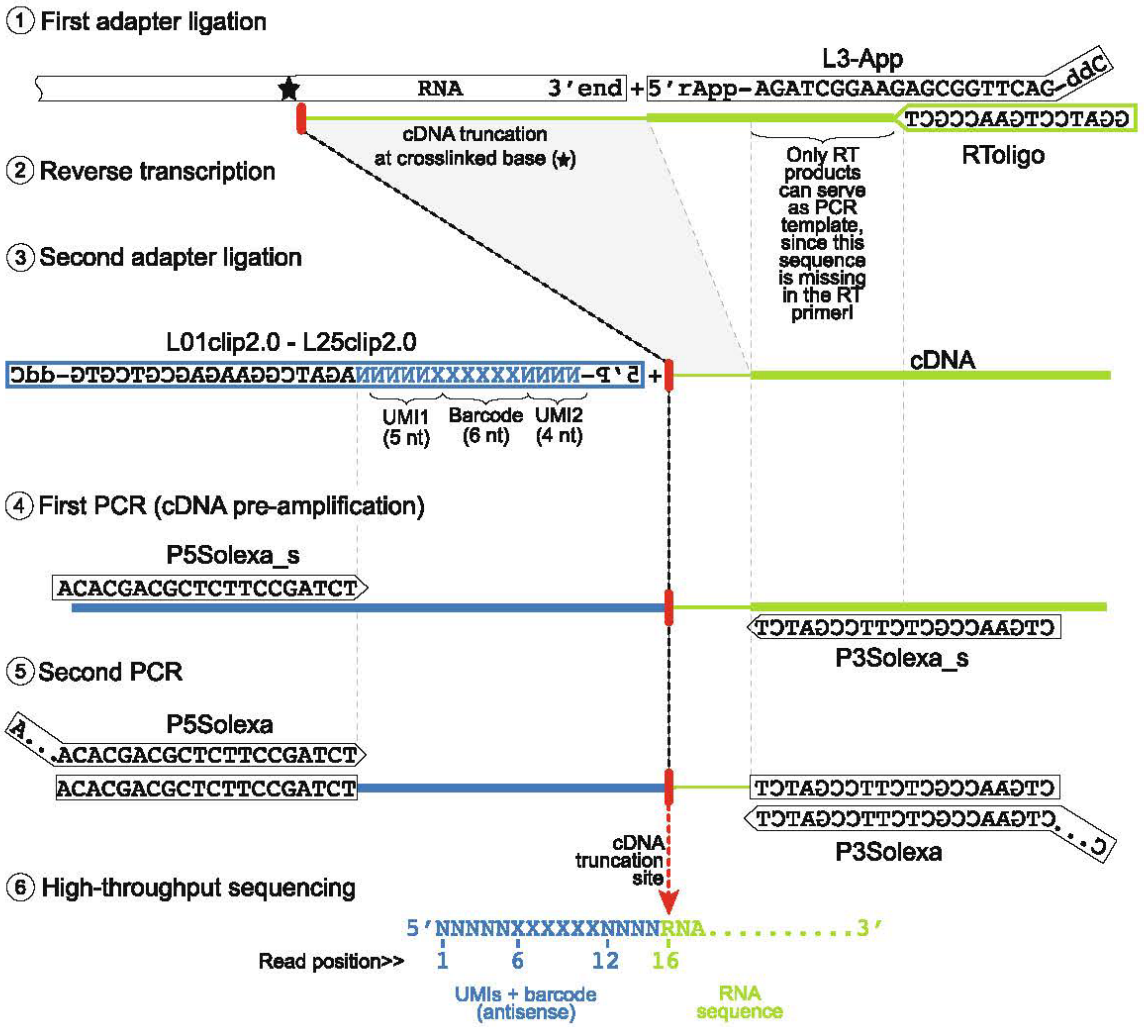


Fig. 2. RNA and DNA oligonucleotide design: (1) In order to enhance the ligation efficiency to the crosslinked RNA, the DNA adapter (L3-App) is preadenylated at the 5′ end (5′ rApp). At the 3′ end, a dideoxycytosin (ddC) modification prevents circularisation and concatemerisation with other L3-App molecules. (2) The reverse transcription primer (RToligo) is complementary to the 3′ end of L3-App, but not to the full L3-App sequence. This guarantees that only reverse transcription (RT) products can be amplified, which acquired the full L3-App sequence. (3) The second adapter (L01clip2.0–L25clip2.0) is phosphorylated at the 5′ end (5′P) to allow ligation to the cDNA products, and is protected by ddC at the 3′ end. UMI1 (5 nt) and UMI2 (4 nt) indicate the bipartite arrangement of unique molecular identifiers (UMIs), while "XXXXXX" is the experiment-specific 6-nt barcode. The experimental barcode is unique for each L#clip2.0 oligonucleotide. (4) P5Solexa_s and P3Solexa_s are used for the cDNA pre-amplification (first PCR). (5) The primers P5Solexa and P3Solexa are used for the second PCR amplification and carry the necessary adapter sequences for high-throughput sequencing. (6) The resulting sequencing reads start with the set of barcodes (antisense) at their 5′ end followed by the cDNA sequence from position 16 onwards. Grey dashed lines indicate complementary sections.

3.2.1. Bead preparation

- Add 100 µl of protein G Dynabeads per experiment to a microtube.

Note – For rabbit antibodies, protein A Dynabeads can work better in some cases. Make sure that the beads do not dry during storage, e.g. keep the storage vessel upright in the fridge.

- Wash beads 2× with lysis buffer.
- Resuspend beads in 100 µl lysis buffer with 2–10 µg antibody per experiment.

Note – The amount of antibody required depends on its quality and purity. This should be optimised in preliminary experiments.

- Rotate tubes at room temperature for 30–60 min (until lysate is ready).
- Wash beads 1× with high-salt wash buffer.
- Wash beads 2× with lysis buffer and leave in 100 µl lysis buffer with added protease inhibitors until ready to proceed to step 3.2.3.

3.2.2. Lysis and partial RNA digestion

- Resuspend cell pellet (from step 3.1) in 1 ml lysis buffer (with protease inhibitors).

Note – We are aiming for a protein concentration of ~2 mg/ml. A way is to determine RNA/protein concentration with a NanoDrop/BCA assay and normalise concentrations to the lowest sample. Take 1 ml from each sample and proceed to next step. This should lead to more reproducible RNase digestions.

- Prepare the optimised RNase I dilution in lysis buffer and add 10 µl to the lysate together with 2 µl Turbo DNase (referred to as low-RNase sample).
- **Digest RNA for exactly 3 min** shaking at 37 °C and 1100 rpm. After incubation transfer to ice for 3 min.

*Note – **Important: The RNase I dilution for the low-RNase sample has to be optimised in initial experiments.** This is critical for the success of the experiment! We aim for a RNase I concentration such that the*

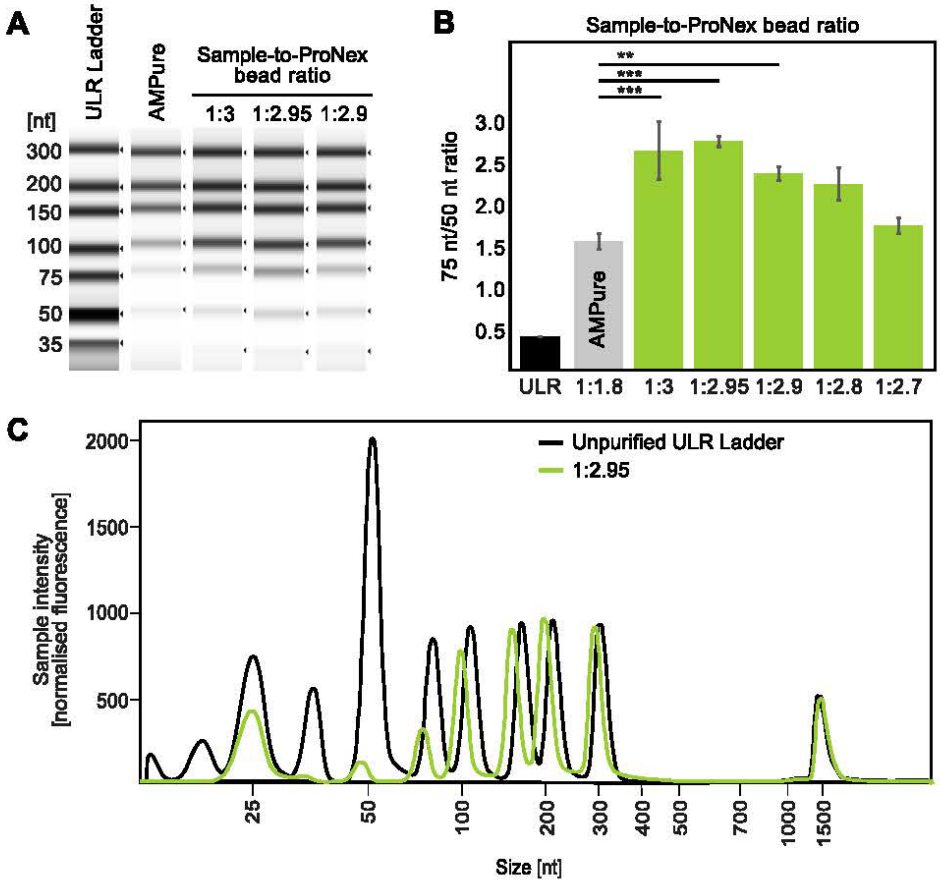


Fig. 3. Optimising library size selection for iCLIP2: Sample-to-ProNex bead ratio of 1:2.95 produces best purification result and outperforms AMPure XP beads. A) Capillary gel electrophoresis image of Solid Phase Reversible Immobilisation (SPRI) bead size-selected ULR Ladder with various sample-to-bead ratios. B) Quantified yield of ULR Ladder size selection with different sample-to-ProNex bead ratios (x-axis) and two types of SPRI beads. High 75 nt/50 nt ratios (y-axis) indicate good purification efficiency (mean ± standard deviation [s.d.], n = 3). Asterisks indicate significantly better purification than AMPure XP beads (**, $P < 0.01$; ***, $P < 0.001$; Student's *t*-test). C) Electropherogram of capillary gel electrophoresis. Unpurified ULR Ladder reference (black) and ProNex size-selected ULR Ladder with optimal sample-to-ProNex bead ratio of 1:2.95 (green). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

majority of RNA fragments is between 50 nt and 200 nt. The optimal dilution factor for the RNase I digestion depends on the batch of RNase, the concentration of the lysate and the type of cells. So in an optimisation experiment, several dilutions should be tested. ***Usually the optimal dilution is in the range of 1:50 to 1:1000.*** *If many RNase I batches are needed, it is recommended to combine several batches and thus save the RNase optimisation for each batch. For a detailed description of this optimisation step, see Huppertz et al.* [13] *and Sutandy et al.* [14]*, Figs. 3 and 2, respectively.*

Note – For an efficient and reproducible heating, use 1.5 ml tubes for a 1.5 ml thermomixer.

Note – Unlike other RNases, RNase I has no strong base preference, and therefore cleaves after all four nucleotides with almost equal efficiency.

Note – Recommended experiment for quality control and initial optimisation: Treat one sample with a high RNase I concentration: prepare a 1:5 to 1:50 (depending on cell type) RNase I dilution in lysis buffer and add 10 µl to the lysate together with 2 µl Turbo DNase. Incubate for 3 min at 37 °C shaking at 1100 rpm and then transfer to ice for 3 min. Since the RNA in this sample will be too short for library preparation, this control can directly continue with isotopic labelling (step 3.5). To minimise the use of reagents, it is possible to take only 1/5 of the cell lysate and all other reagents for this experiment. Under high-RNase conditions, the size of the radioactive band seen in the SDS-PAGE has to change in comparison to low-RNase conditions (optimised concentration for library preparation), confirming that the signal corresponds to protein-RNA complexes. Furthermore, this experiment helps to control the size of the immunoprecipitated RNA-binding protein, as the protein will be bound to short RNAs and thus will migrate as a less diffuse band ~5 kDa above the expected molecular weight of the protein alone. Hence, the high-RNase control is also important to monitor the specificity of the IP.

Note – Other recommended controls include the use of material in which the RNA-binding protein is absent (such as from a knockout animal or knockdown cells), bypassing the UV crosslinking step or omitting the antibody during IP. For further details on these controls, see Huppertz et al. [13] *and Sutandy et al.* [14]*.*

Note – Unlike other DNases, Turbo DNase is active in conditions of up to 200 mM NaCl.

- Spin at 4 °C at top speed 21,100*g* (14,800 rpm) for 10 min and transfer the supernatant to a new 1.5 ml tube.
- Load 500 µl of the lysate onto a Proteus mini clarification spin column. Spin at 4 °C at 16,000*g* (12,900 rpm) for 1 min. Transfer flow-through to a new tube. Repeat with second half of the lysate and combine both.

Note – To test a new antibody, collect 15 µl at this step for Western blot comparison of lysate before and after IP (to visualise depletion of the protein from the lysate).

3.2.3. Immunoprecipitation

- Add the cell extract to the beads.
- Rotate beads-lysate mix for 2 h at 4 °C.

Note – Do not vary the incubation time of the immunoprecipitation between experiments. RNase I is also active at 4 °C and will affect the RNA fragmentation patterns if the incubation time changes.

Note – If monitoring depletion efficiency, place on magnet and save 15 µl supernatant for Western blot analysis.

- Discard the supernatant and wash 2× with high-salt wash buffer (incubate the second wash for at least 2 min on ice).
- Wash 2× with PNK buffer and then resuspend in 1 ml PNK buffer (samples can be left at 4 °C until you are ready to proceed to step 3.3.

3.3. RNA 3′ end dephosphorylation

Note – Step 3.3 can be omitted for initial test experiments such as immunoprecipitation check and RNase optimisation. Also, steps 3.3 and 3.4 do not need to be carried out on no-UV and high-RNase samples. However, they are preformed on the no-antibody control to use as a background estimate for the complete library preparation.

- Discard supernatant. Resuspend the beads in 20 µl of the following mixture:

15 µl	Water
4 µl	5× PNK buffer, pH 6.5
0.5 µl	RNase inhibitor
0.5 µl	T4 PNK enzyme (with 3′ phosphatase activity)

- Incubate for 20 min at 37 °C in a thermomixer at 1100 rpm.
- Wash 1× with PNK buffer.
- Wash 2× with high-salt wash buffer (incubate the second wash for at least 2 min on ice).
- Wash 2× with PNK buffer.

3.4. First adapter ligation to the 3′ end of the RNA

- Carefully remove the supernatant and resuspend the beads in 20 µl of the following mix:

8 µl	Water
5 µl	4× ligation buffer
0.5 µl	RNase inhibitor
1.5 µl	Pre-adenylated L3-App (20 µM)
4 µl	PEG400
1 µl	T4 RNA ligase

- Incubate overnight at 16 °C in a thermomixer at 1100 rpm.
- Add 500 µl PNK buffer.
- Wash 2× with 900 µl high-salt wash buffer (incubate the second wash for at least 2 min on ice).
- Wash 2× with 900 µl PNK buffer and leave in 1 ml of the second wash. **Transfer to new tubes** after the first wash.

3.5. Radioactive RNA 5′ end labelling

Note – We label only 20% of the beads to reduce the amount of used radioactivity and to minimise radiation damage of the RNAs. If working with a lowly abundant protein, the full reaction can be labelled. Then the volume of the hot PNK mix should be increased to at least 16 µl (4 × 4 µl PNK mix).

- Collect 200 µl (20%) of beads from step 3.4 and remove the supernatant.
- Add 4 µl of hot PNK mix:

3 µl	Water
0.4 µl	10× PNK buffer (NEB)
0.4 µl	^{32}P-γ-ATP
0.2 µl	T4 PNK enzyme (with 3′ phosphatase activity)

- Incubate for 5 min at 37 °C in a thermomixer at 1100 rpm.
- Remove the hot PNK mix and wash beads with 1× PNK buffer. Discard the supernatant in solid radioactive waste (in a closed tube).
- After removing the supernatant, add 20 µl of 1× NuPAGE loading buffer to the beads (optional: use reducing agent and antioxidant to avoid potential interference of antibodies).
- Remove the supernatant from remaining cold beads from step 3.4. Then add the radioactively labelled beads to the cold beads and incubate at 70 °C for 5 min.
- Place on magnet to precipitate the beads and transfer supernatant to new tubes. Place the new tubes again on the magnet and load the eluate on the gel (step 3.6).

3.6. SDS-PAGE and nitrocellulose transfer

- Load the samples on a 4–12% NuPAGE Bis-Tris gel according to the manufacturer's instructions (Life Technologies). Use 0.5 L of 1× MOPS running buffer. Also, load 5 µl of a prestained protein size marker.

Note – The standardised Novex NuPAGE gels are critical. A pour-your-own SDS-PAGE gel (Laemmli) changes its pH during the run, which can get to pH ~ 9.5 leading to alkaline hydrolysis of the RNA. The Novex NuPAGE buffer system is close to pH 7. We use 1× MOPS NuPAGE running buffer.

Note – For higher molecular weight proteins, NuPAGE 3–8% Tris-acetate gels (Life Technologies) can be used.

- Run the gel for 50 min at 180 V.
- Remove the ATP-containing dye front and discard as solid radioactive waste.
- Transfer the protein-RNA complexes from the gel to a nitrocellulose membrane using the Novex XCell II Blot Module (Life Technologies) according to the manufacturer's instructions (transfer for 1–2 h at 30 V; add 10% methanol to the transfer buffer).

Note – The pure nitrocellulose membrane can be fragile, but works better for the RNA/protein extraction step.

- After the transfer, rinse the membrane in cold PBS buffer, then wrap it in saran wrap and expose it to a BAS Storage Phosphor Screen (GE Healthcare). Place radioactive and visible dots next to the membrane to later align the picture of the autoradiograph with the membrane. Expose the screen for 1 h, then scan on a Typhoon scanner (GE Healthcare).

Note – Overnight incubation is also possible, but usually indicates that there is little RNA to prepare a library from.

3.7. RNA isolation

Note – Use the high-RNase condition to examine the specificity of the protein-RNA complexes. When performing iCLIP2 for the first time, use the following criteria to check that a specific RNA-protein crosslink and immunoprecipitation has been achieved:

1. *Is there a radioactive band ~5 kDa above the expected molecular weight of the protein in the high-RNase experiment?*
2. *Does the band disappear in the control experiments? These might include no UV crosslink, pulldown with no antibody (beads only or pre-immune serum), samples from a knockout organism or knockdown cells, or an appropriate control for overexpressed tagged proteins.*
3. *Does the band move up and become more diffuse in the low-RNase condition? Because the RNA digestion is random, the RNA sizes vary more in the low-RNase condition and thus the RNA-protein complexes are more heterogeneous in size.*

On this basis, if you are convinced of the veracity of your results, proceed to RNA isolation and amplification. Note the following guidelines:

1. *The average molecular weight of 70 nt of RNA is ~20 kDa. As the tags*

contain an adapter of 21 nt (L3-App), the ideal position of RNA-protein complexes that will generate iCLIP2 tags of sufficient length is ~20–60 kDa above the expected molecular weight of the protein.

2. *The width of the excised band depends on potential other RNA-protein complexes present in the vicinity as seen in the high-RNase experiment. If none is apparent, cut a wide band covering the complete range of protein-RNA complexes. If, however, other contaminant bands are present above the size of the protein, cut only up to the size of those bands. If the contaminating bands run below your RNA-protein complex, you might consider cutting an additional band between the contaminating band and your protein-RNA complex. The RNA sequences cloned from this band can later be used to compare with those purified with your protein-RNA complex to control the specificity of your experiment.*

- Isolate the protein-RNA complexes from the low-RNase experiment using your autoradiograph as a mask for cutting the respective region (see note above) out of the nitrocellulose membrane. Use the radioactive, visible dots to align mask and membrane. The region can be taken either in a single piece or further divided. Place the membrane fragments into 1.5 ml tubes. Since the membrane is too large to fit down to the bottom of the tube, cut it into several pieces before placing it into the tube.
- Optional: Re-expose the membrane after excising the bands to confirm accuracy of cutting.
- Add 10 µl proteinase K in 200 µl PK buffer to the nitrocellulose pieces (all should be submerged). Incubate in a thermomixer for 20 min at 37 °C and 1100 rpm.
- Add 200 µl of PK + urea buffer and incubate in a thermomixer for further 20 min at 37 °C and 1100 rpm.
- Collect the solution and add it together with 400 µl phenol/chloroform to a 2 ml Phase Lock Gel Heavy tube.

Note – Since we ligate a DNA adapter to the RNAs, neutral phenol/chloroform has to be used to efficiently purify the hybrid DNA-RNA molecule. Acidic phenol as commonly used for RNA isolation will capture DNA into the organic phase.

Note – Over 90% of the radioactive signal should be removed after proteinase K treatment. This can be monitored by a Geiger counter measurement of the membrane pieces before adding proteinase K and after removing it.

- Incubate in a thermomixer for 5 min at 30 °C shaking at 1100 rpm (**do not vortex**). Separate the phases by spinning for 5 min at 16,000*g* (12,900 rpm) at room temperature.
- Transfer the aqueous layer into a new tube (be careful not to touch the gel matrix with the pipette). **Spin the supernatant again for 1 min and transfer into a new tube.**
- Precipitate by addition of 0.75 µl GlycoBlue and 40 µl 3 M sodium acetate pH 5.5. Then mix and add 1 ml 100% ethanol, mix by inverting several times and place over night at −20 °C.

Note – GlycoBlue is necessary to efficiently precipitate the small quantity of RNA.

- Spin for 20 min at 21,100*g* (14,800 rpm) at 4 °C. Remove the supernatant; wash the pellet with 900 µl 80% ethanol, and spin again for 5 min. Air-dry for 3 min. Resuspend the pellet in 5 µl water and transfer to a PCR tube.

Note – Remove the wash first with a p1000 pipette tip and then with a p20 or p10. Try not to disturb the pellet, but if you do, spin it down again. Leave on the bench for 3 min, but no longer, with the cap open to dry. When resuspending, make sure to pipette along the back area of the tube.

3.8. Reverse transcription (RT)

- Add the following reagents to the resuspended pellet from step 3.7:

1 µl	primer *RToligo* (0.5 pmol/µl)
1 µl	dNTP mix (10 mM)

Note – Do not forget a negative control. This can either be a reaction in which no RNA was added to the mix, but preferably a control sample that was isolated from a piece of nitrocellulose membrane that should not contain the protein-RNA complex (for example the no-antibody control).

- RT thermal programme:

70 °C	5 min
25 °C	hold

until the RT mix (see below) is added, mix by pipetting:

RT mix	
7 µl	Water
4 µl	5× RT buffer
1 µl	0.1 M DTT
0.5 µl	RNase inhibitor
0.5 µl	SuperScript III

- RT thermal programme continued:

25 °C	5 min
42 °C	20 min
50 °C	40 min
80 °C	5 min
4 °C	Hold

- Add 1.65 µl 1 M NaOH and incubate at 98 °C for 20 min. Then add 20 µl 1 M HEPES-NaOH pH 7.3. This will eliminate radioactivity from strongly labelled samples after the next step and prevent RNA from interfering with subsequent reactions.

3.9. Second adapter ligation to the 3′ end of the cDNA

3.9.1. MyONE clean-up

- From a thoroughly mixed MyONE Silane bead solution, use 10 µl MyONE Silane beads per sample. Attract the beads magnetically and discard the supernatant. Wash the beads with 500 µl RLT buffer and resuspend them in 125 µl RLT buffer.
- Add 125 µl washed beads to each sample and mix. Then add 150 µl 100% ethanol and carefully mix by pipetting. After 5 min at room temperature, mix the sample once more by pipetting and repeat the 5 min incubation step a second time.
- Magnetically attract the beads and discard the supernatant. Resuspend the beads in 900 µl 80% ethanol and transfer the mix to a new tube.
- Magnetically attract the beads, discard the supernatant and wash with another 900 µl 80% ethanol. Let the ethanol incubate on the sample for 30 s at room temperature and repeat the wash with 900 µl 80% ethanol for a third time.
- Briefly spin the mix in a microcentrifuge, magnetically attract the beads and discard the supernatant. Air-dry the beads for 5 min at room temperature and resuspend them in 5 µl water. Incubate the water on the sample for 5 min at room temperature and **proceed**

without removal of beads.

Note – Radiation should be at background levels after the clean-up and samples can be taken out of the radiation-controlled area. If not, continue in the radiation-controlled area.

3.9.2. Second adapter ligation

Note – Differences in the efficiency of library preparation between samples can arise from different quality of the L#clip2.0 oligos (poor oligo synthesis can lead to poor libraries). Hence, we recommend testing all L#clip2.0 oligos once on the same input material to ensure oligo quality.

- Add the second adapter and DMSO to the cDNA-bead solution (step 3.9.1). Heat the mix for 2 min at 75 °C and immediately keep it on ice for < 1 min:

2 µl	L#clip2.0 oligo (10 µM)
1 µl	100% DMSO

- Prepare the following ligation master-mix on ice:

0.3 µl	Water
2 µl	10× NEB RNA ligase buffer (with DTT)
0.2 µl	ATP, 100 mM
9 µl	50% PEG 8000
0.5 µl	High conc. RNA ligase

- To ensure homogeneity, mix the ligation master-mix by vigorous stirring, pipetting and flicking. Briefly centrifuge the mix in a microcentrifuge. Then add 12 µl of ligation master-mix to 8 µl sample-adapter mix and mix it thoroughly.
- Add another 1 µl RNA ligase to each sample (final volume: 21 µl), mix by stirring and agitate it overnight at room temperature and 1100 rpm.

3.9.3. MyONE clean-up

- In this step, **fresh MyONE Silane beads will be added to the cDNA-bead mix from step 3.9.2.** Therefore, magnetically attract 5 µl of fresh MyONE Silane beads, discard the supernatant and wash them with 500 µl RLT buffer. Repeat magnetic attraction and buffer removal in order to resuspend the beads in 60 µl RLT buffer per sample.
- Add 60 µl beads to the cDNA-bead slurry from step 3.9.2 and mix. Then add 60 µl 100% ethanol and mix by pipetting the solution carefully. After incubating the sample for 5 min at room temperature, mix it once more by pipetting and repeat the 5 min incubation step a second time.
- Magnetically attract the beads and discard the supernatant. Resuspend the beads in 900 µl 80% ethanol and transfer the mix to a new tube.
- Magnetically attract the beads, discard the supernatant and wash with another 900 µl 80% ethanol. Let the ethanol incubate on the sample for 30 s at room temperature and repeat the wash with 900 µl 80% ethanol for a third time.
- Briefly spin the mix in a microcentrifuge, magnetically attract the beads and discard the supernatant. Air-dry the beads for 5 min at room temperature and resuspend them in 23 µl water. Incubate the water on the sample for 5 min at room temperature. Magnetically attract the beads and add the eluate to the PCR mix of the next step.

3.10. First PCR (cDNA pre-amplification)

- Prepare the following PCR mix:

22.5 µl	cDNA (from step 3.9.3)
2.5 µl	Primer mix of P5Solexa_s and P3Solexa_s, 10 µM each
25 µl	2× Phusion HF PCR MasterMix

Note – We use shorter primers for the cDNA pre-amplification to obtain smaller PCR products, which allow for efficient size selection with ProNex Chemistry.

- Run the following PCR:

98°C	**30 s**	
98°C	**10 s**	**6 cycles**
65°C	**30 s**	
72°C	**30 s**	
72°C	**3 min**	
16°C	**Hold**	

3.11. First ProNex size selection to remove primer-dimers

To remove excess primer dimers, size-select your samples with ProNex Chemistry. In order to estimate the efficiency of the size selection, we recommend to include two samples with GeneRuler Ultra Low Range Ladder (ULR Ladder; Life Technologies). While one sample is used for size selection, the second serves as a reference in the TapeStation run and can be stored at room temperature. The following scheme will help to prepare the size selection control:

ULR Ladder reference:	ULR Ladder for size selection:
1 µl ULR Ladder	1 µl ULR Ladder
49 µl Water or ProNex Elution Buffer	24 µl Water
	25 µl 2× Phusion PCR MasterMix

- Equilibrate the ProNex Chemistry to room temperature for 30 min and resuspend the beads by vigorous vortexing.
- For 50 µl of sample (either PCR product from step 3.10 or ULR Ladder for size selection), add 147.5 µl ProNex Chemistry (beads). This is a 1:2.95 v/v ratio of sample to beads. Mix by pipetting 10× up and down.
- Incubate the ProNex Chemistry on the samples at room temperature for 10 min.
- Place the samples on a magnetic stand for 2 min. Discard the supernatant.
- **Leave the beads on the magnetic stand** and add 300 µl ProNex Wash Buffer to the samples. If necessary, scale up the volume of ProNex Wash Buffer to cover all beads on the magnet. While the beads are magnetically attracted, incubate the ProNex Wash Buffer for 30–60 s before removal.

Note – Do not remove ProNex beads from magnet for wash resuspension. This can cause up to 20% sample loss. For larger samples, increase the volume of ProNex Wash Buffer proportionally to the volume of sample and beads.

- Repeat the last wash of the magnetically attracted beads with another 300 µl ProNex Wash Buffer for 40–60 s, subsequently discard the supernatant and allow the samples to air-dry for ca. 8–10 min

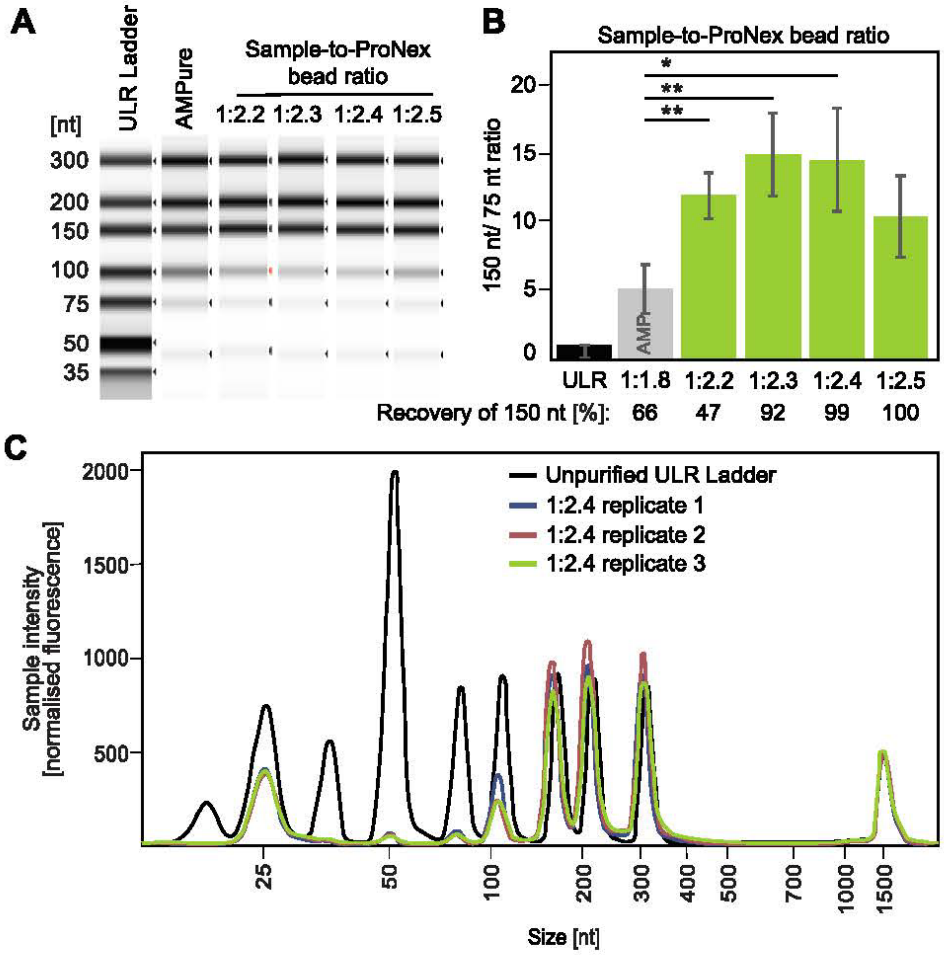


Fig. 4. Optimising the second library size selection: Sample-to-ProNex bead ratio of 1:2.4 produces best purification results. A) Capillary gel electrophoresis image of SPRI bead size-selected ULR Ladder with various sample-to-bead ratios. B) Quantified yield of ULR Ladder size selection with different sample-to-ProNex bead ratios (x-axis) and two types of SPRI beads. High 150 nt/75 nt fragment ratios (y-axis) indicate good purification efficiency (mean ± s.d., n = 3). 'Recovery of 150 nt' indicates which percentage of the 150 nt fragment is recovered after the purification. C) Electropherogram of capillary gel electrophoresis. Unpurified ULR Ladder (black) and three replicates of ProNex size-selected ULR Ladder with the optimal sample-to-bead ratio of 1:2.4 (green, blue and brown). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(< 60 min) until cracking starts.

- Remove the beads from the magnetic stand and start eluting the samples. Be careful to elute the beads of the samples in 23 µl water (or ProNex Elution Buffer), whereas the ULR Ladder sample (for size selection) requires to be eluted in 50 µl. Resuspend all samples by pipetting, and let them stand for 5 min at room temperature.

Note – ProNex Elution Buffer and water elute with the same efficiency.

Note – For the comparison, it is important that both ULR ladders (with and without size selection) end up in the same volume of the same medium (either ProNex Elution Buffer or water).

Note – Depending on the downstream application, you can add more or less ProNex Elution Buffer to elute the sample. Higher elution volumes do not result in significant yield increases. However, elution volumes < 25% of the starting sample volume can be difficult to work with and may result in some yield loss due to the resin void volume.

- Return the samples to the magnetic stand for 1 min, then carefully transfer the eluted cDNA to a clean tube.
- To check the selection efficiency of your samples, compare the ULR Ladder with and without ProNex size selection on a High Sensitivity D1000 TapeStation Kit. The efficiency of this size selection step is estimated from the ratio of intensities of the 75 nt and 50 nt fragments of the ULR Ladder. This ratio should be around 2.5 (Fig. 3).

Note – Instead of the TapeStation, also other systems to visualise low molecular weight DNA size distributions can be used. This includes for example Bioanalyzer (Agilent) or polyacrylamide gels.

TEST: Test sample-to-ProNex bead ratios for the first size selection.

Sequencing reads originating from short cDNAs or primer-dimer products compromise the information content of the final library. Therefore, size selection of the library is required to remove such products. In the original iCLIP protocol, this was done with polyacrylamide gel electrophoresis, which is time-consuming, difficult to perform and can lead to sample loss. To avoid this, we implemented and optimised a cDNA size selection in iCLIP2 that is based on ProNex bead purification.

In the size selection step, we aim to retain inserts longer than 20 nt, while inserts < 20 nt and primer-dimer products need to be excluded. A minimum insert length of 20 nt is desirable to be able to unambiguously map the resulting sequencing reads during data processing. In iCLIP2, the two adapter ligation and amplification steps extend the 20 nt long inserts to cDNA fragments of a final length of 75 nt. On the other hand, primer-dimers and adapter-adapter ligation products appear as side products and lead to fragments of 55 nt and smaller. Hence, in the first ProNex size selection step we aim to discard fragments with 55 nt or less, while retaining fragments longer than 75 nt.

In order to optimise the ProNex bead purification, we size-selected the ULR Ladder. Since the size selection cut-off in a ProNex bead purification depends on the ratio of sample to ProNex beads, we compared different ratios with the unpurified ULR Ladder (see step 3.11). All samples were quantified using the TapeStation system (Agilent). Based on the ratio of 75 nt/50 nt fragments, a sample-to-ProNex bead ratio of 1:2.95 resulted in the best fragment discrimination (Fig. 3). Since the optimal ratio for size selection might depend on the batch of the beads and additional factors, we recommend performing a similar optimisation experiment when setting up iCLIP2. Also, as indicated in the

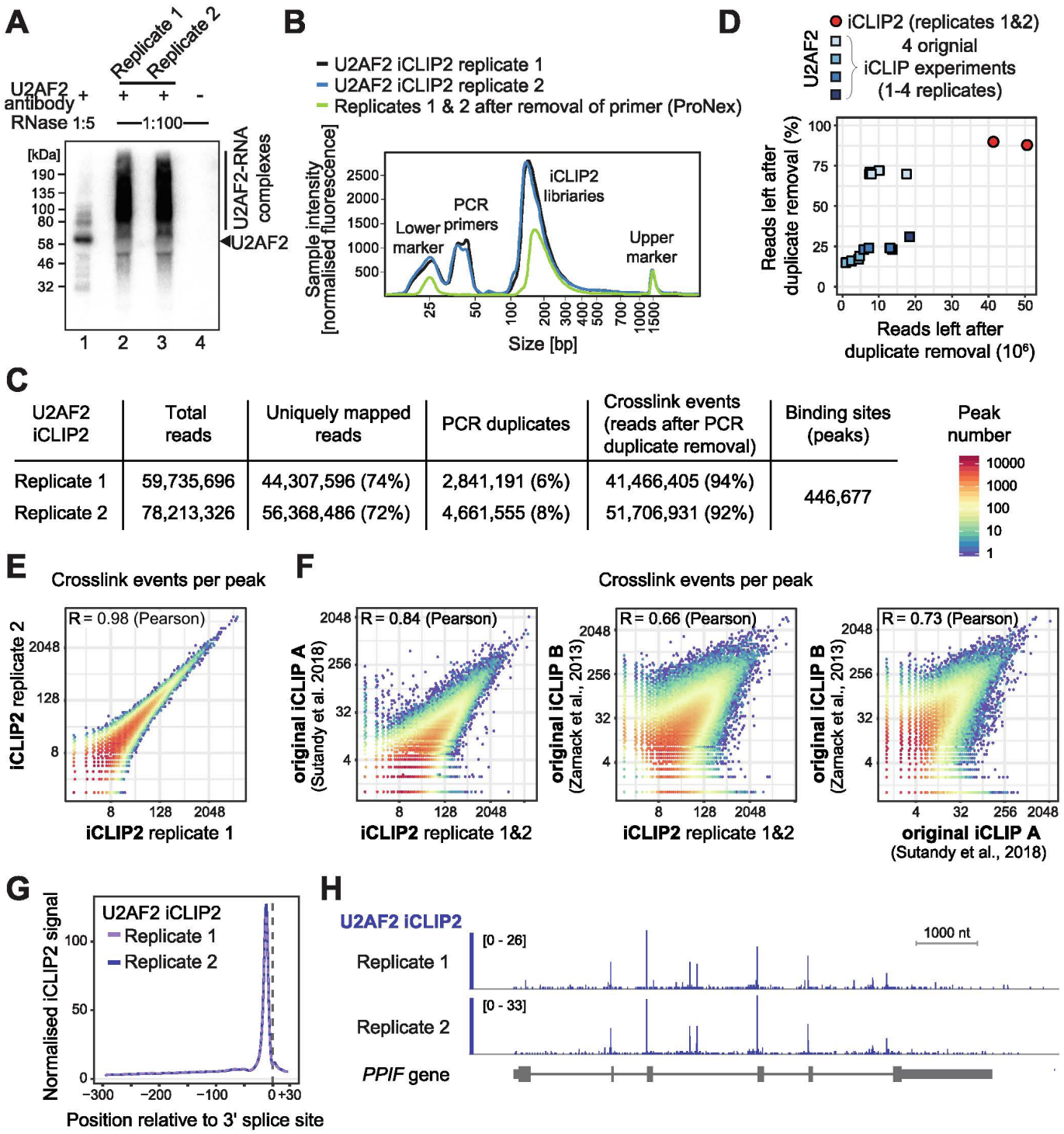


Fig. 5. Mapping U2AF2 binding with the iCLIP2 protocol. A) Autoradiograph of radioactively labelled U2AF2-RNA complexes. B) Electropherogram of capillary gel electrophoresis. U2AF2 iCLIP2 libraries of replicates 1 and 2 (after the second PCR) as well as the combined library of both replicates (after the second ProNex purification) are shown. Note that the final library had been diluted before TapeStation analysis and hence peak heights are not comparable. C) Comparison of the absolute number and fraction of reads kept after PCR duplicate removal for iCLIP2 replicates 1 and 2 (red) against 15 replicates from four independent experiments with the original iCLIP protocol (shades of blue). All experiments were performed with the same antibody and amount of starting material (~2.5 million HeLa cells). D) Table depicting read and peak numbers of the two iCLIP2 replicates after sequencing. E, F) Density scatterplots comparing crosslink events per peak between replicates 1 and 2 of the new U2AF2 iCLIP2 experiment (E) or between the combined iCLIP2 replicates and two previously published U2AF2 iCLIP datasets that were generated with the original protocol [15,17] (F). G) Metaprofile of U2AF2 crosslink events around expressed 3′ splice sites. In order to avoid a dominance of highly abundant transcripts, the signal in each position was first normalised to the total signal within the respective intron over intron length (n = 19,695 and 25,228 introns in replicate 1 and 2, respectively, of ≥300 nt length and sufficient coverage; protein-coding transcripts, transcript support level ≤3, gene support level ≤2, GENCODE annotation version 30). **H)** Genome browser view of the *PPIF* gene locus visualising U2AF2 crosslink events. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

protocol above we recommend monitoring successful size selection in parallel to each iCLIP2 library preparation.

3.12. Second PCR amplification

3.12.1. PCR cycle optimisation

- Prepare the following PCR mix:

3.5 µl	Water
1 µl	cDNA (from step 3.11)
0.5 µl	Primer mix P5Solexa/P3Solexa, 10 µM each
5 µl	2× Phusion HF PCR MasterMix

- Run the following PCR:

98°C	**30 s**	
98°C	**10 s**	**6-11 cycles**
65°C	**30 s**	
72°C	**30 s**	
72°C	**3 min**	
16°C	**hold**	

Note – A good starting point for cycle optimisation is a range of 6–11 cycles. Very good libraries lie within 6–9 cycles, while libraries amplified with 10–13 cycles will contain higher amounts of PCR duplicates. The required cycles reflect the performance of protein-RNA complex purification, which depends on protein abundance, quality of the antibody and UV crosslinking efficiency, among others. It is necessary to continue the optimisation until libraries reach a sufficient concentration for sequencing, but without over-amplification. Over-amplification is apparent from fragments shifting into higher size ranges. For more details, see Fig. 4C in Huppertz et al. [13].

Note – All steps performed after the second PCR must be carried out on a specially designated bench, ideally in a different room. The amplified cDNA must never be taken to an area where work with iCLIP RNA or cDNA is done to prevent cross-contamination between experiments.

- Run 2 µl of the amplified library on capillary gel electrophoresis using the High Sensitivity D1000 Kit in a TapeStation system.

3.12.2. Preparative PCR

- From your results in step 3.12.1, estimate the minimum number of PCR cycles to use to amplify half of the library. Consider that you will amplify a 2.5-times more concentrated cDNA (see PCR mix below), therefore one cycle less is needed than in the optimisation PCR.

Note – We perform this final library amplification in two half experiments. With two half experiments, there is still a backup sample in case the first half fails (e.g. over-amplified or under-amplified).

- Prepare the following PCR mix:

8 µl	Water
10 µl	cDNA (from step 3.11)
2 µl	Primer mix P5Solexa/P3Solexa, 10 µM each
20 µl	2× Phusion HF PCR MasterMix

- Run the same PCR programme as in step 3.12.1, but with adjusted cycle number.
- Test 2 µl of the amplified library with capillary gel electrophoresis using the High Sensitivity D1000 Kit in a TapeStation system. If everything looks fine, amplify the second half of the library and combine with the first half.
- Optional: This is a good point to multiplex different samples or replicates (all must have different barcodes) that should be sequenced together. To this end, determine the concentration of the library (e.g. the concentration under the peak with the TapeStation software) and combine the different samples in equal molarities.

3.13. Second ProNex size selection to remove residual primers

In the second cDNA size selection step, we aim to discard excess primer that would negatively influence sequencing results but to keep the amplified library. Due to the longer primers in the preparative PCR in comparison to the cDNA pre-amplification, a different sample-to-ProNex bead ratio than in the first size selection is required.

- As described in step 3.11, estimate the size selection efficiency of this second size selection by comparing the ULR Ladder with and without size selection on the High Sensitivity D1000 TapeStation Kit.
- Equilibrate the ProNex Chemistry (beads) to room temperature for 30 min and resuspend the beads by vigorous vortexing.
- Based on the optimal sample-to-ProNex (v/v) ratio of 1:2.4, calculate and add the needed volume of beads to your samples (e.g. for 50 µl sample, add 120 µl beads). Mix by 10× pipetting up and down.
- Incubate the ProNex Chemistry on the samples at room temperature for 10 min.
- Place the samples on a magnetic stand for 2 min. Discard the supernatant.
- **Leave the beads on the magnetic stand** and add 300 µl ProNex Wash Buffer to the sample. While the beads are magnetically attracted, incubate the ProNex Wash Buffer for 30–60 s before removal. If necessary, scale up the volume of ProNex Wash Buffer to cover all beads on the magnet.
- Repeat the last wash of the magnetically attracted beads with another 300 µl ProNex Wash Buffer for 40–60 s, subsequently discard the supernatant and allow the samples to air-dry for ca. 8–10 min (< 60 min) until cracking starts.

Note – Do not resuspend the beads in ProNex Wash Buffer as this causes up to 20% sample loss. For larger samples, increase the volume of ProNex Wash Buffer proportionally to the volume of sample and beads.

- Repeat the washing and allow the sample to air-dry for ca. 8–10 min (< 60 min) until cracking starts.
- Remove the beads from the magnetic stand and start eluting the samples. Be careful to elute the beads of the samples in 20 µl water (or ProNex Elution Buffer), whereas the ULR Ladder sample (for size selection) requires to be eluted in 50 µl. Resuspend all samples by pipetting and let them stand for 5 min at room temperature.

Note - Depending on the downstream application, you can add more or less ProNex Elution Buffer to elute the sample. Higher elution volumes do not result in significant yield increases. However, elution volumes < 25% of the starting sample volume can be difficult to work with and may result in some yield loss due to the resin void volume.

- Return the samples to the magnetic stand for 1 min, then carefully transfer the eluted cDNA to a clean tube or well.
- To check for successful primer removal test 2 µl of the purified library with capillary gel electrophoresis using the High Sensitivity D1000 Kit in a TapeStation system. To check the selection efficiency, compare the ULR Ladder with and without ProNex size selection on a High Sensitivity D1000 TapeStation Kit. Calculate the ratio of intensities of the 150 nt and 75 nt fragments of the size-selected ULR Ladder. In a successful size selection, this ratio should be around 15 (Fig. 4).

TEST: Sample-to-ProNex bead ratio for the second size selection

Since residual primers can negatively influence the sequencing results (e.g. due to competition effects on the Illumina Flow Cell), we implemented a second cDNA size selection step. In accordance with the longer primers of the second PCR (P5Solexa and P3Solexa, 60 nt each), the cut-off for the second size selection had to be defined newly. After the second PCR, initial inserts of at least 20 nt will be extended to cDNA fragments of 155 nt and longer. Therefore, we aimed to retain fragments longer than 155 nt, while discarding the 58 nt and 61 nt long PCR primers.

To optimise the second library size selection, we again tested varying sample-to-ProNex bead ratios using the ULR Ladder. For comparison, we also tested AMPure XP beads. Before and after purification, the ladder fragments were quantified by the TapeStation system (Agilent). We used the 150 nt/75 nt fragment ratios to determine the size selection efficiency. In addition, we quantified the recovery of the 150 nt fragment from the input ULR Ladder to monitor for absolute sample loss. For the second size selection, a sample-to-ProNex bead ratio of 1:2.4 was found to be optimal, as it gave high selectivity (150 nt/75 nt ratio of 14.7) while recovering 99% of the 150 nt fragments (Fig. 4). Since the optimal ratio for size selection might depend on the batch of the beads and additional factors, we recommend performing a similar optimisation experiment when setting up iCLIP2. Also, as indicated in the protocol above, we recommend monitoring successful size selection in parallel to each iCLIP2 library preparation.

3.14. Library quantification

- Measure library twice with Qubit High Sensitivity D1000 Kit and calculate the median fragment size [nt] and concentration [ng/µl].
- Together with the concentration and the median fragment size (TapeStation peak), determine the molarity of the library with the help of the following calculation:

$$\text{Library molecular weight} = \frac{\text{median fragment peak [nt]} * 650\ \text{g/mol}}{\text{nt}}$$

$$\text{Molar concentration} = \frac{\text{Qubit concentration [ng/}\mu\text{l]}}{\text{molecular weight}}$$

- Dilute library to 10 nM and submit 20 µl for sequencing. Store at −20 °C.

4. Showcase example: RNA binding of the splicing factor U2AF2

TEST: Preparation of U2AF2 iCLIP2 libraries

In order to demonstrate the performance, we applied iCLIP2 to map the RNA interaction sites of U2 small nuclear RNA auxiliary factor 2 (U2AF2, also known as U2AF65) in human HeLa cells. Being an essential splicing factor, U2AF2 specifically binds to the polypyrimidine tract upstream of exons where it recruits the spliceosome and thus plays a central role in 3′ splice site recognition [15].

Following the iCLIP2 protocol as described above, we performed two replicate experiments with an anti-U2AF2 antibody (Sigma U4758; 10 µg per sample) plus a control omitting the antibody. For each sample, we used a third of a 10 cm cell culture dish at 80% confluency, corresponding to ~2.5 million cells. The RNase concentration (1:100 dilution, corresponding to a final concentration of 1 U/ml Rnase I) had been optimised to obtain the majority of RNA fragments within 30–200 nt. This resulted in protein-RNA complexes ranging from the expected size of U2AF2 (65 kDa) up to 200 kDa (Fig. 5A, lanes 2 and 3). For comparison, we included a high-RNase condition (1:5 dilution, corresponding to 20 U/ml Rnase I) to reduce the crosslinked RNA fragments to a few nucleotides, thereby focussing the protein-RNA complexes mostly around 65 kDa (Fig. 5A, lane 1). As expected, no complexes were detected when omitting the antibody (Fig. 5A, lane 4).

The amplification of the cDNA libraries required 15 cycles (6 cycles for the first PCR, and 9–10 cycles for the second PCR), indicative of high-complexity libraries with low PCR duplication rates (see below). Analysis of the library size distribution with TapeStation (step 3.12) showed that the PCR products peaked around 160 nt and did not exhibit visible amounts of primer-dimers or adapter-adapter ligation products, which would be expected around 135 nt (Fig. 2). The two replicates were then pooled and excess primers were efficiently removed in the second ProNex purification (Fig. 5B; step 3.13).

High-throughput sequencing on an Illumina NextSeq500 platform generated a total of 138 million reads (Fig. 5C). The data was processed as described in Bush et al. in this issue [9], including quality filtering, genomic mapping and removal of PCR duplicates. This identified 41,466,405 and 51,706,931 U2AF2 crosslink events (i.e. reads after PCR duplicate removal) for replicates 1 and 2, respectively. Importantly, < 10% of mapped reads corresponded to PCR duplicates (Fig. 5C, D), suggesting that the libraries were not exhaustively sequenced and that a higher sequencing depth would retrieve even more U2AF2 crosslink events. The complexity of the iCLIP2 libraries was substantially higher compared to previous U2AF2 iCLIP experiments that had been performed in the König lab with the original iCLIP protocol and the same amount of starting material. In these experiments, < 20 million unique crosslink events had been retained after removal of up to 80% of the reads as PCR duplicates (Fig. 5D).

In order to delimit the U2AF2 binding sites in the crosslink profiles, we performed peak calling on the merged replicate data using PureCLIP and post-processing [16] (see also Busch et al. in this issue [9]). This yielded almost 0.5 million high-confidence U2AF2 binding sites (Fig. 5C). Notably, the quantitative information in the iCLIP2 data was highly reproducible between replicates (Fig. 5E). The iCLIP2 data was also in good agreement with two published U2AF2 experiments obtained with the original iCLIP protocol [15,17], achieving a comparable correlation as observed among the previous datasets (Fig. 5F). Moreover, metaprofiles confirmed the strong preference of U2AF2 to bind immediately upstream of 3′ splice sites, as exemplified in the *PPIF* gene (Fig. 5G, H).

Acknowledgments

We would like to thank all members of the König group for constant support and discussion. We would like to thank Kathi Zarnack, Oliver Rossbach and Christoph Dieterich for discussion and comments on the manuscript. We kindly thank members of the IMB Genomics and Bioinformatics Core Facilities for technical assistance and reagents, and for the use of the NextSeq500 Illumina sequencer (INST 247/870-1 FUGG). This work was funded by theGerman Research Foundation (DFG) via SPP1935 to JK (KO 4566/2-1).

References

[1] N.H. Gehring, E. Wahle, U. Fischer, Deciphering the mRNP code: RNA-bound determinants of post-transcriptional gene regulation, Trends Biochem. Sci. 42 (5) (2017) 369–382.
[2] A. Castello, et al., Insights into RNA biology from an atlas of mammalian mRNA-binding proteins, Cell 149 (6) (2012) 1393–1406.
[3] A.G. Baltz, et al., The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts, Mol. Cell 46 (5) (2012) 674–690.
[4] S. Gerstberger, M. Hafner, T. Tuschl, A census of human RNA-binding proteins, Nat. Rev. Genet. 15 (12) (2014) 829–845.
[5] J. Ule, et al., CLIP identifies Nova-regulated RNA networks in the brain, Science 302 (5648) (2003) 1212–1215.
[6] D.D. Licatalosi, et al., HITS-CLIP yields genome-wide insights into brain alternative RNA processing, Nature 456 (7221) (2008) 464–469.
[7] M. Hafner, et al., Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP, Cell 141 (1) (2010) 129–141.
[8] J. König, et al., iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution, Nat. Struct. Mol. Biol. 17 (7) (2010) 909–915.
[9] A. Busch, et al., iCLIP data analysis: A complete pipeline from sequencing reads to RBP binding sites, Methods (2019), https://doi.org/10.1016/j.ymeth.2019.11.008 (in press).

[10] E.L. Van Nostrand, et al., Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP), Nat. Methods 13 (6) (2016) 508–514.
[11] B.J. Zarnegar, et al., irCLIP platform for efficient characterization of protein-RNA interactions, Nat. Methods 13 (6) (2016) 489–492.
[12] F.C.Y. Lee, J. Ule, Advances in CLIP technologies for studies of protein-RNA interactions, Mol. Cell 69 (3) (2018) 354–369.
[13] I. Huppertz, et al., iCLIP: protein-RNA interactions at nucleotide resolution, Methods 65 (3) (2014) 274–287.
[14] F.X.R. Sutandy, A. Hildebrandt, J. König, Profiling the binding sites of RNA-binding proteins with nucleotide resolution using iCLIP, Methods Mol. Biol. 1358 (2016) 175–195.
[15] K. Zarnack, et al., Direct competition between hnRNP C and U2AF65 protects the transcriptome from the exonization of *Alu* elements, Cell 152 (3) (2013) 453–466.
[16] S. Krakau, H. Richard, A. Marsico, PureCLIP: capturing target-specific protein-RNA interaction footprints from single-nucleotide CLIP-seq data, Genome Biol. 18 (1) (2017) 240.
[17] F.X.R. Sutandy, et al., In vitro iCLIP-based modeling uncovers how the splicing factor U2AF2 relies on regulation by cofactors, Genome Res. 28 (5) (2018) 699–713.

2.2 Deep and accurate detection of m^{6}A RNA modifications using miCLIP2 and m6Aboost machine learning

2.2.1 Zusammenfassung

Als die häufigsten vorkommende interne mRNA-Modifikation in Eukaryoten N^6-Methyladenosin (m^{6}A) beeinflusst die meisten Schritte der RNA-Prozessierung und des RNA-Stoffwechsels. Um m^{6}A in Einzel-Nukleotid-Auflösung zu kartieren, wurde der antikörperbasierte UV-Kreuzvernetzungs- und Immunpräzipitations Ansatz miCLIP (engl. m^{6}A individual-nucleotide resolution UV-crosslinking and immunoprecipitation) entwickelt. Die begrenzte Antikörperspezifizität erschwert jedoch die zuverlässige Identifizierung von m^{6}A-Stellen in miCLIP Daten. Um diesen Nachteil zu beheben, haben wir miCLIP2 entwickelt und das verbesserte Protokoll mit einem maschinellen Lernmodell kombiniert, um die Identifizierung von m^{6}A-Stellen zu vereinfachen. Das ursprüngliche miCLIP-Protokoll erforderte große Mengen an Input RNA. Das optimierte miCLIP2 produziert komplexe Sequenzieungsbibliotheken aus geringem Inputmaterial. Um die hohen Anzahl von falsch-positiven Ergebnissen aufgrund des Antikörper-basierten m^{6}A-Identifizierungsansatzes zu überwinden, haben wir außerdem eine robuste Bioinformatische Pipeline entwickelt. Wir verwendeten *Mettl3*-Knockout-Zellen, in denen reduzierte m^{6}A-Modifikationen zu finden sind, um unsere Analysen zu kalibrieren. Auf diese Weise konnten wir Informationen über die Eigenschaften m^{6}A-Stellen generieren und unter anderem m^{6}A-Stellen außerhalb von DRACH-Motiven finden. Um miCLIP2 universell einsetzbar zu machen, trainierten wir einen maschinellen Lernalgorithmus (m6Aboost), der sowohl die experimentellen als auch die RNA-Sequenzmerkmale nutzte. Vorangegangene computergestützte Analysen haben DRACH-Filterungsschritte beinhaltet. Die Verwendung von m6Aboost ermöglicht die Vorhersage von m^{6}A-Stellen aus miCLIP2-Daten, ohne dass DRACH-Filterungsschritte oder die Erzeugung von *Mettl3*-verminderten Zelllinien erforderlich ist. Durch den Einsatz des m6Aboost-Algorithmus konnten wir m^{6}A-Stellen mit hoher Zuverlässlichkeit in verschiedenen Zelllinien von Mensch und Maus identifizieren. Damit konnten wir ein wertvolles Hilfsmittel für zukünftige Studien schaffen. Insgesamt konnten wir durch die Kombination des verbesserten experimentellen Protokolls sowie der computergestützten Methodik die Identifizierung von m^{6}A-Stellen erheblich verbessern.

2.2.2 Abstract

As the most abundant internal mRNA modification in eukaryotes, N^6-methyladenosine (m^6A) influences most aspects of RNA processing and metabolism. In order to map m^6A in an individual-nucleotide resolution manner, the antibody-based approach miCLIP (m^6A individual-nucleotide resolution UV-crosslinking and immunoprecipitation) was developed. Nevertheless, limited antibody specificity inhibits the reliable identification of m^6A sites in data produced by miCLIP. To tackle this drawback, we developed miCLIP2 and combined the improved protocol with a machine learning model to improve m^6A site detection. In the original miCLIP protocol, high amounts of input RNA were required. The optimized miCLIP2 produces high-complexity libraries from low input material. Furthermore, to overcome the issue of high false-positives due to the antibody-based m^6A identification approach, we established a robust computational pipeline. We used *Mettl3* knock out cells that lack m^6A modifications to calibrate our analyses. Therefore, we could learn about the characteristics of m^6A site deposition, accompanied by m^6A sites outside of DRACH motifs. In order to make miCLIP2 universally applicable, we trained a machine learning algorithm using the experimental as well as RNA sequence features, which we subsequently termed m6Aboost. Previous computational analysis included DRACH filtering steps. Using m6Aboost enables m^6A site prediction of genuine sites from miCLIP2 data without the need for DRACH filtering or generation of *Mettl3*-depleted cell lines. Employing the m6Aboost algorithm, we were able to identify high-confidence m^6A sites across different cell lines in human and mice. Therefore, we generated a valuable recourse for future studies. Taken together, combining our experimental advances with our computational methodology significantly enhance m^6A site identification.

2.2.3 Statement of contribution

In order to optimise the miCLIP protocol, we have taken the recently improved iCLIP2 protocol to establish miCLIP2 (Buchbender *et al.,* 2020). I established the miCLIP2 protocol and optimised poly(A) selection, UV-crosslinking and RNA fragmentation steps. I generated all miCLIP2 dataset generated for this manuscript. I performed miCLIP2 in mouse (mESC WT and mESC *Mettl3* KO, mouse heart tissue, RAW 264.7 macrophages) and human (HEK293T and C643) cells. In order to provide m^6A-depletion in human HEK293T cells, I tested and optimised the treatment conditions for METTL3 inhibitor STM2457. After computational analysis and establishing the machine learning algorithm m6Aboost, I validated predicted m^6A sites. To this end, I used the orthogonal antibody-free method SELECT and validated non-m^6A sites and m^6A sites in DRACH and non-DRACH motifs. I prepared the respective figures, wrote respective parts of the manuscript and reviewed the manuscript.

Supervisor confirmation: _____

https://doi.org/10.1093/nar/gkab485

Deep and accurate detection of m^{6}A RNA modifications using miCLIP2 and m6Aboost machine learning

Nadine Körtel[1,†], Cornelia Rücklé[1,†], You Zhou[2,†], Anke Busch[1], Peter Hoch-Kraft[1], FX Reymond Sutandy[1,3], Jacob Haase[4], Mihika Pradhan[1], Michael Musheev[1], Dirk Ostareck[5], Antje Ostareck-Lederer[5], Christoph Dieterich[6,7], Stefan Hüttelmaier[4], Christof Niehrs[1,8], Oliver Rausch[9], Dan Dominissini[10], Julian König[1,*] and Kathi Zarnack[2,*]

[1]Institute of Molecular Biology (IMB), Mainz 55128, Germany, [2]Buchmann Institute for Molecular Life Sciences (BMLS) & Faculty of Biological Sciences, Goethe University Frankfurt, Frankfurt 60438, Germany, [3]Institute of Biochemistry II, Goethe University Frankfurt, Frankfurt 60590, Germany, [4]Institute of Molecular Medicine, Sect. Molecular Cell Biology, Martin Luther University Halle-Wittenberg, Charles Tanford Protein Center, Halle 06120, Germany, [5]Department of Intensive Care Medicine, University Hospital RWTH Aachen, Aachen 52074, Germany, [6]Klaus Tschira Institute for Integrative Computational Cardiology, University Hospital Heidelberg, Heidelberg 69120, Germany, [7]German Centre for Cardiovascular Research (DZHK) - Partner Site Heidelberg/Mannheim, Heidelberg 69120, Germany, [8]Division of Molecular Embryology, DKFZ-ZMBH Alliance, Heidelberg, Germany, [9]STORM Therapeutics Ltd, Cambridge CB22 3AT, UK and [10]Cancer Research Center and Wohl Institute for Translational Medicine, Chaim Sheba Medical Center, Tel HaShomer, and Sackler School of Medicine, Tel Aviv University, Tel Aviv 6997801, Israel

Received January 13, 2021; Revised April 21, 2021; Editorial Decision May 17, 2021; Accepted June 07, 2021

ABSTRACT

N6-methyladenosine (m^{6}A) is the most abundant internal RNA modification in eukaryotic mRNAs and influences many aspects of RNA processing. miCLIP (m^{6}A individual-nucleotide resolution UV crosslinking and immunoprecipitation) is an antibody-based approach to map m^{6}A sites with single-nucleotide resolution. However, due to broad antibody reactivity, reliable identification of m^{6}A sites from miCLIP data remains challenging. Here, we present miCLIP2 in combination with machine learning to significantly improve m^{6}A detection. The optimized miCLIP2 results in high-complexity libraries from less input material. Importantly, we established a robust computational pipeline to tackle the inherent issue of false positives in antibody-based m^{6}A detection. The analyses were calibrated with *Mettl3* knockout cells to learn the characteristics of m^{6}A deposition, including m^{6}A sites outside of DRACH motifs. To make our results universally applicable, we trained a machine learning model, m6Aboost, based on the experimental and RNA sequence features. Importantly, m6Aboost allows prediction of genuine m^{6}A sites in miCLIP2 data without filtering for DRACH motifs or the need for Mettl3 depletion. Using m6Aboost, we identify thousands of high-confidence m^{6}A sites in different murine and human cell lines, which provide a rich resource for future analysis. Collectively, our combined experimental and computational methodology greatly improves m^{6}A identification.

INTRODUCTION

The epitranscriptome collectively describes modifications in RNA and has emerged as a crucial and complex mechanism for the post-transcriptional regulation of gene expression. Pervasively occurring in all three kingdoms of life, N6-methyladenosine (m^{6}A) is the most prevalent internal modification on mRNA (1,2). The emerging interest in RNA modifications revealed m^{6}A as an essential regulator in al-

*To whom correspondence should be addressed. Tel: +49 69 798 42506; Email: kathi.zarnack@bmls.de
Correspondence may also be addressed to Julian König. Email: j.koenig@imb-mainz.de
†The authors wish it to be known that, in their opinion, the first three authors should be regarded as joint First Authors. The order of the first authors was determined by lottery.

most all aspects of mRNA metabolism and uncovered diverse physiological functions (3–8).

m^{6}A is a dynamic modification. It is deposited by writers, recognized by readers and removed by erasers. The writing of m^{6}A in mRNA is mainly carried out by a highly conserved, multicomponent methyltransferase complex that catalyzes the conversion of adenosine to m^{6}A. The methyltransferase like 3 (METTL3) acts as the catalytically active subunit, possessing an S-adenosylmethionine (SAM) binding domain (MTA-70 like domain) with the conserved catalytic DPPW motif (Asp-Pro-Pro-Trp) (9). It installs m^{6}A by transferring a methyl group of a SAM donor to targeted adenosines (10). While methyltransferase like 14 (METTL14) is catalytically inactive, it forms a stable heterodimer with METTL3; simultaneously facilitating RNA interaction and increasing the catalytic activity of METTL3 (9,11). Additionally, different methyltransferases were identified as m^{6}A writers which mainly add m^{6}A to U2 and U6 snRNAs, lncRNA or pre-mRNA (12–14). In mRNA, m^{6}A enriches in a DRACH ([G/A/U][G>A]m^{6}AC[U>A>C]) consensus sequence and occurs in thousands of transcripts, with an average of one to three m^{6}A sites per mRNA transcript (15–17). However, only a fraction of DRACH motifs contain an m^{6}A modification. Furthermore, m^{6}A was found to cluster predominantly within the coding sequence in long internal exons, nearby stop codons and in the 3′ UTR (15,16).

In order to fully capture and understand the cellular impact of m^{6}A, it is essential to precisely locate the modification. Although m^{6}A had been identified over four decades ago, only recent technological breakthroughs allowed a transcriptome-wide mapping of m^{6}A (15,16,18,19). Antibody-based immunoprecipitation followed by high-throughput sequencing (m^{6}A-seq, m^{6}A-MeRIP) enabled mapping of m^{6}A within a ∼100 nucleotide (nt) window and paved the way to further understand and dissect the cellular and physiological functions of m^{6}A (15,16). Further improvements in 2015 led to an individual-nucleotide resolution UV crosslinking and immunoprecipitation (iCLIP)-based method, called m^{6}A iCLIP (miCLIP), which allows the transcriptome-wide mapping of individual m^{6}A residues at single-nucleotide resolution (17).

Despite the novel and important insights these epitranscriptomic sequencing methods uncovered, they also suffered several limitations. A critical disadvantage is the required high amount of input material, which makes transcriptome-wide m^{6}A detection exclusionary for samples with limited input material. Hence, sequencing low input samples using the aforementioned techniques may lead to over-amplified libraries with a high PCR duplication rate and low complexity. Moreover, it is broadly observed that miCLIP data comprise a lot of background signal due to limited antibody specificity, which makes computational analysis for m^{6}A-site identification challenging (20–23).

