Analysis of m^{6}A mRNA modifications and their role in dosage compensation

---

Dissertation
Zur Erlangung des Grades
Doktor der Naturwissenschaften

Am Fachbereich Biologie der
Johannes Gutenberg-Universität Mainz

Cornelia Rücklé
Mainz, April 2023

Dekan:
1. Berichterstatter:
2. Berichterstatter:
Tag der mündlichen Prüfung: 21.07.23

Zusammenfassung

m^6A-mRNA-Modifikationen spielen eine essentielle Rolle in der Regulation von mRNA. Eine wichtige Funktion von m^6A ist der Abbau von m^6A-mRNA über das YTHDF2-Protein. Um genau zu verstehen, welche m^6A-modifizierten mRNAs abgebaut werden, ist es entscheidend, die Modifikation mit hoher Genauigkeit transkriptomweit zu identifizieren. Obwohl m^6A üblicherweise mit antikörperbasierten Methoden nachgewiesen wird, hat sich gezeigt, dass diese Verfahren aufgrund mangelnder Antikörperselektivitäten einige Einschränkungen aufweisen. Um dies zu verbessern, haben wir uns im ersten Teil dieser Arbeit auf den genauen Nachweis der Modifikation konzentriert. Wir haben einige Limitierungen des m^6A-Nachweises mit dem miCLIP-Protokoll überwunden und ein verbessertes miCLIP2-Protokoll etabliert. Dieses haben wir mit einer umfangreichen Bioinformatik-Pipeline und einem maschinellen Lernklassifikator gekoppelt, was die genaue Identifizierung von m^6A transkriptomweit in Nukleotidauflösung ermöglicht. Unter Verwendung unserer neuen m^6A Annotation haben wir im zweiten Projekt eine neue Rolle von m^6A-Modifikationen in der X-zu-Autosom-Dosiskompensation identifiziert. Weibliche und männliche Zellen haben ein aktives X-Chromosom. Dies führt zu einem Ungleichgewicht zwischen dem einen X-Chromosom und den Autosomen, welche in zwei Kopien vorhanden sind. Um dieses Ungleichgewicht auszugleichen, wird vermutet, dass X-chromosomale Gene stärker exprimiert werden und somit ähnliche Expressionsniveaus wie autosomale Gene erreichen. Es hat sich gezeigt, dass X-chromosomale Transkripte eine höhere Halbwertszeit haben als autosomale Transkripte, doch es ist nicht bekannt, wie dies erreicht wird. In dieser Arbeit zeigen wir, dass X-chromosomale Transkripte deutlich weniger m^6A-Modifikationen haben und dadurch stabiler sind als autosomale Transkripte. Bei Verringerung von m^6A werden autosomale Transkripte stabiler und erreichen ähnliche Stabilitäten wie die X-chromosomalen Transkripte. Unsere Arbeit bietet eine verbesserte Methode für den transkriptomweiten Nachweis von m^6A sowie neue Einblicke in eine globale Funktion der Modifikation bei der Regulierung von Genexpression.

Abstract

m^{6}A mRNA modifications play a crucial role in mRNA metabolism, including the degradation of m^{6}A-mRNA via the YTHDF2 reader protein. To specifically understand which mRNAs are degraded in an m^{6}A-dependent way, it is essential to locate the modification in a high-confidence and transcriptome-wide manner. m^{6}A is commonly detected with antibody-based methods. These approaches have been shown to suffer several limitations due to insufficient antibody selectivities. To investigate the distribution of m^{6}A, we focused in the first part of this work on improving the accurate detection of the modification. We specifically overcame limitations of m^{6}A detection using the miCLIP protocol and established an improved miCLIP2 protocol. This was specifically coupled with an extensive bioinformatic pipeline and a machine learning classifier, allowing the accurate detection of m^{6}A transcriptome-wide in a single-nucleotide resolution. Using this novel high-confidence annotation, we present in the second part of the work a novel role of m^{6}A modifications in X-to-autosome dosage compensation. X-to-autosome dosage compensation aims to balance the gene expression of autosomes and the X chromosome: While the X chromosome is present in one copy, autosomes are present in two copies. It has been proposed that X-chromosomal genes are upregulated and reach similar expression levels as autosomal genes. It has been shown that X-chromosomal transcripts have higher half-lives than autosomal transcritps, however, how this is accomplished is not known. Here, we show that X-chromosomal transcripts are significantly depleted of m^{6}A and thereby more stable than autosomal transcripts. When depleting m^{6}A, autosomal transcripts become more stable and reach similar stabilities as the X-chromosomal transcripts. Collectively, our work firstly provides an enhanced tool for the detection of m^{6}A in a transcriptome-wide manner and, secondly, gives novel insights into a global function of the modification in the regulation of gene expression.

Preface

The results presented in this thesis have been published previously or have been accepted for print.

- **Publication 1:** *Körtel, N., ***Rücklé, C.**, *Zhou, Y., Busch, A., Hoch-Kraft, P., Sutandy, F. R., et al. (2021). Deep and accurate detection of m^6A RNA modifications using miCLIP2 and m6Aboost machine learning. *Nucleic Acids Research*, 49(16), e92-e92.
- **Publication 2: *Rücklé, C.**, *Körtel, N., Basilicata, M. F., Busch, A., Yhou, Z., Hoch-Kraft, P., et al. (Accepted for publication). RNA stability controlled by m^6A methylation contributes X-to-autosome dosage compensation in mammals. *Nature Structural & Molecular Biology.*

Both chapters are prefaced by a summary of the work and a statement in which information about my contributions is provided.

* First authors contributed equally to this work.

Contents

List of Figures

Chapter 1

Introduction

1.1 Chromosomes

Genetic information of cells is stored within the deoxyribonucleic acid (DNA). In order to protect and efficiently pack this information, DNA in eukaryotic cells is wrapped around specific proteins that are called histones, and further organized into chromatin. Chromatin folds into nucleosomes and finally into chromosomes (reviewed in Felsenfeld and Groudine, 2003).
Human cells carry 23 chromosome pairs, of which 22 are autosomal chromosomes. Autosomes exist in pairs that are homologous and carry between 750 and 2,900 genes (Nurk et al., 2022). The 23rd pair of chromosomes are sex chromosomes. Sex chromosomes differ between females and males (heteromorphic sex chromosomes): While females carry two X chromosomes that are homologous (homogametic), males carry one X chromosome and one Y chromosome (heterogametic). In fact, the Y chromosome encodes the master sex determination gene *SRY*, which confers maleness (Andrew et al., 1990; reviewed in Schafer and Goodfellow, 1996; Arnold, 2009).

1.1.1 Sex chromosome evolution

The mammalian X and Y chromosomes evolved from a pair of homologous autosomes over the past 200-300 million years after the mammalian ancestor diverged from the ancestor of birds and reptiles (Muller, 1914; Lahn and Page, 1999; reviewed in Livernois et al., 2012; Sangrithi and Turner, 2018). Accordingly, genes on the mammalian sex chromosomes have homologous genes on autosomes in outgroup species. For example, homologous genes of the eutherian (example given (e.g.) human or mouse) X chromosome are located on chromosome 1 and 4 in chicken (Ross et al., 2005; Bellott et al., 2010; reviewed in Sangrithi and Turner, 2018). X-linked genes that are homologous to chromosome 4 in chicken are present in all eutherian mammals

and belong therefore to the X conserved region (XCR). However, some human genes that are homologous to chromosome 1 in chicken are not located on the X chromosome in other mammalian species and therefore are more recent evolutionary "arrivals". Those genes belong to the X added region (XAR) (reviewed in Sangrithi and Turner, 2018; Livernois et al., 2012) (Figure 1.1).

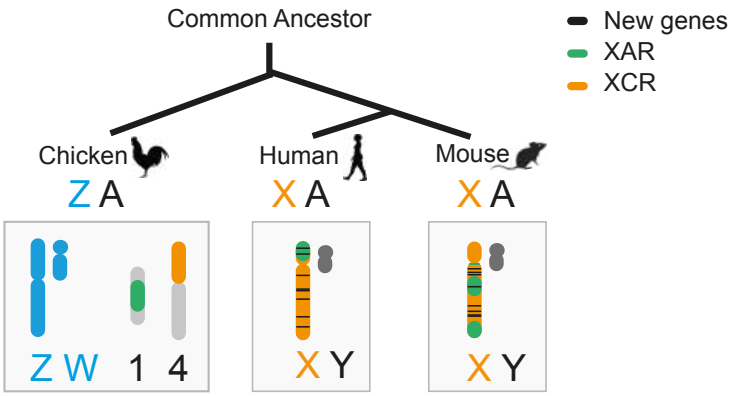


FIGURE 1.1: Mammalian sex chromosome evolution. The eutherian mammalian sex chromosomes evolved from a pair of ancestral autosomes into X and Y. While some regions on the X chromosome are recent arrivals (black), other genes are homologous to genes on chromosome 1 in chicken (X-added region (XAR), green) or to chromosome 4 in chicken (X-conserved region (XCR), orange). Adapted from Sangrithi and Turner, 2018.

It is hypothesized that the ancestral autosomes (proto-X and proto-Y) evolved into a X and Y chromosome when the proto-Y chromosome acquired the sex determining region, the *SRY* locus ('sex-determining region on the Y chromosome'), which drives e.g. testis determination. After the acquisition of *SRY* and further sex-specific genes, recombination with the X chromosome started to become suppressed and led to a degeneration of the Y chromosome (Charlesworth, 1991; reviewed in Bachtrog, 2013). Only a small region on the Y chromosome is able to recombine with the X chromosome, the pseudoautosomal region (Burgoyne, 1982). Thus, the Y chromosome carries low numbers of genes, less than 80 protein-coding genes, and is relatively small compared to the X chromosome (Skaletsky et al., 2003; reviewed in Bachtrog, 2013). In contrast, the X chromosome carries around 800 protein-coding genes (Ross et al., 2005). Since females carry two gene-rich X chromosomes and males a single X chromosome, an imbalance in genetic dosage is possible and might need to be compensated. This is underlined by studies showing that expression of both sets of X-chromosomal genes is lethal in early development in mice (Takagi and Abe, 1990; Marahrens et al., 1997).

1.1.2 X-chromosome inactivation to harmonize gene expression between males and females

To compensate gene expression differences due to heteromorphic sex chromosomes, one X chromosome is randomly inactivated during development in female eutherian mammals. X chromosome inactivation was discovered over 60 years ago, when a dense structure in female nuclei was observed, which was reffered to as the Barr body (Barr and Bertram, 1949) and subsequently was recognized being the inactive X chromosome (Ohno and Hauschka, 1960; Russell, 1961; Lyon, 1961). Hence, genes on one X chromosome are silenced and as a result both females and males carry one active X chromosome per cell. This mechanism is known as X chromosome inactivation and became intensely studied over the years (reviewed in Augui et al., 2011; Loda et al., 2022).

In eutherian mammals, X chromosome inactivation is initiated in the early stage of development and triggered by the dosage of the X inactivation center (Xic), which is present in two copies in females (Rastan, 1983; Rastan and Robertson, 1985). The Xic contains the master regulator of X chromosome inactivation (*Xist*), a non-coding RNA whose expression is sufficient to induce silencing of the chromosome from which it is expressed (Brockdorff et al., 1991; Borsani et al., 1991; Brown et al., 1991; Wutz and Jaenisch, 2000). *Xist* is monoallelically expressed, upregulated along with X chromosome inactivation and during this process able to coat the X chromosome from which it is expressed (reviewed in Loda et al., 2022). Once *Xist* has initiated silencing, multiple mechanisms that then maintain X chromosome inactivation have been described, such as gene silencing through (repressive) DNA methylations (e.g. Norris et al., 1991; Gendrel et al., 2012), histone modifications (e.g. Keohane et al., 1996; Żylicz et al., 2019), formation of supramolecular complexes (Markaki et al., 2021), as well as re-organization of the X chromosome, leading to a nuclear compartment that is silenced (Splinter et al., 2011; Giorgetti et al., 2016, reviewed in Loda et al., 2022).

Interestingly, X chromosome inactivation is not limited to one X chromosome: Individuals with several X chromosomes (e.g. three or four X chromosomes (XXX or XXXX)) have been shown to inactivate all X chromosomes except for a single one, indicating that a "counting" mechanism is in place which ensures that only a single X chromosome is active per cell (Jacobs and Strong, 1959; reviewed in Avner and Heard, 2001; Jégu et al., 2017).

Importantly, a number of genes can escape X chromosome inactivation and

become expressed from both X chromosomes (escapees) (Lyon, 1962; Carrel and Willard, 2005; reviewed in Heard and Disteche, 2006). In humans, about 15-20% of X-chromosomal genes have been described to be able to escape X chromosome inactivation (Carrel and Willard, 2005; Tukiainen et al., 2017); while in mice, only around 3% have been described as escapees (Yang et al., 2010; Berletch et al., 2015).

1.1.3 X-to-autosome dosage compensation

Males and females carry one active X chromosome. This is one copy less than autosomes, which are present in two copies. Consequently, this leads to another layer of dosage imbalance (Figure 1.2). Imbalance in X chromosomal gene dosage can lead to a number of different developmental defects (reviewed in Basilicata and Keller Valsecchi, 2021). The need for a balanced X-to-autosome dosage becomes especially apparent when considering that sex chromosomes have evolved from a pair of ancestral autosomes. This raises the question, *"How can X-chromosomal genes expressed from one chromosome keep up with autosomal genes?"* or more specifically, *"How did the regulation of X-chromosomal gene expression change over the course of evolution when genes were only expressed from one copy?"*.

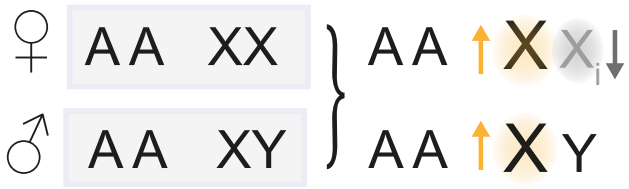


FIGURE 1.2: The two layers of dosage compensation. To equalize heterogametic sex chromosomes between males and females, one X chromosome is inactivated in female cells. Then, to equalize expression between the one active X chromosome and two active autosomes, the X chromosome is upregulated.

In 1966, Susumu Ohno formulated what is known as *"Ohno's Hypothesis"*: while autosomal genes are expressed from two chromosome copies, the single active X chromosome is in an upregulated state, leading to higher expression levels of X-chromosomal genes that are comparable and compatible with those of autosomal genes (Figure 1.2) (Ohno, 1966). The advent of high-throughput and genome-wide approaches allowed the investigation of

Ohno's hypothesis in mammals over the past years. Microarray and RNA seq data allow for powerful and detailed expression analysis, and several studies have challenged the hypothesis of X chromosome upregulation. On the one hand, different studies have claimed that X-to-autosome ratios (X:A ratios) are close to 0.5 indicating that X-chromosomal genes are not expressed in equal amounts (e.g. Xiong et al., 2010; Lin et al., 2012). On the other hand, many other studies reported X:A ratios close to 1 and supported Ohno's hypothesis (e.g. Deng et al., 2011; Lin et al., 2011).

1.1.4 X-to-autosome dosage compensation mechanisms

Several X-to-autosome dosage compensation mechanisms on the transcriptional and post-transcriptional level have been described so far. Among them, X chromosome upregulation has been shown to be supported by higher transcriptional initiation due to higher RNA polymerase II occupancies and more activating histone modifications on X-chromosomal genes (Yildirim et al., 2012; Deng et al., 2013). Furthermore, elevated transcriptional burst frequencies of X-chromosomal genes have been reported. Importantly, increased burst frequencies were not observed in undifferentiated female mouse embryonic stem cells that harbor two active X chromosomes, indicating that this mechanism can act as a molecular switch that is turned on when needed (Larsson et al., 2019; Lentini et al., 2022).

Faucillion and Larsson (2015) motivated their study on X chromosome upregulation based on the hypothesis that the *"progressive degeneration of the proto Y-chromosome must have led to evolutionary pressure at all levels to compensate for losses of functional gene copies."* (Faucillion and Larsson, 2015). This suggests that although many studies have focused on elevated transcription rates of X-chromosomal genes and elevated translation rates, RNA stability and protein stability should also play a role in X chromosome upregulation. Indeed, Faucillion and Larsson (2015) showed that X-chromosomal transcripts have longer mRNA half-lives and higher ribosome densities. Both these characteristics of X-chromosomal transcrips have also been observed in other studies (Deng et al., 2013; Wang et al., 2020b). Consequently, X-to-autosome dosage compensation is achieved through various mechanisms, both on the transcriptional as well as post-transcriptional level.

While several mechanisms have been described for transcriptional X chromosome upregulation, it is largely unknown how higher RNA stabilities of X-chromosomal transcripts are achieved (Deng et al., 2013). In 2009, a study

showed that X-chromosomal transcripts harbor lower numbers of premature translation termination codons (PTC) which can generally be introduced into a transcript by mis-splicing events. These PTCs lead to the degradation of a transcript via the nonsense-mediated decay (NMD) pathway. Due to lower numbers of PTCs on the X chromosome, transcripts are less degraded and thus more stable (Yin et al., 2009). However, no further studies have shed light on differential mRNA stability of autosomal and X-chromosomal transcripts. This raises an interesting question on how post-transcriptional mechanisms can selectively recognize and act on X-chromosomal transcripts.

1.2 RNA as a regulator of gene expression

The central dogma of molecular biology, first stated by Crick (1958), describes the translation of genetic information from DNA into functional proteins via an RNA intermediate. This is not working in a one-to-one manner, meaning that each step in this process harbors opportunities for regulation, leading to genes that are expressed to different levels. The term "gene expression" therefore refers to the idea that the translation of one gene into a protein is dynamically regulated. One way of regulation is the modification of DNA or RNA bases, or protein amino acids. More recently, the field of mRNA modifications has undergone several breakthroughs for the detection of modifications leading to a heavily studied and rapidly evolving research field.

1.2.1 Epitranscriptomics: RNA modifications as a regulator of RNA

RNA modifications are chemical alterations ("decorations") of the RNA nucleotides that influence the fate of an RNA molecule. To date, more than 150 different RNA modifications have been identified, with numerous different functions in the regulation of RNA turnover (Boccaletto and Bagiński, 2021; reviewed in Jonkhout et al., 2017). The advent of high-throughput sequencing techniques allowed the transcriptome-wide detection of mRNA modifications (Dominissini et al., 2012; Meyer et al., 2012) and thus led to a rapidly evolving research field known as *Epitranscriptomics* (He, 2010; Saletore et al., 2012). Due to its direct role in gene expression, several mRNA modifications have been identified to play important roles in human diseases and thus have become a focus of RNA therapeutics and drug discovery research (reviewed in Jonkhout et al., 2017; Cayir, 2022; Barbieri and Kouzarides, 2020).

1.2.2 N^6-methyladenosine

The most abundant internal mRNA modification is N^6-methyladenosine and was identified to be present in mRNAs over 4 decades ago (Desrosiers et al., 1974; Perry et al., 1975). In mammals, estimations of m^6A modifications on transcripts reached from 1-3 modifications per RNA molecule (Perry and Kelley, 1976; Wei et al., 1975; Dominissini et al., 2012; Meyer et al., 2012). Due to its abundance, Desrosieres hypothesized already in 1974 that m^6A plays an essential role in cellular processes. Accordingly, m^6A has been shown to be involved in various different functions in the cell, including splicing, nuclear export, translation as well as RNA degradation (reviewed in Murakami and Jaffrey, 2022). Interestingly, recent work has suggested that m^6A modification numbers were largely underestimated, with new estimations up to 13 modifications per transcript (Uzonyi et al., 2023; Tegowski et al., 2022; Liu et al., 2022).

1.2.3 m^6A deposition

m^6A on mRNA is co-transcriptionally deposited by a methyltransferase "writer" complex and S-adensosylmethionine (SAM) serves as the methyl donor (Bokar et al., 1994; Śledź and Jinek, 2016; reviewed in Jia et al., 2013). The writer complex is estimated to be 1,000 kilodalton large and is comprised of several proteins (reviewed in Garcias Morales and Reyes, 2021). Three proteins, METTL3, METTL14 and WTAP form the core writer complex. METTL3 is the catalytically active subunit and it has been shown that the protein contains a conserved DPPW (Asp-Pro-Pro-Trp) motif in loop 1 of the methyltransferase domain which is coordinating the adenine group of the acceptor mRNA substrate (Śledź and Jinek, 2016). METTL14 stabilizes METTL3 and binds to RNA, but is catalytically inactive (Liu et al., 2014; Wang et al., 2016a; Wang et al., 2016b). Together with WTAP, which was shown to localize the complex into nuclear speckles (Ping et al., 2014), they form the core writer complex (Schöller et al., 2018) (Figure 1.3). Additional proteins of the complex play roles in the stability, localization, substrate binding or catalytic efficiency (reviewed in Garcias Morales and Reyes, 2021).
Since METTL3 has been shown to be the catalytically active subunit required for m^6A deposition, many studies have targeted the gene in knockout (KO) or knock-down (KD) experiments. However, depleting Mettl3 is lethal in early embryonic development in mice (Geula et al., 2015) and most

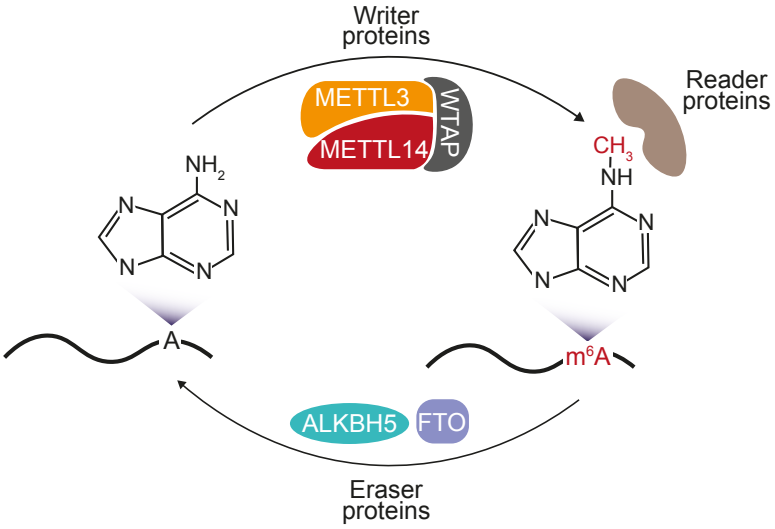


FIGURE 1.3: m^{6}A is deposited by a methyltransferase complex of which METTL3 METTL14 and WTAP form the core complex. It can be removed again by eraser proteins such as FTO and ALKBH5. The biological role is carried out by reader proteins that influence the fate of the mRNA.

successful *METTL3* KO in human and mouse cells have shown a substantial remaining percentage of m^{6}A in cells (Poh et al., 2022). This was first suggested to be METTL3-independent m^{6}A sites, however, more recently, a study showed that many *METTL3* KO cells harbor a specific METTL3 isoform that is still catalytically active and responsible for the majority of remaining m^{6}A in *METTL3* KO cells (Poh et al., 2022). Importantly, a small molecule inhibitor targeting the DPPW motif of METTL3 has recently been shown to successfully inhibit m^{6}A deposition (Yankova et al., 2021). This inhibitor is drastically facilitating the research of m^{6}A by overcoming the limitations of *METTL3* KO or KD studies.

1.2.4 m^{6}A location in mRNAs

m^{6}A is deposited in a specific sequence within a DRACH (D = A, G or T; R = A or G; H = A, C or U) motif and is enriched around stop codons (Dimock and Stoltzfus, 1977; Wei et al., 1976; Dominissini et al., 2012; Meyer et al., 2012). It has been suggested that not all DRACH motifs are methylated and that m^{6}A deposition is non-random and regulated not only through sequence content (Dominissini et al., 2012; Meyer et al., 2012).
The non-random pattern of m^{6}A modifications prompted the hypothesis that

mechanisms must be in place that guide positions on transcripts that are supposed to be methylated. In 2019, a study showed that a specific histone modification, H3K36me3, is co-occurring on DNA positions that are later m^6A methylated positions on RNA (Huang et al., 2019). It was shown that changes in the histone modifications led to alterations of m^6A modifications on RNAs. Together with the finding that METTL14 binds H3K36me3, they hypothesized that the H3K36me3 could mark m^6A deposition by interacting with the m^6A methylttansferase complex (Huang et al., 2019).
Another recent article has shown that m^6A deposition is influenced by the proximity to a splice junction. While positions close to a splice junction are less modified, all other positions carry m^6A modifications (Uzonyi et al., 2023). Another recent paper described a similar model, in which a core component of the exon junction complex, EIF4A3, blocks the binding and m^6A deposition of METTL3 to exon-exon boundaries. Regions that are depleted in the exon junction complex, such as long internal exons or the last exon of a transcript are thus more accessible to METTL3 and more methylated (Yang et al., 2022). These findings could explain the characteristic enrichment of m^6A around the stop codon, which has been observed for nearly a decade (Dominissini et al., 2012; Meyer et al., 2012; Linder et al., 2015).

1.2.5 m^6A dynamics and removal

One finding that raised interest for the epitranscriptomic field is that m^6A is a dynamic modification that is reversible (Jia et al., 2011). This suggests that the modification is dynamically modulated in a signal-dependent manner (reviewed in Mauer and Jaffrey, 2018). Two "eraser" proteins have been identified so far: fat mass and obesity-associated protein (FTO) (Jia et al., 2011) as well as alpha-ketoglutarate-dependent dioxygenase alkB homolog 5 (ALKBH5) (Zheng et al., 2013). While ALKBH5 has been shown to demethylate m^6A in several studies, FTO was recently identified to rather demethylate a similar modification, m^6A_m (Mauer et al., 2017; reviewed in Lesbirel and Wilson, 2019).
m^6A methylation is non-stoichiometric: not all transcripts harbor m^6A at the same positions but only a fraction of transcripts can be methylated (reviewed in Tuck, 1992; Meyer and Jaffrey, 2014). It is tempting to speculate that m^6A stoichiometry is driven by an active and permanent m^6A removal, however, it has been suggested over the years that m^6A removal is a rather

rare event and not influencing m^{6}A dynamics in a broad manner. The non-stoichiometric nature of m^{6}A is suggested to be driven by the writer complex (reviewed in Mauer and Jaffrey, 2018; Lesbirel and Wilson, 2019).

1.2.6 m^{6}A functions via reader proteins

The function of m^{6}A is carried out by RNA-binding proteins that specifically recognize the modification ("reader" proteins). To date, several different proteins have been shown to selectively bind m^{6}A and change mRNA fate in a specific manner. Among them, the YTH-domain containing proteins are the most prominent and well-studied m^{6}A readers. Additionally, IGF2BP1 and HuR have been described as m^{6}A readers (Huang et al., 2018; Visvanathan et al., 2018), however, if they directly recognize m^{6}A or indirectly via other co-factors has been discussed in the field. mRNA stability (reviewed in Boo and Kim, 2020), mRNA export (reviewed in Lesbirel and Wilson, 2019) and splicing (Xiao et al., 2016) as well as mRNA translation (reviewed in Meyer, 2019b) are among the processes that have been shown to be influenced by m^{6}A (reviewed in Jiang et al., 2021). In this thesis, I will focus on YTH domain-containing m^{6}A reader proteins and their function in RNA stability.

1.2.7 YTH domain-containing proteins

YT521-B-homology domain (YTH)-containing proteins contain a close to 140 amino acids RNA-binding domain that is highly conserved among the proteins (Stoilov et al., 2002). This YTH domain has been shown to recognize and bind m^{6}A (Xu et al., 2014; Zhu et al., 2014). Three different YTH-domain-containing protein categories have been identified: the YTH domain-containing protein 1 (YTHDC1) and YTH domain-containing protein 2 (YTHDC2) and the YTH domain-containing family (YTHDF proteins). YTHDC1 was described to have several different functions upon m^{6}A recognition, such as functions in splicing (Xiao et al., 2016) as well as nuclear export of m^{6}A-mRNA (Roundtree et al., 2017). Interestingly, YTHDC1 has also been shown to play a role in X chromosome inactivation (Patil et al., 2016). More specifically, *Xist* is highly m^{6}A methylated and recruits YTHDC1. This binding of YTHDC1 to *Xist* promotes the repression of X-chromosomal genes. This shows that m^{6}A plays an important role in gene expression regulation

by recruiting specific reader proteins that influence cellular processes. In contrast to YTHDC1, YTHDC2 is the most distinct protein to the other YTH proteins (reviewed in Patil et al., 2018; Liao et al., 2018). It has been reported to regulate mammalian spermatogenesis in an m^6A-dependent manner (Hsu et al., 2017) but whether it really binds m^6A has also been challenged (Patil et al., 2016; reviewed in Patil et al., 2018).
The three YTHDF proteins (YTHDF1/2/3) have highly conserved sequences (reviewed in Patil et al., 2018). Early studies of the YTHDF proteins have suggested that the three cytoplasmic proteins have independent roles, such as YTHDF1 in influencing translation (Wang et al., 2015), YTHDF2 promoting m^6A-RNA degradation (Wang et al., 2014) and YTHDF3 acting together with both proteins in translation and mRNA degradation (Shi et al., 2017; Li et al., 2017). This is in contrast to studies that have suggested similar functions for all three proteins (Du et al., 2016; reviewed in Patil et al., 2018). More recently, two studies have further supported the hypothesis of a unified function of YTHDF proteins showing that the three YTHDF proteins bind to the same m^6A-modified mRNAs and that the proteins act together in promoting the degradation of m^6A-modified mRNAs (Zaccara and Jaffrey, 2020; Lasman et al., 2020).

1.2.8 mRNA degradation via YTHDF proteins

Early studies reported that m^6A sites influence mRNA stability (Wang et al., 2014; Schwartz et al., 2014). Further, it was shown that YTHDF2 localizes into processing bodies (P bodies) (Wang et al., 2014; Ries et al., 2019) in which mRNAs are degraded (Sheth and Parker, 2003). These findings have set the ground of the nowadays prominent function of m^6A: mRNA degradation. One leading question in the field is how m^6A and YTHDF proteins mechanistically drive mRNA degradation. So far, YTHDF proteins have been linked to three different mRNA degradation pathways in an m^6A-dependent manner (Figure 1.4) (Gibbs and Chanfreau, 2022).
Firstly, it was reported that YTHDF2 interacts with CCR4-NOT transcription complex subunit 1 (CNOT1) and recruits the Carbon Catabolite Repression-Negative On TATA-less (CCR4-NOT) deadenylase complex, which then induces deadenylation of m^6A-mRNAs (Du et al., 2016). Secondly, it has been shown that in presence of a specific binding motif for the heat-responsive protein 12 (HRSP12), m^6A-mRNAs bound by YTHDF2 can be degraded

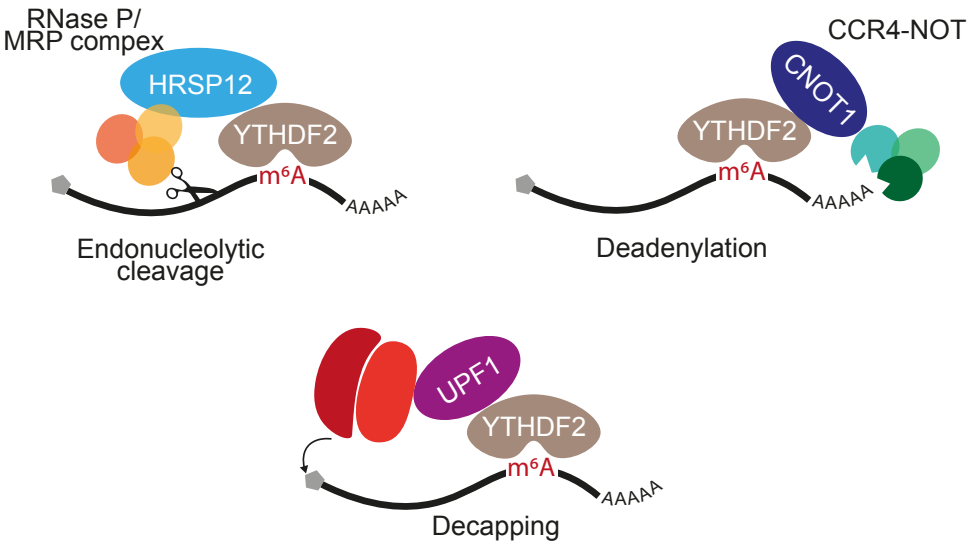


FIGURE 1.4: Three different mRNA degradation mechanisms have been linked to YTHDF2 and m^{6}A-mRNA, endonucleolytic cleavage with the RNase P/MRP complex, deadenylation via the CCR4-NOT complex and decapping of the mRNA via UPF1. Adapted from Gibbs and Chanfreau, 2022

through endonucleolytic cleavage (Park et al., 2019). Thirdly and most recently, a novel mRNA decay pathway has been linked to m^{6}A and YTHDF2 (Boo et al., 2022). YTHDF2 has been previously shown to interacts with the RNA-decay factor Up-frameshift protein 1 (UPF1) (Schweingruber et al., 2016). Boo and colleagues could now show that YTHDF2 directly interacts with UPF1 and leads to the degradation of m^{6}A-mRNA. This is accomplished by the recruitment of enzymes that initiate de-capping and thereby mRNA decay (Boo et al., 2022). Even though mRNA degradation of m^{6}A-modified mRNAs is a prominent function of m^{6}A and the YTHDF proteins, it is not clear if all m^{6}A-mRNAs are degraded or if there is a specific subset of mRNAs that are preferably degraded. This raises the interesting question: *"Which mRNAs are marked with m^{6}A to be degraded and why?"*

1.3 m^{6}A detection methods

To globally identify functions of m^{6}A, the accurate and transcriptome-wide identification of the modification is essential. Even though m^{6}A was discovered nearly 50 years ago, detection limitations have hindered the field from transcriptome-wide analysis of the function and distribution of the modification (Fu and He, 2012).

The development of m^{6}A-specific antibodies (Bringmann and Lührmann, 1987), coupled with advances in next-generation sequencing (NGS) has given rise to novel methods that allow a broad and transcriptome-wide characterizations of m^{6}A: two similar techniques that were developed in 2012, m^{6}A-seq and MeRIP seq, are based on m^{6}A-specific antibodies that first enrich modified mRNAs and then high-throughput sequencing is used to identify the region of the modification. These studies identified thousands of m^{6}A sites transcriptome-wide, demonstrating that m^{6}A is an ubiquitous modification and paving the road for the field of *m^{6}A Epitranscriptomics* (Dominissini et al., 2012; Meyer et al., 2012).

1.3.1 miCLIP to detect m^{6}A

MeRIP seq and m^{6}A-seq enrich for m^{6}A modified mRNAs and allowed a transcriptome-wide mapping of m^{6}A. However, the exact position of the modification cannot be determined by these protocols, since they only give information about m^{6}A in a 24-100 nucleotide window (Dominissini et al., 2012). The need for transcriptome-wide m^{6}A annotations in a single-nucleotide resolution gave rise to the development of Ultraviolet (UV) crosslinking-based approaches (reviewed in Moshitch-Moshkovitz et al., 2022). One of these approaches is m^{6}A-individual nucleotide resolution crosslinking and immunoprecipitation (miCLIP) (Linder et al., 2015).
In miCLIP, an m^{6}A antibody is crosslinked to poly(A) RNA using UV-light, leading to a covalent bond of the antibody and its target-RNA. This complex is then purified. After adapter ligation, the antibody-RNA complex is digested using proteinase K. The antibody is not fully digested but leaves a small polypeptide on the RNA at the position where the crosslink happened. In the next step, the RNA is reverse transcribed into complementary DNA (cDNA). During this process, the remaining polypeptide on the RNA causes the reverse-transcriptase to mostly truncate at this position (Figure 1.5). Of note, depending on the m^{6}A antibody used, the remaining peptide may also cause the reverse transcriptase to pass over (rather than truncate) over the position of the crosslink and remaining peptide. This introduces a mismatch after the position of the crosslink. After polymerase chain reaction (PCR) amplification and high-throughput sequencing, the first position of each read will be the truncation position of the reverse transcriptase (RT), or, in case of a pass-over of the transcriptase, the position of the crosslink can be detected through a mismatch (König et al., 2010; Linder et al., 2015). This protocol is

followed by an extensive bioinformatic analysis to identify the position of m^6A.

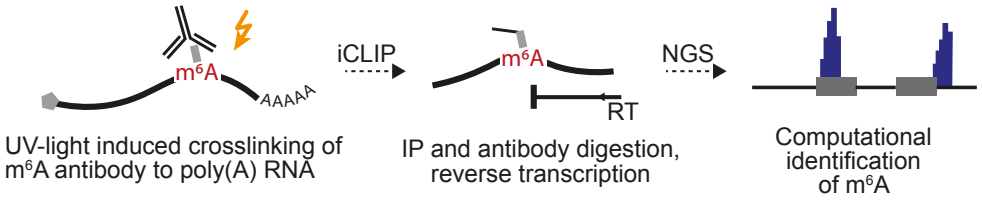


FIGURE 1.5: m^6A detection using miCLIP. Poly(A) RNA is incubated with an m^6A-specific antibody, which is crosslinked to its target mRNA using UV light. After antibody digestion and reverse transcription by the reverse transcriptase (RT), the amplified library can be sequenced and positions of m^6A can be detected.

1.3.2 Limitations of miCLIP

Since its development, miCLIP has been suffering from several limitations: one of them is the need for high-input materials. This makes the protocol exclusive for samples where high amounts of materials are available. Further, high background signals due to limited antibody specificity have been reported. This is partially due to cross-reactivities of the antibody with for example other modifications like m^6A_m (McIntyre et al., 2020). Furthermore, miCLIP does not provide information about the quantity of transcripts that are carrying a specific m^6A site: there is no stoichiometric information. Since m^6A has been proposed to be a low-stoichiometric modification, the need for quantitative m^6A detection has been pointed out.

1.3.3 Bioinformatic approaches to enhance the m^6A detection from miCLIP data

An important part of NGS-based mapping methods such as miCLIP is the bioinformatic analysis to accurately identify the signal from the produced data. This implies to balance the identification of the signal and maximizing the results while minimizing error rates or false positives. In miCLIP, the first steps of the analysis include quality controls, mapping of the sequences to a reference genome and removing PCR duplicates in the data (Busch et al., 2020; Hawley and Jaffrey, 2019). To identify positions in the transcriptome

that are significantly enriched in crosslinking events, specific algorithms designed for the detection of signal in CLIP data can be applied (peak calling, reviewed in Chakrabarti et al., 2018). Peak calling detects positions or regions that are significantly enriched in signal compared to the background signal surrounding those positions. Even though peak calling enriches for positions with high signals (peaks), it is important to further filter these identified peaks, since some of them can stem from unspecific antibody binding, cross-reactivities with other modifications, or general crosslinking biases (reviewed in Chakrabarti et al., 2018). To overcome unspecific signal leading to false positives, the miCLIP protocol only considers identified peaks at adenosines and further filters them for the location within a DRACH motif (Hawley and Jaffrey, 2019). This does not ensure that the signal of these sites stems from true m^6A sites, and further, excludes the possibility that some rare m^6A sites could be occurring outside of DRACH motifs. It will be of interest to overcome these limitations to improve the detection of m^6A in a transcriptome-wide manner.

Limitations in the experimental detection of m^6A have given rise to complementary computational methods. Machine learning algorithms can support or complement bioinformatic analysis in identifying patterns in data that lead to predictions with high accuracies. Based on previously detected m^6A sites, machine learning-based methods have been established to predict m^6A sites. While 26 m^6A-predictors were developed until 2021, only five were based on high quality, single-nucleotide resolution datasets. Since one major caveat of training a machine learning classifier is the quality of the training dataset, many tools only reached poor prediction performances (reviewed in Chen et al., 2020).

1.3.4 Antibody-free m^6A detection

Limitations of antibody-based m^6A detection methods has stimulated the development of antibody-free m^6A detection methods. Several different approaches have been published so far. One method makes use of a RNase called MazF, which is able to cleave unmodified ACA motifs, but does not act on modified ACA motifs (Mazter-seq, Garcia-Campos et al., 2019; m^6A-REF-seq, Chen et al., 2022). The limitation of this method is the exclusive detection of m^6A sites within ACA motifs. Further studies have focused on detecting m^6A through reader or writer proteins: in DART-seq, the YTH domain of

YTHDF2 was fused to apolipoprotein B mRNA editing enzyme catalytic subunit 1 (APOBEC1), which is a cytidine to uridine (C-to-U) deaminase, leading to the positions that are close to the protein binding site to have the characteristic C-to-U pattern (Meyer, 2019a). This method has been recently brought to single-cell level (Tegowski et al., 2022). Further approaches that base on the eraser protein FTO or the methyl-donor SAM have been established (Wang et al., 2020a; Shu et al., 2020; reviewed in Moshitch-Moshkovitz et al., 2022). These methods are limited to the detection of m^6A sites that are recognized by the specific protein. Most recently, two novel methods that are based on the deamination of only unmethylated adenosines were developed (GLORI, Liu et al., 2022; eTAM-seq, Xiao et al., 2023). These techniques allow the accurate detection and at the same time stoichiometric quantification of m^6A without antibodies.
RNA modification detection has also been developed in the field of direct RNA sequencing (dRNA seq): Oxford Nanopore Technologies dRNA seq is fundamentally different from other sequencing techniques: it directly sequences the native RNA by passing it through a membrane-embedded protein pore. During this passing, disruptions in ion currents can be identified, which make it possible to detect the nucleotide sequence based on the specific current pattern (Deamer et al., 2016). Importantly, these currents are different if an RNA base is modified, and in case of m^6A, can be read out as an error profile. Several different machine learning algorithms have been established to detect m^6A from dRNA seq data (Pratanwanich et al., 2021; reviewed in Furlan et al., 2021). Several antibody-free methods have been developed, however, antibody-based m^6A detection is still the most commonly used method to identify m^6A transcriptome-wide.

1.4 Aim of this thesis

m^6A mRNA modifications have been shown to be important key players in many cellular processes. For many identified functions of m^6A it is still unclear how or which individual modification is influencing process changes in detail. This lack of understanding is due to the limitations of an accurate detection of m^6A. Therefore, to study m^6A in more detail, this work aims to enhance the detection of m^6A. Then, using a high confidence m^6A annotation, we investigate the role of m^6A-mediated mRNA stability in X-to-autosome dosage compensation.

In the first part of this thesis, we aim to overcome limitations of m^6A detection using the miCLIP protocol. Coupling an enhanced experimental protocol with an extensive bioinformatic analysis as well as a machine learning classifier will allow to accurately predict m^6A sites from miCLIP data. This will significantly enhance the detection of m^6A and allow to learn more about the positioning and distribution about the modification.
In the second chapter, we aim to use the novel high-confidence m^6A annotation to investigate the m^6A distribution of transcripts of autosomal versus X-chromosomal transcripts. By comparing the distribution of m^6A sites of X-chromosomal and autosomal transcripts, we tackle the questions if m^6A is differentially distributed in a chromosome-wide manner. Then, using RNA sequencing experiments as well as RNA half-live measurements, we aim to understand if differential RNA stabilities of X-chromosomal and autosomal transcripts are mediated by m^6A RNA modifications. Collectively, this work will help to provide a comprehensive and high-confidence m^6A annotation in a transcriptome-wide and single-nucleotide resolution, as well as investigate a novel role of the modification in the regulation of gene expression.

Chapter 2

Publication 1: Deep and accurate detection of m^{6}A RNA modifications using miCLIP2 and m6Aboost machine learning.

2.1 Summary

m^{6}A is the most abundant internal mRNA modification and plays a role in many processes in the cell. One established method to detect the modification is miCLIP (m^{6}A individual-nucleotide resolution UV crosslinking and immunoprecipitation). While miCLIP allows for the transcriptome-wide m^{6}A detection in a single-nucleotide resolution, it suffers from several limitations, such as high input materials and high background noises in the data. To overcome these limitations we present miCLIP2, which is an enhanced protocol that results in high-complexity libraries while using less input material. We couple the protocol with an extensive bioinformatical pipeline and specifically identify unspecific background signals by comparing miCLIP2 WT with miCLIP2 *Mettl3* KO data. Using the resulting high-confidence m^{6}A annotation, we train a machine learning classifier that specifically learns the miCLIP2 footprint of true m^{6}A sites. Using this classifier, we can predict m^{6}A signal of miCLIP2 data in a high-confidence manner without the need of a complementary *Mettl3* KO experiment.

2.2 Zusammenfassung

m^{6}A ist die am häufigsten vorkommende interne mRNA-Modifikation und spielt in vielen Prozessen in der Zelle eine Rolle. Eine etablierte Methode zur

Detektion der Modifikation ist miCLIP (m^6A individual-nucleotide resolution UV crosslinking and immunoprecipitation). miCLIP ermöglicht die transkriptomweite m^6A-Detektion in Nukleotid-Auflösung, aber diese Methode leidet aber auch unter mehreren Einschränkungen, wie zum Beispiel hohes benötigtes Inputmaterial sowie starkem unspezifischem Hintergrundsignal in den Daten. Um diese Einschränkungen zu überwinden, stellen wir miCLIP2 vor, ein verbessertes Protokoll, das mit weniger Inputmaterial zu hochkomplexen Libraries führt. Wir koppeln das Protokoll mit einer umfangreichen bioinformatischen Pipeline und identifizieren gezielt unspezifische Hintergrundsignale durch den Vergleich von miCLIP2 WT mit miCLIP2 *Mettl3* KO-Daten. Anhand der daraus resultierenden hoch zuverlässigen m^6A-Annotation trainieren wir einen maschinellen Lernklassifikator, der speziell den miCLIP2-Fußabdruck echter m^6A-Stellen lernt. Mit diesem Klassifikator können wir das m^6A-Signal von miCLIP2-Daten mit hoher Zuverlässigkeit vorhersagen, ohne dass ein komplementäres *Mettl3* KO-Experiment benötigt wird.

2.3 Statement of contribution

In this work I contributed to the miCLIP2 analysis pipeline for the processing of sequencing reads and detecting initial signals as well as differential methylation analysis. Further, I contributed to the computational validation of predicted m^6A sites in human datasets by comparing them to RBP-binding data and other m^6A annotations. I participated in all project and paper-related meetings and developed ideas and directions of the project. I contributed to the manuscript preparation and revision in all stages.

Supervisor confirmation:

_____

Published online 22 June 2021

Nucleic Acids Research, 2021, Vol. 49, No. 16 ***e92***
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.

© The Author(s) 2021. Published by Oxford University Press on behalf of Nucleic Acids Research.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from https://academic.oup.com/nar/article/49/16/e92/6307904 by Institut für Recht und Wirtschaft user on 28 February 2023

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

m^6A is a dynamic modification. It is deposited by writers, recognized by readers and removed by erasers. The writing of m^6A in mRNA is mainly carried out by a highly conserved, multicomponent methyltransferase complex that catalyzes the conversion of adenosine to m^6A. 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^6A 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^6A writers which mainly add m^6A to U2 and U6 snRNAs, lncRNA or pre-mRNA (12–14). In mRNA, m^6A enriches in a DRACH ([G/A/U][G>A]m^6AC[U>A>C]) consensus sequence and occurs in thousands of transcripts, with an average of one to three m^6A sites per mRNA transcript (15–17). However, only a fraction of DRACH motifs contain an m^6A modification. Furthermore, m^6A 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^6A, it is essential to precisely locate the modification. Although m^6A had been identified over four decades ago, only recent technological breakthroughs allowed a transcriptome-wide mapping of m^6A (15,16,18,19). Antibody-based immunoprecipitation followed by high-throughput sequencing (m^6A-seq, m^6A-MeRIP) enabled mapping of m^6A within a ∼100 nucleotide (nt) window and paved the way to further understand and dissect the cellular and physiological functions of m^6A (15,16). Further improvements in 2015 led to an individual-nucleotide resolution UV crosslinking and immunoprecipitation (iCLIP)-based method, called m^6A iCLIP (miCLIP), which allows the transcriptome-wide mapping of individual m^6A 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^6A 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^6A-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^6A sites by an orthogonal method. The resulting high-confidence m^6A sites within DRACH and non-DRACH motifs were used to train a machine learning model, named m6Aboost, which recognizes the specific characteristics of m^6A 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^6A sites in miCLIP2 datasets independently of a *Mettl3* KO.

MATERIALS AND METHODS

LC–MS/MS analysis of m^6A levels

The experiments were performed as described in (25). Ribonucleoside (A, m^6A) 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^6A 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^6A) and 285.1→153.1 ($^2H_3$-N6-mrA). Quantification of all samples utilized biological triplicates, and averaged values of m^6A 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

Downloaded from https://academic.oup.com/nar/article/49/16/e92/6307904 by Institut für Recht und Wirtschaft user on 28 February 2023

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.