Here, we present the optimized miCLIP2 protocol, along with the machine learning-based analysis tool m6Aboost to overcome these limitations. Experimental improvements comprise two separately ligated adapters, two independent cDNA amplification steps and a bead-based size selection (24). These advances result in high-complexity miCLIP2 libraries using less input material at less effort. We performed miCLIP2 in murine embryonic stem cells (mESC), using wild-type (WT) and *Mettl3* knockout (KO) cells to identify peaks that are significantly depleted upon *Mettl3* KO and validated selected m^{6}A sites by an orthogonal method. The resulting high-confidence m^{6}A sites within DRACH and non-DRACH motifs were used to train a machine learning model, named m6Aboost, which recognizes the specific characteristics of m^{6}A sites in miCLIP2 data. We applied m6Aboost to multiple miCLIP2 datasets from human and mouse. Thus, our new miCLIP2 protocol in combination with our m6Aboost machine learning model allow to globally predict m^{6}A sites in miCLIP2 datasets independently of a *Mettl3* KO.

MATERIALS AND METHODS

LC–MS/MS analysis of m^{6}A levels

The experiments were performed as described in (25). Ribonucleoside (A, m^{6}A) standards, ammonium acetate, and LC/MS grade acetonitrile were purchased from Sigma-Aldrich. $^{13}C_9$-A was purchased from Silantes, GmbH (Munich, Germany). $^2H_3$-m^{6}A was obtained from TRC, Inc. (Toronto, Canada). All solutions were prepared using ultrapure water (Barnstead GenPure xCAD Plus, Thermo Scientific). 0.1–1 μg of poly(A)+ RNA was degraded to nucleosides with 0.003 U nuclease P1 (Roche), 0.01 U snake venom phosphodiesterase (Worthington), and 0.1 U alkaline phosphatase (Fermentas). Separation of the nucleosides from the digested RNA samples was performed with an Agilent 1290 UHPLC system equipped with RRHD Eclipse Plus C18 (95Å, 2.1 × 50 mm, 1.8 μm, Zorbax, USA) with a gradient of 5 mM ammonium acetate (pH 7, solvent A) and acetonitrile (solvent B). Separations started at a flow rate of 0.4 ml/min and linearly increased to 0.5 ml/min during first 7 min. Then, washing and re-conditioning was done at 0.5 ml/min for an additional 3 min and linearly decrease to 0.4 ml/min during the last minute. The gradients were as follows: solvent B linear increase from 0 to 7% for first 3 min, followed by isocratic elution at 7% solvent B for another 4 min; then switching to 0% solvent B for last 4 min, to recondition the column. Quantitative MS/MS analysis was performed with an Agilent 6490 triple quadrupole mass spectrometer in positive ion mode. Details of the method and instrument settings are described in (26). MRM transitions used in this study were 269.2→137.2 (A), 278.2→171.2 ($^{13}C_9$-A), 282.1→150.1 (m^{6}A) and 285.1→153.1 ($^2H_3$-N6-mrA). Quantification of all samples utilized biological triplicates, and averaged values of m^{6}A normalized to A, with the respective standard deviation are shown.

Cell culture and RNA samples

The HEK293T cell line was cultured in Dulbecco's Modified Eagle Medium (DMEM, Life Technologies) containing 10% fetal bovine serum (FBS, Life Technologies), 1% L-glutamine (Life Technologies) and 1% penicillin-streptomycin (Life Technologies) at 37°C with 5% CO_2. All cell lines were monitored for mycoplasma contamination. Anaplastic thyroid carcinoma-derived C643 cells (CLS, RRID:CVCL_5969) were cultured on 15 cm dishes

in DMEM (Thermo Fisher Scientific) supplemented with GlutaMAX (Thermo Fisher Scientific) and 10% FBS at 37°C and 5% CO_2.

Mouse embryonic stem cells (mESC) with wild-type and *Mettl3* KO genotype were taken from a previous publication (27) and cultured under FBS/LIF conditions as described therein. RAW 264.7 cells (ATCC, Wesel, Germany, TIB-71) were cultured in DMEM (Thermo Fisher Scientific, 12430054) supplemented with 10% heat inactivated FBS (Biochrom, Berlin, Germany, S0613) and 1× penicillin/streptomycin (Thermo Fisher Scientific, 15140–122).

m^{6}A depletion by METTL3 inhibitor treatment

For m^{6}A validation in HEK293T cells using SELECT, m^{6}A was depleted by using the METTL3 inhibitor STM2457 (STORM Therapeutics) (28). STM2457 was titrated to test for optimal m^{6}A depletion quantified by liquid chromatography–tandem mass spectrometry (LC–MS/MS). To this end, HEK293T cells were treated with 2–20 μM STM2457 in DMSO 0.05–0.2% (v/v) or DMSO alone 0.2% (v/v) as a negative control. After 16 h of treatment, the cells were washed with ice-cold PBS and collected on ice.

RNA isolation and poly(A) selection

For RNA extraction from HEK293T and mESC cells, cells were washed in ice-cold PBS and collected on ice for the isolation of total RNA using the RNeasy Plus Mini Kit (Qiagen) following the manufacturer's recommended protocol. For C643 and RAW 264.7 cells, cells were washed with PBS, and total RNA was extracted using TRIzol reagent (Thermo Fisher Scientific) according to the manufacturer's instructions. Prior to isolation of poly(A)+ RNA, total RNA samples were treated with DNase I (New England Biolabs) according to the manufacturer's protocol, and subsequently cleaned up again by using TRIzol LS (Thermo Fisher Scientific).

For HEK293T and C643 cells, poly(A)+ RNA was extracted using Oligo d(T)$_{25}$ Magnetic Beads using the manufacturer's recommended protocol (Thermo Fisher Scientific, 61002). Poly(A)+ concentration was measured using Qubit™ RNA HS Assay Kit (Thermo Fisher Scientific). For RAW 264.7 cells, poly(A)+ RNA was extracted by incubating 100 μg total RNA with 200 μl Dynabeads solution (Dynabeads mRNA Direct Purification Kit, Thermo Fisher Scientific, 61012) and purified following the manufacturer's protocols.

The quality of poly(A)+ RNA was ensured using High Sensitivity RNA ScreenTapes for the 2200 TapeStation system (Agilent). If a predominant peak for ribosomal RNA was still detectable, an additional round of poly(A) selection was performed, resulting in one round of selection for mESC and RAW 264.7 cells, and two rounds for HEK293T and C643 cells.

RNA fragmentation

Poly(A)+ RNA was fragmented using RNA fragmentation reagents from Thermo Fisher Scientific. 1 μg of poly(A)+ RNA was filled up to 22 μl with H_2O for each condition. 1 μl of 0.1–0.4× diluted fragmentation buffer was added (always prepared freshly). The mixture was incubated for 7–12 min at 70°C in thermomixer at 1,100 rpm and put immediately on ice. 1 μl of 0.1–0.4× diluted STOP solution was added. The solution was mixed and placed back on ice until use. Time of fragmentation and dilution of fragmentation reaction solutions were optimized prior to miCLIP2 experiments for each new batch of RNA.

miCLIP2 experiments

All miCLIP2 experiments were performed with rabbit anti-m^{6}A antibody purchased from Synaptic Systems (order number 202 003).

UV crosslinking and immunoprecipitation. 50 μl of protein A Dynabeads (Dynal, 100.02) were magnetically separated, washed two times in 900 μl IP buffer (50 mM Tris, pH 7.4, 100 mM NaCl, 0.05% NP-40) and then resuspended in 50 μl IP buffer and put at 4°C until use. 6 μg of m^{6}A antibody was added to the 24 μl of fragmented RNA and rotated for 2 h at 4°C. The IP mixture was placed on a parafilm-coated dish and UV irradiated with 2 × 150 mJ/cm^2 of UV 254 nm. The mixture was placed back into the tube, another 500 μl of IP buffer and 50 μl of washed protein A beads were added. The mixture was rotated at 4°C for 1 h. The beads were magnetically separated and the supernatant was discarded. The beads were washed two times with high-salt wash (50 mM Tris–HCl, pH 7.4, 1 M NaCl, 1 mM EDTA, 1% Igepal CA-630 [Sigma I8896], 0.1% SDS, 0.5% sodium deoxycholate). The second wash was rotated for at least 1 min at 4°C. Subsequently, the beads were washed two times with PNK buffer (20 mM Tris–HCl, pH 7.4, 10 mM $MgCl_2$, 0.2% Tween-20) and resuspended in 1 ml PNK buffer (the samples can be left at 4°C until ready to proceed).

3′ End RNA dephosphorylation. The beads were magnetically separated and resuspended in 20 μl of 3′ end RNA dephosphorylation mixture (4 μl 5× PNK pH 6.5 buffer, 0.5 μl PNK [New England Biolabs; with 3′ phosphatase activity], 0.5 μl RNasin, 15 μl water). The mixture was incubated for 20 min at 37°C in a thermomixer at 1,100 rpm. The beads were washed once with PNK buffer, once with high-salt wash (rotate wash for at least 1 min at 4°C) and again washed two times with PNK buffer.

L3 DNA linker ligation. The supernatant was magnetically removed and the beads were resuspended in 20 μl of L3 DNA linker ligation mixture (8 μl water, 5 μl 4× ligation buffer, 1 μl RNA ligase [New England Biolabs], 0.5 μl RNasin [N2615, Promega GmbH], 1.5 μl pre-adenylated DNA linker L3-App [20 μM; 5′-/rApp/AGATCGGAAG AGCGGTTCAG/ddC/-3′], 4 μl PEG400 [202398, Sigma]). The mixture was incubated overnight at 16°C at 1,100 rpm in a thermomixer. Subsequently, 500 μl of PNK buffer was added. The beads were washed two times with 1 ml high-salt buffer and two times with 1 ml PNK buffer. After the first wash, the mixture was transferred to a new tube.

5′ End labelling. The beads were magnetically separated and 4 μl of hot PNK mix (0.2 μl PNK [New England Biolabs], 0.4 μl ^{32}P-γ-ATP, 0.4 μl 10× PNK buffer [New England Biolabs], 3 μl H_2O) was added and incubated for 5 min at 37°C in a thermomixer at 1,100 rpm. Next, the supernatant was removed and 20 μl of 1× NuPAGE loading buffer (4× stock was mixed with water and reducing agent and antioxidant was used to avoid potential interference of antibodies) was added to the beads and incubated at 70°C for 5 min.

SDS-PAGE and nitrocellulose transfer. The beads were magnetically separated and the eluate was loaded on a 4–12% NuPAGE Bis-Tris gel (Invitrogen). 0.5 l of 1× MOPS running buffer (Invitrogen) was used. Additionally, 5 μl of a pre-stained protein size marker was loaded. The gel was run for 50 min at 180 V. The dye front was cut and discarded as solid radioactive waste. For transferring the protein–RNA complexes to a Protan BA85 Nitrocellulose Membrane, a Novex wet transfer apparatus was used according to the manufacturer's instructions. The transfer was performed for 1 h at 30 V in 1× transfer buffer with 10% methanol. After the transfer, the membrane was rinsed in 1× PBS buffer. Afterwards, it was wrapped in saran wrap and exposed to a Fuji film at 4°C for 30 min, 1 h, or overnight. The film was exposed to a Typhoon phosphoimager.

RNA isolation. The protein-RNA complexes were isolated by using the autoradiograph as a mask by cutting the respective regions out of the nitrocellulose membrane. The fragments were placed in a 1.5 ml tube and 10 μl proteinase K (Roche, 03115828001) in 200 μl PK buffer (100 mM Tris–HCl, pH 7.4, 50 mM NaCl, 10 mM EDTA) was added and incubated at 37°C for 20 min at 1,100 rpm. 200 μl of PK buffer + 7 M urea (100 mM Tris–HCl pH 7.4, 50 mM NaCl) was added and incubated at 37°C for 20 min at 1,100 rpm. The solution was collected and added together with 400 μl phenol/chloroform (Sigma P3803) to a 2 ml Phase Lock Gel Heavy tube (713-2536, VWR). The mixture was incubated for 5 min at 30°C at 1,100 rpm. The phases were separated by spinning for 5 min at 13,000 rpm at room temperature. Next, the aqueous layer was transferred into a new tube. Precipitation was performed by addition of 0.75 μl glycoblue (Ambion, 9510), 40 μl 3 M sodium acetate pH 5.5 and addition of 1 ml 100% ethanol. After mixing, the mixture was placed at –20°C overnight. The mixture was spun for 20 min at 15,000 rpm at 4°C. After removing the supernatant, the pellet was washed with 0.9 ml 80% ethanol and spun again for 5 min. After removing the supernatant, the pellet was resuspended in 5 μl H_2O and transferred to a PCR tube.

Reverse transcription. RT primers and dNTPs (1 μl primer Rtclip2.0 [5′-GGATCCTGAACCGCT-3′], 0.5 pmol/μl and 1 μl dNTP mix, 10 mM) were added to the resuspended pellet and incubated in a thermocycler (70°C, 5 min, 25°C hold until RT mix is added). After adding the RT mix (7 μl H_2O, 4 μl 5× RT buffer [Invitrogen], 1 μl 0.1 M DTT, 0.5 μl RNasin, 0.5 μl Superscript III) the mixture was incubated in a thermocycler (25°C, 5 min; 42°C, 20 min; 50°C, 40 min; 80°C, 5 min; 4°C, hold). 1.65 μl of 1 M NaOH was added and incubated at 98°C for 20 min. Subsequently, 20 μl of 1 M HEPES-NaOH pH 7.3 was added. This will eliminate radioactivity from strongly labelled samples after the next step and prevent RNA from interfering with subsequent reactions.

Silane clean-up. For bead preparation: 10 μl MyONE Silane beads were magnetically separated per sample and the supernatant was removed. The beads were washed with 500 μl RLT buffer and resuspended in 93 μl RLT buffer. For cDNA binding the beads in 93 μl were added to each sample. After mixing, 111.6 μl of 100% ethanol was added. The mixture was carefully mixed and incubated for 5 min at RT. After incubation, the mixture was again mixed and incubated for 5 min further. After magnetically separating the beads and removing the supernatant, 1 ml of 80% ethanol was added and the mixture was transferred to a new tube. The beads were washed twice in 80% ethanol. The beads were magnetically separated and the supernatant was removed. The tube was briefly mixed in a picoFuge and the remaining supernatant was removed. The beads were air-dried for 5 min at RT. The beads were resuspended in 5 μl H_2O and incubated for 5 min at RT before performing the on-bead ligation. Radioactivity should be removed. If radioactivity is still detected, continue in hot-lab area.

Ligation of 5′ linker to cDNA (on-bead). The linker was prepared by heating the linker mix (2 μl L##clip2.0 (10 μM stock) 1 μl 100% DMSO) for 2 min at 75°C and keeping it on ice afterwards for > 1 min. The DNA linker L##clip2.0 has the sequence 5′-/5Phos/NNNNXXXXXXNNNNNAGATCGGAAGAGCGTCGTG/3ddC/-3′, where N's are the 4-nt and 5-nt random nucleotides from the unique molecular identifier (UMI) and X's are the 6-nt the sample-specific experimental barcode given in Supplementary Table S1. After adding the linker mix to the bead containing sample, the ligation mixture (2.0 μl 10× RNA Ligase Buffer [with DTT; New England Biolabs], 0.2 μl 0.1 M ATP, 9.0 μl 50% PEG 8000, 0.3 μl H_2O, 0.5 μl high conc. RNA Ligase [New England Biolabs]) was pipetted on ice. To ensure homogeneity, the ligation-master-mix was mixed by flicking and spinning it down and was subsequently added with the linker-sample-mix. After vigorous stirring, another 1 μl RNA ligase was added to each sample and mixed by stirring. The mixture was incubated at RT at 1,100 rpm overnight.

Silane cleanup of linker-ligated cDNA. Per sample, 5 μl MyONE Silane beads were prepared. The MyONE Silane clean-up was performed as described in the previous Silane clean-up step with following modification: After washing the beads in 500 μl RLT, the beads were resuspended in 60 μl RLT buffer and added to the already bead-containing sample. After the precipitation was performed as previously described, the dried beads were resuspended in 22.5 μl H_2O.

First PCR amplification. The PCR mixture (2.5 μl primer mix 1st PCR [P5Solexa_s, 5′-ACACGACGCTCTTCCGATCT-3′ and P3Solexa_s, 5′-CTGAACCGCTCTTCCGATCT-3], 10 μM each, 25 μl Phusion High Fidelity PCR Master Mix [New England Biolabs, M0531S] was prepared

and added to the 22.5 μl of sample from the previous step. A 6-cycle PCR was performed in a thermocycler (98°C, 30 s; 6× [98°C, 10 s; 65°C, 30 s; 72°C, 30 s]; 72°C, 3 min; 16°C, hold).

First ProNex size selection. In order to remove primer and primer-dimers, a bead-based size selection was performed prior to preparative PCRs. In addition to the samples, 50 μl of 'Ultra Low Range Ladder' (ULR, Thermo Fisher Scientific) will be size selected in parallel to monitor ProNex size selection efficiency. ProNex chemistry was adjusted to RT by keeping it for 30 min at RT. 50 μl of ULR-Phusion mix (1.2 μl ULR Ladder, 28.8 μl H_2O, 30 μl Phusion PCR mastermix [New England Biolabs] and the samples were mixed with 147.5 μl ProNex chemistry. This is a 1:2.95 (v/v) ratio of sample:beads. This was optimized in previous experiments (24). The mixture was mixed ten times by pipetting and incubated for 10 min at RT. The sample-bead mixture was placed on a magnetic stand for 2 min and the supernatant was removed. While leaving the bead on the magnetic stand, 200 μl ProNex wash buffer was added to the sample. The buffer was incubated for 60 s before removal. The washes were repeated for a total of two washes. After removal of the supernatant, the beads were air-dried for 8–10 min (< 60 min) until cracking starts. The beads were eluted in 23 μl H_2O. After 5 min of incubation, the mixture was returned to the magnetic stand for 1 min and the supernatant was carefully transferred to a new tube. The size selection efficiency was monitored for the ULR sample on a High Sensitivity D1000 TapeStation Kit. For comparison, the selected and unselected ULR Phusion mix was analyzed. The 75-nt/50-nt ladder fragment ratio was compared which should be around 2.5.

Optimize PCR amplification. In order to prevent over-amplification of the library, the PCR cycle has to be optimized to a minimum. Therefore, optimize PCR amplification reactions have to be performed for each sample with each 6 and 10 cycles. The PCR mixture (0.5 μl primer mix P5Solexa [5′-AATGATACGGCGACCACCGAGATCTACACTCT TTCCCTACACGACGCTCTTCCGATCT-3′]/P3Solexa [5′-CAAGCAGAAGACGGCATACGAGATCGGTCT CGGCATTCCTGCTGAACCGCTCTTCCGATCT-3′], 10 μM each, 5 μl Phusion High Fidelity PCR Master Mix [New England Biolabs, M0531S], 3.5 μl water) was added to 1 μl of the pre-amplified library. The PCR reaction was performed in a thermocycler (98°C, 30 s; 6 or 10× (98°C, 10 s; 65°C, 30 s; 72°C, 30 s); 72°C, 3 min; 16°C, hold). 2 μl of the amplified library was run on a High Sensitivity D1000 Kit in a TapeStation system. Repeat this step until libraries are seen without over-amplification.

Preparative PCR. From previous results of the PCR cycle optimization, the minimum of PCR cycles was used to amplify $\frac{1}{2}$ of the library. Here, 2.5 times more concentrated cDNA is used, therefore one cycle less is needed than in the preliminary PCR. The PCR mix (8 μl H_2O, 2 μl primer mix P5Solexa/P3Solexa, 10 μM each, 20 μl Phusion HF Mix [New England Biolabs]) was added to 10 μl cDNA. The PCR was performed in a thermocycler using the same program as in the optimization PCR with the optimized cycle number. 2 μl of the amplified library was run on a High Sensitivity D1000 Kit in a TapeStation system. If the results looked fine, the second half of the library was also amplified and combined with the first half. Finally, the concentration under the peak was determined using TapeStation software, and replicates were combined either in equal molarities or equal volumes.

Second size selection by ProNex. Before submitting the samples for sequencing, another round of bead-based size selection was performed to remove residual primers. This ProNex size selection was performed as described above with the following modifications: After ULR preparation, the samples and beads were mixed in a 1:2.4 (v/v) ratio of sample:beads. This was optimized in previous experiments in (24). After the incubation and washing steps, the dried beads were eluted in 20 μl H_2O. Again, for comparison the selected and unselected ULR Phusion mix was analyzed as described previously. The 100-nt/75-nt ladder fragment ratio should be around 4.5.

SELECT experiments to validate m^{6}A modifications

We used the elongation and ligation-based qPCR amplification method SELECT (29) to independently test for m^{6}A modifications at several putative m^{6}A sites identified from our miCLIP2 data. Experiments for mESC cells were performed with RNA from mESC WT cells and compared to RNA from mESC *Mettl3* KO cells. Experiments for HEK293T cells were performed with RNA from cells treated with 20 μM METTL3 inhibitor STM2457 (STORM Therapeutics) (28) or DMSO alone as control (see above).

Normalization of input RNA. For *Mettl3* KO or METTL3 inhibitor-treated cell lines, the amount of m^{6}A is greatly reduced. Due to m^{6}A-mediated RNA degradation or stabilization processes, absence of m^{6}A may influence the abundance of specific transcripts. To ensure usage of same RNA amounts, Qubit (Thermo Fisher Scientific) with Qubit™ RNA HS Assay Kit (Thermo Fisher Scientific) was used to precisely measure RNA concentrations. To ensure usage of equal amounts of transcripts, qPCR experiments were performed for normalization of input RNA amounts in WT versus m^{6}A-depleted cell lines.

Elongation and ligation-based qPCR amplification. For the quantitative real-time PCR (qPCR)-based validation of a presumed m^{6}A site (termed X site), two primers (Up and Down primer) were designed flanking the site of interest. To precisely measure RNA concentrations before each experiment, Qubit™ RNA HS Assay Kit (Thermo Fisher Scientific) was used. An influence of m^{6}A on transcript stability may lead to a difference in transcript abundance upon *Mettl3* KO. Therefore, qPCR for the respective transcript was performed and the amount of total RNA for each SELECT experiment was normalized. To further monitor usage of equal amounts of input material, an Up and Down primer were designed flanking an adjacent nucleotide (termed N site). N sites between X-8 and X+4 were used as

input control. According to the previously published SELECT method, 20 ng of poly(A)+ RNA was used per experiment. The RNA was mixed in a total volume of 17 μl in 1xCutSmart buffer containing 40 nM Up primer, 40 nM Down primer and 5 μM dNTPs. The RNA and primers were annealed by incubation in a thermocycler (90°C to 40°C with a decrease of –10°C after 1 min, then left at 40°C for 6 min). 0.02 U Bst 2.0 DNA polymerase, 0.5 U SplintR ligase and 10 nmol ATP in a volume of 3 μl in 1× CutSmart buffer was added and incubated at 40°C for 20 min. After denaturation at 80°C for 20 min, the mixture was kept at 4°C. Using the Applied Biosystems ViiA7 Real-Time PCR system, qPCR was performed. The 20 μl qPCR reaction mixture contained 2 μl of the final reaction mixture after denaturation, 0.2 nM per qPCR primer, 2x Luminaris HiGreen Lox Rox (Thermo Fisher Scientific) and ddH_2O. The quantitative qPCR reaction condition was run as follows: 95°C, 5 min; (95°C, 10 s; 60°C, 35 s) x 40 cycles; 95°C, 15 s; 60°C, 1 min; 95°C, 15 s (collect fluorescence at a ramping rate of 0.05°C/s); 4°C hold. qPCR data analysis was performed using QuantStudio Real-Time PCR Software v1.3. All experiments were performed in three technical replicates (separate SELECT reactions). Oligonucleotides used for SELECT are listed in Supplementary Table S2.

RT-PCR quantification of intron retention isoforms

Reverse transcription followed by polymerase chain reaction (RT-PCR) was performed to validate changes in isoform frequencies of selected transcripts (*Ythdc1*, *Mif4gd*) comparing *Mettl3* KO and WT mESCs. Cells were grown on irradiated CF1 mouse embryonic fibroblasts (A34181, Gibco) under normal FCS/LIF conditions, as described before (27). Total RNA was isolated from feeder-depleted mESCs using the RNeasy Plus Kit after removal of genomic DNA with gDNA eliminator columns (Qiagen). Random hexamer primers were used to reverse transcribe 1 μg of total RNA into cDNA using the RevertAid First Strand cDNA Synthesis Kit (Thermo Fisher Scientific) in a thermocycler at 65°C for 5 min, 25°C for 5 min, 42°C for 60 min, 45°C for 10 min, and 70°C for 5 min. Three-primer PCR reactions were performed with One*Taq* DNA Polymerase (New England Biolabs) in a 25 μl reaction, according to the recommended protocol, using 0.5 μl cDNA as template, a shared forward primer located in the upstream exon and two isoform-specific reverse primers in the intron (IR) and the downstream exon (spliced), respectively. All three primers were used in a final concentration of 200 nM each, rendering the shared primer as a rate-limiting factor in the reaction. Primer sequences were: Ythdc1_shared (5′-CCATCCCGTCGAGAACCAG-3′), Ythdc1_IR (5′-CCAACGTGACCATGTGAAATCC-3′), Ythdc1_exonic (5′-TGGTCTCTGGTGAAACTCAG G-3′), Mif4gd_shared (5′-CCTGAGAGTCTGAGCAGG GA-3′), Mif4gd_IR (5′-AAGCCTTGGCCTCTATGTGC-3′) and Mif4gd_exonic (5′-AGCCGTCCCGGATTAGGA TA-3′). The PCR reaction was carried out in a thermocycler at 94 °C for 30 s, 30 cycles of [94°C for 30 s, 55°C (*Mif4gd*) or 54°C (*Ythdc1*) for 1 min, 68°C for 1 min] and final extension at 68°C for 5 min. PCR products were analyzed by capillary gel electrophoresis on the TapeStation 2200 system using D1000 ScreenTapes (Agilent) according to the manufacturer's recommendations. Band intensities were quantified using the TapeStation Analysis Software and frequency was calculated as the relative proportion of IR and spliced transcript abundance.

miCLIP2 read processing

Multiplexed miCLIP2 libraries were sequenced as 91-nt or 92-nt single-end reads on an Illumina NextSeq500 sequencing system including a 6-nt sample barcode as well as 5-nt+4-nt unique molecular identifiers (UMIs).

Initial data processing was done as described in Chapters 3 and 4.1 of (30) for iCLIP data. In short, after checking the sequencing qualities with FastQC (v0.11.8) (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and filtering reads based on sequencing qualities (Phred score) of the barcode region (FASTX-Toolkit v0.0.14) (http://hannonlab.cshl.edu/fastx_toolkit/), seqtk v1.3 (https://github.com/lh3/seqtk/), reads were de-multiplexed based on the experimental barcode (positions 6 to 11 of the reads) and adapter sequences were removed from the read ends (Flexbar v3.4.0) (31). UMIs were trimmed as well and added to the read names. Reads shorter than 15 nt were removed from further analysis. Individual samples were then mapped to the respective genome (assembly version GRCh38.p12 for all human samples, GRCm38.p6 for all mouse samples) and its annotation (GENCODE release 31 for all human samples, GENCODE release M23 for all mouse samples) (32) using STAR (v2.7.3a) (33). When running STAR (with parameter --outSAMattributes All), up to 4% mismatches were allowed per read, soft-clipping was prohibited on the 5′ end of reads and only uniquely mapping reads were kept for further analysis. Following mapping, sorted BAM files were indexed (SAMtools v1.9) (34) and duplicate reads were removed (UMI-tools v1.0.0) (35). Reads were defined duplicates if their 5′ ends map to the same position and strand in the genome and they have identical UMIs.

After removing duplicates, all mutations found in reads were extracted using the Perl script parseAlignment.pl of the CLIP Tool Kit (CTK, v1.1.3) (36). The list of all found mutations specifies the mutations, their locations in the genome as well as the names of the reads in which they were found. The list was filtered for C-to-T mutations using basic Bash commands and kept in BED file format as described in (37). Based on the filtered list of C-to-T mutations, deduplicated reads were separated into two BAM files holding reads with and without C-to-T mutation, respectively, using SAMtools and basic Bash commands. The BAM file of reads without C-to-T mutation was transformed to a BED file using bedtools bamtobed (BEDTools v2.27.1) (38) and considering only the 5′ mapping position of each read. Afterwards, the BED file was sorted and summarized to strand-specific bedGraph files which were shifted by one base pair upstream (since this nucleotide is considered as the cross-linked nucleotide) using bedtools genomecov (BEDTools v2.27.1). Similarly, the BED files of C-to-T mutations were also sorted and summarized to strand-specific bedGraph files using bedtools genomecov. Finally, all bed-

Graph files were transformed to bigWig track files using bedGraphToBigWig of the UCSC tool suite (v365) (39).

The code for miCLIP2 data processing as described here is available from two recent data analysis publications (30,37).

Peak calling, transcript assignment and relative signal strength

BAM files with reads without C-to-T mutation were used for peak calling with PureCLIP (v1.3.1) (40) individually on each replicate for each condition. PureCLIP significant sites per replicate were then filtered for presence in at least two replicates for a given condition (PureCLIP peaks in Supplementary Table S1). For assigning a host gene to each PureCLIP peak, transcript annotations were taken from GENCODE (release 31, GRCh38.p12 for human and release M23, GRCm38.p6 for mouse), and filtered for a transcript support level ≤ 3 and support level ≤ 2. For overlapping transcripts, the longest annotation was chosen. We next assigned the miCLIP2 peaks to the transcripts.

In order to calculate the relative signal strengths of all peaks within a transcript, we calculated the mean number of truncation events for all peaks in the same transcript. Then, we divided the individual truncation read number of each peak by the mean of the peak strength in the corresponding transcript, leading to a value representing the relative peak strength.

Differential methylation analysis to identify Mettl3-dependent m^{6}A sites

Similar to iCLIP, the miCLIP2 signal is strongly influenced by the underlying transcript abundance (41,42). Therefore, when applying DESeq2 (43) collectively to all peaks (*one-run*), any change of transcript abundance will lead to incorrect fold change and FDR estimations, resulting in false positive calls in down-regulated genes. We tested four different approaches to overcome this, namely separately running DESeq2 on peaks of individual genes (*gene-wise*) or groups of genes with similar abundance change (*bin-based*), by building a combined DESeq2 model on peak signals and transcript counts using interaction terms (*2-factor*) as well as by using DEXSeq (*dexseq-run*) (44) instead of DESeq2. The different approaches are explained in more detail in the Supplementary Material. The best performance was seen for the *bin-based* approach, which was used for all following analyses.

Training and evaluation of the machine learning model m6Aboost

Based on the $\log_2$-transformed fold change (log2FC) and the false discovery rate (FDR) from the *bin-based* differential methylation analysis between WT and *Mettl3* KO cells, we used peaks at A to compile a positive ($\log_2$FC < 0, FDR ≤ 0.01; $n = 11{,}707$) and negative ($\log_2$FC ≥ 0, FDR > 0.5; $n = 42{,}090$) set. Both were combined and then randomly split into a training set (80%) and an independent test set (20%). We then extracted 27 features, including the nucleotide sequence in a 21-nt window around the central A, the transcript region as well as the relative signal strength ($\log_2$) and the number of associated C-to-T transitions ($\log_2$). We initially tested three different machine learning algorithms (AdaBoost, support vector machine [SVM], random forest) and evaluated their performance based on precision-recall curves and area under the curve (AUC) as well as by comparing F1-score, Matthews correlation coefficient (MCC), precision, accuracy, sensitivity and specificity on the independent test set. Based on these measures, we selected the AdaBoost-based predictor, which we named m6Aboost (see Supplementary Material, Section B for details).

RNA-seq read processing

RNA sequencing (RNA-seq) libraries were sequenced on an Illumina NextSeq500 as 84-nt single-end reads, yielding 31–35 million reads per sample. Basic sequencing quality checks were applied to all reads using FastQC (v0.11.8) (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Reads were mapped to the mouse genome (assembly version GRCm38.p6) and its annotation based on GENCODE release M23 using STAR (v2.6.1b) (33). When running STAR, up to 4% mismatches were allowed per read and only one location was kept for multi-mapping reads. Coverage tracks for visualization were obtained by merging BAM files for each condition using SAMtools (v1.11). Coverage was calculated with bamCoverage (v3.5.0) from the deepTools suite (45) using RPGC normalization and --effectiveGenomeSize calculated by faCount of the UCSC tool suite (v377).

For differential gene expression analysis, mapped reads were counted with htseq-count (v0.12.4, -s reverse) (46) into gene annotation based on GENCODE release M23. Differential expression analysis was performed with DESeq2 (v1.30.0) (43) using the method 'apeglm' for shrinkage of $\log_2$-transformed fold changes.

Intron retention (IR) analysis was done with IRFinder (v1.3.0) (47) using built-in script analysisWithLowReplicates.pl for differential analysis (48). We adapted some built-in filtering steps by overwriting line 179 of analysisWithLowReplicates.pl into:

my $ok = ($pA[8] > 0 || $pB[8] > 0) && ($pA[19] > 0 || $pB[19] > 0) && separatedAB(\@repsIR, $repsA, $repsB);

and line 186 into:

if (($pA[8] > 0 || max($pA[16],$pA[17]) > 0) && ($pB[8] > 0 || max($pB[16],$pB[17]) > 0)) {

For downstream analysis, IR events were filtered for IR-ratio ≥ 0.03 in at least one condition and mean IntronDepth ≥ 3. *P* values were corrected using Benjamini-Hochberg adjustment.

Overlap with MAZTER-seq

Processed MAZTER-seq data from (21) were downloaded from Gene Expression Omnibus (GEO) via accession number GSE122956. The m^{6}A sites therein were filtered for a difference in MazF cleavage efficiency > 0.1 between WT

and *Mettl3* KO, yielding a total of 580 reliably identified m^{6}A sites from mESC cells. Two hundred of these (34.5%) overlapped at single-nucleotide resolution with the 4,464 predicted m^{6}A sites at ACA from our mESC miCLIP2 data.

YTHDF1 iCLIP processing and overlap with predicted m^{6}A sites

YTHDF1 iCLIP reads were quality filtered and processed as in Busch *et al*. (30), used tools versions are as described above for miCLIP2. For peak calling with PureCLIP (40) reads from the four replicates were merged. Resulting peaks were filtered to be present in at least two out of four replicates. To generate binding sites, peaks closer than 4 nt were merged, allowing no overlapping binding sites. Finally, binding sites were centred at the position with the highest truncation read number as described in (30). All predicted m^{6}A sites were aligned and spanned with a 21-nt window to count the presence of YTHDF1 binding sites in that area.

RESULTS

The miCLIP2 protocol allows profiling of m^{6}A RNA modifications

In order to allow for deep m^{6}A profiling, we combined the miCLIP procedure with our recently optimized iCLIP2 protocol, termed miCLIP2 (Figure 1A) (17,24). Experiments were performed with poly(A)+ RNA from mouse embryonic stem cells (mESCs). We first performed two consecutive rounds of poly(A)+ RNA enrichment for total RNA samples (Supplementary Figure S1A) and optimized the RNA fragmentation time required for each sample (Supplementary Figure S1B). The RNA was then incubated with an m^{6}A-specific antibody (Synaptic Systems), which was previously shown to yield highest truncation efficiency in miCLIP experiments (Figure 1A) (17). After optimizing UV irradiation (254 nm twice with 150 mJ/cm^2 strength; Supplementary Figure S1C), crosslinked antibody-RNA complexes were immunoprecipitated using protein A beads. Co-purified RNAs were 3′dephosphorylated with T4 polynucleotide kinase (PNK) prior to first adapter ligation (L3-APP) and radioactive labelling. After SDS-PAGE gel and transfer, the respective nitrocellulose membrane fragment was excised (Supplementary Figure S1D). Transferred RNA was recovered by proteinase K treatment, leaving a polypeptide at the crosslinking site. Reverse transcription generally truncates at this polypeptide, thus encoding the positional information about m^{6}A sites within resulting cDNA fragments (17,49). The residual readthrough events usually incorporate C-to-T transitions (17), which provide additional confidence for truncation-identified crosslink sites (see below). After bead-based clean-up and second linker ligation, a pre-amplification PCR (6 cycles) was employed to minimize loss of information by potential material loss in the following steps. This was followed by size selection to remove primer dimers and a second PCR which was optimized for a minimal number of PCR cycles to obtain sufficient material for sequencing (here 11 cycles). After a second size selection to remove remaining primers, the library was subjected to high-throughput sequencing (Supplementary Figure S1E).

The majority of miCLIP2 peaks are not sensitive to *Mettl3* KO

In order to test whether miCLIP2 peaks are dependent on Mettl3, we performed miCLIP2 experiments ($n = 3$ replicates) from wild-type (WT) as well as *Mettl3* knockout (KO) mESCs. The latter lack the primary m^{6}A methyltransferases Mettl3 and hence, lost most of m^{6}A mRNA methylation (Figure 1B) (27,50). Reads with C-to-T transitions (6%) were removed for later usage (Supplementary Table S1). The remaining reads corresponded to a total of 261 million putative truncation events (Supplementary Table S1). Peak calling on the data from WT mESC cells identified $>$ 500,000 peaks that exceeded the local background signal (peaks on all samples are reported in Supplementary Table S1). The number of truncation events in called peaks were highly reproducible between replicates (Figure 1C and Supplementary Figure S2A). To allow for quantitative comparisons between transcripts, we calculated the relative signal strength of all peaks, which was independent of the underlying transcript abundance (see Materials and Methods; Supplementary Figure S2B).

Analysis of the underlying sequence showed that most peaks resided on uridine rather than adenosine and only 25% of these adenosines were part of a DRACH motif (Figure 1D–G), reflecting UV crosslinking biases and limited antibody specificity as reported previously (20,21). Nevertheless, the strongest peaks frequently coincided with AC and were located precisely on the A nucleotide (Supplementary Figure S2C). We noted an additional enrichment of AC downstream of the peaks. However, these particular peaks did not harbor a DRACH motif and their signal was not reduced in the *Mettl3* KO, indicating that they are part of the unspecific background signal of the employed antibody or m^{6}A sites independent of Mettl3 (Supplementary Figure S2C). Importantly, peaks at A, AC and DRACH motifs were specifically lost in the *Mettl3* KO, supporting that miCLIP2 detects Mettl3-dependent m^{6}A modifications (Figure 1E–G and Supplementary Figure S2D). In addition to the putative m^{6}A sites, we observed an accumulation of miCLIP2 truncation events at transcript start sites which did not respond to the *Mettl3* KO (Supplementary Figure S2E and F). This likely reflected the related RNA modification N6,2′-O-dimethyladenosine (m^{6}Am) which is known to reside at 5′ cap structures and is also recognized by the m^{6}A-specific antibody (17). Overall, the high amount of non-specific background and cross-reactivity in the miCLIP2 data required more precise measures to define true Mettl3-dependent m^{6}A sites.

Differential methylation analysis detects Mettl3-dependent m^{6}A sites at DRACH and non-DRACH motifs

In order to learn about the features of genuine m^{6}A sites in the miCLIP2 data, we sought to extract all miCLIP2 peaks that significantly changed in the *Mettl3* KO mESCs. However, changes at individual peaks were overshadowed by massive shifts in gene expression in *Mettl3* KO cells, with

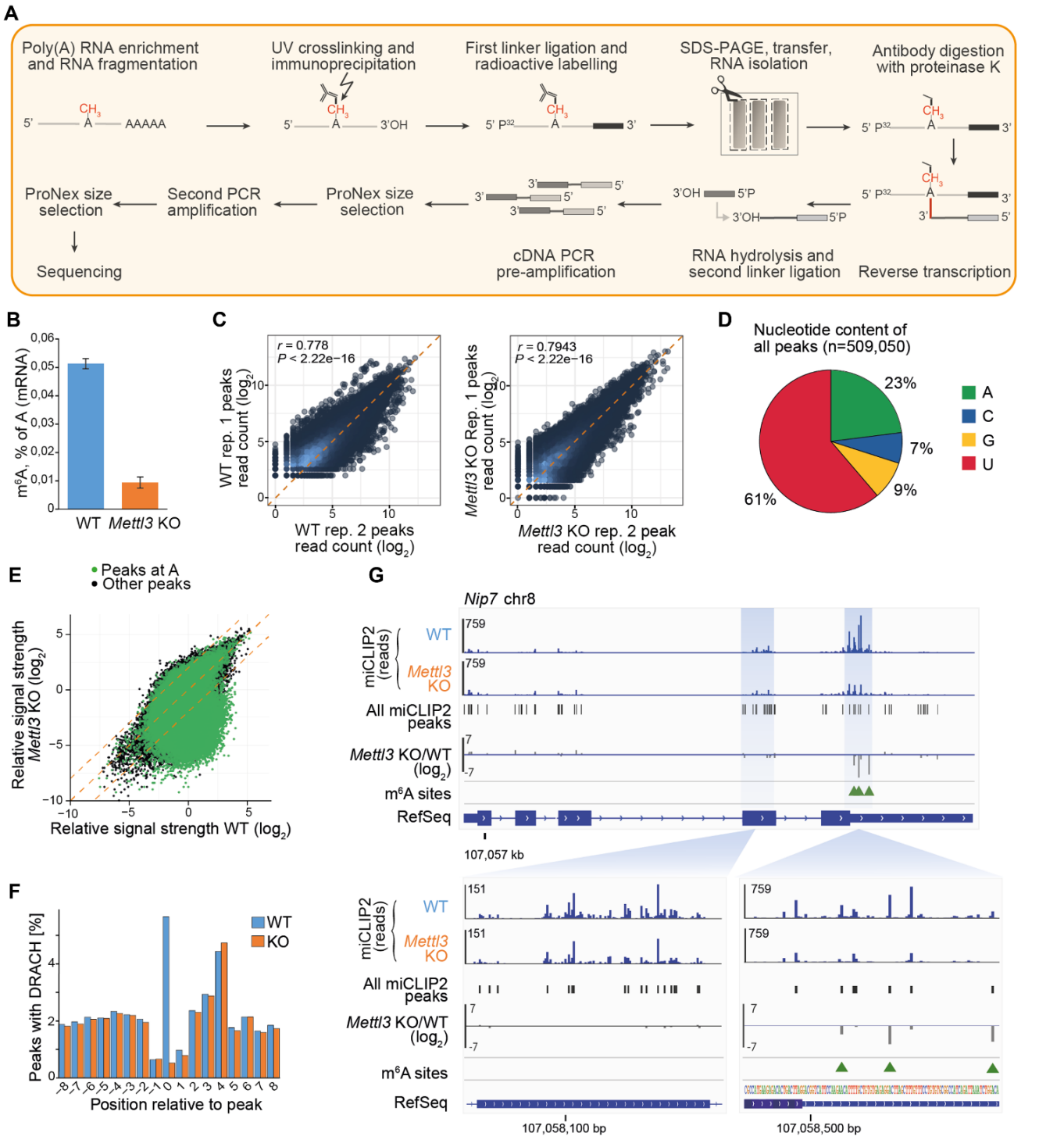


Figure 1. The optimized miCLIP2 protocol produces high complexity libraries with high reproducibility. (**A**) An overview of the miCLIP2 protocol. (**B**) mESC *Mettl3* KO cells show a significant depletion of m^{6}A on mRNAs. m^{6}A levels measured by liquid chromatography-tandem mass spectrometry (LC-MS/MS) for poly(A)+ RNA from WT and *Mettl3* KO mESCs. Quantification of m^{6}A as percent of A in mRNA. Error bars indicate standard deviation of mean (s.d.m.), $n = 3$. (**C**) miCLIP2 data are highly reproducible between replicates. Pairwise comparison of the miCLIP2 truncation reads within peaks from two miCLIP2 replicates from WT and *Mettl3* KO mESCs. Pearson correlation coefficients (r) and associated P values are given. Additional replicates are shown in Supplementary Figure S2A. (**D**) Most peaks are located at uridines and adenines. Pie chart representing the nucleotide distribution of all miCLIP2 WT peaks. (**E**) The majority of peaks are unchanged in a *Mettl3* KO miCLIP2 experiment, indicating high background signal. Scatterplot of the $\log_2$-transformed relative signal strength (corrected for transcript abundance) of all miCLIP2 peaks in WT and *Mettl3* KO mESC. Peaks located at an A are highlighted in green. Dotted lines indicate diagonal and 4-fold change. (**F**) DRACH motifs are enriched at miCLIP2 WT peaks. Metaprofile of DRACH motifs around aligned miCLIP2 peaks (position 0). Percentage of DRACH motifs (counted at position of A within DRACH) around the miCLIP2 peaks of WT and *Mettl3* KO mESCs are shown. (**G**) *Mettl3* KO miCLIP2 signal is reduced at specific positions in the *Nip7* 3′ UTR. Genome browser view of miCLIP2 data (blue) from WT and *Mettl3* KO mESCs and fold change between conditions (grey). Identified miCLIP2 peaks (black bars) and m6Aboost-predicted m^{6}A sites (green arrowheads) are given. Zoom-ins (bottom) show more detailed views of an exonic region without m^{6}A sites and a 3′ UTR region with three m^{6}A sites.

more than 2,809 genes altered at least 2-fold in comparison to WT mESCs (false discovery rate [FDR] ≤ 0.01; Figure 2A). These massive shifts in the underlying transcript abundances meant that miCLIP2 read counts at individual peaks could not be compared directly. In order to overcome this shortcoming, we tested several strategies for differential methylation analysis to account for the substantial gene expression changes in the *Mettl3* KO cells (see Supplementary Material, Section A). Best performance was achieved with the *bin-based* approach, in which genes were stratified according to their expression change upon *Mettl3* KO (Figure 2B and Supplementary Figure S3A-C). All miCLIP2 peaks within the genes of the same bin, i.e., with a similar change in gene expression, were then tested collectively using DESeq2 (43) (see Supplementary Material, Section A). As expected, the changing peaks almost exclusively showed a loss of miCLIP2 signal in the *Mettl3* KO (Figure 2C), and 85.3% of these downregulated peaks were located at A (Figure 2D), supporting that our differential methylation analysis enriched for m^{6}A sites. From these, we compiled a stringent set of 11,707 sites at A with reduced signal in the *Mettl3* KO ($\log_2$-transformed fold change [log2FC] < 0, FDR ≤ 0.01), which served as 'positive set' of true m^{6}A sites in the following analyses (see Supplementary Material, Section A). As previously described, the positive sites accumulated nearby stop codons and in 3′ UTRs, and the underlying sequences resembled the DRACH motif (16,51) (Figure 2E and F), supporting that they indeed represented Mettl3-dependent m^{6}A sites. For comparison, we selected a 'negative set' of 42,090 peaks that were also located at A but unchanged or even mildly increased upon *Mettl3* KO (log2FC ≥ 0, FDR > 0.5) and hence represented the nonspecific background in the data.

Among the DRACH motifs identified in the positive set, the most frequent pentamer was GGACT, followed by GAACT and AGACT (17) (Figure 2G). Surprisingly, however, we also detected 741 m^{6}A sites (6.3%) at non-DRACH motifs (non-DRACH m^{6}A). While most of these non-DRACH motifs still contained the AC dinucleotide (52), some also diverged from this, such as GGATT (Figure 2G). We used SELECT (single-base elongation- and ligation-based qPCR amplification) (29) as an orthogonal antibody-independent m^{6}A detection method to test the reliability of our approach. To this end, we compared SELECT qPCR amplification curves from WT versus *Mettl3* KO samples for an exemplary non-DRACH m^{6}A site from the positive set, located in the last exon of the *Trim27* gene (A at position chr13:21192298:+, GGATT). Indeed, we detected Mettl3-dependent methylation at A in the GGATT motif, reflected in a reduced efficiency of the qPCR amplification when the m^{6}A mark is present (Figure 2H). As a control, we tested an adjacent A in the same gene (position chr13:21192294:+), which remained unchanged upon *Mettl3* KO (Supplementary Figure S3D). We similarly validated two out of two additional non-DRACH m^{6}A sites in the genes *Palm3* (chr8:84029842:+, GTACT) and *Hic2* (chr16:17257755:+, GGACG) (Figure 2H and Supplementary Figure S3D). For comparison, we also confirmed three out of three m^{6}A sites at *bona fide* DRACH motifs in the genes *Eif4ebp1* (chr8:27275332:+, TGACT), *Ccnt2* (chr1:127802764:+, GAACA) and *Phb2* (chr6:124716745:+, GAACT) (Figure 2I and Supplementary Figure S3D).

DRACH motifs were also present at 1,043 peaks (2.5%) in the negative set. The miCLIP2 signal at these peaks did not decrease in the *Mettl3* KO, indicating that the antibody may show a residual background activity against the DRACH motif itself. SELECT experiments for two out of two selected sites in the genes *Nanog* (chr6:122711605:+) and *Zfp710* (chr7:8008671:+) confirmed that the respective A indeed did not carry an N6-methyl modification (Figure 2J).

All together, we defined a positive set of > 10,000 m^{6}A sites, that are modified in a Mettl3-dependent manner. In addition to canonical DRACH motifs, we identified a fraction of m^{6}A modifications at non-DRACH motifs which show the same characteristics and Mettl3 dependency as m^{6}A sites at DRACH motifs.

Machine learning allows to reliably predict m^{6}A sites from miCLIP2 data

To allow for m^{6}A detection independently of an accompanying KO dataset, we built a predictive machine learning model to discriminate true m^{6}A sites from background signal in the miCLIP2 data (Figure 3A). For model training, we combined the positive (n = 11,707) and negative (n = 42,090) sets identified in the differential methylation analysis upon *Mettl3* KO. The unbalanced setup was chosen to reflect the predominance of nonspecific background in the miCLIP2 data (Figure 1D–G). We randomly split the data into a training set (80%) and an independent test set (20%). The input variables for training included 10-nt flanking nucleotide sequence to either side of A, the transcript region and the relative signal strength. We further added, as orthogonal information, the number of coinciding C-to-T transitions in the read-through reads, which we initially removed from the data (Figure 3B, see Supplementary Material, Section B).

We tested three different machine learning algorithms, which consistently reached high predictive accuracy (support vector machine, random forest, and adaptive boosting [AdaBoost]; Supplementary Figure S4A–E, see Supplementary Material, Section B). Following a series of benchmarks, we chose the AdaBoost-based predictor, which we named m6Aboost. AdaBoost is a boosting ensemble algorithm that weights the input for each iteration by the misclassification errors from previous iterations, and thereby improves the accuracy of the final predictions (53). The error rate of m6Aboost on the independent test set reached 0.99%, with > 99% area under the curve (AUC) in a precision–recall curve (Figure 3C and Supplementary Figure S4A and D). Evaluation on an independent test set showed that 99% of sites were correctly classified (Figure 3D). The performance was confirmed by five-fold cross-validation (Supplementary Figure S4C). The highest informative content was attributed to the immediate sequence around the modified A nucleotide, the relative signal intensity of peaks, and orthogonal information on C-to-T transitions (Figure 3B). Baseline models trained only on sequence information (position –10 to +10; 'sequence-only') or experimental features (relative signal strength, C-to-T transitions, and

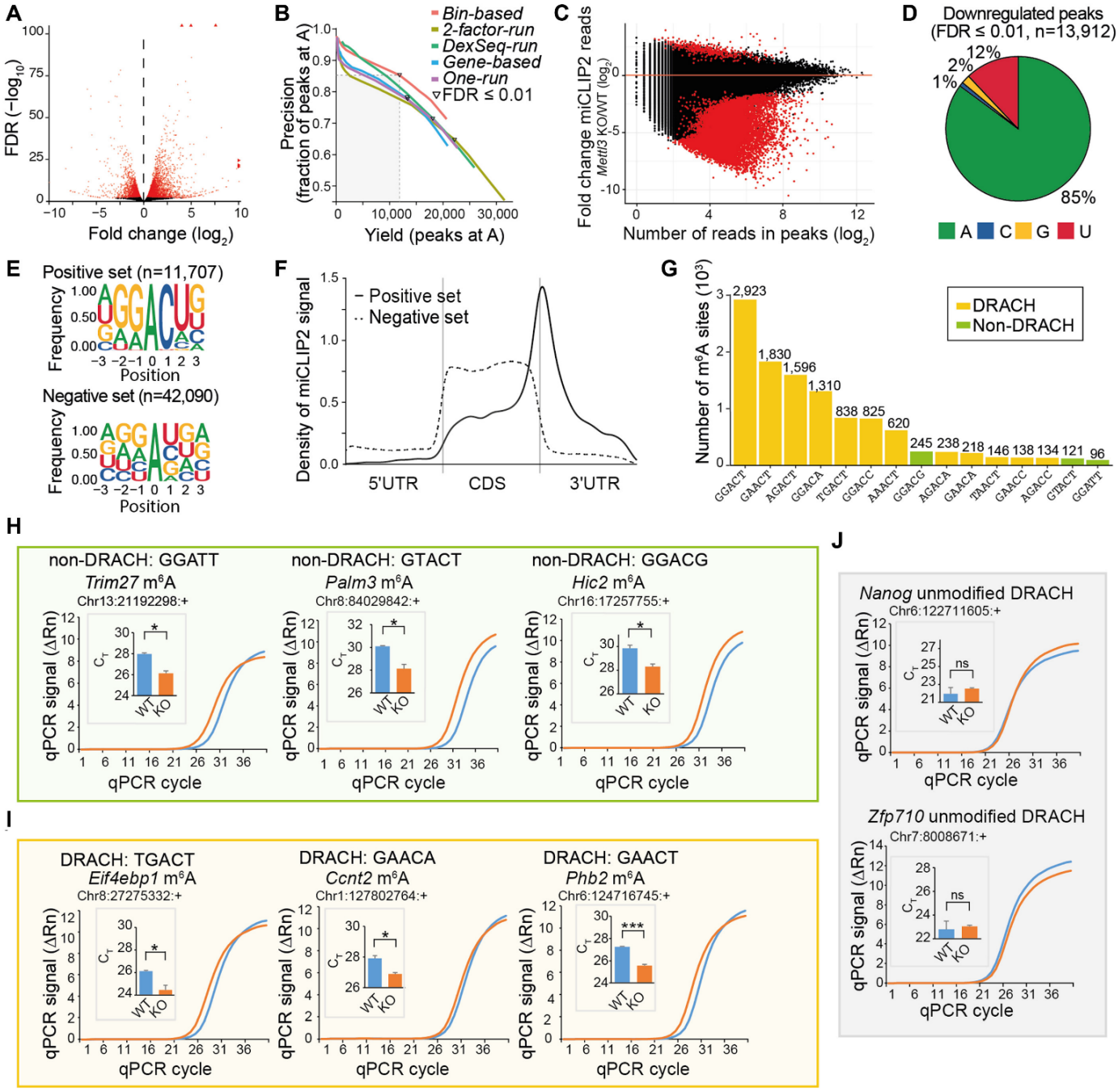


Figure 2. Differential peak analysis allows to identify true m^{6}A sites from miCLIP2 data. (**A**) *Mettl3* KO causes drastic changes in gene expression. Volcano plot shows negative $\log_2$-transformed fold change (log2FC) of gene expression between *Mettl3* KO and WT against $\log_{10}$-transformed false discovery rate (FDR). Significantly changing genes are highlighted in red (FDR $\leq$ 0.01). (**B**) The *bin-based* approach for differential methylation analysis outperforms other tested strategies. Number of identified peaks at A (x-axis) and fraction of peaks at A (y-axis) are given for different approaches (see Supplementary Material, Section A). Curves were generated by step-wise increases in stringency (FDR). FDR $\leq$ 0.01 is marked for each approach. (**C**) Most changing peaks go down upon *Mettl3* KO. Comparison of log2FC in miCLIP2 signal per peak between *Mettl3* KO and WT (y-axis) against reads per peak ($\log_2$-transformed, x-axis). Significantly regulated peaks are highlighted in red ($|\log_2 FC| > 1$, FDR $\leq$ 0.01). (**D**) Most significantly downregulated peaks are located at adenosines. Pie chart represents nucleotide distribution of downregulated peaks. (**E**) Sequence motifs of peaks in the positive (top) and negative (bottom) set. Logos show relative frequency of nucleotides at positions –3 to +3 around central A. (**F**) Peaks in the positive set accumulate around stop codons. Density plot shows distribution of peaks in scaled transcript regions. UTR, untranslated region, CDS, coding sequence. (**G**) The most frequent pentamers include non-DRACH motifs. Number of peaks (positive set) located at specific pentamer at DRACH (orange) and non-DRACH (olive) motifs. (**H–J**) Selected m^{6}A sites were validated by SELECT experiments. Exemplary real-time fluorescence amplification curves (normalized reporter value, ΔRn) and quantifications of threshold cycle (C_T) values (technical replicates) for mESC WT versus *Mettl3* KO samples are shown for m^{6}A sites at non-DRACH (H) and DRACH (I) motifs as well as unmodified DRACH motifs with a miCLIP2 peak (J). Neighboring unmodified A nucleotides as control for each tested site are given in Supplementary Figure S3D. *** P value < 0.001, * $P < 0.05$, ns, not significant, two-sided Student's t-test, $n = 3$.

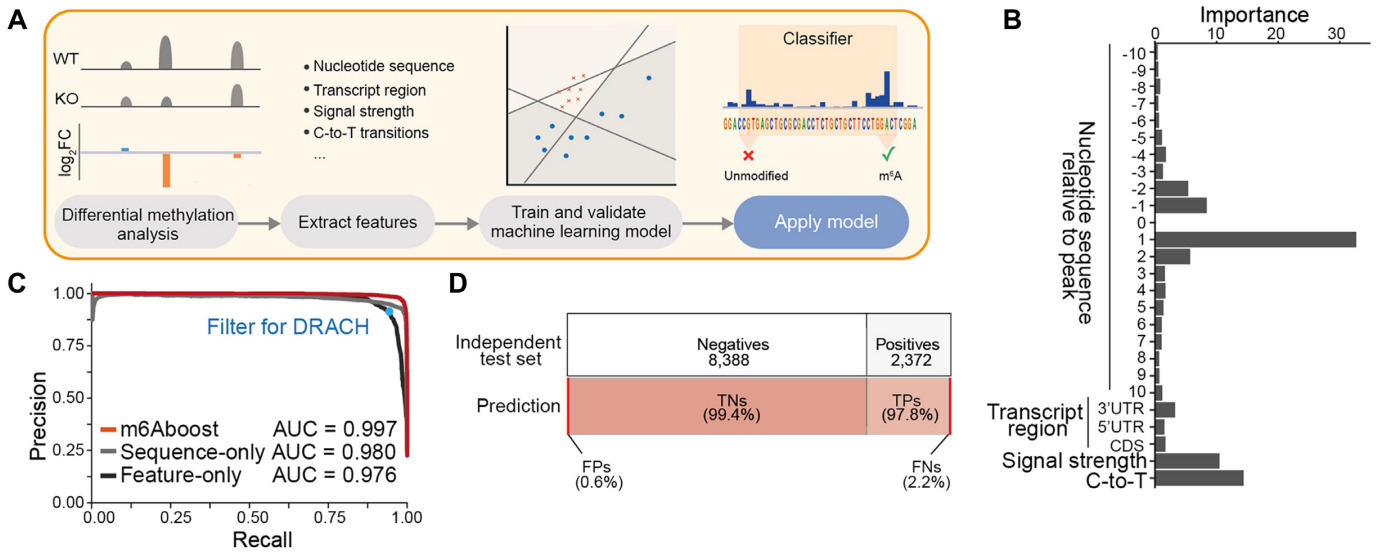


Figure 3. The machine learning classifier m6Aboost reliably predicts m^{6}A sites from miCLIP2 data. (**A**) Overview of the machine learning approach. First, miCLIP2 WT and *Mettl3* KO datasets are analyzed for differential methylation to identify Mettl3-dependent m^{6}A sites. The resulting positive and negative sets are used to extract features and train a machine learning classifier. The model is validated on an independent test set. Finally, the model can be applied to new miCLIP2 datasets to classify the miCLIP2 peaks as modified m^{6}A sites versus unmodified background signal. (**B**) Highest informative content lies in the nucleotide sequence, the relative signal strength of the peak and the number of C-to-T transitions. Bar plot shows the features used for m6Aboost prediction and their associated importance ranking. UTR, untranslated region, CDS, coding sequence. (**C**) m6Aboost outperforms baseline models trained only on sequence (sequence-only) or experimental features (feature-only). Precision-recall curve shows performance of m6Aboost compared to baseline models with the corresponding area under the curve (AUC). Precision and recall when solely filtering for DRACH motifs are shown for comparison (blue dot). (**D**) m6Aboost achieves 99% accuracy on an independent test set. Bars visualize composition of independent test set ($n = 10{,}760$) from positive (22%) and negative (78%) peaks (top) and the resulting m6Aboost predictions (bottom). In total, 10,658 peaks (99%) were correctly predicted, while 102 peaks were misclassified. TNs, true negatives, TPs, true positives, FNs, false negatives, FPs, false positives.

transcript region; 'feature-only') achieved worse classification results (Figure 3C), supporting that both types of features are required for optimal performance. Consistently, our m6Aboost outperformed a simple filter for DRACH motifs (Figure 3C, blue dot).

m6Aboost predicts m^{6}A sites also in lowly expressed transcripts

To test the algorithm on a complete miCLIP2 dataset, we applied m6Aboost to all peaks on A nucleotides in the mESC WT miCLIP2 data (n = 117,142). In total, m6Aboost extracted 25,456 putative m^{6}A sites in 9,363 genes (Figure 4A). These included 11,548 sites from our initial positive set (98.6% of positive set) plus 13,908 additional m^{6}A sites. The latter were enriched in lowly expressed genes and most likely failed to reach significance in the differential methylation analysis due to low read counts (Supplementary Figure S4F). The miCLIP2 signal in all sites coherently went down in the *Mettl3* KO (94% with $\log_2$FC < –1; Figure 4B), supporting that they are indeed true m^{6}A sites.

Of note, 1,813 out of 25,456 (7.1%) predicted m^{6}A sites resided at non-DRACH motifs (Figure 4A). These non-DRACH m^{6}A sites showed an enrichment nearby stop codons similar to the positive set and the vast majority were depleted in the *Mettl3* KO (Figure 4C and D), supporting that predicted non-DRACH sites are indeed true m^{6}A sites. On the other hand, m6Aboost predicted that not all peaks at DRACH motifs corresponded to true m^{6}A sites. Indeed, about half of these sites did not respond to *Mettl3* KO and distributed similarly to the negative set (Figure 4D and Supplementary Figure S4G), suggesting that the m^{6}A-specific antibody shows a residual activity towards the unmodified DRACH motif. The other half had low read counts and preferentially resided in lowly expressed genes (Supplementary Figure S4G), possibly leading to their misclassification. Importantly, m6Aboost associates a prediction score with each site that allows to minimize the number of false positives, at the expense of false negatives, by tightening the prediction score threshold (Supplementary Figure S4H and I). Altogether, we conclude that m6Aboost efficiently discriminates relevant signal from nonspecific background, offering a reliable prediction of genuine m^{6}A sites from miCLIP2 data.

As an orthogonal support, we compared our predicted m^{6}A sites to those detected by the antibody-independent method MAZTER-seq in the same cell line (21). MAZTER-seq relies on the methylation-sensitive RNase MazF which cleaves at unmethylated ACA motifs. We found that 34.5% of the reliably identified m^{6}A sites from MAZTER-seq (200 out of 580 sites) were also present in our data, further supporting the validity of our approach.