Downloaded from https://academic.oup.com/nar/article/49/16/e92/6307904 by Institut für Recht und Wirtschaft user on 28 February 2023

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/NNNNXXXXXXNNNNNA GATCGGAAGAGCGTCGTG/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′-ACACGACGCTCTTCCG ATCT-3′ and P3Solexa_s, 5′-CTGAACCGCTCTTCCG ATCT-3], 10 μM each, 25 μl Phusion High Fidelity PCR Master Mix [New England Biolabs, M0531S] was prepared

Downloaded from https://academic.oup.com/nar/article/49/16/e92/6307904 by Institut für Recht und Wirtschaft user on 28 February 2023

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

Downloaded from https://academic.oup.com/nar/article/49/16/e92/6307904 by Institut für Recht und Wirtschaft user on 28 February 2023

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-

Downloaded from https://academic.oup.com/nar/article/49/16/e92/6307904 by Institut für Recht und Wirtschaft user on 28 February 2023

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

Downloaded from https://academic.oup.com/nar/article/49/16/e92/6307904 by Institut für Recht und Wirtschaft user on 28 February 2023

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

Downloaded from https://academic.oup.com/nar/article/49/16/e92/6307904 by Institut für Recht und Wirtschaft user on 28 February 2023

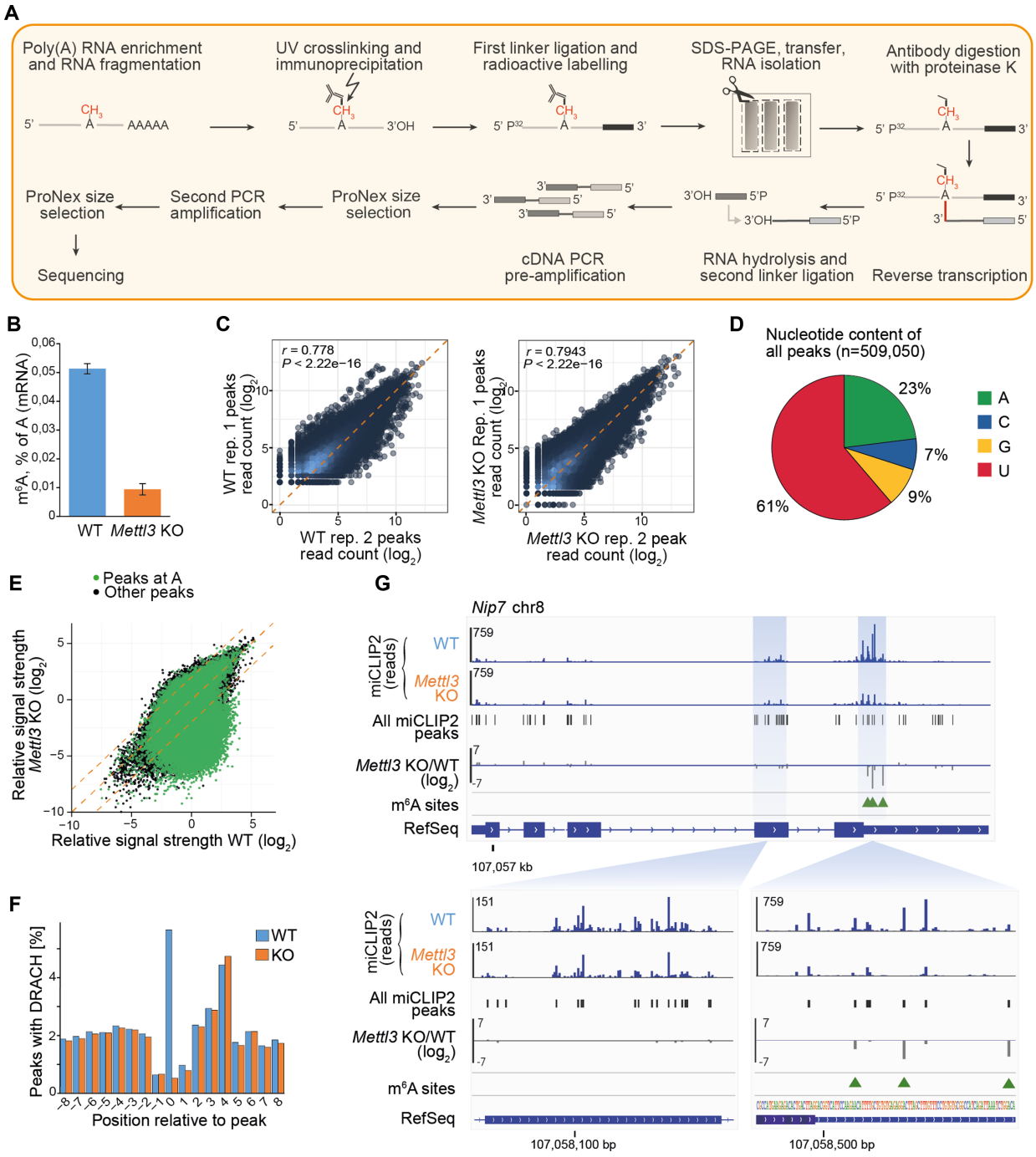


Figure 1. The optimized miCLIP2 protocol produces high complexity libraries with high reproducibility. (**A**) An overview of the miCLIP2 protocol. (**B**) mESC *Mettl3* KO cells show a significant depletion of m^{6}A on mRNAs. m^{6}A levels measured by liquid chromatography-tandem mass spectrometry (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.

Downloaded from https://academic.oup.com/nar/article/49/16/e92/6307904 by Institut für Recht und Wirtschaft user on 28 February 2023

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

Downloaded from https://academic.oup.com/nar/article/49/16/e92/6307904 by Institut für Recht und Wirtschaft user on 28 February 2023

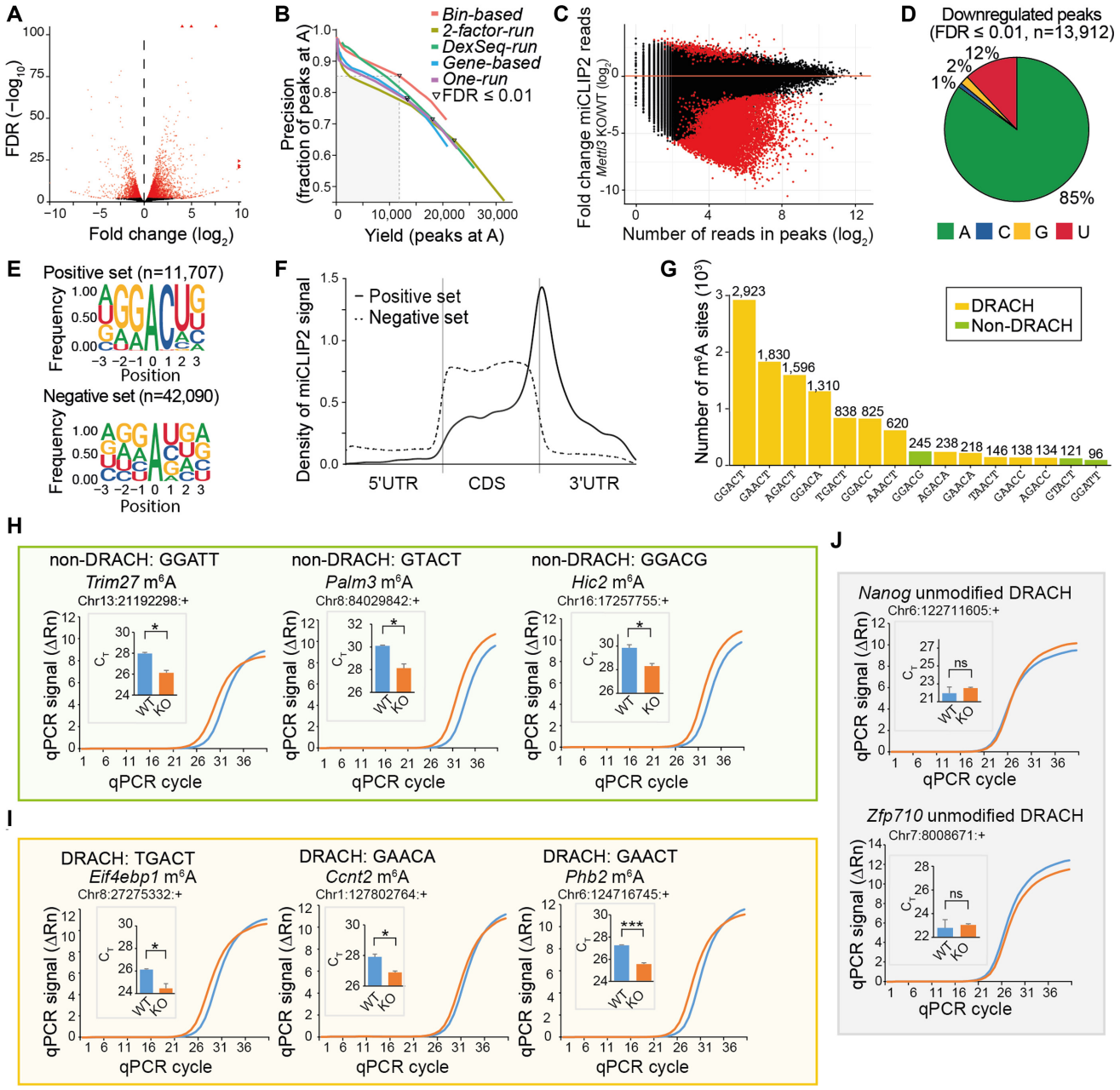


Figure 2. Differential peak analysis allows to identify true m^{6}A sites from miCLIP2 data. (**A**) *Mettl3* KO causes drastic changes in gene expression. Volcano plot shows negative log$_2$-transformed fold change (log2FC) of gene expression between *Mettl3* KO and WT against log$_{10}$-transformed 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, $n = 3$.

Downloaded from https://academic.oup.com/nar/article/49/16/e92/6307904 by Institut für Recht und Wirtschaft user on 28 February 2023

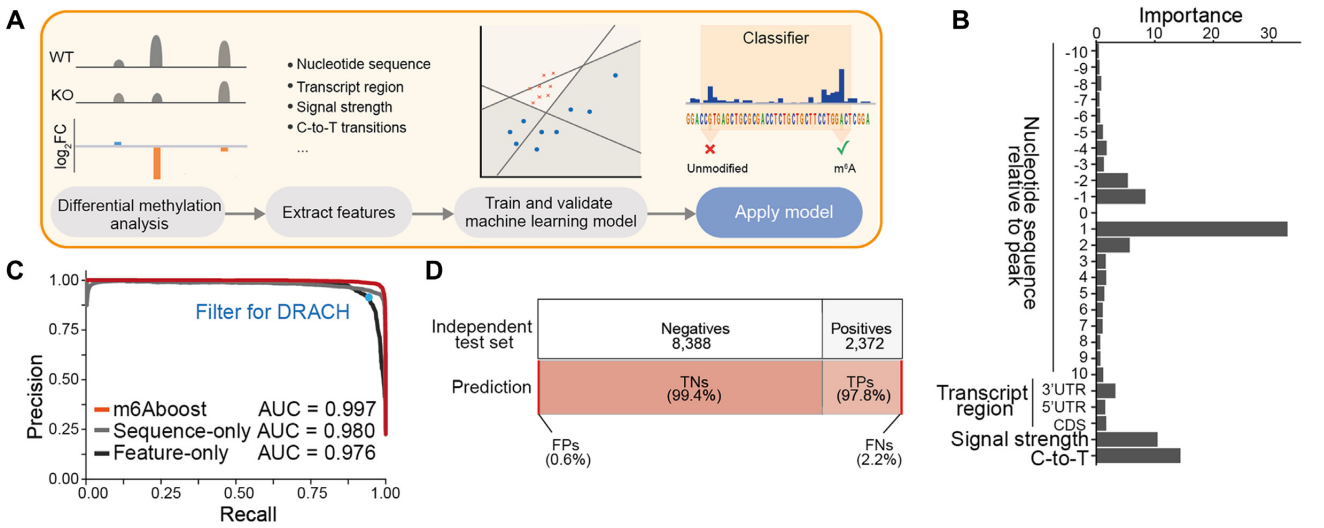


Figure 3. The machine learning classifier m6Aboost reliably predicts m^{6}A sites from miCLIP2 data. (**A**) Overview of the machine learning approach. First, miCLIP2 WT and *Mettl3* KO datasets are analyzed for differential methylation to identify Mettl3-dependent m^{6}A sites. The resulting positive 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

Downloaded from https://academic.oup.com/nar/article/49/16/e92/6307904 by Institut für Recht und Wirtschaft user on 28 February 2023

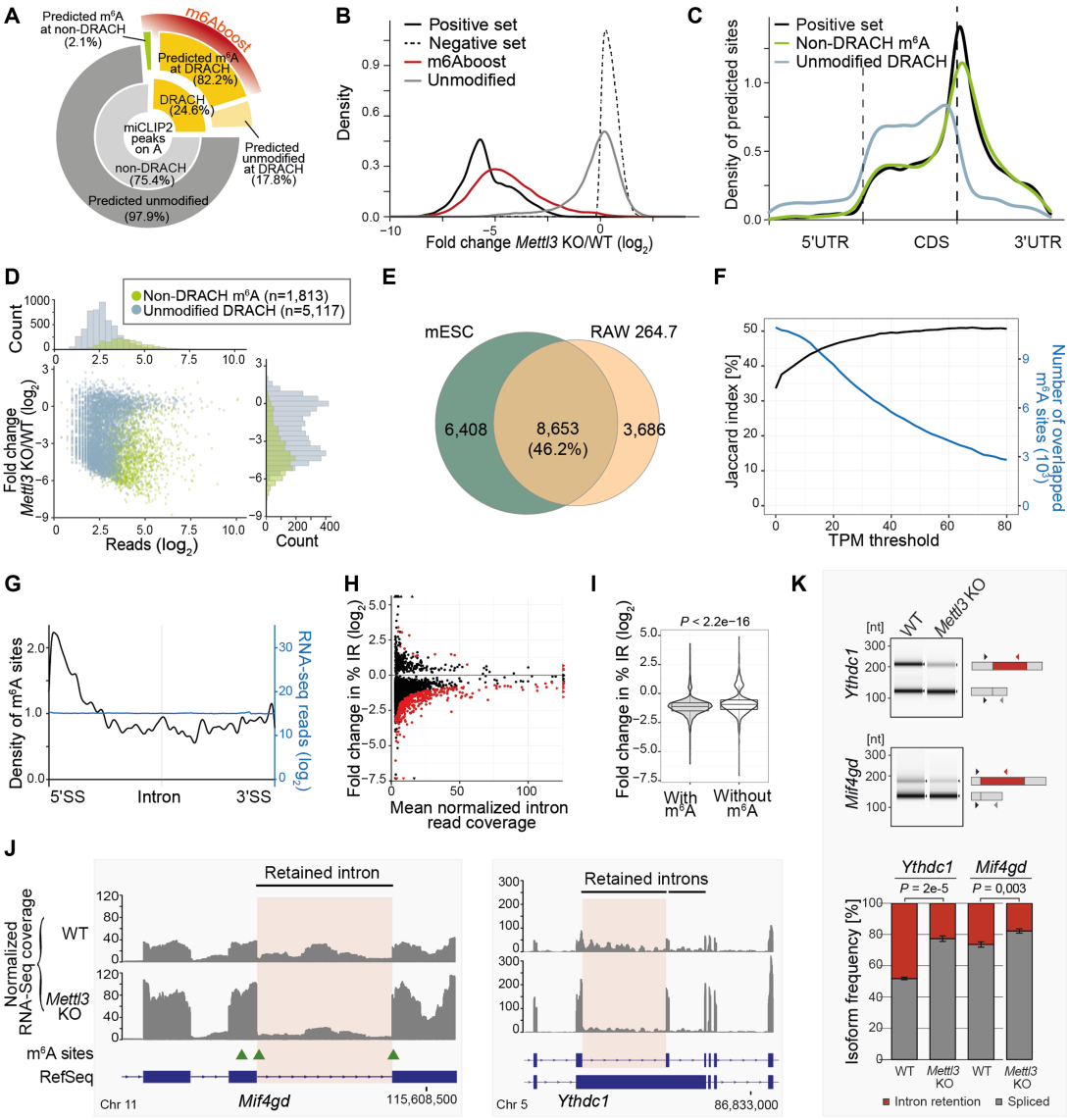


Figure 4. m^{6}A sites occur at non-DRACH motifs and accumulate in retained introns. (**A**) m6Aboost predicts m^{6}A sites at DRACH and non-DRACH motifs in mESC WT miCLIP2 data. Inner circle of donut chart shows occurrence of DRACH ($n = 28,760$, 24.6%) and non-DRACH ($n = 88,382$, 75.4%) motifs for 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 ≥ 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 ≤ 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.

Downloaded from https://academic.oup.com/nar/article/49/16/e92/6307904 by Institut für Recht und Wirtschaft user on 28 February 2023

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).

Downloaded from https://academic.oup.com/nar/article/49/16/e92/6307904 by Institut für Recht und Wirtschaft user on 28 February 2023

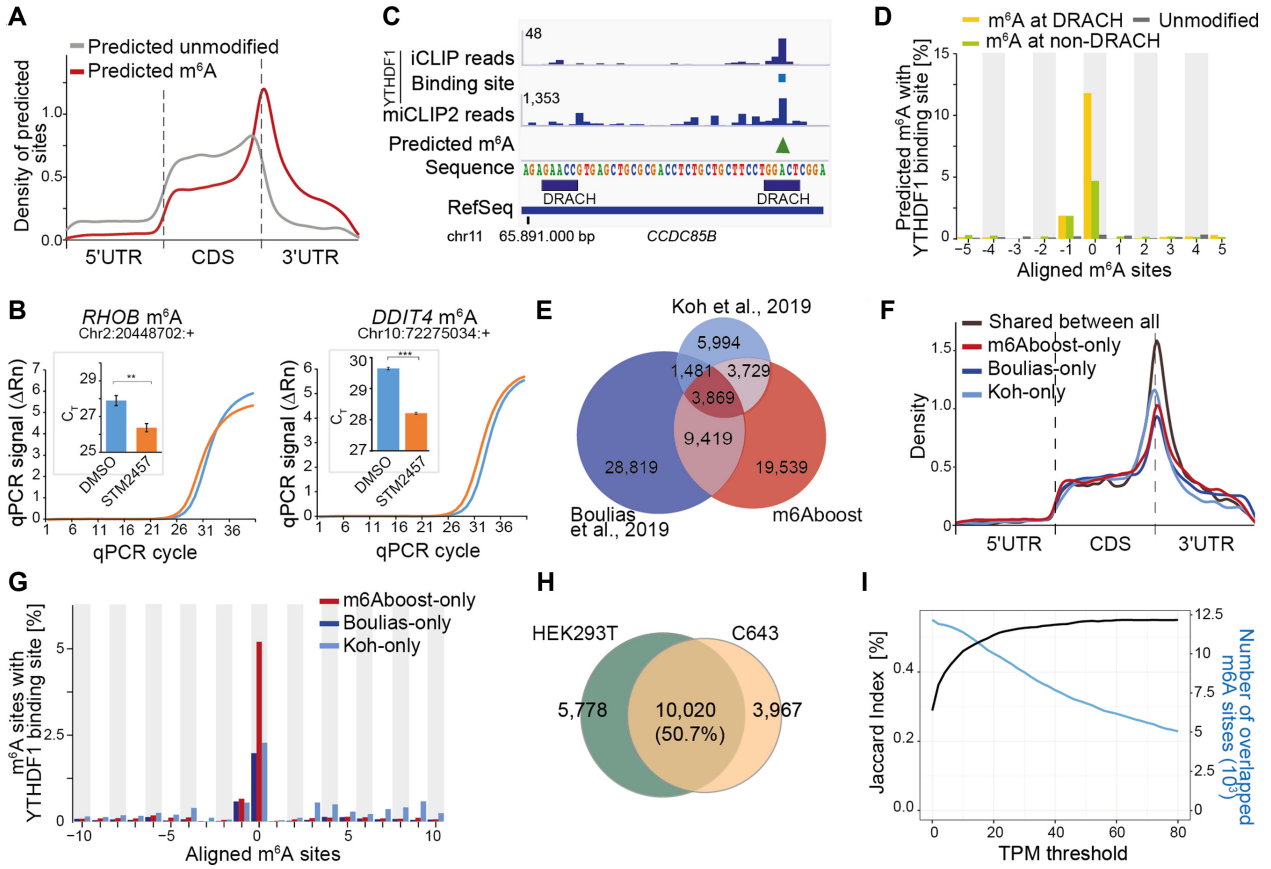


Figure 5. m6Aboost predicts 36,556 m^{6}A sites from HEK293T miCLIP2 data. (**A**) Predicted m^{6}A sites are located around the stop codon. Visualization as in Figure 2F. (**B**) Selected m^{6}A sites were validated by SELECT with HEK293T cells treated with METTL3 inhibitor (STM2457) or DMSO control. 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 ≥ 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

Downloaded from https://academic.oup.com/nar/article/49/16/e92/6307904 by Institut für Recht und Wirtschaft user on 28 February 2023

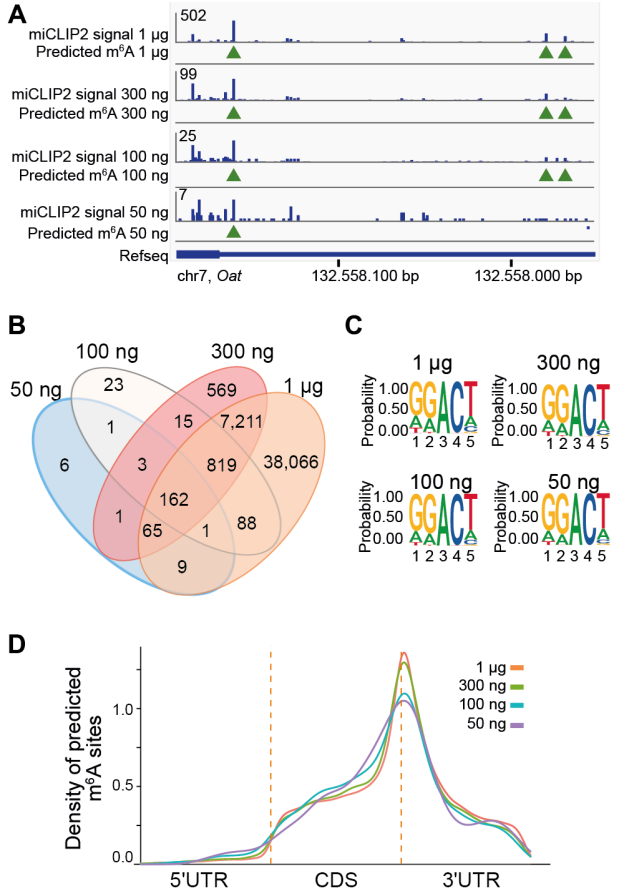


Figure 6. miCLIP2 allows to map m^{6}A sites from low input material (**A**) m6Aboost predicts overlapping m^{6}A sites from miCLIP2 data for different RNA input concentrations. Example genome browser view of the *Oat* gene shows miCLIP2 signals and corresponding m6Aboost predictions (green arrowheads) 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.

Downloaded from https://academic.oup.com/nar/article/49/16/e92/6307904 by Institut für Recht und Wirtschaft user on 28 February 2023

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

In this study, we generated m^6A 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^6A sites are shared between cell lines in either species. Moreover, we confirm that m^6A 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^6A 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^6A 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^6A modifications on alternative splicing. While some studies reported on extensive splicing alterations upon *Mettl3* depletion, others rebutted a strong connection between m^6A 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^6A 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^6A 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

Downloaded from https://academic.oup.com/nar/article/49/16/e92/6307904 by Institut für Recht und Wirtschaft user on 28 February 2023

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.

Downloaded from https://academic.oup.com/nar/article/49/16/e92/6307904 by Institut für Recht und Wirtschaft user on 28 February 2023

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.

Downloaded from https://academic.oup.com/nar/article/49/16/e92/6307904 by Institut für Recht und Wirtschaft user on 28 February 2023

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 ≥ 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 ≤ 3 and support level ≤ 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:

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

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

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

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

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

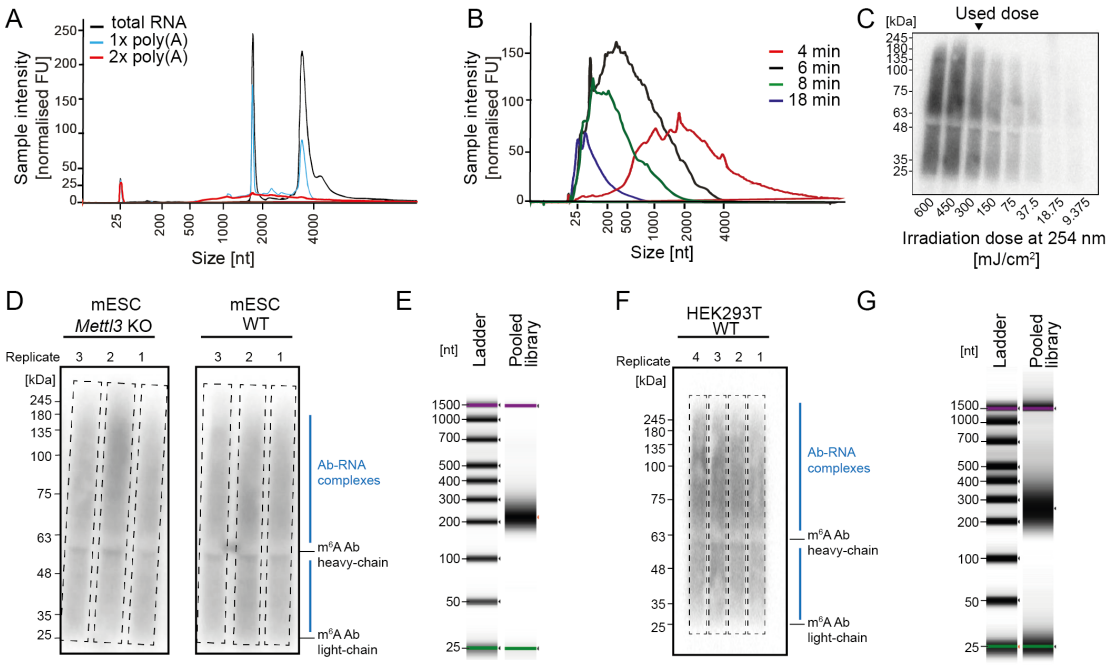
$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (6)$$

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**).

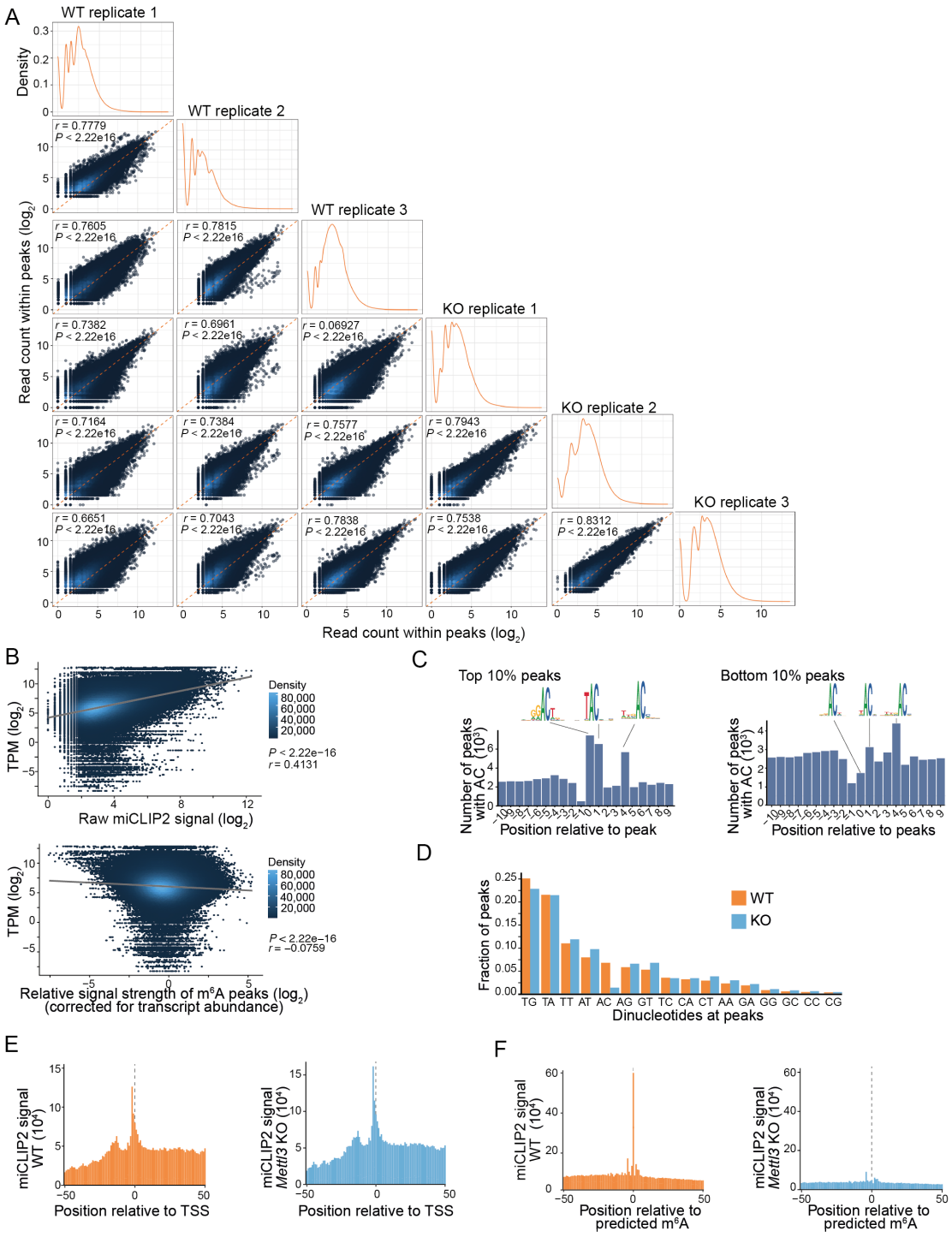
Normalisation of numerical features

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

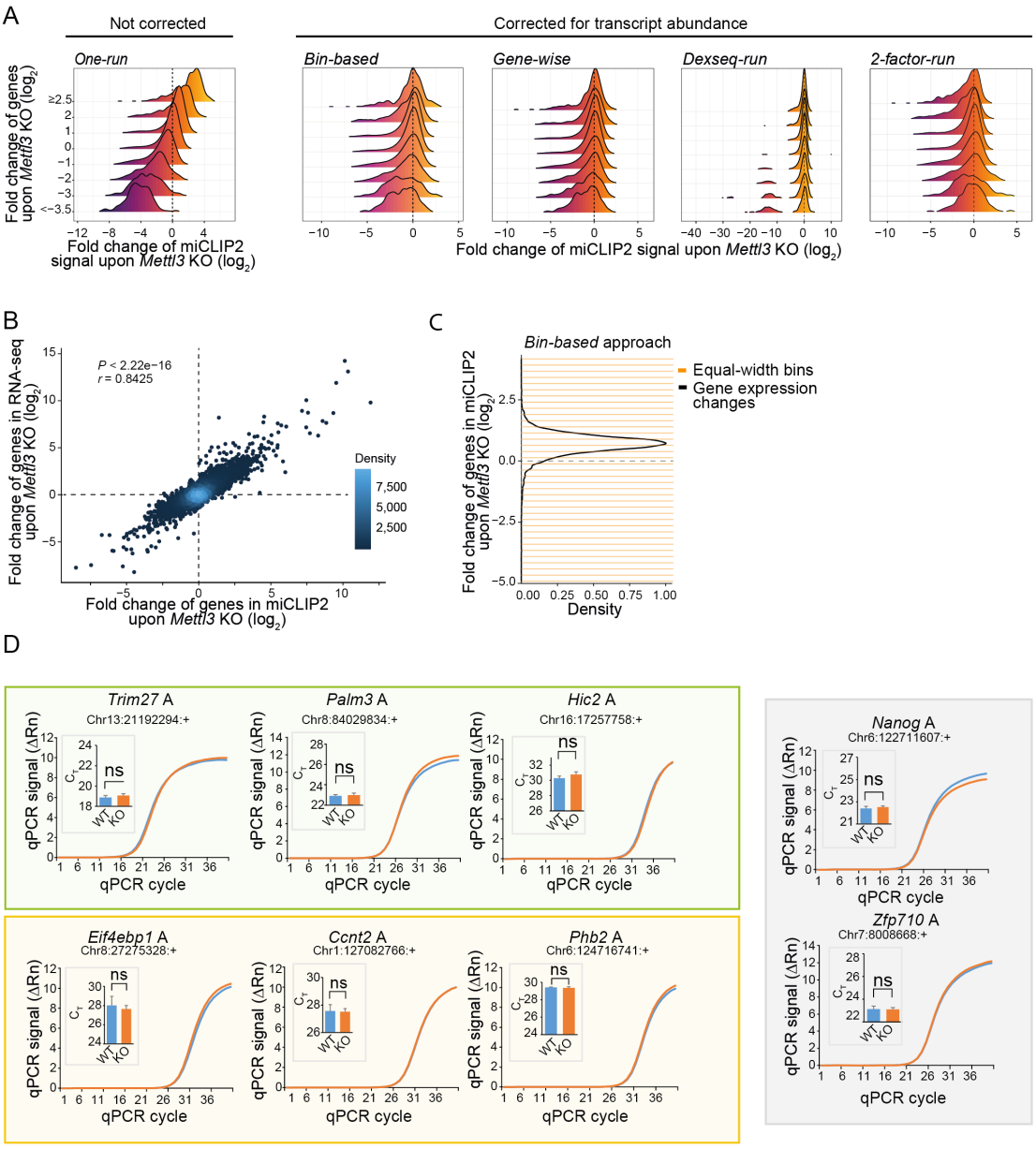


Supplementary Figure S1. miCLIP2 library preparation. ***A-C.*** Optimisation of the miCLIP2 protocol. ***A.*** Ribosomal RNA is fully depleted after two rounds of poly(A) selection. Electropherogram illustrating poly(A) enrichment for 1 µg total RNA of HEK293T cells after one (1x, blue) and two (2x, 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.



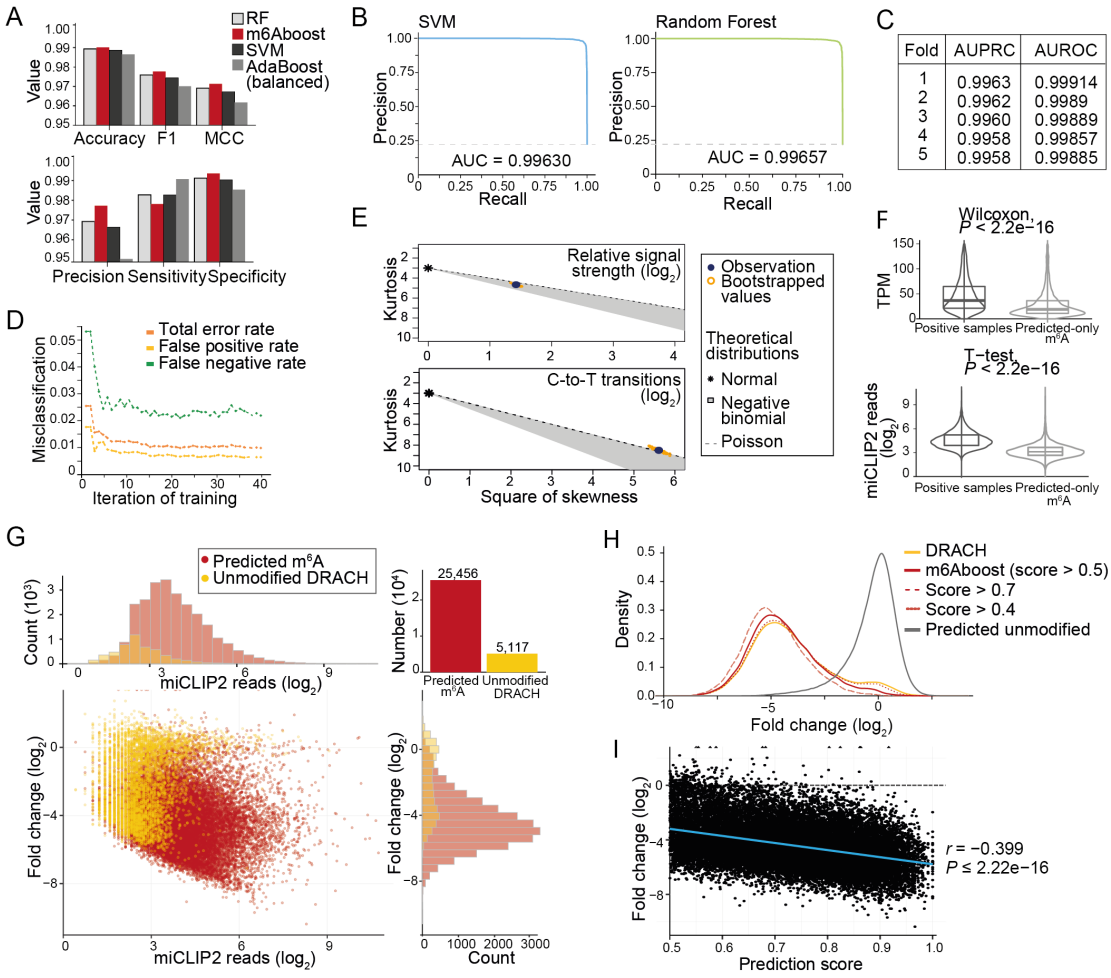
Supplementary Figure S2. Data quality. ***A.*** miCLIP2 libraries are highly reproducible between replicates. Pairwise comparison of truncation read counts within peaks for all replicates from WT and *Mettl3* KO mESCs are shown as an extension of **Figure 1C**. Pearson correlation coefficients (*r*) 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.



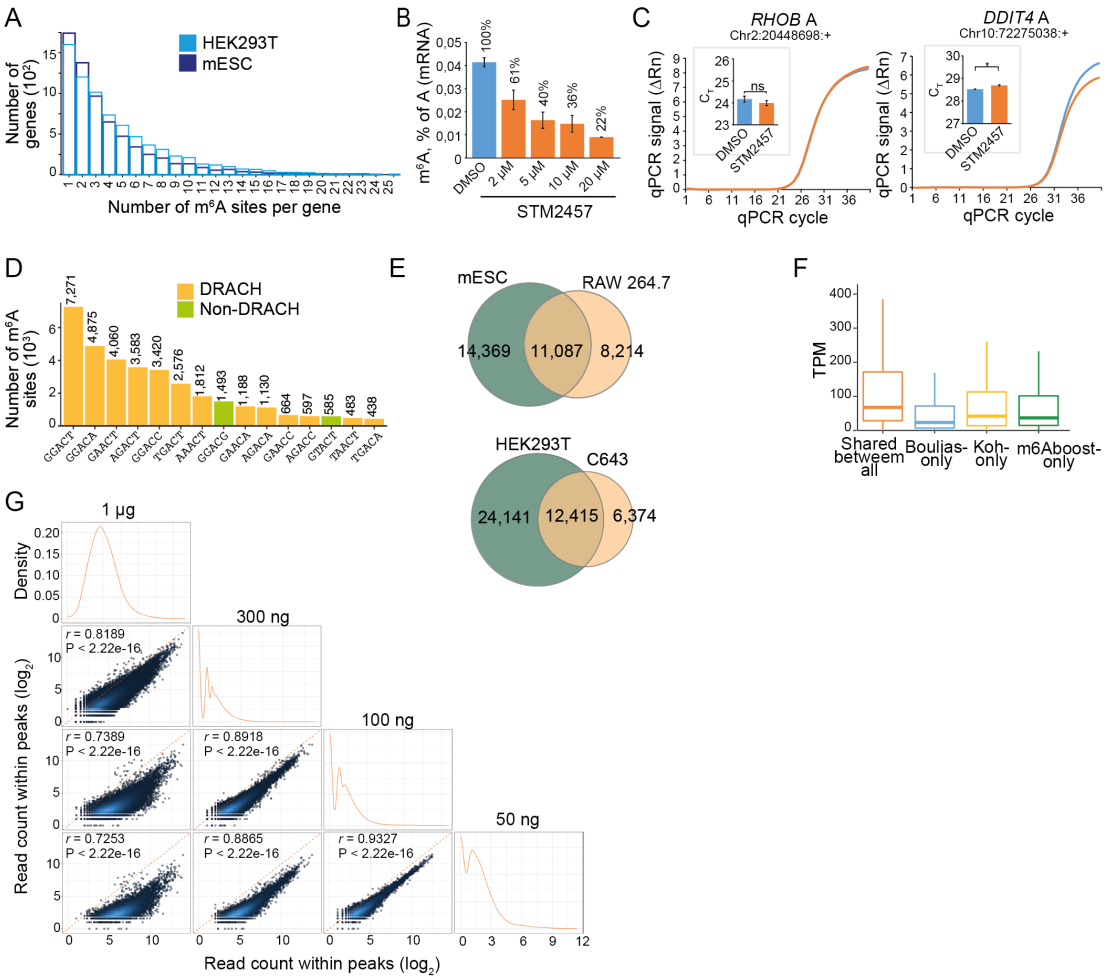
Supplementary Figure S3. Differential methylation analysis and validation. ***A.*** The *bin-based* approach allows to correct for expression changes between WT and *Mettl3* KO. For each tested differential methylation method, the fold change of genes upon *Mettl3* KO (y-axis) and the estimated fold 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.

Chapter 3

Publication 2: RNA stability controlled by m^6A methylation contributes X-to-autosome dosage compensation in mammals.

3.1 Summary

X-to-autosome dosage compensation is a mechanism that is in place to balance the gene expression of one active X chromosome compared to gene products of two active copies of autosomes. Different mechanisms driving X-to-autosome dosage compensation have been identified. Among them, it has been shown that X-chromosomal transcripts reach higher stabilities than autosomal transcripts. How this is accomplished is unknown. m^6A is the most abundant internal mRNA modification and has been shown to play a role in mRNA degradation. Here, we identify that X-chromosomal transcripts are significantly depleted in m^6A sites and thereby are less degraded and reach higher stabilities than autosomal transcripts. When depleting m^6A, we can show that autosomal transcripts become more stable, which indicates that m^6A globally destabilizes those transcripts. Hence, we present m^6A as a novel player that mediates X-to-autosome dosage compensation by destabilizing autosomal transcripts in a selective manner.

3.2 Zusammenfassung

Die X-zu-Autosom-Dosierungs Kompensation ist ein Mechanismus, der die Genexpression eines aktiven X-Chromosoms zu den Genprodukten von zwei aktiven Kopien der Autosomen ausgleicht. Es wurden verschiedene

Mechanismen entdeckt, die den X-zu-Autosom-Dosierungsausgleich bewirken. Unter anderem wurde gezeigt, dass X-chromosomale Transkripte eine höhere Stabilität als autosomale Transkripte erreichen. Wie dies erreicht wird, ist unbekannt. m^6A ist die häufigste interne mRNA-Modifikation und es wurde gezeigt, dass sie eine Rolle im mRNA-Abbau spielt. In diesem Teil der Arbeit können wir zeigen, dass X-chromosomale Transkripte deutlich weniger m^6A-Modifikationen als autosomale Trankripte aufweisen und dadurch weniger abgebaut werden sowie eine höhere Stabilität erreichen. Wenn m^6A in den Zellen verringert wird können wir zeigen, dass autosomale Transkripte stabiler werden, was darauf hindeutet, dass m^6A diese Transkripte global destabilisiert. Folglich präsentieren wir m^6A als neuen Faktor, der den X-zu-Autosom-Dosierungsausgleich vermittelt, indem er autosomale Transkripte auf selektive Weise destabilisiert.

3.3 Statement of contribution

In this project I performed the majority of bioinformatic analysis for this project. I performed all analysis regarding the quantification of m^6A sites on transcripts of different chromosomes, analysis of RNA sequencing and DNA sequencing data as well as all SLAM seq analysis. Furthermore, I contributed to the design of the study and interpretation of results and participated in all project related meetings. I contributed to the text and figures of the manuscript and was involved in all steps of the manuscript preparation and revision.

This paper has been accepted for publication in Nature Structural & Molecular Biology on the 6^{th} of April 2023.

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^{6}A modifications affect mRNA fate in the cytoplasm upon their deposition in the nucleus.

Here, we show that m^{6}A RNA modifications play a key role in X-to-autosome dosage compensation. We find that the m^{6}A 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.

Figure Legends

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)

455 estimated by RNA-seq at different timepoints of m^{6}A depletion (STM2457, 3, 6, 9 and
456 12 h). Boxes represent quartiles, centre lines denote medians, and whiskers extend
457 to most extreme values within 1.5x interquartile range. **D.** Same as ***A***. for human
458 primary fibroblasts (STM2457, 9 h). *P* value = 6.24e-06, two-tailed Wilcoxon rank-sum
459 test. Effect sizes are shown as the shift in medians of the two distributions, expressed
460 as percent of the average IQR of autosomal and X-chromosomal genes (see
461 Methods). **E.** Same as ***B*** for human cell lines. (*P* value = 0.0000803 [human
462 fibroblasts], *P* value = 0.0000379 [HEK293T], *P* value = 0.0003284 [C643], *P* value =
463 0.0002982 [RPE1]. *P* values were calculated as in ***A***, multiple testing correction.

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

490

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.

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^6A 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^6A 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^6A 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^4U concentrations. For optimal incorporation, the s^4U-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^6A 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 are available via the EGA European Genome-Phenome Archive under the accession number EGAS00001007112.

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

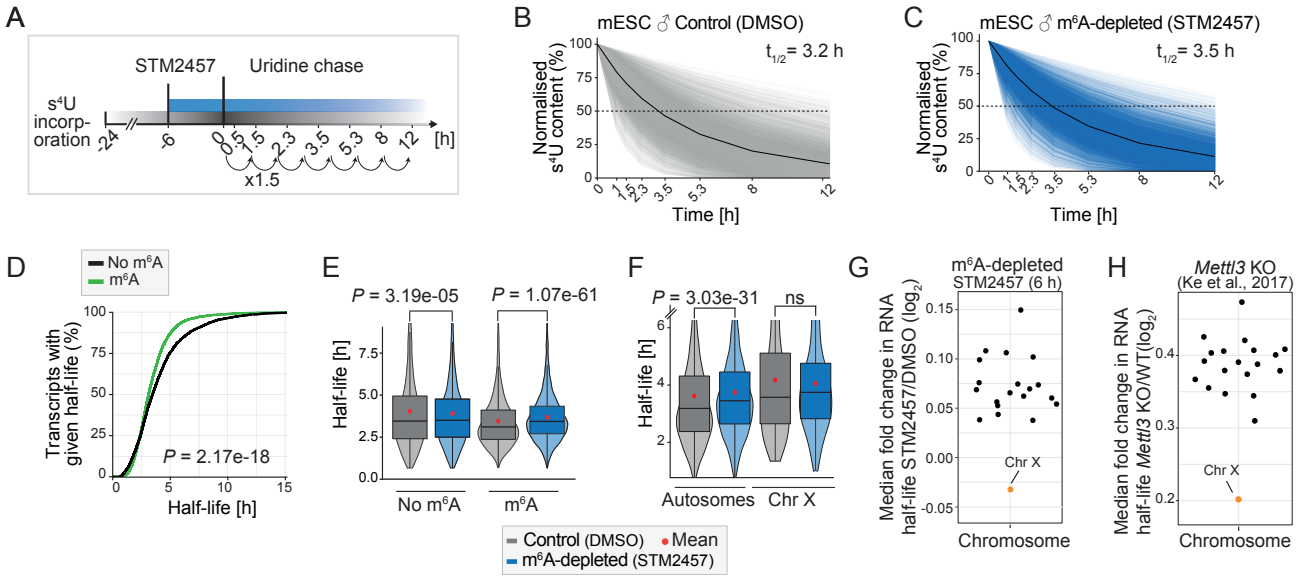


Figure 1. X-chromosomal transcripts are more stable upon m^{6}A depletion. A. Experimental setup for SLAM-seq experiment. **B, C.** Transcripts (*n* = 7,310) in ***B.*** control and ***C.*** m^{6}A-depleted conditions show a median half-life of 3.2 h and 3.5 h, respectively (*P* value = 5.25e-29, two-tailed 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 log2-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

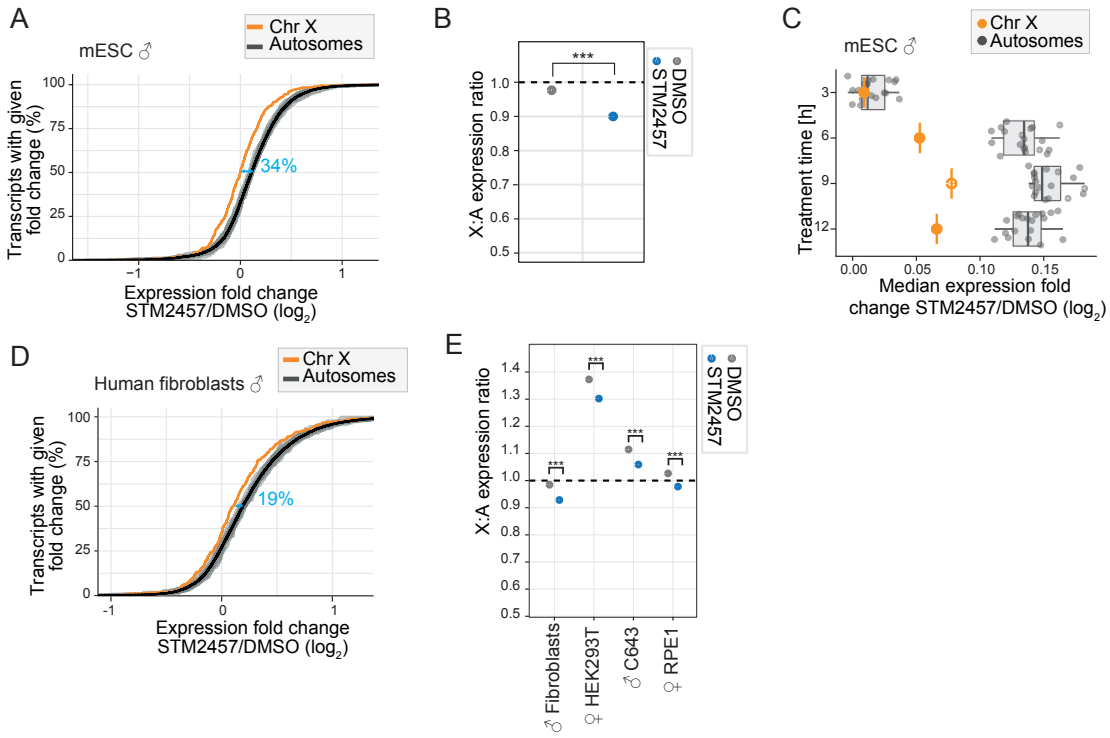


Figure 2. X-chromosomal transcripts are more stable and less upregulated upon m^{6}A depletion. A. X-chromosomal transcripts are less upregulated upon m^{6}A depletion in male mESC (*P* value = 1.86e-17, two-tailed Wilcoxon rank-sum test). Cumulative fraction of transcripts (RPKM > 1) on individual autosomes (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 [HEK-293T], *P* value = 0.0003284 [C643], P value = 0.0002982 [RPE1]. *P* values were calculated as in ***A***, multiple testing correction.