For comparison, we also performed miCLIP2 experiments on poly(A)+ RNA from RAW 264.7 cells, a mouse macrophage cell line (three biological replicates, 29.8 million truncation events on average). Out of 462,073 miCLIP2 peaks, m6Aboost identified a total of 19,301 m^{6}A sites (Supplementary Table S1). Overlay with the mESC data showed that a third of the predicted m^{6}A sites were shared between both cell lines, rising to about 50% when focussing

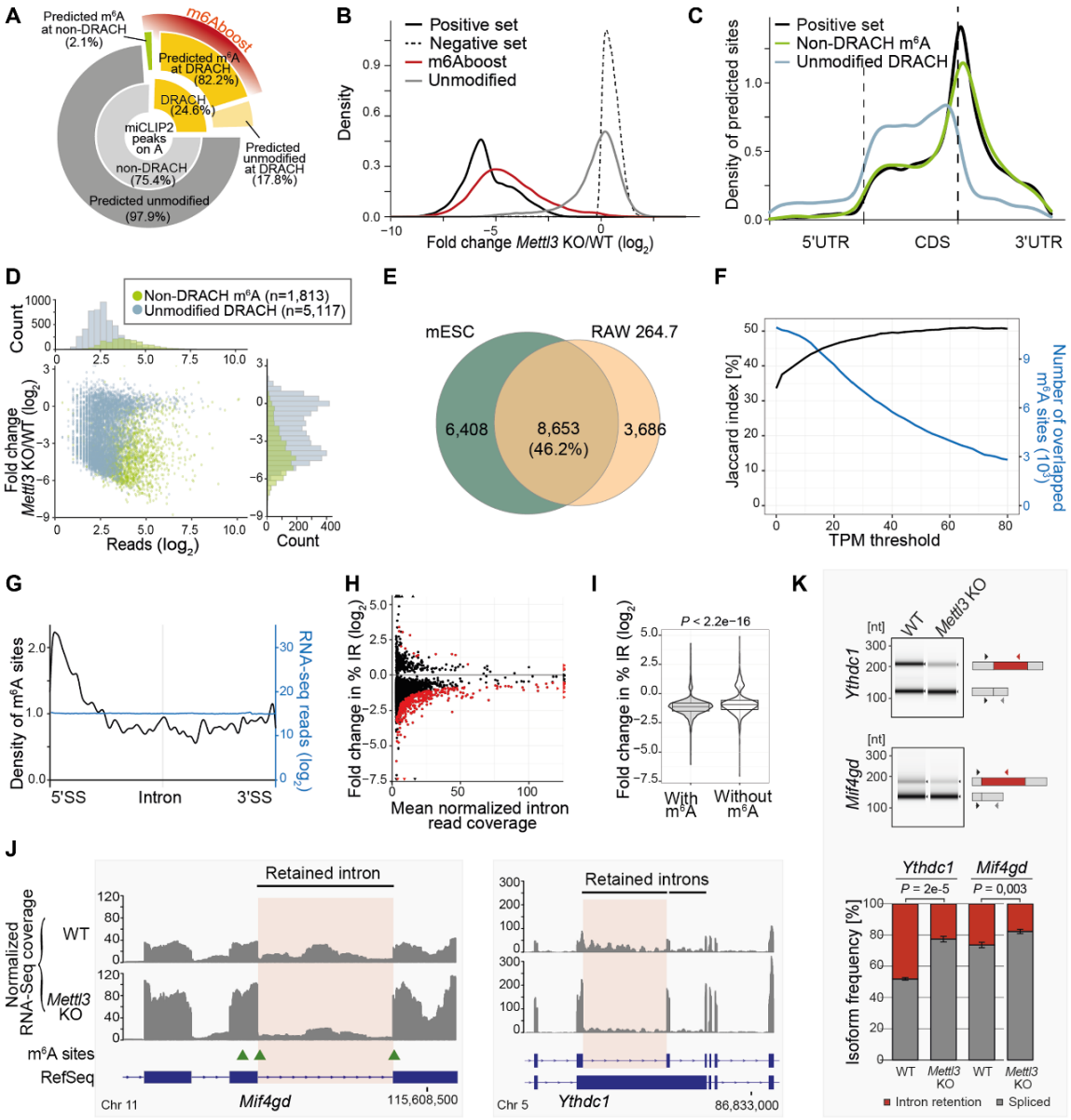


Figure 4. m^{6}A sites occur at non-DRACH motifs and accumulate in retained introns. (**A**) m6Aboost predicts m^{6}A sites at DRACH and non-DRACH motifs in mESC WT miCLIP2 data. Inner circle of donut chart shows occurrence of DRACH (n = 28,760, 24.6%) and non-DRACH (n = 88,382, 75.4%) motifs for all miCLIP2 peaks at A. Outer circle shows m6Aboost prediction results (marked in red) with 23,643 m^{6}A sites and 5,117 unmodified sites at DRACH (82.2% and 17.8%, respectively, of all peaks at DRACH) as well as 1,813 m^{6}A sites and 86,569 unmodified sites at non-DRACH (2.1% and 97.9%, respectively, of all peaks at non-DRACH). (**B**) Predicted m^{6}A sites (n = 25,456) go down upon *Mettl3* KO, whereas predicted unmodified sites (n = 91 686) remain unchanged. Density plot shows distribution of log$_2$-transformed fold changes in miCLIP2 signal between *Mettl3* KO and WT samples. Positive and negative set are shown for comparison. (**C**) m^{6}A sites at non-DRACH motifs (n = 1,813) show a similar accumulation at stop codons as the positive set. Visualization as in Figure 2F. (**D**) m6Aboost predicts that not all peaks at DRACH motifs are m^{6}A sites. Scatter plot and histograms show fold change in miCLIP2 signal (log$_2$-transformed, y-axis) against number of reads per peak (log$_2$-transformed, x-axis) for 5,117 peaks at DRACH motifs (light blue) that are predicted to be unmodified by m6Aboost. Predicted m^{6}A sites at non-DRACH motifs (olive) are shown for comparison. (**E**) Most m^{6}A sites are shared between two mouse cell lines. Venn diagram shows overlap of predicted m^{6}A sites in expressed genes (TPM $\geq$ 20, n = 4,490) from mESC WT and RAW 264.7 cells. Venn diagram without expression filter is shown in Supplementary Figure S5E. (**F**) Overlap of m^{6}A sites between two mouse cell lines increases in higher expressed genes. TPM threshold representing the gene expression (x-axis) against the Jaccard index (y-axis). Numbers of overlapping m^{6}A sites are shown as comparison (blue). (**G**) m^{6}A sites accumulate towards the 5′ splice sites of introns. Metaprofile shows density of m^{6}A sites along scaled introns (n = 3,509 m^{6}A sites on 1,465 different introns). Coverage of RNA-seq reads on the same introns is shown for comparison (blue). SS, splice site. (**H**) Intron retention (IR) is globally reduced in the *Mettl3* KO cells. Scatter plot shows fold change in relative IR (%IR, y-axis) against mean normalized RNA-seq reads on the introns across all samples (x-axis) for 4,925 measured IR events. 401 significantly changed IR events are highlighted in red (FDR $\leq$ 0.05). (**I**) Introns harboring m^{6}A sites show a significant trend towards IR reduction. Violin plot compares fold changes in %IR for retained introns with (n = 4,098) and without (n = 827) m^{6}A sites. P value < 2.22e−16, Wilcoxon rank-sum test. (**J**) IR is reduced in the *Mif4gd* and *Ythdc1* transcripts. Genome browser views of RPGC-normalized RNA-seq coverage are shown for merged replicates from WT and *Mettl3* KO mESCs. Predicted m^{6}A sites are indicated with green arrowheads. IR events validated in (K) are highlighted. (**K**) Frequency of *Ythdc1* and *Mif4gd* IR isoforms is lower in *Mettl3* KO mESCs. Semiquantitative three-primer RT-PCR to quantify isoform frequencies in WT and *Mettl3* KO cells, with shared forward and isoform-specific reverse primers displayed next to corresponding PCR products in capillary gel electrophoresis (top). Quantification of relative band intensities (bottom) is displayed as mean ± s.d.m., n = 3, unpaired two-sided Student's t-test.

on genes that were highly expressed in both cell lines (TPM ≥ 20 or more; Figure 4E and F and Supplementary Figure S5E).

m^{6}A depletion triggers efficient splicing of retained introns

Since our miCLIP2 data was generated for poly(A)-selected RNA, most identified m^{6}A sites were located in exons. However, we also detected a number of m^{6}A sites in retained introns. Interestingly, the intronic m^{6}A sites showed a strong accumulation towards the 5′ splice sites (Figure 4G), suggesting that they might impact intron splicing. Indeed, using IRFinder (47), we could identify 401 significantly changed intron retention (IR) events in the RNA-seq data of *Mettl3* KO mESCs (change in IR [|ΔIR|] > 3%, FDR < 0.05; Figure 4H and I). 384 out of 401 significantly changed introns showed reduced coverage in the *Mettl3* KO, as seen for intron 5 in *Mif4gd* and intron 11 in *Ythdc1* (Figure 4J), indicating increased splicing efficiency. Isoform-specific semi-quantitative RT-PCR confirmed a lower frequency of the *Ythdc1* and *Mif4gd* isoforms with retained introns in *Mettl3* KO mESCs (Figure 4K). This trend was also reflected in a global reduction in IR across the transcriptome, as 4,563 out of 4,925 measured IR events (92.7%) showed a ΔIR < 0 (Figure 4H). Generally, introns harboring m^{6}A modifications showed a significant trend towards more IR reduction compared to unmodified introns (Figure 4I), indicating that modifications on retained introns may directly influence splicing efficiency.

m6Aboost can be applied to predict m^{6}A sites in human cells

To test m6Aboost on miCLIP2 data from a different species, we performed miCLIP2 experiments with poly(A)+ RNA from human HEK293T cells ($n = 4$ replicates with 30 million truncation events on average, Supplementary Figure S1F and G). Starting from > 788,758 miCLIP2 peaks, m6Aboost identified 36,556 m^{6}A sites in 7,552 genes, corresponding to 21% of all peaks at A (Supplementary Table S1). The m^{6}A sites occurred with a median of three sites per gene and accumulated around stop codons (Figure 5A and Supplementary Figure S5A), mirroring the distribution in the mouse cells.

We used SELECT to validate the presence of m^{6}A modifications in HEK293T cells in an antibody-independent manner (29). In order to deplete m^{6}A, we employed a specific METTL3 inhibitor (STM2457, STORM Therapeutics) (28), which progressively reduced the relative m^{6}A levels with increasing concentration, down to 22% (Supplementary Figure S5B). We then compared SELECT qPCR amplification curves from inhibitor-treated HEK293T cells against DMSO control samples for three exemplary m^{6}A sites. This confirmed the presence of m^{6}A in two out of three sites in the genes *DDIT4* (chr10:72275034:+) and *RHOB* (chr2:20448702:+) (Figure 5B). As a control, we tested adjacent A sites in the same genes which remained unchanged upon METTL3 inhibition (*DDIT4*: chr10:72275038:+; *RHOB*: chr2:20448698:+; Supplementary Figure S5C). A third putative m^{6}A site could not be validated (*ABT1*: chr6:26598621:+).

As an independent line of evidence, we overlapped the m6Aboost-predicted m^{6}A sites with binding sites of the cytoplasmic m^{6}A reader protein YTHDF1 from published iCLIP data (54). Metaprofiles showed a sharp peak in YTHDF1 binding precisely at the predicted m^{6}A sites at DRACH motifs (Figure 5C and D and Supplementary Figure S5D). Although less pronounced, we detected considerable YTHDF1 binding also at predicted m^{6}A sites at non-DRACH motifs, further supporting that these indeed represent genuine m^{6}A sites.

We compared our predicted m^{6}A sites in HEK293T with published validated m^{6}A sites in the same cell line by the antibody-independent method SCARLET that uses thin-layer chromatography (52). We found that all m^{6}A sites with > 5% methylation in HEK293T cells were also present in our data, whereas sites that were not validated by SCARLET (< 5% methylation) were not detected by miCLIP2 (Supplementary Table S3). To further support the predicted m^{6}A sites, we compared our miCLIP2 data with published miCLIP and m6ACE-seq data for the same cell line (51,55). m6A-Crosslinking-Exonuclease-sequencing (m6ACE-seq)-seq is a recently developed tool which incorporates 5′ to 3′ exoribonuclease treatment after m^{6}A-antibody crosslinking to increase the resolution and omit radioactive gel electrophoresis (55). We found that almost half of our m^{6}A sites overlapped at single-nucleotide level with at least one further dataset (Figure 5E). The remaining sites occurred on lowly expressed genes, but still showed an m^{6}A-typical distribution along transcripts and overlapped with YTHDF1 binding (Figure 5F and G and Supplementary Figure S5F). This suggests that these m^{6}A sites were missed in other studies due to experimental variability and technical limitations rather than lack of modification.

As a second human cell line, we performed miCLIP2 experiments on poly(A)+ RNA from C643 cells, a human thyroid cancer cell line (three biological replicates, Supplementary Table S1). Here, m6Aboost predicted a total of 18,789 m^{6}A sites. Comparison with HEK293T showed that similar to mouse, 50.7% of all m^{6}A sites on highly expressed genes were shared between HEK293T and C643 cells (TPM ≥ 20 or higher; Figure 5H and I and Supplementary Figure S5E), an estimate that is stable with increasing expression. We therefore conclude that about half of all m^{6}A modifications are constitutively present in different cell types in human and mouse.

miCLIP2 allows to map m^{6}A sites from low input material

Most current protocols for antibody-based m^{6}A detection start from 5 to 10 μg of poly(A)+ mRNA (37,56). In our standard setup, we use just 1 μg, from which we obtain more than 30 million unique miCLIP2 reads on average with low PCR duplication rates (Supplementary Table S1). However, when working with scarce material such as tissue samples, the amount of extractable RNA is often limited. We therefore tested whether miCLIP2 can be applied with even lower RNA input. To this end, we used poly(A)+ mRNA from mouse heart tissue samples and titrated the amount of input RNA down to 50 ng. The resulting miCLIP2 libraries contained 2–50 million truncation events (Supplementary Table S1).

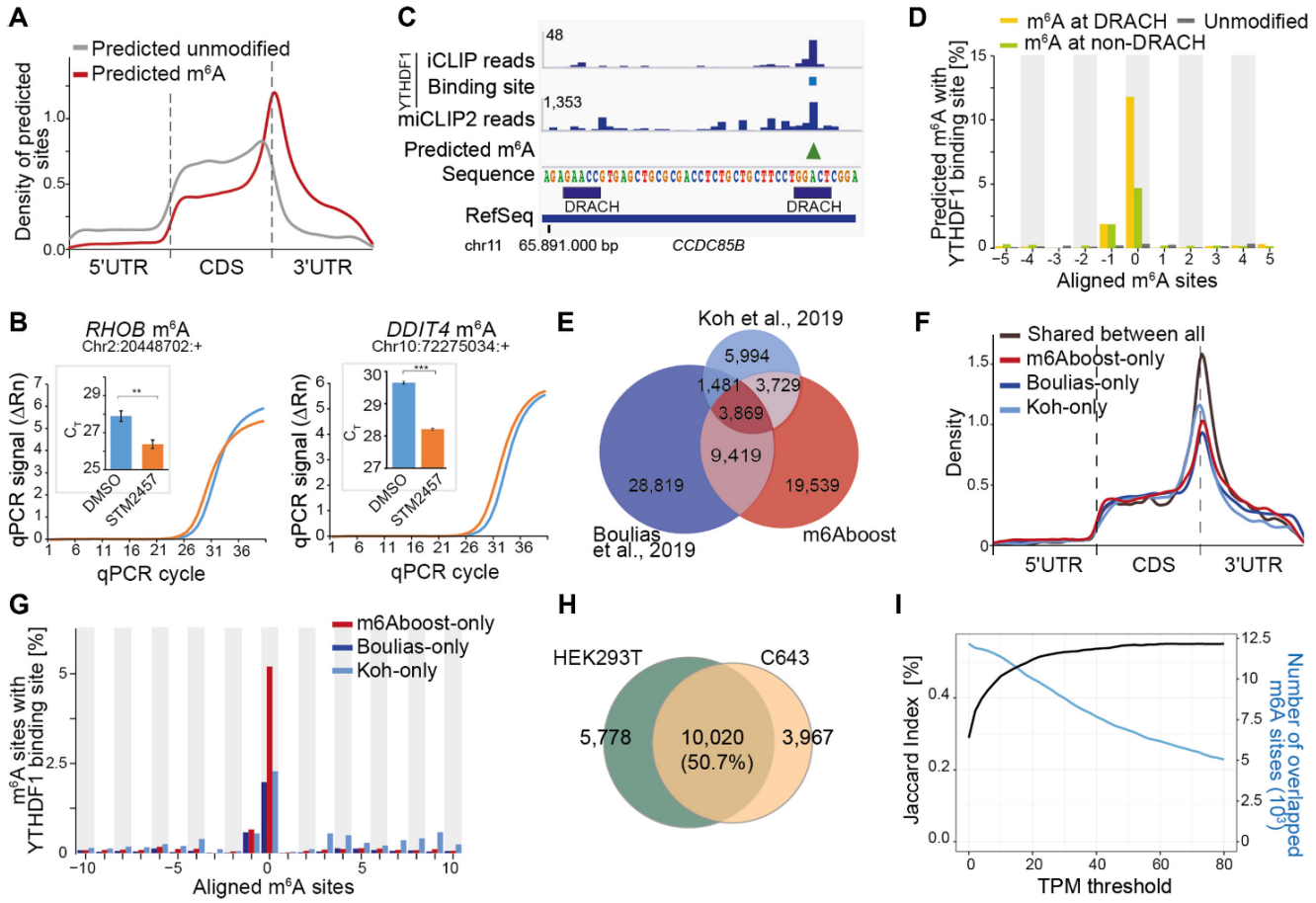


Figure 5. m6Aboost predicts 36,556 m^{6}A sites from HEK293T miCLIP2 data. (**A**) Predicted m^{6}A sites are located around the stop codon. Visualization as in Figure 2F. (**B**) Selected m^{6}A sites were validated by SELECT with HEK293T cells treated with METTL3 inhibitor (STM2457) or DMSO control. Visualization as in Figure 2H. Neighboring unmodified A nucleotides as control are given in Supplementary Figure S5C. *** P value < 0.001, ** $P < 0.01$, two-sided Student's t-test, $n = 3$. (**C**) Predicted m^{6}A sites overlap with YTHDF1 binding sites. Genome browser view of the gene *CCDC85B* shows crosslink events from published YTHDF1 iCLIP data for HEK293T together with miCLIP2 signal (merge of four replicates) and m6Aboost-predicted m^{6}A site (green arrowhead) from our HEK293T miCLIP2 data. (**D**) YTHDF1 precisely binds at the predicted m^{6}A sites. Percentage of m^{6}A sites (position 0) with YTHDF1 binding sites (y-axis) in a 21-nt window are given for predicted m^{6}A sites at DRACH (yellow) and non-DRACH (green), as well as predicted unmodified sites (grey). (**E**) Predicted m^{6}A sites from HEK293T miCLIP2 overlap with published m^{6}A data. Venn diagram shows single-nucleotide overlap with miCLIP and m6ACE-seq data (m^{6}A antibody by Synaptic Systems and Abcam, respectively). Note that m^{6}A sites in Boulias *et al.*, 2019 had been filtered for DRACH motifs. (**F, G**) Analysis of m^{6}A sites that are unique to one of the three datasets compared in (E). (**F**) Unique m^{6}A sites accumulate around stop codons. Visualization as in Figure 2F. (**G**) Unique m^{6}A sites are enriched in YTHDF1 binding sites. Visualization as in (D). (**H**) Most m^{6}A sites are shared between two different human cell lines. Venn diagram shows overlap of predicted m^{6}A sites in expressed genes (TPM $\geq$ 20, n = 3,298) from HEK293T and C643 cells. A Venn diagram without expression filter is shown in Supplementary Figure S5E. (**I**) More m^{6}A sites are shared between two human cell lines in higher expressed genes. TPM threshold representing the gene expression (x-axis) against the Jaccard index (y-axis). Numbers of overlapping m^{6}A sites are shown as comparison (blue).

We found that even with these small amounts of input RNA, the miCLIP2 signals were still reproducible at nucleotide level (Figure 6A and Supplementary Figure S5G). As expected, the sensitivity of miCLIP2 progressively decreased with lower input material. The precision, however, was hardly compromised, since the identified sites were highly overlapping at all concentrations (Figure 6B). Moreover, m^{6}A sites from all RNA input concentrations were consistently enriched at DRACH motifs and nearby stop codons (Figure 6C and D). Together, these results suggest that our approach can be used to identify m^{6}A modifications even from a limited amount of input RNA.

DISCUSSION

Knowledge on the precise location of m^{6}A sites is essential to unravel the molecular effects and biological functions of this universal RNA modification. With the advent of next-generation sequencing, new experimental protocols allow for a systematic mapping of m^{6}A sites, often with single-nucleotide resolution (57). Although alternative methods recently became available (21,22,58,59), the most widely used approaches rely on a set of available antibodies against the modified nucleotide (57). These methods suffer from the broad reactivity of these antibodies, which cross-react with unmodified adenosines or related modifications such as m^{6}Am, thereby generating excessive false positives (17). Moreover, many protocols require high amounts of starting material, or target only a restricted subset of m^{6}A sites that occur for instance in a specific sequence context (21,22,37,56). In this study, we tackle these limitations by combining the optimized miCLIP2 protocol and the machine learning model m6Aboost to reliably map m^{6}A modifications at high resolution and depth. Our

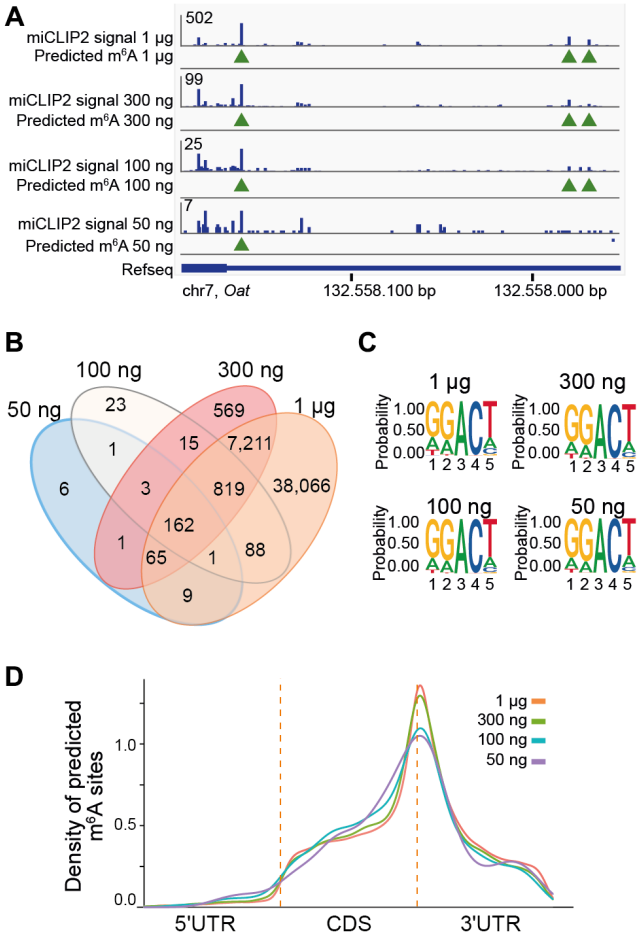


Figure 6. miCLIP2 allows to map m^{6}A sites from low input material (**A**) m6Aboost predicts overlapping m^{6}A sites from miCLIP2 data for different RNA input concentrations. Example genome browser view of the *Oat* gene shows miCLIP2 signals and corresponding m6Aboost predictions (green arrowheads) for 1 µg, 300 ng, 100 ng and 50 ng of input RNA. (**B**) The majority of m^{6}A sites predicted from low-input libraries overlap with the 1 µg input library. Overview of the overlap of predicted m^{6}A sites from different concentrations. (**C**) All predicted sites from different concentrations resemble a DRACH motif. Sequence logo of the predicted m^{6}A sites from miCLIP2 from different RNA input concentrations including the surrounding four nucleotides. (**D**) Predicted m^{6}A sites from miCLIP2 with different RNA input concentrations cluster around the stop codon. Visualization as in Figure 2F.

approach builds on three major experimental and computational innovations that are critical for its efficiency and accuracy.

First, we improved the efficiency of the experimental miCLIP2 protocol by incorporating the recently published iCLIP2 library preparation (24), including separately ligated adapters, two rounds of PCR amplification and a bead-based clean-up strategy. This reduces the processing time to just four days and provides high-complexity datasets without PCR duplicates. With this setup, we now routinely obtain more than 30 million unique miCLIP2 reads from 1 µg input RNA – twenty times less than in the original protocol (17). The moderate duplication rate (Supplementary Table S1) indicates the miCLIP2 libraries in this study were not sequenced to saturation, suggesting that many more m^{6}A sites could still be identified from the same libraries. Moreover, it is possible to obtain reproducible data down to 100 ng and less of input RNA. The reduced input requirement will be particularly useful for studies on nascent RNA or clinical samples and *in vivo* disease models where starting material is limiting.

Second, we tackled the high false positive rate from the m^{6}A-specific antibodies, which is inherent to antibody-based approaches, through the direct comparison with *Mettl3* KO cells. Using a custom-tailored differential methylation analysis strategy, we identified >10 000 Mettl3-dependent m^{6}A sites in the WT mESC miCLIP2 data that constituted the positive set of high-confidence m^{6}A sites for subsequent model training (see below). Of note, we find that m^{6}A modifications occur outside of DRACH motifs (6.3% of all predicted m^{6}A sites) and validate selected m^{6}A sites at non-DRACH motifs using an orthogonal antibody-independent method. Similar motifs were previously reported and recently confirmed in direct RNA sequencing data (Oxford Nanopore Technologies) (17,60). Importantly, since the m^{6}A sites at non-DRACH motifs were included in the m6Aboost model training, similar sites can be readily identified in future miCLIP2 experiments. In addition, we propose that the sequence composition of the high-confidence m^{6}A sites from the differential methylation analysis (Figure 2E), captured for instance in a position weight matrix, could be used to filter other datasets in a more effective way. Moreover, our strategies to account for changes in transcript abundance in order to identify differentially methylated sites will be applicable for other RNA modifications, such as 5-methylcytosine (m^{5}C) in m^{5}C-miCLIP (57,61).

Third, we trained a machine learning model, termed m6Aboost, to accurately extract Mettl3-dependent m^{6}A sites from any miCLIP2 dataset. Several machine learning approaches have been developed to predict m^{6}A sites from the primary RNA sequences (62–64). However, most existing models were trained on data of limited resolution and size, and consequently perform poorly for single-nucleotide predictions. Here, we apply machine learning to predict m^{6}A sites in miCLIP2 data based on a high-confidence positive set of Mettl3-dependent m^{6}A sites. We therefore tackle the inherent problem of false positives that impair most antibody-based m^{6}A detection protocols (57). The resulting m6Aboost model allows to transfer our gained knowledge to other miCLIP datasets without the need for an accompanying *Mettl3* KO, which is not feasible in many biological settings. Because m6Aboost allows for m^{6}A sites at non-DRACH motifs and sorts out false positive miCLIP2 signals, even at DRACH, it outperforms the commonly used DRACH motif filter (37,51,59). The stringency against false positives can be tuned according to the requirements of the user by adjusting the prediction score of m6Aboost.

We note that our model was trained on miCLIP2 data that was obtained with a specific m^{6}A antibody (Synaptic Systems). It is known that certain biochemical features such as the truncation rate at the crosslinked antibody and the distribution of C-to-T transitions vary with each antibody (17,57). We envision that our machine learning model can be retrained on data for other antibodies against m^{6}A and other RNA modifications that can be mapped via miCLIP2, if an accompanying depletion dataset is available.

This includes the related RNA modification m^{6}Am, which is present in the miCLIP2 data due to cross-reactivity of the m^{6}A antibody, and could be recognized and specifically discriminated from m^{6}A after retraining upon depletion of the m^{6}Am-specific methyltransferase PCIF1 (51,65).

In this study, we generated m^{6}A profiles for four human and mouse cell lines that will serve as a resource for future studies. Comparing the methylation profiles revealed that about half of all m^{6}A sites are shared between cell lines in either species. Moreover, we confirm that m^{6}A is mainly deposited around stop codons and within the 3′ UTR (15,16). Interestingly, we also observe an accumulation near the 5′ splice sites of retained introns. Further, our data indicates that m^{6}A can promote intron retention. Previous studies rather described an increase in intron retention events in *Mettl3* KO mESC cells (27), or in null mutants of the Mettl3 orthologue Ime4 in *Drosophila melanogaster* (4,66,67). In contrast, a recent study found that TARBP2-dependent m^{6}A deposition in introns prevents splice factor recruitment and efficient intron excision (68), in line with our observations. This adds a new angle to the controversy surrounding the impact of m^{6}A modifications on alternative splicing. While some studies reported on extensive splicing alterations upon *Mettl3* depletion, others rebutted a strong connection between m^{6}A and splicing (69–72). Consistent with the latter view, we generally observe very few changes in cassette exon splicing in the *Mettl3* KO mESCs. Intron retention, however seemed to be systemically affected, with retained introns being spliced more efficiently throughout the transcriptome of *Mettl3* KO cells.

In essence, the combination of miCLIP2 and m6Aboost allows for a deep and accurate detection of m^{6}A sites. Our study illustrates how artificial intelligence helps to eliminate background signals in order to decode high-throughput data and thereby aids to improve the precise analysis of m^{6}A sites with nucleotide resolution.

DATA AVAILABILITY

The computational code is available in the GitHub repository (https://github.com/ZarnackGroup/Publications/tree/main/Koertel_et_al_2021).

All miCLIP2 and RNA-seq data sets generated in this study were submitted to the Gene Expression Omnibus (GEO) under the SuperSeries accession number GSE163500 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?&acc=GSE163500).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We gratefully acknowledge Tina Lence and Jean-Yves Roignant for help with miCLIP, Lina Worpenberg with initial help with SELECT, Yannic Schumacher for preparing poly(A)+ RNA from RAW 264.7 macrophages and Eric Miska for discussion. We are grateful to Marcel Schulz for advice on machine learning. Support by the IMB Genomics Core Facility and the use of its NextSeq500 (INST 247/870-1FUGG) are gratefully acknowledged. We kindly thank members of the IMB Genomics and Bioinformatics Core Facilities for technical assistance and reagents as well as all members of the König and Zarnack groups for lively discussions. N.K., C.R. and M.P. were supported by the International PhD Programme on Gene Regulation, Epigenetics & Genome Stability, Mainz, Germany.

FUNDING

Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) [SPP1935, KO4566/3-2 to J.K., SPP1935, ZA881/5-2 to K.Z., SPP1935, OS290/6-1 to A.O.L., INST 247/870-1FUGG]. Funding for open access charge: Institutional core funding (to J.K.).

Conflict of interest statement. Oliver Rausch is an employee of STORM Therapeutics Ltd.

REFERENCES

1. Wei,C.M., Gershowitz,A. and Moss,B. (1975) Methylated nucleotides block 5′ terminus of HeLa cell messenger RNA. *Cell*, **4**, 379–386.
2. Yue,Y., Liu,J. and He,C. (2015) RNA N6-methyladenosine methylation in post-transcriptional gene expression regulation. *Genes Dev.*, **29**, 1343–1355.
3. Lin,S. and Gregory,R.I. (2014) Methyltransferases modulate RNA stability in embryonic stem cells. *Nat. Cell Biol.*, **16**, 129–131.
4. Lence,T., Akhtar,J., Bayer,M., Schmid,K., Spindler,L., Ho,C.H., Kreim,N., Andrade-Navarro,M.A., Poeck,B., Helm,M. *et al.* (2016) m(6)A modulates neuronal functions and sex determination in *Drosophila*. *Nature*, **540**, 242–247.
5. Zheng,G., Dahl,J.A., Niu,Y., Fedorcsak,P., Huang,C.M., Li,C.J., Vagbo,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.
6. Li,A., Chen,Y.S., Ping,X.L., Yang,X., Xiao,W., Yang,Y., Sun,H.Y., Zhu,Q., Baidya,P., Wang,X. *et al.* (2017) Cytoplasmic m(6)A reader YTHDF3 promotes mRNA translation. *Cell Res.*, **27**, 444–447.
7. 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.
8. Zaccara,S., Ries,R.J. and Jaffrey,S.R. (2019) Reading, writing and erasing mRNA methylation. *Nat. Rev. Mol. Cell Biol.*, **20**, 608–624.
9. Sledz,P. and Jinek,M. (2016) Structural insights into the molecular mechanism of the m(6)A writer complex. *Elife*, **5**, e18434.
10. 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.
11. Wang,P., Doxtader,K.A. and Nam,Y. (2016) Structural basis for cooperative function of Mettl3 and Mettl14 methyltransferases. *Mol. Cell*, **63**, 306–317.
12. Gu,L., Wang,L., Chen,H., Hong,J., Shen,Z., Dhall,A., Lao,T., Liu,C., Wang,Z., Xu,Y. *et al.* (2020) CG14906 (mettl4) mediates m(6)A methylation of U2 snRNA in Drosophila. *Cell Discov*, **6**, 44.
13. van Tran,N., Ernst,F.G.M., Hawley,B.R., Zorbas,C., Ulryck,N., Hackert,P., Bohnsack,K.E., Bohnsack,M.T., Jaffrey,S.R., Graille,M. *et al.* (2019) The human 18S rRNA m6A methyltransferase METTL5 is stabilized by TRMT112. *Nucleic Acids Res.*, **47**, 7719–7733.
14. Warda,A.S., Kretschmer,J., Hackert,P., Lenz,C., Urlaub,H., Höbartner,C., Sloan,K.E. and Bohnsack,M.T. (2017) Human METTL16 is a N(6)-methyladenosine (m(6)A) methyltransferase that targets pre-mRNAs and various non-coding RNAs. *EMBO Rep.*, **18**, 2004–2014.
15. 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.
16. Meyer,K.D., Saletore,Y., Zumbo,P., Elemento,O., Mason,C.E. and Jaffrey,S.R. (2012) Comprehensive analysis of mRNA methylation reveals enrichment in 3′ UTRs and near stop codons. *Cell*, **149**, 1635–1646.
17. Linder,B., Grozhik,A.V., Olarerin-George,A.O., Meydan,C., Mason,C.E. and Jaffrey,S.R. (2015) Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. *Nat. Methods*, **12**, 767–772.
18. Desrosiers,R., Friderici,K. and Rottman,F. (1974) Identification of methylated nucleosides in messenger RNA from Novikoff hepatoma cells. *Proc. Natl. Acad. Sci. U.S.A.*, **71**, 3971–3975.
19. Schäfer,K.P. (1982) RNA synthesis and processing reactions in a subcellular system from mouse L cells. *Hoppe Seylers Z. Physiol. Chem.*, **363**, 33–43.
20. Mauer,J. and Jaffrey,S.R. (2018) FTO, m(6)Am, and the hypothesis of reversible epitranscriptomic mRNA modifications. *FEBS Lett.*, **592**, 2012–2022.
21. Garcia-Campos,M.A., Edelheit,S., Toth,U., Safra,M., Shachar,R., Viukov,S., Winkler,R., Nir,R., Lasman,L., Brandis,A. *et al.* (2019) Deciphering the "m(6)A Code" via Antibody-Independent Quantitative Profiling. *Cell*, **178**, 731–747.
22. Meyer,K.D. (2019) DART-seq: an antibody-free method for global m(6)A detection. *Nat. Methods*, **16**, 1275–1280.
23. McIntyre,A.B.R., Gokhale,N.S., Cerchietti,L., Jaffrey,S.R., Horner,S.M. and Mason,C.E. (2020) Limits in the detection of m(6)A changes using MeRIP/m(6)A-seq. *Sci. Rep.*, **10**, 6590.
24. Buchbender,A., Mutter,H., Sutandy,F.X.R., Körtel,N., Hänel,H., Busch,A., Ebersberger,S. and König,J. (2020) Improved library preparation with the new iCLIP2 protocol. *Methods*, **178**, 33–48.
25. Leismann,J., Spagnuolo,M., Pradhan,M., Wacheul,L., Vu,M.A., Musheev,M., Mier,P., Andrade-Navarro,M.A., Graille,M., Niehrs,C. *et al.* (2020) The 18S ribosomal RNA m(6) A methyltransferase Mettl5 is required for normal walking behavior in Drosophila. *EMBO Rep.*, **21**, e49443.
26. Schomacher,L., Han,D., Musheev,M.U., Arab,K., Kienhöfer,S., von Seggern,A. and Niehrs,C. (2016) Neil DNA glycosylases promote substrate turnover by Tdg during DNA demethylation. *Nat. Struct. Mol. Biol.*, **23**, 116–124.
27. Geula,S., Moshitch-Moshkovitz,S., Dominissini,D., Mansour,A.A., Kol,N., Salmon-Divon,M., Hershkovitz,V., Peer,E., Mor,N., Manor,Y.S. *et al.* (2015) Stem cells. m6A mRNA methylation facilitates resolution of naive pluripotency toward differentiation. *Science*, **347**, 1002–1006.
28. Yankova,E., Blackaby,W., Albertella,M., Rak,J., De Braekeleer,E., Tsagkogeorga,G., Pilka,E.S., Aspris,D., Leggate,D., Hendrick,A.G. *et al.* (2021) Small molecule inhibition of METTL3 as a strategy against myeloid leukaemia. *Nature*, **593**, 597–601.
29. Xiao,Y., Wang,Y., Tang,Q., Wei,L., Zhang,X. and Jia,G. (2018) An elongation- and ligation-based qPCR amplification method for the radiolabeling-free detection of locus-specific N(6)-methyladenosine modification. *Angew. Chem. Int. Ed. Engl.*, **57**, 15995–16000.
30. Busch,A., Brüggemann,M., Ebersberger,S. and Zarnack,K. (2020) iCLIP data analysis: a complete pipeline from sequencing reads to RBP binding sites. *Methods*, **178**, 49–62.
31. Roehr,J.T., Dieterich,C. and Reinert,K. (2017) Flexbar 3.0 - SIMD and multicore parallelization. *Bioinformatics*, **33**, 2941–2942.
32. Frankish,A., Diekhans,M., Ferreira,A.M., Johnson,R., Jungreis,I., Loveland,J., Mudge,J.M., Sisu,C., Wright,J., Armstrong,J. *et al.* (2019) GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res.*, **47**, D766–D773.
33. 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.
34. Li,H., Handsaker,B., Wysoker,A., Fennell,T., Ruan,J., Homer,N., Marth,G., Abecasis,G., Durbin,R. and Genome Project Data Processing, S. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
35. Smith,T., Heger,A. and Sudbery,I. (2017) UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res.*, **27**, 491–499.
36. Shah,A., Qian,Y., Weyn-Vanhentenryck,S.M. and Zhang,C. (2017) CLIP Tool Kit (CTK): a flexible and robust pipeline to analyze CLIP sequencing data. *Bioinformatics*, **33**, 566–567.
37. Hawley,B.R. and Jaffrey,S.R. (2019) Transcriptome-wide mapping of m(6)A and m(6)Am at single-nucleotide resolution using miCLIP. *Curr Protoc Mol Biol*, **126**, e88.
38. Quinlan,A.R. and Hall,I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.
39. Kent,W.J., Zweig,A.S., Barber,G., Hinrichs,A.S. and Karolchik,D. (2010) BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics*, **26**, 2204–2207.
40. Krakau,S., Richard,H. and Marsico,A. (2017) PureCLIP: capturing target-specific protein-RNA interaction footprints from single-nucleotide CLIP-seq data. *Genome Biol.*, **18**, 240.
41. König,J., Zarnack,K., Luscombe,N.M. and Ule,J. (2012) Protein-RNA interactions: new genomic technologies and perspectives. *Nat. Rev. Genet.*, **13**, 77–83.
42. Chakrabarti,A.M., Haberman,N., Praznik,A., Luscombe,N.M. and Ule,J. (2018) Data science issues in studying protei.n–RNA interactions with CLIP technologies. *Annu. Rev. Biomed. Data Sci.*, **1**, 235–261.
43. Love,M.I., Huber,W. and Anders,S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, **15**, 550.
44. Anders,S., Reyes,A. and Huber,W. (2012) Detecting differential usage of exons from RNA-seq data. *Genome Res.*, **22**, 2008–2017.
45. Ramirez,F., Ryan,D.P., Gruning,B., Bhardwaj,V., Kilpert,F., Richter,A.S., Heyne,S., Dundar,F. and Manke,T. (2016) deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.*, **44**, W160–W165.
46. Anders,S., Pyl,P.T. and Huber,W. (2015) HTSeq – a Python framework to work with high-throughput sequencing data. *Bioinformatics*, **31**, 166–169.
47. Middleton,R., Gao,D., Thomas,A., Singh,B., Au,A., Wong,J.J., Bomane,A., Cosson,B., Eyras,E., Rasko,J.E. *et al.* (2017) IRFinder: assessing the impact of intron retention on mammalian gene expression. *Genome Biol.*, **18**, 51.
48. Audic,S. and Claverie,J.M. (1997) The significance of digital gene expression profiles. *Genome Res.*, **7**, 986–995.
49. König,J., Zarnack,K., Rot,G., Curk,T., Kayikci,M., Zupan,B., Turner,D.J., Luscombe,N.M. and Ule,J. (2010) iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat. Struct. Mol. Biol.*, **17**, 909–915.
50. Bokar,J.A., Shambaugh,M.E., Polayes,D., Matera,A.G. and Rottman,F.M. (1997) Purification and cDNA cloning of the AdoMet-binding subunit of the human mRNA (N6-adenosine)-methyltransferase. *RNA*, **3**, 1233–1247.
51. Boulias,K., Toczydlowska-Socha,D., Hawley,B.R., Liberman,N., Takashima,K., Zaccara,S., Guez,T., Vasseur,J.J., Debart,F., Aravind,L. *et al.* (2019) Identification of the m(6)Am Methyltransferase PCIF1 Reveals the Location and Functions of m(6)Am in the Transcriptome. *Mol. Cell*, **75**, 631–643.
52. Liu,N., Parisien,M., Dai,Q., Zheng,G., He,C. and Pan,T. (2013) Probing N6-methyladenosine RNA modification status at single nucleotide resolution in mRNA and long noncoding RNA. *RNA*, **19**, 1848–1856.
53. Freund,Y. and Schapire,R.E. (1995) In: Vitányi,P. (ed). *Computational Learning Theory. EuroCOLT 1995. Lecture Notes in Computer Science (Lecture Notes in Artificial Intelligence)*. Vol. **904**, Springer, Berlin, Heidelberg.
54. Patil,D.P., Chen,C.K., Pickering,B.F., Chow,A., Jackson,C., Guttman,M. and Jaffrey,S.R. (2016) m(6)A RNA methylation promotes XIST-mediated transcriptional repression. *Nature*, **537**, 369–373.
55. Koh,C.W.Q., Goh,Y.T. and Goh,W.S.S. (2019) Atlas of quantitative single-base-resolution N(6)-methyl-adenine methylomes. *Nat. Commun.*, **10**, 5636.
56. Grozhik,A.V., Linder,B., Olarerin-George,A.O. and Jaffrey,S.R. (2017) Mapping m(6)A at individual-nucleotide resolution using crosslinking and immunoprecipitation (miCLIP). *Methods Mol. Biol.*, **1562**, 55–78.
57. Anreiter,I., Mir,Q., Simpson,J.T., Janga,S.C. and Soller,M. (2020) New twists in detecting mRNA modification dynamics. *Trends Biotechnol.*, **39**, 72–89.

58. Garalde,D.R., Snell,E.A., Jachimowicz,D., Sipos,B., Lloyd,J.H., Bruce,M., Pantic,N., Admassu,T., James,P., Warland,A. *et al.* (2018) Highly parallel direct RNA sequencing on an array of nanopores. *Nat. Methods*, **15**, 201–206.
59. Liu,H., Begik,O., Lucas,M.C., Ramirez,J.M., Mason,C.E., Wiener,D., Schwartz,S., Mattick,J.S., Smith,M.A. and Novoa,E.M. (2019) Accurate detection of m(6)A RNA modifications in native RNA sequences. *Nat. Commun.*, **10**, 4079.
60. Pratanwanich,P.N., Zao,F., Chen,Y., Koh,C.W.Q., Hendra,C., Poon,P., Goh,Y.T., Yap,P.M.L., Yuan,C.J., Chng,W.J. *et al.* (2020) Detection of differential RNA modifications from direct RNA sequencing of human cell lines. bioRxiv doi: https://doi.org/10.1101/2020.06.18.160010, 20 June 2020, preprint: not peer reviewed.
61. Hussain,S., Sajini,A.A., Blanco,S., Dietmann,S., Lombard,P., Sugimoto,Y., Paramor,M., Gleeson,J.G., Odom,D.T., Ule,J. *et al.* (2013) NSun2-mediated cytosine-5 methylation of vault noncoding RNA determines its processing into regulatory small RNAs. *Cell Rep.*, **4**, 255–261.
62. Chen,Z., Zhao,P., Li,F., Wang,Y., Smith,A.I., Webb,G.I., Akutsu,T., Baggag,A., Bensmail,H. and Song,J. (2020) Comprehensive review and assessment of computational methods for predicting RNA post-transcriptional modification sites from RNA sequences. *Brief. Bioinform.*, **21**, 1676–1696.
63. Zhao,Z., Peng,H., Lan,C., Zheng,Y., Fang,L. and Li,J. (2018) Imbalance learning for the prediction of N(6)-Methylation sites in mRNAs. *BMC Genomics*, **19**, 574.
64. Zhou,Y., Zeng,P., Li,Y.H., Zhang,Z. and Cui,Q. (2016) SRAMP: prediction of mammalian N6-methyladenosine (m6A) sites based on sequence-derived features. *Nucleic Acids Res.*, **44**, e91.
65. Sendinc,E., Valle-Garcia,D., Dhall,A., Chen,H., Henriques,T., Navarrete-Perea,J., Sheng,W., Gygi,S.P., Adelman,K. and Shi,Y. (2019) PCIF1 Catalyzes m6Am mRNA Methylation to Regulate Gene Expression. *Mol. Cell*, **75**, 620–630.
66. Haussmann,I.U., Bodi,Z., Sanchez-Moran,E., Mongan,N.P., Archer,N., Fray,R.G. and Soller,M. (2016) m(6)A potentiates Sxl alternative pre-mRNA splicing for robust Drosophila sex determination. *Nature*, **540**, 301–304.
67. Gehring,N.H. and Roignant,J.Y. (2021) Anything but ordinary - emerging splicing mechanisms in eukaryotic gene regulation. *Trends Genet.*, **37**, 355–372.
68. Fish,L., Navickas,A., Culbertson,B., Xu,Y., Nguyen,H.C.B., Zhang,S., Hochman,M., Okimoto,R., Dill,B.D., Molina,H. *et al.* (2019) Nuclear TARBP2 drives oncogenic dysregulation of RNA splicing and decay. *Mol. Cell*, **75**, 967–981.
69. Ke,S., Pandya-Jones,A., Saito,Y., Fak,J.J., Vagbo,C.B., Geula,S., Hanna,J.H., Black,D.L., Darnell,J.E. Jr and Darnell,R.B. (2017) 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.
70. Zhao,X., Yang,Y., Sun,B.F., Shi,Y., Yang,X., Xiao,W., Hao,Y.J., Ping,X.L., Chen,Y.S., Wang,W.J. *et al.* (2014) FTO-dependent demethylation of N6-methyladenosine regulates mRNA splicing and is required for adipogenesis. *Cell Res.*, **24**, 1403–1419.
71. 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.
72. Price,A.M., Hayer,K.E., McIntyre,A.B.R., Gokhale,N.S., Abebe,J.S., Della Fera,A.N., Mason,C.E., Horner,S.M., Wilson,A.C., Depledge,D.P. *et al.* (2020) Direct RNA sequencing reveals m(6)A modifications on adenovirus RNA are necessary for efficient splicing. *Nat. Commun.*, **11**, 6016.

Deep and accurate detection of m^{6}A RNA modifications using miCLIP2 and m6Aboost machine learning

Nadine Körtel[1,#], Cornelia Rücklé[1,#], You Zhou[2,#], Anke Busch[1], Peter Hoch-Kraft[1], FX Reymond Sutandy[1,3], Jacob Haase[4], Mihika Pradhan[1], Michael Musheev[1], Dirk Ostareck[5], Antje Ostareck-Lederer[5], Christoph Dieterich[6,7], Stefan Hüttelmaier[4], Christof Niehrs[1,8], Oliver Rausch[9], Dan Dominissini[10], Julian König[1,*], and Kathi Zarnack[2,*]

[1] Institute of Molecular Biology (IMB), Mainz, 55128, Germany. [2] Buchmann Institute for Molecular Life Sciences (BMLS) & Faculty of Biological Sciences, Goethe University Frankfurt, Frankfurt, 60438, Germany. [3] Institute of Biochemistry II, Goethe University Frankfurt, Frankfurt, 60590, Germany. [4] Institute of Molecular Medicine, Sect. Molecular Cell Biology, Martin Luther University Halle-Wittenberg, Charles Tanford Protein Center, Halle, 06120, Germany. [5] Department of Intensive Care Medicine, University Hospital RWTH Aachen, Aachen, 52074, Germany. [6] Klaus Tschira Institute for Integrative Computational Cardiology, University Hospital Heidelberg, Heidelberg, 69120, Germany. [7] German Centre for Cardiovascular Research (DZHK) - Partner Site Heidelberg/Mannheim, Heidelberg, 69120, Germany. [8] Division of Molecular Embryology, DKFZ-ZMBH Alliance, Heidelberg, Germany. [9] STORM Therapeutics Ltd, Cambridge, CB22 3AT, UK. [10] Cancer Research Center and Wohl Institute for Translational Medicine, Chaim Sheba Medical Center, Tel HaShomer, and Sackler School of Medicine, Tel Aviv University, Tel Aviv, 6997801, Israel.

[#] These authors contributed equally. Order of first authors was determined by lottery.

* Corresponding authors: Kathi Zarnack (kathi.zarnack@bmls.de) & Julian König (j.koenig@imb-mainz.de)

SUPPLEMENTARY MATERIAL

Content:

Supplementary Methods

A. Differential methylation analysis

In order to discriminate true m^{6}A sites from background in the miCLIP2 data, we compared miCLIP2 profiles from wildtype (WT) and *Mettl3* knockout (KO) mouse embryonic stem cells (mESCs) which deplete m^{6}A modifications from mRNAs (**Figure 1B**). However, this analysis was confounded by broad changes in gene expression in response to the *Mettl3* KO, which resulted in 5,372 differentially regulated genes including 3,005 up- and 2,367 down-regulated genes (false discovery rate [FDR] $\leq$ 0.01; **Figure 2A**). Since the miCLIP2 signal, similar like regular iCLIP (1,2), is strongly dependent on the underlying transcript abundance (**Supplementary Figure S2B**, top panel), this means that if not corrected for, differential analysis will erroneously pick up many peaks with reduced signal in downregulated genes. In order to illustrate this, we applied DESeq2 (3) collectively to all peaks in the dataset (approach termed *one-run*) which thus tests each peak independently and omits the underlying transcript level changes. These and all following analyses are based on miCLIP2 truncation reads in peaks identified by peak calling with PureCLIP (4) (see Methods). As expected, the changes in transcript abundance were mirrored in estimated fold changes in miCLIP2 signal of the associated peaks, such that almost all peaks in strongly downregulated genes (log2FC < -3.5) went down in the *Mettl3* KO, whereas peaks in upregulated genes tended to go up (**Supplementary Figure S3A**, left).

In order to overcome this, we tested three different approaches based on (i) separately running DESeq2 on the peaks of each gene (*gene-wise*), (ii) combining peaks for groups of genes with similar abundance change (*bin-based*), and (iii) adopting DEXSeq (*dexseq-run*) (5) instead of DESeq2 (**Supplementary Figure S3A**). The three approaches worked as follows:

Gene-wise approach. Here, we ran an individual DESeq2 analysis for each gene. To this end, we first assigned all peaks to their host gene based on GENCODE gene annotation (release M23) (6). Overlapping genes were resolved by the genes' support level and length, prioritising better support and longer genes. All peaks of a given gene were then used for a collective DESeq2 analysis.

Bin-based approach. A disadvantage of testing for individual genes, as in the *gene-wise* approach and in *dexseq-run* below, is that it relies on a sufficient number of peaks per gene to estimate the required parameters, such as the dispersion, correctly. In order to overcome this, we stratified all genes based on their expression changes upon *Mettl3* KO into equally sized bins. We then collectively tested the signal changes of the peaks from all genes within each bin using DESeq2. Gene expression changes were calculated on the total number of miCLIP2 truncation reads in each gene, acquired by htseq-count with default parameters (7) using GENCODE

gene annotation. Comparison with a parallel DESeq2 analysis on matching RNA-seq data for the same samples showed a high correlation of log2FC values (**Supplementary Figure S3B**), supporting that the summed miCLIP2 signal on the genes allows for a reliable estimation of gene expression changes. We next used the miCLIP2-derived log2FC values to stratify all genes into equal bins (width of Δlog2FC = 0.3, 47 bins; **Supplementary Figure S3C**). Subsequently, we applied DESeq2 collectively to all peaks from all genes of the same bin.

2-factor approach. In this approach, we built a combined DESeq2 model on the miCLIP2 signal in the peaks and the summed miCLIP2 signal on the genes as a proxy for gene expression. To account for gene expression changes, we tested for an interaction term for the peak signals and gene counts (design = ~condition + condition:geneExpression) to the design formula of the DESeq2 model.

Dexseq-run. We adopted DEXSeq (5) (version 1.36.0), an R/Bioconductor package that was developed to test for alternative splicing in RNA-seq data. Originally, DEXSeq2 models RNA-seq read counts in exonic bins, which are grouped by genes, in a generalised linear model to test for differential exon abundance. Additional parameters in the model account for congruent changes across the exons of the same gene to estimate changes in overall gene expression. To run DEXSeq on the miCLIP2 data, we treated each peak as an exonic bin and grouped them by their assigned host gene. DEXSeq was then run with the formula "~sample + peak + condition:peak".

We benchmarked the performance of the different approaches based on the assumptions that true m^6A sites should reside at A and show reduced miCLIP2 signal in the *Mettl3* KO cells. To this end, we compared the differential peaks identified with increasing stringency (FDR) with respect to the total number (yield) and fraction (precision) at A (**Figure 2B**). With more stringent FDR thresholds, the proportion of significantly differential peaks at A continuously increased for all approaches. The best performance was seen for the *bin-based* approach, which yielded the highest number of significant peaks with the highest proportion at A at most FDR thresholds (**Figure 2B**). Besides its accuracy, the *bin-based* approach also showed the best run-time performance by consuming just 1% of CPU time compared to the *gene-wise* or *dexseq-run* approaches.

With the *bin-based* approach at a threshold of FDR < 0.01, we identified total of 14,282 significantly differential peaks, out of which 13,912 peaks (97.4%) went down upon *Mettl3* KO (**Figure 2C**). 11,862 (85.3%) of the decreased peaks resided at A (**Figure 2D**). These were further filtered for the following analyses as described below.

B. AdaBoost machine learning to identify true m^{6}A sites

Compiling the positive and negative sets

We next sought to build a machine learning classifier to distinguish true m^{6}A signals from background in any miCLIP2 datasets, without accompanying *Mettl3* KO. Starting from the reduced peaks at A from the differential methylation analysis (*bin-based* approach, FDR $\leq$ 0.01, log2FC < 0), we calculated the frequency of pentamer motifs at the putative modification site and removed 155 peaks (1.3%) with the most rarely occurring pentamers (present in less than four peaks). This yielded a stringent positive set of 11,707 peaks that were treated as true m^{6}A sites, including 10,966 at DRACH (93.7%) and 741 at non-DRACH motifs (6.3%).

For the corresponding negative set, we required that peaks were not depleted and did not show significant regulation upon the *Mettl3* KO and hence are likely to be part of the nonspecific background in the miCLIP2 data. Based on the *bin-based* approach, we filtered for peaks at A with log2FC $\geq$ 0 and FDR > 0.5. This yielded a negative set of 42,090 peaks, including 1,043 at DRACH (2.5%) and 41,047 at non-DRACH motifs (97.5%). The unbalanced ratio of roughly 4:1 between the negative and the positive set reflected the observed contribution of true m^{6}A sites among all miCLIP2 peaks (**Figure 4A**).

We combined the positive and negative set and then randomly selected 80% of the sites as our training set. The remaining 20% of the sites were kept as an independent test set.

Feature selection

For all peaks in the training and test set, we extracted the following positional and experimental features:

(i) Surrounding nucleotide sequence: We used a 21-nt window around the putatively modified A nucleotide. This is less than what is used by most existing algorithms that predict m^{6}A sites solely on sequence information, including the popular tools SRAMP and DeepM6ASeq (8-10).

(ii) Transcript region: Since m^{6}A sites accumulate in certain transcript regions (11,12) (**Figure 2F**), we included their location within 5' UTR, CDS and 3' UTR as features for the prediction. Transcript annotations were taken from GENCODE (release M23) and filtered for a transcript support level $\leq$ 3 and support level $\leq$ 2. Since the same position can residue in different regions of different isoforms, we separately extracted whether a peak overlapped with at least one 5' UTR, CDS and 3' UTR, and then used this information as three features for the prediction.

(iii) Relative signal strength: In our initial characterisation of the miCLIP2 data, we found that the relative signal strength offered means to enrich for putative m^{6}A sites (**Supplementary Figure S2C**). The relative signal strength is calculated as the number

of truncation events in each peak divided by the mean number of truncation events for all peaks in the same gene (see Methods). For the prediction, values were increased by a pseudo-count of 1 and then $\log_2$-transformed.

(iv) C-to-T transitions: It was previously shown that in the case of readthrough, C-to-T transitions appear at the sites of m^{6}A modifications (13). We therefore included the number of C-to-T transitions 1 nt downstream of each peak as orthogonal feature from the miCLIP2 data. The values were increased by a pseudo-count of 1 and then $\log_2$-transformed.

The importance of the features in the final m6Aboost model is shown in **Figure 3B**.

Machine learning approaches tested for m^{6}A site prediction from miCLIP2 data

We initially tested three different machine learning algorithms. (i) AdaBoost: Adaptive boosting (AdaBoost) formulated by Yoav Freund and Robert Schapire (14) is one of the most widely known boosting algorithms. This tree boosting algorithm combines decision stumps (weak learners) and turns them into a strong learner via applying the boosting method. Moreover, AdaBoost performs exceptionally well for dichotomous tasks. In this project, we used the R package adabag (15) to construct the AdaBoost-based m^{6}A predictor m6Aboost. (ii) Support vector machine (SVM) is a popular machine learning algorithm in bioinformatics which transfers the data to a higher dimension and then finds the hyperplanes to best classify the samples. It was previously used for the prediction of mammalian m^{6}A modification sites (16,17). In this project, we used an interface (e1071) of LIBSVM (18) in the R language to construct the SVM-based predictor. (iii) Random Forest (RF) is a decision tree-based algorithm which shows an excellent performance in supervised learning. It is used in SRAMP (8), one of the earliest and most commonly used m^{6}A predictors. We used an R package of randomForest (19) to build the RF-based m^{6}A predictor.

For all three classifiers, we evaluated the prediction performance on the independent test set using precision-recall (PR) curves (**Figure 3C and Supplementary Figure S4B**). We also calculated sensitivity, specificity, accuracy, precision, F1-score and Matthews correlation coefficient (MCC) (**Supplementary Figure S4A**) as follows:

$$Sensitivity = Recall = \frac{TP}{TP+FN} \quad (1)$$

$$Specificity = \frac{TN}{TN+FP} \quad (2)$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

$$F1 = 2 * \frac{Precision * Recall}{Precision+Recall} \quad (5)$$

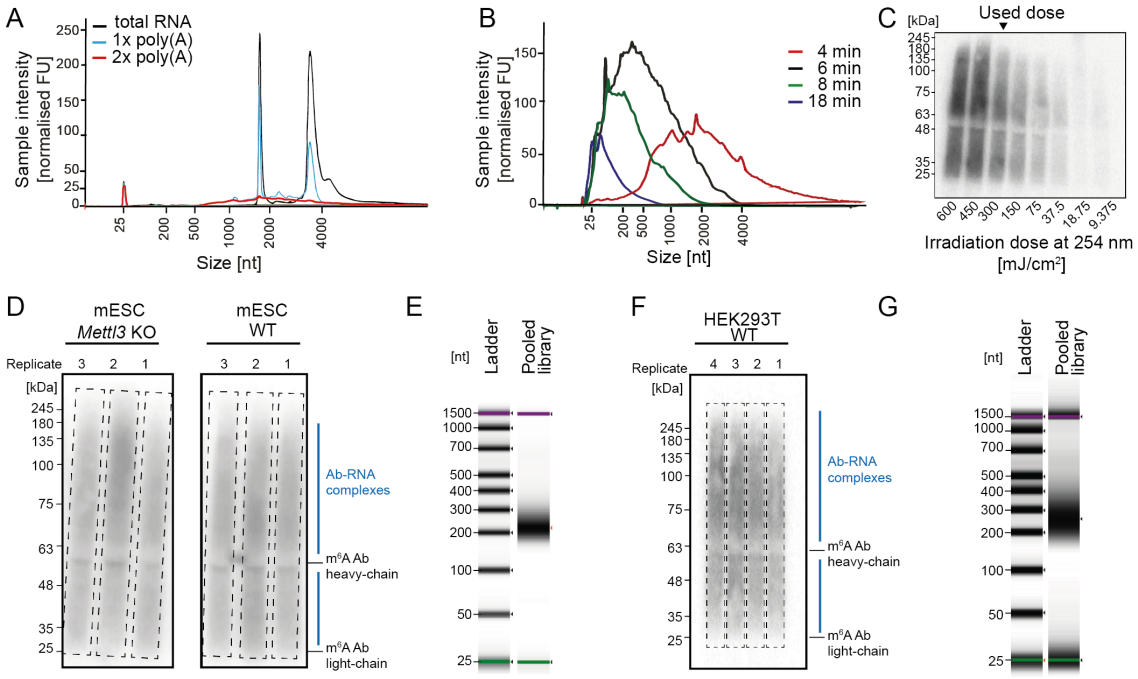
$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (6)$$

where TP, TN, FP and FN represent the counts of true positive, true negative, false positive and false negative predictions, respectively. For m6Aboost, we additionally employed 5-fold cross-validation using the area under the curve (AUC) of PR curves and receiver operating characteristic (ROC) curves to measure the prediction performance (**Supplementary Figure S4C**). We also tested a variant of the AdaBoost model that was trained and tested on a balanced setup (**Supplementary Figure S4A**). For this, we randomly subsampled the negative set to 11,707 sites to match the positive set. Based on the employed measures, we selected the AdaBoost-based predictor m6Aboost (**Supplementary Figure S4D**).

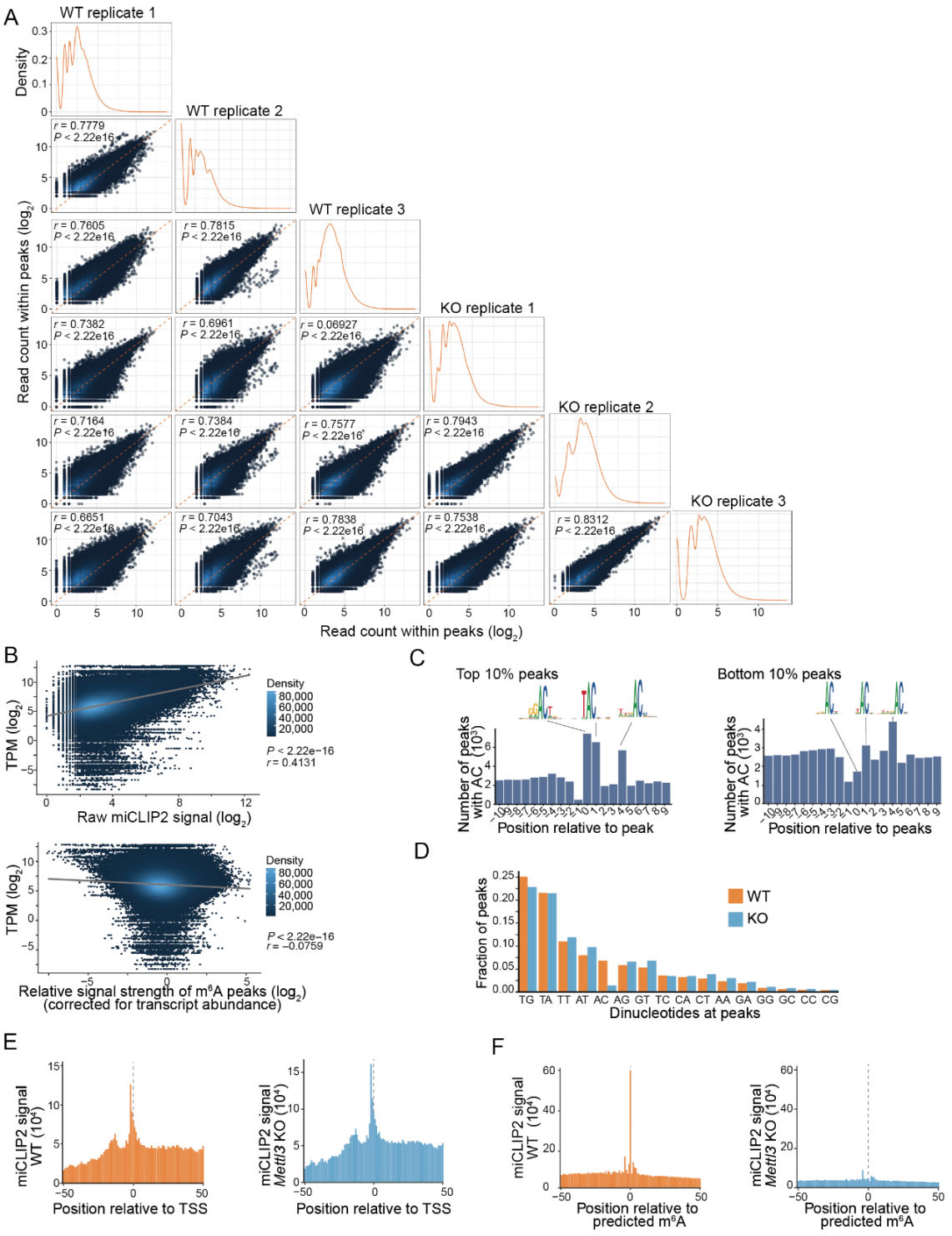
Normalisation of numerical features

Application of the machine learning model to new datasets requires that the data were generated by the same protocol and thus show an independent and identical distribution. The m6Aboost model includes two numerical features from the miCLIP2 data, namely relative signal strength and C-to-T transitions, which could systematically vary between experiments. Since in the training set, both features approximated a Poisson distribution (**Supplementary Figure S4E**), we normalised the values of each features in the input samples by the ratio of the mean for this feature between the input dataset and the training set.

Supplementary Figures

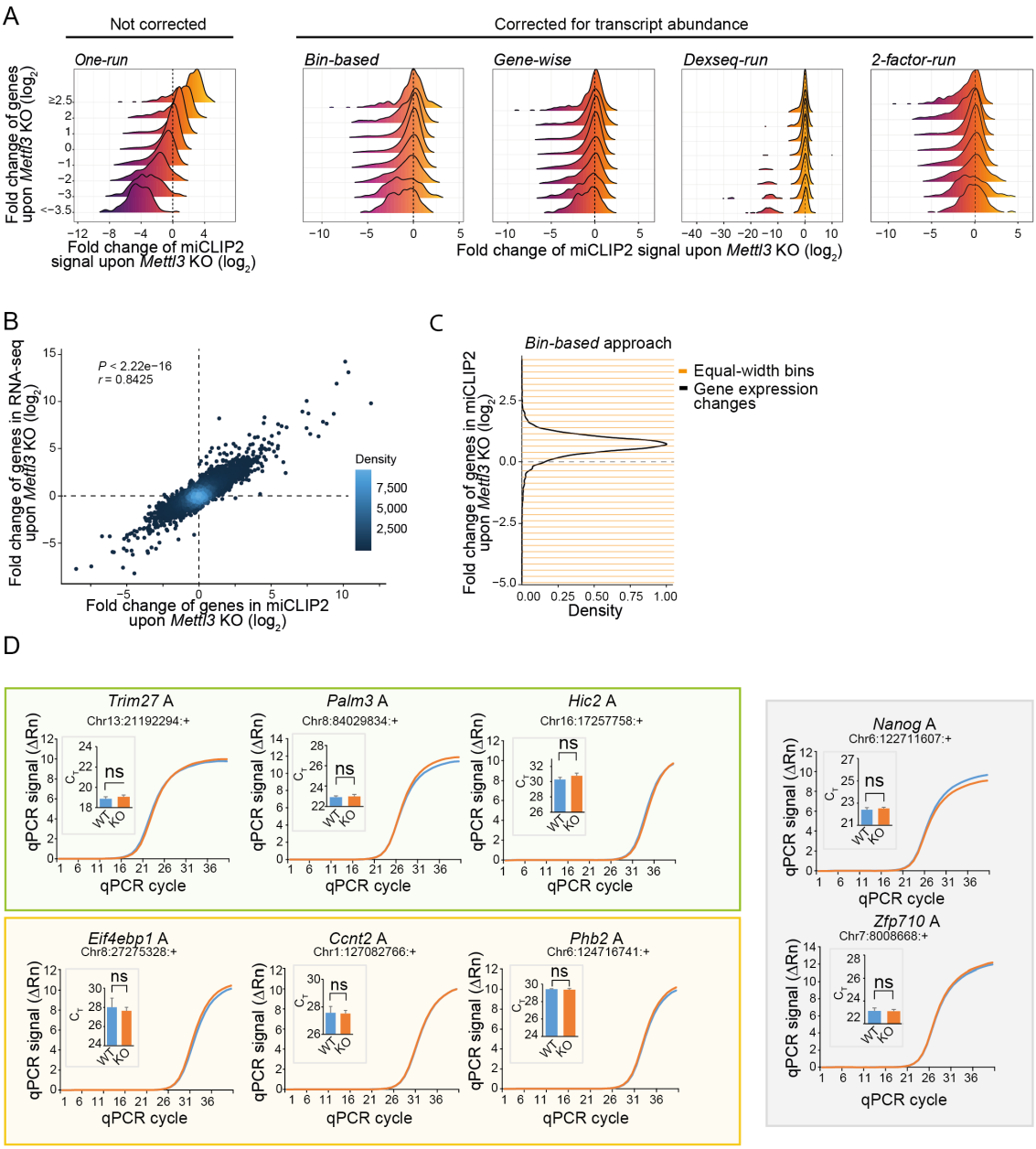


Supplementary Figure S1. miCLIP2 library preparation. ***A-C.*** Optimisation of the miCLIP2 protocol. ***A.*** Ribosomal RNA is fully depleted after two rounds of poly(A) selection. Electropherogram illustrating poly(A) enrichment for 1 µg total RNA of HEK293T cells after one (1x, blue) and two (2x, red) rounds of poly(A) selection. ***B.*** An incubation time of 8 min results in optimal RNA fragmentation. For an optimal RNA fragment spectrum between 50-200 nt, fragmentation periods of 4-18 minutes were compared. ***C.*** Autoradiograph illustrating comparison of RNA crosslinked to the m^{6}A antibody at different irradiation doses in mJ/cm² at 254 nm UV light. The expected molecular weight of the m^{6}A antibody is 50 kDa. ***D-G.*** Visualisation of the miCLIP2 libraries from mouse embryonic stem cells (mESCs) (D,E) and human HEK293T cells (F,G). ***D.*** Autoradiograph illustrating radioactively labelled m^{6}A-antibody-RNA complexes from wild-type (WT) and *Mettl3* knockout (KO) mESCs. The expected molecular weight of the anti-m^{6}A antibody is 50 kDa for the heavy-chain and 25 kDa for the light chain. Excised regions are indicated with dotted lines. ***E.*** Final miCLIP2 pooled library of three biological replicates each for mESC WT and mESC *Mettl3* KO. Note that the final library contained independent samples from an unrelated experiment that were multiplexed for high-throughput sequencing. ***F.*** Autoradiograph illustrating radioactively labelled m^{6}A-antibody-RNA complexes from HEK293T cells. Excised regions are indicated with dotted lines. ***G.*** Final miCLIP2 pooled library of four biological replicates from HEK293T cells. Note that the pooled library also includes material from an additional experiment which is not part of this study.