Figure 3

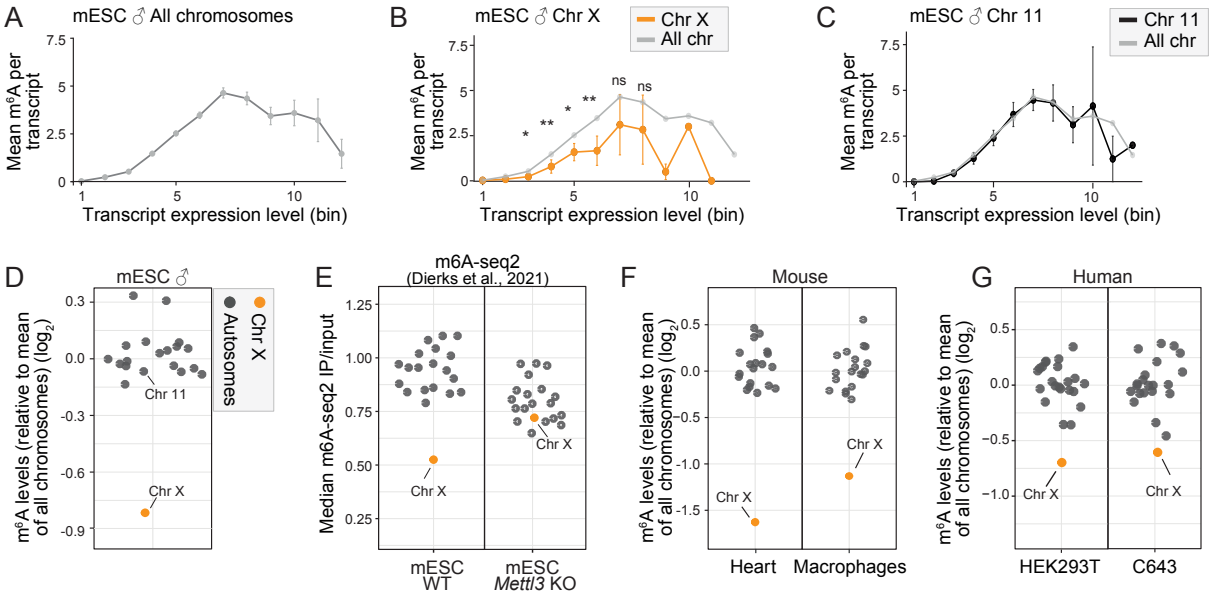


Figure 3. m^{6}A sites are reduced on transcripts from the X chromosome. A. The number of detected m^{6}A sites varies with the expression level. Mean m^{6}A sites per transcript were quantified for transcripts with each expression bin (*n* = 12,034 transcripts, see **Extended Data Fig. 6A** for *n* 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

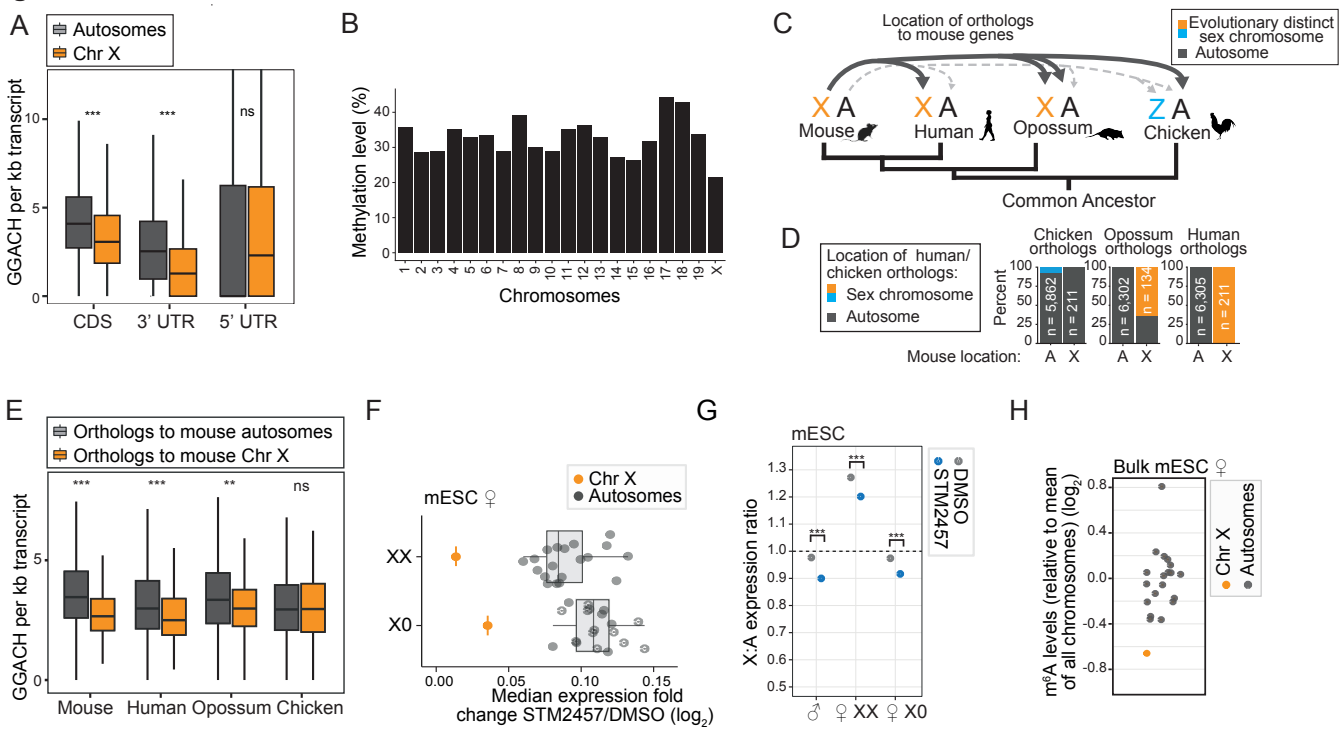


Figure 4. Reduced m^{6}A on X-chromosomal transcripts is intrinsically encoded. A. GGACH motifs (normalised to region length) in different transcript regions of autosomal (grey) and X-chromosomal transcripts (orange) in mouse (*P* value = 1.38e-29 [CDS, *n* = 16,631 annotations], *P* value = 1.06e-40 [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

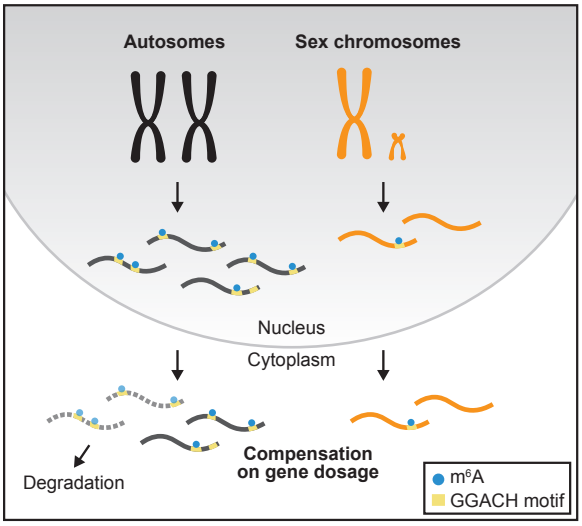
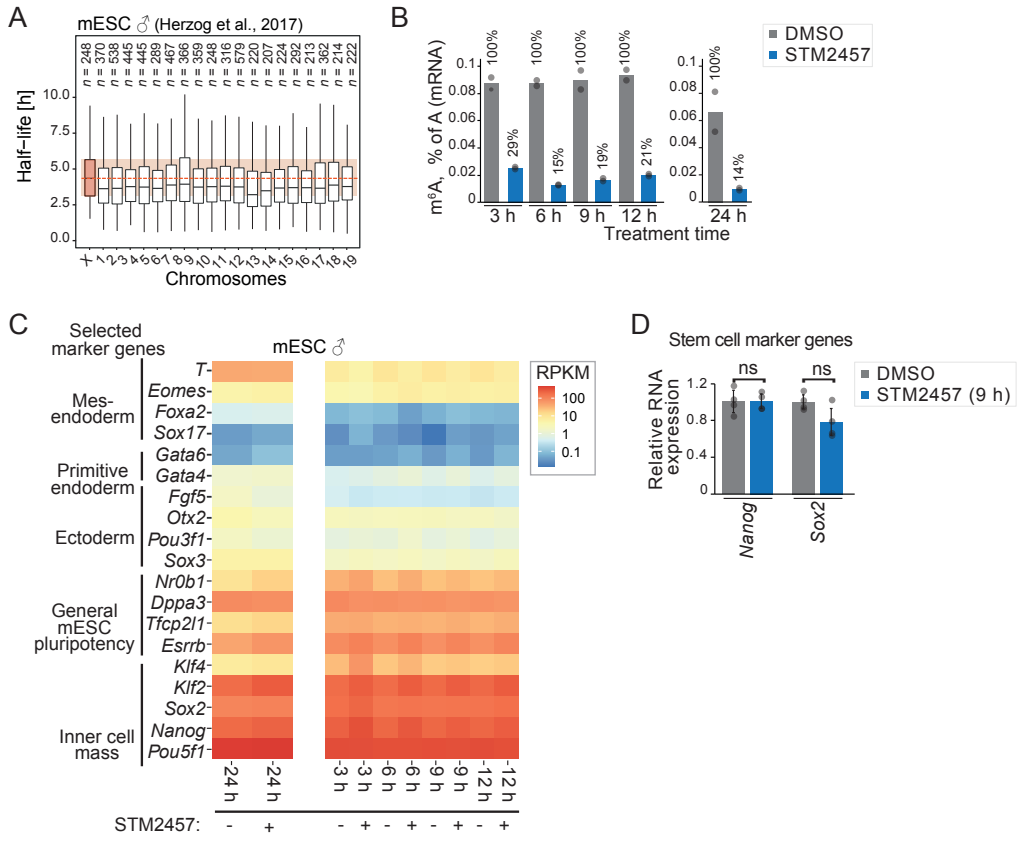
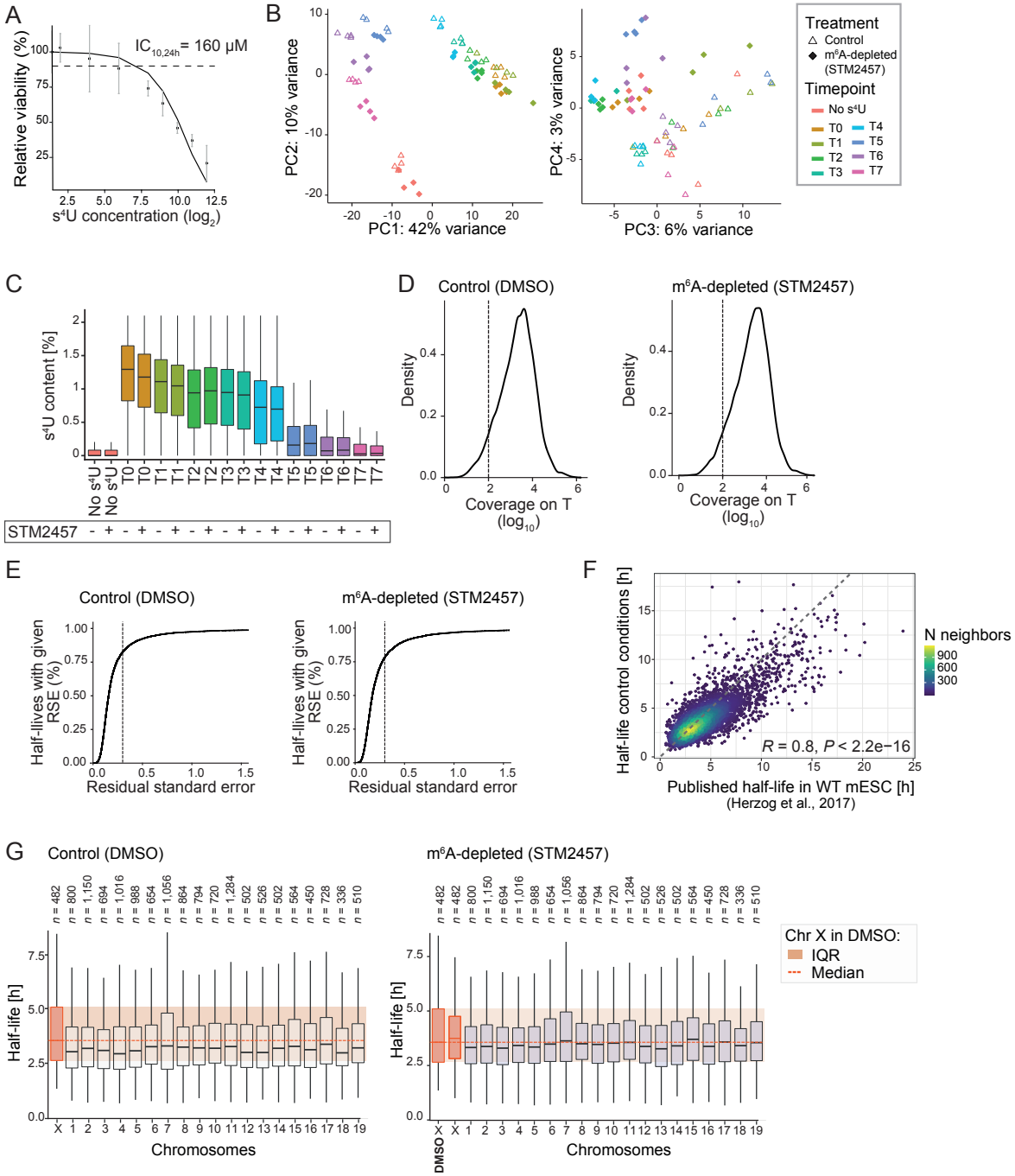


Figure 5. The role of m^{6}A in X-to-autosome dosage compensation. m^{6}A acts as a selective degradation signal on autosomal transcripts and thereby contributes to X-to-autosome dosage compensation. Transcripts from the autosomes are transcribed from two active chromosomes, leading to higher transcript 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.