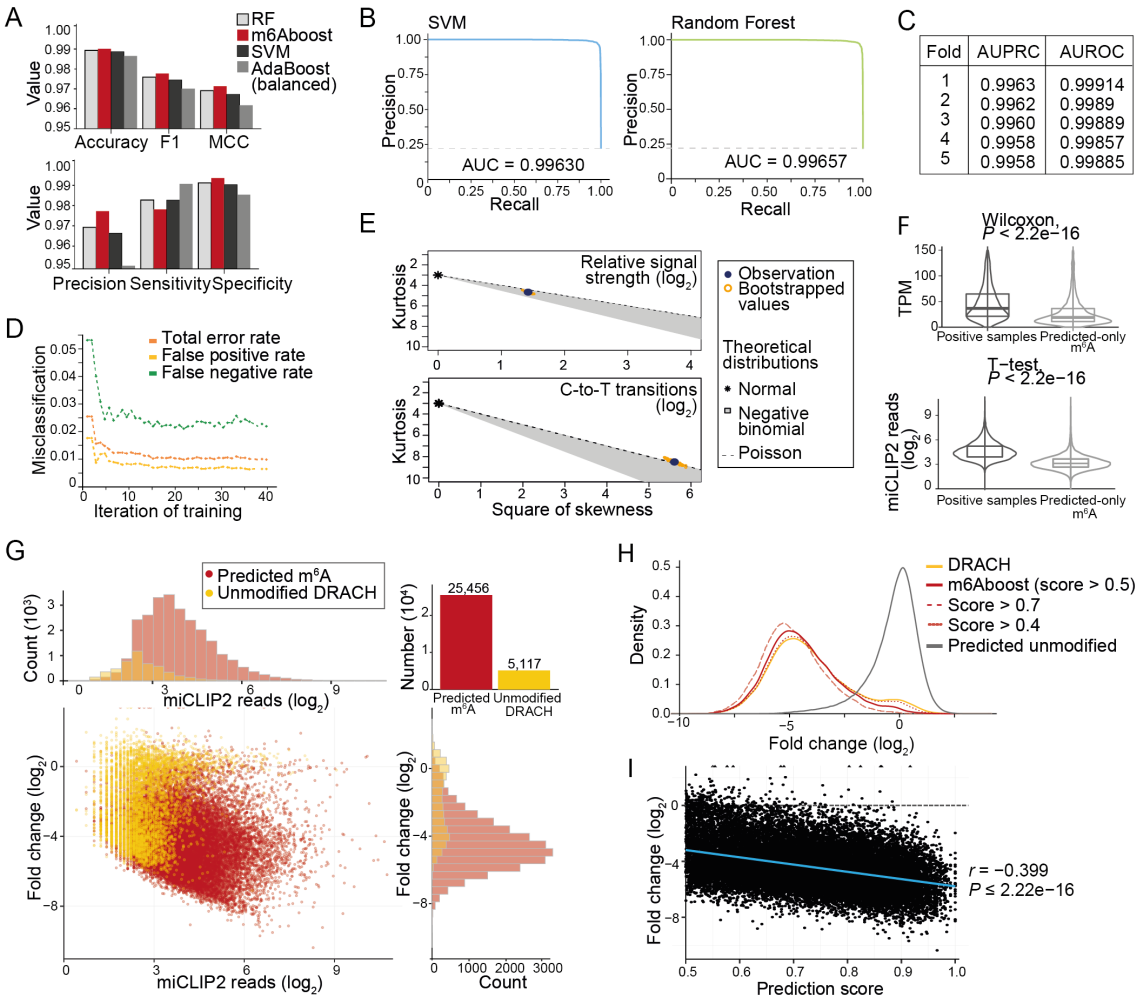
Supplementary Figure S2. Data quality. ***A.*** miCLIP2 libraries are highly reproducible between replicates. Pairwise comparison of truncation read counts within peaks for all replicates from WT and *Mettl3* KO mESCs are shown as an extension of **Figure 1C**. Pearson correlation coefficients (*r*) and associated *P* values are given. ***B.*** Relative signal strength corrects for the effect of gene expression on the miCLIP2 signal. Scatter plots show correlation between miCLIP2 truncation reads (top) or relative signal strength (bottom) and expression of the respective gene (in transcripts per million, TPM, $\log_2$) for all peaks from the WT miCLIP2 data. Colour gradient shows point density. Pearson correlation coefficients (*r*) and associated *P* values are given. ***C.*** Stronger peaks are more often located at AC dinucleotides than weaker peaks. AC

dinucleotide content in a 21-nt window around the 10% strongest peaks (relative signal strength) compared to the 10% weakest peaks from miCLIP2 WT data. ***D.*** Less peaks are located at AC dinucleotides in the *Mettl3* KO miCLIP2 data. Dinucleotide distribution of all peaks from the miCLIP2 WT (orange) and *Mettl3* KO (blue) experiment. ***E.*** Transcript start sites (TSS) accumulate miCLIP2 WT signal which is not reduced upon *Mettl3* KO. miCLIP2 truncation events in a 101-nt window relative to the TSS from WT (upper, orange) and *Mettl3* KO (lower, blue) data. ***F.*** miCLIP2 WT signal enriches around predicted m^6A sites and is depleted in the *Mettl3* KO. miCLIP2 truncation events in a 101-nt window relative to the predicted m^6A sites from WT (upper, orange) and *Mettl3* KO (lower, blue) data.



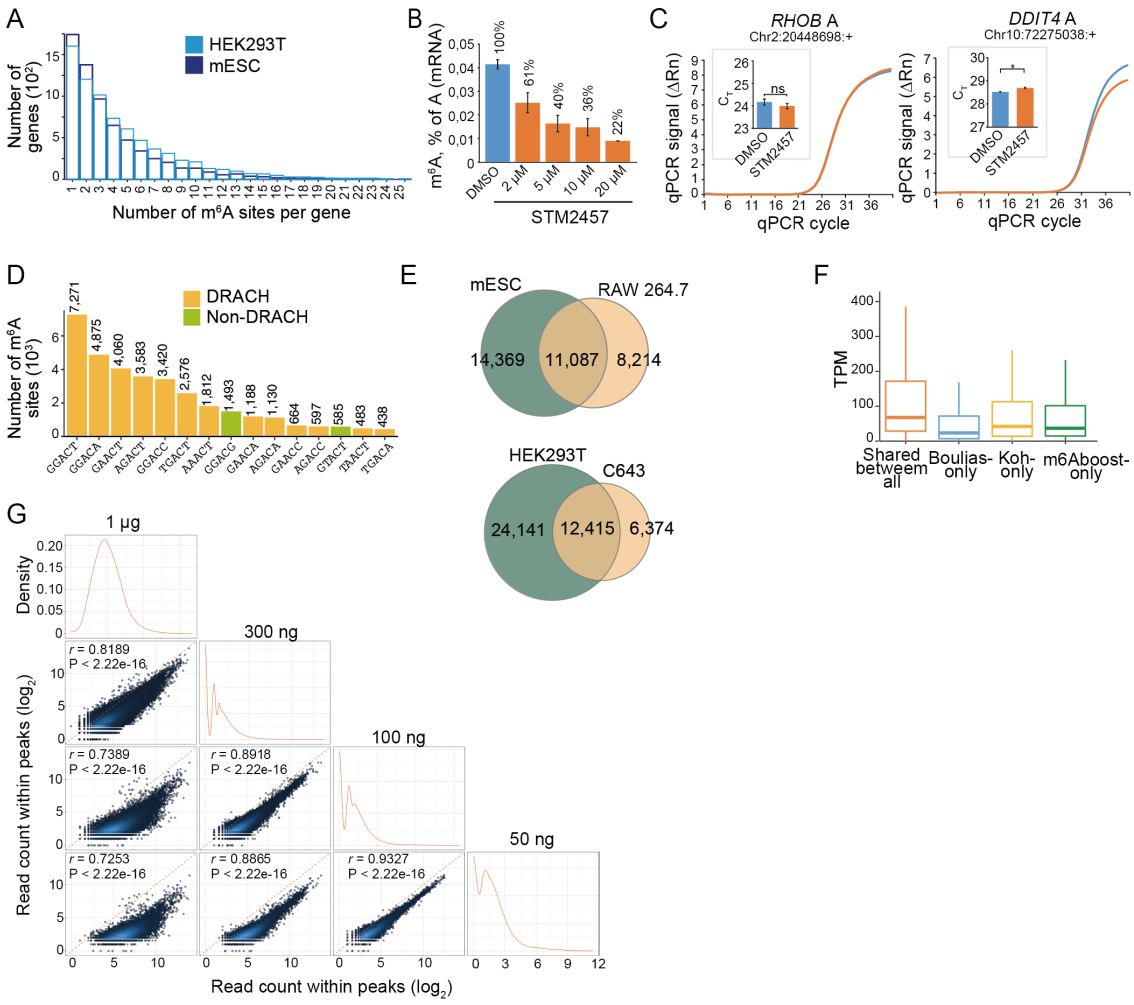
Supplementary Figure S3. Differential methylation analysis and validation. ***A.*** The *bin-based* approach allows to correct for expression changes between WT and *Mettl3* KO. For each tested differential methylation method, the fold change of genes upon *Mettl3* KO (y-axis) and the estimated fold change in the miCLIP2 signal upon the *Mettl3* KO (x-axis) are compared. The *one-run* approach, which does not correct for gene expression changes, is shown for comparison. ***B.*** Changes in gene expression can be estimated from the miCLIP2 data. Comparison of fold changes of genes (n=14,989) calculated from the miCLIP2 data (x-axis) and matching RNA-seq data (y-axis). Pearson correlation coefficients (*r*) and associated *P* value are given. ***C.*** For the *bin-based* approach, genes were stratified based on their expression changes upon *Mettl3* KO (log_2-transformed fold change) into equal-width bins (width of Δlog2FC = 0.3, highlighted in orange). The peaks on all genes within each bin were then collectively tested for differential methylation. ***D.*** Complemental control positions of unmodified A sites neighbouring the validates m^6A sites at non-DRACH (olive background) and DRACH (orange background) motifs (**Figure 2H and I**) as well

as two unmodified sites (grey) at a DRACH motif (**Figure 2J**). Exemplary real-time fluorescence amplification curves (normalised reporter value, ΔRn) and quantification of threshold cycle (C_T) values (technical replicates) for SELECT experiments with mESC WT versus *Mettl3* KO samples are shown for neighbouring unmodified A nucleotides (gene name and genomic coordinates given above). ns, not significant, two-sided Student's *t*-test, n=3.



Supplementary Figure S4. Machine learning to predict m^{6}A sites from miCLIP2 data. ***A.*** Performance measures to compare four different machine learning models. Accuracy, F1-score, Matthews correlation coefficient (MCC), precision, sensitivity and specificity (see **Supplementary Material**) are given for models based on AdaBoost (m6Aboost), support vector machine (SVM) and random forest (RF). An AdaBoost models trained on a balanced test set is shown for comparison. ***B.*** Precision-recall curves for the support vector machine (SVM; left) and random forest (right) models. PR curve for m6Aboost is shown in **Figure 3C**. The corresponding area under the curve (AUC) is given. ***C.*** Results of five-fold cross-validation for m6Aboost. AUC are given for receiver operating characteristic (AUROC) and precision recall curve (AUPRC). ***D.*** Progressive training of m6Aboost. Graph displays misclassification (y-axis) per iteration of training (x-axis). Total error rate (orange), false positive rate (yellow) and false negative rate (green) are shown. ***E.*** Log_2-transformed relative signal strength values (top) and C-to-T transitions per peak follow a Poisson distribution. Cullen and Frey graphs compare square of skewness (x-axis) against kurtosis (y-axis) of the two experimental features (observation and 100 bootstrapped values) against normal, negative binominal and Poisson distribution. ***F.*** m^{6}A sites that are predicted by m6Aboost but not part of the positive set preferentially occur in lowly expressed genes. Top, boxplot shows gene expression values (in transcripts per million, TPM) for 4,292 genes with m^{6}A sites from the positive set against 5,640 genes which exclusively harbour m^{6}A sites that

were only predicted by m6Aboost. *P* value < 2.2e-16, Wilcoxon rank-sum test. Bottom, boxplot compares number of miCLIP2 reads (log_2-transformed) in 11,707 m^6A sites from the positives set against 13,908 m^6A sites that were only predicted by m6Aboost. *P* value < 2.2e-16, Student's *t*-test. ***G.*** Scatter plot and associated histograms show fold change in miCLIP2 signal (log_2-transformed, x-axis) against number of miCLIP2 reads per peak (log_2-transformed, y-axis) for 5,117 peaks at DRACH motifs (yellow) that are predicted to be unmodified by m6Aboost. m^6A sites predicted by m6Aboost (red) are shown for comparison. ***H.*** m6Aboost associates a probability with each predicted m^6A site which can be used to filter more stringently. Density of predicted m6A sites identified with various probability scores against the log_2-transformed fold change between WT and *Mettl3* KO of the corresponding sites. Results of different prediction scores are shown (score [s] = 0.5, red, s=0.7, dashed, s=0.4, dotted) and filtering for a DRACH motif only (orange), as well as the distribution of unmodified sites. ***I.*** The m6Aboost prediction score correlates with the change in miCLIP2 signal upon *Mettl3* KO. Scatterplot showing log_2-transformed fold change in miCLIP2 read counts in WT versus *Mettl3* KO mESC (y-axis) against m6Aboost prediction score (x-axis).



Supplementary Figure S5. m^{6}A sites predicted in miCLIP2 data from human HEK293T and mouse heart tissue. ***A.*** Most genes carry up to three m^{6}A sites. Barplot shows the number of genes (y-axis) with the given number of m^{6}A sites (x-axis) in the miCLIP data from mESC WT (orange) and HEK293T (blue) cells. ***B.*** Titration with increasing concentrations of the METTL3 inhibitor (STM2457)(20) on HEK293T cells shows a gradual reduction of m^{6}A levels on mRNAs. m^{6}A levels measured by liquid chromatography-tandem mass spectrometry (LC-MS/MS) for poly(A)+ RNA from untreated (DMSO) and 2 µM, 5 µM, 10 µM and 20 µM STM2457-treated HEK293T cells. Quantification of m^{6}A as percent of A in RNA. Error bars indicate standard deviation of mean (s.d.m.), n = 3. ***C.*** Complementary control positions of unmodified A sites neighbouring the validates m^{6}A sites from HEK293T cells in the genes *RHOB* and *DDIT4* (**Figure 5B**). Exemplary real-time fluorescence amplification curves (normalised reporter value, ΔRn) and quantification of threshold cycle (C_T) values (technical replicates) for SELECT experiments with untreated (DMSO) versus STM2457-treated samples are shown for neighbouring unmodified A nucleotides (gene name and genomic coordinates given above). ns, not significant, two-sided Student's *t*-test, n=3. ***D.*** The most frequent pentamers at m6Aboost-predicted m^{6}A sites in HEK293T include DRACH and non-DRACH motifs. Same as **Figure 2G** for predicted m^{6}A sites from HEK293T cells. ***E.*** About half of all m^{6}A sites are shared between the two cell lines. Venn diagram as **Figure 4E** (mouse) and **Figure 5H** (human) for m^{6}A sites on all genes (not filtered for expression). ***F.*** m^{6}A sites that

were only found in one out of three datasets from HEK293T cells (this study, (21,22)) are located in lowly expressed genes. Boxplot summarises expression of genes (in transcripts per million, TPM, $\log_2$) harbouring m^6A sites unique to one dataset. ***G.*** miCLIP2 data from different amounts of RNA input are reproducible. Pairwise comparison of truncation read counts within peaks for miCLIP2 libraries from decreasing amounts of RNA input material from mouse heart tissue. Pearson correlation coefficients (*r*) and associated *P* values are given.

Supplementary Tables

Supplementary Table S1. Summary of miCLIP2 experiments. Table includes information on all conducted miCLIP2 experiments including sample names, cell or tissue type, and employed barcodes. It further specifies number of uniquely mapped reads, how many of those were truncation reads (no C-to-T transition) or harboured C-to-T transitions. For each condition, the number of identified PureCLIP peaks and m6Aboost-predicted m^{6}A sites are given. [provided as Excel file]

Supplementary Table S2. Oligonucleotides used in SELECT experiments in **Figure 2H-J and 5B and Supplementary Figure S3D and S5C**. qPCR oligonucleotides for target genes were used for normalisation of input material. Names indicate target and position relative to targeted m^{6}A site. Oligonucleotides were designed complementary anneal to RNA leaving a gap at targeted m^{6}A site or adjacent A site (UP and DOWN probe). Lowercase letters represent adapter sequences for qPCR as described in (23). Uppercase letters represent complementary sequence to target site. Phos indicates 5' phosphorylation.

Name	Sequence [5' - 3']
qPCR_fwd for SELECT	ATGCAGCGACTCAGCCTCTG
qPCR_rev for SELECT	TAGCCAGTACCGTAGTGCGTG
m^{6}A sites from mESC:	
Eif4ebp1_qPCR_fwd	ACTCACCTGTGGCCAAAACA
Eif4ebp1_qPCR_rev	TTGTGACTCTTCACCGCCT
Eif4ebp1_m^{6}A_UP	tagccagtaccgtagtgcgtgGGGAGGGTGTGAGTGAGA G
Eif4ebp1-m^{6}A_DOWN	[Phos]CATTCCCCTGCAGTAGCAGcagaggctgagtcgctgcat
Eif4ebp1_m^{6}A-4 UP	tagccagtaccgtagtgcgtgGGGTGTGAGTGAGAGTCAT
Eif4ebp1_m^{6}A-4 DOWN	[Phos]CCCCTGCAGTAGCAGCTCGcagaggctgagtcgctgcat
Ccnt2_qPCR_fwd	GGGCAACGTCTCAATGTCTCT
Ccnt2_qPCR_rev	AAGCTTTCGAGCCTGCTCTT
Ccnt2_m^{6}A_UP	tagccagtaccgtagtgcgtgGCCCATGCTTGTGCTGCTG
Ccnt2_m^{6}A_DOWN	[Phos]TCTGCATGGGCAGCTAGATcagaggctgagtcgctgcat
Ccnt2_m^{6}A+2_UP	tagccagtaccgtagtgcgtgCGGCCCATGCTTGTGCTGC
Ccnt2_m^{6}A+2_DOWN	[Phos]GTTCTGCATGGGCAGCTAGcagaggctgagtcgctgcat
Phb2_qPCR_fwd	ATCCGTGTTCACCGTGGAAG
Phb2_qPCR_rev	ACCAGGGGATCCTGAAGTGA
Phb2_m^{6}A_UP	tagccagtaccgtagtgcgtgGAGGGCAGATACAGAAAAG
Phb2_m^{6}A_DOWN	[Phos]CCATCACATGATGCCTGGGcagaggctgagtcgctgcat
Phb2_m^{6}A-4_UP	tagccagtaccgtagtgcgtgGCAGATACAGAAAAGTCCA
Phb2_m^{6}A-4_DOWN	[Phos]CACATGATGCCTGGGGCAGcagaggctgagtcgctgcat
Trim27_qPCR_fwd	GGAGGGCTTCAAGGAGCAAA
Trim27_qPCR_rev	AGCTGCTCAAACTCCCAGAC
Trim27_m^{6}A_UP	tagccagtaccgtagtgcgtgACAATGACACTGCCCAGAA
Trim27_m^{6}A_DOWN	[Phos]CCATTCTGGGGGGCTGAGGcagaggctgagtcgctgcat
Trim27_m^{6}A-4_UP	tagccagtaccgtagtgcgtgTGACACTGCCCAGAATCCA
Trim27_m^{6}A-4_DOWN	[Phos]TCTGGGGGGCTGAGGTCACcagaggctgagtcgctgcat

[continued on next page]

Supplementary Table S2. Oligonucleotides used in SELECT experiments (continued from previous page).

Name	Sequence [5' - 3']
Palm3_qPCR_fwd	TACAGCTGTTGCAAAGTGCG
Palm3_qPCR_rev	CACATCAGTCGGGGCGGTA
Palm3_m^{6}A_UP	tagccagtaccgtagtgcgtgTGGGGGACCCTCTCGCTCAG
Palm3_m^{6}A_DOWN	[Phos]ACAGGGCTCAGGCTTACTGcagaggctgagtcgctgcat
Palm3_m6A-8_UP	tagccagtaccgtagtgcgtgCTCTCGCTCAGTACAGGGC
Palm3_m6A-8_DOWN	[Phos]CAGGCTTACTGGCTGCCCCcagaggctgagtcgctgcat
Hic2_qPCR_fwd	CTGGCAGGCACCTGAGGTAA
Hic2_qPCR_rev	AGCTGTAGCAGGAGCTGTTT
Hic2_m^{6}A_UP	tagccagtaccgtagtgcgtgTGCCAGCAGTACCCACTCG
Hic2_m^{6}A_DOWN	[Phos]CCAGGGCCAAAGGGCTTGCcagaggctgagtcgctgcat
Hic2_m^{6}A+3_UP	tagccagtaccgtagtgcgtgCTATGCCAGCAGTACCCAC
Hic2_m^{6}A+3_DOWN	[Phos]CGTCCAGGGCCAAAGGGCTcagaggctgagtcgctgcat
Unmodified DRACH sites:	
Nanog_qPCR_fwd	ACCTGAGCTATAAGCAGGTTAAGAC
Nanog_qPCR_rev	CCCTGGGGATAGCTGCAATG
Nanog_nom^{6}A_UP	tagccagtaccgtagtgcgtgCAGGACTTGAGAGCTTTTG
Nanog_nom^{6}A_DOWN	[Phos]TTGGGACTGGTAGAAGAATcagaggctgagtcgctgcat
Nanog_nom^{6}A+2_UP	tagccagtaccgtagtgcgtgCTCAGGACTTGAGAGCTTT
Nanog_nom^{6}A+2_DOWN	[Phos]GTTTGGGACTGGTAGAAGAacagaggctgagtcgctgcat
Zfp710_qPCR_fwd	TACCGCAGCCAGCTACAAAA
Zfp710_qPCR_rev	CTCCTTCACACCCTTGTGGG
Zfp710_nom^{6}A_UP	tagccagtaccgtagtgcgtgGTTTGCTTCTGCACGAAGG
Zfp710_nom^{6}A_DOWN	[Phos]CTTGAAGCAGATGTGGCACcagaggctgagtcgctgcat
Zfp710_nom6A-3_UP	tagccagtaccgtagtgcgtgTGCTTCTGCACGAAGGTCT
Zfp710_nom6A-3_DOWN	[Phos]GAA GCA GAT GTG GCA CTG Gcagaggctgagtcgctgcat
m^{6}A sites from HEK293T:	
DDIT4_qPCR_fwd	TCGTCGTCCACCTCCTCTTC
DDIT4_qPCR_rev	GGTAAGCCGTGTCTTCCTCC
DDIT4_m6A_UP	tagccagtaccgtagtgcgtgCTTGGGCCAGAGTCGTGAG
DDIT4_m6A_DOWN	[Phos]CCAGGCGCAGCACGAGGGTcagaggctgagtcgctgcat
DDIT4_m6A+4_UP	tagccagtaccgtagtgcgtgGGATCTTGGGCCAGAGTCG
DDIT4_m6A+4_DOWN	[Phos]GAGTCCAGGCGCAGCACGAcagaggctgagtcgctgcat
RHOB_qPCR_fwd	CAGTAAGGACGAGTTCCCCG
RHOB_qPCR_rev	GTCCACCGAGAAGCACATGA
RHOB_m6A_UP	tagccagtaccgtagtgcgtgAAGCTGTGTCCTCCCCAAG
RHOB_m6A_DOWN	[Phos]CAGTTGCAAATGTCTTCCCcagaggctgagtcgctgcat
RHOB_m6A-4_UP	tagccagtaccgtagtgcgtgTGTGTCCTCCCCAAGTCAG
RHOB_m6A-4_DOWN	[Phos]TGCAAATGTCTTCCCCAGGcagaggctgagtcgctgcat
Not validated site:	
ABT1_qPCR_fwd	AAGAAACGGGTAGTGCCAGG
ABT1_qPCR_rev	GTCTCACGAACCGGTCCTC
ABT1_m6A_UP	tagccagtaccgtagtgcgtgAGTCCCTGACAAGGGAAGG
ABT1_m6A_DOWN	CCCTCCATGCTCTCTGAGGcagaggctgagtcgctgcat
ABT1_m6A-4_UP	tagccagtaccgtagtgcgtgCCTGACAAGGGAAGGTCCC
ABT1_m6A-4_DOWN	CCATGCTCTCTGAGGGTGGccagaggctgagtcgctgcat

Supplementary Table S3. Overlap of predicted m^6A sites and SCARLET-validated sites in HEK293T cells taken from (24). m^6A sites with >5% modification are shown in bold. Genomic coordinates are relative to human genome version GRCh38.p13.

Genomic coordinate	Motif	Percent methylation according to SCARLET	Predicted by m6Aboost for HEK293T miCLIP2
Chr11:65500276	**GGACU**	**0.41**	**yes**
Chr11:65500338	**GGACU**	**0.51**	**yes**
Chr11:65500372	**GGACU**	**0.13**	**yes**
Chr11:65500435	AGACU	0.03	no
Chr11:65500445	AGACA	0.02	no
Chr11:65500459	GAACC	0.03	no
Chr11:65500481	**GGACU**	**0.07**	**yes**

Supplementary References

1. Chakrabarti, A.M., Haberman, N., Praznik, A., Luscombe, N.M. and Ule, J. (2018) Data Science Issues in Studying Protein–RNA Interactions with CLIP Technologies. *Annu Rev Biomed Data Sci*, **1**, 235–261.
2. König, J., Zarnack, K., Luscombe, N.M. and Ule, J. (2012) Protein-RNA interactions: new genomic technologies and perspectives. *Nat Rev Genet*, **13**, 77-83.
3. Love, M.I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*, **15**, 550.
4. Krakau, S., Richard, H. and Marsico, A. (2017) PureCLIP: capturing target-specific protein-RNA interaction footprints from single-nucleotide CLIP-seq data. *Genome Biol*, **18**, 240.
5. Anders, S., Reyes, A. and Huber, W. (2012) Detecting differential usage of exons from RNA-seq data. *Genome Res*, **22**, 2008-2017.
6. Frankish, A., Diekhans, M., Ferreira, A.M., Johnson, R., Jungreis, I., Loveland, J., Mudge, J.M., Sisu, C., Wright, J., Armstrong, J. *et al.* (2019) GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res*, **47**, D766-D773.
7. Anders, S., Pyl, P.T. and Huber, W. (2015) HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics*, **31**, 166-169.
8. Zhou, Y., Zeng, P., Li, Y.H., Zhang, Z. and Cui, Q. (2016) SRAMP: prediction of mammalian N6-methyladenosine (m6A) sites based on sequence-derived features. *Nucleic Acids Res*, **44**, e91.
9. Chen, Z., Zhao, P., Li, F., Wang, Y., Smith, A.I., Webb, G.I., Akutsu, T., Baggag, A., Bensmail, H. and Song, J. (2020) Comprehensive review and assessment of computational methods for predicting RNA post-transcriptional modification sites from RNA sequences. *Brief Bioinform*, **21**, 1676-1696.
10. Zhang, Y. and Hamada, M. (2018) DeepM6ASeq: prediction and characterization of m6A-containing sequences using deep learning. *BMC Bioinformatics*, **19**, 524.
11. 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.
12. Meyer, K.D., Saletore, Y., Zumbo, P., Elemento, O., Mason, C.E. and Jaffrey, S.R. (2012) Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell*, **149**, 1635-1646.
13. Linder, B., Grozhik, A.V., Olarerin-George, A.O., Meydan, C., Mason, C.E. and Jaffrey, S.R. (2015) Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. *Nat Methods*, **12**, 767-772.
14. Freund, Y. and Schapire, R.E. (1995) In Vitányi, P. (ed.), *Computational Learning Theory. EuroCOLT 1995. Lecture Notes in Computer Science (Lecture Notes in Artificial Intelligence)*. Springer, Berlin, Heidelberg, Vol. 904.
15. Alfaro, E., Gamey, M. and Garcia, N. (2013) adabag: An R Package for Classification with Boosting and Bagging. *J Stat Softw*, **54**, 1-35.
16. Chen, W., Xing, P. and Zou, Q. (2017) Detecting N(6)-methyladenosine sites from RNA transcriptomes using ensemble Support Vector Machines. *Sci Rep*, 7, 40242.

17. Xiang, S., Liu, K., Yan, Z., Zhang, Y. and Sun, Z. (2016) RNAMethPre: A Web Server for the Prediction and Query of mRNA m6A Sites. *PLoS One*, **11**, e0162707.
18. Chang, C.-C. and Lin, C.-J. (2011) LIBSVM: A library for support vector machines. *ACM Trans Intell Syst Technol*, 27.
19. Liaw, A. and Wiener, M. (2002) Classification and Regression by randomForest. *R News*, **2**, 18-22.
20. Yankova, E., Blackaby, W., Albertella, M., Rak, J., De Braekeleer, E., Tsagkogeorga, G., Pilka, E.S., Aspris, D., Leggate, D., Hendrick, A.G. *et al.* (2021) Small molecule inhibition of METTL3 as a strategy against myeloid leukaemia. *Nature*, **in press**.
21. Boulias, K., Toczydlowska-Socha, D., Hawley, B.R., Liberman, N., Takashima, K., Zaccara, S., Guez, T., Vasseur, J.J., Debart, F., Aravind, L. *et al.* (2019) Identification of the m(6)Am Methyltransferase PCIF1 Reveals the Location and Functions of m(6)Am in the Transcriptome. *Mol Cell*, **75**, 631-643 e638.
22. Koh, C.W.Q., Goh, Y.T. and Goh, W.S.S. (2019) Atlas of quantitative single-base-resolution N(6)-methyl-adenine methylomes. *Nat Commun*, **10**, 5636.
23. Xiao, Y., Wang, Y., Tang, Q., Wei, L., Zhang, X. and Jia, G. (2018) An Elongation- and Ligation-Based qPCR Amplification Method for the Radiolabeling-Free Detection of Locus-Specific N(6)-Methyladenosine Modification. *Angew Chem Int Ed Engl*, **57**, 15995-16000.
24. Liu, N., Parisien, M., Dai, Q., Zheng, G., He, C. and Pan, T. (2013) Probing N6-methyladenosine RNA modification status at single nucleotide resolution in mRNA and long noncoding RNA. *RNA*, **19**, 1848-1856.

2.3 RNA stability controlled by m^{6}A methylation mediates X-to-autosome dosage compensation in mammals

The following manuscript has been revised and "accepted in principal" by *Nature Structural & Molecular Biology*.

2.3.1 Zusammenfassung

Männliche Säugetiere besitzen ein X- und ein Y-Chromosom (XY), während weibliche zwei X-Chromosomen (XX) besitzen. Da X-chromosomale Gene ausschließlich von dem einzigen vorhandenem männlichen X-Chromosom exprimiert werden, führte das genetische Ungleichgewicht zwischen den Geschlechtern zur Inaktivierung eines X-Chromosoms bei den Weibchen. Das daraus resultierende Ungleichgewicht zwischen dem verbleibenden aktiven X-Chromosom und zwei aktiven Kopien der Autosomen muss ausgeglichen werden. Daher wurde ein Dosierungsausgleich für X-chromosomale Gene vorgeschlagen. Der Mechanismus des Dosierungsausgleichs zwischen X-Chromosom und Autosomen wird jedoch noch immer kontrovers diskutiert. Mithilfe von miCLIP2 konnten wir nachweisen, dass die m^{6}A-Modifikation in X-chromosomalen Transkripten geringer ist. Darüber hinaus zeigen wir, dass X-chromosomale Transkripte im Vergleich zu ihren autosomalen Pendants stabiler sind. Unter Verwendung eines METTL3-Inhibitors, der zu einer akuten m^{6}A-Abreicherung führt, wurden autosomale Transkripte selektiv stabilisiert, was zu einer gestörten Dosierungskompensation führte. Wir vermuten, dass eine höhere RNA-Stabilität von X-chromosomalen Transkripten durch reduzierte m^{6}A-Spiegel erreicht wird. Daher zeigen wir einen epitranskriptomischen Mechanismus für den Dosiskompensationsausgleich.

2.3.2 Abstract

Male mammals possess one X and one Y chromosome (XY), while females possess two X chromosomes (XX). Since X-chromosomal genes are solely expressed from the single male X chromosome, genetic imbalance between the sexes resulted in X-chromosome inactivation of one X chromosome in females. The resulting imbalance between the remaining active X chromosome and two active copies of autosomes requires compensation. Therefore, dosage compensation for X-chromosomal genes has been proposed. However, the X-to-autosome dosage compensation mechanism is still under active debate. Using miCLIP2, we demonstrate lower m^{6}A modification levels in X-chromosomal transcripts. Furthermore, we demonstrate that X-chromosomal transcripts are more stable compared to their autosomal counterparts. Using a METTL3 inhibitor leading to acute m^{6}A depletion, autosomal transcripts were selectively stabilized, leading to perturbed dosage compensation. We suggest that higher RNA stability of X-chromosomal transcripts is achieved by reduced m^{6}A levels. Therefore, we demonstrate an epitranscriptomic mechanism for dosage compensation

2.3.3 Statement of contribution

In previously generated miCLIP2 data from mESC cells, we have observed lower levels of m^6A modifications on X-chromosomal transcripts. To test whether higher half-lives of X-chromosomal transcripts derive from lower m^6A levels, we tested RNA half-lives upon m^6A depletion. To this end, I used METTL3 inhibitor STM2457 and performed SLAM-seq experiments in mES cells. In order to test whether expression of X-chromosomal and autosomal transcripts change differently upon m^6A loss, I performed STM2457 treatment for different mouse and human cells lines and subjected the samples to RNA sequencing. Following the observation that lower m^6A-levels on X-chromosomal transcripts are driven by lower GGACH motif, which could indicate a hardcoded mechanism, I performed RNA sequencing and miCLIP2 experiments in female mESC. Since female mESC are prone to lose one X chromosome during culturing, I picked single clones. I performed DNA sequencing experiments to genotype the clones and subsequently performed RNAseq and miCLIP2 for XX and X0 colonies. I prepared the respective figures, wrote respective parts of the manuscript and reviewed the manuscript.

Supervisor confirmation: _____

RNA stability controlled by m^{6}A methylation contributes to X-to-autosome dosage compensation in mammals

Cornelia Rücklé[1,*], Nadine Körtel[1,*], M. Felicia Basilicata[1,3], Anke Busch[1], You Zhou[2], Peter Hoch-Kraft[1], Kerstin Tretow[1], Fridolin Kielisch[1], Marco Bertin[3], Mihika Pradhan[1], Michael Musheev[1], Susann Schweiger[1,3], Christof Niehrs[1,4], Oliver Rausch[5], Kathi Zarnack[2], Claudia Keller Valsecchi[1], Julian König[1,#]

[1]Institute of Molecular Biology (IMB), Mainz 55128, Germany

[2]Buchmann Institute for Molecular Life Sciences (BMLS) & Institute of Molecular Biosciences, Goethe University Frankfurt, Frankfurt 60438, Germany

[3]Institute of Human Genetics, University Medical Center of the Johannes Gutenberg University Mainz, 55131, Germany

[4]Division of Molecular Embryology, DKFZ-ZMBH Alliance, Heidelberg 69120, Germany

[5]STORM Therapeutics Ltd, Cambridge CB22 3AT, UK.

* These authors contributed equally. The order of the first authors was determined by lottery.

[#] Corresponding author: j.koenig@imb-mainz.de (J.K.)

Abstract

In mammals, X-chromosomal genes are expressed from a single copy since males (XY) possess a single X chromosome, while females (XX) undergo X inactivation. To compensate for this reduction in dosage compared to two active copies of autosomes, it has been proposed that genes from the active X chromosome exhibit dosage compensation. However, the existence and mechanism of X-to-autosome dosage compensation are still under debate. Here, we show that X-chromosomal transcripts are reduced in m^6A modifications and more stable compared to their autosomal counterparts. Acute depletion of m^6A selectively stabilises autosomal transcripts, resulting in perturbed dosage compensation in mouse embryonic stem cells. We propose that higher stability of X-chromosomal transcripts is directed by lower levels of m^6A, indicating that mammalian dosage compensation is partly regulated by epitranscriptomic RNA modifications.

Main text

Sex chromosomes evolved from a pair of autosomes. During this process, the chromosome only present in the heterogametic sex (i.e., the Y chromosome in male mammals) acquires mutations, undergoes recurrent chromosomal rearrangements and eventually becomes highly degenerated, gene-poor and heterochromatic[1]. Consequently, the X chromosome and most of its genes are present in a single copy in males, whereas two X chromosomes are present in females. To equalise expression between sexes in eutherian female mammals, one randomly chosen X chromosome is inactivated (X_i) early in development at around the implantation stage. Therefore, XY males and X_iX_a females exhibit an imbalance of gene dosage between sex chromosomes and autosomes, which are present in one and two active copies, respectively[2]. To restore the balance between X chromosomes and autosomes, Susumu Ohno hypothesised that the expression of X-chromosomal genes is upregulated by two-fold[3]. Indeed, there are several mechanisms conceivable for how this could be achieved. For instance, previous studies proposed that higher RNA polymerase II occupancy as well as more activating epigenetic marks and gains in chromatin accessibility on the X chromosome play a role in dosage compensation[4-7]. Additionally, higher RNA stability of X-chromosomal transcripts was observed[6,8]. There is evidence that nonsense-mediated mRNA decay (NMD) targets are enriched for autosomal transcripts[9], which could partially explain the higher RNA stability of X-chromosomal transcripts. Another recent study proposed that dosage compensation could also be mediated by elevated translation of X-chromosomal transcripts[10]. Eventually, dosage compensation may only be required for a certain subset of transcripts which are dosage-sensitive, for instance, if stoichiometry with transcripts from other chromosomes is necessary for proper complex formation[11]. Some dosage-sensitive transcripts may also be protected from the degeneration process occurring on the Y chromosome and thus, be retained in two copies[12]. However, Ohno's hypothesis is still under investigation and both transcriptional and post-transcriptional mechanisms could play a role or act together[10,13-17]. If the latter would be the case, this creates the conundrum of how the chromosomal origin of a transcript is "remembered" in downstream steps of gene expression that occur at the RNA level.

RNA modifications are increasingly recognised for their role in post-transcriptional gene regulation. By their "epitranscriptomic" nature, they have the potential to bridge DNA context to mRNA fate. N6-methyladenosine (m^6A) is the most abundant internal mRNA modification, with estimates ranging from one up to thirteen modifications present per transcript[18-21]. Conserved adenine methyltransferases, such as Mettl3, co-transcriptionally modify nascent mRNAs in the nucleus. The majority of m^6A sites occur within a DRACH motif (i.e., [G/A/U][G>A]m6AC[U>A>C]) with GGACH as the predominantly methylated sequence[22-24]. m^6A-methylated transcripts recruit different reader proteins. Most prominently, Ythdf proteins (Ythdf1, 2 and 3) reduce the stability of m^6A-modified transcripts in the cytoplasm by promoting their degradation[25-27]. Hence, m^6A modifications affect mRNA fate in the cytoplasm upon their deposition in the nucleus.

Here, we show that m^6A RNA modifications play a key role in X-to-autosome dosage compensation. We find that the m^6A content is reduced in transcripts from the X

chromosome, leading to more stable transcripts and longer half-lives. This is crucial to equalise the imbalance in gene dosage between autosomes and the X chromosome.

Results

Autosomal transcripts are stabilised by m^{6}A depletion

One of the most prominent functions of m^{6}A lies in regulating mRNA levels via promoting RNA decay[25]. Since it has been proposed that X-chromosomal transcripts are more stable than autosomal transcripts[6,8], we hypothesised that m^{6}A-mediated RNA stability may be involved in X-to-autosome dosage compensation. To investigate this, we first confirmed the chromosomal differences in RNA stability in published mRNA half-lives from mouse embryonic stem cells (mESC), measured by thiol(SH)-linked alkylation for the metabolic sequencing of RNA (SLAM-seq)[28]. Indeed, transcripts originating from the X chromosome had significantly longer half-lives than autosomal transcripts (**Extended Data Fig. 1A**).

To investigate the direct impact of m^{6}A depletion, we employed the small molecule inhibitor STM2457 that specifically targets the major mRNA m^{6}A methyltransferase Mettl3[29]. We corroborated in a time course experiment that the m^{6}A levels showed a strong reduction already after 3 hours (h) and reached the low point after 6 h of inhibitor treatment (**Extended Data Fig. 1B**). Compared to a *Mettl3* knock-out (KO), this acute m^{6}A depletion enabled us to investigate the immediate response to m^{6}A depletion, while minimising secondary effects[30]. Expression analysis of marker genes[31] and qPCR validations showed that the pluripotent state of the mESC remained unimpaired throughout the treatment (**Extended Data Fig. 1C,D**).

To determine the effect of m^{6}A depletion on mRNA half-lives, we performed SLAM-seq in m^{6}A-depleted and control conditions (6 h STM2457-treated or DMSO-treated as control, **Fig. 1A and Extended Data Fig. 2A,B**). We achieved a stable s^{4}U incorporation rate of 1.36% after 24 h of labelling, which gradually decreased upon washout (**Extended Data Fig. 2C**). By fitting the SLAM-seq data using an exponential decay model and filtering for expression and a sufficient goodness-of-fit (see Methods)[28], we obtained half-life estimates for 7,310 transcripts (**Table S1, Fig. 1B,C and Extended Data Fig. 2D,E**). The estimated half-lives in the control condition correlated well with previously published mRNA half-lives[28] (**Extended Data Fig. 2F**).

Consistent with the role of m^{6}A in destabilising transcripts[25,32], the median half-life of mRNAs significantly increased upon acute m^{6}A depletion (**Fig. 1B,C**). Using high-confidence m^{6}A sites, which we had previously mapped in the same cell line using miCLIP2 (m^{6}A individual-nucleotide resolution UV crosslinking and immunoprecipitation) and m6Aboost[33], we confirmed that in control conditions, transcripts with m^{6}A sites showed significantly shorter half-lives than unmethylated transcripts[28] (**Fig. 1D**). Furthermore, the transcripts with m^{6}A sites were significantly stabilised upon acute m^{6}A depletion (8% median increase), whereas unmethylated transcripts were largely unaffected (0.3% median decrease, **Fig. 1E**).

Having ensured the high quality of our dataset, we turned to chromosomal differences in mRNA stability. X-chromosomal transcripts had significantly longer half-lives than autosomal transcripts under control conditions (**Extended Data Fig. 2G**, left). Importantly, the half-lives of autosomal transcripts significantly increased after acute m^{6}A depletion (5% median increase), whereas the stability of X-chromosomal transcripts remained unchanged (0.2% median decrease, **Fig. 1F**). Transcripts on all autosomes responded similarly, while the X

chromosome was the only chromosome that appeared excluded from this increase (**Fig. 1G and Extended Data Fig. 2G**). These results indicated that m^{6}A-mediated RNA stability could play a direct role in X-to-autosome dosage compensation in mESC. To further support this, we reanalysed published mRNA half-lives for wild-type (WT) and *Mettl3* KO mESC[34] and observed the same difference in RNA stabilisation between X-chromosomal and autosomal transcripts (**Fig. 1H**). The difference between m^{6}A-depleted and *Mettl3* KO condition may result from chromosomal differences or from compensatory mechanisms after KO generation, such as induced expression of alternatively spliced *Mettl3* isoforms[30]. Collectively, the intersection between our experiments and published data conclusively shows that m^{6}A modifications destabilise autosomal transcripts, while X-chromosomal transcripts are largely excluded from such regulation.

X-chromosomal transcripts are less affected by m^{6}A depletion

To test whether the chromosomal differences in RNA stability contribute to balancing expression levels between X chromosome and autosomes, we performed RNA-seq experiments to measure the transcript expression levels after m^{6}A depletion (24 h STM2457, **Extended Data Fig. 3A and Table S2**). The degree of upregulation correlated with the number of m^{6}A sites, such that the most heavily methylated transcripts showed the strongest upregulation (**Extended Data Fig. 3C**). Strikingly, we observed a marked difference in the response to m^{6}A depletion between X-chromosomal and autosomal transcripts. On autosomes, we found more upregulated genes relative to the X chromosome, whereas the X-chromosomal transcripts showed by far the lowest median fold change of all chromosomes (**Fig. 2A**). Between autosomes, observed changes were very similar, suggesting that transcripts on all autosomes were equally affected by acute m^{6}A depletion.

To directly assess the balance between X-chromosomal and autosomal transcript levels, we determined the X-chromosomal-to-autosomal (X:A) expression ratio[5,35]. In DMSO-treated cells, the median X:A ratio approximated 1 when excluding silent or lowly expressed genes, illustrating that X-to-autosome dosage compensation is functional in male mESC (**Extended Data Fig. 3D,E**). Importantly, the X:A ratio significantly went down in the m^{6}A-depleted conditions, indicating that m^{6}A depletion leads to an imbalance in X-to-autosome dosage compensation (**Fig. 2B**). We note that the X:A ratio does not reach 0.5, suggesting that m^{6}A acts in addition to other regulatory mechanisms in X-to-autosome dosage compensation.

The differential effects of m^{6}A depletion on X-chromosomal and autosomal genes was further supported in a time course RNA-seq experiment with 3 to 12 h STM2457 treatment (**Extended Data Fig. 1B,C and Table S2**). Of note, autosomal transcripts showed a distinct response from X-chromosomal transcripts already after 6 h of m^{6}A depletion, which persisted throughout 9 h and 12 h treatment (**Fig. 2C and Extended Data Fig. 4A,B**). This was validated by qPCR for five autosomal and five X-chromosomal transcripts after 9 h of m^{6}A depletion (**Extended Data Fig. 4C**). The clear separation of X-chromosomal and autosomal transcripts at around 6 h was in line with the observed mRNA stability changes after the same treatment duration (**Fig. 1G**) and supported a direct effect of m^{6}A in transcript destabilisation.

Next, we investigated whether m^{6}A similarly regulates X-chromosomal transcripts in humans. To this end, we performed RNA-seq of primary human fibroblasts (male) after 9 h of m^{6}A depletion (**Fig. 2D and Extended Data Fig. 5A**). As in mESC, we observed a clear separation of X chromosome and autosomes, such that X-chromosomal transcripts displayed significantly lower changes (**Fig. 2D**). This was further corroborated by RNA-seq data upon m^{6}A depletion in human HEK293T (female), C643 (male) and RPE1 (female) cells, which consistently demonstrated the same effect across all cell types (**Extended Data Fig. 5A,B**). Similar to mESC, we found X:A expression ratios close to 1 for human fibroblasts and RPE1 cells, whereas higher median X:A ratios were obtained for HEK293T and C643 cells, possibly due to aneuploidies (**Fig. 2E**). Importantly, the X:A ratio went significantly down in all cases in response to m^{6}A depletion, indicating that m^{6}A depletion results in an imbalance of X-chromosomal to autosomal transcript expression. We conclude that the same mechanism we observe in mouse is also active in humans, whereby autosomal and X-chromosomal transcripts are differentially affected by m^{6}A depletion. Our data thus supports a conserved role for m^{6}A in X-to-autosome dosage compensation in mammals.

m^{6}A is reduced on transcripts from the X chromosome

Our RNA-seq data showed that autosomal transcripts are more susceptible to m^{6}A depletion compared to X-chromosomal transcripts. To test whether these differences are driven by differential methylation levels, we analysed the distribution of m^{6}A sites across chromosomes in male mESC using miCLIP2 data[33]. Since m^{6}A detection in miCLIP2 experiments partially depends on the underlying RNA abundance[33], we quantified m^{6}A sites within expression bins (**Extended Data Fig. 6A**). Remarkably, 74.5% of all transcripts with intermediate expression (bins #4-8) harboured at least one m^{6}A site, with an average of 1-5 m^{6}A sites per transcript. In contrast, on lowly expressed transcripts (bins #1-3), we found no m^{6}A sites in most cases, most likely due to detection limits (**Fig. 3A and Extended Data Fig. 6B**).

Intriguingly, separation by chromosomes revealed a significantly lower level of m^{6}A modifications on X-chromosomal transcripts, which were reduced by almost half compared to the genomic average (56% remaining, **Fig. 3B**). In contrast, transcripts on all autosomes showed similar numbers of m^{6}A sites (**Fig. 3C and Extended Data Fig. 6C**). For further quantification, we calculated the average fold change in m^{6}A numbers on a given chromosome relative to all chromosomes. Importantly, this confirmed that all autosomes showed a similar level of m^{6}A modifications and that X-chromosomal transcripts were unique in carrying less m^{6}A (**Fig. 3D and Extended Data Fig. 6D**). As a control, we ensured that this observation was independent of differences in the numbers or lengths of transcripts between chromosomes (see Methods, **Extended Data Fig. 6E,F**). We observed the same reduction in m^{6}A levels on X-chromosomal transcripts in recently published m6A-seq2 data from mESC[36] (**Fig. 3E**).

This phenomenon was not restricted to mESC, since we found a similar reduction in m^{6}A levels on X-chromosomal transcripts in high-confidence m^{6}A sites from mouse heart (female) samples and mouse macrophages (male)[33] (**Fig. 3F**). The distinct m^{6}A patterns also extend to human cells, since human HEK293T (female) and C643 (male) cells displayed a consistent reduction of X-chromosomal m^{6}A sites (**Fig. 3G**). The strength of the reduction was to some degree tissue- and species-dependent. Collectively, we find that X-chromosomal transcripts

show fewer m^{6}A modifications than autosomal transcripts across different tissues and cell lines from mouse and human, further supporting that m^{6}A-mediated dosage compensation is a conserved mechanism.

Reduced m^{6}A levels are due to GGACH motif depletion

m^{6}A in mammals occurs mainly in a DRACH consensus sequence, with GGACH being the most frequently methylated DRACH motif[23,24]. To test whether sequence composition plays a role in the observed differences in m^{6}A levels between chromosomes, we counted the occurrence of GGACH motifs for transcripts on all chromosomes. Remarkably, transcripts on the X chromosome harboured significantly fewer GGACH motifs in their coding sequence (CDS) and 3' untranslated region (3' UTR) than autosomal transcripts (**Fig, 4A and Extended Data Fig. 7A**). Within 3' UTRs, autosomal transcripts contained on average 3.1 GGACH per kilobase of sequence, while this value dropped to 1.7 in X-chromosomal transcripts. This suggests that the lower levels of m^{6}A modifications in X-chromosomal transcripts are intrinsically encoded by lower numbers of GGACH motifs. To further investigate this, we compared strongly and weakly methylated DRACH motifs (**Extended Data Fig. 7B**). While the strong DRACH motifs were depleted on X-chromosomal transcripts, the weak DRACH motifs were equally abundant on X-chromosomal and autosomal transcripts (**Extended Data Fig. 7C,D**). This supports that the lower m^{6}A levels on X-chromosomal transcripts are a consequence of a reduced number of strongly methylated DRACH motifs. In addition, we observed that among the GGACH motifs that are present, the fraction that was methylated in mESC was slightly lower in X-chromosomal compared to autosomal transcripts (**Fig. 4B and Extended Data Fig. 7E-G**), possibly indicating that methylation efficiency of GGACH motifs is also reduced on the X chromosome. To investigate whether this is accompanied by less binding of Mettl3 to X-chromosomal genes, we analysed published Mettl3 ChIP-seq data from mESC[37]. We observed slightly fewer Mettl3 peaks on the X chromosome, indicating that the co-transcriptional recruitment of Mettl3 to X-chromosomal genes may be reduced (**Extended Data Fig. 8A**).

Previous reports suggested that X-to-autosome dosage compensation may be more relevant for certain gene sets than others. For instance, housekeeping genes have been suggested to rely more heavily on upregulation than tissue-specific genes or recently and independently evolved genes on the X chromosome[5,38,39]. However, we did not observe significant differences in GGACH motifs for different gene sets suggested from literature (**Extended Data Fig. 8B**). Furthermore, X-chromosomal genes that have been reported to escape X chromosome inactivation (escaper genes) did not show a significant difference in GGACH motifs, suggesting that they are equally depleted in m^{6}A sites as other X-chromosomal genes[40]. Nonetheless, judging from general variability in GGACH motif content, not all X-chromosomal genes appeared to be equally dependent on dosage compensation. To further dissect this, we performed gene ontology (GO) analyses on the 200 genes with least GGACH motifs, revealing functionalities related to nucleosomes/DNA packaging and ribosomes as most significantly enriched (**Extended Data Fig. 8C**). Indeed, X-chromosomal genes encoding for ribosomal proteins and histones harboured almost no GGACH motifs and thereby clearly differed from their autosomal counterparts (**Extended Data Fig. 8D**), suggesting that proteostasis of these important cellular complexes may be controlled by differential X-to-

autosomal m^{6}A methylation. This fits with previous reports showing that the majority of the *Minute* phenotypes in *Drosophila* are caused by haploinsufficiency of ribosomal proteins[41] and that ribosomal protein stoichiometry is tightly controlled in the mouse brain[42].

Next, we wanted to investigate whether GGACH motifs evolved in a sex chromosome-specific manner. Sex chromosomes are derived from ancestral autosomes. If the selective upregulation of X-chromosomal genes occurs by the reduction of GGACH motifs, outgroup species in which these genes are located on autosomes should not display such a motif disparity. For mammals, chicken is an informative outgroup to investigate the evolution of sex chromosome expression patterns, since the ancestral eutherian X chromosome corresponds to chromosomes 1 and 4 in chicken[43]. Consequently, the orthologs of X-chromosomal mouse genes are located on autosomes in chicken and are not subjected to sex chromosome-linked evolutionary changes[17] (**Fig. 4C,D**). It will be interesting to generate m^{6}A maps in different mammalian species to disentangle the contribution of m^{6}A to the evolution of mammalian dosage compensation. This will also enable the investigation of X-chromosomal regions of different evolutionary origin such as X-added region (XAR), X-conserved region (XCR) and pseudoautosomal region (PAR).

To investigate whether the reduction of GGACH motifs on the X chromosome in mouse is a sex chromosome-linked feature, we compared the GGACH motif content in chicken genes that are orthologous to mouse X-chromosomal or autosomal genes. Of note, given that almost all of these genes reside on autosomes in chicken (**Fig. 4D**), we observed no difference in GGACH content irrespective of whether the orthologs in mouse located to autosomes or the X chromosome (**Fig. 4E**). This parity of GGACH motifs in the chicken orthologs indicated that the reduced number of GGACH motifs on the mouse X chromosome has evolved specifically as a characteristic of a sex chromosome, in line with the resulting need for X-to-autosome dosage compensation.

m^{6}A contributes to dosage compensation in both sexes

The finding that GGACH motifs are less abundant on the X chromosome suggests that reduced m^{6}A levels are an intrinsic feature of X-chromosomal transcripts, which occurs in both sexes independently of X chromosome dosage. To analyse this, we performed RNA-seq experiments in female mESC in which both X chromosomes are still active and hence dosage compensation is not required. Female mESC were cultured under standard conditions to ensure maintenance of their naive state of pluripotency[32]. Since female mESC in cell culture are prone to lose one X chromosome, clonal populations of XX and X0 cells were derived from a given culture plate as matched controls[44-46]. We performed m^{6}A depletion (9 h) on 20 colonies and then determined their chromosome content by DNA-seq to choose three XX and three X0 colonies for RNA-seq analyses (**Extended Data Fig. 9A-C**). Expression analysis revealed that in female mESC with two X chromosomes, the median X:A ratio rose above 1, indicating that with two active X chromosomes, genes reach higher levels of expression than autosomes (**Fig. 4G**). This supports that one X chromosome is sufficient to obtain a median X:A ratio of 1, whereas two active X chromosomes lead to an excess of X-chromosomal gene expression. Again, the X:A ratio significantly went down upon m^{6}A depletion, further supporting that the depletion of m^{6}A impairs X-to-autosome dosage compensation.

We found that in both XX and X0 colonies, X-chromosomal transcripts significantly differed in their response to m^{6}A depletion compared to autosomal transcripts (**Fig, 4F and Extended Data Fig. 9D**). Subsequently, we identified m^{6}A sites in female bulk mESC using miCLIP2[33]. In line with our RNA-seq results, and similar to male mESC, female mESC showed a lower m^{6}A content on X-chromosomal transcripts (**Fig. 4H, Table S3**). This indicated that although both X chromosomes are still active in female mESC, the cells may be able to tolerate higher levels of X-chromosomal transcripts during very early development. The reduced X-chromosomal m^{6}A content in female mESC further supported our finding that the reduced m^{6}A levels are intrinsically encoded in the GGACH motif content. Altogether, our results indicate that m^{6}A-dependent destabilisation of autosomal transcripts also occurs in female mESC prior to X chromosome inactivation.

Discussion

X-chromosomal genes are expressed from only one active chromosome copy in mice and humans. To balance the genetic input between dual-copy autosomal and single-copy X-chromosomal transcripts, Susumo Ohno hypothesised over 50 years ago that compensating mechanisms are required for balancing gene expression[3]. Here, we uncover that differential m^{6}A methylation adds a layer of complexity to X-to-autosomal dosage compensation in eutherian mammals. This causes a global destabilisation of m^{6}A-containing autosomal transcripts, while X-chromosomal transcripts bypass this regulatory mechanism (**Fig. 5**). Importantly, we show that the inhibition of m^{6}A methylation predominantly stabilises autosomal transcripts and thereby affects the X-to-autosome balance of gene expression.

Several sex chromosome-compensating mechanisms identified so far, including X inactivation in mammals, XX dampening in *Caenorhabditis elegans* and X-chromosomal upregulation in *Drosophila melanogaster*, act on the chromatin environment of the sex chromosomes and have been shown to influence RNA polymerase II occupancy and transcription of X-chromosomal genes[7,16,47-52]. On top, RNA-regulatory mechanisms were described as X-to-autosome dosage compensation pathways. These include a higher RNA stability and translational efficiency of X-chromosomal transcripts as well as an enrichment of NMD targets and miRNA targeting sites among autosomal transcripts[4,6,8-10,53,54].

In contrast to the previously described regulatory mechanisms, m^{6}A-mediated dosage compensation acts globally at the epitranscriptomic level and adds an additional layer of regulation to X-to-autosome dosage compensation. Importantly, by inhibiting m^{6}A methylation, we can interfere experimentally with this process, thereby partly disrupting X-to-autosomal dosage compensation. We propose that m^{6}A-mediated dosage compensation is co-transcriptionally initiated in the nucleus, where m^{6}A deposition is catalysed[22], and then executed in the cytoplasm, where m^{6}A-modified transcripts are presumably degraded[25-27]. Multiple reasons are conceivable why mammals evolved an epitranscriptomic mechanism for dosage compensation. For instance, such a mechanism might be most compatible with the epigenetically installed X chromosome inactivation in females. In contrast, installing two epigenetic pathways that antagonistically affect the two X chromosomes at the same time might be more difficult to evolve. Interestingly, X chromosome inactivation has also been shown to depend on m^{6}A methylation of the non-coding RNA *Xist*[55], suggesting that dosage compensation and X chromosome inactivation might be coordinated. Furthermore, RNA-based gene regulation is often used for fine-tuning gene expression[56]. This meets the needs of dosage compensation where maximally a two-fold regulation is required. Hence, m^{6}A regulation might be ideally suited to establish and maintain small changes. Finally, RNA-based mechanisms offer an elegant means to uncouple X-to-autosome dosage compensation from other levels of gene expression regulation. Since it globally affects all X-chromosomal and autosomal transcripts that are expressed at a given moment, it facilitates genetic equilibrium between chromosomes without interfering with transcriptional regulation per se, whereby for instance, cell type-specific regulation remains unaffected.

Our data suggest that differential m^{6}A methylation evolved via a loss and/or gain of m^{6}A consensus motifs (GGACH) on X-chromosomal and autosomal transcripts during mammalian

sex chromosome evolution, respectively. This means that m^{6}A dosage compensation is hardcoded in the individual transcripts and consistently acts on both male and female cells. On top of this, there could be mechanisms that globally modulate m^{6}A methylation on X-chromosomal or autosomal transcripts, such as Mettl3 recruitment via the chromatin mark H3K36me3[57] or a local sequestration of Mettl3 via Line-1 transposons that are heavily m^{6}A-methylated and enriched on the X chromosome[58,59]. Moreover, the m^{6}A-mediated effects may be linked to the previously suggested role of NMD in X-to-autosome dosage compensation[9], since the NMD key factor UPF1 was found to associate with YTHDF2[60].

An exciting question for future research is how the hardcoding of m^{6}A-mediated dosage compensation evolved. Here, the short and redundant m^{6}A consensus sequence could enable its easy generation or removal. However, why would evolution globally select for m^{6}A sites to differentially affect transcripts from different chromosomes? We think that using predominantly hardcoded m^{6}A sites allows to globally modulate dosage compensation, for instance via the overall methylation levels or the expression of the m^{6}A reader proteins that control RNA decay under certain conditions. Even though m^{6}A levels appear to be relatively stable between tissues in mice and humans[61], it will be interesting to decipher how dosage compensation is globally modulated in different tissues, developmental stages or pathological conditions.

Acknowledgements

We kindly thank Edith Heard (EMBL Heidelberg, Germany) for providing female mESC (TX1072). We kindly thank Dan Dominissini (Tel Aviv University, Israel) for providing male mESC. We gratefully acknowledge the support of the IMB Genomics Core Facility and the use of the NextSeq 500 (funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – INST 247/870-1FUGG) and members of the Genomics and Bioinformatics Core Facilities for technical support. C.R., N.K., K.T., M.B., and M.P. were supported by the International PhD Programme on Gene Regulation, Epigenetics & Genome Stability, Mainz, Germany. Animal shapes in **Fig. 4C** were obtained from PhyloPic and are used under the Creative Common Attribution-NonCommercial-ShareAlike 3.0 Unported license.

This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) [SPP 1935 (Projektnummer 273941853), KO4566/3-2 and TRR 319 (Projektnummer 439669440) to J.K.; SPP 1935 (Projektnummer 273941853), ZA881/5-2 to K.Z; INST 247/870-1FUGG]. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions statement

The majority of bioinformatic analyses were performed by CR with help of AB and YZ. The majority of experiments were performed by NK with help of FMB, PH-K, KT, MB, MP, and MM. Statistical analyses were performed by CR and FK. CR, NK, FMB, KZ, CK-V, and JK designed the study and wrote the manuscript. CR, NK, FMB, AB, YZ, PH-K, KT, FK, MB, MP, MM, SS, CN, OR, KZ, CK-V, JK contributed to the design of the study, read and commented on the manuscript.

Competing interests statement

Oliver Rausch is an employee of STORM Therapeutics Ltd. The remaining authors declare that they have no competing interests.

References

1. Graves, J. A. Evolution of vertebrate sex chromosomes and dosage compensation. *Nat Rev Genet* **17**, 33-46 (2016).
2. Lyon, M. F. Gene action in the X-chromosome of the mouse (Mus musculus L.). *Nature* **190**, 372-373 (1961).
3. Ohno, S. *Sex Chromosomes and Sex-Linked Genes* ENDOCRINOLOGY edn, Vol. 1 (1966).
4. Yildirim, E., Sadreyev, R. I., Pinter, S. F. & Lee, J. T. X-chromosome hyperactivation in mammals via nonlinear relationships between chromatin states and transcription. *Nat Struct Mol Biol* **19**, 56-61 (2011).
5. Deng, X. *et al.* Evidence for compensatory upregulation of expressed X-linked genes in mammals, *Caenorhabditis elegans* and *Drosophila melanogaster*. *Nat Genet* **43**, 1179-1185 (2011).
6. Deng, X. *et al.* Mammalian X upregulation is associated with enhanced transcription initiation, RNA half-life, and MOF-mediated H4K16 acetylation. *Dev Cell* **25**, 55-68 (2013).
7. Talon, I. *et al.* Enhanced chromatin accessibility contributes to X chromosome dosage compensation in mammals. *Genome Biol* **22**, 302 (2021).
8. Faucillion, M. L. & Larsson, J. Increased expression of X-linked genes in mammals is associated with a higher stability of transcripts and an increased ribosome density. *Genome Biol Evol* **7**, 1039-1052 (2015).
9. Yin, S. *et al.* Evidence that the nonsense-mediated mRNA decay pathway participates in X chromosome dosage compensation in mammals. *Biochem Biophys Res Commun* **383**, 378-382 (2009).
10. Wang, Z. Y. *et al.* Transcriptome and translatome co-evolution in mammals. *Nature* **588**, 642-647 (2020).
11. Birchler, J. A. Claims and counterclaims of X-chromosome compensation. *Nat Struct Mol Biol* **19**, 3-5 (2012).
12. Bellott, D. W. & Page, D. C. Dosage-sensitive functions in embryonic development drove the survival of genes on sex-specific chromosomes in snakes, birds, and mammals. *Genome Res*, 198-210 (2021).
13. Nguyen, D. K. & Disteche, C. M. Dosage compensation of the active X chromosome in mammals. *Nat Genet* **38**, 47-53 (2006).
14. Xiong, Y. *et al.* RNA sequencing shows no dosage compensation of the active X-chromosome. *Nat Genet* **42**, 1043-1047 (2010).
15. Lin, F., Xing, K., Zhang, J. & He, X. Expression reduction in mammalian X chromosome evolution refutes Ohno's hypothesis of dosage compensation. *Proc Natl Acad Sci U S A* **109**, 11752-11757 (2012).
16. Lentini, A. *et al.* Elastic dosage compensation by X-chromosome upregulation. *Nat Commun* **13**, 1854 (2022).
17. Julien, P. *et al.* Mechanisms and evolutionary patterns of mammalian and avian dosage compensation. *PLoS Biol* **10**, e1001328 (2012).

18. 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**, 3971-3975 (1974).

19. Zaccara, S., Ries, R. J. & Jaffrey, S. R. Reading, writing and erasing mRNA methylation. *Nat Rev Mol Cell Biol* **20**, 608-624 (2019).

20. Tegowski, M., Flamand, M. N. & Meyer, K. D. scDART-seq reveals distinct m(6)A signatures and mRNA methylation heterogeneity in single cells. *Mol Cell* **82**, 868-878 e810 (2022).

21. Uzonyi, A., Slobodin, B. & Schwartz, S. Exon-intron architecture determines mRNA stability by dictating m6A deposition. *bioRxiv*, 2022.2006.2029.498130 (2022).

22. Slobodin, B. *et al.* Transcription Impacts the Efficiency of mRNA Translation via Co-transcriptional N6-adenosine Methylation. *Cell* **169**, 326-337 e312 (2017).

23. Linder, B. *et al.* Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. *Nat Methods* **12**, 767-772 (2015).

24. Pratanwanich, P. N. *et al.* Identification of differential RNA modifications from nanopore direct RNA sequencing with xPore. *Nat Biotechnol* **39**, 1394-1402 (2021).