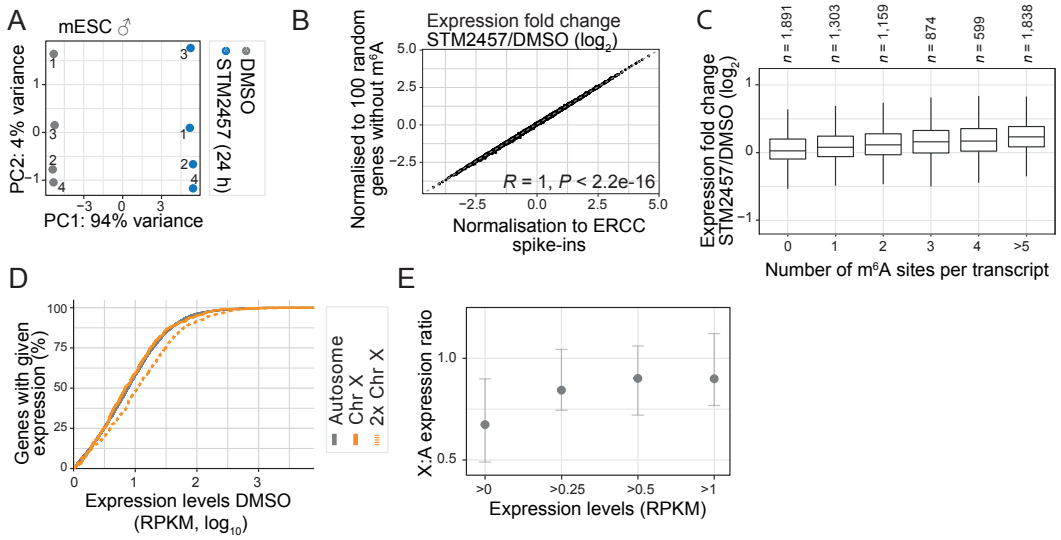
Extended Data Figure 1



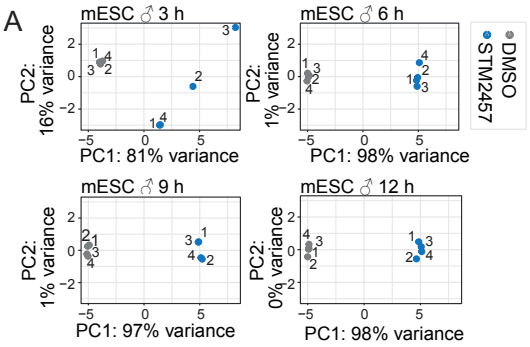
Extended Data Figure 2

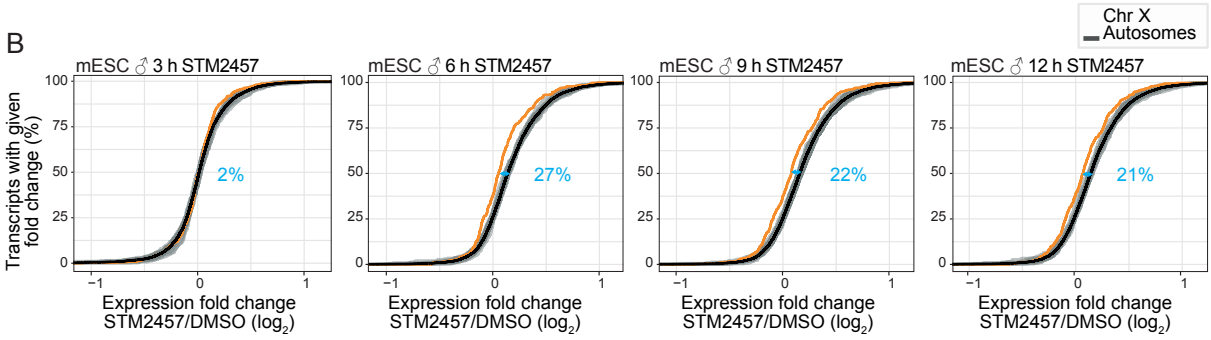


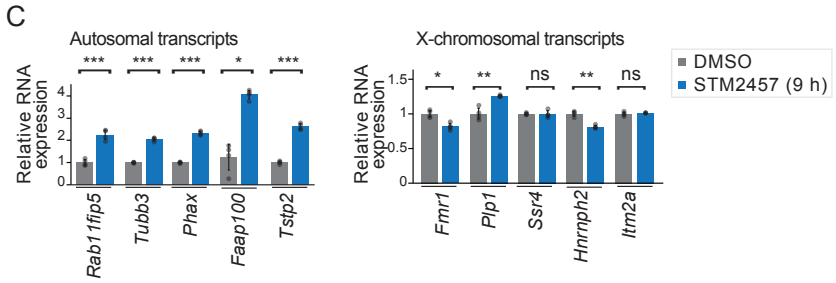
Extended Data Figure 3



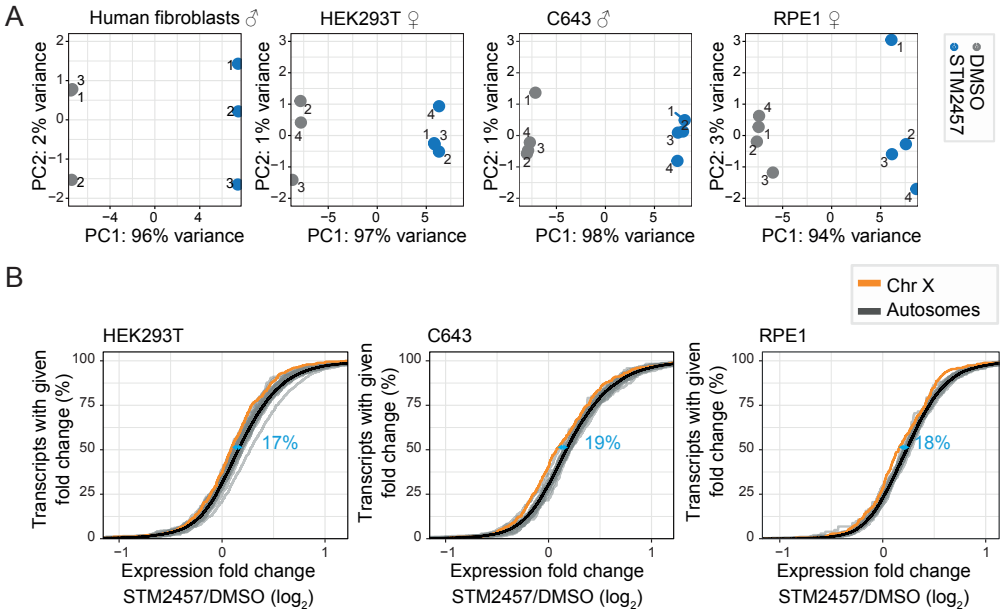
Extended Data Figure 4



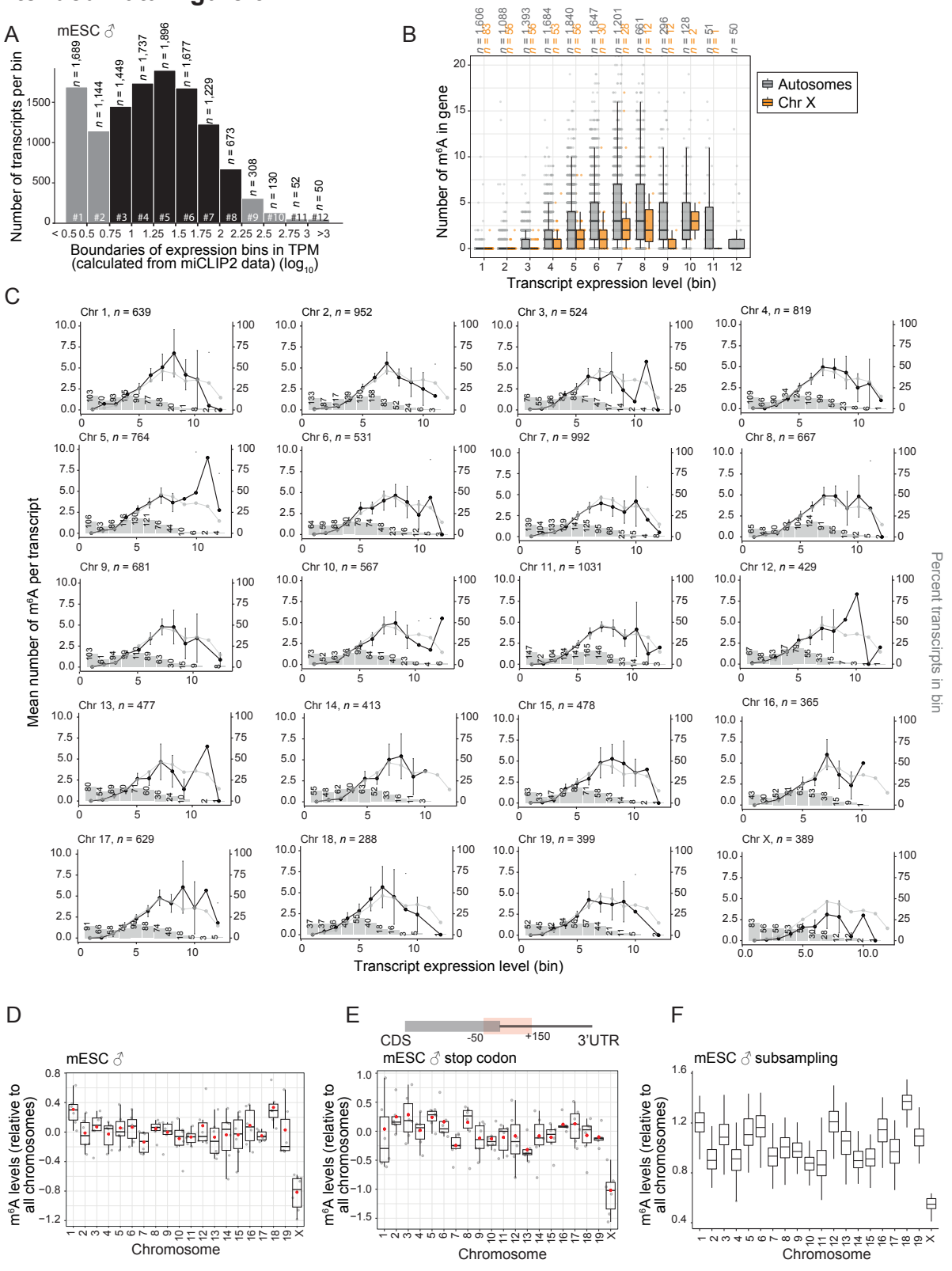




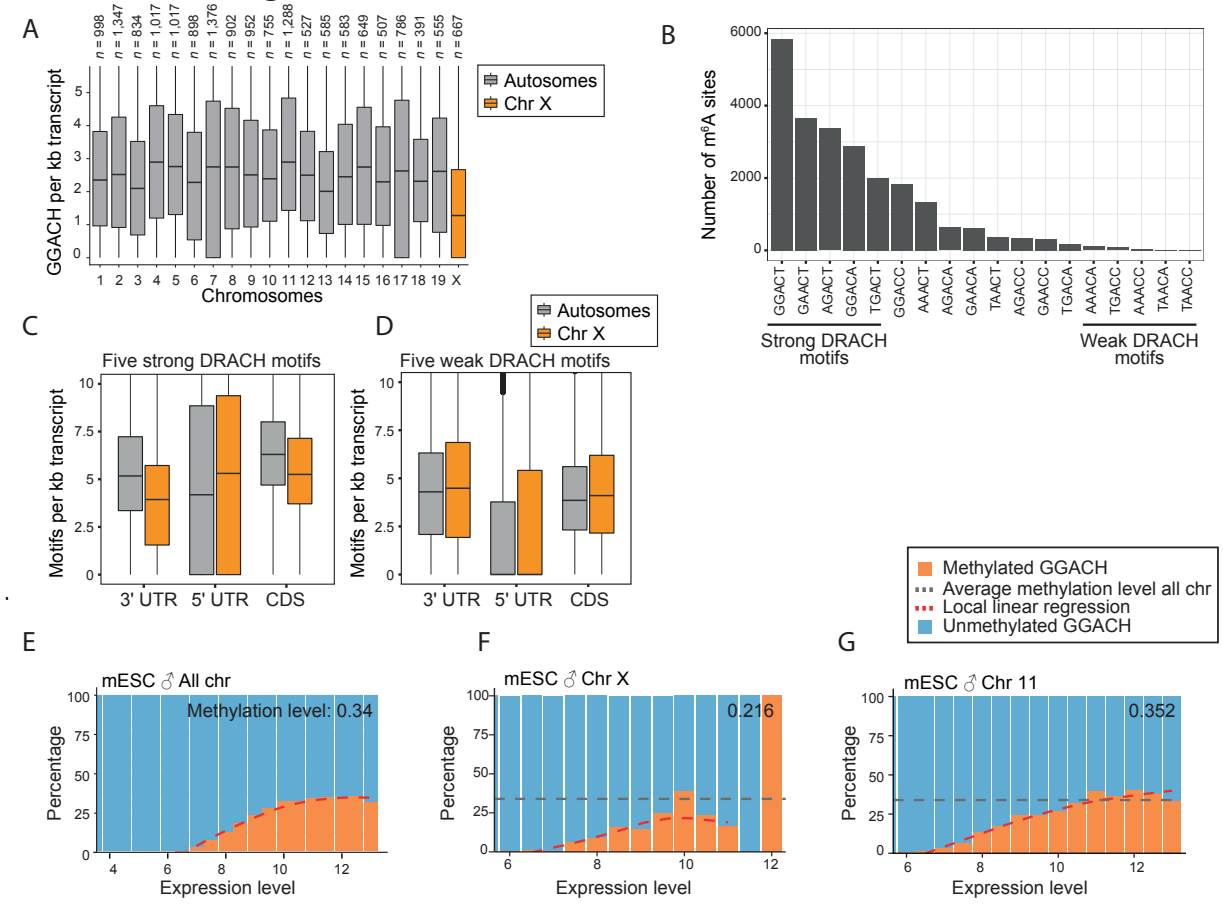
Extended Data Figure 5



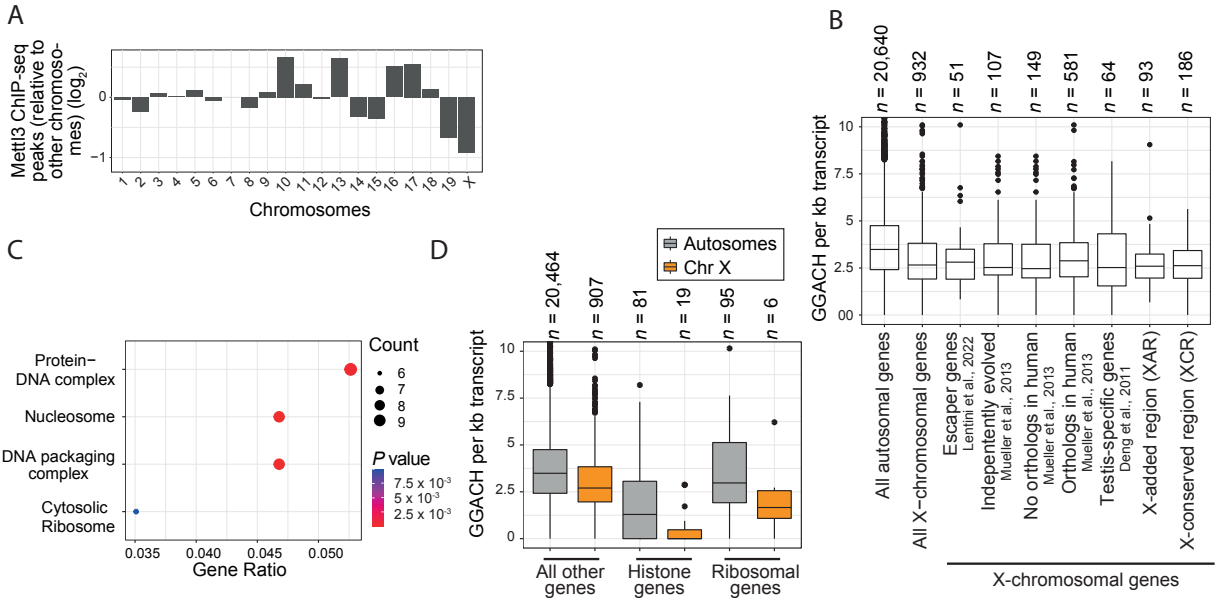
Extended Data Figure 6



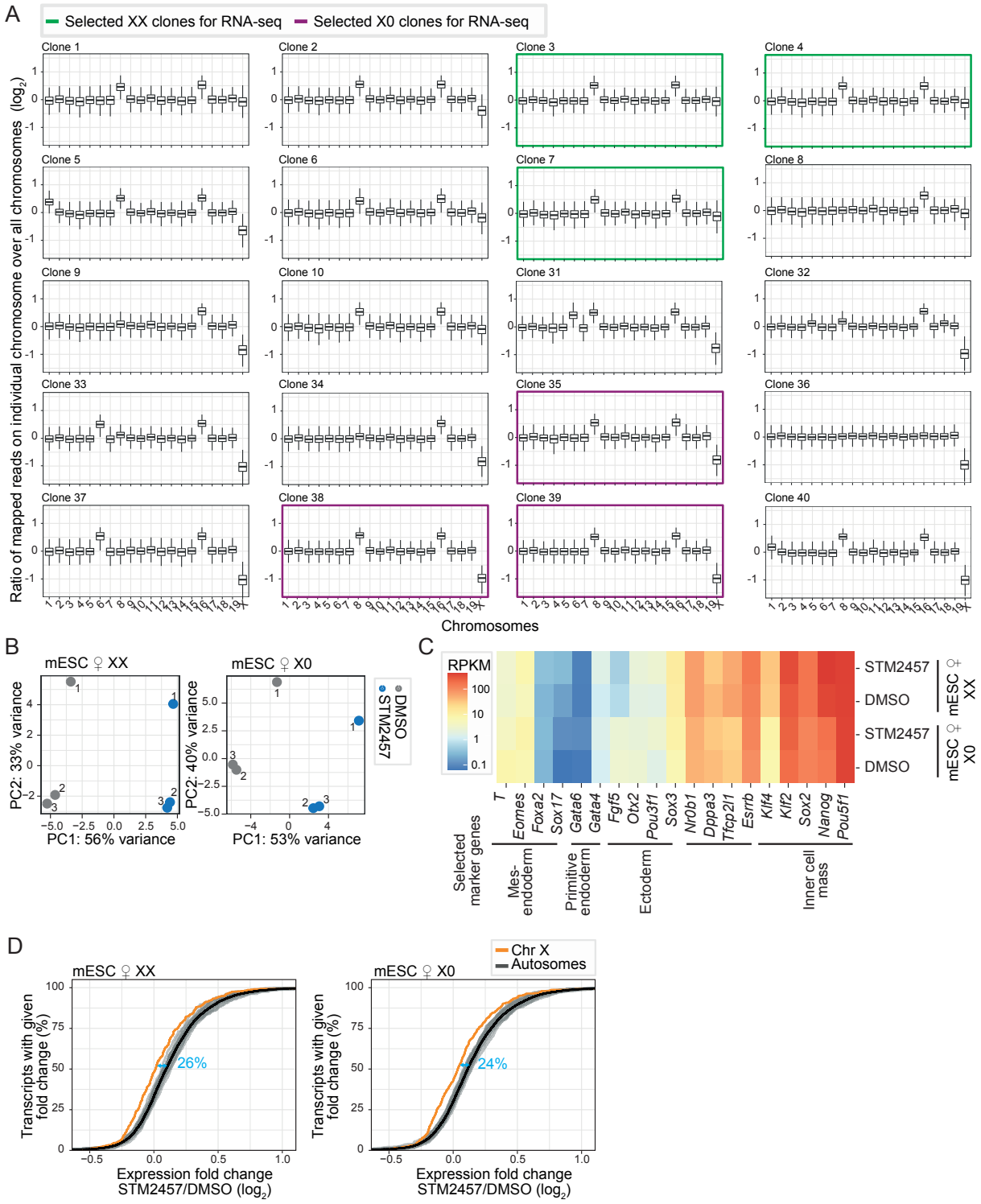
Extended Data Figure 7



Extended Data Figure 8



Extended Data Figure 9



Extended Data Figure Legends

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^4U 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^6A-depleted conditions (symbols). **C.** T-to-C conversions on T's by the overall T coverage per 3' UTR. Maximum s^4U rate is achieved after 24 h of labelling (T0) and steadily decreases after s^4U washout and uridine chase (T1-T7). Unlabelled samples (No s^4U) 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^6A-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^6A-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^{6}A sites per transcript for all mouse chromosomes. Data points indicate mean number of m^{6}A 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^{6}A sites per transcripts for expression bins #3-8 (*n* of mean of expression bins = 6) on an individual chromosome over the mean m^{6}A 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^{6}A 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^{6}A 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^6A sites from mESC miCLIP2 data across different DRACH motifs. Barplot shows the number of m^6A 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^6A 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).

Supplementary Material

Table of content:

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 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^6A sites predicted by m6Aboost was calculated. Since m^6A 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^6A 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^6A 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^6A sites in transcripts. To analyse the general influence of the chromosome type on the number of m^6A sites in transcripts, categorical generalised linear mixed models for negative binomial data were fitted to the data (see Methods "Statistical analyses of m^6A sites in transcripts"). This table summarises the fitted values and 95% confidence intervals (Wald type) of the fold changes (log_2) of expected m^6A 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).

Chapter 4

Discussion

In this work, we improved the accurate detection of m^6A modifications and discovered a novel role of the modification in the regulation of gene expression. We overcame several limitations of the m^6A detection by improving the experimental miCLIP protocol as well as coupling it with an extensive bioinformatic pipeline and machine learning classifier. Using the resulting new high-confidence m^6A annotations, we discovered a novel role of the modification in X-to-autosome dosage compensation.
In Chapter 2 (publication 1), we developed an enhanced miCLIP2 protocol and were able to distinguish true m^6A signal from unspecific background signals in the data. We could identify over 20,000 m^6A modifications transcriptome-wide in a single-nucleotide resolution. Using this annotation, we discovered in Chapter 3 (publication 2) that X-chromosomal transcripts are significantly depleted of the modification and that those transcripts are therefore more stable than m^6A-rich autosomal transcripts. We thereby provide new insights into a novel X-to-autosome dosage compensation that acts on a post-transcriptional level. Collectively, we enhanced the detection of m^6A modifications and described a novel role of the modification in X-to-autosome dosage compensation, which will be discuss in the following section.

4.1 An improved miCLIP2 protocol allows for the accurate detection of m^6A

Transcriptome-wide m^6A detection has been a challenge for several years. Ten years ago, the first high-throughput m^6A detection methods have been established (Dominissini et al., 2012; Meyer et al., 2012). These techniques paved the way for a rapidly-evolving m^6A research field. The development

of miCLIP in 2015 (Linder et al., 2015) allowed the transcriptome-wide detection of m^{6}A in a single-nucleotide resolution, but also suffered several limitations such as high costs, high input materials and high background signals due to insufficient m^{6}A-antibody specificity (McIntyre et al., 2020; Meyer, 2019a; Garcia-Campos et al., 2019). To overcome these limitations, firstly, we modified the miCLIP protocol and adapted the recently developed iCLIP2 protocol which uses lower input materials and produces high-complexity libraries at less costs (Buchbender et al., 2020). Secondly, we performed miCLIP2 in mouse embryonic stem cell (mESC) WT as well as *Mettl3* KO cells to differentially identify signals that are lost in the *Mettl3* KO (Geula et al., 2015). This allowed us to specifically distinguish true m^{6}A sites from unspecific background signals stemming from, for example, antibody cross-reactivities. Our identified high confidence m^{6}A sites were mostly located within DRACH motifs as well as around the stop codon as described before (Dominissini et al., 2012; Meyer et al., 2012; Linder et al., 2015). Thirdly, we used the identified high-confidence m^{6}A sites to train a machine learning classifier based on the footprint of true m^{6}A sites in miCLIP2 datasets. We applied the machine learning classifier to human miCLIP2 datasets which resulted in a m^{6}A prediction which was largely overlapping with previously identified m^{6}A sites (Boulias et al., 2019; Koh et al., 2019). Thus, our classifier can be applied to new miCLIP2 datasets and specifically distinguish between the signal of true m^{6}A sites and unspecific background signal. Without this machine learning classifier, it is only possible to filter out background signal in the data by comparing it to complementary experiments that were conducted in cells lacking m^{6}A. This was for a long time a limiting factor, since generating *METTL3* KO has been challenging (Geula et al., 2015; Poh et al., 2022, see subsection 1.2.3). Further, the machine learning classifier also overcomes the limitations of the high background signals in the miCLIP2 data. It is specifically able to distinguish m^{6}A signal from unspecific background signals. This enhances the detection of m^{6}A and makes the use of a complementary *METTL3* KO dataset obsolete.

4.1.1 The m^{6}A code

m^{6}A is deposited in a DRACH-sequence dependent manner. In fact, previous miCLIP protocols have filtered for signals only at positions within DRACH motifs and excluded all other signals as background (Linder et al., 2015; Hawley and Jaffrey, 2019). Due to comparing miCLIP2 signals from mESC

WT as well as *Mettl3* KO samples, we identified positions that are significantly depleted of signals in the *Mettl3* KO independently of the underlying sequence motif. This revealed that some m^{6}A signals are located outside of DRACH motifs. Importantly, non-DRACH motifs have also been identified in Oxford Nanopore Technologies sequencing-based m^{6}A detection data as well as recent antibody-independent m^{6}A detections (Pratanwanich et al., 2021; Liu et al., 2022). This indicates that m^{6}A can, in rare cases, occur outside of DRACH motifs and that filtering miCLIP signal only for DRACH motifs excludes those m^{6}A sites which are detected by our novel machine learning classifier.

The DRACH motif encodes for several different sequence motifs. It is commonly observed that not all DRACH motifs are methylated in a similar manner, but that some motifs carry more often m^{6}A sites than others (Linder et al., 2015; Pratanwanich et al., 2021; Liu et al., 2022). It remains unclear why some motifs are more likely to be modified and whether different DRACH motifs are important for different functions of m^{6}A. One possible explanation could be a stronger sequence preference by the methyltransferease complex towards specific DRACH motifs.

One question about the m^{6}A deposition in the field remains: *Is there an m^{6}A-determining code beyond the DRACH motif?* In addition to the sequence preference of m^{6}A within the DRACH motifs, m^{6}A is specifically enriched in specific locations of the transcript: our novel annotation resembles the previously identified m^{6}A pattern that accumulates around the stop-codon (Dominissini et al., 2012; Meyer et al., 2012; Schwartz et al., 2014). Furthermore, a recent study detected that m^{6}A occurs within m^{6}A-clusters that are specific enrichments of several m^{6}A sites in a close window. These m^{6}A clusters were also enriched around the stop-codon and shown to be influencing RNA stability as well as translation (Liu et al., 2022). This demonstrates that in addition to a sequence preference within a subset of specific DRACH motifs, m^{6}A is also preferably deposited in specific regions of the transcripts. Consequently, further factors determining m^{6}A deposition might contribute to the decision of m^{6}A deposition. Recently, it was shown that a 50 nucleotides downstream motif of the m^{6}A site contributes to the m^{6}A deposition decision. However, which factors are influenced by this motif and how the motif deposition is regulated remains unclear (Luo et al., 2022).

In contrast, another recent study proposed a model in which no underlying additional sequence to the DRACH motif or other structure influences the

m^6A methylation. They propose that rather all methylation motifs are methylated except for those close to a splice junction (Uzonyi et al., 2023). This would be resembling DNA methylations, which are often methylating all target sequences by default, unless hindered (Uzonyi et al., 2023; Hsieh, 2000). This observation was supported by Yang et al., 2023, who have identified that the m^6A deposition by METTL3 is hindered by the exon-junction complex. Therefore, m^6A modifications accumulate in long exons and in the last exon around the stop codon (Yang et al., 2022). Furthermore, it has been suggested that the histone modification H3K36me3 plays an important role in marking the positions for the m^6A deposition globally (Huang et al., 2019). Additionally, post-translational modifications of METTL3 or WTAP have been shown to influence the methyltransferase activity (Du et al., 2018; Sun et al., 2020). It will be interesting to dissect if the m^6A deposition is influenced by several different factors such as histone modifications, internal sequence motifs or by the exon architecture of a transcript. Investigating different transcript constructions in reporter assays would allow to understand the influencing factors of m^6A deposition further.

4.1.2 m^6A quantities on mRNAs: largely underestimated?

Recently, besides miCLIP2, several novel m^6A-detection techniques have been developed. With these techniques, new estimations of m^6A sites have been far higher than previous ones: our m^6A estimations using m6Aboost range between 20,000-40,000 m^6A sites, which was slightly higher than previous estimations of 10,00-15,000 m^6A sites using miCLIP or other techniques (Dominissini et al., 2012; Meyer et al., 2012; Linder et al., 2015). New studies however, are describing over 100,000 m^6A sites on mRNAs suggesting that previous reports have massively underestimated the m^6A content on mRNAs (Uzonyi et al., 2023; Tegowski et al., 2022; Liu et al., 2022). One possible reason for such underestimations could be the lack of signal-detection in lowly expressed transcripts (McIntyre et al., 2020). Lowly expressed transcripts lead to low numbers of sequencing reads and techniques that rely on e.g. peak calling do not detect signals in those transcripts and hence, no m^6A is detected. This underestimation becomes apparent when we quantified m^6A sites with regard to their expression levels in the second part of our project: lowly expressed transcripts had less m^6A sites predicted than higher expressed transcripts. This is in line with previous estimations that a sufficient read coverage is necessary to reliably identify m^6A in a transcript using

m^{6}Aseq (McIntyre et al., 2020). Hence, it is indeed likely that a large fraction of m^{6}A sites is not detected when using miCLIP or m^{6}A-seq/MeRIP seq.
New quantifications of m^{6}A are not only higher than estimations based on miCLIP or m^{6}A-seq/MeRIP seq but also higher than the first estimations from the 1970s. Those estimations were not based on high-throughput sequencing data or antibody enriched datasets and were in the range of 1-2 m^{6}A sites per mRNA (Desrosiers et al., 1974; Wei et al., 1975; Perry and Kelley, 1976). A possible explanation for this has been described recently: although early analyses were not influenced by detection biases due to mRNA expression, these studies have based their m^{6}A estimations on the assumption that m^{6}A is a stoichiometric modification. However, later it has been shown that m^{6}A is non-stoichiometric: m^{6}A sites are only modified in a small subset of mRNAs of one kind and therefore many more positions can be modified in a low-stoichiometric manner (reviewed in Murakami and Jaffrey, 2022). Thus, early estimations reached only from 1-2 m^{6}A sites per mRNA, but it is likely that m^{6}A numbers per transcript are much higher than previously thought (Uzonyi et al., 2023).

4.1.3 miCLIP2 and m6Aboost: remaining limitations and the need of a gold standard of m^{6}A detection

Despite improving the detection of m^{6}A using miCLIP2, the method still suffers from limitations. One major drawback of deep-sequencing-based techniques is the reliance on transcript abundance. Therefore, m^{6}A numbers remain underestimated in lowly expressed transcripts (McIntyre et al., 2020). Furthermore, miCLIP2 does not give information about the stoichiometry of m^{6}A. m^{6}A is a low-stoichiometric modification and there is a need for the individual quantification of m^{6}A sites (reviewed in Murakami and Jaffrey, 2022). Using miCLIP2, it is only possible to obtain a binary information if a nucleotide was modified or unmodified lacking quantitative information. Other methods are able to detect the stoichiometry of different m^{6}A sites (e.g. Garcia-Campos et al., 2019; Liu et al., 2022). Using this information, it would be interesting to investigate the miCLIP2 data in more detail. Looking at a subset of m^{6}A sites that have been shown by other detection methods to be low-stoichiometric could reveal a different pattern of those sites miCLIP2 data and might indicate that we can estimate m^{6}A stoichiometry in miCLIP2 data to a small extent. Another limitation is that our machine-learning classifier is specifically trained and suited for specific data that was produced by

the miCLIP2 protocol. Different m^{6}A antibodies introduce a variety of truncation and mutation patterns (Linder et al., 2015) and m6Aboost might not reach similarly high performances when miCLIP2 is performed using different antibodies. Additionally, m6Aboost was trained on mESC data and might therefore not be as well-suited for the detection of m^{6}A data derived from different species in which the m^{6}A distribution can be different, such as in *Drosophila*, where m^{6}A has been shown to be mostly located in 5'UTRs (Kan et al., 2017). One possibility to overcome this limitation would be to train m6Aboost on additional data obtained by performing miCLIP2 data using different antibodies or in different species.

m^{6}A research is a rapidly evolving research field with numerous new insights each month (226 new publications on pubmed 2022/11/1-2022/11/30). Despite the enormous numbers of studies on m^{6}A and datasets that quantify the modification in different cell types and conditions, it remains difficult to compare m^{6}A quantifications of different studies among each other due to the usage of a wide repertoire of different detection methods. Since antibody-based m^{6}A detection has been shown to suffer from several limitations, novel antibody-free techniques have been developed recently. However, antibody-based methods such as MeRIP seq and miCLIP2 still remain the most commonly used techniques (McIntyre et al., 2020) owing to the fact that many antibody-free methods have their own set of limitations. For example, the recently developed MAZTER-seq relies on a restriction enzyme which exclusively detects ACA motifs and thereby is restricted to those m^{6}A sites that lie within this motif (Garcia-Campos et al., 2019). Another antibody-free technique, DART-seq, is based on the fusion of the YTH-domain to APOBEC which specifically introduces C-to-U mutations close to the YTH-domain-recognized m^{6}A sites, allowing for the detection of m^{6}A. This method has the limitation of only detecting m^{6}A sites that are recognized by YTH-domains (Meyer, 2019a).

More recent techniques however might overcome the limitations of antibody-based methods as well as limitations of antibody-free detection methods. Two very recently developed techniques are GLORI (Liu et al., 2022) and eTAM-seq (Xiao et al., 2023), both based on deamination of unmodified adenosines and thereby detecting all modified m^{6}A sites transcriptome-wide without the need for antibodies, restriction enzymes or fusion proteins. These techniques are promising new directions in the field and could potentially become the gold standard for m^{6}A detection. Furthermore, m^{6}A detection by direct RNA sequencing of the Oxford Nanopore Technologies is a rapidly

evolving field (Furlan et al., 2021). Several groups have developed machine learning or deep learning-based algorithms to specifically identify m^6A in direct RNA sequencing data (e.g. Cozzuto et al., 2020; Pratanwanich et al., 2021). m^6A detection from direct RNA overcomes experimental preparation steps such as cDNA synthesis, RNA fragmentation or PCR amplifications and also allows the detection of several RNA modifications at the same time (Mateos et al., 2022; Moshitch-Moshkovitz et al., 2022). This will not only advance the detection of m^6A but of the entire epitranscriptomic landscape of RNAs. Collectively, many m^6A-related studies published rely on different m^6A detection methods. This makes it difficult to compare m^6A annotations of different studies to each other. Finding the gold standard for the detection of m^6A is still an ongoing process but will benefit the field and make m^6A-based studies more comparable among each other.