25. Zaccara, S. & Jaffrey, S. R. A Unified Model for the Function of YTHDF Proteins in Regulating m(6)A-Modified mRNA. *Cell* **181**, 1582-1595 e1518 (2020).

26. Wang, X. *et al.* N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* **505**, 117-120 (2014).

27. Shi, H. *et al.* YTHDF3 facilitates translation and decay of N(6)-methyladenosine-modified RNA. *Cell Res* **27**, 315-328 (2017).

28. Herzog, V. A. *et al.* Thiol-linked alkylation of RNA to assess expression dynamics. *Nat Methods* **14**, 1198-1204 (2017).

29. Yankova, E. *et al.* Small-molecule inhibition of METTL3 as a strategy against myeloid leukaemia. *Nature* **593**, 597-601 (2021).

30. Poh, H. X., Mirza, A. H., Pickering, B. F. & Jaffrey, S. R. Understanding the source of METTL3-independent m6A in mRNA. *bioRxiv*, 2021.2012.2015.472866 (2021).

31. Boroviak, T. *et al.* Lineage-Specific Profiling Delineates the Emergence and Progression of Naive Pluripotency in Mammalian Embryogenesis. *Dev Cell* **35**, 366-382 (2015).

32. Geula, S. *et al.* Stem cells. m6A mRNA methylation facilitates resolution of naive pluripotency toward differentiation. *Science* **347**, 1002-1006 (2015).

33. Körtel, N. *et al.* Deep and accurate detection of m6A RNA modifications using miCLIP2 and m6Aboost machine learning. *Nucleic Acids Res* **49**, e92 (2021).

34. 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).

35. Sangrithi, M. N. *et al.* Non-Canonical and Sexually Dimorphic X Dosage Compensation States in the Mouse and Human Germline. *Dev Cell* **40**, 289-301 e283 (2017).

36. Dierks, D. *et al.* Multiplexed profiling facilitates robust m6A quantification at site, gene and sample resolution. *Nat Methods* **18**, 1060-1067 (2021).

37. Xu, W. *et al.* METTL3 regulates heterochromatin in mouse embryonic stem cells. *Nature* **591**, 317-321 (2021).

38. Pessia, E., Makino, T., Bailly-Bechet, M., McLysaght, A. & Marais, G. A. Mammalian X chromosome inactivation evolved as a dosage-compensation mechanism for dosage-sensitive genes on the X chromosome. *Proc Natl Acad Sci U S A* **109**, 5346-5351 (2012).

39. Mueller, J. L. *et al.* Independent specialization of the human and mouse X chromosomes for the male germ line. *Nat Genet* **45**, 1083-1087 (2013).

40. Fang, H., Disteche, C. M. & Berletch, J. B. X Inactivation and Escape: Epigenetic and Structural Features. *Front Cell Dev Biol* **7**, 219 (2019).

41. Marygold, S. J. *et al.* The ribosomal protein genes and *Minute* loci of *Drosophila melanogaster*. *Genome Biol* **8**, R216 (2007).

42. Amirbeigiarab, S. *et al.* Invariable stoichiometry of ribosomal proteins in mouse brain tissues with aging. *Proc Natl Acad Sci U S A* **116**, 22567-22572 (2019).

43. Bellott, D. W. *et al.* Convergent evolution of chicken Z and human X chromosomes by expansion and gene acquisition. *Nature* **466**, 612-616 (2010).

44. Schulz, E. G. *et al.* The two active X chromosomes in female ESCs block exit from the pluripotent state by modulating the ESC signaling network. *Cell Stem Cell* **14**, 203-216 (2014).

45. Zvetkova, I. *et al.* Global hypomethylation of the genome in XX embryonic stem cells. *Nat Genet* **37**, 1274-1279 (2005).

46. Choi, J. *et al.* Prolonged Mek1/2 suppression impairs the developmental potential of embryonic stem cells. *Nature* **548**, 219-223 (2017).

47. Samata, M. & Akhtar, A. Dosage Compensation of the X Chromosome: A Complex Epigenetic Assignment Involving Chromatin Regulators and Long Noncoding RNAs. *Annu Rev Biochem* **87**, 323-350 (2018).

48. Basilicata, M. F. & Keller Valsecchi, C. I. The good, the bad, and the ugly: Evolutionary and pathological aspects of gene dosage alterations. *PLoS Genet* **17**, e1009906 (2021).

49. Galupa, R. & Heard, E. X-Chromosome Inactivation: A Crossroads Between Chromosome Architecture and Gene Regulation. *Annu Rev Genet* **52**, 535-566 (2018).

50. Strome, S., Kelly, W. G., Ercan, S. & Lieb, J. D. Regulation of the X chromosomes in Caenorhabditis elegans. *Cold Spring Harb Perspect Biol* **6**, a018366 (2014).

51. Dossin, F. & Heard, E. The Molecular and Nuclear Dynamics of X-Chromosome Inactivation. *Cold Spring Harb Perspect Biol* **14**, a040196 (2022).

52. Larsson, A. J. M., Coucoravas, C., Sandberg, R. & Reinius, B. X-chromosome upregulation is driven by increased burst frequency. *Nat Struct Mol Biol* **26**, 963-969 (2019).

53. Naqvi, S., Bellott, D. W., Lin, K. S. & Page, D. C. Conserved microRNA targeting reveals preexisting gene dosage sensitivities that shaped amniote sex chromosome evolution. *Genome Res* **28**, 474-483 (2018).

54. Warnefors, M. *et al.* Sex-biased microRNA expression in mammals and birds reveals underlying regulatory mechanisms and a role in dosage compensation. *Genome Res* **27**, 1961-1973 (2017).

55. Patil, D. P. *et al.* m(6)A RNA methylation promotes XIST-mediated transcriptional repression. *Nature* **537**, 369-373 (2016).

56. Dykes, I. M. & Emanueli, C. Transcriptional and Post-transcriptional Gene Regulation by Long Non-coding RNA. *Genomics Proteomics Bioinformatics* **15**, 177-186 (2017).

57. Huang, H. *et al.* Histone H3 trimethylation at lysine 36 guides m(6)A RNA modification co-transcriptionally. *Nature* **567**, 414-419 (2019).

58. Xiong, F. *et al.* RNA m(6)A modification orchestrates a LINE-1-host interaction that facilitates retrotransposition and contributes to long gene vulnerability. *Cell Res* **31**, 861-885 (2021).

59. Chelmicki, T. *et al.* m(6)A RNA methylation regulates the fate of endogenous retroviruses. *Nature* **591**, 312-316 (2021).

60. Boo, S. H. *et al.* UPF1 promotes rapid degradation of m(6)A-containing RNAs. *Cell Rep* **39**, 110861 (2022).

61. Liu, J. *et al.* Landscape and Regulation of m(6)A and m(6)Am Methylome across Human and Mouse Tissues. *Mol Cell* **77**, 426-440 e426 (2020).

Methods

Cell culture

All cell culture was performed in a humidified incubator at 37°C and 5% CO_2. All cell lines were routinely monitored for mycoplasma contamination.

Parental male and female mouse embryonic stem cells (mESC)[32,44] were provided by Dan Dominissini (Tel Aviv University, Israel) and Edith Heard (EMBL Heidelberg, Germany). mESC lines were further authenticated by RNA-seq. Standard tissue culture was performed in 2i/LIF medium. Briefly 235 ml of each DMEM/F12 and neurobasal (Gibco, 21331020, 21103049) was mixed with 7.5 ml BSA solution (7.5%, Thermo Fisher Scientific, 11500496), 5 ml penicillin-streptomycin (P/S, Thermo Fisher Scientific, 10378016), 2 mM L-Glutamin (Thermo Fisher Scientific, 25030024), 100 µM β-mercaptoethanol (Gibco, 21985023), 5 ml mM nonessential amino acids (Gibco, 11140050), 2.5 ml N2 supplement (Gibco, 17502048), 5 ml B27 supplement (Gibco, 17504044), 3 µM CHIR99021 (Sigma, SML1046), 1 µM PD 0325901 (Biomol, 13034-1), 10 ng/ml LIF (IMB Protein Production core facility). Cell culture dishes were coated using 0.1% gelatine (Sigma, ES-006-B). The medium was exchanged every day and cells were passaged every second day. Single colonies of female mESC were picked under the microscope using a pipette tip and cultured under standard conditions in 96-wells until confluency was reached.

HEK293T (ATCC, CRL-3216) and C643 (CLS, RRID:CVCL_5969) cells were cultured in DMEM (Thermo Fisher Scientific, 21969035) supplemented with 10% fetal bovine serum (FBS, Pan Biotech, P40-47500), 1% penicillin/streptomycin (P/S, Thermo Fisher Scientific, 10378016) and 1% L-Glutamine. RPE1 (ATCC, CRL-4000) cells were cultured in DMEM/F12 (Thermo Fisher Scientific, 21331020) supplemented with 10% FBS (Pan Biotech, P40-47500), 1% P/S (Thermo Fisher Scientific, 10378016), 1% L-Glutamine and 0.04% Hygromycin B (Fisher scientific, 10453982).

Human primary dermal fibroblasts were provided by Susann Schweiger (University Medicine Mainz, Germany). Cells were grown in IMDM media (Thermo Fisher Scientific, 12440053) supplemented with 15% FBS and 1% P/S.

Primary human dermal fibroblasts derivation

Primary human dermal fibroblasts were isolated from skin punch biopsies obtained in the Children's Hospital of the University Medical Center in Mainz as previously described with small adjustment[62]. Briefly, 4 mm skin biopsies were processed in small pieces and transferred into a 6-well plate coated with 0.1% gelatine. DMEM (Thermo Fisher Scientific, 21969035) supplemented with 20% fetal bovine serum (FBS, Pan Biotech, P40-47500) and 1% penicillin/streptomycin (P/S, Thermo Fisher Scientific, 10378016) was used for culturing the skin biopsies and medium was exchanged every other day. After 3-4 weeks, when the 6-well plate was full of dermal fibroblasts that migrated out of the skin biopsies, cells were transferred to T75 flasks and cultured in standard conditions. Human dermal fibroblasts were further expanded or frozen in liquid nitrogen for long-term storage. Ethical approval by the

local ethical committee was obtained (No. 4485), and consent for research use in an anonymised way was given.

Mettl3 inhibitor treatment

For acute m^{6}A depletion in mESC, the Mettl3 inhibitor STM2457 (STORM Therapeutics) was used. Cells were treated with medium supplemented with 20 µM STM2457 in DMSO 0.2% (v/v) or with DMSO 0.2% (v/v) alone as control. m^{6}A depletion was monitored by liquid chromatography with tandem mass spectrometry (LC-MS/MS). After 3-24 h of treatment, cells were washed twice with ice-cold 1x PBS and collected on ice for further analysis

RNA isolation and poly(A) selection

Cells were washed twice with ice-cold 1x PBS and collected on ice. For total RNA isolation, the RNeasy Plus Mini Kit (Qiagen, 74136) was used following the manufacturer's instructions. For poly(A) selection, Oligo d(T)25 Magnetic Beads (Thermo Fisher Scientific, 61002) were used following the manufacturer's instructions.

qPCR

For quantification of mRNA levels, 500 ng total RNA was reverse transcribed into cDNA using the RevertAid Reverse Transcriptase (Thermo Fisher Scientific, 10161310) using Oligo(dT)18 primer (Thermo Fisher Scientific, SO131) following the manufacturer's instructions. In accordance to the manufacturer's instruction, qPCR reactions were performed in technical triplicates using the Luminaris HiGreen qPCR Master Mix, low ROX (Thermo Fisher Scientific, K0971) with forward and reverse primer (0.3 µM each) and 2 µl of 1:10 diluted cDNA as template. All qPCR reactions were run on a ViiA 7 Real-Time PCR System (Applied Biosystems). All qPCR primers are listed in **Table S4**.

LC/MS-MS

LC/MS-MS experiments were performed as described in [33]. Quantification of all samples utilised biological duplicates and averaged values of m^{6}A normalised to A, with the respective standard deviations shown.

SLAM-seq

Cell viability for optimisation

For determining the 10% maximal inhibitory concentration in a determined time window ($IC_{10,ti}$), the Cell Viability Titration Module from LeXogen (059.24) was used following the manufacturer's recommended protocol. In brief, 5,000 cells were plated in a 96-well plate one day prior to the experiment. Cells were incubated for 24 h with media supplemented with varying s^{4}U concentrations. For optimal incorporation, the s^{4}U-supplemented media were exchanged every 3 h. Cell viability was assessed using the CellTiter-Glo Luminescent Cell Viability Assay Kit from Promega (G7570) following the manufacturer's recommended protocol. The luminescence was measured using Tecan Infinite M200 Pro plate reader. Cell

doubling time of male mESC in the presence of 100 µM s^4U was 13.3 h as determined by cell counting.

SLAM-seq experiment

mRNA half-lives were determined by SLAM-seq using the Catabolic Kinetics LeXogen Kit (062.24). In brief, mESC were seeded one day prior to the experiment in a 24-well plate to reach full confluency, according to the doubling time, at the time of sample collection. The metabolic labelling was performed by addition of 100 µM s^4U to the mESC medium for 24 h. The medium was exchanged every 3 h. After the metabolic labelling, cells were washed twice with 1x PBS and fresh medium was supplemented with a 100x excess of uridine. At timepoints increasing in a 1.5x rate, medium was removed and cells were directly lysed in TRIzol (Thermo Fisher Scientific,15596026) reagent in reducing conditions. Total RNA was resuspended in the elution buffer provided by the Lexogen catabolic kit. The iodoacetamide treatment was performed using 5 µg of RNA. The library preparation for sequencing was performed using the QuantSeq 3‘ mRNA-Seq Library Prep Kit for Illumina (FWD) from Lexogen following the recommended protocol.

For stable m^6A depletion, STM2457 or DMSO was supplemented 6 h prior to the uridine chase. The media for the uridine chase were supplemented with STM2457 and DMSO for continuous m^6A depletion.

SLAM-seq library preparation

Library preparation for next-generation sequencing was performed with QuantSeq 3′ mRNA-Seq Library Prep Kit FWD (Lexogen, 015) following the manufacturer’s standard protocol (015UG009V0252). Prepared libraries were profiled on a 2100 Bioanalyzer (Agilent Technologies) and quantified using the Qubit dsDNA HS Assay Kit, in a Qubit 2.0 Fluorometer (Life Technologies). All samples were pooled together in equimolar ratio and sequenced on an Illumina NextSeq 500 sequencing device using three High Output flow cells as 84 nt single-end reads.

Data processing

Published SLAM-seq data was taken from [28]. 3’ UTR annotations were taken from [28] and filtered to match the GENCODE annotation[63] release M23. Non-overlapping annotations were discarded.

Raw data was quality checked using FastQC (v0.11.8) (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Sequencing data was processed using SLAM-DUNK (v0.4.3)[64] with the following parameters: Mapping was performed allowing multiple mapping to up to 100 genomic positions for a given read (-n 100). Reads were filtered using SLAM-DUNK -filter with default parameters. For annotation of single nucleotide polymorphisms (SNPs), all unlabelled samples were merged and SNPs were called using SLAM-DUNK snp with default parameters and -f 0.2. Transition rates were calculated using SLAM-DUNK count with default parameters, providing the SNP annotation of unlabelled

samples (-v). If more than one 3' UTR per gene remained, they were collapsed using SLAM-DUNK collapse[64]. Only genes on canonical chromosomes 1-19 and X were considered.

Principle component analysis

Principal component analysis (PCA) of SLAM-seq data was performed by estimating size factors based on read counts using the R/Bioconductor package DESeq2[65] (v1.26.0) in an R environment (v3.6.0). PCA was the performed based on the number of T-to-C reads per gene for 500 genes with the highest variance, corrected by the estimated size factors.

Incorporation rate

s^4U incorporation rates were calculated by dividing the number of T-to-C conversions on T's per 3' UTR by the overall T coverage.

Half-life calculation

To calculate mRNA half-lives, T-to-C background conversion rates (no s^4U labelling) were subtracted from T-to-C conversion rates of s^4U-labelled data. Only 3' UTRs with reads covering over 100 T's (T-coverage > 100) were kept (**Extended Data Fig. 2D**). For each timepoint, T-to-C conversion rates were normalised to the timepoint after 24 h s^4U labelling (i.e., the onset of the uridine chase) which corresponds to the highest amount of s^4U incorporation in the RNA (24 h s^4U labelling, T0) and fitted using an exponential decay model for a first-order reaction using the lm.package (as described in [28], adapted from [66]). Half-lives > 18 h (1.5 times of the last timepoint) and < 0.67 h as well as fitted values with a residual standard error > 0.3 were filtered out (**Extended Data Fig. 2E**). Only transcripts with a valid half-life calculation in both conditions were kept for further analysis. For statistical analysis of half-life fold changes, see Supplementary Methods.

RNA-seq library preparation and data processing

RNA-seq library preparation

RNA-seq library preparation was performed with Illumina's Stranded mRNA Prep Ligation Kit following Stranded mRNA Prep Ligation Reference Guide (June 2020) (Document # 1000000124518 v00). Libraries were profiled on a 2100 Bioanalyzer (Agilent technologies) and quantified using the Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific, Q32851), in a Qubit 2.0 Fluorometer (Life technologies) following the manufacturer's recommended protocols. Samples were pooled in equimolar ratios and sequenced on an Illumina NextSeq 500 sequencing device with one or two dark cycles upfront as 79, 80 or 155 nt single-end reads.

Data processing

Basic quality controls were done for all RNA-seq samples using FastQC (v0.11.8) (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Prior to mapping, possibly remaining adapter sequences were trimmed using Cutadapt[67] (v1.18). A minimal overlap of 3 nt between read and adapter was required and only reads with a length of at least 50 nt after

trimming (--minimum-length 50) were kept for further analysis. For samples sequenced with only one dark cycle at the start of the reads, 1 nt was trimmed in addition at their 5' ends (--cut 1).

Reads were mapped using STAR[68] (v2.7.3a) allowing up to 4% of the mapped bases to be mismatched (--outFilterMismatchNoverLmax 0.04 --outFilterMismatchNmax 999) and with a splice junction overhang (--sjdbOverhang) of 1 nt less than the maximal read length. Genome assembly and annotation of GENCODE[63] release 31 (human) or release M23 (mouse) were used during mapping. In the case that ERCC spike-ins were added during library preparation, their sequences and annotation (http://tools.thermofisher.com/content/sfs/manuals/ERCC92.zip) were used in combination with those from GENCODE. Subsequently, secondary hits were removed using SAMtools[69] (v1.9). Exonic reads per gene were counted using featureCounts from the Subread tool suite[70] (v2.0.0) with non-default parameters --donotsort -s2.

Differential gene expression analysis

Differential gene expression between conditions was performed using the R/Bioconductor package DESeq2 (v1.34.0) (*57*) in an R environment (v4.1.2; https://www.R-project.org/). DESeq2 was used with significance threshold of adjusted *P* value < 0.01 (used also for optimising the independent filtering). Since normalisation to total transcript abundance can introduce biases, especially when the majority of genes are affected by the treatment, we included spike-ins in our initial RNA-seq dataset. As an alternative normalisation strategy to spike-ins, we tested 100 randomly chosen genes without any m^{6}A sites but noticeable expression (reads per kilobase of transcript per million mapped reads [RPKM] > 10) for normalisation. To validate this normalisation approach, the calculated fold changes were compared with spike-in normalised data. Since the correlation between both normalisation strategies was very high, we used the 100 genes for normalisation in all further analyses (**Extended Data Fig. 3B**). For RNA seq expression change analysis see supplementary methods and **Table S5**.

miCLIP2

miCLIP2 experiments were performed as described in [33]. For a detailed description of analyses, see Supplementary Methods.

Quantification of m^{6}A sites in transcripts

m^{6}A sites from miCLIP2 for male mESC, mouse heart samples, mouse macrophages, human HEK293T, and C643 cells were taken from [33] (Gene Expression Omnibus [GEO] accession number GSE163500). m^{6}A sites were predicted using m6Aboost as described in [33]. For miCLIP2 mouse heart data, only m^{6}A sites that were predicted by m6Aboost in both considered datasets (1 µg and 300 ng) were considered for the analysis.

Comparison of m^{6}A sites per transcripts

Numbers of m^{6}A were counted for each protein-coding transcript. Only transcripts on canonical chromosomes 1-19 and X were considered. To account for expression differences, transcripts were stratified according to their expression levels based on the respective miCLIP2 data. Expression levels were estimated using htseq-count[71] (v0.11.1) and genome annotation of GENCODE[63] release M23 on the truncation reads from miCLIP2 data (noC2T reads)[33]. The derived transcript per million (TPM) values for all replicates (n = 3) were averaged, $\log_{10}$-transformed and then used to stratify all transcripts into 12 equal-width bins (step size of $\log_{10}$(TPM) = 0.25), collecting all transcripts with $\log_{10}$(TPM) < 0.5 or > 3 into the outer bins (**Extended Data Fig. 6A**). A minimum of TPM > 1 was set. For each expression bin, the mean and 95% confidence interval of the number of m^{6}A sites per transcript were calculated (**Fig. 3A-C and Extended Data Fig. 6C**). To estimate the fold change of m^{6}A sites per chromosome compared to all other chromosomes (**Fig. 3D,F,G**), only transcripts with intermediate expression (bins #3-8) were taken into account (mouse). For HEK293T data, bins #4-9, and for C643 data, bins #5-10 were used. For each bin, the difference of m^{6}A levels of the respective chromosome to all chromosomes was calculated. For this, the mean m^{6}A sites on transcripts of the respective chromosome was divided by the mean number of m^{6}A sites on transcripts of all chromosomes in the given bin (e.g., orange dots [X chromosome] over grey dots [all transcripts] in **Fig. 3B**). This resulted in a fold change of m^{6}A sites of the respective chromosome over all chromosomes for each of the six considered bins (**Extended Data Fig. 6D**). For comparison with other chromosomes (**Fig. 3D,F,G**), the mean fold change per chromosome over all expression bins was calculated (**Extended Data Fig. 6D**, red dot).

Control for transcript length biases

To exclude biases from different transcript lengths, we repeated the analysis using only m^{6}A sites within a 201-nt window (-50 nt to +150 nt) around the stop codon, where a large fraction of m^{6}A sites accumulate[23]. To obtain stop codon positions, transcript annotations from GENCODE[63] release M23 were filtered for the following parameters: transcript support level ≤ 3, level ≤ 2 and the presence of a Consensus Coding Sequence (CCDS) ID (ccdsid). If more than one transcript per gene remained, the longer isoform was chosen. Repeating the analyses with this subset as described above supported our observation that X-chromosomal transcripts harbour fewer m^{6}A sites is not influenced by differences in transcript lengths (**Extended Data Fig. 6E**).

Subsampling of transcripts in expression bins

To account for potential biases from different numbers of transcripts in the expression bins for each chromosome, we randomly picked 30 genes for each expression bin (using bins #3-5, 90 genes in total) and calculated the fold change of m^{6}A content on transcripts for each chromosome compared to all other chromosomes as described above. The procedure was repeated 100 times. The distribution of resulting fold change values supports that X-chromosomal transcripts harbour fewer m^{6}A sites, irrespective of the number of transcripts considered (**Extended Data Fig. 6F**).

Statistical analysis of m^{6}A sites in transcripts

See Supplementary Methods and **Table S6**.

Analysis of published m6A-seq2 data

Published m6A-seq2 data for wildtype (WT) and *Mettl3* KO mESC were retrieved from [36]. We used the so-called gene index, i.e., the ratio of m^{6}A IP values over IP, for whole genes as a measure of the transcripts methylation level as described in [36] (**Fig. 3E**). Chromosome locations of the genes (*n* = 6,278) were assigned using the provided gene name via the R/Bioconductor package biomaRt in an R environment[72,73].

DRACH motif analyses

GGACH motifs in mouse transcripts

Mouse transcript annotations from GENCODE[63] release M23 were filtered for the following parameters: transcript support level ≤ 3, level ≤ 2 and the presence of a CCDS ID. If more than transcript annotation remained for a gene, the longest transcript was chosen. Different transcript regions (3' UTR, 5' UTR, CDS) were grouped per gene and GGACH motifs were counted per base pair in different transcript regions, e.g., the sum of GGACH motifs in CDS fragments of a given gene, divided by sum of CDS fragment lengths.

GGACH motifs in chicken, opossum and human orthologs

Orthologs of mouse genes in chicken (*Gallus gallus*), human (*Homo sapiens*) and opossum (*Monodelphis domestica*) were retrieved from the orthologous matrix (OMA) browser[74] (accessed on 21/03/2022, for opossum 28/07/22). Only 1-to-1 orthologs were kept. Genes were filtered to have orthologs in all three species (*n* = 6,520). Then, numbers of GGACH motifs per base pair of all protein-coding exons were quantified based on GENCODE annotation (release 31)[63] for human and ENSEMBL annotation (release 107, genome assembly GRCg6a)[75] for chicken and opossum annotation (ASM229v1). GGACH motifs per base pair were quantified and visualised as described above.

Estimation of methylation levels

See Supplementary Methods.

GGACH in gene sets from literature

Independently evolved gene sets and genes with or without ortholog on the human X chromosome were taken from [39]. Escaper genes were taken from [16]. Testis-specific genes were taken from [5]. Genes from the X-added region (XAR) and X-conserved region (XCR) were annotated by identifying X-chromosomal genes in mouse with the location of chicken orthologs on chromosome 1 (XAR) and chromosome 4 (XCR).

ChIP-seq analysis

ChIP-seq peaks were obtained from [37]. The numbers of peaks per chromosome were divided by chromosome lengths. To calculate the peak ratio per chromosome compared to all other

chromosomes, the normalised peak number per chromosome was divided by the median peak number of all chromosomes.

GO analysis

GO term enrichment were performed using the enrichGO function of clusterProfiler[76] (v.4.2.2). Cellular components (ont="CC") were enriched using a *P* value cutoff of 0.01, a q value cutoff of 0.05 and *P* values were corrected using Benjamini-Hochberg correction (pAdjustMethod = "BH").

DNA-seq to determine copy number variation

See Supplementary Methods.

Statistics & Reproducibility

All statistical analyses were performed using R. All boxplots shown in this study are defined as follows: Boxes represent quartiles, centre lines denote medians, and whiskers extend to most extreme values within 1.5x interquartile range. All statistical tests performed in this study were two-tailed. All indicated replicate numbers refer to independent biological replicates. No statistical method was used to predetermine sample size. The experiments were not randomised. No data were excluded from the analysis unless stated otherwise. The Investigators were not blinded to allocation during experiments and outcome assessment.

Data availability

All high-throughput sequencing datasets generated in this study were submitted to the Gene Expression Omnibus (GEO) under the SuperSeries accession GSE203653 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?&acc=GSE203653). RNA-seq data for human primary fibroblasts is available upon request.

Code availability statement

The scripts used to process the files are accessible under the GitHub repository located at: github.com/crueckle/Rueckle_et_al_2023**.**

References Methods

62. Vangipuram, M., Ting, D., Kim, S., Diaz, R. & Schule, B. Skin punch biopsy explant culture for derivation of primary human fibroblasts. *J Vis Exp*, e3779 (2013).
63. Frankish, A. *et al.* GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res* **47**, D766-D773 (2019).
64. Neumann, T. *et al.* Quantification of experimentally induced nucleotide conversions in high-throughput sequencing datasets. *BMC Bioinformatics* **20**, 258 (2019).
65. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).
66. Rothamel, K. *et al.* ELAVL1 primarily couples mRNA stability with the 3' UTRs of interferon-stimulated genes. *Cell Rep* **35**, 109178 (2021).
67. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10-12 (2011).
68. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
69. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008 (2021).
70. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930 (2014).
71. Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166-169 (2015).
72. Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat Protoc* **4**, 1184-1191 (2009).
73. Durinck, S. *et al.* BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics* **21**, 3439-3440 (2005).
74. Altenhoff, A. M. *et al.* OMA orthology in 2021: website overhaul, conserved isoforms, ancestral gene order and more. *Nucleic Acids Res* **49**, D373-D379 (2021).
75. Cunningham, F. *et al.* Ensembl 2022. *Nucleic Acids Res* **50**, D988-D995 (2022).
76. Wu, T. *et al.* clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Camb)* **2**, 100141 (2021).

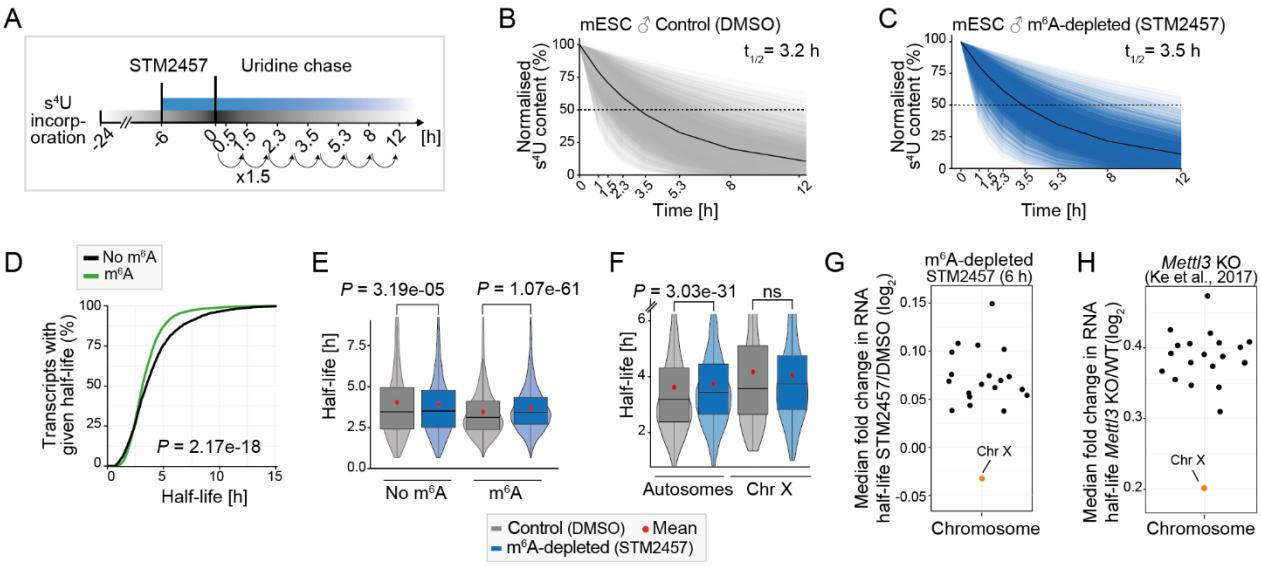


Figure 1. X-chromosomal transcripts are more stable upon m^{6}A depletion. A. Experimental setup for SLAM-seq experiment. **B, C.** Transcripts (*n* = 7,310) in ***B.*** control and ***C.*** m^{6}A-depleted conditions show a median half-life of 3.2 h and 3.5 h, respectively (*P* value = 5.25e-29, two-tailed Wilcoxon signed-rank test). Median s^{4}U content for all transcripts shown in black. **D.** Transcripts with m^{6}A sites have significantly shorter half-lives (*P* value = 2.17e-18, two-tailed Wilcoxon rank-sum test). Cumulative fractions of transcripts with given half-lives for transcripts with (*n* = 2,342, green) or without (*n* = 4,967, black) m^{6}A sites. **E.** Transcripts with m^{6}A sites (*n* = 2,342) significantly increase in half-life upon m^{6}A depletion (8% median increase, *P* value = 1.07e-61, two-tailed Wilcoxon signed-rank test), unmethylated transcripts (*n* = 4,967) were largely unaffected (0.3% median decrease, *P* value = 3.186e-05) (same gene set in both conditions). Mean half-life in each group is shown as red dot. Boxes represent quartiles, centre lines denote medians, and whiskers extend to most extreme values within 1.5x interquartile range. **F.** Half-lives of autosomal transcripts significantly increase upon m^{6}A depletion (*P* value = 3.03e-31, two-tailed Wilcoxon signed-rank test), while X-chromosomal transcripts remain unchanged (*P* value = 0.2121, two-tailed Wilcoxon signed-rank test). Distribution of half-lives for autosomal (*n* = 7,069) and X-chromosomal transcripts (*n* = 241) (same gene set in both conditions). Mean half-life in each group is shown as red dot. Boxes as in ***E***. **G.** X-chromosomal transcripts show the lowest half-life increase upon m^{6}A depletion (*P* value = 0.005486, mean difference in $\log_2$-fold changes = -0.0945, linear mixed model, two-tailed *t*-test on fixed effects, see Methods). Median fold change ($\log_2$) in mRNA half-lives for each chromosome in m^{6}A-depleted over control conditions. **H.** Same as ***G*** for half-lives from *Mettl3* KO over WT mESC[34] (*P* value = 0.000225, X-chromosomal vs. autosomal transcripts, mean difference in $\log_2$-transformed fold changes = -0.22057). The absolute differences between m^{6}A depletion and *Mettl3* KO conditions may result from differences in the experimental setup, including the mode of Mettl3 inactivation and the method to determine transcript half-lives.

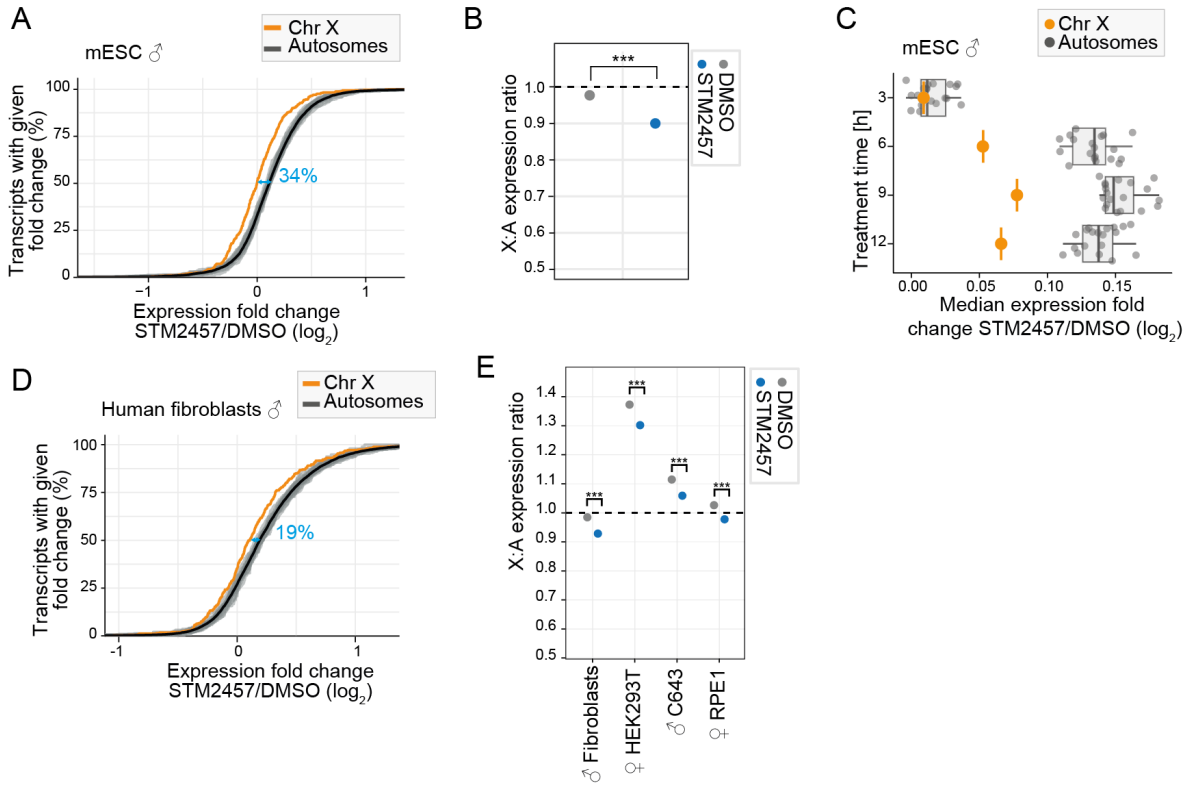


Figure 2. X-chromosomal transcripts are more stable and less upregulated upon m^{6}A depletion. A. X-chromosomal transcripts are less upregulated upon m^{6}A depletion in male mESC (*P* value = 1.86e-17, two-tailed Wilcoxon rank-sum test). Cumulative fraction of transcripts (RPKM > 1) on individual autosomes (grey) and the X chromosome (orange) that show a given expression fold change (log_2, RNA-seq) upon m^{6}A depletion (STM2457, 24 h). Mean expression changes for all autosomes are shown as black line. Effect sizes (blue) shown the shift in medians, expressed as percent of the average interquartile range of autosomal and X-chromosomal genes (IQR, see Methods). **B.** X:A expression ratios show a significant reduction upon m^{6}A depletion (*P* = 1.4e-15 two-tailed *t*-test of linear contrasts in mixed effect Gaussian model in log-scale). **C.** Differential effects on autosomal and X-chromosomal transcripts occur already after 6 h of m^{6}A depletion. Median fold changes (log_2) of transcripts from autosomes (*n* = 19, grey) and the X chromosome (*n* = 1, orange) estimated by RNA-seq at different timepoints of m^{6}A depletion (STM2457, 3, 6, 9 and 12 h). Boxes represent quartiles, centre lines denote medians, and whiskers extend to most extreme values within 1.5x interquartile range. **D.** Same as ***A***. for human primary fibroblasts (STM2457, 9 h). *P* value = 6.24e-06, two-tailed Wilcoxon rank-sum test. Effect sizes are shown as the shift in medians of the two distributions, expressed as percent of the average IQR of autosomal and X-chromosomal genes (see Methods). **E.** Same as ***B*** for human cell lines. (*P* value = 0.0000803 [human fibroblasts], *P* value = 0.0000379 [HEK293T], *P* value = 0.0003284 [C643], *P* value = 0.0002982 [RPE1]. *P* values were calculated as in ***A***, multiple testing correction.

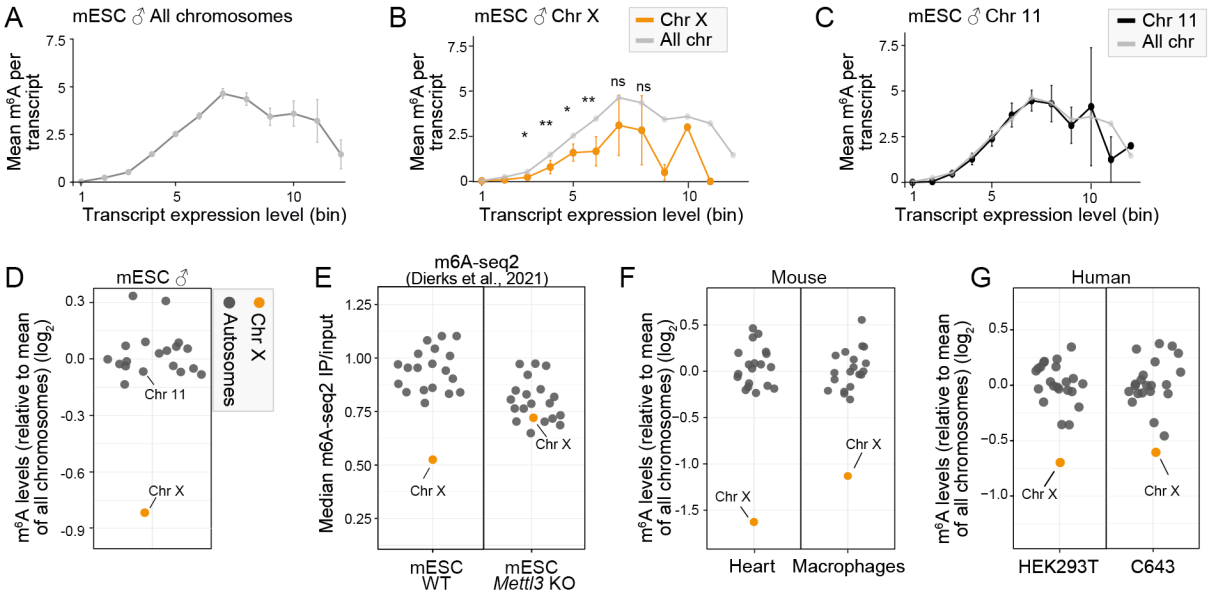


Figure 3. m^{6}A sites are reduced on transcripts from the X chromosome. A. The number of detected m^{6}A sites varies with the expression level. Mean m^{6}A sites per transcript were quantified for transcripts with each expression bin (*n* = 12,034 transcripts, see **Extended Data Fig. 6A** for *n* in each bin). Error bars indicate 95% confidence interval. **B.** X-chromosomal transcripts harbour less m^{6}A sites across expression levels. Visualisation as in ***A*** for transcripts from the X chromosome (orange, *n* = 389 transcripts) compared to the mean of all chromosomes (grey). Numbers of transcripts in expression bins are shown in **Extended Data Fig. 6C**. Significance values for bins #3-8 are indicated by asterisks (autosomes vs. X chromosome, two-tailed Wald tests in a generalized linear model for negative binomial data, multiple testing correction, ns, not significant, **P* value < 0.05, ***P* value < 0.01, exact values given in Source Data). **C.** The m^{6}A content of transcripts from chromosome 11 (*n* = 1,031 transcripts) follows the mean of all chromosomes across all expression levels. Visualisation as in ***A***. for transcripts from chromosome 11 (black) compared to the mean of all chromosomes (grey). Analyses for individual chromosomes are shown in **Extended Data Fig. 6C**. **D-G.** X-chromosomal transcripts exhibit significantly less m^{6}A sites in ***D*** male mESC (*P* = 4.1e-09, generalised linear model for negative binominal data), ***E*** published m6A-seq2 data from mESC[36], ***F*** mouse heart samples (*P* = 8.34e-11) and macrophages (*P* value = 1.38e-08), and ***G*** human HEK293T (*P* = 0.000203) and C643 cell lines (*P* value = 0.001030). Mean fold change (log_2) of m^{6}A sites per transcript on respective chromosomes relative to all chromosomes (**Extended Data Fig. 6D**). For mouse data, transcripts of intermediate expression (bins #3-8) are used. For HEK293T data, bins #4-9, and for C643 data, bins #5-10 were used. X-chromosomal and autosomal transcripts are shown in grey and orange, respectively. Chromosomes 11 and X are labelled for comparison with ***B*** and ***C***. *P* values for comparisons autosomal vs. X-chromosomal transcripts as in ***B***.

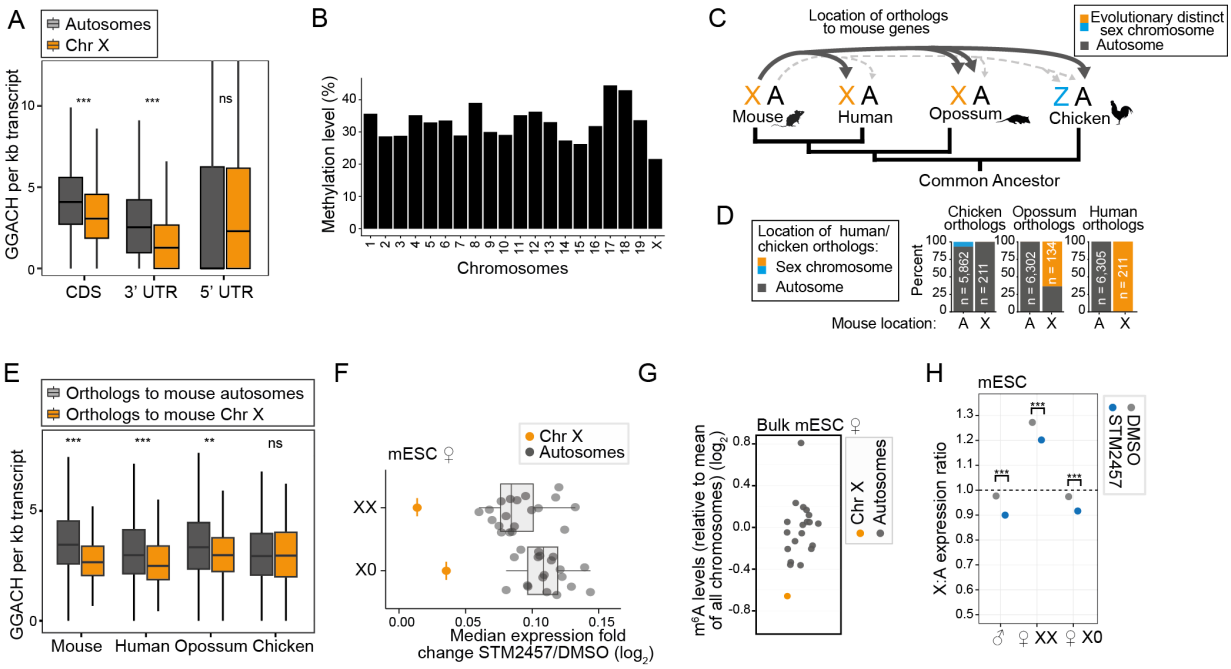


Figure 4. Reduced m^{6}A on X-chromosomal transcripts is intrinsically encoded. A. GGACH motifs (normalised to region length) in different transcript regions of autosomal (grey) and X-chromosomal transcripts (orange) in mouse (*P* value = 1.38e-29 [CDS, *n* = 16,631 annotations], *P* value = 1.06e-40 [3‘ UTR, *n* = 16,484 annotations] and 0.2707 [5’ UTR, *n* = 16,490 annotations], two-tailed Wilcoxon rank-sum test). **B.** Methylation levels of GGACH motifs are slightly reduced on X-chromosomal transcripts. Fraction of m^{6}A sites per chromosome with methylation in miCLIP2 data from male mESC. Boxes represent quartiles, centre lines denote medians, and whiskers extend to most extreme values within 1.5x interquartile range. **C.** Location of mouse X-chromosomal orthologs in human, opossum, and chicken **D.** Percentage of orthologs of X-chromosomal or autosomal genes in mouse that are located on autosomes or sex chromosomes in human, opossum, and chicken. **E.** GGACH motifs in transcripts (exons) from mouse genes and corresponding orthologs in chicken, opossum, and human (*n* = 6,520). Orthologs to mouse X-chromosomal and autosomal genes are indicated in orange and grey, respectively (two-tailed Wilcoxon rank-sum test, ns, not significant, **P* value < 0.05, ***P* value < 0.01, ****P* value < 0.001, *P* value = 1.2e-18 [mouse], 2.7e-06 [human], 0.001227 [opossum], 0.8602 [chicken]). Boxes as in *A*. **F.** Effects of m^{6}A depletion on autosomal and X-chromosomal transcripts in XX and X0 clones of female mESC (*P* value = 1.64e-12 and 3.5e-11, respectively, two-tailed Wilcoxon rank-sum test, **Extended Data Fig. 9A-C**). Median fold changes (log_2) of transcripts from autosomes (n = 19, grey) and the X chromosome (n = 1, orange) estimated by RNA-seq after m^{6}A depletion (STM2457, 9 h). Boxes as in ***A***. **G.** X:A expression ratios are significantly reduced upon m^{6}A depletion (*P* value = 4.12e-15 [mESC], *P* value = 2.06e-11 [female mESC XX], *P* value = 1.08e-10 [female mESC X0]. *P* values as in **Fig. 2B**, multiple testing correction). **H.** Median fold change (log_2) of m^{6}A sites per transcript on respective chromosome relative to all chromosomes (*P* = 0.0018, autosomal (grey) vs. X-chromosomal (orange) transcripts, two-tailed Wald test in generalised linear mixed model for negative binominal data).

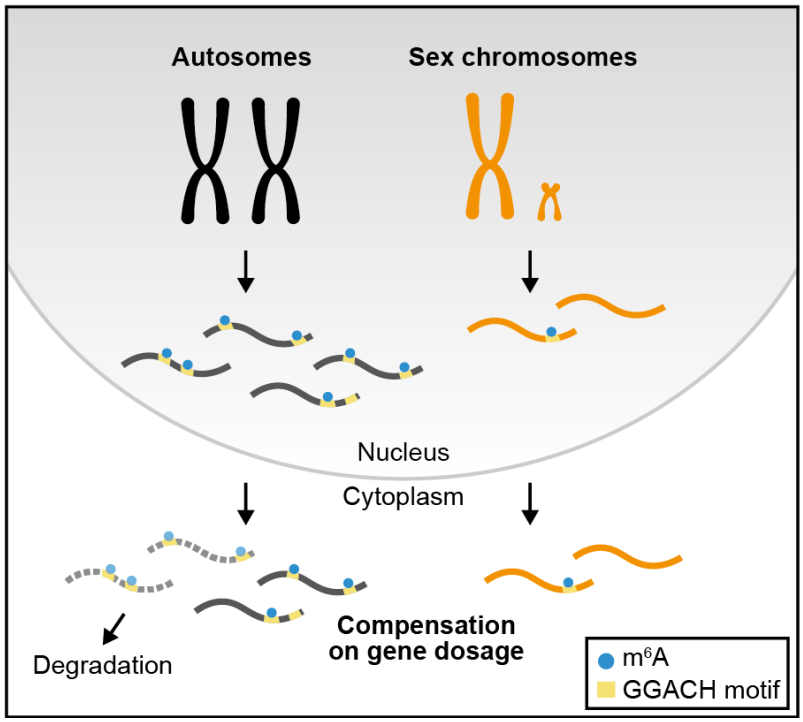


Figure 5. The role of m^{6}A in X-to-autosome dosage compensation. m^{6}A acts as a selective degradation signal on autosomal transcripts and thereby contributes to X-to-autosome dosage compensation. Transcripts from the autosomes are transcribed from two active chromosomes, leading to higher transcript copy numbers per autosomal gene than for X-chromosomal genes. m^{6}A is selectively enriched on transcripts from autosomes, leading to their destabilisation and degradation. Since m^{6}A is not enriched on X-chromosomal transcripts, this leads to an equal dosage between autosomal and X-chromosomal transcripts. m^{6}A thereby contributes to X-to-autosome dosage compensation.

Supplementary Methods

SLAM-seq

Statistical analysis of half-life fold changes

The influence of chromosome type on log_2-transformed fold changes in mRNA half-lives upon m^6A depletion (**Fig. 1G**) or *Mettl3* KO (**Fig. 1H**) was analysed using a categorical Gaussian linear mixed model. Distributional assumptions (normal distribution and homoscedasticity) were checked with Q-Q plots and by comparing empirical standard deviations. The factor *chromosome type* (autosome / X chromosome) was implemented as a fixed effect. To account for differences between individual chromosomes, the factor *chromosome number* (1 - 19, X) was included as a random effect. We used the R packages lme4 (v1.1.29) and lmerTest (v3.1.3). In both datasets, the fits of the random effect's variance were singular, meaning that the effect of individual chromosomes was negligible compared to the effect of chromosome type and that autosomal log_2-transformed fold changes could be pooled to form one group. Inference using Wald tests in the resulting models is equivalent to unpaired Student's *t*-tests for autosomal and X-chromosomal log_2-transformed fold changes.

Analysis of expression changes (RNA-seq)

For comparison of expression changes between groups, log_2-transformed fold changes were used. Only genes with a mean RPKM > 1 over all samples were considered. Effect sizes between groups were calculated as follows: The median log_2-transformed fold change of all autosomal genes was subtracted from the median log_2-transformed fold change of all X-chromosomal genes. This value was divided by the mean interquartile range (IQR) of both distributions, reported as the corresponding IQR of the median shift. The median shifts and IQR values for all datasets are summarised in **Table S5**.

Median X:A expression ratios were calculated using the pairwiseCI package in R using 'Median.ratio' with 10,000 bootstrap replications as described before[1]. We used categorical weighted mixed-effect Gaussian models for the analysis of RPKM levels in different cell lines (mESC male / XX / X0 and human fibroblasts / HEK293T / C643 / RPE1). We fitted the models with the R package *lme4*[2] (v1.1.29) and performed statistical inference with the R packages *lmerTest*[3] (v3.1.3) and *emmeans* (v1.8.0). A separate model was fitted for each cell line. The response variable was log-transformed mean RPKM values, filtered for mean values > 1. The factors *treatment* (DMSO and STM2457) and *chromosome type* (autosomal and X) were implemented as fixed effects. The *factor* gene ID was implemented as a random effect to account for the correlation of RPKM values belonging to the same gene. We used inverse variance weighting to account for heteroscedasticity. We used tests based on the multivariate *t*-distribution to assess for both treatments if the RPKM log-ratio between X-chromosomal and autosomal genes was different from 1 and if the ratios were different between treatments. The P values are adjusted for multiple testing per model.

miCLIP2 to map m^{6}A sites

miCLIP2 experiment

miCLIP2 experiments in female mESC were performed as described in [4] using 1 µg of input material per replicate. For all experiments, the m^{6}A-specific polyclonal antibody from SynapticSystems (cat. 202 003) was used. 6 µg m^{6}A-specific antibody was used per 1 µg of RNA.

The miCLIP2 libraries were sequenced on an Illumina NextSeq 500 sequencing machine as 92-nt single-end reads including a 6-nt sample barcode as well as 5+4-nt unique molecular identifiers (UMIs) yielding between 32 and 46 million reads. Basic quality controls were done using FastQC (v0.11.8) (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and reads were filtered based on sequencing qualities (Phred score) in the barcode and UMI regions using the FASTX-Toolkit (v0.0.14) (http://hannonlab.cshl.edu/fastx_toolkit/) and seqtk (v1.3) (https://github.com/lh3/seqtk/). Flexbar[5] (v3.4.0) was used to de-multiplex reads based on the sample barcode on positions 6 to 11 of the reads. Subsequently, UMI and barcode regions as well as adapter sequences were trimmed from read ends using Flexbar requiring a minimal overlap of 1 nt of read and adapter and adding UMIs to the read names. Reads shorter than 15 nt were removed from further analysis. The downstream analysis was done as described in Chapters 3.4 and 4.1 of Busch et al.[6] with an additional step to remove reads directly mapped to the chromosome ends. Those reads do not have an upstream position and, thus, no crosslink position can be extracted. Genome assembly and annotation of GENCODE[7] (release M23) were used during mapping with STAR[8] (v2.7.3a). Information on possibly occurring mutations was collected through the MD tag by running STAR with option "--outSAMattributes All".

After removing duplicates, all mutations found in reads were extracted using the Perl script parseAlignment.pl of the CLIP Tool Kit[9] (CTK, v1.1.3). The list of all found mutations was filtered for C-to-T mutations using basic Bash commands and kept in BED file format as described in [10]. Reads in this list (i.e., reads with C-to-T mutations) were removed from the de-duplicated BAM file using SAMtools[11] (v1.9) and basic Bash commands. The resulting BAM file with the truncation reads (noC2T) was transformed to a BED file using bedtools bamtobed[12] (BEDTools v2.27.1) considering only the 5' mapping position of each read. Afterwards, the BED file was sorted and summarised to strand-specific bedGraph files, which were shifted by one base pair upstream (since this nucleotide is considered as the cross-linked nucleotide) using bedtools genomecov (BEDtools v2.27.1). All bedGraph files were transformed to bigWig track files using bedGraphToBigWig of the UCSC tool suite[13] (v365).

m^{6}A sites were predicted as described in [4]. In brief, peaks were called on noC2T reads (BAM files) using PureCLIP[14] (v1.3.1) and filtered for the presence in 3 out of 4 replicates. Then, m^{6}A sites were predicted using the machine learning model m6Aboost which we previously trained to discriminate m^{6}A sites from background in miCLIP2 data, based on data from *Mettl3* KO and control mESC. A detailed description of the method can be found in [4].

Statistical analysis of half-life fold changes

The influence of chromosome type on log_2-transformed fold changes in mRNA half-lives upon m^{6}A depletion (**Fig. 1G**) or *Mettl3* KO (**Fig. 1H**) was analysed using a categorical Gaussian linear mixed model. Distributional assumptions (normal distribution and homoscedasticity) were checked with Q-Q plots and by comparing empirical standard deviations. The factor *chromosome type* (autosome / X chromosome) was implemented as a fixed effect. To account for differences between individual chromosomes, the factor *chromosome number* (1 - 19, X) was included as a random effect. We used the R packages lme4 (v1.1.29) and lmerTest (v3.1.3). In both datasets, the fits of the random effect's variance were singular, meaning that the effect of individual chromosomes was negligible compared to the effect of chromosome type and that autosomal log_2-transformed fold changes could be pooled to form one group. Inference using Wald tests in the resulting models is equivalent to unpaired Student's *t*-tests for autosomal and X-chromosomal log_2-transformed fold changes.

Analysis of expression changes (RNA-seq)

For comparison of expression changes between groups, log_2-transformed fold changes were used. Only genes with a mean RPKM > 1 over all samples were considered. Effect sizes between groups were calculated as follows: The median log_2-transformed fold change of all autosomal genes was subtracted from the median log_2-transformed fold change of all X-chromosomal genes. This value was divided by the mean interquartile range (IQR) of both distributions, reported as the corresponding IQR of the median shift. The median shifts and IQR values for all datasets are summarised in **Table S5**.

Median X:A expression ratios were calculated using the pairwiseCI package in R using 'Median.ratio' with 10,000 bootstrap replications as described before[1]. We used categorical weighted mixed-effect Gaussian models for the analysis of RPKM levels in different cell lines (mESC male / XX / X0 and human fibroblasts / HEK293T / C643 / RPE1). We fitted the models with the R package *lme4*[2] (v1.1.29) and performed statistical inference with the R packages *lmerTest*[3] (v3.1.3) and *emmeans* (v1.8.0). A separate model was fitted for each cell line. The response variable was log-transformed mean RPKM values, filtered for mean values > 1. The factors *treatment* (DMSO and STM2457) and *chromosome type* (autosomal and X) were implemented as fixed effects. The *factor* gene ID was implemented as a random effect to account for the correlation of RPKM values belonging to the same gene. We used inverse variance weighting to account for heteroscedasticity. We used tests based on the multivariate *t*-distribution to assess for both treatments if the RPKM log-ratio between X-chromosomal and autosomal genes was different from 1 and if the ratios were different between treatments. The P values are adjusted for multiple testing per model.

miCLIP2 to map m^{6}A sites

miCLIP2 experiment

miCLIP2 experiments in female mESC were performed as described in [4] using 1 µg of input material per replicate. For all experiments, the m^{6}A-specific polyclonal antibody from SynapticSystems (cat. 202 003) was used. 6 µg m^{6}A-specific antibody was used per 1 µg of RNA.

The miCLIP2 libraries were sequenced on an Illumina NextSeq 500 sequencing machine as 92-nt single-end reads including a 6-nt sample barcode as well as 5+4-nt unique molecular identifiers (UMIs) yielding between 32 and 46 million reads. Basic quality controls were done using FastQC (v0.11.8) (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and reads were filtered based on sequencing qualities (Phred score) in the barcode and UMI regions using the FASTX-Toolkit (v0.0.14) (http://hannonlab.cshl.edu/fastx_toolkit/) and seqtk (v1.3) (https://github.com/lh3/seqtk/). Flexbar[5] (v3.4.0) was used to de-multiplex reads based on the sample barcode on positions 6 to 11 of the reads. Subsequently, UMI and barcode regions as well as adapter sequences were trimmed from read ends using Flexbar requiring a minimal overlap of 1 nt of read and adapter and adding UMIs to the read names. Reads shorter than 15 nt were removed from further analysis. The downstream analysis was done as described in Chapters 3.4 and 4.1 of Busch et al.[6] with an additional step to remove reads directly mapped to the chromosome ends. Those reads do not have an upstream position and, thus, no crosslink position can be extracted. Genome assembly and annotation of GENCODE[7] (release M23) were used during mapping with STAR[8] (v2.7.3a). Information on possibly occurring mutations was collected through the MD tag by running STAR with option "--outSAMattributes All".

After removing duplicates, all mutations found in reads were extracted using the Perl script parseAlignment.pl of the CLIP Tool Kit[9] (CTK, v1.1.3). The list of all found mutations was filtered for C-to-T mutations using basic Bash commands and kept in BED file format as described in [10]. Reads in this list (i.e., reads with C-to-T mutations) were removed from the de-duplicated BAM file using SAMtools[11] (v1.9) and basic Bash commands. The resulting BAM file with the truncation reads (noC2T) was transformed to a BED file using bedtools bamtobed[12] (BEDTools v2.27.1) considering only the 5' mapping position of each read. Afterwards, the BED file was sorted and summarised to strand-specific bedGraph files, which were shifted by one base pair upstream (since this nucleotide is considered as the cross-linked nucleotide) using bedtools genomecov (BEDtools v2.27.1). All bedGraph files were transformed to bigWig track files using bedGraphToBigWig of the UCSC tool suite[13] (v365).

m^{6}A sites were predicted as described in [4]. In brief, peaks were called on noC2T reads (BAM files) using PureCLIP[14] (v1.3.1) and filtered for the presence in 3 out of 4 replicates. Then, m^{6}A sites were predicted using the machine learning model m6Aboost which we previously trained to discriminate m^{6}A sites from background in miCLIP2 data, based on data from *Mettl3* KO and control mESC. A detailed description of the method can be found in [4].

Statistical analysis of m^{6}A sites in transcripts

To analyse the m^{6}A sites in autosomes and the X chromosome, stratified by expression bins, a categorical generalized linear model for negative binomial data was fitted using the core R routine glm.nb (R version 4.1.2). The factors *chromosome type* (autosome / X chromosome) and *expression bin* (#3-8), as well as their interaction, were implemented. Based on visual assessment of the fits and on chi-squared tests for goodness of fit, the negative binomial model was selected in preference to a Poisson model. For each expression bin, Wald tests were used to test the difference between autosomes and the X chromosome. The *P* values were corrected for multiple testing (FWER-control) using the single step method implemented in the R package multcomp (v1.4.19).

To analyse the general influence of the factor *chromosome type* on m^{6}A sites, categorical generalized linear mixed models for negative binomial data were fitted using the R packages lme4 (v1.1.29) and lmerTest (v3.1.3). The factor *chromosome type* was implemented as a fixed main effect. The influences of expression bins and chromosome number were included as random effects. For the analysis of the mouse data sets, expression bins #3-8 were considered (**Figs. 3D,F and 4G**). Bins #4-9 were analysed in the HEK293T data set and bins #5-10 were analysed in the C643 data set (**Fig. 3G**). For each data set, the negative binomial models were preferable to Poisson models (visual assessment and chi-squared tests for fit of distribution). For the mouse heart data set, the likelihood ratio test and AIC comparison showed that the random effect *chromosome number* was not necessary to explain the data. The model was therefore fitted for the factors *chromosome type* and *expression bin*. The influence of the factor *chromosome type* on the m^{6}A counts was tested with Wald tests. The fitted values and 95% confidence intervals (Wald type) of the fold changes (log_2) of expected m^{6}A counts in X-chromosomal over autosomal transcripts for all figures are reported in **Table S6**.

Estimation of methylation levels

Transcript annotations were taken from GENCODE (genome release M23, release 31), selecting one transcript per gene with the following hierarchy: (i) highest transcript support level, (ii) highest gene support level, and (iii) longest transcript. GGACH motifs were identified in each transcript using the R/Bioconductor package Biostrings (v2.59.2) and grep. To take into account only GGACH motifs in transcript regions with sufficient expression, we calculated the local read coverage in the miCLIP2 data. For this, the truncation reads from miCLIP2 data (noC2T reads) were converted into a single nucleotide coverage using bamCoverage (v3.5.1) from the deepTools suite[15]. The local read coverage was estimated as the median single nucleotide coverage in a 21-nt window centred on each GGACH motif. The GGACH motifs were binned by their log_2-transformed local coverage, adding a pseudo-count of 1 before log_2 transformation. Within each bin, the percentage of GGACH motifs harbouring high-confidence m^{6}A sites predicted by m6Aboost was calculated. Since m^{6}A detection partly depends on expression, this value increases steadily with increasing expression bins and then levels off at a certain methylation level. To determine this, a local linear regression curve was fitted using loess.smooth and used to identify the point at which the slope drops below 0.01 (**Extended Data Fig. 7E,F,G**). The corresponding percentage of GGACH motifs with an m^{6}A site was used as an estimate of the methylation level on a given chromosome. If the slope for a given chromosome did not drop below 0.01 due to coverage limitations, the percentage of methylated GGACH motifs at the transition point between bins #11 and #12 was taken to estimate the methylation level for this chromosome.

DNA-seq to determine chromosome copy numbers

DNA isolation

Cells were washed twice with ice-cold 1x PBS and collected on ice. For DNA isolation, the PureLink Genomic DNA MINI Kit (Invitrogen, 10593245) was used following the manufacturer's instructions.

DNA-seq library preparation

DNA-seq library preparation was performed by using genomic DNA, which was sheared with a Covaris E220 focused ultrasonicator. NGS library preparation was performed using half of the reaction of NEBNext Ultra II DNA Library Prep Kit for Illumina Version 6.0, 3/20 following the manufacturer's recommended protocol. Libraries were profiled on a 2100 Bioanalyzer (Agilent technologies) and quantified using the Qubit dsDNA HS Assay Kit, in a Qubit 2.0 Fluorometer (Life technologies). All samples were pooled in equimolar ratio and sequenced on an Illumina NextSeq500 sequencing device using a Mid Output flow cell as 159-nt single-end reads.

DNA-seq data processing

Basic quality controls were done for all DNA-seq samples using FastQC (v0.11.8) (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Possibly remaining adapter sequences were trimmed using Cutadapt[16] (v2.4) prior to mapping. A minimal overlap of 3 nt between reads and adapter was required and only reads with a length of at least 20 nt after trimming (--minimum-length 20) were kept for further analysis. Reads were mapped from start to end (--end-to-end) using Bowtie2[17] (v2.3.4.3) without allowing any mismatches in a seed alignment (-N 0) of length 31 (-L 31). Additional parameters specifying the behaviour of multi-seed alignments were set as -i S,1,0.50 -D 20 -R 3. Genome assembly of GENCODE[7] release 31 (human) or release M23 (mouse) were used during mapping. Subsequently, multi-mapping or low-quality alignments were removed using SAMtools[11] (v1.9). Since sequencing of DNA samples was very shallow, detected duplicates are very likely PCR duplicates rather than real duplicates. Thus, they were removed using Picard (v2.20.3) (https://github.com/broadinstitute/picard).

To determine copy number variations, mapped reads were counted in 100 kilobase bins for each chromosome and normalised by library size. The ratio for each bin was calculated by dividing the number of mapped reads per bin by the median of mapped reads of all bins and chromosomes. Only the canonical chromosomes 1-19 and X were considered.

Supplementary Tables

Table S1. Half-lives measured by SLAM-seq in male mESC under m^{6}A-depleted (STM2457) and control conditions. Half-lives for control and m^{6}A-depleted conditions are given for each gene with the corresponding residual standard error which indicates the goodness of the fit (see Methods). Additionally, the mean T coverage over all replicates and samples which was used for expression estimations is given for each condition.

Table S2. Summary of SLAM-seq, RNA-seq, and DNA-seq experiments conducted in this study. Table summarises the numbers of reads for all high-throughput sequencing experiments conducted in this study. For RNA-seq and DNA-seq experiments, the numbers of total sequenced reads and uniquely mapped reads are given. For SLAM-seq, the numbers of sequenced and retained read (SLAM-DUNK) are given. For miCLIP2, the numbers of uniquely mapped reads and reads after duplicate removal are given.

Table S3. Identified m^{6}A sites for miCLIP2 data on bulk female mESC. Table provides information on all m6Aboost-predicted m^{6}A sites (*n* = 33,371) in the miCLIP2 data performed on bulk female mESC. Coordinates are given in a bed file-compatible format, i.e., as 0-based, right-open intervals.

Provided as worksheets in Excel file Supplementary Tables.

Table S4. List of qPCR primers used to validate RNA expression upon m^{6}A depletion in male mESC. Oligonucleotides used as primers for qPCR experiments in **Extended Data Figs. 1D and 4C** are given. For each oligonucleotide, the sequence and target transcript are given together with the primer orientation (forward or reverse).

Name	Sequence 5' - 3'	Transcript	Orientation
qPCR_mNanog-for	CCTCCAGCAGATGCAAGAACTC	*Nanog*	Forward
qPCR_mNanog-rev	CTTCAACCACTGGTTTTTCTGCC	*Nanog*	Reverse
qPCR_mSox2-for	ACAGATGCAACCGATGCACC	*Sox2*	Forward
qPCR_mSox2-rev	TGGAGTTGTACTGCAGGGCG	*Sox2*	Reverse
Plp1_qPCR_for	CCAGAATGTATGGTGTTCTCCC	*Plp1*	Forward
Plp1_qPCR_rev	GGCCCATGAGTTTAAGGACG	*Plp1*	Reverse
Fmr1_qPCR_for	GGTCAAGGAATGGGTCGAGG	*Fmr1*	Forward
Fmr1_qPCR_rev	AGTTCGTCTCTGTGGTCAGAT	*Fmr1*	Reverse
Ssr4_qPCR_for	ACCACAGATCACCCCTTCTTAC	*Ssr4*	Forward
Ssr4_qPCR_rev	CCACTAACGTCGGCATAAAGAG	*Ssr4*	Reverse
Hnrnph2_qPCR_for	GGAGGGGTTCGTGGTGAAG	*Hnrnph2*	Forward
Hnrnph2_qPCR_rev	GAACACCTGATGTGCCATTTTG	*Hnrnph2*	Reverse
Itm2a_qPCR_for	TTGCCTCATACTTATGTGGTTCG	*Itm2a*	Forward
Itm2a_qPCR_rev	GCGGAAGGATTTTCGGTTGTTG	*Itm2a*	Reverse

(Continued on next page)

Table S4. List of qPCR primers used to validate RNA expression upon m^{6}A depletion in male mESC. (Continued from previous page)

Name	Sequence 5' - 3'	Transcript	Orientation
Rab11fip5_qPCR_for	CTCTGGACGAGGTCTTCCG	*Rab11fip5*	Forward
Rab11fip5_qPCR_rev	TGTTCCGTGTGAACTGGATGG	*Rab11fip5*	Reverse
Tubb3_qPCR_for	TAGACCCCAGCGGCAACTAT	*Tubb3*	Forward
Tubb3_qPCR_rev	GTTCCAGGTTCCAAGTCCACC	*Tubb3*	Reverse
Phax_qPCR_for	CGATGACGATTGCTCTCTTTGG	*Phax*	Forward
Phax_qPCR_rev	CGCATCTTGATTCTGTTCCTGG	*Phax*	Reverse
Faap100_qPCR_for	GGACGCGAGTTCGTCTATGTG	*Faap100*	Forward
Faap100_qPCR_rev	ACAGGACGTAGAGTGCCCT	*Faap100*	Reverse
Tpst2_qPCR_for	CGTGCTGTGTAACAAGGACC	*Tpst2*	Forward
Tpst2_qPCR_rev	CGTCACGCACCATTAGCAG	*Tpst2*	Reverse
qPCR_mGapdh-for	TCACCACCATGGAGAAGGC	*Gapdh*	Forward
qPCR_mGapdh-rev	CCCTTTTGGCTCCACCCT	*Gapdh*	Reverse

Table S5. Additional information for estimated effect sizes. Effect sizes for comparisons of fold changes between groups, e.g., differences in expression fold changes upon m^{6}A depletion between X-chromosomal and autosomal transcripts (**Fig. 2A**) are reported as the difference in medians of both distributions, divided by the mean interquartile range (IQR) of both distributions (see Methods). This table summarises the corresponding values for all effect sizes reported in this study, including the corresponding figure, the dataset analysed, the median shift between X-chromosomal and autosomal transcripts, the effect size, and the IQRs of distributions.

Figure	Dataset	Effect size	Median shift	IQR Chr X	IQR autosomes
Fig. 2A	mESC male RNA-seq	34%	0.11	0.31	0.35
Fig. 2D	Human primary fibroblasts RNA-seq	19%	0.08	0.4	0.47
Extended Data Fig. 4B	mESC male (3 h STM2457) RNA-seq	2%	0.0045	0.21	0.25
Extended Data Fig. 4B	mESC male (6 h STM2457) RNA-seq	27%	0.09	0.29	0.34
Extended Data Fig. 4B	mESC male (9 h STM2457) RNA-seq	22%	0.08	0.36	0.35
Extended Data Fig. 4B	mESC male (12 h STM2457) RNA-seq	21%	0.07	0.35	0.35
Extended Data Fig. 5B	Human HEK293T RNA-seq	17%	0.07	0.4	0.46
Extended Data Fig. 5B	Human C643 RNA-seq	19%	0.097	0.52	0.49
Extended Data Fig. 5B	Human RPE1 RNA-seq	18%	0.08	0.44	0.43
Extended Data Fig. 9D	mESC female X0 RNA-seq	24%	0.08	0.33	0.33
Extended Data Fig. 9D	mESC female XX RNA-seq	26%	0.08	0.31	0.32

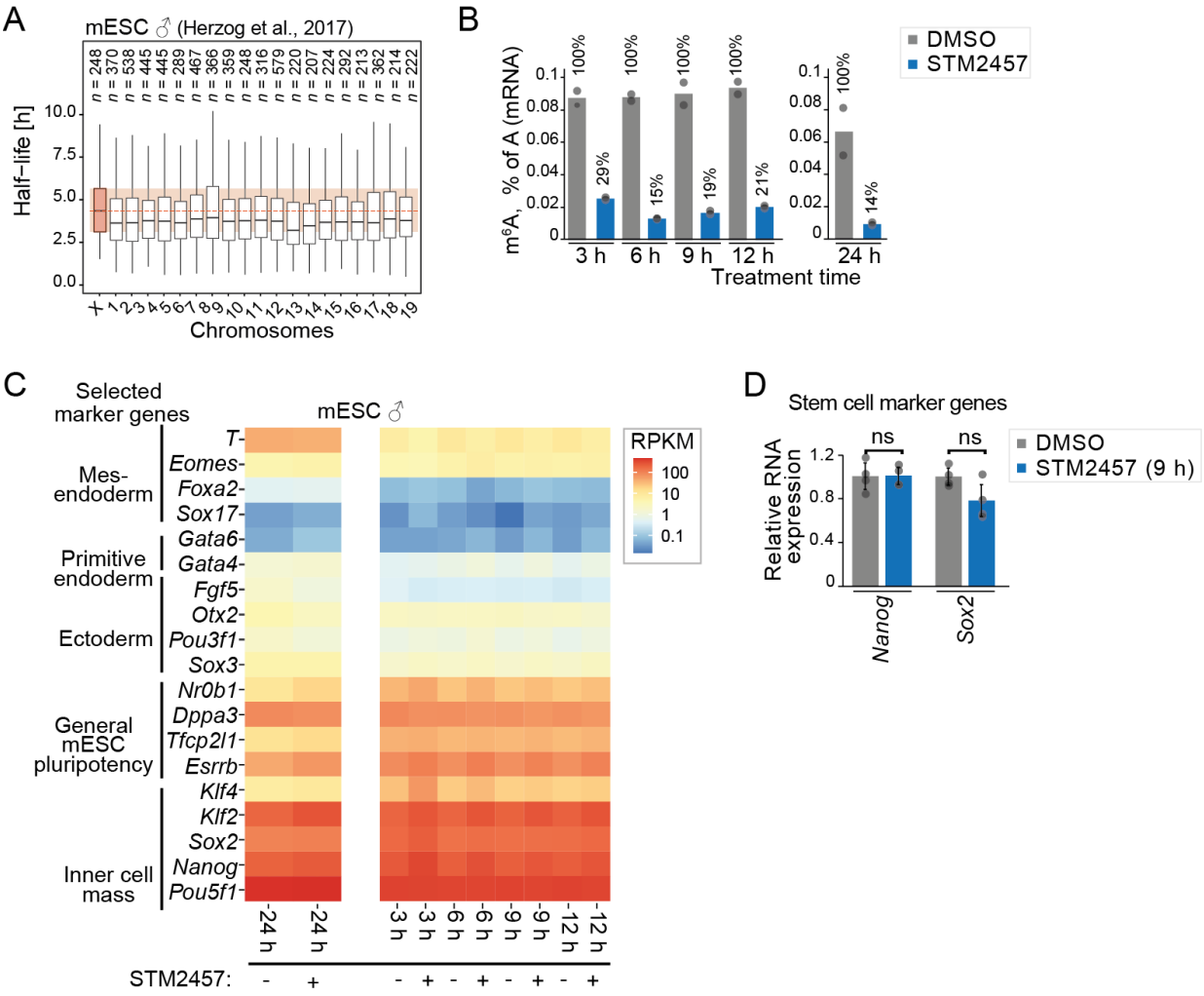
Table S6. Additional information for statistical analyses of m^{6}A sites in transcripts. To analyse the general influence of the chromosome type on the number of m^{6}A sites in transcripts, categorical generalised linear mixed models for negative binomial data were fitted to the data (see Methods "Statistical analyses of m^{6}A sites in transcripts"). This table summarises the fitted values and 95% confidence intervals (Wald type) of the fold changes (log_2) of expected m^{6}A counts in X-chromosomal over autosomal transcripts as well as the two-tailed Wald test *P* values. The confidence intervals and *P* values in this table are not corrected for multiple testing.

Figure	Fold change (log_2)	95% confidence interval	*P* value
Fig. 3D (male mESC)	-0.8178638	[-1.0904474, -0.5452803]	4.1e-09
Fig. 3F (heart)	-1.586387	[-2.065105, -1.107670]	8.34e-11
Fig. 3F (macrophages)	-1.0423472	[-1.4023045, -0.6823898]	1.38e-08
Fig. 3G (HEK293T)	-0.5777994	[-0.8826179, -0.2729808]	0.000203
Fig. 3G (C643)	-0.6506555	[-1.0391719, -0.2621391]	0.001030
Fig. 4H (bulk female mESC)	-0.6324775	[-1.0297596, -0.2351954]	0.0018

Supplementary References

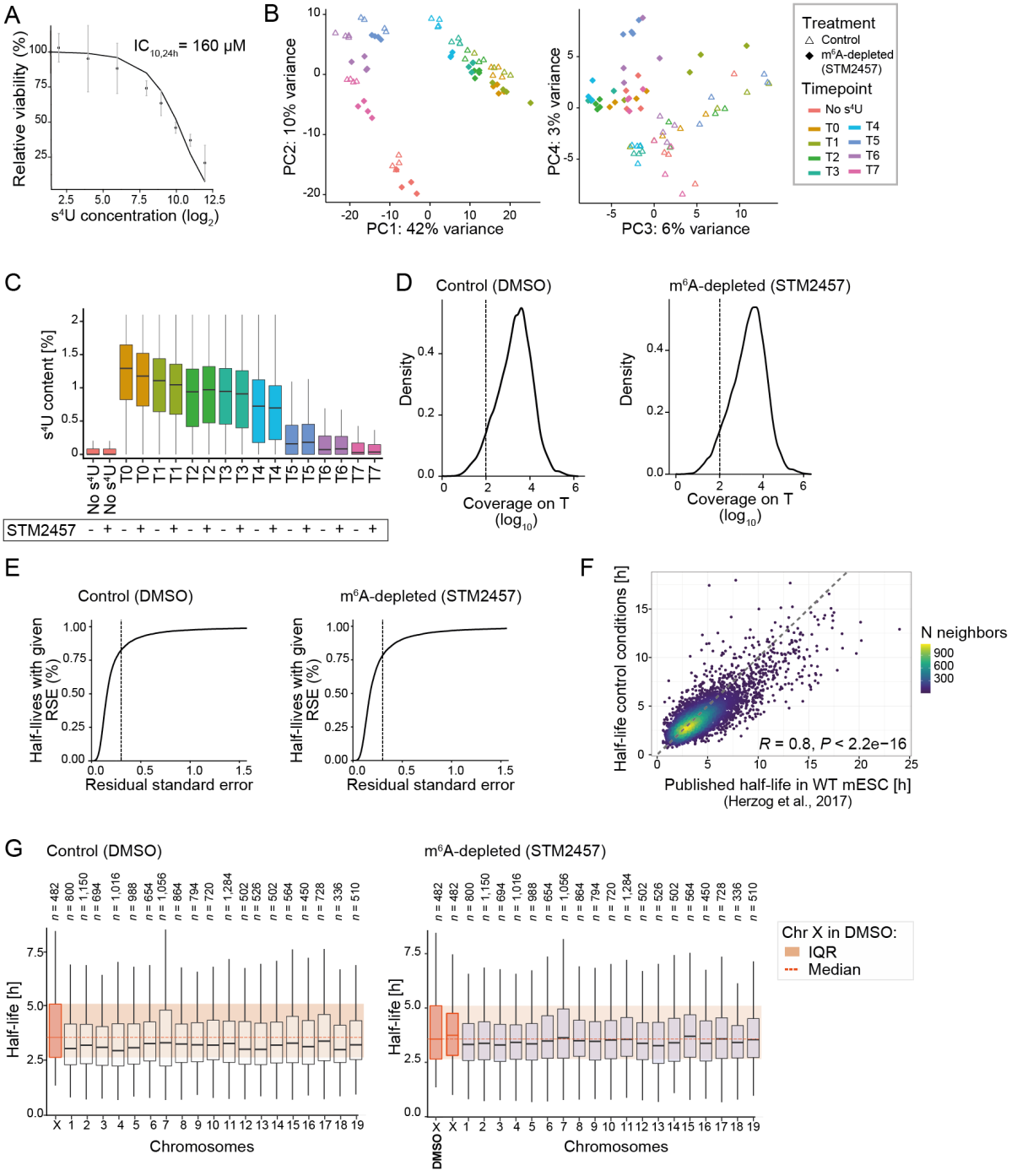
1. Sangrithi, M. N. *et al.* Non-Canonical and Sexually Dimorphic X Dosage Compensation States in the Mouse and Human Germline. *Dev Cell* **40**, 289-301 e283 (2017).
2. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* **67**, 1-48 (2015).
3. Kuznetsova, A., Brockhoff, P. B. & Christensen, R. H. B. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* **82**, 1-26 (2017).
4. Körtel, N. *et al.* Deep and accurate detection of m6A RNA modifications using miCLIP2 and m6Aboost machine learning. *Nucleic Acids Res* **49**, e92 (2021).
5. Roehr, J. T., Dieterich, C. & Reinert, K. Flexbar 3.0 - SIMD and multicore parallelization. *Bioinformatics* **33**, 2941-2942 (2017).
6. Busch, A., Brüggemann, M., Ebersberger, S. & Zarnack, K. iCLIP data analysis: A complete pipeline from sequencing reads to RBP binding sites. *Methods* **178**, 49-62 (2020).
7. Frankish, A. *et al.* GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res* **47**, D766-D773 (2019).
8. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
9. Shah, A., Qian, Y., Weyn-Vanhentenryck, S. M. & Zhang, C. CLIP Tool Kit (CTK): a flexible and robust pipeline to analyze CLIP sequencing data. *Bioinformatics* **33**, 566-567 (2017).
10. Hawley, B. R. & Jaffrey, S. R. Transcriptome-Wide Mapping of m(6) A and m(6) Am at Single-Nucleotide Resolution Using miCLIP. *Curr Protoc Mol Biol* **126**, e88 (2019).
11. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008 (2021).
12. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842 (2010).
13. Kent, W. J., Zweig, A. S., Barber, G., Hinrichs, A. S. & Karolchik, D. BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics* **26**, 2204-2207 (2010).
14. Krakau, S., Richard, H. & Marsico, A. PureCLIP: capturing target-specific protein-RNA interaction footprints from single-nucleotide CLIP-seq data. *Genome Biol* **18**, 240 (2017).
15. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**, W160-165 (2016).
16. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10-12 (2011).
17. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359 (2012).

Extended Figures



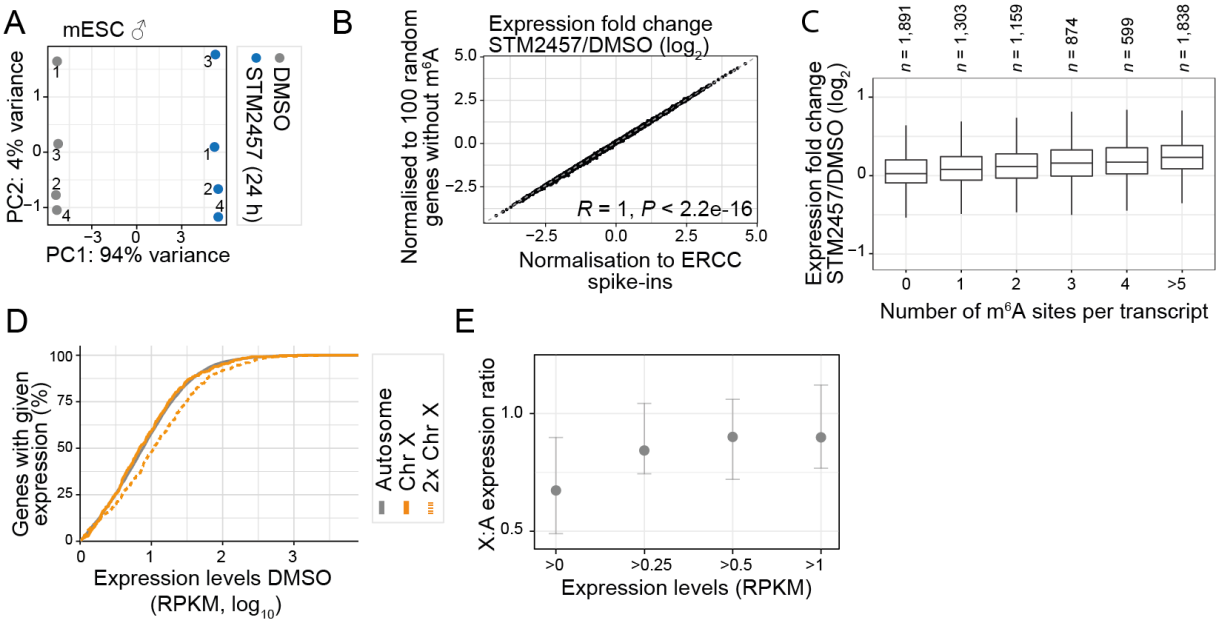
Extended Data Figure 1. Mettl3 inhibitor treatment of mouse embryonic stem cells (mESC) depletes m^{6}A levels. A. X-chromosomal transcripts are more stable than autosomal transcripts (median half-life = 3.72 h [autosomes] vs. 4.35 h [X chromosome], *P* value = 1.02e-05, two-sided Wilcoxon rank-sum test). Distribution of half-lives from published SLAM-seq data for mESC for transcripts on each individual chromosome. Dashed red line and red box indicate median and inter-quartile range of X-chromosomal transcripts, respectively, for comparison. Boxes represent quartiles, centre lines denote medians, and whiskers extend to most extreme values within 1.5x interquartile range. **B.** Time course experiments shows that treatment of male mESC with the Mettl3 inhibitor (STM2457, 20 µM) results in a gradual reduction of m^{6}A levels on mRNAs. m^{6}A levels were measured by liquid chromatography-tandem mass spectrometry (LC-MS/MS) for poly(A)+ RNA from m^{6}A-depleted (STM2457, 3-24 h) and control conditions. Quantification of m^{6}A as percent of A in poly(A)+ RNA. *n* = 2 independent biological replicates. **C.** Expression levels of marker genes confirm the pluripotent state of the male mESC throughout the time course experiment. Gene expression levels (RNA-seq) are shown as reads per kilobase of transcript per million mapped reads (RPKM, mean over all replicates, $\log_{10}$) in m^{6}A-depleted (STM2457, 3-24 h) and control conditions. **D.** Quantitative real-time PCR (qPCR) to quantify expression changes of stem cell marker genes in m^{6}A-depleted (STM2457, 9 h) and control conditions. Normalised C_T values

(ΔC_T, normalised to *Gapdh* expression) are compared between conditions. Fold changes are displayed as mean ± s.d.m., two-sided Student's *t*-test on $\log_2$-transformed data, n = 4 independent biological samples, ns, not significant. *P* value = 0.8 [*Sox2*]; 0.96 [*Nanog*].

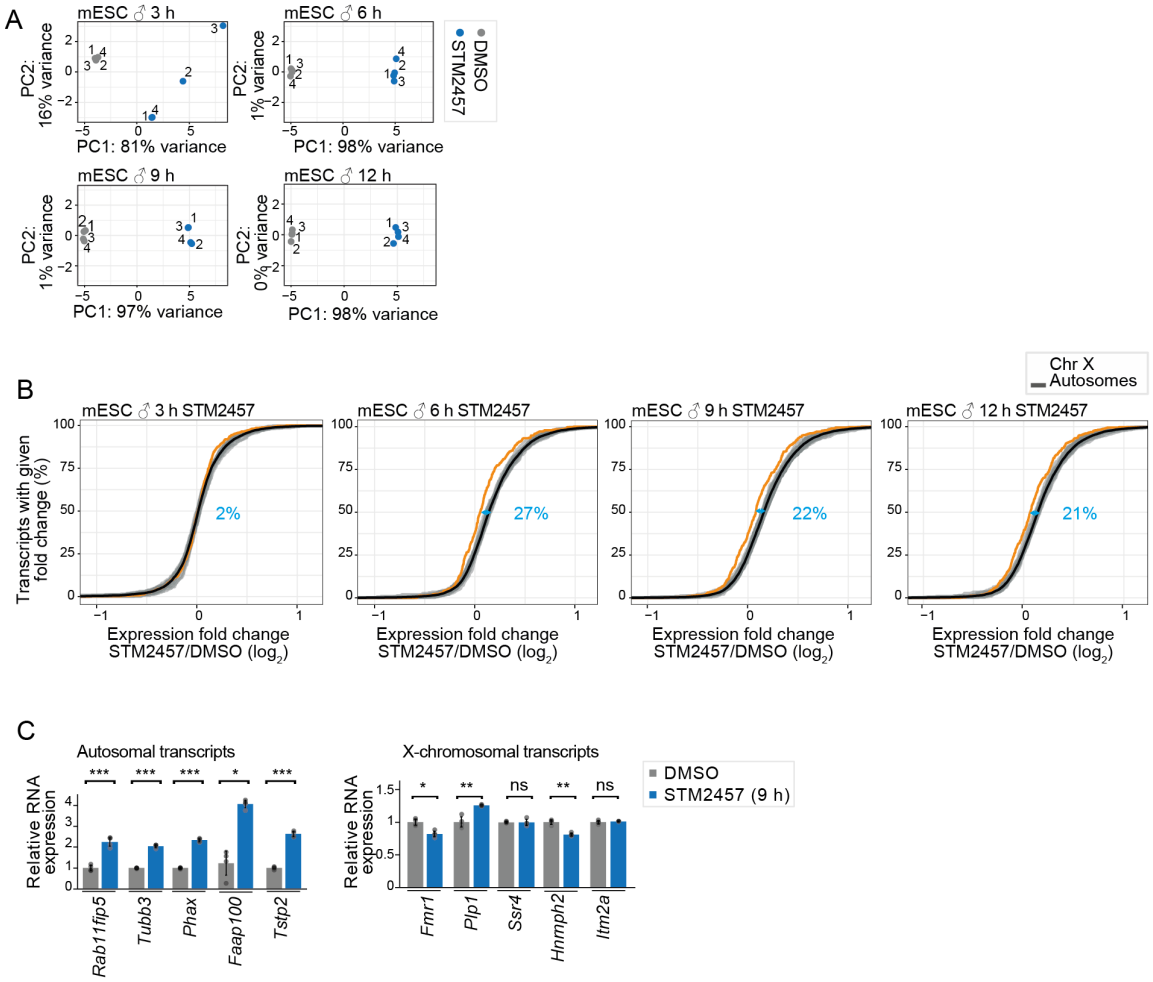


Extended Data Figure 2. SLAM-seq measures mRNA half-lives in mESC. A. Cell viability assessed for male mESC cultured with s^{4}U for 24 h in varying concentrations (x-axis, $\log_2$-transformed). Viability of labelled cells in relation to unlabelled cells is shown as mean ± s.d.m., n = 3 biologically independent samples. $IC_{10,24h}$ is indicated as dashed line. **B.** Principal

component analysis of SLAM-seq replicates based on numbers of reads with T-to-C conversions. Principal component (PC) 1 and PC2 (left) separate the different timepoints of the experiment (colours), PC3 and PC4 (right), separate control and m^{6}A-depleted conditions (symbols). **C.** T-to-C conversions on T's by the overall T coverage per 3' UTR. Maximum s^{4}U rate is achieved after 24 h of labelling (T0) and steadily decreases after s^{4}U washout and uridine chase (T1-T7). Unlabelled samples (No s^{4}U) are shown for comparison. *n* = 21,527 UTRs with incorporation rates per replicate. Boxes represent quartiles, centre lines denote medians, and whiskers extend to most extreme values within 1.5x interquartile range. **D.** Expression estimates based on $\log_{10}$-transformed coverage on T's per 3' UTR (mean over all replicates and timepoints per condition). Only 3' UTRs with SLAM-seq reads covering at least 100 T's (indicated by dotted line) were used for subsequent fitting. **E.** Cumulative distribution of the goodness-of-fit (residual standard error, RSE) of half-lives calculated from SLAM-seq data. Dotted lines indicate filtering cut-off (RSE > 0.3). **F.** Correlation of half-lives determined in this study (male mESC, control condition) with previously published half-lives in male mESC (two-sided Pearson correlation coefficient [*R*] = 0.8, *P* value < 2.2e-16). **G.** Distribution of half-lives of transcripts on individual chromosomes in control (left) or m^{6}A-depleted conditions (right). In control conditions, half-lives of X-chromosomal transcripts differ significantly from autosomal transcripts (median half-life 3.19 h [autosomes] vs. 3.57 [X chromosome], *P* value = 7.63e-05, two-sided Wilcoxon rank-sum test). In m^{6}A-depleted conditions, autosomal transcript half-lives approximate X-chromosomal transcript half-lives in control conditions (*P* value = 0.06228, two-sided Wilcoxon rank-sum test). Red lines and boxes indicate median and interquartile range, respectively, of half-lives of X-chromosomal transcripts in control conditions. Boxes as in ***C***.

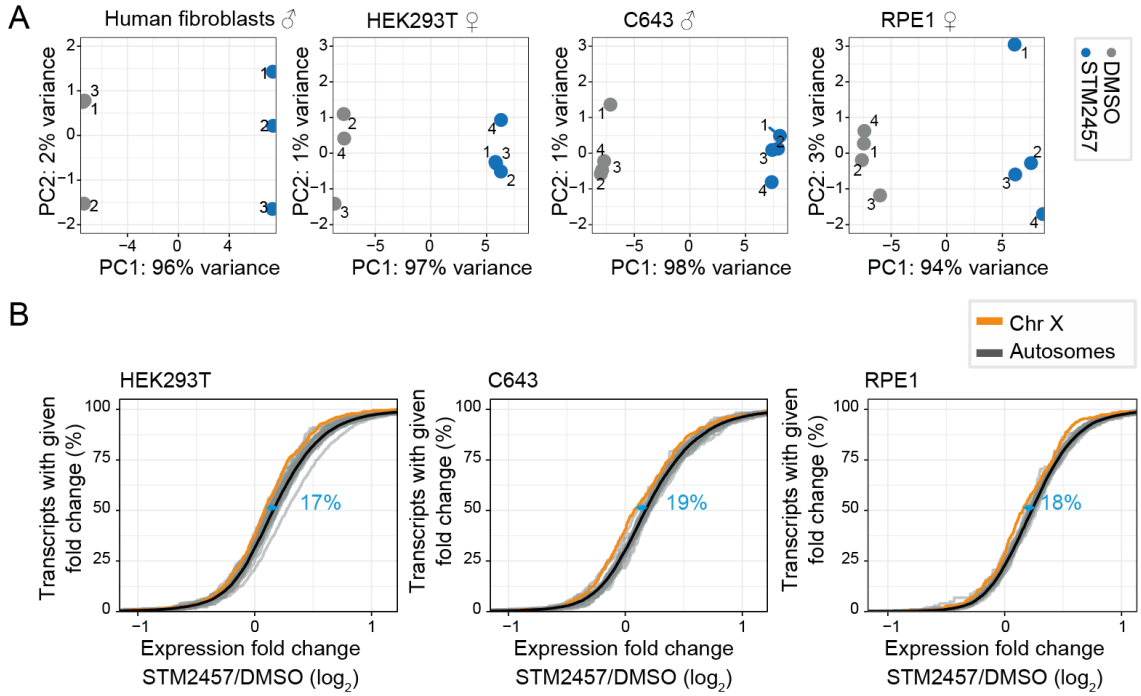


Extended Data Figure 3. RNA-seq upon m^{6}A depletion reveals upregulation of autosomal but not X-chromosomal transcripts. A. Principal component analysis indicates high reproducibility of RNA-seq data for male mESC in control and m^{6}A-depleted conditions (STM2457, 24 h, 4 replicates per condition, total of 398 million uniquely mapped reads). Replicate number given next to each data point. **B.** Correlation of expression fold changes (log_2) of RNA seq data in m^{6}A-depleted (STM2457, 24 h) over control conditions using normalisation to ERCC spike-ins (x-axis) or 100 randomly chosen genes without m^{6}A sites (y-axis, see Methods; two-sided Pearson correlation coefficient [*R*] = 1, *P* value < 2.2e-16). **C.** Upregulation upon m^{6}A depletion increases with the number of m^{6}A sites in the transcripts. Distribution of fold changes (log_2) in m^{6}A-depleted (STM2457, 24 h) over control conditions in expressed transcripts (transcripts per million [TPM] > 1, based on total miCLIP2 signal) stratified by their number of m^{6}A sites. Numbers of transcripts in each category are indicated above. Boxes represent quartiles, centre lines denote medians, and whiskers extend to most extreme values within 1.5x interquartile range. **D.** Cumulative distribution of expressed autosomal (grey) and X-chromosomal (orange) transcripts (RPKM > 1) with a given expression level (RPKM, x-axis). The expression distributions of X-chromosomal and autosomal transcripts are largely identical, supporting a X:A ratio close to 1 across the full expression range. For comparison, a theoretical doubling of the X-chromosomal expression is shown (orange, dotted) which would exceed autosomal expression levels. **E.** Median X-to-autosome (X:A) expression ratios increase with higher RPKM cut-offs (>0 , *n* [genes] = 26,291, ≥0.25, *n* = 13,795, ≥0.5 *n* = 12,255, ≥1 , *n* = 10,849). Median X:A ratios for male mESC and 95% confidence intervals are given.

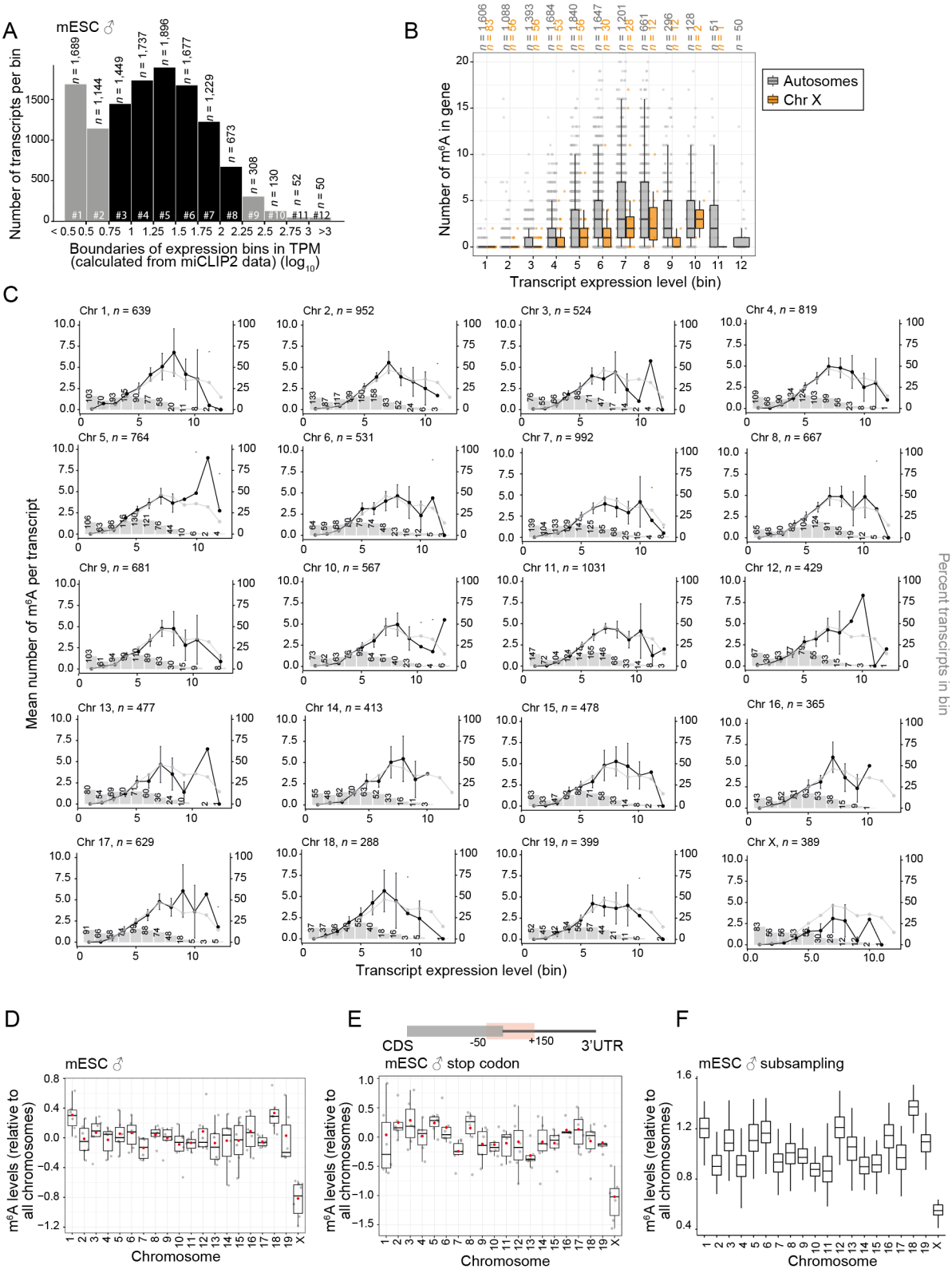


Extended Data Figure 4. Time-course RNA-seq upon m^{6}A depletion reveals upregulation of autosomal genes after 6 h of inhibitor treatment. A. Principal component analyses of RNA-seq replicates of control and m^{6}A-depleted male mESC at different time points (STM2457, 3-12 h) based on numbers of reads or the 500 genes with highest variance across all samples for a given time point. Replicate number given next to each data point. **B.** After 6 h of m^{6}A depletion, X-chromosomal transcripts show significantly lower fold changes (log_2) compared to autosomal transcripts (*P* value = 0.48 [3 h], *P* value = 1.02e-12 [6 h], *P* value = 5.12e-10 [9 h], *P* value = 1.69e-08 [12 h], two-sided Wilcoxon rank-sum test). Cumulative fraction of transcripts on individual autosomes (grey) and the X chromosome (orange) that show a given expression fold change (log_2, RNA-seq) at different timepoints of m^{6}A depletion (STM2457, 3-12 h) in male mESC. Mean expression changes for all autosomes are shown as black line. Effect sizes (blue) show the shift in medians, expressed as percent of the average interquartile range (IQR) of autosomal and X-chromosomal genes (see Methods). **C.** qPCR to quantify expression changes of five autosomal (left) and five X-chromosomal (right) transcripts in control and m^{6}A-depleted (STM2457, 9 h) male mESC cells. Normalised C_T values (ΔC_T, normalised to *Gapdh* expression) are compared between conditions. Fold changes are displayed as mean ± s.d.m., two-sided Student's *t*-test on log_2-transformed data, *n* = 4 biologically independent samples, **P* value < 0.05, ***P* value < 0.01, ****P* value < 0.001, ns, not significant. *P* value = 0.00017

[*Rab11fip5*], 8.57e-07 [*Tubb3*], 8.08e-08 [*Phax*], 0.049 [*Faap100*], 1.46e-06 [*Tstp2*]; 0.56 [*Itm2a*], 0.001 [*Hnrnph2*], 0.95 [*Ssr4*], 0.007 [*Plp1*], 0.01 [*Fmr1*].

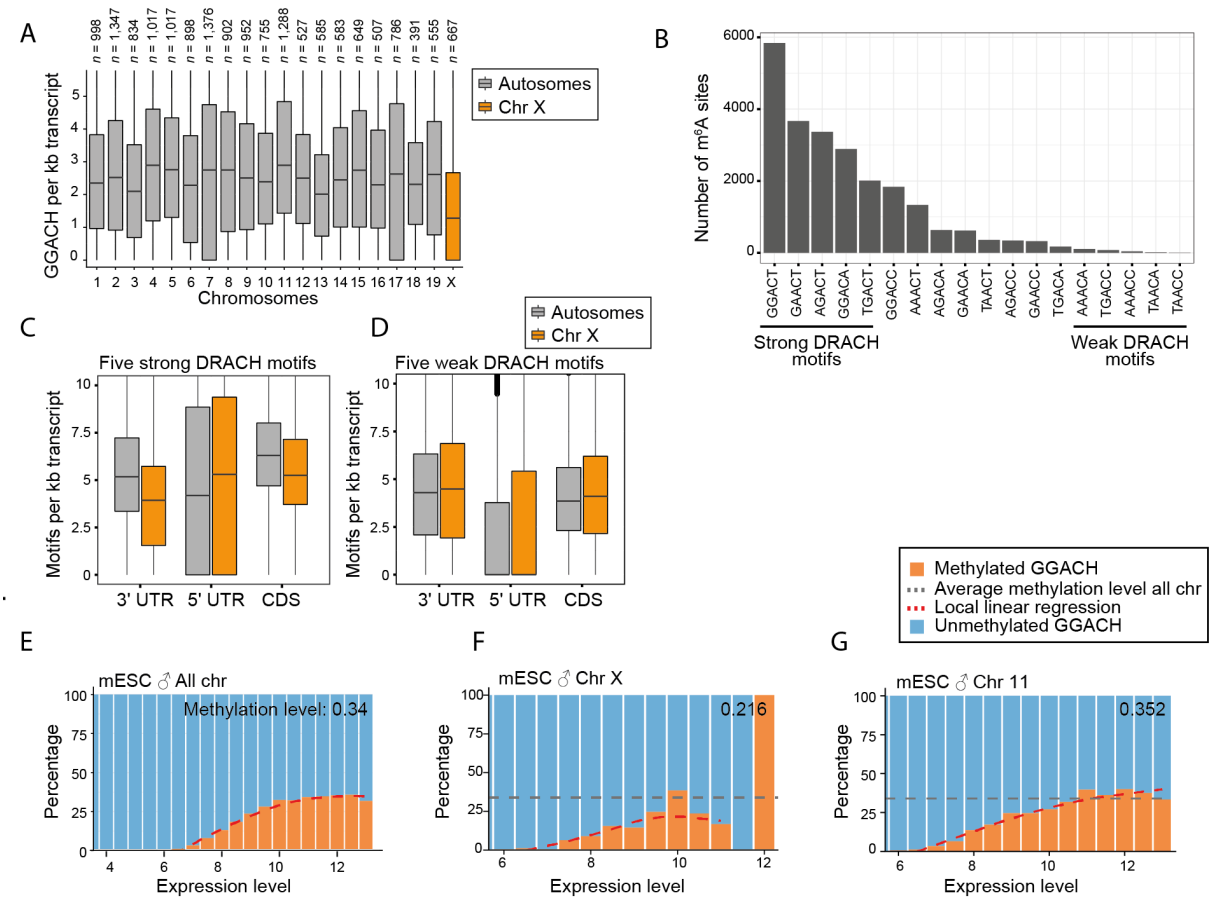


Extended Data Figure 5. RNA-seq upon m^{6}A depletion reveals upregulation of autosomal transcripts in human cell lines. A. Principal component analyses for replicates of RNA-seq experiments under m^{6}A-depleted and control conditions for human primary fibroblasts (STM2457, 9 h), HEK293T cells, C643 cells and RPE1 cells (STM2457, 24 h). Replicate number given next to each data point. **B.** X-chromosomal transcripts show significantly lower fold changes upon m^{6}A depletion than autosomal transcripts (*P* value = 6.92e-06 [HEK293T, *n* = 12,856 of autosomal transcripts, *n* = 443 of X-chromosomal transcripts], *P* value = 4.53e-05 [C643, *n* = 11,109 of autosomal transcripts, *n* = 383 of X-chromosomal transcripts], *P* value = 0.0001901 [RPE1, *n* = 10,732 of autosomal transcripts, *n* = 347 of X-chromosomal transcripts], Wilcoxon rank-sum test). Cumulative fraction of transcripts on individual autosomes (grey) and the X chromosome (orange) that show a given fold change (log_2) in m^{6}A-depleted (STM2457, 24 h) over control conditions for HEK293T, C643, and RPE1 cells. Mean expression changes for all autosomes are shown as black line. Effect sizes (blue) shown the shift in medians, expressed as percent of the average IQR of autosomal and X-chromosomal genes (see Methods).



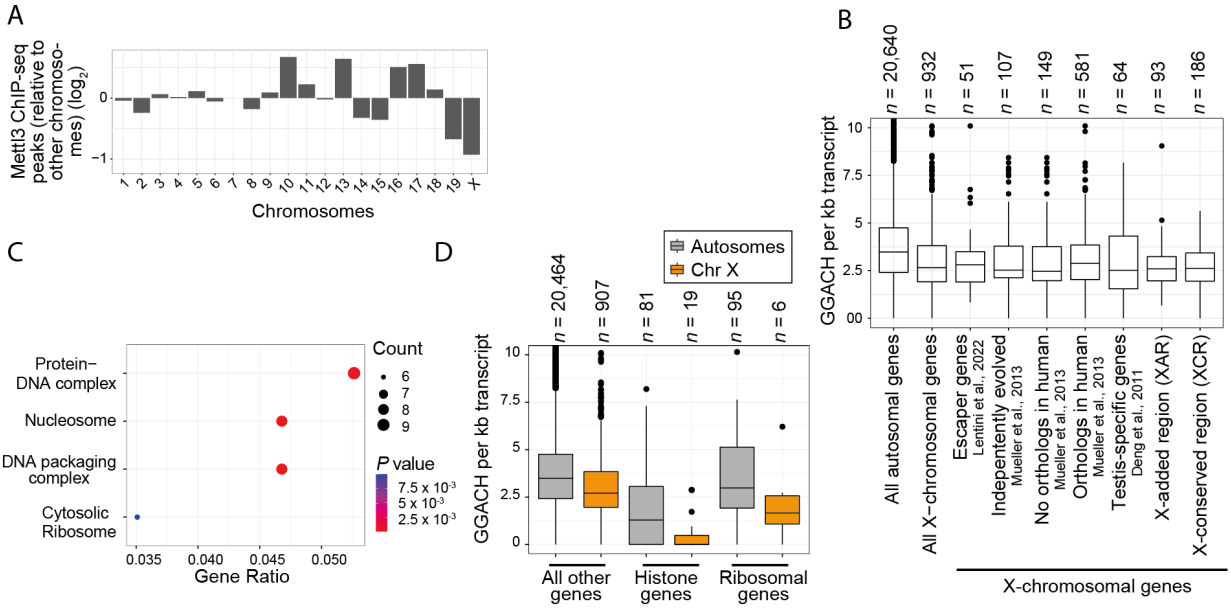
Extended Data Figure 6. X-chromosomal transcripts harbour less m^{6}A sites than autosomal transcripts in male mESC. A. Transcripts were stratified into 12 bins (#1-12) according to their expression in male mESC (transcripts per million [TPM, $\log_{10}$], see Methods). x-axis depicts boundaries between bins (in TPM). Bin number (#) and number of transcripts therein are given below and above each bar, respectively. Bins #3-8 that were used for quantifications of m^{6}A sites per transcripts are highlighted in black. **B.** Quantification of m^{6}A for each transcript in the

different expression bins of autosomal (grey) and X-chromosomal (orange) transcripts. Boxes represent quartiles, centre lines denote medians, and whiskers extend to most extreme values within 1.5x interquartile range. **C.** Quantification of m^6A sites per transcript for all mouse chromosomes. Data points indicate mean number of m^6A sites per transcript and 95% confidence interval (left y-axis) in each expression bin (x-axis, bins as defined in ***A***.) for all chromosomes (chromosome name and total number of expressed transcripts given above). Grey bars indicate the percentage of transcripts in each expression bin (right y-axis) relative to all expressed transcripts on the chromosome. Absolute numbers of transcripts in each bin are given above the bars. Only genes with a mean TPM > 1 over all samples were considered. **D.** Fold change (log_2, grey dots) in mean m^6A sites per transcripts for expression bins #3-8 (*n* of mean of expression bins = 6) on an individual chromosome over the mean m^6A sites per transcripts across all chromosomes. Red dots indicate mean fold change of the six bins on the given chromosome. Boxes as in ***B***. **E.** Same as ***D***. using only m^6A sites in a fixed window around stop codons (-50 nt to +150 nt) to exclude confounding effects of transcript length differences. Boxes as in ***B***. **F.** Same as ***C***. after randomly subsampling *n* = 30 genes from expression bins #3-5 to exclude potential biases from different numbers of transcripts in the expression bins for each chromosome. Shown is the distribution of mean m^6A sites per transcript for each chromosome from 100 repeats of subsampling. Boxes as in ***B***.

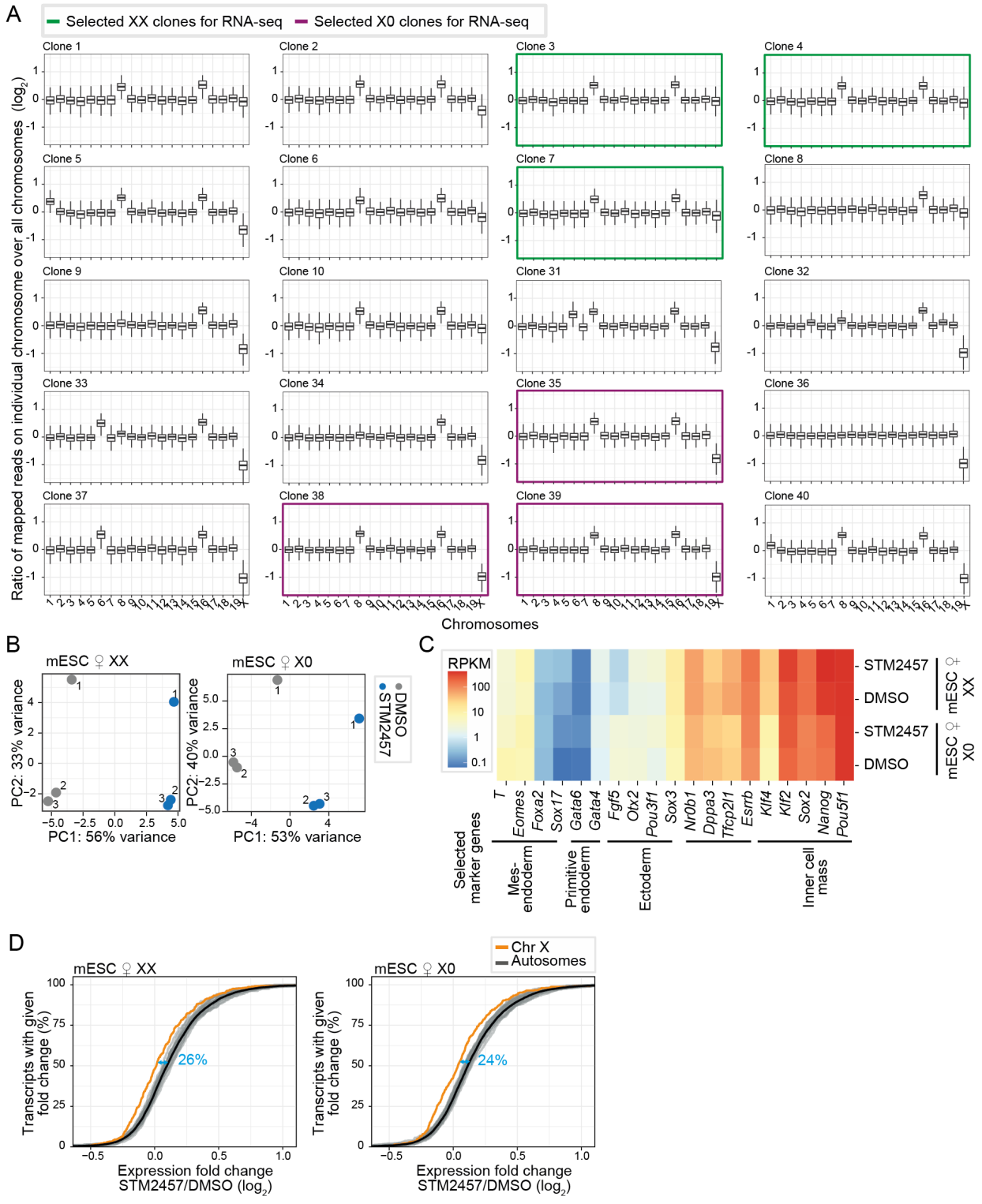


Extended Data Figure 7. The number of GGACH motifs and their methylation level are reduced on X-chromosomal transcripts compared to autosomal transcripts. A. X-chromosomal transcripts harbour fewer GGACH motifs than autosomal transcripts. Distribution of GGACH (H = [A|C|U]) per kilobase (kb) transcript sequence for individual chromosomes (corresponding to **Fig. 4A**). Boxes represent quartiles, centre lines denote medians, and whiskers extend to most extreme values within 1.5x interquartile range. **B.** Distribution of m^{6}A sites from mESC miCLIP2 data across different DRACH motifs. Barplot shows the number of m^{6}A sites for a given type of DRACH motif in mESC. The five most often methylated ("strong") and least often methylated ("weak") DRACH motifs are labelled below. **C.** Autosomal transcripts harbour more frequently methylated DRACH motifs in CDS and 3' UTR. Quantification of strong DRACH motifs in different transcript regions (normalised to region length) of autosomal (grey) and X-chromosomal transcripts (orange) in mouse. CDS *n* of annotations = 16,631, 3' UTR *n* of annotations = 16,484 and 5' UTR *n* of annotations = 16,490. Boxes as in ***A***. **D.** Autosomal transcripts harbour similar numbers of the least methylated DRACH motifs ("weak") in CDS and 3' UTR. Quantification of the five least methylated DRACH motifs as in *(**C.**)*. CDS *n* of annotations = 16,631, 3' UTR *n* of annotations = 16,484 and 5' UTR *n* of annotations = 16,490. Boxes as in ***A***. **E-G.** The methylation level of GGACH motifs in male mESC, i.e., the percentage of GGACH motifs that are methylated, is slightly reduced in X-chromosomal transcripts (***F.***), compared to transcripts across all chromosomes (***E.***) or from chromosome 11 (***G.***). To take into account only GGACH motifs in transcript regions with sufficient expression, GGACH motifs in transcripts were stratified into

bins by the local miCLIP2 read coverage (see Methods) and overlayed with m6Aboost-predicted m^{6}A sites from the same data. Dashed red line indicates local linear regression to estimate the methylation level (shown in **Fig. 4B**), i.e., the point at which the slope drops below 0.01. Dashed grey lines in ***F*** and ***G*** show estimated GGACH methylation level for transcripts across all chromosomes (***E.***) for comparison.



Extended Data Figure 8. The number of GGACH motifs is reduced on transcripts encoding histones and ribosomal proteins. A. The X chromosome harbours fewer Mettl3 ChIP-seq peaks. The number of published ChIP-seq peaks (normalised by chromosome length) per chromosome relative to peaks on all other chromosomes (log_2). **B.** Different gene sets on the X-chromosome are similarly depleted in GGACH motifs. Quantification of GGACH motifs of all autosomal or X-chromosomal genes is compared to the following gene sets: escaper genes, independently evolved genes, genes with or without orthologs on the human X chromosome, testis-specific genes or genes residing in the X-added region (XAR) and X-conserved region (XCR). Numbers of genes are given in the figure (*n*). Boxes represent quartiles, centre lines denote medians, and whiskers extend to most extreme values within 1.5x interquartile range. **C.** X-chromosomal genes with low GGACH motif numbers are associated with DNA packaging or the cytosolic ribosome. Gene ontology (GO) enrichment analysis of the 200 genes with the lowest density of GGACH motifs on the X chromosome. *P* values were calculated by overrepresentation analysis (see Methods). **D.** Histone and ribosomal protein-encoding genes on the X chromosome are depleted in GGACH motifs. Quantification of GGACH motifs for histone-encoding and ribosomal protein-encoding genes on autosomes or on the X chromosome. Numbers of genes are given in the figure (*n*). Boxes as in ***B***.



Extended Data Figure 9. X-chromosomal and autosomal transcripts differ in their response to m^{6}A depletion in both XX or X0 clones of female mESC. A. The majority of clones lost one copy of the X chromosome (X0). 20 single colonies of female mESC were picked and cultured under standard conditions until confluency was reached. To determine chromosome copy number, DNA-seq reads were counted into 100 kb bins along the chromosome and divided by the median mapped reads of all bins along the genome. Shown is the distribution of the

resulting ratios for the bins on each chromosome. Six clones that were selected for RNA-seq in control and m^{6}A-depleted (STM2457, 9 h) condition are highlighted in green. Boxes represent quartiles, centre lines denote 50th percentiles (medians), and whiskers extend to most extreme values within 1.5x interquartile range. **B.** Principal component analysis of RNA-seq replicates from female X0 (left) and XX (right mESC clones under m^{6}A-depleted (STM2457, 9 h) and control conditions. Analysis based on numbers of reads for the 500 genes with highest variance across all samples. **C.** Expression levels (RNA-seq) of marker genes confirm the pluripotent state of the female XX and X0 mESC under m^{6}A-depleted (STM2457, 9 h) and control conditions. Expression is shown as RPKM (mean over replicates, $\log_{10}$). **D.** X-chromosomal transcripts are less upregulated than autosomal transcripts upon m^{6}A depletion in female X0 and XX mESC (*P* value = 3.51e-11 [mESC X0], *P* value = 1.64e-12 [mESC XX], two-sided Wilcoxon rank-sum test). Cumulative fraction of transcripts (RPKM > 1) on individual autosomes (grey) and the X chromosome (orange) that show a given expression fold change ($\log_2$, RNA-seq) upon m^{6}A depletion (STM2457, 9 h). Mean expression changes for all autosomes are shown as black line. Effect sizes (blue) shown the shift in medians, expressed as percent of the average IQR of autosomal and X-chromosomal transcripts (see Methods).

3. Discussion

3.1 Improved library preparation using iCLIP2

To date, the epitranscriptome comprises more than 170 chemical modifications (http://modomics.genesilico.pl/sequences/, http://mods.rna.albany.edu). The importance of epitranscriptomic marks have been demonstrated in a plethora of studies, describing the role of different modifications in cellular and biological processes in all branches of life and beyond. However, the biological functions of most RNA modifications remain enigmatic. In addition to the importance of characterizing RMPs, the precise localization of modified nucleotides is essential to characterize the functional consequences.

In order to characterize RMP-RNA interactions, CLIP-seq techniques are commonly used to identify RMP target transcripts. Although CLIP-seq and its variations can provide high quality transcriptome-wide datasets, library preparation efficiency can be a limiting factor (Buchbender *et al.,* 2020). Limited input material, e.g. small cell number or proteins in low abundance can lead to low complexity libraries with high PCR duplicates. Hence, we overcame these limitations by improving the iCLIP protocol. Here, I contributed to the development of iCLIP2 to significantly enhance transcriptome-wide mapping of RNA-protein interactions (Buchbender *et al.,* 2020). We combined the best features of different CLIP-seq-based techniques to improve mapping of RBP binding sites. To improve library preparation, we took advantage of features of iCLIP, irCLIP and eCLIP and provide a sensitive and robust method for transcriptome-wide RBP-binding site mapping (Buchbender *et al.,* 2020). The most important changes comprise two separately ligated adapters, bead-based RNA clean-up strategies, pre-PCR amplification steps and optimized size-selection (Buchbender *et al.,* 2020). Compared to the original iCLIP protocol, iCLIP2 makes iCLIP experiments more sensitive and robust while minimising costs and time of experiment.

Taken together, using an iCLIP-based approach to map RMP-RNA interactions in a transcriptome-wide manner paved the way to broaden our knowledge about RNA targets of RMPs. Therefore, developing an iCLIP-based approach to map RNA modifications in single-nucleotide resolution provided an important tool for RNA modification research. Nevertheless, the inevitably inherited limitations introduced by the iCLIP-based technique are commonly acknowledged in the field as elaborated before. On top, limited antibody specificity of the m^6A antibody introduces high background-noise. Therefore, this study combined our recently developed iCLIP2 protocol with the miCLIP approach to overcome these limitations and to provide a robust and sensitive method for precise m^6A-site mapping.

3.1.1 Enhanced method for transcriptome-wide m^{6}A detection - miCLIP2

One critical disadvantage of the original miCLIP and m^{6}A-CLIP protocols is the high-required amount of input material. This made the transcriptome-wide m^{6}A detection exclusionary for rare input materials such as clinical samples or *in vivo* tissues with restricted material. Due to the improvements in library preparation of our recently developed iCLIP2 protocol, we reasoned to combine the original miCLIP protocol with iCLIP2 to reduce the amount of input material (Buchbender *et al.,* 2020; Körtel *et al.,* 2021). Thus, the improvements of the miCLIP2 protocol comprise the two separately ligated adapters, bead-based clean-up strategies, a pre-PCR amplification step and an optimized size selection procedure (Körtel *et al.,* 2021). Furthermore, several steps have been optimized. miCLIP is an *in vitro* CLIP approach and hence, purified RNA is used and incubated with an m^{6}A-specific antibody and rRNA and tRNA are heavily m^{6}A-modified. Thus, to reduce the excessive amount of sequenced rRNA and tRNA, poly(A) selection is critical. We have optimized poly(A) selection and perform two consecutive rounds. Therefore, we find only moderate rRNA and tRNA levels in miCLIP2 data. Moreover, we optimized RNA fragmentation and UV-crosslinking steps, allowing to routinely perform miCLIP2 experiments without pre-optimization steps. However, concentration of fragmentation buffer has to be optimized for each separate isolated RNA batch. Due to remaining EDTA after final washing steps during poly(A) selection, the RNA fragmentation buffer concentration requirements might vary. This could be further optimized by omitting EDTA in the final washing step during poly(A) selection.

Our improvements reduced the time of experiment from around seven days down to four days of immediate experiment. Importantly, compared to previous protocols, our standard set-up requires only 1 µg per replicate. Titration experiments further showed that our improvements allow to perform miCLIP2 for limited input material. We titrated the amount of required input material down to 50 ng per replicate. Performing miCLIP2 experiments in our routine 1 µg RNA input per replicate set-up in different mouse and human cell lines, we are able to routinely obtain more than 30 million unique reads with low duplication rates. Our titration experiment showed that we can we can reduce the input material while still obtaining 2-50 million truncation events and the mapped m^{6}A sites. We were able to show that miCLIP2 signals were still reproducible at the single-nucleotide level. Although the sensitivity progressively decreased with less input material, the precision was not minimised. At all tested concentrations, the identified m^{6}A sites largely overlapped and sites were consistently enriched at DRACH motifs and around stop codons. The moderate duplication rates further indicate that the sequencing libraries were not sequenced to full saturation, meaning that many more m^{6}A sites potentially could be identified.

Since the original Linder *et al.,* protocol 20 µg of input RNA per replicate was required, we were able to significantly reduce the amount of input material. This provides an m^{6}A-mapping method for rare and low input material. In comparison, recently developed antibody-free methods also allowed mapping of m^{6}A in low input material. For instance, while MAZTER-seq requires 100 ng, DART-seq uses 10 ng of input material per replicate (Garcia-Campos *et al.,*

2019; Meyer, 2019). Nevertheless, also these methods come with certain limitations as elaborated in chapter 1.4.4. While MAZTER-seq only allows mapping of 16% - 25% of all m^{6}A sites, DART-seq induces sequence preferences and possibly induce high false-positive and false-negative rates (Zhang *et al.,* 2019; Garcia-Campos *et al.,* 2019; Capitanchik *et al.,* 2020). Hence, the optimised miCLIP2 allows mapping of m^{6}A in rare input material without inheriting limitations that arise from antibody-independent methods.
While tackling the issue of low input material, miCLIP2 is still unable to provide stoichiometric information. Newly developed m^{6}A-site mapping using pore-based direct RNA sequencing methods, could provide stoichiometric, as well as positional information for multiple modifications at once as reviewed in Capitanchik *et al.,* 2020. However, high input requirements and the required no-methylation control condition are drawbacks that still need to be overcome. Moreover, the currently low throughput and high base-error rates limit the usage of direct-sequencing based m^{6}A-mapping methods (Zaccara *et al.,* 2019; Capitanchik *et al.,* 2020). Therefore, although miCLIP2 data provide no stoichiometric information, the high precision even in low input material generates a deep and accurate detection method for m^{6}A.

3.1.2 m6Aboost allows reliable prediction of m^{6}A outside of DRACH motifs without the need of a Mettl3 *KO condition*

3.1.2.1 Differential methylation analysis detects Mettl3-dependent m^{6}A sites

Using our newly established miCLIP2 protocol, we performed experiments in mESC cells. Surprisingly, we find that most peaks reside on uridine rather than on adenine. This reflects the strong UV-crosslinking bias that is broadly observed using UV-crosslinking in CLIP-seq techniques (Sugimoto *et al.,* 2012). Furthermore, only 25% of peaks on adenosine were located within DRACH motifs. In accordance with previous studies, this reflects the limited antibody-specificity of the m^{6}A antibody (Helm et al., 2019). Hence, the crosslinking bias and limited antibody-specificity raised the need for an m^{6}A-depleted condition. Here, comparison of an m^{6}A-depleted to a control condition could identify sites that are changed. Since Mettl3 is the catalytic active subunit of the methyltransferase complex and the main m^{6}A-writer on mRNA, KO of *Mettl3* hypothetically should result in the loss of the majority of miCLIP2 peaks. In order to test whether we find Mettl3-responsive m^{6}A sites using our miCLIP2 protocol, we performed miCLIP2 in mESC wild-type (WT) and *Mettl3* KO cell lines. We find that the majority of miCLIP2 peaks are not responsive to *Mettl3* KO. We found an enrichment for miCLIP2 peaks around the transcription start site (TSS), which were also not responsive to *Mettl3* KO, indicating the cross-reactivity with the m^{6}Am modification. Nevertheless, by performing differential methylation analysis using the WT and the *Mettl3* KO dataset, we identified 11,707 genuine *Mettl3*-responsive m^{6}A sites. In accordance with previous observations, these sites

were found to be located around the stop codon, in the 3'UTR and in the DRACH consensus motif (Linder *et al.,* 2015, Meyer *et al.,* 2012). This further provides proof that miCLIP2 can identify genuine m^{6}A sites. Interestingly, we were also able to find m^{6}A sites outside of the DRACH motif. Using an orthogonal method, we validated the *Mettl3*-responsive non-DRACH m^{6}A sites. Most of the non-DRACH m^{6}A sites contained the AC dinucleotide. However, some m^{6}A sites diverged and were found in an AT context, for instance. In a previous study and using Nanopore technology, non-DRACH have been reported before. Here, we provide the first validation using an orthogonal method. Importantly, previous computational DRACH filtering steps will not allow to capture non-DRACH m^{6}A sites.

Nevertheless, using differential methylation analysis requires the availability of a Mettl3-depleted condition. It is not feasible to generate a *Mettl3* KO for most cell lines since it may cause lethality. Considering the unavailability of *Mettl3* KO especially for clinical samples or *in vivo* disease models, further computational measures were required. Moreover, a prominent function of Mettl3-dependent m^{6}A methylation is promoting RNA degradation. Therefore, massive shifts in gene expression in KO condition hinder computational analysis. Even if a *Mettl3* KO is achieved in certain cell lines, the loss of *Mettl3* may lead to unwanted compensatory effects or compensatory mechanisms arising during culturing cells for several generations. Hence, circumventing the need for a *Mettl3* KO condition is desired for various experimental setups. Furthermore, the high amount of detected non-specific background signal and the m^{6}Am cross-reactivity of the m^{6}A antibody, further demonstrated that more precise computational measures are required to identify genuine m^{6}A sites.

3.1.2.2 Machine learning allows reliable m^{6}A prediction without the need of Mettl3 *KO*

To tackle the drawback of a required Mettl3-depleted condition, we built m6Aboost, an adaptive boosting (adaboost)-based machine learning classifier. This allows m^{6}A identification in any miCLIP2 dataset without the need of an accompanying *Mettl3* KO. Testing three different predictive models, all tested models reached high predictive accuracy. However, the adaboost-based model showed best performance among the tested algorithms. Hence, we chose the adaboost-based predictor and subsequently named it m6Aboost. We trained and validated the machine learning model using our m^{6}A sites and a non-m^{6}A testing set identified by differential methylation analysis. Applying m6Aboost, we identified 25,456 m^{6}A sites and validated individual sites using an orthogonal method.

While training our predictive model in mESC, we wondered whether our model could be applied across species. Thus, we produced miCLIP2 data from human HEK293T cells and tested our model. The m6Aboost algorithm was able to predict 36,556 m^{6}A sites in 7,552 genes, which greatly overlapped with iCLIP data from m^{6}A reader protein YTHDF2. Using the orthogonal antibody-free method SELECT, we validated several individual m^{6}A sites. Hence, m6Aboost allows m^{6}A site prediction across species. However, whether m6Aboost allows m^{6}A prediction from miCLIP2 data from species, which are evolutionary further away from

mammals, remains to be tested. For instance, in Drosophila m^{6}A resides rather in the 5'UTR and occurs in an RRACH consensus motif. Since training of our model was performed in a mammalian species, different m^{6}A distribution patterns present in different species could inhibit the performance of m6Aboost. Training our model with miCLIP2 data from mouse cells allowed learning about specific distribution patterns from the respective organism and hence, m^{6}A prediction in another species may require re-training of the model.

Previous studies used *in silico* prediction models to identify m^{6}A sites. Several approaches use either support vector machine (SVM) or random forest models (Chen *et al.,* 2017; Xiang *et al.,* 2016; Chen (A) *et al.,* 2019; Zhou *et al.,* 2016). Two critical factors are the key for *in silico* predictions. First, the quality and reliability of the training data set and secondly, the models ability to identify critical features (Capitanchik *et al.,* 2020). Therefore, previous miCLIP datasets may have restricted the quality of *in silico* predictions due to low-complexity libraries with high PCR duplication rates. Moreover, some models include DRACH filtering steps, which prohibits m^{6}A site identification at non-DRACH sites (Xiang *et al.,* 2016; Chen (A) *et al.,* 2019; Zhou *et al.,* 2016). Therefore, we developed an improved machine learning model using our generated miCLIP2 data and without including DRACH filtering steps to tackle this issue.

Since our initial test uses Mettl3-dependent m^{6}A sites, the machine learning model does not allow to identify m^{6}A sites deposited by different methyltransferases like Mettl5 or Mettl16 (Pendleton *et al.,* 2017; Shima *et al.,* 2017; Warda *et al.,* 2017). Most of the m^{6}A sites introduced by methyltransferases other than Mettl3 rather reside in different types of ncRNAs, hence, the poly(A) selection step during the miCLIP2 protocol nonetheless hinders identification of these sites.