4.2 A novel regulator of X-to-autosome dosage compensation

To compensate for sex chromosome dimorphism, one X chromosome is randomly inactivated in female eutherian mammals. This leads to one active X chromosome in both females and males, while autosomal chromosomes are present in two copies. To compensate for this, Ohno's hypothesis states that the X chromosome is upregulated which leads to X-chromosomal genes reaching similar expression levels as autosomal genes (Ohno, 1966). The advent of high-throughput sequencing techniques allowed to test this hypothesis. Controversially, several studies have been conducted and showed conflicting evidence. While some studies calculate a X-to-autosome ratio (X:A ratio) close to 0.5 and state that X-to-autosome dosage compensation (X:A DC) does not exist (Xiong et al., 2010; Lin et al., 2012), others have shown that X:A DC is in place and that the X:A ratio is close to 1 (Nguyen and Disteche, 2006; Borensztein et al., 2017; Deng et al., 2011; Lin et al., 2011).

4.2.1 Bioinformatical challenges: Does X-to-autosome dosage compensation exist?

The conflicting evidences of X:A ratios have left an open question whether X:A DC is in place or not. Using our expression data, we could show that X:A ratios are similar to 1 and that X-to-autosome dosage compensation is

in place in male mESCs, as well as different human cell lines. This is supported by several recent studies which have shown that calculating X:A expression ratios is strongly influenced by a subset of genes on the X chromosome: tissue-specific genes such as testis-specific genes that are enriched on the X chromosome and not expressed in somatic cells (Khil et al., 2004; Mueller et al., 2008). Thus, when calculating the X:A ratios in cell types in which those genes are not expressed, the X:A ratios are close to 0.5 due to many genes on the X chromosome with a low expression value. However, when calculating the X:A ratio only based on expressed genes, the ratio is close to 1, indicating that X-chromosomal transcripts have similar expression levels as autosomal ones and that X:A DC is in place. This shows that correct filtering of expression datasets is crucial for a correct X:A ratio calculation. Several studies have addressed this and shown that X-to-autosome dosage compensation is in place (Deng et al., 2011; Kharchenko et al., 2011; Yildirim et al., 2012; Sangrithi et al., 2017).
Additionally, we also observed that the X-to-autosome ratio is above 1 in female mESCs with two active X chromosomes. This indicates that also before X chromosome inactivation dosage compensation is in place. This is in contrast to other studies that have shown that X chromosome upregulation is triggered by X chromosome inactivation (see subsection 4.2.5). Collectively, we and others have shown that the X:A ratio is close to 1 and that Ohno's hypothesis holds true when applying sufficient parameters to filter out unexpressed genes.

4.2.2 X-to-autosome dosage compensation: a multiple-layer mechanisms

How is the upregulation of X-chromosomal transcripts accomplished? Several X:A DC mechanisms have been identified so far. Among them, increased transcription rates of X-chromosomal genes as well as higher translational efficiencies of X-chromosomal transcripts have been described (see subsection 1.1.4). Thus, X:A DC is not driven by the alteration of one mechanism but rather an orchestrated combination of different processes that are altered for X-chromosomal genes and transcripts. Our finding supports this hypothesis by complementing previously identified X:A DC mechanisms. Although higher X-chromosomal transcript stabilities have been described previously, the underlying mechanisms was unknown (Deng et al., 2013; Faucillion and Larsson, 2015). We could show that through depleted m^6A sites on

X-chromosomal transcripts, they are selectively stabilized and reach higher half-lives. Therefore, we present m^{6}A as a novel regulator of X-to-autosome dosage compensation. Of note, while the initial hypothesis by Ohno (Ohno, 1966) as well as several identified mechanisms support the upregulation of X-chromosomal genes, we identified a mechanism that promotes an autosomal downregulation. Thus, we can now update the current model that mainly focuses on X chromosome upregulation into an extended new model that includes both, X chromosome upregulation as well as an autosomal downregulation to balance X-to-autosome gene expression. This is in line with the hypothesis that a gradual degeneration of the Y chromosome and an evolving X chromosome inactivation must have led to a selective pressure at all levels of gene expression to acquire and ensure dosage compensation (Faucillion and Larsson, 2015). Of note, the effect size of m^{6}A-driven RNA stability does not account for the entire effect size of X chromosome upregulation, indicating that X:A DC is indeed achieved through several different layers altering the gene expression of X-chromosomal genes.

4.2.3 Mechanisms behind m^{6}A-mediated mRNA decay

A depletion of m^{6}A motifs (GGACH, H = A, C or T) on X-chromosomal transcripts accounts for higher transcript stabilities, the underlying mechanisms however are not clear yet. Since m^{6}A acts as a global degradation signal, we hypothesize that X-chromosomal transcripts are less prone to being degraded through the m^{6}A-dependent mRNA degradation pathway. To date, three m^{6}A-dependent mRNA degradation mechanisms have been described (see subsection 1.2.8; Gibbs and Chanfreau, 2022). If the depletion of m^{6}A on X-chromosomal transcripts inhibits all three degradation mechanisms equally remains to be explored.

One degradation pathway that has been linked to YTHDF2 m^{6}A-mRNA degradation is mediated by UPF1, a key player in NMD. NMD is triggered by premature termination of translation, which is induced by premature termination codons (PTCs) (reviewed in Kervestin and Jacobson, 2012). X-chromosomal transcripts have been shown to harbor lower numbers of PTC-induced NMD target sites (Yin et al., 2009). It was suggested that this might lead to less PTC-induced NMD of X-chromosomal transcripts. This is supported by the observation that upon UPF1 KD, the X:A ratio is perturbed, indicating that UPF1-mediated mRNA degradation affects X-chromosomal transcripts differently than autosomal transcripts and plays a role in X:A DC

(Yin et al., 2009). Although m^6A-mediated binding of YTHDF2 has been connected to UPF1-mediated mRNA degradation, it has been shown at the same time that this is an NMD-independent mechanism (Boo et al., 2022). Therefore, it remains unclear if the depletion of PTCs from X-chromosomal transcripts plays a role in the m^6A-mediated mRNA degradation via UPF1 or if it could be a separate mechanism that evolved independently to increase X-chromosomal transcript stabilities.
Two additional YTHDF2-related m^6A-mRNA degradation mechanisms have been identified. It is unclear if also the HRSP12-mediated, and CCR4/NOT-mediated mRNA degradation play a role in higher X-chromosomal stabilities. To investigate this, it would be interesting to quantify the specific HRSP12 binding motifs which have been shown to guide HRSP12 recruitment and m^6A-dependent endoribonucleolytic cleavage (Park et al., 2019). This could shed light on whether X-chromosomal genes are depleted of this motif and therefore less bound by HRSP12. Collectively, it will be an interesting future research field to identify the mechanisms underlying the selective-m^6A mediated degradation of X-chromosomal transcripts.

4.2.4 Motif depletion guiding less m^6A sites

We showed that lower numbers of m^6A on X-chromosomal transcripts are guided by an intrinsically ("hard-coded") mechanism of low numbers of a subset of DRACH motifs, the GGACH motifs on X-chromosomal genes. A subset of motifs within the DRACH motif family is more often methylated than others. The five most often methylated DRACH motifs are depleted from X-chromosomal genes, while the five motifs that are the least often methylated motifs are not depleted. This indicates on the one hand that a specific evolutionary pressure led to a depletion of strongly methylated DRACH motifs on X-chromosomal transcripts. On the other hand, this could also suggest that the most often methylated DRACH motifs are especially important for transcript stability and other motifs might carry out different functions that do not play a role in X:A DC and are therefore not depleted on X-chromosomal genes. To ivestigate this further, one could mutate different methylated DRACH motifs on a transcript and evaluate which DRACH motif mutation have an effect on transcript stability.
It was previously shown that m^6A sites within the coding sequence (CDS) as well as the 3′ untranslated region (3′UTR) have an effect on mRNA stability, while m^6A located in the 5′ untranslated region (5′UTR) of transcripts has

less effect on transcript stability (Herzog et al., 2017). Consistently, we could show that GGACH motifs are depleted CDS and 3'UTR, while it is not depleted in the 5'UTR. This further indicates that m^6A sites located in CDS and 3'UTR indeed influence transcript stability.
We confirmed that GGACH motif depletion of X-chromosomal genes arose in a sex-chromosome-specific manner. Dosage compensation evolved independently in different species: for example, in *Drosophila*, females do not undergo X chromosome inactivation, but an X chromosome upregulation in males (reviewed in Gelbart and Kuroda, 2009). Investigating whether m^6A is also contributing to X:A DC in *Drosophila* via the upregulation of X-chromosomal genes could shed light on the conservation of the function of m^6A in dosage compensation.
In addition to the depletion of GGACH motifs guiding less m^6A on X-chromosomal transcripts, it is possible that other mechanisms also ensure lower methylation levels of the X chromosome. For example, METTL3 could be less recruited to the X chromosome than other chromosomes. This could be mediated by a depleted number of H3K36me3 methylations which have been shown to recruit METTL14 to methylate mRNAs (Huang et al., 2019). A quantification of the methylation levels on the X chromosome versus autosomes could shed light on such a mechanism. Furthermore, it has been recently shown that members of the exon junction complex block METTL3-mediated m^6A methylation (Yang et al., 2022). To investigate this further, one could analyze the binding of the exon junction complex in existing datasets (e.g. Saulière et al., 2012) and whether it is different on X-chromosomal transcripts and could thereby play a role in lower numbers of m^6A sites on those transcripts.

4.2.5 Is dosage compensation always active?

In early developmental states, both X chromosomes are active in female cells. Logically, X:A DC would not be necessary if two X chromosomes are active. This raises the question if X-to-autosome dosage compensation is always in place or specifically activated when needed, after X chromosome inactivation. We showed that m^6A-mediated dosage compensation is hard coded via a depletion of GGACH motifs on X-chromosomal genes. Thus, this mechanism is most likely active at all times. This was supported by our X:A expression calculations in female mESC that have two active X chromosomes.

Here, we calculated an X:A ratio of over 1, indicating that two active X chromosomes lead to higher gene expression levels than two active autosomal chromosomes. As a consequence, a dosage excess in females before X chromosome inactivation occurs, which was previously hypothesized to be tolerated (reviewed in Snell and Turner, 2018). This is in line with previous studies that have shown that X:A DC is active before X chromosome inactivation (Lin et al., 2007; Borensztein et al., 2017), but in contrast with others that have shown that X chromosome upregulation is "switched on" when needed, and thus the expression of the active X chromosome increases after X chromosome inactivation (Lentini et al., 2022). This might be due to the fact that X:A DC is a multi-layer process and some of the mechanisms are hard-coded and always active while other mechanisms can be specifically switched on followed by individual triggers.

4.2.6 Dosage compensation more important for some genes?

Some genes escape X chromosome inactivation and stay expressed from both X chromosomes. These "escapees" would therefore not need an X:A DC mechanism. In our data, escapees showed a significant depletion of GGACH motifs, similar to other genes on the X chromosome. This indicates that the m^6A mediated dosage compensation also applies to those genes. Why are escaper genes subjected to X:A DC? Escaper genes have been shown to escape X chromosome inactivation in a tissue-specific and cell-type specific manner (Berletch et al., 2015). Therefore, it is possible that X chromosome upregulation for escaper genes has to be individually modulated in different tissues. It is an ongoing question in the field whether all X-chromosomal genes are similarly affected by dosage compensation or if X:A DC is more important for certain subgroups of genes. For example, it has been suggested that X:A DC is of little importance for genes on the X chromosome that are tissue-specific or independently evolved genes (Deng et al., 2011; Mueller et al., 2013). In our analysis, we did not find a difference in GGACH motif content in the sequence of different gene groups of the X chromosome, such as conserved genes, recently evolved genes or testis-specific genes. This indicates that dosage compensation is equally important for differently evolved genes on the X chromosome. When screening X-chromosomal genes for particularly low numbers of GGACH motifs, we detected an enrichment of ribosomal genes as well as histone genes. Interestingly, for histone genes, we could also detect a depletion of GGACH motifs on autosomal genes showing

that not only X-chromosomal histone genes might lack m^{6}A modifications but also histone genes located on autosomes. Histone mRNAs do not carry a poly(A) tail like other mRNAs, instead, they carry a conserved stem loop at their transcript end (reviewed in Marzluff et al., 2008). This stem loop is recognized by a specific protein which recruits UPF1 to induce mRNA decay (reviewed in Schoenberg and Maquat, 2012). Since histone mRNAs have the specific stem loop mediated recruitment of UPF1, it is possible that they do not need an m^{6}A-mediated recruitment of degradation factors via YTHDF2. This could explain why histone mRNAs lack DRACH motifs.

4.2.7 m^{6}A - a regulator of both layers of dosage compensation?

Dosage compensation is a two layer process. To compensate genetical differences between females and males, one X chromosome is inactivated in females. Then, to compensate the imbalance of one active chromosome to the two copies of autosomes, the X chromosome is upregulated. In this work, we have identified m^{6}A as a contributor to X-to-autosome dosage compensation. m^{6}A has also been identified to contribute to X chromosome inactivation (Patil et al., 2016). The long non coding RNA *Xist* is regulating X chromosome inactivation and harbors several m^{6}A modifications (Linder et al., 2015; Nesterova et al., 2019). These modifications were suggested to recruit YTHDC1 to *Xist*, and that this is a required step in X chromosome inactivation (Patil et al., 2016). While the exact mechanism of m^{6}A and YTHDFC1 in X chromosome inactivation remains to be identified, it has been confirmed that m^{6}A is playing a role in X chromosome inactivation (Coker et al., 2020). This indicates that m^{6}A is also playing a role in both layers of dosage compensation. This could indicate that X chromosome inactivation and upregulation are a coordinated event that might be achieved through a cross-talk of m^{6}A.

4.3 Conclusion and Outlook

In this work, we presented an improved method to detect m^{6}A mRNA modifications in a transcriptome-wide and single-nucleotide resolution. Additionally, our work provides comprehensive annotations of m^{6}A sites for humans and mice across different cell-lines. Recently, a large number of novel techniques for the detection of m^{6}A have been developed, with each new technique overcoming limitations of previously presented methods. With novel

techniques at hand, it will only be a question of time to find a m^6A detection method that overcomes current limitations and will become the gold standard applied by the vast majority of researchers in the field. While m^6A is the best studied mRNA modification, continuous ongoing research will also focus on the detection of functions of other mRNA modifications or cross-talks of different modifications. This will shed light on the open questions of the deposition of m^6A within different DRACH motifs or which other factors might influence the deposition.

By presenting a novel function of m^6A in the global regulation of X-to-autosome dosage compensation, we have discovered a before unknown function of m^6A. This highlights the importance of mRNA modifications and their potency in gene-expression regulation. Future directions will detangle the exact mechanisms that drive m^6A-related mRNA degradation. One crucial question to answer will be whether UPF1 is the main player in m^6A-mediated X-to-autosome dosage compensation or if other degradation mechanisms are acting differently on X-chromosomal transcripts. In line with this question, it will be interesting to investigate if the depletion of m^6A from X-chromosomal transcripts indeed leads to less binding and thus less degradation mediated by YTHDF2. It would be crucial to validate this hypothesis by performing *YTHDF2* KD or KO experiments to investigate if X-chromosomal transcripts are differentially affected by the loss of the protein.

X-to-autosome dosage compensation has been explained by a set of different molecular mechanisms, of which some have been identified to work in a highly dynamic manner and others to be hard coded. Future work will investigate how different X:A DC mechanisms orchestrate X:A DC and resolve some unanswered questions about the dynamics of the mechanism itself as well as the cross-talk of X chromosome inactivation and X chromosome upregulation.

List of Abbreviations

3'UTR	3 prime untranslated region
5'UTR	5 prime untranslated region
ALKBH5	Alpha-ketoglutarate-dependent dioxygenase alkB homolog 5
APOBEC1	Apolipoprotein B mRNA editing enzyme catalytic subunit 1
C	Cytidine
CCR4-NOT	Carbon Catabolite Repression—Negative On TATA-less
cDNA	Complementary Deoxyribonucleic acid
CNOT1	CCR4-NOT transcription complex subunit 1
DART-seq	Deamination adjacent to RNA modification targets
DNA	Deoxyribonucleic acid
DPPW motif	Aspartic acid, proline, proline, tryptophan motif
DRACH	D = adenine, guanine or thymine; R = adenine or guanine; H = adenine, cytosine or thymine
dRNA seq	Direct RNA sequencing
e.g.	Example given
EIF4A3	Eukaryotic translation initiation factor 4A3
eTAM-seq	TadA-assisted N^6-methyladenosine sequencing
FTO	Fat mass and obesity-associated protein
GLORI	Glyoxal and nitrite-mediated deamination of unmethylated adenosines
H3K36me3	Histone three lysine 36 trimethylation
HuR	Hu-antigen R
iCLIP	Individual-nucleotide-resolution cross-linking and immunoprecipitation
IGF2BP1	Insulin like growth factor 2 mRNA binding protein 1
KD	Knock down
KO	Knock out
m^6A	N^6-methyladenosine
m^6A-seq	m^6A sequencing

METTL3	Methyltransferase like 3
METTL14	Methyltransferase like 14
meRIP-seq	Methylated RNA immunoprecipitation sequencing
miCLIP	m^6A individual-nucleotide-resolution cross-linking and immunoprecipitation
mRNA	Messenger ribonucleic acid
NGS	Next-generation sequencing
NMD	Nonsense-mediated decay
P bodies	Processing bodies
PCR	Polymerase chain reaction
PTC	premature translation termination codons
RNA	Ribonucleic acid
RNA seq	RNA sequencing
RT	Reverse transcriptase
SAM	S-adensosylmethionine
SLAM seq	Thiol(SH)-linked alkylation for the metabolic sequencing of RNA
Sry	Sex-determining region on the Y chromoome
U	Uridine
UPF1	Up-frameshift protein 1
UV	Ultraviolet
WTAP	Wilms' tumor 1-associating protein
X:A ratios	X-to-autosome ratios
XAR	X added region
XCR	X conserved region
Xic	X-linked locus
Xist	X-inactive specific transcript
YTHDC	YTH domain-containing protein
YTHDF	YT521-B homology domain-containing family

Bibliography

Andrew, H Sinclair, Berta Philippe, S Palmer Mark, et al. (1990). "A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif". In: *Nature* 346.6281, pp. 240–244.

Arnold, Arthur P (2009). "Mouse models for evaluating sex chromosome effects that cause sex differences in non-gonadal tissues". In: *Journal of neuroendocrinology* 21.4, pp. 377–386.

Augui, Sandrine, Elphège P Nora, and Edith Heard (2011). "Regulation of X-chromosome inactivation by the X-inactivation centre". In: *Nature Reviews Genetics* 12.6, pp. 429–442.

Avner, Philip and Edith Heard (2001). "X-chromosome inactivation: counting, choice and initiation". In: *Nature Reviews Genetics* 2.1, pp. 59–67.

Bachtrog, Doris (2013). "Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration". In: *Nature Reviews Genetics* 14.2, pp. 113–124.

Barbieri, Isaia and Tony Kouzarides (2020). "Role of RNA modifications in cancer". In: *Nature Reviews Cancer* 20.6, pp. 303–322.

Barr, Murray L and Ewart G Bertram (1949). "A morphological distinction between neurones of the male and female, and the behaviour of the nucleolar satellite during accelerated nucleoprotein synthesis". In: *Nature* 163.4148, pp. 676–677.

Basilicata, M Felicia and Claudia Isabelle Keller Valsecchi (2021). "The good, the bad, and the ugly: Evolutionary and pathological aspects of gene dosage alterations". In: *PLoS Genetics* 17.12, e1009906.

Bellott, Daniel W, Helen Skaletsky, Tatyana Pyntikova, Elaine R Mardis, Tina Graves, et al. (2010). "Convergent evolution of chicken Z and human X chromosomes by expansion and gene acquisition". In: *Nature* 466.7306, pp. 612–616.

Berletch, Joel B, Wenxiu Ma, Fan Yang, Jay Shendure, William S Noble, Christine M Disteche, and Xinxian Deng (2015). "Escape from X inactivation varies in mouse tissues". In: *PLoS genetics* 11.3, e1005079.

Boccaletto, Pietro and Błażej Bagiński (2021). "MODOMICS: an operational guide to the use of the RNA modification pathways database". In: *RNA Bioinformatics*. Springer, pp. 481–505.

Bokar, Joseph A, Mary Eileen Rath-Shambaugh, Rachael Ludwiczak, Prema Narayan, and Fritz Rottman (1994). "Characterization and partial purification of mRNA N6-adenosine methyltransferase from HeLa cell nuclei. Internal mRNA methylation requires a multisubunit complex". In: *Journal of Biological Chemistry* 269.26, pp. 17697–17704.

Boo, Sung Ho and Yoon Ki Kim (2020). "The emerging role of RNA modifications in the regulation of mRNA stability". In: *Experimental & molecular medicine* 52.3, pp. 400–408.

Boo, Sung Ho, Hongseok Ha, Yujin Lee, Min-Kyung Shin, Sena Lee, and Yoon Ki Kim (2022). "UPF1 promotes rapid degradation of m6A-containing RNAs". In: *Cell Reports* 39.8, p. 110861.

Borensztein, Maud, Laurène Syx, Katia Ancelin, Patricia Diabangouaya, Christel Picard, et al. (2017). "Xist-dependent imprinted X inactivation and the early developmental consequences of its failure". In: *Nature structural & molecular biology* 24.3, pp. 226–233.

Borsani, Giuseppe, Rossana Tonlorenzi, M Christine Simmler, Luisa Dandolo, Danielle Arnaud, et al. (1991). "Characterization of a murine gene expressed from the inactive X chromosome". In: *Nature* 351.6324, pp. 325–329.

Boulias, Konstantinos, Diana Toczydłowska-Socha, Ben R Hawley, Noa Liberman, Ken Takashima, et al. (2019). "Identification of the m6Am methyltransferase PCIF1 reveals the location and functions of m6Am in the transcriptome". In: *Molecular cell* 75.3, pp. 631–643.

Bringmann, Peter and Reinhard Lührmann (1987). "Antibodies specific for N6-methyladenosine react with intact snRNPs U2 and U4/U6". In: *FEBS letters* 213.2, pp. 309–315.

Brockdorff, Neil, Alan Ashworth, Graham F Kay, Penny Cooper, Sandy Smith, et al. (1991). "Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome". In: *Nature* 351.6324, pp. 329–331.

Brown, Carolyn J, Ronald G Lafreniere, Vicki E Powers, Gianfranco Sebastio, Andrea Ballabio, et al. (1991). "Localization of the X inactivation centre on the human X chromosome in Xq13". In: *Nature* 349.6304, pp. 82–84.

Buchbender, Andreas, Holger Mutter, FX Reymond Sutandy, Nadine Körtel, Heike Hänel, et al. (2020). "Improved library preparation with the new iCLIP2 protocol". In: *Methods* 178, pp. 33–48.

Burgoyne, Paul S (1982). "Genetic homology and crossing over in the X and Y chromosomes of mammals". In: *Human genetics* 61.2, pp. 85–90.

Busch, Anke, Mirko Brüggemann, Stefanie Ebersberger, and Kathi Zarnack (2020). "iCLIP data analysis: A complete pipeline from sequencing reads to RBP binding sites". In: *Methods* 178, pp. 49–62.

Carrel, Laura and Huntington F Willard (2005). "X-inactivation profile reveals extensive variability in X-linked gene expression in females". In: *Nature* 434.7031, pp. 400–404.

Cayir, Akin (2022). "RNA modifications as emerging therapeutic targets". In: *Wiley Interdisciplinary Reviews: RNA* 13.4, e1702.

Chakrabarti, Anob M, Nejc Haberman, Arne Praznik, Nicholas M Luscombe, and Jernej Ule (2018). "Data science issues in studying protein–RNA interactions with CLIP technologies". In: *Annual Review of Biomedical Data Science* 1, pp. 235–261.

Charlesworth, Brian (1991). "The evolution of sex chromosomes". In: *Science* 251.4997, pp. 1030–1033.

Chen, Hong-Xuan, Zhang Zhang, Dong-Zhao Ma, Li-Qian Chen, and Guan-Zheng Luo (2022). "Mapping single-nucleotide m6A by m6A-REF-seq". In: *Methods* 203, pp. 392–398.

Chen, Zhen, Pei Zhao, Fuyi Li, Yanan Wang, A Ian Smith, et al. (2020). "Comprehensive review and assessment of computational methods for predicting RNA post-transcriptional modification sites from RNA sequences". In: *Briefings in bioinformatics* 21.5, pp. 1676–1696.

Coker, Heather, Guifeng Wei, Benoit Moindrot, Shabaz Mohammed, Tatyana Nesterova, and Neil Brockdorff (2020). "The role of the Xist 5'm6A region and RBM15 in X chromosome inactivation". In: *Wellcome open research* 5.