Taken together, miCLIP2 in combination with the m6Aboost predictor allows deep and accurate m^{6}A site identification in single-nucleotide resolution. Thus, generation of miCLIP2 datasets and m^{6}A site identification using m6Aboost, helps to further unravel the biological function of m^{6}A.

3.1.3 Efficient splicing of retained intron upon m^{6}A depletion

The m^{6}A RNA modification was found to be involved in every aspect of RNA metabolism. A crucial step during mRNA maturation is the excision of intronic sequences. The process called "pre-mRNA splicing" removes intronic sequences and joins the remaining exonic sequences to form the mature mRNA. The so-called process of alternative splicing allows the exons to be combined in different combinations. Here, exons can be skipped, introns can be retained or alternative 3' or 5' splice sites can be used. Several studies proposed a role for m^{6}A in splicing. In order to find further support for a role in splicing, we analysed our miCLIP2 data. We find that most identified m^{6}A sites reside in exons. However, this is due to the poly(A) selection step included to our protocol, since most unprocessed and immature mRNAs are removed and thus, it is not ruled out that m^{6}A sites may be more frequent in intronic sequences. Using our miCLIP2 and m6Aboost approach, we generated several datasets along with RNA-seq datasets. Despite finding the majority of m^{6}A sites in exons, we also identify several m^{6}A sites

in retained introns close to the 5' splice site. In the absence of Mettl3, we found 401 significantly changes IR events in RNA-seq data. Interestingly, in *Mettl3* KO we find significantly changed introns with reduced coverage, which is indicating an increase in the efficiency of splicing. Therefore, the presence of m^{6}A may promote intron retention.

Previous work provided conflicting evidence for the exact impact that m^{6}A inflicts in regards to splicing. In order to study the role of m^{6}A in splicing, the distribution of m^{6}A was mapped on nascent RNA or chromatin-associated RNA (CA-RNA). Analysing m^{6}A on CA-RNA, Ke *et al.,* found that around 93% of m^{6}A sites are found in exons (Ke *et al.,* 2017; Yang *et al.,* 2022). Using TNT-seq (Transient N-6-methyladenosine transcriptome sequencing) to identify m^{6}A sites in newly synthesized RNAs, Louloupi *et al*., found that most m^{6}A sites reside in introns (Louloupi *et al.,* 2018; Yang *et al.,* 2022). Another recent publication proposed that m^{6}A deposition is exclusion based. Every possible DRACH consensus motif is modified, but m^{6}A is excluded in the proximity of splice sites (Uzonyi *et al.,* 2023). This was proposed to be mediated by the exon junction complex (Uzonyi *et al.,* 2023, Yang *et al.,* 2022). To resolve the contradiction on whether m^{6}A rather reside in exonic or intronic sequences, further research is necessary.

Nevertheless, in order to unravel the impact of m^{6}A on splicing, several studies provided evidence for either promotion or inhibition of intron retention (IR) upon m^{6}A modification. Several studies showed an increase in IR events in *Mettl3* KO in mESC or in *Ime4* (Inducer of Meiosis; Mettl3 orthologue in *Drosophila*) mutants in *Drosophila melanogaster* (Geula *et al.,* 2015; Lence *et al.,* 2016; Haussmann *et al.,* 2016; Gehring *et al.,* 2021). On the contrary, another study proposed that the double-stranded RNA binding protein TARBP2 (Trans-Activation-Responsive RNA-Binding Protein 2) recruits the m^{6}A methyltransferase machinery to guide intron methylation and subsequently induces intron retention (Fish *et al.,* 2019). Here, the authors showed that upon m^{6}A methylation, the splicing-factor recruitment is inhibited and intron excision is impaired (Fish *et al.,* 2019). Therefore, introns are retained when m^{6}A is present. Our data supports the finding that intron inclusion is promoted upon m^{6}A modification and is in contrast to the study from the Schwartz group (Uzonyi *et al.,* 2023). To further dissect the discrepancy in current literature, miCLIP2 on nascent RNA could shed light on this issue.

3.2 The link between dosage compensation, RNA stability and m^{6}A

The precise balancing of gene expression throughout the genome is crucial for the fitness of a cell or a given organism. The local and global disturbance of gene dosage can cause detrimental consequences and potentially leads to lethality. As elaborated above, loss of larger chromosome fractions or variation of copy numbers of chromosomes (aneuploidy) lead to reduced fitness and can cause embryonic lethality (Torres *et al.,* 2008; Tang *et al.,* 2013; Epstein, 1990). Therefore, the imbalance introduced by sex chromosomes need compensatory

mechanisms. In brief, during the process of sex chromosome evolution in many species, a homogametic and a heterogametic sex evolved. The chromosome that is exclusively present in the heterogametic sex (Y chromosome in males) is highly degenerated, which causes an imbalance between males and females (Charlesworth, 1996; Charlesworth *et al.,* 2005; Rice *et al.,* 1996; Heard and Disteche 2006). Therefore, the heterogametic sex is in need for mechanisms to account for the gene dosage disturbance. Upon X-chromosome inactivation of one X chromosome in females, the gene dosage is balanced between both sexes. However, the gene dosage between two active copies of autosomes and one single remaining X chromosome in both sexes is imbalanced and further mechanisms are required to accomplish X:AA dosage compensation. Since Susumu Ohno hypothesized X-upregulation to accomplish X-to-autosome dosage compensation, many studies found evidence to support the hypothesis (Ohno, 1966; Nguyen and Disteche, 2006; Lin *et al.,* 2007; Deng *et al.,* 2011; Kharchenko *et al.,* 2011; Lin *et al.,* 2011). However, as elaborated in chapter 1.5.3, Ohno's hypothesis has also been challenged (Xiong *et al.,* 2010). Nevertheless, several dosage compensation mechanisms had been identified. The identified mechanisms range from transcriptional to post-transcriptional regulation (Yildirim *et al.,* 2011; Talon *et al.,* 2021; Deng *et al.,* 2011; Yildirim *et al.,* 2011; Yin *et al.,* 2009; Deng *et al.,* 2013; Faucillion and Larsson, 2015; Wang (B) *et al.,* 2020; Larrson *et al.,* 2019). Interestingly, several studies provided evidence for increased RNA stability of X-chromosomal transcripts (Deng *et al.,* 2013; Faucillion and Larsson, 2015). Despite the potential contribution of the NMD pathway, it remains enigmatic how chromosomal origin is remembered. Interestingly, although epitranscriptomic marks have been described to play a crucial role in gene expression regulation, they have not previously been linked to X:AA dosage compensation mechanisms. However, epigenetic marks inherit the potential to bridge DNA context to the fate of an mRNA.

As the most abundant internal mRNA modification and the diverse functions of m^{6}A, this study sought to uncover a link between dosage compensation and the m^{6}A RNA modification. The most prominent function of m^{6}A is promoting RNA degradation. Hence, we sought to uncover the potential interplay between dosage compensation regulation by RNA stability mediated by m^{6}A.

3.2.1 RNA stability controlled by m^{6}A mediates dosage compensation

To uncover the potential interplay between epitranscriptomic gene expression regulation and dosage compensation achieved by RNA stability, we first confirmed higher RNA half-lives of X-chromosomal transcripts in mESC in a published dataset (Herzog *et al.,* 2017). Moreover, we confirmed higher RNA half-lives by performing SLAM-seq experiments. Therefore, we provided further evidence for differences in RNA stability as a dosage compensation mechanism. In order to link m^{6}A to dosage compensation, we used a small molecule inhibitor for Mettl3 to induce acute depletion and rapid loss of m^{6}A (Yankova *et al.,* 2021). By using the predicted m^{6}A sites derived from mESC miCLIP2 and m6Aboost data, we found an increase in

half-life upon Mettl3 inhibition for transcripts carrying m^{6}A. Dividing transcripts according to their chromosomal origin, we found that acute Mettl3 depletion led to a significant increase in half-lives of autosomal transcripts, while X-chromosomal transcripts largely remained unchanged. Using published half-life data comparing mESC WT and *Mettl3* KO condition, we observed the same effect of differential RNA stability between X-chromosomal and autosomal transcripts (Ke *et al.,* 2017).
Here, we showed that RNA stability is mediated by m^{6}A in a chromosome-dependent manner. Nevertheless, a potential interplay of the NMD and the m^{6}A pathway could be possible. X-chromosomal transcripts were found to have the least hitchhiking of NMD targets (Yin *et al.,* 2009). This indicates a weaker NMD-mediated mRNA decay for X-chromosomal transcripts. Upon depletion of the NMD key factor *UPF1*, a skewed gene expression between X- and autosomal transcripts was found, underlining the role in X-to-autosome dosage compensation (Yin *et al.,* 2009). Interestingly, UPF1 was also found to associate with m^{6}A reader protein YTHDF2 (Boo *et al.,* 2022). The interaction between YTHDF2 and UPF1 was shown to induce rapid decay of m^{6}A-methylated transcripts (Boo *et al.,* 2022). Therefore, the destabilization of autosomal transcripts may be a result of a potential interplay between the m^{6}A and the NMD pathway which could further contribute to X-to-autosome dosage compensation.

3.2.2 Differential gene expression on different chromosomes upon acute m^{6}A loss

The differential changes in RNA stability between autosomes and the X chromosome directly link the RNA modification to dosage compensation. In order to strengthen our hypothesis that m^{6}A mediates dosage compensation by regulating RNA stability, we performed RNA-seq experiments upon Mettl3 inhibition. Since the most prominent function of m^{6}A is promoting degradation, the loss results in massive shifts in gene expression. Comparison of changes in expression of transcripts deriving from different chromosomes, showed a marked difference between X-chromosomal and autosomal transcripts. Thus, transcripts deriving from the X chromosome behave differently upon loss of m^{6}A. The overall observed upregulation of autosomal genes reflect the increase in RNA half-life observed by SLAM-seq. Therefore, autosomal transcripts are less degraded due to loss of m^{6}A. Further supporting the observed differential behaviour of X-chromosomal and autosomal transcripts, we performed a time-course experiment from 3 h – 12 h using the small inhibitor molecule for acute m^{6}A. The autosomal transcripts showed a clear separation from X-chromosomal transcripts after 6 h of treatment. This provides further evidence for a direct effect of m^{6}A in the destabilisation of transcripts. Thus, due to the usage of an acute depletion method, no secondary effects account for the observed differential response upon loss of m^{6}A. This could be further supported by usage of a degron tag system, providing another immediate loss of m^{6}A (Nishimura *et al.,* 2009). However, the small inhibitor molecule provides an elegant method,

which can be used across different cell line and species without prior laborious cell line generation.

All tested cell lines, including mESC, human primary fibroblasts and various commonly used human cell lines consistently exhibit the same differentially behaviour of X-chromosomal and autosomal transcripts upon m^{6}A loss. However, the rather mild effect seen in human HEK293T and C632 cells could be reasoned by severe aneuploidies, which are commonly observed in these cell lines. HEK293T cells have an unstable karyotype and the number of chromosomal aberrations may severely differ between different distributors and labs (Stepanenko *et al.,* 2015). C643 is a human thyroid cancer cell line. Cancer cells are typically characterized by unstable and complex karyotypes (Nicholson *et al.,* 2013). DNA-sequencing confirmed severe copy number variations for both cell lines. However, when compared to the effect in HEK293T and C643, RPE1 cells and human primary fibroblast show only a moderate effect as well. RPE1 cells are chromosomally stable with only distinct chromosomal duplications. Human primary fibroblast are cultured directly after taking cells from a donor and exhibit a normal karyotype due to small cell passage numbers. Therefore, the mild effect seen in human cells may only be partially a consequence of chromosomal aberration. Therefore, we might observe generally smaller effects in humans. Nevertheless, the consistently observable effect seen in different cell lines and across mammalian species implicates a conserved function of m^{6}A in dosage compensation in mammals.

To see whether the chromosomal differences in RNA stability and transcript abundance directly result in balanced gene expression, loss of m^{6}A should result in an imbalanced expression ratio. To investigate the relative contribution of m^{6}A, we calculated X:AA transcript expression ratios in pre- and absence of m^{6}A. According to Ohno's hypothesis, the X:AA expression ratios should be ~1.0 if dosage compensation is existent. Consequently, ratios should be ~0.5 when dosage compensation is fully disturbed. As described in chapter 1.5.3, the interpretation of RNA-seq data requires careful consideration since the X chromosome contains higher fractions of tissue-specific expressed genes. Thus, we excluded silent or lowly expressed genes. We demonstrate that all tested cell lines exhibit an expression ratio approximating ~1 in control conditions, implicating that dosage compensation is in place. This further provides evidence for Ohno's hypothesis holding true. HEK293T cells and C643 cells showed an expression ratio of ~1.4 and ~1.1. This may reflect the severe karyotype changes.

Upon m^{6}A-depleted condition, the overall balance is disturbed and the expression ratios are significantly reduced. However, the expression ratios do not fully reach 0.5. This indicates that the m^{6}A pathway might be an additional regulatory pathway, which adds an additional layer to RNA stability mediated dosage compensation.

Interestingly, in female mESC, where both X chromosomes are still active, the expression ratio is above 1. In undifferentiated female mESC, X-chromosome inactivation is not yet in place. Therefore, dosage compensation may not be required in the earliest developmental stages of an organism. Single-clone picking allowed culturing XX and X0 female mESC. During female mESC culturing, one X chromosome commonly gets lost. Consequently, expression ratios in X0 cells approximated down towards 1. This indicates that a single X chromosome is sufficient to reach an expression ratio of ~1 while two X chromosomes lead to an excess of X-

chromosomal transcripts, which is tolerated by cells in the earliest developmental stages. Previous work analysed public RNA sequencing data, where the authors calculated an X:AA ratio of 0.5 in human and 0.3 in mice (Xiong *et al.,* 2010). Here, genes that are silent or lowly expressed were not excluded. Hence, a subsequent study reanalysed the data and calculated X:AA expression ratios over >0.5 and approximating 1.0 (Deng *et al.,* 2011; Kharchenko *et al.,* 2011; Lin *et al.,* 2011). In accordance with our finding of X:AA ratios approximating 1.0, Ohno's hypothesis is further supported. Moreover, we find that depletion of m^{6}A disrupts global gene dosage balance.

3.2.3 X-chromosomal transcripts harbour less m^{6}A modifications

Higher stabilities and differential effects upon m^{6}A loss for X-chromosomal transcripts implicate a direct role for m^{6}A in RNA stability mediated dosage compensation. Thus, we used our miCLIP2 and m6Aboost derived transcriptome-wide m^{6}A maps to assess methylation numbers on transcripts deriving from different chromosomes. To account for different expression levels across all transcripts, we divided transcripts in expression bins according to their expression level. Along all tested expression bins, only the transcripts from the X chromosome harbour significantly lower m^{6}A modifications. We found that compared to autosomal transcripts, transcripts from the X chromosome are reduced in m^{6}A modifications by almost half. These results were independent from transcript length and numbers of transcripts encoded on different chromosomes. All autosomes showed similar m^{6}A modification numbers. We tested all generated miCLIP2 and m6Aboost m^{6}A maps that we have previously generated for mESC, mouse heart samples, mouse macrophages, human HEK293T and C643 cells. Consistent with our observation that X-chromosomal transcripts behave different upon loss of m^{6}A in different mammalian cell lines, all tested cells harbour lower m^{6}A levels on X-chromosomal transcripts. However, the extend of reduction was partially tissue and species dependent. For instance, human X-chromosomal transcripts generally showed a lower m^{6}A reduction compared to mouse X-chromosomal transcripts. This is in accordance with the finding that human cells showed a milder effect on differential expression changes between X-chromosomal and autosomal transcripts upon m^{6}A loss. Nevertheless, the consistent reduction of m^{6}A numbers on X-chromosomal transcripts further supports that m^{6}A-mediated dosage compensation is a conserved mechanism in mammals. To control that our findings are not biased by using our miCLIP2 detection method and m6Aboost, we used the recently published m6A-seq2 data and consistently observed a reduction of m^{6}A numbers on X-chromosomal transcripts (Dierks *et al.,* 2021). However, since m6A-seq2 is also an m^{6}A-antibody based method, an antibody-independent method could be used to further confirm that these findings are not method-biased.

As elaborated in chapter 3.1.2, m^{6}A site identification using miCLIP2 and m6Aboost partially depends on the underlying transcript abundance. Hence, transcriptome-wide mapping is hindered for lowly expressed genes and many more m^{6}A sites could potentially account for dosage compensation. Therefore, the vast majority of autosomal transcripts with low

expression harbour no m^{6}A sites, while the 74.5% of moderately expressed genes harbour at least one m^{6}A site. Recent studies propose that many more m^{6}A sites are present than previously anticipated. Moreover, detecting low stoichiometry m^{6}A sites may be hindered as well. As suggested in a recent finding, all potential consensus motif may get methylated and thus, the effect on m^{6}A-mediated dosage compensation is potentially higher than found during this study (Uzonyi *et al.,* 2023). Using pore-based direct sequencing methods or GLORI-seq, which found more m^{6}A sites than previously anticipated and detect m^{6}A sites with low stoichiometry, could shed light on this in future studies.

Given that m^{6}A controls RNA stability to mediate dosage compensation, it would be interesting to investigate the downstream degradation processes to see whether autosomal transcripts are more likely found to associate with the deadenylation degradation machinery. YTHDF proteins were found to have a unified function in promoting degradation of methylated transcripts (Zaccara *et al.,* 2020). Moreover, all three DF proteins were found to localize to P bodies and to associate with the deadenylation complex CCR4/NOT (Zaccara *et al.,* 2020; Du *et al.,* 2016). Moreover, it was shown that m^{6}A-modified transcripts are degraded by CCR4/NOT complex-mediated deadenylation (Du *et al.,* 2016). DF1-3 iCLIP experiments could be used to determine whether the DF proteins differentially bind to autosomal and X-chromosomal transcripts. Differential DF1-3 binding could indicate whether methylated transcripts are potentially differentially targeted for subsequent degradation processes. Furthermore, since the CCR4/NOT deadenylation complex localises to P bodies, proximity labelling experiments could be used to investigate differential mRNA subcellular localisation between autosomal and X-chromosomal transcripts. Here, proximity-CLIP or CAP-seq (chromophore-assisted proximity labelling and sequencing) experiments could be used (Benhalevy *et al.,* 2018; Wang *et al.,* 2019). Proximity-CLIP makes use of the APEX2 (Apurinic/apyrimidinic endodeoxyribonuclease 2) protein, which induces biotinylation of proteins and RNA in close proximity in subcellular compartments (Benhalevy *et al.,* 2018). The CAP-seq technique uses miniSOG, which mediates the proximity-dependent photo-oxidation of the guanine RNA nucleobase in a proximity dependent-manner upon blue light excitation (Wang *et al.,* 2019; Weissinger *et al.,* 2021). APEX2 and miniSOG can be targeted to specific subcellular compartments. Both approaches could provide insights into differences of subcellular localisation of autosomal and X-chromosomal transcripts due to m^{6}A methylations. Acute m^{6}A depletion could further unravel how autosomal and X-chromosomal transcripts are differentially affected in their changes of subcellular localization upon loss of m^{6}A modifications.

3.2.4 Reduced m^{6}A levels on X-chromosomal transcripts internally hardcoded by fewer GGACH motifs

Since the observation of higher RNA half-lives of X-chromosomal transcripts was made, the question how this is achieved needed further consideration. In this study, we found that

reduced m^{6}A numbers on X-chromosomal transcripts control dosage compensation by reducing RNA stability of autosomal transcripts. How the differential methylation of transcripts from different chromosomes is achieved and chromosomal origin is remembered remained to be answered. Since m^{6}A occurs in a DRACH consensus sequence-dependent manner in mammals, we assessed the sequence composition of all transcripts from all chromosomes. The GGACH motif is the most frequently methylated DRACH motif and thus, we counted GGACH motif throughout the transcriptome. Interestingly, the 3'UTR and CDS harbour significantly less GGACH motifs compared to autosomal transcripts. Throughout literature, m^{6}A is reported to accumulate in the CDS and 3'UTR while it is less frequent in the 5'UTR (Linder *et al.,* 2015, Meyer *et al.,* 2012). Consistently, no difference in GGACH motifs was found in the 5'UTR. The reduction of GGACH motifs is around half, which is consistent with the finding that X-chromosomal transcripts are reduced to around half of the autosomal m^{6}A numbers. To further support this finding, we compared strongly and weakly methylated DRACH motifs. We found that the strong DRACH motifs are reduced on X-chromosomal transcripts, while weak motifs showed no difference between autosomal and X-chromosomal transcripts. This further supports the notion that m^{6}A levels are reduced on X-chromosomal transcripts as a result of reduced strong DRACH motifs and provides evidence for an internally hardcoded mechanism. Reanalysing Mettl3 CLIP-seq data further revealed reduced occupancy on X-chromosomal chromosomes. Moreover, we found that the fraction of methylated DRACH motifs is lower on X-chromosomal transcripts. Next to the internally hardcoded mechanism, different mechanisms could be in place to ensure lower m^{6}A levels on X-chromosomal transcripts. For instance, since m^{6}A is deposited co-transcriptionally and guided by histone mark H3K36me3, the X chromosome could be less occupied by both, Mettl3 and the histone mark (Huang *et al.,* 2019). Another possible explanation could be a Mettl3 repelling mechanism. To test this, X-chromosomal genes could be translocated to autosomes. If more m^{6}A modifications appear on transcripts from X-chromosomal genes that are located on autosomes would indicate a repelling mechanism and could explain the lower fraction of m^{6}A methylated DRACH motifs. Moreover, ChIP-seq experiments could provide further insights into histone mark-guided m^{6}A methylations or less Mettl3 occupied X chromosomes. Moreover, since the Schwartz group recently proposed that all DRACH motifs are methylated, also an eraser protein-guided mechanism could be possible (Uzonyi *et al.,* 2023). In a case where all DRACH motifs are methylated, a higher demethylation efficiency for X-chromosomal transcripts could be a possible explanation. Therefore, ChIP and CLIP experiments for the eraser proteins FTO and ALKBH5 could give further insights into the possible mechanism.

Mammalian sex-determining chromosomes evolved from ancestral autosomes (Lahn and Page, 1999; Livernois and Waters, 2012). Before the evolution of sex chromosomes, numerous genes located rather to autosomes than to the respective sex chromosome. Therefore, we tested whether reduced GGACH motifs evolved in a sex-chromosome specific manner. Here, chicken displays an informative outgroup where genes from mammalian sex chromosomes primarily locate to autosomes 1 and 4. We found that the orthologs to X-chromosomal mouse genes show no reduction in GGACH motifs. This suggests that reduced m^{6}A modification specifically evolved in a sex chromosome-dependent manner to mediate dosage

compensation. The sex chromosomes in chicken are denoted as Z and W. Males possess two Z chromosomes while females possess one Z and one W chromosome. Dosage compensation was shown to generally be less effective in chicken (Itoh *et al.,* 2007; Ellegren *et al.,* 2007). Thus, how reduction of GGACH motifs was evolutionary achieved and tracking the m^{6}A levels throughout evolution an exciting question for future research.

Interestingly, dosage compensation was reported to be more important for certain sets of genes. Housekeeping genes were reported to be dosage sensitive and rely on the full two-fold upregulation compared to tissue-specific expressed genes as well as recently evolved X-chromosomal genes (Deng *et al.,* 2011; Pessia *et al.,* 2012; Mueller *et al.,* 2013). While some genes on the X chromosome still contain m^{6}A sites, many other may require full dosage compensation and strictly rely on low GGACH motif resulting in low m^{6}A levels. To investigate this, we performed gene ontology (GO) analysis and found that specifically ribosomal genes and histones contain almost no GGACH motif and m^{6}A modifications. Ribosomal genes encode for ribosomal proteins and are located on both, autosomes and the X chromosome. We found that the GGACH motifs on ribosomal genes located on the X chromosome significantly differed from their autosomal counterparts. As previously suggested, especially for multisubunit protein complexes the correct stoichiometry of proteins and therefore their dosage compensation is crucial (Papp *et al.,* 2003). Underlining this, the *Minute* phenotype in *Drosophila* is caused by the haploinsufficiency of ribosomal genes, demonstrating the importance of correct stoichiometry and proteostasis of these genes (Marygold *et al.,* 2007; Amirbeigiarab *et al.,* 2019). Moreover, the importance of correct ribosomal protein stoichiometry has been reported to be crucial for correct mouse brain development, demonstrating the importance of correct dosage compensation regulated by m^{6}A (Amirbeigiarab *et al.,* 2019). Analysing GGACH motifs and m^{6}A levels in *Drosophila* ribosomal genes could shed light on the evolutionary conservation of reduced m^{6}A levels on these genes. Taken together, the observation that m^{6}A is less frequent on X-chromosomal transcripts as a result of an internally hardcoded mechanism. Until now, it remained enigmatic how chromosomal origin was remembered and how transcripts from different chromosomes can achieve differential RNA stability. We show that less m^{6}A is a direct result of less frequent GGACH motif on X-chromosomal transcripts.

3.2.5 m^{6}A guided dosage compensation in males and females

Lower GGACH motifs on X-chromosomal transcripts indicate an internally hardcoded mechanism. Therefore, this suggests that in both males and females, m^{6}A-mediated dosage compensation by RNA stability is an intrinsic feature of X-chromosomal transcripts. As detailed in chapter 3.2.2, female mESC have two active X chromosomes and hence, an elevated X:AA expression ratio. To strengthen the hypothesis of an internally hardcoded mechanism, we generated RNA-seq data for X0 and XX female subclones in pre- and absence of m^{6}A. Although the expression ratio between X0 and XX clones is reduced, we find X-chromosomal transcripts differentially responded to m^{6}A loss compared to their autosomal counterparts in both, XX

and X0 subclones. In line with this, miCLIP2 and m6Aboost showed lower m^{6}A numbers on X-chromosomal transcripts. This further provides evidence for a hardcoded mechanism driven by lower GGACH motifs and the mechanism is in place even before X chromosome inactivation.

Upon differentiation, one X chromosome is inactivated in females (Lyon, 1961). Interestingly, escaper genes are able to escape the X-chromosome inactivation and are expressed from both X chromosomes. Hence, full dosage compensation potentially is not required for these genes. Therefore, analysing m^{6}A numbers and the X-chromosomal transcript response upon m^{6}A loss exhibits an exciting future question. Moreover, in human around 15% of X-chromosomal genes are escapees, while in mouse only around 3% are considered escapees (Berletch *et al.,* 2010). Therefore, the evolutionary aspect of escapees between mouse and human and differences in m^{6}A content on these genes are exciting future research opportunities.

-

4. Outlook and Conclusion

Since m^{6}A was discovered in the 1970s, the modification has emerged as a crucial and complex regulator of post-transcriptional gene expression regulation. In order to fully understand the biological consequences of m^{6}A, it is crucial to precisely locate the modification in a confident and transcriptome-wide manner. However, the mapping of m^{6}A has only recently been enabled and remains challenging, to date.

The aim of my PhD was to improve the efficiency of iCLIP library preparation by developing iCLIP2. Furthermore, the aim was to further improve the m^{6}A detection method by combining our improved iCLIP2 protocol with the m^{6}A-antibody-based detection method miCLIP. Moreover, we combined our experimental advances with a robust computational pipeline and trained a machine learning model to reliably extract genuine m^{6}A sites from miCLIP2 data. In this study, we show that miCLIP2 produces high-complexity libraries from low input material. The machine learning classifier m6Aboost identifies high-confidence Mettl3-dependent m^{6}A sites from miCLIP2 data across tissues and species. In order to get a complete m^{6}A map without Mettl3-dependency, future projects could entail to map m^{6}A sites that are deposited by different methyltransferases. Therefore, the machine learner m6Aboost could be complemented by KO data sets from different methyltransferases like Mettl16 or Mettl5. Moreover, it would be exciting to test miCLIP2 in combination with m6Aboost in species that are evolutionary further away from mammals. Here, miCLIP2 could be applied to different tissues or cell lines of different species. Another exciting question would be whether miCLIP2 in combination with a machine learner could be used for different RNA modifications. Therefore, the miCLIP2 protocol could be adapted and antibodies against different RNA modifications could be used.
Additionally, we found that m^{6}A sites accumulate towards the 5’ splice site and the depletion of m^{6}A triggers the efficient splicing of retained introns. To resolve the discrepancy in literature whether the majority of m^{6}A site occur in intronic or exonic sequences, nascent miCLIP2 could shed light in this issue. Moreover, this could help to understand the functional impact on splicing.
Taken together, our miCLIP2 protocol in combination with m6Aboost provides a robust method for m^{6}A site identification and opens a reliable tool for future research. Importantly, we generated several transcriptome-wide m^{6}A maps in different species and cell lines, which provides a rich resource for future studies.

The second aim of this study comprised the usage of our generated high-confidence transcriptome-wide m^{6}A maps to uncover novel biological functions of m^{6}A. Since differences in RNA stability were found as a mechanism for dosage compensation and the degrading nature of m^{6}A, we aimed to uncover the potential contribution of m^{6}A to X-to-autosome dosage compensation in mammals. We found that X-chromosomal transcripts are more stable compared to autosomes because they harbour less m^{6}A sites. Thus, autosomal transcripts are

selectively destabilized in an m^{6}A-dependent manner. Moreover, we found X-chromosomal and autosomal transcripts are differentially affected by acute m^{6}A-depletion. We further found that the reduced m^{6}A levels are internally driven by fewer GGACH motifs on X-linked transcripts. Additionally, we showed that the m^{6}A-mediated dosage compensation pathway is present in both sexes. Hence, this study uncovers a novel function of m^{6}A in mediating mammalian dosage compensation. Future studies could unravel the potential interplay with other dosage compensation mechanisms. Here, the potential interplay between the NMD and the m^{6}A pathway could be further investigated. The interaction between the main NMD player UPF1 and m^{6}A reader protein YTHDF2 strongly indicates a connection between both pathways and exhibits an exciting direction for future research.

Since the X chromosome is less occupied by Mettl3 and we find a lower fraction of methylated DRACH motifs on X-chromosomal transcripts, other mechanisms besides the hardcoded mechanism could be in place. To check whether an epigenetic mechanism is in place, differences of the H3K36me3 histone marks on different chromosomes could be analysed. Furthermore, translocating X-chromosomal genes to the autosomes could shed light on the question whether a Mettl3 repelling mechanism on the X chromosome is existent. Another possible explanation could be an eraser-protein guided mechanism. Here, ChIP-seq experiments for the m^{6}A eraser proteins could shed light on this idea.

Another exiting direction is tracking the sex-chromosome specific evolution of fewer GGACH motifs and m^{6}A modifications. Hence, m^{6}A could be mapped using miCLIP2 and m6Aboost in species that are evolutionary further away from mammals. This could shed light on when selection towards lower GGACH motifs on X-linked genes evolved. In line with this, the evolution of X-chromosome inactivation escaper genes could be further analysed. Since dosage compensation may not be required, it could be exciting question whether these genes are similar in m^{6}A levels compared to autosomal genes. Hence, this could further shed light on how selection of fewer GGACH motifs evolved.

An important consideration is that dosage compensation might be especially important for the stoichiometry of multisubunit protein complexes. Therefore, it would be interesting to investigate the direct changes on translation of X chromosome versus autosome encoded genes upon m^{6}A depletion. Here, the nascent transcriptome could be analysed upon immediate loss of m^{6}A, which could directly be linked to stoichiometry of multicomponent complexes.

Taken together, we find a novel function for m^{6}A in mammalian X-to-autosome dosage compensation. This unravels that chromosomal origin of an RNA transcript is remembered due to an internally hardcoded epitranscriptomic mechanism. Thus, the epitranscriptome bridges DNA context to the fate of an RNA molecule. This opens exciting questions for future research.

5. List of Abbreviations

%	Percent
>	Over
Ψ	Pseudouridine
2'Om	2'-O-methyladenosine
4SU	4-thiouridine
5mC	5-Methylcytosine
6SG	6-thioguanosine
A	Adenine
a^{6}A	N^6-allyladenosines
ac^{4}C	N^4-acetylcytidine
AD	Alzheimer disease
ADAR	Adenosine deaminases acting on RNA
ALKBH5	Alkb homolog 5
ALS	Amyotrophic lateral sclerosis
AML	Acute myeloid leukaemia
APOBEC1	Apolipoprotein B mRNA editing enzyme catalytic subunit 1
APEX2	Apurinic/apyrimidinic endodeoxyribonuclease 2
C	Cyosine
C643	Human Thyroidea Carcinoma cell line
CA-RNA	Chromatin-associated RNA
CAP-seq	Chromophore-assisted proximity labelling and sequencing
CCR4	C-C Motif Chemokine Receptor 4
cDNA	Complementary DNA
CDS	Coding sequence
CH$_3$	Methyl group
ChIP-seq	Chromatin immunoprecipitation followed by sequencing
circRNA	Circular RNA
CLIP	Cross-linking and immunoprecipitation
Ctcf	CCCTC-binding factor
DART-seq	Deamination adjacent to RNA modification targets
DCC	Dosage compensation complex
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase
dm^{6}A	N^6-dithiolsitolmethyladenosine
dTAG	Degradation tag
DTT	Dithiothreitol
EWSR1	EWS RNA binding protein 1
f^{6}A	N^6-formyladenosine
e.g.	exempli gratia (for example)
eCLIP	Enhanced CLIP
EDTA	Ethylenediaminetetraacetic acid
FMRP	Fragile X mental retardation protein
FTO	Fat mass and obesity-associated protein
G	Guanine

G3BP	Stress granule assembly factor
GLORI-seq	Glyoxal and nitrite-mediated deamination of unmethylated adenosine
GO	Gene ontology
H	Hour
H2K119ub1	H2A lysine 119-monoubiquititinated
H3K9me3	H3 trimethylation of lysine 9
H3K27me3	H3 trimethylation of lysine 27
H3K36me3	H3 trimethylation at Lysine 36
H4K20me1	H4 monomethylation of lysine 20
HAKAI	E3 ubiquitin ligase Hakai
HCV	Hepatitis C virus
HEK293T	Human embryonic kidney 293T cells
HIV-1	Human immunodeficiency virus type 1
hm^{5}C	5-hydroxymethylcytosine
hm^{6}A	N^6-hydroxymethyladenosine
HNRNPA2B1	Heterogeneous nuclear ribonucleoprotein A2B1
HNRNPC	Heterogeneous nuclear ribonucleoprotein C
hPUM2	Human RNA-binding protein Pumilio 2
I	Inosine
IGF2BP	insulin like growth factor proteins
Ime4	Inducer of meiosis
IP	Immunoprecipitation
iPSCs	Induced pluripotent stem cells
IR	Intron retention
irCLIP	Infrared CLIP
KH	K homology
KO	Knockout
LLPS	Liquid-liquid phase separation
lncRNA	Long non-coding RNA
µg	Micro gram
m^{1}A	N^1-methyladenosine
m^{3}C	3-methylcytosine
m^{5}C	5-methylcytidine
m^{6}A	N^6-methyladenosine
m^{6}A-SEAL-seq	FTO-assisted m^{6}A selective chemical labeling method
m^{6}Am	2'-O-dimethyladenosine
m^{7}G	7-methylguanosine
MAZTER-seq	RNA digestion via m^{6}A sensitive RNase
MazF	Endoribonuclease toxin MazF
mESC	Mouse embryonic stem cells
MEF	Mouse embryonic fibroblasts
MeRIP-seq	m^{6}A-specific methylated RNA immunoprecipitation with next generation sequencing
METTL3	Methyltransferase Like 3

METTL5	Methyltransferase Like 5
METTL14	Methyltransferase Like 14
METTL16	Methyltransferase Like 16
miCLIP	m^{6}A-individual nucleotide resolution UV-crosslinking and immunoprecipitation
MINES	m^{6}A Identification using Nanopore Sequencing
miRNA	Micro RNA
MOF	Males absent on the first
mRNA	Messenger RNA
MSL	Male-specific-lethal
ncRNA	non-coding RNA
nM	Nano meter
Nm	2'O-methylations
NMD	Nonsense-mediated decay
NOT	Negative on TATA-less
NPC	Neuronal progenitor cells
NSUN	RNA cytosine C(5)-methyltransferase NSUN2NOL1/NOP2/Sun domain family member
nt	Nucleotide
NXF1	Nuclear RNA export factor 1
o^{8}G	8-oxoguanine
P bodies	Processing bodies
PAR	Pseudoautosomal region
PAR-CLIP	Photoactivatable-ribonucleoside-enhanced crosslinking and immunoprecipitation
PCR	Polymerase chain reaction
PolII-S5p	RNA pol II, which is phosphorylated at serine 5
qPCR	Quantitative polymerase chain reaction
RBM15/15B	RNA binding motif protein 15/15B
RBP	RNA-binding protein
Rex1	Zinc finger protein 42
RMP	RNA-modifying protein
RNA	Ribonucleic acic
RNA PolII	RNA polymerase II
Rnf12	E3 ubiquitin-protein ligase RLIM
rRNA	Ribosomal RNA
RT	Reverse transcription
SAM	S-adenosylmethionine
SCARLET	Site-specific cleavage and radioactive labelling followed by ligation-assisted extraction and thin-layer chromatography
scRNA-seq	Single cell RNA sequencing
SDS-PAGE	Sodium dodecyl-sulfate polyacrylamide gel electrophoresis
Seq	Sequencing
SG	Stress granules
SLAM-seq	Thiol(SH)-linked alkylation for the metabolic sequencing of RNA
snoRNA	Small nucleolar RNA
snRNA	Small nuclear RNA
ssDNA	Single-stranded DNA

SRSF3	Serine and Arginine rich Splicing factor 3
SRSF10	Serine and Arginine rich Splicing factor 10
SVM	Support vector machine
T	Thymine
T2D	Type 2 Diabetes mellitus
TARBP2	Trans-Activation-Responsive RNA-Binding Protein
TNT-seq	Transient N-6-methyladenosine transcriptome sequencing
tRNA	Transfer RNA
TRMT112	TRNA Methyltransferase Activator Subunit 11-2
Tsix	XIST Antisense RNA
TSS	Transcription start site
UMI	Unique molecular identifier
UPF1	Up-Frameshift Suppressor 1 Homolog
UTR	Untranslated region
UV	Ultraviolet
VIRMA	Vir like m^6A methylation associated
WT	Wild type
WTAP	Wilms' tumor 1-associating protein
X:AA	X-to-autosome expression
X	Xanthosine
X_a	Active X chromosome
X_i	Inactive X chromosome
XIST	X Inactive Specific Transcript
Xrn1	5'-3' Exoribonuclease 1
YTH	YT521-B homology
YTHDC	YTH domain containing
YTHDF	YTH domain family
ZC3H13	Zinc finger CCCH-type containing 13
ZIKV	Zika virus

6. References

Aas, A., Isakson, P., Bindesbøll, C., Alemu, E. A., Klungland, A., & Simonsen, A. (2017). Nucleocytoplasmic Shuttling of FTO Does Not Affect Starvation-Induced Autophagy. *PloS one*, *12*(3), e0168182. https://doi.org/10.1371/journal.pone.0168182

Adams, J. M., & Cory, S. (1975). Modified nucleosides and bizarre 5'-termini in mouse myeloma mRNA. *Nature*, *255*(5503), 28–33. https://doi.org/10.1038/255028a0

Alarcón, C. R., Goodarzi, H., Lee, H., Liu, X., Tavazoie, S., & Tavazoie, S. F. (2015). HNRNPA2B1 Is a Mediator of m(6)A-Dependent Nuclear RNA Processing Events. *Cell*, *162*(6), 1299–1308. https://doi.org/10.1016/j.cell.2015.08.011

Alaskhar Alhamwe, B., Khalaila, R., Wolf, J., von Bülow, V., Harb, H., Alhamdan, F., Hii, C. S., Prescott, S. L., Ferrante, A., Renz, H., Garn, H., & Potaczek, D. P. (2018). Histone modifications and their role in epigenetics of atopy and allergic diseases. *Allergy, asthma, and clinical immunology : official journal of the Canadian Society of Allergy and Clinical Immunology*, *14*, 39. https://doi.org/10.1186/s13223-018-0259-4

ALLFREY, V. G., FAULKNER, R., & MIRSKY, A. E. (1964). ACETYLATION AND METHYLATION OF HISTONES AND THEIR POSSIBLE ROLE IN THE REGULATION OF RNA SYNTHESIS. *Proceedings of the National Academy of Sciences of the United States of America*, *51*(5), 786–794. https://doi.org/10.1073/pnas.51.5.786

Amirbeigiarab, S., Kiani, P., Velazquez Sanchez, A., Krisp, C., Kazantsev, A., Fester, L., Schlüter, H., & Ignatova, Z. (2019). Invariable stoichiometry of ribosomal proteins in mouse brain tissues with aging. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(45), 22567–22572. https://doi.org/10.1073/pnas.1912060116

Arango, D., Sturgill, D., Alhusaini, N., Dillman, A. A., Sweet, T. J., Hanson, G., Hosogane, M., Sinclair, W. R., Nanan, K. K., Mandler, M. D., Fox, S. D., Zengeya, T. T., Andresson, T., Meier, J. L., Coller, J., & Oberdoerffer, S. (2018). Acetylation of Cytidine in mRNA Promotes Translation Efficiency. *Cell*, *175*(7), 1872–1886.e24. https://doi.org/10.1016/j.cell.2018.10.030

Ashworth, A., Rastan, S., Lovell-Badge, R., & Kay, G. (1991). X-chromosome inactivation may explain the difference in viability of XO humans and mice. *Nature*, *351*(6325), 406–408. https://doi.org/10.1038/351406a0

Bartosovic, M., Molares, H. C., Gregorova, P., Hrossova, D., Kudla, G., & Vanacova, S. (2017). N6-methyladenosine demethylase FTO targets pre-mRNAs and regulates alternative splicing and 3'-end processing. *Nucleic acids research*, *45*(19), 11356–11370. https://doi.org/10.1093/nar/gkx778

Bass, B. L., & Weintraub, H. (1988). An unwinding activity that covalently modifies its double-stranded RNA substrate. *Cell*, *55*(6), 1089–1098. https://doi.org/10.1016/0092-8674(88)90253-x

Baltimore D. (1970). RNA-dependent DNA polymerase in virions of RNA tumour viruses. *Nature*, *226*(5252), 1209–1211. https://doi.org/10.1038/2261209a0

Bawankar, P., Lence, T., Paolantoni, C., Haussmann, I. U., Kazlauskiene, M., Jacob, D., Heidelberger, J. B., Richter, F. M., Nallasivan, M. P., Morin, V., Kreim, N., Beli, P., Helm, M., Jinek, M., Soller, M., & Roignant, J. Y. (2021). Hakai is required for stabilization of core components of the m^{6}A mRNA methylation machinery. *Nature communications*, *12*(1), 3778. https://doi.org/10.1038/s41467-021-23892-5

Benhalevy, D., Anastasakis, D. G., & Hafner, M. (2018). Proximity-CLIP provides a snapshot of protein-occupied RNA elements in subcellular compartments. *Nature methods*, *15*(12), 1074–1082. https://doi.org/10.1038/s41592-018-0220-y

Berletch, J. B., Yang, F., & Disteche, C. M. (2010). Escape from X inactivation in mice and humans. *Genome biology*, *11*(6), 213. https://doi.org/10.1186/gb-2010-11-6-213

Bird A. (2002). DNA methylation patterns and epigenetic memory. *Genes & development*, *16*(1), 6–21. https://doi.org/10.1101/gad.947102

Bokar, J. A., Shambaugh, M. E., Polayes, D., Matera, A. G., & Rottman, F. M. (1997). Purification and cDNA cloning of the AdoMet-binding subunit of the human mRNA (N6-adenosine)-methyltransferase. *RNA (New York, N.Y.)*, *3*(11), 1233–1247.

Boo, S. H., Ha, H., Lee, Y., Shin, M. K., Lee, S., & Kim, Y. K. (2022). UPF1 promotes rapid degradation of m^{6}A-containing RNAs. *Cell reports*, *39*(8), 110861. https://doi.org/10.1016/j.celrep.2022.110861

Brockdorff, N., & Turner, B. M. (2015). Dosage compensation in mammals. *Cold Spring Harbor perspectives in biology*, *7*(3), a019406. https://doi.org/10.1101/cshperspect.a019406

Brown, C. J., Ballabio, A., Rupert, J. L., Lafreniere, R. G., Grompe, M., Tonlorenzi, R., & Willard, H. F. (1991). A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature*, *349*(6304), 38–44. https://doi.org/10.1038/349038a0

Buchbender, A., Mutter, H., Sutandy, F. X. R., Körtel, N., Hänel, H., Busch, A., Ebersberger, S., & König, J. (2020). Improved library preparation with the new iCLIP2 protocol. *Methods (San Diego, Calif.)*, *178*, 33–48. https://doi.org/10.1016/j.ymeth.2019.10.003

Bujnicki, J. M., Feder, M., Radlinska, M., & Blumenthal, R. M. (2002). Structure prediction and phylogenetic analysis of a functionally diverse family of proteins homologous to the MT-A70 subunit of the human mRNA:m(6)A methyltransferase. *Journal of molecular evolution*, *55*(4), 431–444. https://doi.org/10.1007/s00239-002-2339-8

Burgoyne P. S. (1982). Genetic homology and crossing over in the X and Y chromosomes of Mammals. *Human genetics*, *61*(2), 85–90. https://doi.org/10.1007/BF00274192

Canaani, D., Kahana, C., Lavi, S., & Groner, Y. (1979). Identification and mapping of N6-methyladenosine containing sequences in simian virus 40 RNA. *Nucleic acids research*, *6*(8), 2879–2899. https://doi.org/10.1093/nar/6.8.2879

Carrel, L., & Willard, H. F. (2005). X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature*, *434*(7031), 400–404. https://doi.org/10.1038/nature03479

Chang, M., Lv, H., Zhang, W., Ma, C., He, X., Zhao, S., Zhang, Z. W., Zeng, Y. X., Song, S., Niu, Y., & Tong, W. M. (2017). Region-specific RNA m^{6}A methylation represents a new layer of control in the gene regulatory network in the mouse brain. *Open biology*, *7*(9), 170166. https://doi.org/10.1098/rsob.170166

Charlesworth B. (1996). The evolution of chromosomal sex determination and dosage compensation. *Current biology : CB*, *6*(2), 149–162. https://doi.org/10.1016/s0960-9822(02)00448-7

Charlesworth, D., Charlesworth, B., & Marais, G. (2005). Steps in the evolution of heteromorphic sex chromosomes. *Heredity*, *95*(2), 118–128. https://doi.org/10.1038/sj.hdy.6800697

Cantara, W. A., Crain, P. F., Rozenski, J., McCloskey, J. A., Harris, K. A., Zhang, X., Vendeix, F. A., Fabris, D., & Agris, P. F. (2011). The RNA Modification Database, RNAMDB: 2011 update. *Nucleic acids research*, *39*(Database issue), D195–D201. https://doi.org/10.1093/nar/gkq1028

Capitanchik, C., Toolan-Kerr, P., Luscombe, N. M., & Ule, J. (2020). How Do You Identify m^6 A Methylation in Transcriptomes at High Resolution? A Comparison of Recent Datasets. *Frontiers in genetics*, *11*, 398. https://doi.org/10.3389/fgene.2020.00398

Carlile, T. M., Rojas-Duran, M. F., Zinshteyn, B., Shin, H., Bartoli, K. M., & Gilbert, W. V. (2014). Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature*, *515*(7525), 143–146. https://doi.org/10.1038/nature13802

Charette, M., & Gray, M. W. (2000). Pseudouridine in RNA: what, where, how, and why. *IUBMB life*, *49*(5), 341–351. https://doi.org/10.1080/152165400410182

Chelmicki, T., Roger, E., Teissandier, A., Dura, M., Bonneville, L., Rucli, S., Dossin, F., Fouassier, C., Lameiras, S., & Bourc'his, D. (2021). m^6A RNA methylation regulates the fate of endogenous retroviruses. *Nature*, *591*(7849), 312–316. https://doi.org/10.1038/s41586-020-03135-1

(A) Chen, W., Tran, H., Liang, Z., Lin, H., & Zhang, L. (2015). Identification and analysis of the N(6)-methyladenosine in the Saccharomyces cerevisiae transcriptome. *Scientific reports*, *5*, 13859. https://doi.org/10.1038/srep13859

(B) Chen, T., Hao, Y. J., Zhang, Y., Li, M. M., M., Han, W., Wu, Y., Lv, Y., Hao, J., Wang, L., Li, A., Yang, Y., Jin, K. X., Zhao, X., Li, Y., Ping, X. L., Lai, W. Y., Wu, L. G., Jiang, G., Wang, H. L., … Zhou, Q. (2015). m(6)A RNA methylation is regulated by microRNAs and promotes reprogramming to pluripotency. *Cell stem cell*, *16*(3), 289–301. https://doi.org/10.1016/j.stem.2015.01.016

Chen, W., Tang, H., & Lin, H. (2017). MethyRNA: a web server for identification of N^6-methyladenosine sites. *Journal of biomolecular structure & dynamics*, *35*(3), 683–687. https://doi.org/10.1080/07391102.2016.1157761

(A) Chen, K., Wei, Z., Zhang, Q., Wu, X., Rong, R., Lu, Z., Su, J., de Magalhães, J. P., Rigden, D. J., & Meng, J. (2019). WHISTLE: a high-accuracy map of the human N6-methyladenosine (m6A) epitranscriptome predicted using a machine learning approach. *Nucleic acids research*, *47*(7), e41.

(B) Chen, X. Y., Zhang, J., & Zhu, J. S. (2019). The role of m^6A RNA methylation in human cancer. *Molecular cancer*, *18*(1), 103. https://doi.org/10.1186/s12943-019-1033-z

Chu, C., Zhang, Q. C., da Rocha, S. T., Flynn, R. A., Bharadwaj, M., Calabrese, J. M., Magnuson, T., Heard, E., & Chang, H. Y. (2015). Systematic discovery of Xist RNA binding proteins. *Cell*, *161*(2), 404–416. https://doi.org/10.1016/j.cell.2015.03.025

Clemson, C. M., McNeil, J. A., Willard, H. F., & Lawrence, J. B. (1996). XIST RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure. *The Journal of cell biology*, *132*(3), 259–275. https://doi.org/10.1083/jcb.132.3.259

COHN, W., VOLKIN, E. Nucleoside-5'-Phosphates from Ribonucleic Acid. *Nature* **167**, 483–484 (1951). https://doi.org/10.1038/167483a0

Collart M. A. (2016). The Ccr4-Not complex is a key regulator of eukaryotic gene expression. *Wiley interdisciplinary reviews. RNA*, *7*(4), 438–454. https://doi.org/10.1002/wrna.1332

CRICK F. H. (1958). On protein synthesis. *Symposia of the Society for Experimental Biology*, *12*, 138–163.

Crick F. (1970). Central dogma of molecular biology. *Nature*, *227*(5258), 561–563. https://doi.org/10.1038/227561a0

Czerwoniec, A., Dunin-Horkawicz, S., Purta, E., Kaminska, K. H., Kasprzak, J. M., Bujnicki, J. M., Grosjean, H., & Rother, K. (2009). MODOMICS: a database of RNA modification pathways. 2008 update. *Nucleic acids research*, *37*(Database issue), D118–D121. https://doi.org/10.1093/nar/gkn710

Dai, B., & Rasmussen, T. P. (2007). Global epiproteomic signatures distinguish embryonic stem cells from differentiated cells. *Stem cells (Dayton, Ohio)*, *25*(10), 2567–2574. https://doi.org/10.1634/stemcells.2007-0131

Daoud, H., Zhang, D., McMurray, F., Yu, A., Luco, S. M., Vanstone, J., Jarinova, O., Carson, N., Wickens, J., Shishodia, S., Choi, H., McDonough, M. A., Schofield, C. J., Harper, M. E., Dyment, D. A., & Armour, C. M. (2016). Identification of a pathogenic FTO mutation by next-generation sequencing in a newborn with growth retardation and developmental delay. *Journal of medical genetics*, *53*(3), 200–207. https://doi.org/10.1136/jmedgenet-2015-103399

Davis D. R. (1995). Stabilization of RNA stacking by pseudouridine. *Nucleic acids research*, *23*(24), 5020–5026. https://doi.org/10.1093/nar/23.24.5020

Deamer, D., Akeson, M., & Branton, D. (2016). Three decades of nanopore sequencing. *Nature biotechnology*, *34*(5), 518–524. https://doi.org/10.1038/nbt.3423

Debrand, E., Chureau, C., Arnaud, D., Avner, P., & Heard, E. (1999). Functional analysis of the DXPas34 locus, a 3' regulator of Xist expression. *Molecular and cellular biology*, *19*(12), 8513–8525. https://doi.org/10.1128/MCB.19.12.8513

Delatte, B., Wang, F., Ngoc, L. V., Collignon, E., Bonvin, E., Deplus, R., Calonne, E., Hassabi, B., Putmans, P., Awe, S., Wetzel, C., Kreher, J., Soin, R., Creppe, C., Limbach, P. A., Gueydan, C., Kruys, V., Brehm, A., Minakhina, S., Defrance, M., … Fuks, F. (2016). RNA biochemistry. Transcriptome-wide distribution and function of RNA hydroxymethylcytosine. *Science (New York, N.Y.)*, *351*(6270), 282–285. https://doi.org/10.1126/science.aac5253

Deng, X., Hiatt, J. B., Nguyen, D. K., Ercan, S., Sturgill, D., Hillier, L. W., Schlesinger, F., Davis, C. A., Reinke, V. J., Gingeras, T. R., Shendure, J., Waterston, R. H., Oliver, B., Lieb, J. D., & Disteche, C. M. (2011). Evidence for compensatory upregulation of expressed X-linked genes in mammals, Caenorhabditis elegans and Drosophila melanogaster. *Nature genetics*, *43*(12), 1179–1185. https://doi.org/10.1038/ng.948

Deng, X., Berletch, J. B., Ma, W., Nguyen, D. K., Hiatt, J. B., Noble, W. S., Shendure, J., & Disteche, C. M. (2013). Mammalian X upregulation is associated with enhanced transcription initiation, RNA half-life, and MOF-mediated H4K16 acetylation. *Developmental cell*, *25*(1), 55–68. https://doi.org/10.1016/j.devcel.2013.01.028

Deng, X., Chen, K., Luo, G. Z., Weng, X., Ji, Q., Zhou, T., & He, C. (2015). Widespread occurrence of N6-methyladenosine in bacterial mRNA. *Nucleic acids research*, *43*(13), 6557–6567. https://doi.org/10.1093/nar/gkv596

Desaulniers, J. P., Chang, Y. C., Aduri, R., Abeysirigunawardena, S. C., SantaLucia, J., Jr, & Chow, C. S. (2008). Pseudouridines in rRNA helix 69 play a role in loop stacking interactions. *Organic & biomolecular chemistry*, *6*(21), 3892–3895. https://doi.org/10.1039/b812731j

Desrosiers, R., Friderici, K., & Rottman, F. (1974). Identification of methylated nucleosides in messenger RNA from Novikoff hepatoma cells. *Proceedings of the National Academy of Sciences of the United States of America*, *71*(10), 3971–3975. https://doi.org/10.1073/pnas.71.10.3971

Dierks, D., Garcia-Campos, M. A., Uzonyi, A., Safra, M., Edelheit, S., Rossi, A., Sideri, T., Varier, R. A., Brandis, A., Stelzer, Y., van Werven, F., Scherz-Shouval, R., & Schwartz, S. (2021). Multiplexed profiling facilitates robust m6A quantification at site, gene and sample resolution. *Nature methods*, *18*(9), 1060–1067. https://doi.org/10.1038/s41592-021-01242-z

Disteche C. M. (1999). Escapees on the X chromosome. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(25), 14180–14182. https://doi.org/10.1073/pnas.96.25.14180

Dominissini, D., Moshitch-Moshkovitz, S., Schwartz, S., Salmon-Divon, M., Ungar, L., Osenberg, S., Cesarkas, K., Jacob-Hirsch, J., Amariglio, N., Kupiec, M., Sorek, R., & Rechavi, G. (2012). Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature*, *485*(7397), 201–206. https://doi.org/10.1038/nature11112

Dominissini, D., Nachtergaele, S., Moshitch-Moshkovitz, S., Peer, E., Kol, N., Ben-Haim, M. S., Dai, Q., Di Segni, A., Salmon-Divon, M., Clark, W. C., Zheng, G., Pan, T., Solomon, O., Eyal, E., Hershkovitz, V., Han, D., Doré, L. C., Amariglio, N., Rechavi, G., & He, C. (2016). The dynamic N(1)-methyladenosine methylome in eukaryotic messenger RNA. *Nature*, *530*(7591), 441–446. https://doi.org/10.1038/nature16998

Du, H., Zhao, Y., He, J., Zhang, Y., Xi, H., Liu, M., Ma, J., & Wu, L. (2016). YTHDF2 destabilizes m(6)A-containing RNA through direct recruitment of the CCR4-NOT deadenylase complex. *Nature communications*, *7*, 12626. https://doi.org/10.1038/ncomms12626

Dubin, D. T., & Stollar, V. (1975). Methylation of Sindbis virus "26S" messenger RNA. *Biochemical and biophysical research communications*, *66*(4), 1373–1379. https://doi.org/10.1016/0006-291x(75)90511-2

Dubin, D. T., & Taylor, R. H. (1975). The methylation state of poly A-containing messenger RNA from cultured hamster cells. *Nucleic acids research*, *2*(10), 1653–1668. https://doi.org/10.1093/nar/2.10.1653

DUNN D. B. (1961). The occurrence of 1-methyladenine in ribonucleic acid. *Biochimica et biophysica acta*, *46*, 198–200. https://doi.org/10.1016/0006-3002(61)90668-0

Edelheit, S., Schwartz, S., Mumbach, M. R., Wurtzel, O., & Sorek, R. (2013). Transcriptome-wide mapping of 5-methylcytidine RNA modifications in bacteria, archaea, and yeast reveals m5C within archaeal mRNAs. *PLoS genetics*, *9*(6), e1003602. https://doi.org/10.1371/journal.pgen.1003602

Edens, B. M., Vissers, C., Su, J., Arumugam, S., Xu, Z., Shi, H., Miller, N., Rojas Ringeling, F., Ming, G. L., He, C., Song, H., & Ma, Y. C. (2019). FMRP Modulates Neural Differentiation through m^6A-Dependent mRNA Nuclear Export. *Cell reports*, *28*(4), 845–854.e5. https://doi.org/10.1016/j.celrep.2019.06.072

Edupuganti, R. R., Geiger, S., Lindeboom, R. G. H., Shi, H., Hsu, P. J., Lu, Z., Wang, S. Y., Baltissen, M. P. A., Jansen, P. W. T. C., Rossa, M., Müller, M., Stunnenberg, H. G., He, C., Carell, T., & Vermeulen, M. (2017). N^6-methyladenosine (m^6A) recruits and repels proteins to regulate mRNA homeostasis. *Nature structural & molecular biology*, *24*(10), 870–878. https://doi.org/10.1038/nsmb.3462

Ellegren, H., Hultin-Rosenberg, L., Brunström, B., Dencker, L., Kultima, K., & Scholz, B. (2007). Faced with inequality: chicken do not have a general dosage compensation of sex-linked genes. *BMC biology*, *5*, 40. https://doi.org/10.1186/1741-7007-5-40

Epstein C. J. (1990). The consequences of chromosome imbalance. *American journal of medical genetics. Supplement*, *7*, 31–37. https://doi.org/10.1002/ajmg.1320370706

Faucillion, M. L., & Larsson, J. (2015). Increased expression of X-linked genes in mammals is associated with a higher stability of transcripts and an increased ribosome density. *Genome biology and evolution*, *7*(4), 1039–1052. https://doi.org/10.1093/gbe/evv054

Filippova, G. N., Cheng, M. K., Moore, J. M., Truong, J. P., Hu, Y. J., Nguyen, D. K., Tsuchiya, K. D., & Disteche, C. M. (2005). Boundaries between chromosomal domains of X inactivation and escape bind CTCF and lack CpG methylation during early development. *Developmental cell*, *8*(1), 31–42. https://doi.org/10.1016/j.devcel.2004.10.018

Fish, L., Navickas, A., Culbertson, B., Xu, Y., Nguyen, H. C. B., Zhang, S., Hochman, M., Okimoto, R., Dill, B. D., Molina, H., Najafabadi, H. S., Alarcón, C., Ruggero, D., & Goodarzi, H. (2019). Nuclear TARBP2 Drives Oncogenic Dysregulation of RNA Splicing and Decay. *Molecular cell*, *75*(5), 967–981.e9. https://doi.org/10.1016/j.molcel.2019.06.001

Furuichi Y. (2015). Discovery of m(7)G-cap in eukaryotic mRNAs. *Proceedings of the Japan Academy. Series B, Physical and biological sciences*, *91*(8), 394–409. https://doi.org/10.2183/pjab.91.394

Fustin, J. M., Doi, M., Yamaguchi, Y., Hida, H., Nishimura, S., Yoshida, M., Isagawa, T., Morioka, M. S., Kakeya, H., Manabe, I., & Okamura, H. (2013). RNA-methylation-dependent RNA processing controls the speed of the circadian clock. *Cell*, *155*(4), 793–806. https://doi.org/10.1016/j.cell.2013.10.026

Fustin, J. M., Kojima, R., Itoh, K., Chang, H. Y., Ye, S., Zhuang, B., Oji, A., Gibo, S., Narasimamurthy, R., Virshup, D., Kurosawa, G., Doi, M., Manabe, I., Ishihama, Y., Ikawa, M., & Okamura, H. (2018). Two *Ck1δ* transcripts regulated by m6A methylation code for two antagonistic kinases in the control of the circadian clock. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(23), 5980–5985. https://doi.org/10.1073/pnas.1721371115

Galupa, R., & Heard, E. (2018). X-Chromosome Inactivation: A Crossroads Between Chromosome Architecture and Gene Regulation. *Annual review of genetics*, *52*, 535–566. https://doi.org/10.1146/annurev-genet-120116-024611

Garcia-Campos, M. A., Edelheit, S., Toth, U., Safra, M., Shachar, R., Viukov, S., Winkler, R., Nir, R., Lasman, L., Brandis, A., Hanna, J. H., Rossmanith, W., & Schwartz, S. (2019). Deciphering the "m^6A Code" via Antibody-Independent Quantitative Profiling. *Cell*, *178*(3), 731–747.e16. https://doi.org/10.1016/j.cell.2019.06.013

Gehring, N. H., & Roignant, J. Y. (2021). Anything but Ordinary - Emerging Splicing Mechanisms in Eukaryotic Gene Regulation. *Trends in genetics : TIG*, *37*(4), 355–372. https://doi.org/10.1016/j.tig.2020.10.008

Geula, S., Moshitch-Moshkovitz, S., Dominissini, D., Mansour, A. A., Kol, N., Salmon-Divon, M., Hershkovitz, V., Peer, E., Mor, N., Manor, Y. S., Ben-Haim, M. S., Eyal, E., Yunger, S., Pinto, Y., Jaitin, D. A., Viukov, S., Rais, Y., Krupalnik, V., Chomsky, E., Zerbib, M., … Hanna, J. H. (2015). Stem cells. m6A mRNA methylation facilitates resolution of naïve pluripotency toward differentiation. *Science (New York, N.Y.)*, *347*(6225), 1002–1006. https://doi.org/10.1126/science.1261417

Goto, Y., Gomez, M., Brockdorff, N., & Feil, R. (2002). Differential patterns of histone methylation and acetylation distinguish active and repressed alleles at X-linked genes. *Cytogenetic and genome research*, *99*(1-4), 66–74. https://doi.org/10.1159/000071576

Greenberg, H., & Penman, S. (1966). Methylation and processing of ribosomal RNA in HeLa cells. *Journal of molecular biology*, *21*(3), 527–535. https://doi.org/10.1016/0022-2836(66)90025-8

Greer, E. L., & Shi, Y. (2012). Histone methylation: a dynamic mark in health, disease and inheritance. *Nature reviews. Genetics*, *13*(5), 343–357. https://doi.org/10.1038/nrg3173

Grosjean, H. Modification and editing of RNA: historical overview and important facts to remember. In: Grosjean, H. (eds) Fine-Tuning of RNA Functions by Modification and Editing. Topics in Current Genetics, vol 12. Springer, Berlin, Heidelberg. https://doi.org/10.1007/b106848

Grosjean H. (2015). RNA modification: the Golden Period 1995-2015. *RNA (New York, N.Y.)*, *21*(4), 625–626. https://doi.org/10.1261/rna.049866.115

Gulati, P., Avezov, E., Ma, M., Antrobus, R., Lehner, P., O'Rahilly, S., & Yeo, G. S. (2014). Fat mass and obesity-related (FTO) shuttles between the nucleus and cytoplasm. *Bioscience reports*, *34*(5), e00144. https://doi.org/10.1042/BSR20140111

Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jr, Jungkamp, A. C., Munschauer, M., Ulrich, A., Wardle, G. S., Dewell, S., Zavolan, M., & Tuschl, T. (2010). Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell*, *141*(1), 129–141. https://doi.org/10.1016/j.cell.2010.03.009

Hafner, M., Katsantoni, M., Köster, T. *et al.* CLIP and complementary methods. *Nat Rev Methods Primers* **1**, 20 (2021). https://doi.org/10.1038/s43586-021-00018-1

Han, M., Liu, Z., Xu, Y., Liu, X., Wang, D., Li, F., Wang, Y., & Bi, J. (2020). Abnormality of m6A mRNA Methylation Is Involved in Alzheimer's Disease. *Frontiers in neuroscience*, *14*, 98. https://doi.org/10.3389/fnins.2020.00098

Harper, J. E., Miceli, S. M., Roberts, R. J., & Manley, J. L. (1990). Sequence specificity of the human mRNA N6-adenosine methylase in vitro. *Nucleic acids research*, *18*(19), 5735–5741. https://doi.org/10.1093/nar/18.19.5735

Hartmann, A. M., Nayler, O., Schwaiger, F. W., Obermeier, A., & Stamm, S. (1999). The interaction and colocalization of Sam68 with the splicing-associated factor YT521-B in nuclear dots is regulated by the Src family kinase p59(fyn). *Molecular biology of the cell*, *10*(11), 3909–3926. https://doi.org/10.1091/mbc.10.11.3909

Haussmann, I. U., Bodi, Z., Sanchez-Moran, E., Mongan, N. P., Archer, N., Fray, R. G., & Soller, M. (2016). m^6A potentiates Sxl alternative pre-mRNA splicing for robust Drosophila sex determination. *Nature*, *540*(7632), 301–304. https://doi.org/10.1038/nature20577

He C. (2010). Grand challenge commentary: RNA epigenetics?. *Nature chemical biology*, *6*(12), 863–865. https://doi.org/10.1038/nchembio.482

He, F., & Jacobson, A. (2015). Nonsense-Mediated mRNA Decay: Degradation of Defective Transcripts Is Only Part of the Story. *Annual review of genetics*, *49*, 339–366. https://doi.org/10.1146/annurev-genet-112414-054639

Heard, E., Rougeulle, C., Arnaud, D., Avner, P., Allis, C. D., & Spector, D. L. (2001). Methylation of histone H3 at Lys-9 is an early mark on the X chromosome during X inactivation. *Cell*, *107*(6), 727–738. https://doi.org/10.1016/s0092-8674(01)00598-0

Heard, E., & Disteche, C. M. (2006). Dosage compensation in mammals: fine-tuning the expression of the X chromosome. *Genes & development*, *20*(14), 1848–1867. https://doi.org/10.1101/gad.1422906

Helm, M., Lyko, F., & Motorin, Y. (2019). Limited antibody specificity compromises epitranscriptomic analyses. *Nature communications*, *10*(1), 5669. https://doi.org/10.1038/s41467-019-13684-3

Herzog, V. A., Reichholf, B., Neumann, T., Rescheneder, P., Bhat, P., Burkard, T. R., Wlotzka, W., von Haeseler, A., Zuber, J., & Ameres, S. L. (2017). Thiol-linked alkylation of RNA to assess expression dynamics. *Nature methods*, *14*(12), 1198–1204. https://doi.org/10.1038/nmeth.4435

Hess, M. E., Hess, S., Meyer, K. D., Verhagen, L. A., Koch, L., Brönneke, H. S., Dietrich, M. O., Jordan, S. D., Saletore, Y., Elemento, O., Belgardt, B. F., Franz, T., Horvath, T. L., Rüther, U., Jaffrey, S. R., Kloppenburg, P., & Brüning, J. C. (2013). The fat mass and obesity associated gene (Fto) regulates activity of the dopaminergic midbrain circuitry. *Nature neuroscience*, *16*(8), 1042–1048. https://doi.org/10.1038/nn.3449

Hideyama, T., & Kwak, S. (2011). When Does ALS Start? ADAR2-GluA2 Hypothesis for the Etiology of Sporadic ALS. *Frontiers in molecular neuroscience*, *4*, 33. https://doi.org/10.3389/fnmol.2011.00033

Hideyama, T., Yamashita, T., Aizawa, H., Tsuji, S., Kakita, A., Takahashi, H., & Kwak, S. (2012). Profound downregulation of the RNA editing enzyme ADAR2 in ALS spinal motor neurons. *Neurobiology of disease*, *45*(3), 1121–1128. https://doi.org/10.1016/j.nbd.2011.12.033

Hillman, R. T., Green, R. E., & Brenner, S. E. (2004). An unappreciated role for RNA surveillance. *Genome biology*, *5*(2), R8. https://doi.org/10.1186/gb-2004-5-2-r8

Hongay, C. F., & Orr-Weaver, T. L. (2011). Drosophila Inducer of MEiosis 4 (IME4) is required for Notch signaling during oogenesis. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(36), 14855–14860. https://doi.org/10.1073/pnas.1111577108

Horiuchi, K., Kawamura, T., Iwanari, H., Ohashi, R., Naito, M., Kodama, T., & Hamakubo, T. (2013). Identification of Wilms' tumor 1-associating protein complex and its role in alternative splicing and the cell cycle. *The Journal of biological chemistry*, *288*(46), 33292–33302. https://doi.org/10.1074/jbc.M113.500397

Horowitz, S., Horowitz, A., Nilsen, T. W., Munns, T. W., & Rottman, F. M. (1984). Mapping of N6-methyladenosine residues in bovine prolactin mRNA. *Proceedings of the National Academy of Sciences of the United States of America*, *81*(18), 5667–5671. https://doi.org/10.1073/pnas.81.18.5667

HOTCHKISS R. D. (1948). The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. *The Journal of biological chemistry*, *175*(1), 315–332.

Hsu, P. J., Zhu, Y., Ma, H., Guo, Y., Shi, X., Liu, Y., Qi, M., Lu, Z., Shi, H., Wang, J., Cheng, Y., Luo, G., Dai, Q., Liu, M., Guo, X., Sha, J., Shen, B., & He, C. (2017). Ythdc2 is an N^6-methyladenosine binding protein that regulates mammalian spermatogenesis. *Cell research*, *27*(9), 1115–1127. https://doi.org/10.1038/cr.2017.99

Huang, H., Weng, H., Sun, W., Qin, X., Shi, H., Wu, H., Zhao, B. S., Mesquita, A., Liu, C., Yuan, C. L., Hu, Y. C., Hüttelmaier, S., Skibbe, J. R., Su, R., Deng, X., Dong, L., Sun, M., Li, C., Nachtergaele, S., Wang, Y., … Chen, J. (2018). Recognition of RNA N^6-methyladenosine by IGF2BP proteins enhances mRNA

stability and translation. *Nature cell biology*, *20*(3), 285–295. https://doi.org/10.1038/s41556-018-0045-z

Huang, H., Weng, H., Zhou, K., Wu, T., Zhao, B. S., Sun, M., Chen, Z., Deng, X., Xiao, G., Auer, F., Klemm, L., Wu, H., Zuo, Z., Qin, X., Dong, Y., Zhou, Y., Qin, H., Tao, S., Du, J., Liu, J., … Chen, J. (2019). Histone H3 trimethylation at lysine 36 guides m^{6}A RNA modification co-transcriptionally. *Nature*, *567*(7748), 414–419. https://doi.org/10.1038/s41586-019-1016-7

Huppertz, I., Attig, J., D'Ambrogio, A., Easton, L. E., Sibley, C. R., Sugimoto, Y., Tajnik, M., König, J., & Ule, J. (2014). iCLIP: protein-RNA interactions at nucleotide resolution. *Methods (San Diego, Calif.)*, *65*(3), 274–287. https://doi.org/10.1016/j.ymeth.2013.10.011

Imanishi, M., Tsuji, S., Suda, A., & Futaki, S. (2017). Detection of N^6-methyladenosine based on the methyl-sensitivity of MazF RNA endonuclease. *Chemical communications (Cambridge, England)*, *53*(96), 12930–12933. https://doi.org/10.1039/c7cc07699a

Isken, O., & Maquat, L. E. (2008). The multiple lives of NMD factors: balancing roles in gene and genome regulation. *Nature reviews. Genetics*, *9*(9), 699–712. https://doi.org/10.1038/nrg2402

Itoh, Y., Melamed, E., Yang, X., Kampf, K., Wang, S., Yehya, N., Van Nas, A., Replogle, K., Band, M. R., Clayton, D. F., Schadt, E. E., Lusis, A. J., & Arnold, A. P. (2007). Dosage compensation is less effective in birds than in mammals. *Journal of biology*, *6*(1), 2. https://doi.org/10.1186/jbiol53

Ivanova, I., Much, C., Di Giacomo, M., Azzi, C., Morgan, M., Moreira, P. N., Monahan, J., Carrieri, C., Enright, A. J., & O'Carroll, D. (2017). The RNA m^{6}A Reader YTHDF2 Is Essential for the Post-transcriptional Regulation of the Maternal Transcriptome and Oocyte Competence. *Molecular cell*, *67*(6), 1059–1067.e4. https://doi.org/10.1016/j.molcel.2017.08.003

Jegalian, K., & Page, D. C. (1998). A proposed path by which genes common to mammalian X and Y chromosomes evolve to become X inactivated. *Nature*, *394*(6695), 776–780. https://doi.org/10.1038/29522

Jeppesen, P., & Turner, B. M. (1993). The inactive X chromosome in female mammals is distinguished by a lack of histone H4 acetylation, a cytogenetic marker for gene expression. *Cell*, *74*(2), 281–289. https://doi.org/10.1016/0092-8674(93)90419-q

Jia, G., Fu, Y., Zhao, X., Dai, Q., Zheng, G., Yang, Y., Yi, C., Lindahl, T., Pan, T., Yang, Y. G., & He, C. (2011). N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nature chemical biology*, *7*(12), 885–887. https://doi.org/10.1038/nchembio.687

Jonkhout, N., Tran, J., Smith, M. A., Schonrock, N., Mattick, J. S., & Novoa, E. M. (2017). The RNA modification landscape in human disease. *RNA (New York, N.Y.)*, *23*(12), 1754–1769. https://doi.org/10.1261/rna.063503.117

Kane, S. E., & Beemon, K. (1985). Precise localization of m6A in Rous sarcoma virus RNA reveals clustering of methylation sites: implications for RNA processing. *Molecular and cellular biology*, *5*(9), 2298–2306. https://doi.org/10.1128/mcb.5.9.2298-2306.1985

Kasowitz, S. D., Ma, J., Anderson, S. J., Leu, N. A., Xu, Y., Gregory, B. D., Schultz, R. M., & Wang, P. J. (2018). Nuclear m6A reader YTHDC1 regulates alternative polyadenylation and splicing during mouse oocyte development. *PLoS genetics*, *14*(5), e1007412. https://doi.org/10.1371/journal.pgen.1007412

Kay, G. F., Penny, G. D., Patel, D., Ashworth, A., Brockdorff, N., & Rastan, S. (1993). Expression of Xist during mouse development suggests a role in the initiation of X chromosome inactivation. *Cell*, *72*(2), 171–182. https://doi.org/10.1016/0092-8674(93)90658-d

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., Vågbø, C. B., Kusśnierczyk, A., Klungland, A., Darnell, J. E., Jr, & Darnell, R. B. (2015). A majority of m6A residues are in the last exons, allowing the potential for 3' UTR regulation. *Genes & development*, *29*(19), 2037–2053. https://doi.org/10.1101/gad.269415.115

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, & Darnell, R. B. (2017). m^{6}A mRNA modifications are deposited in nascent pre-mRNA and are not required for splicing but do specify cytoplasmic turnover. *Genes & development*, *31*(10), 990–1006. https://doi.org/10.1101/gad.301036.117

Kennedy, E. M., Bogerd, H. P., Kornepati, A. V., Kang, D., Ghoshal, D., Marshall, J. B., Poling, B. C., Tsai, K., Gokhale, N. S., Horner, S. M., & Cullen, B. R. (2016). Posttranscriptional m(6)A Editing of HIV-1 mRNAs Enhances Viral Gene Expression. *Cell host & microbe*, *19*(5), 675–685. https://doi.org/10.1016/j.chom.2016.04.002

Kharchenko, P. V., Xi, R., & Park, P. J. (2011). Evidence for dosage compensation between the X chromosome and autosomes in mammals. *Nature genetics*, *43*(12), 1167–1172. https://doi.org/10.1038/ng.991

Kierzek, E., & Kierzek, R. (2003). The thermodynamic stability of RNA duplexes and hairpins containing N6-alkyladenosines and 2-methylthio-N6-alkyladenosines. *Nucleic acids research*, *31*(15), 4472–4480. https://doi.org/10.1093/nar/gkg633

Kierzek, E., Malgowska, M., Lisowiec, J., Turner, D. H., Gdaniec, Z., & Kierzek, R. (2014). The contribution of pseudouridine to stabilities and structure of RNAs. *Nucleic acids research*, *42*(5), 3492–3501. https://doi.org/10.1093/nar/gkt1330

Kohlmaier, A., Savarese, F., Lachner, M., Martens, J., Jenuwein, T., & Wutz, A. (2004). A chromosomal memory triggered by Xist regulates histone methylation in X inactivation. *PLoS biology*, *2*(7), E171. https://doi.org/10.1371/journal.pbio.0020171

Korlach, J., & Turner, S. W. (2012). Going beyond five bases in DNA sequencing. *Current opinion in structural biology*, *22*(3), 251–261. https://doi.org/10.1016/j.sbi.2012.04.002

Körtel, N., Rücklé, C., Zhou, Y., Busch, A., Hoch-Kraft, P., Sutandy, F. X. R., Haase, J., Pradhan, M., Musheev, M., Ostareck, D., Ostareck-Lederer, A., Dieterich, C., Hüttelmaier, S., Niehrs, C., Rausch, O., Dominissini, D., König, J., & Zarnack, K. (2021). Deep and accurate detection of m6A RNA modifications using miCLIP2 and m6Aboost machine learning. *Nucleic acids research*, *49*(16), e92. https://doi.org/10.1093/nar/gkab485

König, J., Zarnack, K., Rot, G., Curk, T., Kayikci, M., Zupan, B., Turner, D. J., Luscombe, N. M., & Ule, J. (2010). iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nature structural & molecular biology*, *17*(7), 909–915. https://doi.org/10.1038/nsmb.1838

Kretschmer, J., Rao, H., Hackert, P., Sloan, K. E., Höbartner, C., & Bohnsack, M. T. (2018). The m^{6}A reader protein YTHDC2 interacts with the small ribosomal subunit and the 5'-3' exoribonuclease XRN1. *RNA (New York, N.Y.)*, *24*(10), 1339–1350. https://doi.org/10.1261/rna.064238.117

Lahn, B. T., & Page, D. C. (1999). Four evolutionary strata on the human X chromosome. *Science (New York, N.Y.)*, *286*(5441), 964–967. https://doi.org/10.1126/science.286.5441.964

Larsson, A. J. M., Coucoravas, C., Sandberg, R., & Reinius, B. (2019). X-chromosome upregulation is driven by increased burst frequency. *Nature structural & molecular biology*, *26*(10), 963–969. https://doi.org/10.1038/s41594-019-0306-y

Lavi, S., & Shatkin, A. J. (1975). Methylated simian virus 40-specific RNA from nuclei and cytoplasm of infected BSC-1 cells. *Proceedings of the National Academy of Sciences of the United States of America*, *72*(6), 2012–2016. https://doi.org/10.1073/pnas.72.6.2012

Lee, J. T., & Lu, N. (1999). Targeted mutagenesis of Tsix leads to nonrandom X inactivation. *Cell*, *99*(1), 47–57. https://doi.org/10.1016/s0092-8674(00)80061-6

Lee, F. C. Y., & Ule, J. (2018). Advances in CLIP Technologies for Studies of Protein-RNA Interactions. *Molecular cell*, *69*(3), 354–369. https://doi.org/10.1016/j.molcel.2018.01.005

Leger, A., Amaral, P. P., Pandolfini, L., Capitanchik, C., Capraro, F., Miano, V., Migliori, V., Toolan-Kerr, P., Sideri, T., Enright, A. J., Tzelepis, K., van Werven, F. J., Luscombe, N. M., Barbieri, I., Ule, J., Fitzgerald, T., Birney, E., Leonardi, T., & Kouzarides, T. (2021). RNA modifications detection by comparative Nanopore direct RNA sequencing. *Nature communications*, *12*(1), 7198. https://doi.org/10.1038/s41467-021-27393-3

Lence, T., Akhtar, J., Bayer, M., Schmid, K., Spindler, L., Ho, C. H., Kreim, N., Andrade-Navarro, M. A., Poeck, B., Helm, M., & Roignant, J. Y. (2016). m^{6}A modulates neuronal functions and sex determination in Drosophila. *Nature*, *540*(7632), 242–247. https://doi.org/10.1038/nature20568

Li, F., Zhao, D., Wu, J., & Shi, Y. (2014). Structure of the YTH domain of human YTHDF2 in complex with an m(6)A mononucleotide reveals an aromatic cage for m(6)A recognition. *Cell research*, *24*(12), 1490–1492. https://doi.org/10.1038/cr.2014.153

Li, X., Xiong, X., Wang, K., Wang, L., Shu, X., Ma, S., & Yi, C. (2016). Transcriptome-wide mapping reveals reversible and dynamic N(1)-methyladenosine methylome. *Nature chemical biology*, *12*(5), 311–316. https://doi.org/10.1038/nchembio.2040

(A) Li, Z., Weng, H., Su, R., Weng, X., Zuo, Z., Li, C., Huang, H., Nachtergaele, S., Dong, L., Hu, C., Qin, X., Tang, L., Wang, Y., Hong, G. M., Huang, H., Wang, X., Chen, P., Gurbuxani, S., Arnovitz, S., Li, Y., … Chen, J. (2017). FTO Plays an Oncogenic Role in Acute Myeloid Leukemia as a N^6-Methyladenosine RNA Demethylase. *Cancer cell*, *31*(1), 127–141. https://doi.org/10.1016/j.ccell.2016.11.017

(B) Li, A., Chen, Y. S., Ping, X. L., Yang, X., Xiao, W., Yang, Y., Sun, H. Y., Zhu, Q., Baidya, P., Wang, X., Bhattarai, D. P., Zhao, Y. L., Sun, B. F., & Yang, Y. G. (2017). Cytoplasmic m^{6}A reader YTHDF3 promotes mRNA translation. *Cell research*, *27*(3), 444–447. https://doi.org/10.1038/cr.2017.10

Lichinchi, G., Gao, S., Saletore, Y., Gonzalez, G. M., Bansal, V., Wang, Y., Mason, C. E., & Rana, T. M. (2016). Dynamics of the human and viral m(6)A RNA methylomes during HIV-1 infection of T cells. *Nature microbiology*, *1*, 16011. https://doi.org/10.1038/nmicrobiol.2016.11

Lin, H., Gupta, V., Vermilyea, M. D., Falciani, F., Lee, J. T., O'Neill, L. P., & Turner, B. M. (2007). Dosage compensation in the mouse balances up-regulation and silencing of X-linked genes. *PLoS biology*, *5*(12), e326. https://doi.org/10.1371/journal.pbio.0050326

Lin, H., Halsall, J. A., Antczak, P., O'Neill, L. P., Falciani, F., & Turner, B. M. (2011). Relative overexpression of X-linked genes in mouse embryonic stem cells is consistent with Ohno's hypothesis. *Nature genetics*, *43*(12), 1169–1172. https://doi.org/10.1038/ng.992

Lin, F., Xing, K., Zhang, J., & He, X. (2012). Expression reduction in mammalian X chromosome evolution refutes Ohno's hypothesis of dosage compensation. *Proceedings of the National Academy*

of Sciences of the United States of America, *109*(29), 11752–11757. https://doi.org/10.1073/pnas.1201816109

Linder, B., Grozhik, A. V., Olarerin-George, A. O., Meydan, C., Mason, C. E., & Jaffrey, S. R. (2015). Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. *Nature methods*, *12*(8), 767–772. https://doi.org/10.1038/nmeth.3453

Liu, J., Yue, Y., Han, D., Wang, X., Fu, Y., Zhang, L., Jia, G., Yu, M., Lu, Z., Deng, X., Dai, Q., Chen, W., & He, C. (2014). A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. *Nature chemical biology*, *10*(2), 93–95. https://doi.org/10.1038/nchembio.1432

Liu, N., Dai, Q., Zheng, G., He, C., Parisien, M., & Pan, T. (2015). N(6)-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions. *Nature*, *518*(7540), 560–564. https://doi.org/10.1038/nature14234

Liu, N., & Pan, T. (2015). Probing RNA Modification Status at Single-Nucleotide Resolution in Total RNA. *Methods in enzymology*, *560*, 149–159. https://doi.org/10.1016/bs.mie.2015.03.005

Liu, N., & Pan, T. (2016). N6-methyladenosine–encoded epitranscriptomics. *Nature structural & molecular biology*, *23*(2), 98–102. https://doi.org/10.1038/nsmb.3162

Liu, N., Zhou, K. I., Parisien, M., Dai, Q., Diatchenko, L., & Pan, T. (2017). N6-methyladenosine alters RNA structure to regulate binding of a low-complexity protein. *Nucleic acids research*, *45*(10), 6051–6063. https://doi.org/10.1093/nar/gkx141

Liu, H., Begik, O., Lucas, M. C., Ramirez, J. M., Mason, C. E., Wiener, D., Schwartz, S., Mattick, J. S., Smith, M. A., & Novoa, E. M. (2019). Accurate detection of m^6A RNA modifications in native RNA sequences. *Nature communications*, *10*(1), 4079. https://doi.org/10.1038/s41467-019-11713-9

Liu, C., Sun, H., Yi, Y., Shen, W., Li, K., Xiao, Y., Li, F., Li, Y., Hou, Y., Lu, B., Liu, W., Meng, H., Peng, J., Yi, C., & Wang, J. (2022). Absolute quantification of single-base m^6A methylation in the mammalian transcriptome using GLORI. *Nature biotechnology*, 10.1038/s41587-022-01487-9. Advance online publication. https://doi.org/10.1038

Livernois, A. M., Graves, J. A., & Waters, P. D. (2012). The origin and evolution of vertebrate sex chromosomes and dosage compensation. *Heredity*, *108*(1), 50–58. https://doi.org/10.1038/hdy.2011.106

Lorenz, D. A., Sathe, S., Einstein, J. M., & Yeo, G. W. (2020). Direct RNA sequencing enables m^6A detection in endogenous transcript isoforms at base-specific resolution. *RNA (New York, N.Y.)*, *26*(1), 19–28. https://doi.org/10.1261/rna.072785.119

Louloupi, A., Ntini, E., Conrad, T., & Ørom, U. A. V. (2018). Transient N-6-Methyladenosine Transcriptome Sequencing Reveals a Regulatory Role of m6A in Splicing Efficiency. *Cell reports*, *23*(12), 3429–3437. https://doi.org/10.1016/j.celrep.2018.05.077

Lovejoy, A. F., Riordan, D. P., & Brown, P. O. (2014). Transcriptome-wide mapping of pseudouridines: pseudouridine synthases modify specific mRNAs in S. cerevisiae. *PloS one*, *9*(10), e110799. https://doi.org/10.1371/journal.pone.0110799

Luikenhuis, S., Wutz, A., & Jaenisch, R. (2001). Antisense transcription through the Xist locus mediates Tsix function in embryonic stem cells. *Molecular and cellular biology*, *21*(24), 8512–8520. https://doi.org/10.1128/MCB.21.24.8512-8520.2001

Luo, G. Z., MacQueen, A., Zheng, G., Duan, H., Dore, L. C., Lu, Z., Liu, J., Chen, K., Jia, G., Bergelson, J., & He, C. (2014). Unique features of the m6A methylome in Arabidopsis thaliana. *Nature communications*, *5*, 5630. https://doi.org/10.1038/ncomms6630

Lykke-Andersen, S., & Jensen, T. H. (2015). Nonsense-mediated mRNA decay: an intricate machinery that shapes transcriptomes. *Nature reviews. Molecular cell biology*, *16*(11), 665–677. https://doi.org/10.1038/nrm4063

LYON M. F. (1961). Gene action in the X-chromosome of the mouse (Mus musculus L.). *Nature*, *190*, 372–373. https://doi.org/10.1038/190372a0

Ma, H., Wang, X., Cai, J., Dai, Q., Natchiar, S. K., Lv, R., Chen, K., Lu, Z., Chen, H., Shi, Y. G., Lan, F., Fan, J., Klaholz, B. P., Pan, T., Shi, Y., & He, C. (2019). N^{6-}Methyladenosine methyltransferase ZCCHC4 mediates ribosomal RNA methylation. *Nature chemical biology*, *15*(1), 88–94. https://doi.org/10.1038/s41589-018-0184-3

Machnicka, M. A., Milanowska, K., Osman Oglou, O., Purta, E., Kurkowska, M., Olchowik, A., Januszewski, W., Kalinowski, S., Dunin-Horkawicz, S., Rother, K. M., Helm, M., Bujnicki, J. M., & Grosjean, H. (2013). MODOMICS: a database of RNA modification pathways--2013 update. *Nucleic acids research*, *41*(Database issue), D262–D267. https://doi.org/10.1093/nar/gks1007

Martin, J. L., & McMillan, F. M. (2002). SAM (dependent) I AM: the S-adenosylmethionine-dependent methyltransferase fold. *Current opinion in structural biology*, *12*(6), 783–793. https://doi.org/10.1016/s0959-440x(02)00391-3

Marygold, S. J., Roote, J., Reuter, G., Lambertsson, A., Ashburner, M., Millburn, G. H., Harrison, P. M., Yu, Z., Kenmochi, N., Kaufman, T. C., Leevers, S. J., & Cook, K. R. (2007). The ribosomal protein genes and Minute loci of Drosophila melanogaster. *Genome biology*, *8*(10), R216. https://doi.org/10.1186/gb-2007-8-10-r216

Mauer, J., Luo, X., Blanjoie, A., Jiao, X., Grozhik, A. V., Patil, D. P., Linder, B., Pickering, B. F., Vasseur, J. J., Chen, Q., Gross, S. S., Elemento, O., Debart, F., Kiledjian, M., & Jaffrey, S. R. (2017). Reversible methylation of m^6A_m in the 5' cap controls mRNA stability. *Nature*, *541*(7637), 371–375. https://doi.org/10.1038/nature21022

Mauer, J., Sindelar, M., Despic, V., Guez, T., Hawley, B. R., Vasseur, J. J., Rentmeister, A., Gross, S. S., Pellizzoni, L., Debart, F., Goodarzi, H., & Jaffrey, S. R. (2019). FTO controls reversible m^6Am RNA methylation during snRNA biogenesis. *Nature chemical biology*, *15*(4), 340–347. https://doi.org/10.1038/s41589-019-0231-8

McHugh, C. A., Chen, C. K., Chow, A., Surka, C. F., Tran, C., McDonel, P., Pandya-Jones, A., Blanco, M., Burghard, C., Moradian, A., Sweredoski, M. J., Shishkin, A. A., Su, J., Lander, E. S., Hess, S., Plath, K., & Guttman, M. (2015). The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature*, *521*(7551), 232–236. https://doi.org/10.1038/nature14443

Meissner, A., Mikkelsen, T. S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B. E., Nusbaum, C., Jaffe, D. B., Gnirke, A., Jaenisch, R., & Lander, E. S. (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature*, *454*(7205), 766–770. https://doi.org/10.1038/nature07107

Merkurjev, D., Hong, W. T., Iida, K., Oomoto, I., Goldie, B. J., Yamaguti, H., Ohara, T., Kawaguchi, S. Y., Hirano, T., Martin, K. C., Pellegrini, M., & Wang, D. O. (2018). Synaptic N^6-methyladenosine (m^6A) epitranscriptome reveals functional partitioning of localized transcripts. *Nature neuroscience*, *21*(7), 1004–1014. https://doi.org/10.1038/s41593-018-0173-6

Meroueh, M., Grohar, P. J., Qiu, J., SantaLucia, J., Jr, Scaringe, S. A., & Chow, C. S. (2000). Unique structural and stabilizing roles for the individual pseudouridine residues in the 1920 region of Escherichia coli 23S rRNA. *Nucleic acids research*, *28*(10), 2075–2083. https://doi.org/10.1093/nar/28.10.2075

Meyer, K. D., Saletore, Y., Zumbo, P., Elemento, O., Mason, C. E., & Jaffrey, S. R. (2012). Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell*, *149*(7), 1635–1646. https://doi.org/10.1016/j.cell.2012.05.003

Meyer K. D. (2019). DART-seq: an antibody-free method for global m^6A detection. *Nature methods*, *16*(12), 1275–1280. https://doi.org/10.1038/s41592-019-0570-0

Mohandas, T., Sparkes, R. S., & Shapiro, L. J. (1981). Reactivation of an inactive human X chromosome: evidence for X inactivation by DNA methylation. *Science (New York, N.Y.)*, *211*(4480), 393–396. https://doi.org/10.1126/science.6164095

Moindrot, B., Cerase, A., Coker, H., Masui, O., Grijzenhout, A., Pintacuda, G., Schermelleh, L., Nesterova, T. B., & Brockdorff, N. (2015). A Pooled shRNA Screen Identifies Rbm15, Spen, and Wtap as Factors Required for Xist RNA-Mediated Silencing. *Cell reports*, *12*(4), 562–572. https://doi.org/10.1016/j.celrep.2015.06.053

Morrill, S. A., & Amon, A. (2019). Why haploinsufficiency persists. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(24), 11866–11871. https://doi.org/10.1073/pnas.1900437116

Motorin, Y., & Helm, M. (2022). RNA nucleotide methylation: 2021 update. *Wiley interdisciplinary reviews. RNA*, *13*(1), e1691. https://doi.org/10.1002/wrna.1691

Mu, H., Zhang, T., Yang, Y., Zhang, D., Gao, J., Li, J., Yue, L., Gao, D., Shi, B., Han, Y., Zhong, L., Chen, X., Wang, Z. B., Lin, Z., Tong, M. H., Sun, Q. Y., Yang, Y. G., & Han, J. (2021). METTL3-mediated mRNA N^6-methyladenosine is required for oocyte and follicle development in mice. *Cell death & disease*, *12*(11), 989. https://doi.org/10.1038/s41419-021-04272-9

Mueller, J. L., Skaletsky, H., Brown, L. G., Zaghlul, S., Rock, S., Graves, T., Auger, K., Warren, W. C., Wilson, R. K., & Page, D. C. (2013). Independent specialization of the human and mouse X chromosomes for the male germ line. *Nature genetics*, *45*(9), 1083–1087. https://doi.org/10.1038/ng.2705

Nabet, B., Roberts, J. M., Buckley, D. L., Paulk, J., Dastjerdi, S., Yang, A., Leggett, A. L., Erb, M. A., Lawlor, M. A., Souza, A., Scott, T. G., Vittori, S., Perry, J. A., Qi, J., Winter, G. E., Wong, K. K., Gray, N. S., & Bradner, J. E. (2018). The dTAG system for immediate and target-specific protein degradation. *Nature chemical biology*, *14*(5), 431–441. https://doi.org/10.1038/s41589-018-0021-8

Navaratnam, N., Morrison, J. R., Bhattacharya, S., Patel, D., Funahashi, T., Giannoni, F., Teng, B. B., Davidson, N. O., & Scott, J. (1993). The p27 catalytic subunit of the apolipoprotein B mRNA editing enzyme is a cytidine deaminase. *The Journal of biological chemistry*, *268*(28), 20709–20712.

Navarro, P., & Avner, P. (2010). An embryonic story: analysis of the gene regulative network controlling Xist expression in mouse embryonic stem cells. *BioEssays : news and reviews in molecular, cellular and developmental biology*, *32*(7), 581–588. https://doi.org/10.1002/bies.201000019

Navarro, P., Oldfield, A., Legoupi, J., Festuccia, N., Dubois, A., Attia, M., Schoorlemmer, J., Rougeulle, C., Chambers, I., & Avner, P. (2010). Molecular coupling of Tsix regulation and pluripotency. *Nature*, *468*(7322), 457–460. https://doi.org/10.1038/nature09496

Nguyen, D. K., & Disteche, C. M. (2006). Dosage compensation of the active X chromosome in mammals. *Nature genetics*, *38*(1), 47–53. https://doi.org/10.1038/ng1705

Nicholson, J. M., & Cimini, D. (2013). Cancer karyotypes: survival of the fittest. *Frontiers in oncology*, *3*, 148. https://doi.org/10.3389/fonc.2013.00148

Nishimura, K., Fukagawa, T., Takisawa, H., Kakimoto, T., & Kanemaki, M. (2009). An auxin-based degron system for the rapid depletion of proteins in nonplant cells. *Nature methods*, *6*(12), 917–922. https://doi.org/10.1038/nmeth.1401

Nora, E. P., Lajoie, B. R., Schulz, E. G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N. L., Meisig, J., Sedat, J., Gribnau, J., Barillot, E., Blüthgen, N., Dekker, J., & Heard, E. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*, *485*(7398), 381–385. https://doi.org/10.1038/nature11049

Ohno, S. (1966). Time and Place of Action of Sex-determining Factors in Ontogeny. In: Sex Chromosomes and Sex-Linked Genes. Monographs on Endocrinology, vol 1. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-88178-7_15

Ougland, R., Zhang, C. M., Liiv, A., Johansen, R. F., Seeberg, E., Hou, Y. M., Remme, J., & Falnes, P. Ø. (2004). AlkB restores the biological function of mRNA and tRNA inactivated by chemical methylation. *Molecular cell*, *16*(1), 107–116. https://doi.org/10.1016/j.molcel.2004.09.002

Pan T. (2018). Modifications and functional genomics of human transfer RNA. *Cell research*, *28*(4), 395–404. https://doi.org/10.1038/s41422-018-0013-y

Panning B. (2008). X-chromosome inactivation: the molecular basis of silencing. *Journal of biology*, *7*(8), 30. https://doi.org/10.1186/jbiol95

Papp, B., Pál, C., & Hurst, L. D. (2003). Dosage sensitivity and the evolution of gene families in yeast. *Nature*, *424*(6945), 194–197. https://doi.org/10.1038/nature01771

Patil, D. P., Chen, C. K., Pickering, B. F., Chow, A., Jackson, C., Guttman, M., & Jaffrey, S. R. (2016). m(6)A RNA methylation promotes XIST-mediated transcriptional repression. *Nature*, *537*(7620), 369–373. https://doi.org/10.1038/nature19342

Patil, D. P., Pickering, B. F., & Jaffrey, S. R. (2018). Reading m^6A in the Transcriptome: m^6A-Binding Proteins. *Trends in cell biology*, *28*(2), 113–127. https://doi.org/10.1016/j.tcb.2017.10.001

Payer, B., & Lee, J. T. (2008). X chromosome dosage compensation: how mammals keep the balance. *Annual review of genetics*, *42*, 733–772. https://doi.org/10.1146/annurev.genet.42.110807.091711

Pendleton, K. E., Chen, B., Liu, K., Hunter, O. V., Xie, Y., Tu, B. P., & Conrad, N. K. (2017). The U6 snRNA m^6A Methyltransferase METTL16 Regulates SAM Synthetase Intron Retention. *Cell*, *169*(5), 824–835.e14. https://doi.org/10.1016/j.cell.2017.05.003

Penny, G. D., Kay, G. F., Sheardown, S. A., Rastan, S., & Brockdorff, N. (1996). Requirement for Xist in X chromosome inactivation. *Nature*, *379*(6561), 131–137. https://doi.org/10.1038/379131a0

Perry, R.P., Kelley, D.E., 1974. Existence of methylated messenger RNA in mouse L cells.Cell 1, 37–42. https://doi.org/10.1016/0092-8674(74)90153-6

Pessia, E., Makino, T., Bailly-Bechet, M., McLysaght, A., & Marais, G. A. (2012). Mammalian X chromosome inactivation evolved as a dosage-compensation mechanism for dosage-sensitive genes on the X chromosome. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(14), 5346–5351. https://doi.org/10.1073/pnas.1116763109

Phatnani, H. P., & Greenleaf, A. L. (2006). Phosphorylation and functions of the RNA polymerase II CTD. *Genes & development*, *20*(21), 2922–2936. https://doi.org/10.1101/gad.1477006

Ping, X. L., Sun, B. F., Wang, L., Xiao, W., Yang, X., Wang, W. J., Adhikari, S., Shi, Y., Lv, Y., Chen, Y. S., Zhao, X., Li, A., Yang, Y., Dahal, U., Lou, X. M., Liu, X., Huang, J., Yuan, W. P., Zhu, X. F., Cheng, T., … Yang, Y. G. (2014). Mammalian WTAP is a regulatory subunit of the RNA N6-methyladenosine methyltransferase. *Cell research*, *24*(2), 177–189. https://doi.org/10.1038/cr.2014.3

Pinheiro, I., & Heard, E. (2017). X chromosome inactivation: new players in the initiation of gene silencing. *F1000Research*, *6*, F1000 Faculty Rev-344. https://doi.org/10.12688/f1000research.10707.1

Plenefisch, J. D., DeLong, L., & Meyer, B. J. (1989). Genes that implement the hermaphrodite mode of dosage compensation in Caenorhabditis elegans. *Genetics*, *121*(1), 57–76. https://doi.org/10.1093/genetics/121.1.57

Pogo, B. G., Allfrey, V. G., & Mirsky, A. E. (1966). RNA synthesis and histone acetylation during the course of gene activation in lymphocytes. *Proceedings of the National Academy of Sciences of the United States of America*, *55*(4), 805–812. https://doi.org/10.1073/pnas.55.4.805

Probst, A. V., Santos, F., Reik, W., Almouzni, G., & Dean, W. (2007). Structural differences in centromeric heterochromatin are spatially reconciled on fertilisation in the mouse zygote. *Chromosoma*, *116*(4), 403–415. https://doi.org/10.1007/s00412-007-0106-8

Ramazi, S., & Zahiri, J. (2021). Posttranslational modifications in proteins: resources, tools and prediction methods. *Database : the journal of biological databases and curation*, *2021*, baab012. https://doi.org/10.1093/database/baab012

Rastan S. (1983). Non-random X-chromosome inactivation in mouse X-autosome translocation embryos--location of the inactivation centre. *Journal of embryology and experimental morphology*, *78*, 1–22.

Rastan, S., & Robertson, E. J. (1985). X-chromosome deletions in embryo-derived (EK) cell lines associated with lack of X-chromosome inactivation. *Journal of embryology and experimental morphology*, *90*, 379–388.

Rice, W. R. (1996). Evolution of the Y Sex Chromosome in Animals: Y chromosomes evolve through the degeneration of autosomes, *BioScience*, Volume 46, Issue 5, May 1996, Pages 331–343, https://doi.org/10.2307/1312947

Ries, R. J., Zaccara, S., Klein, P., Olarerin-George, A., Namkoong, S., Pickering, B. F., Patil, D. P., Kwak, H., Lee, J. H., & Jaffrey, S. R. (2019). m^{6}A enhances the phase separation potential of mRNA. *Nature*, *571*(7765), 424–428. https://doi.org/10.1038/s41586-019-1374-1

Roost, C., Lynch, S. R., Batista, P. J., Qu, K., Chang, H. Y., & Kool, E. T. (2015). Structure and thermodynamics of N6-methyladenosine in RNA: a spring-loaded base modification. *Journal of the American Chemical Society*, *137*(5), 2107–2115. https://doi.org/10.1021/ja513080v

Rottman, F., Shatkin, A. J., & Perry, R. P. (1974). Sequences containing methylated nucleotides at the 5' termini of messenger RNAs: possible implications for processing. *Cel*

Roundtree, I. A., Luo, G. Z., Zhang, Z., Wang, X., Zhou, T., Cui, Y., Sha, J., Huang, X., Guerrero, L., Xie, P., He, E., Shen, B., & He, C. (2017). YTHDC1 mediates nuclear export of N^6-methyladenosine methylated mRNAs. *eLife*, *6*, e31311. https://doi.org/10.7554/eLife.31311*l*, *3*(3), 197–199. https://doi.org/10.1016/0092-8674(74)90131-7

Sanchez-Pulido, L., & Andrade-Navarro, M. A. (2007). The FTO (fat mass and obesity associated) gene codes for a novel member of the non-heme dioxygenase superfamily. *BMC biochemistry*, *8*, 23. https://doi.org/10.1186/1471-2091-8-23

Santos, F., Peters, A. H., Otte, A. P., Reik, W., & Dean, W. (2005). Dynamic chromatin modifications characterise the first cell cycle in mouse embryos. *Developmental biology*, *280*(1), 225–236. https://doi.org/10.1016/j.ydbio.2005.01.025

Saxonov, S., Berg, P., & Brutlag, D. L. (2006). A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(5), 1412–1417. https://doi.org/10.1073/pnas.0510310103

SCHUSTER, H., & WILHELM, R. C. (1963). Reaction differences between tobacco mosaic virus and its free ribonucleic acid with nitrous acid. *Biochimica et biophysica acta*, *68*, 554–560. https://doi.org/10.1016/0006-3002(63)90184-7

Schwartz, S., Agarwala, S. D., Mumbach, M. R., Jovanovic, M., Mertins, P., Shishkin, A., Tabach, Y., Mikkelsen, T. S., Satija, R., Ruvkun, G., Carr, S. A., Lander, E. S., Fink, G. R., & Regev, A. (2013). High-resolution mapping reveals a conserved, widespread, dynamic mRNA methylation program in yeast meiosis. *Cell*, *155*(6), 1409–1421. https://doi.org/10.1016/j.cell.2013.10.047

Schwartz, S., Bernstein, D. A., Mumbach, M. R., Jovanovic, M., Herbst, R. H., León-Ricardo, B. X., Engreitz, J. M., Guttman, M., Satija, R., Lander, E. S., Fink, G., & Regev, A. (2014). Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. *Cell*, *159*(1), 148–162. https://doi.org/10.1016/j.cell.2014.08.028

Sears, R. M., May, D. G., & Roux, K. J. (2019). BioID as a Tool for Protein-Proximity Labeling in Living Cells. *Methods in molecular biology (Clifton, N.J.)*, *2012*, 299–313. https://doi.org/10.1007/978-1-4939-9546-2_15

Shafik, A. M., Zhang, F., Guo, Z., Dai, Q., Pajdzik, K., Li, Y., Kang, Y., Yao, B., Wu, H., He, C., Allen, E. G., Duan, R., & Jin, P. (2021). N6-methyladenosine dynamics in neurodevelopment and aging, and its potential role in Alzheimer's disease. *Genome biology*, *22*(1), 17. https://doi.org/10.1186/s13059-020-02249-z

Shan, X., Tashiro, H., & Lin, C. L. (2003). The identification and characterization of oxidized RNAs in Alzheimer's disease. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *23*(12), 4913–4921. https://doi.org/10.1523/JNEUROSCI.23-12-04913.2003

Shapiro, R., & Pohl, S. H. (1968). The reaction of ribonucleosides with nitrous acid. Side products and kinetics. *Biochemistry*, *7*(1), 448–455. https://doi.org/10.1021/bi00841a057

Shen, F., Huang, W., Huang, J. T., Xiong, J., Yang, Y., Wu, K., Jia, G. F., Chen, J., Feng, Y. Q., Yuan, B. F., & Liu, S. M. (2015). Decreased N(6)-methyladenosine in peripheral blood RNA from diabetic patients is associated with FTO expression rather than ALKBH5. *The Journal of clinical endocrinology and metabolism*, *100*(1), E148–E154. https://doi.org/10.1210/jc.2014-1893

Shi, H., Wang, X., Lu, Z., Zhao, B. S., Ma, H., Hsu, P. J., Liu, C., & He, C. (2017). YTHDF3 facilitates translation and decay of N^6-methyladenosine-modified RNA. *Cell research*, *27*(3), 315–328. https://doi.org/10.1038/cr.2017.15

Shi, H., Wei, J., & He, C. (2019). Where, When, and How: Context-Dependent Functions of RNA Methylation Writers, Readers, and Erasers. *Molecular cell*, *74*(4), 640–650. https://doi.org/10.1016/j.molcel.2019.04.025

Shima, H., Matsumoto, M., Ishigami, Y., Ebina, M., Muto, A., Sato, Y., Kumagai, S., Ochiai, K., Suzuki, T., & Igarashi, K. (2017). S-Adenosylmethionine Synthesis Is Regulated by Selective N^6-Adenosine Methylation and mRNA Degradation Involving METTL16 and YTHDC1. *Cell reports*, *21*(12), 3354–3363. https://doi.org/10.1016/j.celrep.2017.11.092

Shu, X., Cao, J., Cheng, M., Xiang, S., Gao, M., Li, T., Ying, X., Wang, F., Yue, Y., Lu, Z., Dai, Q., Cui, X., Ma, L., Wang, Y., He, C., Feng, X., & Liu, J. (2020). A metabolic labeling method detects m^6A transcriptome-wide at single base resolution. *Nature chemical biology*, *16*(8), 887–895. https://doi.org/10.1038/s41589-020-0526-9

Slama, K., Galliot, A., Weichmann, F., Hertler, J., Feederle, R., Meister, G., & Helm, M. (2019). Determination of enrichment factors for modified RNA in MeRIP experiments. *Methods (San Diego, Calif.)*, *156*, 102–109. https://doi.org/10.1016/j.ymeth.2018.10.020

Slobodin, B., Han, R., Calderone, V., Vrielink, J. A. F. O., Loayza-Puch, F., Elkon, R., & Agami, R. (2017). Transcription Impacts the Efficiency of mRNA Translation via Co-transcriptional N6-adenosine Methylation. *Cell*, *169*(2), 326–337.e12. https://doi.org/10.1016/j.cell.2017.03.031

Smith, A. M., Jain, M., Mulroney, L., Garalde, D. R., & Akeson, M. (2019). Reading canonical and modified nucleobases in 16S ribosomal RNA using nanopore native RNA sequencing. *PloS one*, *14*(5), e0216709. https://doi.org/10.1371/journal.pone.0216709

Sprinzl, M., & Vassilenko, K. S. (2005). Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic acids research*, *33*(Database issue), D139–D140. https://doi.org/10.1093/nar/gki012

Squires, J. E., Patel, H. R., Nousch, M., Sibbritt, T., Humphreys, D. T., Parker, B. J., Suter, C. M., & Preiss, T. (2012). Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA. *Nucleic acids research*, *40*(11), 5023–5033. https://doi.org/10.1093/nar/gks144

Stepanenko, A. A., & Dmitrenko, V. V. (2015). HEK293 in cell biology and cancer research: phenotype, karyotype, tumorigenicity, and stress-induced genome-phenotype evolution. *Gene*, *569*(2), 182–190. https://doi.org/10.1016/j.gene.2015.05.065

Stewart, C. L., Stuhlmann, H., Jähner, D., & Jaenisch, R. (1982). De novo methylation, expression, and infectivity of retroviral genomes introduced into embryonal carcinoma cells. Proceedings of the National Academy of Sciences of the United States of America, 79(13), 4098–4102. https://doi.org/10.1073/pnas.79.13.4098

Sugimoto, Y., König, J., Hussain, S., Zupan, B., Curk, T., Frye, M., & Ule, J. (2012). Analysis of CLIP and iCLIP methods for nucleotide-resolution studies of protein-RNA interactions. *Genome biology*, *13*(8), R67. https://doi.org/10.1186/gb-2012-13-8-r67

Sun, L., Fazal, F. M., Li, P., Broughton, J. P., Lee, B., Tang, L., Huang, W., Kool, E. T., Chang, H. Y., & Zhang, Q. C. (2019). RNA structure maps across mammalian cellular compartments. *Nature structural & molecular biology*, *26*(4), 322–330. https://doi.org/10.1038/s41594-019-0200-7

Takagi, N., Wake, N., & Sasaki, M. (1978). Cytologic evidence for preferential inactivation of the paternally derived X chromosome in XX mouse blastocysts. *Cytogenetics and cell genetics*, *20*(1-6), 240–248. https://doi.org/10.1159/000130856

Talon, I., Janiszewski, A., Theeuwes, B., Lefevre, T., Song, J., Bervoets, G., Vanheer, L., De Geest, N., Poovathingal, S., Allsop, R., Marine, J. C., Rambow, F., Voet, T., & Pasque, V. (2021). Enhanced chromatin accessibility contributes to X chromosome dosage compensation in mammals. *Genome biology*, *22*(1), 302. https://doi.org/10.1186/s13059-021-02518-5

Tang, Y. C., & Amon, A. (2013). Gene copy-number alterations: a cost-benefit analysis. *Cell*, *152*(3), 394–405. https://doi.org/10.1016/j.cell.2012.11.043

Taoka, M., Nobe, Y., Yamaki, Y., Sato, K., Ishikawa, H., Izumikawa, K., Yamauchi, Y., Hirota, K., Nakayama, H., Takahashi, N., & Isobe, T. (2018). Landscape of the complete RNA chemical modifications in the human 80S ribosome. *Nucleic acids research*, *46*(18), 9289–9298. https://doi.org/10.1093/nar/gky811

Tegowski, M., Flamand, M. N., & Meyer, K. D. (2022). scDART-seq reveals distinct m^{6}A signatures and mRNA methylation heterogeneity in single cells. *Molecular cell*, *82*(4), 868–878.e10. https://doi.org/10.1016/j.molcel.2021.12.038

Teixeira, D., & Parker, R. (2007). Analysis of P-body assembly in Saccharomyces cerevisiae. *Molecular biology of the cell*, *18*(6), 2274–2287. https://doi.org/10.1091/mbc.e07-03-0199

Temin, H. M., & Mizutani, S. (1970). RNA-dependent DNA polymerase in virions of Rous sarcoma virus. *Nature*, *226*(5252), 1211–1213. https://doi.org/10.1038/2261211a0

Tirumuru, N., Zhao, B. S., Lu, W., Lu, Z., He, C., & Wu, L. (2016). N(6)-methyladenosine of HIV-1 RNA regulates viral infection and HIV-1 Gag protein expression. *eLife*, *5*, e15528. https://doi.org/10.7554/eLife.15528

Torres, E. M., Williams, B. R., & Amon, A. (2008). Aneuploidy: cells losing their balance. *Genetics*, *179*(2), 737–746. https://doi.org/10.1534/genetics.108.090878

Ule, J., Jensen, K. B., Ruggiu, M., Mele, A., Ule, A., & Darnell, R. B. (2003). CLIP identifies Nova-regulated RNA networks in the brain. *Science (New York, N.Y.)*, *302*(5648), 1212–1215. https://doi.org/10.1126/science.1090095

Uzonyi, A., Dierks, D., Nir, R., Kwon, O. S., Toth, U., Barbosa, I., Burel, C., Brandis, A., Rossmanith, W., Le Hir, H., Slobodin, B., & Schwartz, S. (2022). Exclusion of m6A from splice-site proximal regions by the exon junction complex dictates m6A topologies and mRNA stability. *Molecular cell*, S1097-2765(22)01205-9. Advance online publication. https://doi.org/10.1016/j.molcel.2022.12.026

Vaidyanathan, P. P., AlSadhan, I., Merriman, D. K., Al-Hashimi, H. M., & Herschlag, D. (2017). Pseudouridine and *N*^{6}-methyladenosine modifications weaken PUF protein/RNA interactions. *RNA (New York, N.Y.)*, *23*(5), 611–618. https://doi.org/10.1261/rna.060053.116

Van Nostrand, E. L., Pratt, G. A., Shishkin, A. A., Gelboin-Burkhart, C., Fang, M. Y., Sundararaman, B., Blue, S. M., Nguyen, T. B., Surka, C., Elkins, K., Stanton, R., Rigo, F., Guttman, M., & Yeo, G. W. (2016). Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nature methods*, *13*(6), 508–514. https://doi.org/10.1038/nmeth.3810

van Tran, N., Ernst, F. G. M., Hawley, B. R., Zorbas, C., Ulryck, N., Hackert, P., Bohnsack, K. E., Bohnsack, M. T., Jaffrey, S. R., Graille, M., & Lafontaine, D. L. J. (2019). The human 18S rRNA m6A methyltransferase METTL5 is stabilized by TRMT112. *Nucleic acids research*, *47*(15), 7719–7733. https://doi.org/10.1093/nar/gkz619

Vicoso, B., & Charlesworth, B. (2006). Evolution on the X chromosome: unusual patterns and processes. *Nature reviews. Genetics*, *7*(8), 645–653. https://doi.org/10.1038/nrg1914

Wan, Y., Tang, K., Zhang, D., Xie, S., Zhu, X., Wang, Z., & Lang, Z. (2015). Transcriptome-wide high-throughput deep m(6)A-seq reveals unique differential m(6)A methylation patterns between three organs in Arabidopsis thaliana. *Genome biology*, *16*, 272. https://doi.org/10.1186/s13059-015-0839-2

Wang, X., Lu, Z., Gomez, A., Hon, G. C., Yue, Y., Han, D., Fu, Y., Parisien, M., Dai, Q., Jia, G., Ren, B., Pan, T., & He, C. (2014). N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature*, *505*(7481), 117–120. https://doi.org/10.1038/nature12730

Wang, X., Zhao, B. S., Roundtree, I. A., Lu, Z., Han, D., Ma, H., Weng, X., Chen, K., Shi, H., & He, C. (2015). N(6)-methyladenosine Modulates Messenger RNA Translation Efficiency. *Cell*, *161*(6), 1388–1399. https://doi.org/10.1016/j.cell.2015.05.014

(A) Wang, P., Doxtader, K. A., & Nam, Y. (2016). Structural Basis for Cooperative Function of Mettl3 and Mettl14 Methyltransferases. *Molecular cell*, *63*(2), 306–317. https://doi.org/10.1016/j.molcel.2016.05.041

(B) Wang, X., Feng, J., Xue, Y., Guan, Z., Zhang, D., Liu, Z., Gong, Z., Wang, Q., Huang, J., Tang, C., Zou, T., & Yin, P. (2016). Structural basis of N(6)-adenosine methylation by the METTL3-METTL14 complex. *Nature*, *534*(7608), 575–578. https://doi.org/10.1038/nature18298

Wang, P., Tang, W., Li, Z., Zou, Z., Zhou, Y., Li, R., Xiong, T., Wang, J., & Zou, P. (2019). Mapping spatial transcriptome with light-activated proximity-dependent RNA labeling. *Nature chemical biology*, *15*(11), 1110–1119. https://doi.org/10.1038/s41589-019-0368-5

(A) Wang, Y., Xiao, Y., Dong, S., Yu, Q., & Jia, G. (2020). Antibody-free enzyme-assisted chemical approach for detection of N^6-methyladenosine. *Nature chemical biology*, *16*(8), 896–903. https://doi.org/10.1038/s41589-020-0525-x

(B) Wang, Z. Y., Leushkin, E., Liechti, A., Ovchinnikova, S., Mößinger, K., Brüning, T., Rummel, C., Grützner, F., Cardoso-Moreira, M., Janich, P., Gatfield, D., Diagouraga, B., de Massy, B., Gill, M. E., Peters, A. H. F. M., Anders, S., & Kaessmann, H. (2020). Transcriptome and translatome co-evolution in mammals. *Nature*, *588*(7839), 642–647. https://doi.org/10.1038/s41586-020-2899-z

Warda, A. S., Kretschmer, J., Hackert, P., Lenz, C., Urlaub, H., Höbartner, C., Sloan, K. E., & Bohnsack, M. T. (2017). Human METTL16 is a *N*^{6}-methyladenosine (m^6A) methyltransferase that targets pre-mRNAs and various non-coding RNAs. *EMBO reports*, *18*(11), 2004–2014. https://doi.org/10.15252/embr.201744940

A) Wei, C. M., Gershowitz, A., & Moss, B. (1975). Methylated nucleotides block 5' terminus of HeLa cell messenger RNA. *Cell*, *4*(4), 379–386. https://doi.org/10.1016/0092-8674(75)90158-0

B) Wei, C., Gershowitz, A., & Moss, B. (1975). N6, O2'-dimethyladenosine a novel methylated ribonucleoside next to the 5' terminal of animal cell and virus mRNAs. *Nature*, *257*(5523), 251–253. https://doi.org/10.1038/257251a0

Wei, C. M., Gershowitz, A., & Moss, B. (1976). 5'-Terminal and internal methylated nucleotide sequences in HeLa cell mRNA. *Biochemistry*, *15*(2), 397–401. https://doi.org/10.1021/bi00647a024

Wei, W., Ji, X., Guo, X., & Ji, S. (2017). Regulatory Role of N^6 -methyladenosine (m^6 A) Methylation in RNA Processing and Human Diseases. *Journal of cellular biochemistry*, *118*(9), 2534–2543. https://doi.org/10.1002/jcb.25967

Wei, J., Liu, F., Lu, Z., Fei, Q., Ai, Y., He, P. C., Shi, H., Cui, X., Su, R., Klungland, A., Jia, G., Chen, J., & He, C. (2018). Differential m^{6}A, m^{6}A$_m$, and m^{1}A Demethylation Mediated by FTO in the Cell Nucleus and Cytoplasm. *Molecular cell*, *71*(6), 973–985.e5. https://doi.org/10.1016/j.molcel.2018.08.011

Wei, G., Almeida, M., Pintacuda, G., Coker, H., Bowness, J. S., Ule, J., & Brockdorff, N. (2021). Acute depletion of METTL3 implicates *N*^{6}-methyladenosine in alternative intron/exon inclusion in the nascent transcriptome. *Genome research*, *31*(8), 1395–1408. https://doi.org/10.1101/gr.271635.120

Weissinger, R., Heinold, L., Akram, S., Jansen, R. P., & Hermesh, O. (2021). RNA Proximity Labeling: A New Detection Tool for RNA-Protein Interactions. *Molecules (Basel, Switzerland)*, *26*(8), 2270. https://doi.org/10.3390/molecules26082270

Wen, J., Lv, R., Ma, H., Shen, H., He, C., Wang, J., Jiao, F., Liu, H., Yang, P., Tan, L., Lan, F., Shi, Y. G., He, C., Shi, Y., & Diao, J. (2018). Zc3h13 Regulates Nuclear RNA m^{6}A Methylation and Mouse Embryonic Stem Cell Self-Renewal. *Molecular cell*, *69*(6), 1028–1038.e6. https://doi.org/10.1016/j.molcel.2018.02.015

Wojtas, M. N., Pandey, R. R., Mendel, M., Homolka, D., Sachidanandam, R., & Pillai, R. S. (2017). Regulation of m^{6}A Transcripts by the 3'→5' RNA Helicase YTHDC2 Is Essential for a Successful Meiotic Program in the Mammalian Germline. *Molecular cell*, *68*(2), 374–387.e12. https://doi.org/10.1016/j.molcel.2017.09.021

Workman, R. E., Tang, A. D., Tang, P. S., Jain, M., Tyson, J. R., Razaghi, R., Zuzarte, P. C., Gilpatrick, T., Payne, A., Quick, J., Sadowski, N., Holmes, N., de Jesus, J. G., Jones, K. L., Soulette, C. M., Snutch, T. P., Loman, N., Paten, B., Loose, M., Simpson, J. T., … Timp, W. (2019). Nanopore native RNA sequencing of a human poly(A) transcriptome. *Nature methods*, *16*(12), 1297–1305. https://doi.org/10.1038/s41592-019-0617-2

Wu, G., Adachi, H., Ge, J., Stephenson, D., Query, C. C., & Yu, Y. T. (2016). Pseudouridines in U2 snRNA stimulate the ATPase activity of Prp5 during spliceosome assembly. *The EMBO journal*, *35*(6), 654–667. https://doi.org/10.15252/embj.201593113

Wu, B., Su, S., Patil, D. P., Liu, H., Gan, J., Jaffrey, S. R., & Ma, J. (2018). Molecular basis for the specific and multivariant recognitions of RNA substrates by human hnRNP A2/B1. *Nature communications*, *9*(1), 420. https://doi.org/10.1038/s41467-017-02770-z

WYATT G. R. (1950). Occurrence of 5-methylcytosine in nucleic acids. *Nature*, *166*(4214), 237–238. https://doi.org/10.1038/166237b0

Xia, H., Zhong, C., Wu, X., Chen, J., Tao, B., Xia, X., Shi, M., Zhu, Z., Trudeau, V. L., & Hu, W. (2018). *Mettl3* Mutation Disrupts Gamete Maturation and Reduces Fertility in Zebrafish. *Genetics*, *208*(2), 729–743. https://doi.org/10.1534/genetics.117.300574

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., Wang, X., Ma, H. L., Huang, C. M., Yang, Y., Huang, N., Jiang, G. B., Wang, H. L., Zhou, Q., Wang, X. J., Zhao, Y. L., … Yang, Y. G. (2016). Nuclear m(6)A Reader YTHDC1 Regulates mRNA Splicing. *Molecular cell*, *61*(4), 507–519. https://doi.org/10.1016/j.molcel.2016.01.012

Xiang, S., Liu, K., Yan, Z., Zhang, Y., & Sun, Z. (2016). RNAMethPre: A Web Server for the Prediction and Query of mRNA m6A Sites. *PloS one*, *11*(10), e0162707. https://doi.org/10.1371/journal.pone.0162707

Xiao, Y., Wang, Y., Tang, Q., Wei, L., Zhang, X., & Jia, G. (2018). An Elongation- and Ligation-Based qPCR Amplification Method for the Radiolabeling-Free Detection of Locus-Specific N^6 -

Methyladenosine Modification. *Angewandte Chemie (International ed. in English)*, *57*(49), 15995–16000. https://doi.org/10.1002/anie.201807942

Xiong, Y., Chen, X., Chen, Z., Wang, X., Shi, S., Wang, X., Zhang, J., & He, X. (2010). RNA sequencing shows no dosage compensation of the active X-chromosome. *Nature genetics*, *42*(12), 1043–1047. https://doi.org/10.1038/ng.711

Xu, C., Liu, K., Ahmed, H., Loppnau, P., Schapira, M., & Min, J. (2015). Structural Basis for the Discriminative Recognition of N6-Methyladenosine RNA by the Human YT521-B Homology Domain Family of Proteins. *The Journal of biological chemistry*, *290*(41), 24902–24913. https://doi.org/10.1074/jbc.M115.680389

Xu, L., Liu, X., Sheng, N., Oo, K. S., Liang, J., Chionh, Y. H., Xu, J., Ye, F., Gao, Y. G., Dedon, P. C., & Fu, X. Y. (2017). Three distinct 3-methylcytidine (m^3C) methyltransferases modify tRNA and mRNA in mice and humans. *The Journal of biological chemistry*, *292*(35), 14695–14703. https://doi.org/10.1074/jbc.M117.798298

Yang, X., Triboulet, R., Liu, Q., Sendinc, E., & Gregory, R. I. (2022). Exon junction complex shapes the m^6A epitranscriptome. *Nature communications*, *13*(1), 7904. https://doi.org/10.1038/s41467-022-35643-1

Yankova, E., Blackaby, W., Albertella, M., Rak, J., De Braekeleer, E., Tsagkogeorga, G., Pilka, E. S., Aspris, D., Leggate, D., Hendrick, A. G., Webster, N. A., Andrews, B., Fosbeary, R., Guest, P., Irigoyen, N., Eleftheriou, M., Gozdecka, M., Dias, J. M. L., Bannister, A. J., Vick, B., … Kouzarides, T. (2021). Small-molecule inhibition of METTL3 as a strategy against myeloid leukaemia. *Nature*, *593*(7860), 597–601. https://doi.org/10.1038/s41586-021-03536-w

Yi, J., Gao, R., Chen, Y., Yang, Z., Han, P., Zhang, H., Dou, Y., Liu, W., Wang, W., Du, G., Xu, Y., & Wang, J. (2017). Overexpression of NSUN2 by DNA hypomethylation is associated with metastatic progression in human breast cancer. *Oncotarget*, *8*(13), 20751–20765. https://doi.org/10.18632/oncotarget.10612

Yildirim, E., Sadreyev, R. I., Pinter, S. F., & Lee, J. T. (2011). X-chromosome hyperactivation in mammals via nonlinear relationships between chromatin states and transcription. *Nature structural & molecular biology*, *19*(1), 56–61. https://doi.org/10.1038/nsmb.2195

Yin, S., Deng, W., Zheng, H., Zhang, Z., Hu, L., & Kong, X. (2009). Evidence that the nonsense-mediated mRNA decay pathway participates in X chromosome dosage compensation in mammals. *Biochemical and biophysical research communications*, *383*(3), 378–382. https://doi.org/10.1016/j.bbrc.2009.04.021

Yoon, K. J., Ringeling, F. R., Vissers, C., Jacob, F., Pokrass, M., Jimenez-Cyrus, D., Su, Y., Kim, N. S., Zhu, Y., Zheng, L., Kim, S., Wang, X., Doré, L. C., Jin, P., Regot, S., Zhuang, X., Canzar, S., He, C., Ming, G. L., & Song, H. (2017). Temporal Control of Mammalian Cortical Neurogenesis by m^6A Methylation. *Cell*, *171*(4), 877–889.e17. https://doi.org/10.1016/j.cell.2017.09.003

Youn, J. Y., Dunham, W. H., Hong, S. J., Knight, J. D. R., Bashkurov, M., Chen, G. I., Bagci, H., Rathod, B., MacLeod, G., Eng, S. W. M., Angers, S., Morris, Q., Fabian, M., Côté, J. F., & Gingras, A. C. (2018). High-Density Proximity Mapping Reveals the Subcellular Organization of mRNA-Associated Granules and Bodies. *Molecular cell*, *69*(3), 517–532.e11. https://doi.org/10.1016/j.molcel.2017.12.020

YU, C. T., & ALLEN, F. W. (1959). Studies on an isomer of uridine isolated from ribonucleic acids. *Biochimica et biophysica acta*, *32*, 393–406. https://doi.org/10.1016/0006-3002(59)90612-2

Yue, Y., Liu, J., Cui, X., Cao, J., Luo, G., Zhang, Z., Cheng, T., Gao, M., Shu, X., Ma, H., Wang, F., Wang, X., Shen, B., Wang, Y., Feng, X., He, C., & Liu, J. (2018). VIRMA mediates preferential m^{6}A mRNA methylation in 3'UTR and near stop codon and associates with alternative polyadenylation. *Cell discovery*, *4*, 10. https://doi.org/10.1038/s41421-018-0019-0

Zaccara, S., Ries, R. J., & Jaffrey, S. R. (2019). Reading, writing and erasing mRNA methylation. *Nature reviews. Molecular cell biology*, *20*(10), 608–624. https://doi.org/10.1038/s41580-019-0168-5

Zaccara, S., & Jaffrey, S. R. (2020). A Unified Model for the Function of YTHDF Proteins in Regulating m^{6}A-Modified mRNA. *Cell*, *181*(7), 1582–1595.e18. https://doi.org/10.1016/j.cell.2020.05.012

Zarnegar, B. J., Flynn, R. A., Shen, Y., Do, B. T., Chang, H. Y., & Khavari, P. A. (2016). irCLIP platform for efficient characterization of protein-RNA interactions. *Nature methods*, *13*(6), 489–492. https://doi.org/10.1038/nmeth.3840

Zentner, G. E., & Henikoff, S. (2013). Regulation of nucleosome dynamics by histone modifications. *Nature structural & molecular biology*, *20*(3), 259–266. https://doi.org/10.1038/nsmb.2470

Zhang, F., Kang, Y., Wang, M., Li, Y., Xu, T., Yang, W., Song, H., Wu, H., Shu, Q., & Jin, P. (2018). Fragile X mental retardation protein modulates the stability of its m6A-marked messenger RNA targets. *Human molecular genetics*, *27*(22), 3936–3950. https://doi.org/10.1093/hmg/ddy292

Zhang, Z., Chen, L. Q., Zhao, Y. L., Yang, C. G., Roundtree, I. A., Zhang, Z., Ren, J., Xie, W., He, C., & Luo, G. Z. (2019). Single-base mapping of m^{6}A by an antibody-independent method. *Science advances*, *5*(7), eaax0250. https://doi.org/10.1126/sciadv.aax0250

(A) Zhang, Z., Chen, T., Chen, H. X., Xie, Y. Y., Chen, L. Q., Zhao, Y. L., Liu, B. D., Jin, L., Zhang, W., Liu, C., Ma, D. Z., Chai, G. S., Zhang, Y., Zhao, W. S., Ng, W. H., Chen, J., Jia, G., Yang, J., & Luo, G. Z. (2021). Systematic calibration of epitranscriptomic maps using a synthetic modification-free RNA library. *Nature methods*, *18*(10), 1213–1222. https://doi.org/10.1038/s41592-021-01280-7

(B) Zhang, W., Qian, Y., & Jia, G. (2021). The detection and functions of RNA modification m^{6}A based on m^{6}A writers and erasers. *The Journal of biological chemistry*, *297*(2), 100973. https://doi.org/10.1016/j.jbc.2021.100973

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., Lu, Z., Bosmans, R. P., Dai, Q., Hao, Y. J., Yang, X., Zhao, W. M., Tong, W. M., Wang, X. J., Bogdan, F., Furu, K., … He, C. (2013). ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Molecular cell*, *49*(1), 18–29. https://doi.org/10.1016/j.molcel.2012.10.015

Zhong, S., Li, H., Bodi, Z., Button, J., Vespa, L., Herzog, M., & Fray, R. G. (2008). MTA is an Arabidopsis messenger RNA adenosine methylase and interacts with a homolog of a sex-specific splicing factor. *The Plant cell*, *20*(5), 1278–1288. https://doi.org/10.1105/tpc.108.058883

Zhou, Y., Zeng, P., Li, Y. H., Zhang, Z., & Cui, Q. (2016). SRAMP: prediction of mammalian N6-methyladenosine (m6A) sites based on sequence-derived features. *Nucleic acids research*, *44*(10), e91.

Zhu, T., Roundtree, I. A., Wang, P., Wang, X., Wang, L., Sun, C., Tian, Y., Li, J., He, C., & Xu, Y. (2014). Crystal structure of the YTH domain of YTHDF2 reveals mechanism for recognition of N6-methyladenosine. *Cell research*, *24*(12), 1493–1496. https://doi.org/10.1038/cr.2014.152

Zug R. (2022). Developmental disorders caused by haploinsufficiency of transcriptional regulators: a perspective based on cell fate determination. *Biology open*, *11*(1), bio058896. https://doi.org/10.1242/bio.058896

7. Acknowledgements