Cozzuto, Luca, Huanle Liu, Leszek P Pryszcz, Toni Hermoso Pulido, Anna Delgado-Tejedor, Julia Ponomarenko, and Eva Maria Novoa (2020). "MasterOfPores: a workflow for the analysis of oxford nanopore direct RNA sequencing datasets". In: *Frontiers in genetics* 11, p. 211.

Crick, Francis H (1958). "On protein synthesis". In: *Symp Soc Exp Biol*. Vol. 12. 138-63, p. 8.

Deamer, David, Mark Akeson, and Daniel Branton (2016). "Three decades of nanopore sequencing". In: *Nature biotechnology* 34.5, pp. 518–524.

Deng, Xinxian, Joseph B Hiatt, Di Kim Nguyen, Sevinc Ercan, David Sturgill, et al. (2011). "Evidence for compensatory upregulation of expressed X-linked genes in mammals, Caenorhabditis elegans and Drosophila melanogaster". In: *Nature genetics* 43.12, pp. 1179–1185.

Deng, Xinxian, Joel B Berletch, Wenxiu Ma, Joseph B Hiatt, William S Noble, Jay Shendure, Christine M Disteche, et al. (2013). "Mammalian X upregulation is associated with enhanced transcription initiation, RNA half-life, and MOF-mediated H4K16 acetylation". In: *Developmental cell* 25.1, pp. 55–68.

Desrosiers, Ronald, Karen Friderici, and Fritz Rottman (1974). "Identification of methylated nucleosides in messenger RNA from Novikoff hepatoma cells". In: *Proceedings of the National Academy of Sciences* 71.10, pp. 3971–3975.

Dimock, Kenneth and C Martin Stoltzfus (1977). "Sequence specificity of internal methylation in B77 avian sarcoma virus RNA subunits". In: *Biochemistry* 16.3, pp. 471–478.

Dominissini, Dan, Sharon Moshitch-Moshkovitz, Schraga Schwartz, Mali Salmon-Divon, Lior Ungar, et al. (2012). "Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq". In: *Nature* 485.7397, pp. 201–206.

Du, Hao, Ya Zhao, Jinqiu He, Yao Zhang, Hairui Xi, et al. (2016). "YTHDF2 destabilizes m6A-containing RNA through direct recruitment of the CCR4–NOT deadenylase complex". In: *Nature communications* 7.1, pp. 1–11.

Du, Yuzhang, Guofang Hou, Hailong Zhang, Jinzhuo Dou, Jianfeng He, et al. (2018). "SUMOylation of the m6A-RNA methyltransferase METTL3 modulates its function". In: *Nucleic acids research* 46.10, pp. 5195–5208.

Faucillion, Marie-Line and Jan Larsson (2015). "Increased expression of X-linked genes in mammals is associated with a higher stability of transcripts and an increased ribosome density". In: *Genome biology and evolution* 7.4, pp. 1039–1052.

Felsenfeld, Gary and Mark Groudine (2003). "Controlling the double helix". In: *Nature* 421.6921, pp. 448–453.

Fu, Ye and Chuan He (2012). "Nucleic acid modifications with epigenetic significance". In: *Current opinion in chemical biology* 16.5-6, pp. 516–524.

Furlan, Mattia, Anna Delgado-Tejedor, Logan Mulroney, Mattia Pelizzola, Eva Maria Novoa, and Tommaso Leonardi (2021). "Computational methods for RNA modification detection from nanopore direct RNA sequencing data". In: *RNA biology* 18.sup1, pp. 31–40.

Garcia-Campos, Miguel Angel, Sarit Edelheit, Ursula Toth, Modi Safra, Ran Shachar, et al. (2019). "Deciphering the "m6A code" via antibody-independent quantitative profiling". In: *Cell* 178.3, pp. 731–747.

Garcias Morales, David and José L Reyes (2021). "A birds'-eye view of the activity and specificity of the mRNA m6A methyltransferase complex". In: *Wiley Interdisciplinary Reviews: RNA* 12.1, e1618.

Gelbart, Marnie E and Mitzi I Kuroda (2009). "Drosophila dosage compensation: a complex voyage to the X chromosome". In: *Development* 136 (9), pp. 1399–1410.

Gendrel, Anne-Valerie, Anwyn Apedaile, Heather Coker, Ausma Termanis, Ilona Zvetkova, et al. (2012). "Smchd1-dependent and-independent pathways determine developmental dynamics of CpG island methylation on the inactive X chromosome". In: *Developmental cell* 23.2, pp. 265–279.

Geula, Shay, Sharon Moshitch-Moshkovitz, Dan Dominissini, Abed AlFatah Mansour, Nitzan Kol, et al. (2015). "m6A mRNA methylation facilitates resolution of naïve pluripotency toward differentiation". In: *Science* 347.6225, pp. 1002–1006.

Gibbs, Michelle R and Guillaume F Chanfreau (2022). "UPF1 adds an m6A feather to its (de) cap". In: *Cell Reports* 39.8, p. 110898.

Giorgetti, Luca, Bryan R Lajoie, Ava C Carter, Mikael Attia, Ye Zhan, et al. (2016). "Structural organization of the inactive X chromosome in the mouse". In: *Nature* 535.7613, pp. 575–579.

Hawley, Ben R and Samie R Jaffrey (2019). "Transcriptome-Wide Mapping of m6A and m6Am at Single-Nucleotide Resolution Using miCLIP". In: *Current protocols in molecular biology* 126.1, e88.

He, Chuan (2010). "Grand challenge commentary: RNA epigenetics?" In: *Nature chemical biology* 6.12, pp. 863–865.

Heard, Edith and Christine M Disteche (2006). "Dosage compensation in mammals: fine-tuning the expression of the X chromosome". In: *Genes & development* 20.14, pp. 1848–1867.

Herzog, Veronika A, Brian Reichholf, Tobias Neumann, Philipp Rescheneder, Pooja Bhat, et al. (2017). "Thiol-linked alkylation of RNA to assess expression dynamics". In: *Nature methods* 14.12, pp. 1198–1204.

Hsieh, Chih-Lin (2000). "Dynamics of DNA methylation pattern". In: *Current opinion in genetics & development* 10.2, pp. 224–228.

Hsu, Phillip J, Yunfei Zhu, Honghui Ma, Yueshuai Guo, Xiaodan Shi, et al. (2017). "Ythdc2 is an N6-methyladenosine binding protein that regulates mammalian spermatogenesis". In: *Cell research* 27.9, pp. 1115–1127.

Huang, Huilin, Hengyou Weng, Wenju Sun, Xi Qin, Hailing Shi, et al. (2018). "Recognition of RNA N6-methyladenosine by IGF2BP proteins enhances mRNA stability and translation". In: *Nature cell biology* 20.3, pp. 285–295.

Huang, Huilin, Hengyou Weng, Keren Zhou, Tong Wu, Boxuan Simen Zhao, et al. (2019). "Histone H3 trimethylation at lysine 36 guides m6A RNA modification co-transcriptionally". In: *Nature* 567.7748, pp. 414–419.

Jacobs, Patricia A and John A Strong (1959). "A case of human intersexuality having a possible XXY sex-determining mechanism". In: *Nature* 183.4657, pp. 302–303.

Jégu, Teddy, Eric Aeby, and Jeannie T Lee (2017). "The X chromosome in space". In: *Nature Reviews Genetics* 18.6, pp. 377–389.

Jia, Guifang, YE Fu, XU Zhao, Qing Dai, Guanqun Zheng, et al. (2011). "N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO". In: *Nature chemical biology* 7.12, pp. 885–887.

Jia, Guifang, Ye Fu, and Chuan He (2013). "Reversible RNA adenosine methylation in biological regulation". In: *Trends in Genetics* 29.2, pp. 108–115.

Jiang, Xiulin, Baiyang Liu, Zhi Nie, Lincan Duan, Qiuxia Xiong, et al. (2021). "The role of m6A modification in the biological functions and diseases". In: *Signal transduction and targeted therapy* 6.1, pp. 1–16.

Jonkhout, Nicky, Julia Tran, Martin A Smith, Nicole Schonrock, John S Mattick, and Eva Maria Novoa (2017). "The RNA modification landscape in human disease". In: *Rna* 23.12, pp. 1754–1769.

Kan, Lijuan, Anya V Grozhik, Jeffrey Vedanayagam, Deepak P Patil, Nan Pang, et al. (2017). "The m6A pathway facilitates sex determination in Drosophila". In: *Nature communications* 8.1, pp. 1–16.

Keohane, Ann M, Laura P O'neill, Nikolai D Belyaev, Jayne S Lavender, and Bryan M Turner (1996). "X-Inactivation and histone H4 acetylation in embryonic stem cells". In: *Developmental biology* 180.2, pp. 618–630.

Kervestin, Stephanie and Allan Jacobson (2012). "NMD: a multifaceted response to premature translational termination". In: *Nature reviews Molecular cell biology* 13.11, pp. 700–712.

Kharchenko, Peter V, Ruibin Xi, and Peter J Park (2011). "Evidence for dosage compensation between the X chromosome and autosomes in mammals". In: *Nature genetics* 43.12, pp. 1167–1169.

Khil, Pavel P, Natalya A Smirnova, Peter J Romanienko, and R Daniel Camerini-Otero (2004). "The mouse X chromosome is enriched for sex-biased genes not subject to selection by meiotic sex chromosome inactivation". In: *Nature genetics* 36.6, pp. 642–646.

Koh, Casslynn WQ, Yeek Teck Goh, and WS Sho Goh (2019). "Atlas of quantitative single-base-resolution N 6-methyl-adenine methylomes". In: *Nature communications* 10.1, p. 5636.

König, Julian, Kathi Zarnack, Gregor Rot, Tomaž Curk, Melis Kayikci, et al. (2010). "iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution". In: *Nature structural & molecular biology* 17.7, pp. 909–915.

Lahn, Bruce T and David C Page (1999). "Four evolutionary strata on the human X chromosome". In: *Science* 286.5441, pp. 964–967.

Larsson, Anton JM, Christos Coucoravas, Rickard Sandberg, and Björn Reinius (2019). "X-chromosome upregulation is driven by increased burst frequency". In: *Nature Structural & Molecular Biology* 26.10, pp. 963–969.

Lasman, Lior, Vladislav Krupalnik, Sergey Viukov, Nofar Mor, Alejandro Aguilera-Castrejon, et al. (2020). "Context-dependent functional compensation between Ythdf m6A reader proteins". In: *Genes & development* 34.19-20, pp. 1373–1391.

Lentini, Antonio, Huaitao Cheng, JC Noble, Natali Papanicolaou, Christos Coucoravas, et al. (2022). "Elastic dosage compensation by X-chromosome upregulation". In: *Nature communications* 13.1, pp. 1–12.

Lesbirel, Simon and Stuart A Wilson (2019). "The m6A-methylase complex and mRNA export". In: *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms* 1862.3, pp. 319–328.

Li, Ang, Yu-Sheng Chen, Xiao-Li Ping, Xin Yang, Wen Xiao, et al. (2017). "Cytoplasmic m6A reader YTHDF3 promotes mRNA translation". In: *Cell research* 27.3, pp. 444–447.

Liao, Shanhui, Hongbin Sun, and Chao Xu (2018). "YTH domain: a family of N6-methyladenosine (m6A) readers". In: *Genomics, proteomics & bioinformatics* 16.2, pp. 99–107.

Lin, Fangqin, Ke Xing, Jianzhi Zhang, and Xionglei He (2012). "Expression reduction in mammalian X chromosome evolution refutes Ohno's hypothesis of dosage compensation". In: *Proceedings of the National Academy of Sciences* 109.29, pp. 11752–11757.

Lin, Hong, Vibhor Gupta, Matthew D VerMilyea, Francesco Falciani, Jeannie T Lee, Laura P O'Neill, and Bryan M Turner (2007). "Dosage compensation in the mouse balances up-regulation and silencing of X-linked genes". In: *PLoS biology* 5.12, e326.

Lin, Hong, John A Halsall, Philipp Antczak, Laura P O'Neill, Francesco Falciani, and Bryan M Turner (2011). "Relative overexpression of X-linked genes in mouse embryonic stem cells is consistent with Ohno's hypothesis". In: *Nature genetics* 43.12, pp. 1169–1170.

Linder, Bastian, Anya V Grozhik, Anthony O Olarerin-George, Cem Meydan, Christopher E Mason, and Samie R Jaffrey (2015). "Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome". In: *Nature methods* 12.8, pp. 767–772.

Liu, Cong, Hanxiao Sun, Yunpeng Yi, Weiguo Shen, Kai Li, et al. (2022). "Absolute quantification of single-base m6A methylation in the mammalian transcriptome using GLORI". In: *Nature Biotechnology*, pp. 1–12.

Liu, Jianzhao, Yanan Yue, Dali Han, Xiao Wang, Ye Fu, et al. (2014). "A METTL3–METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation". In: *Nature chemical biology* 10.2, pp. 93–95.

Livernois, AM, JAM Graves, and PD Waters (2012). "The origin and evolution of vertebrate sex chromosomes and dosage compensation". In: *Heredity* 108.1, pp. 50–58.

Loda, Agnese, Samuel Collombet, and Edith Heard (2022). "Gene regulation in time and space during X-chromosome inactivation". In: *Nature Reviews Molecular Cell Biology* 23.4, pp. 231–249.

Luo, Zhiyuan, Jiacheng Zhang, Jingyi Fei, and Shengdong Ke (2022). "Deep learning modeling m6A deposition reveals the importance of downstream cis-element sequences". In: *Nature communications* 13.1, pp. 1–16.

Lyon, Mary F (1961). "Gene action in the X-chromosome of the mouse (Mus musculus L.)" In: *nature* 190.4773, pp. 372–373.

Lyon, Mary F (1962). "Sex chromatin and gene action in the mammalian X-chromosome". In: *American journal of human genetics* 14.2, p. 135.

Marahrens, York, Barbara Panning, Jessica Dausman, William Strauss, and Rudolf Jaenisch (1997). "Xist-deficient mice are defective in dosage

compensation but not spermatogenesis." In: *Genes & development* 11.2, pp. 156–166.

Markaki, Yolanda, Johnny Gan Chong, Yuying Wang, Elsie C Jacobson, Christy Luong, et al. (2021). "Xist nucleates local protein gradients to propagate silencing across the X chromosome". In: *Cell* 184.25, pp. 6174–6192.

Marzluff, William F, Eric J Wagner, and Robert J Duronio (2008). "Metabolism and regulation of canonical histone mRNAs: life without a poly (A) tail". In: *Nature Reviews Genetics* 9.11, pp. 843–854.

Mateos, P Acera, AJ Sethi, M Guarnacci, A Ravindran, A Srivastava, et al. (2022). "Identification of m6A and m5C RNA modifications at single-molecule resolution from Nanopore sequencing". In: *bioRxiv (Mar. 2022)* 14.

Mauer, Jan and Samie R Jaffrey (2018). "FTO, m6Am, and the hypothesis of reversible epitranscriptomic mRNA modifications". In: *FEBS letters* 592.12, pp. 2012–2022.

Mauer, Jan, Xiaobing Luo, Alexandre Blanjoie, Xinfu Jiao, Anya V Grozhik, et al. (2017). "Reversible methylation of m6Am in the 5 cap controls mRNA stability". In: *Nature* 541.7637, pp. 371–375.

McIntyre, Alexa BR, Nandan S Gokhale, Leandro Cerchietti, Samie R Jaffrey, Stacy M Horner, and Christopher E Mason (2020). "Limits in the detection of m6A changes using MeRIP/m6A-seq". In: *Scientific reports* 10.1, pp. 1–15.

Meyer, Kate D (2019a). "DART-seq: an antibody-free method for global m6A detection". In: *Nature methods* 16.12, pp. 1275–1280.

Meyer, Kate D (2019b). "m6A-mediated translation regulation". In: *Biochimica et Biophysica Acta (Bba)-Gene Regulatory Mechanisms* 1862.3, pp. 301–309.

Meyer, Kate D and Samie R Jaffrey (2014). "The dynamic epitranscriptome: N6-methyladenosine and gene expression control". In: *Nature reviews Molecular cell biology* 15.5, pp. 313–326.

Meyer, Kate D, Yogesh Saletore, Paul Zumbo, Olivier Elemento, Christopher E Mason, and Samie R Jaffrey (2012). "Comprehensive analysis of mRNA methylation reveals enrichment in 3 UTRs and near stop codons". In: *Cell* 149.7, pp. 1635–1646.

Moshitch-Moshkovitz, Sharon, Dan Dominissini, and Gideon Rechavi (2022). "The epitranscriptome toolbox". In: *Cell* 185.5, pp. 764–776.

Mueller, Jacob L, Shantha K Mahadevaiah, Peter J Park, Peter E Warburton, David C Page, and James Turner (2008). "The mouse X chromosome is

enriched for multicopy testis genes showing postmeiotic expression". In: *Nature genetics* 40.6, pp. 794–799.

Mueller, Jacob L, Helen Skaletsky, Laura G Brown, Sara Zaghlul, Susan Rock, et al. (2013). "Independent specialization of the human and mouse X chromosomes for the male germ line". In: *Nature genetics* 45.9, pp. 1083–1087.

Muller, Hermann J (1914). "A gene for the fourth chromosome of Drosophila". In: *Journal of Experimental Zoology* 17.3, pp. 325–336.

Murakami, Shino and Samie R Jaffrey (2022). "Hidden codes in mRNA: Control of gene expression by m6A". In: *Molecular Cell* 82.12, pp. 2236–2251.

Nesterova, Tatyana B, Guifeng Wei, Heather Coker, Greta Pintacuda, Joseph S Bowness, et al. (2019). "Systematic allelic analysis defines the interplay of key pathways in X chromosome inactivation". In: *Nature communications* 10.1, p. 3129.

Nguyen, Di Kim and Christine M Disteche (2006). "Dosage compensation of the active X chromosome in mammals". In: *Nature genetics* 38.1, pp. 47–53.

Norris, Dominic P, Neil Brockdorff, and Sohaila Rastan (1991). "Methylation status of CpG-rich islands on active and inactive mouse X chromosomes". In: *Mammalian Genome* 1.2, pp. 78–83.

Nurk, Sergey, Sergey Koren, Arang Rhie, Mikko Rautiainen, Andrey V Bzikadze, et al. (2022). "The complete sequence of a human genome". In: *Science* 376.6588, pp. 44–53.

Ohno, Susumu (1966). "Sex chromosome and sex-linked genes". In: *Springer*.

Ohno, Susumu and TS Hauschka (1960). "Allocycly of the X-chromosome in tumors and normal tissues". In: *Cancer Research* 20.4, pp. 541–545.

Park, Ok Hyun, Hongseok Ha, Yujin Lee, Sung Ho Boo, Do Hoon Kwon, Hyun Kyu Song, and Yoon Ki Kim (2019). "Endoribonucleolytic cleavage of m6A-containing RNAs by RNase P/MRP complex". In: *Molecular cell* 74.3, pp. 494–507.

Patil, Deepak P, Chun-Kan Chen, Brian F Pickering, Amy Chow, Constanza Jackson, Mitchell Guttman, and Samie R Jaffrey (2016). "m6A RNA methylation promotes XIST-mediated transcriptional repression". In: *Nature* 537.7620, pp. 369–373.

Patil, Deepak P, Brian F Pickering, and Samie R Jaffrey (2018). "Reading m6A in the transcriptome: m6A-binding proteins". In: *Trends in cell biology* 28.2, pp. 113–127.

Perry, Robert P and Dawn E Kelley (1976). "Kinetics of formation of 5 terminal caps in mRNA". In: *Cell* 8.3, pp. 433–442.

Perry, Robert P, Dawn E Kelley, Karen Friderici, and Fritz Rottman (1975). "The methylated constituents of L cell messenger RNA: evidence for an unusual cluster at the 5 terminus". In: *Cell* 4.4, pp. 387–394.

Ping, Xiao-Li, Bao-Fa Sun, Lu Wang, Wen Xiao, Xin Yang, et al. (2014). "Mammalian WTAP is a regulatory subunit of the RNA N6-methyladenosine methyltransferase". In: *Cell research* 24.2, pp. 177–189.

Poh, Hui Xian, Aashiq H Mirza, Brian F Pickering, and Samie R Jaffrey (2022). "Alternative splicing of METTL3 explains apparently METTL3-independent m6A modifications in mRNA". In: *PLoS biology* 20.7, e3001683.

Pratanwanich, Ploy N, Fei Yao, Ying Chen, Casslynn WQ Koh, Yuk Kei Wan, et al. (2021). "Identification of differential RNA modifications from nanopore direct RNA sequencing with xPore". In: *Nature biotechnology* 39.11, pp. 1394–1402.

Rastan, Sohaila (1983). "Non-random X-chromosome inactivation in mouse X-autosome translocation embryos—location of the inactivation centre". In: *J Embryol Exp Morphol.* 78, pp. 1–22.

Rastan, Sohaila and Elizabeth J Robertson (1985). "X-chromosome deletions in embryo-derived (EK) cell lines associated with lack of X-chromosome inactivation". In: *J Embryol Exp Morphol.* 90, pp. 379–388.

Ries, Ryan J, Sara Zaccara, Pierre Klein, Anthony Olarerin-George, Sim Namkoong, et al. (2019). "m6A enhances the phase separation potential of mRNA". In: *Nature* 571.7765, pp. 424–428.

Ross, Mark T, Darren V Grafham, Alison J Coffey, Steven Scherer, Kirsten McLay, et al. (2005). "The DNA sequence of the human X chromosome". In: *Nature* 434.7031, pp. 325–337.

Roundtree, Ian A, Guan-Zheng Luo, Zijie Zhang, Xiao Wang, Tao Zhou, et al. (2017). "YTHDC1 mediates nuclear export of N6-methyladenosine methylated mRNAs". In: *elife* 6, e31311.

Russell, Liane Brauch (1961). "Genetics of Mammalian Sex Chromosomes: Mouse studies throw light on the functions and on the occasionally aberrant behavior of sex chromosomes." In: *Science* 133.3467, pp. 1795–1803.

Saletore, Yogesh, Kate Meyer, Jonas Korlach, Igor D Vilfan, Samie Jaffrey, and Christopher E Mason (2012). "The birth of the Epitranscriptome: deciphering the function of RNA modifications". In: *Genome biology* 13.10, pp. 1–12.

Sangrithi, Mahesh N and James MA Turner (2018). “Mammalian X chromosome dosage compensation: perspectives from the germ line”. In: *Bioessays* 40.6, p. 1800024.

Sangrithi, Mahesh N, Helene Royo, Shantha K Mahadevaiah, Obah Ojarikre, Leena Bhaw, et al. (2017). “Non-canonical and sexually dimorphic X dosage compensation states in the mouse and human germline”. In: *Developmental cell* 40.3, pp. 289–301.

Saulière, Jérôme, Valentine Murigneux, Zhen Wang, Emélie Marquenet, Isabelle Barbosa, et al. (2012). “CLIP-seq of eIF4AIII reveals transcriptome-wide mapping of the human exon junction complex”. In: *Nature structural & molecular biology* 19.11, pp. 1124–1131.

Schafer, Alan J and Peter N Goodfellow (1996). “Sex determination in humans”. In: *Bioessays* 18.12, pp. 955–963.

Schoenberg, Daniel R and Lynne E Maquat (2012). “Regulation of cytoplasmic mRNA decay”. In: *Nature Reviews Genetics* 13.4, pp. 246–259.

Schöller, Eva, Franziska Weichmann, Thomas Treiber, Sam Ringle, Nora Treiber, et al. (2018). “Interactions, localization, and phosphorylation of the m6A generating METTL3–METTL14–WTAP complex”. In: *Rna* 24.4, pp. 499–512.

Schwartz, Schraga, Maxwell R Mumbach, Marko Jovanovic, Tim Wang, Karolina Maciag, et al. (2014). “Perturbation of m6A writers reveals two distinct classes of mRNA methylation at internal and 5 sites”. In: *Cell reports* 8.1, pp. 284–296.

Schweingruber, Christoph, Paolo Soffientini, Marc-David Ruepp, Angela Bachi, and Oliver Mühlemann (2016). “Identification of interactions in the NMD complex using proximity-dependent biotinylation (BioID)”. In: *PLoS One* 11.3, e0150239.

Sheth, Ujwal and Roy Parker (2003). “Decapping and decay of messenger RNA occur in cytoplasmic processing bodies”. In: *Science* 300.5620, pp. 805–808.

Shi, Hailing, Xiao Wang, Zhike Lu, Boxuan S Zhao, Honghui Ma, et al. (2017). “YTHDF3 facilitates translation and decay of N6-methyladenosine-modified RNA”. In: *Cell research* 27.3, pp. 315–328.

Shu, Xiao, Jie Cao, Mohan Cheng, Siying Xiang, Minsong Gao, et al. (2020). “A metabolic labeling method detects m6A transcriptome-wide at single base resolution”. In: *Nature Chemical Biology* 16.8, pp. 887–895.

Skaletsky, Helen, Tomoko Kuroda-Kawaguchi, Patrick J Minx, Holland S Cordum, LaDeana Hillier, et al. (2003). “The male-specific region of the

human Y chromosome is a mosaic of discrete sequence classes". In: *Nature* 423.6942, pp. 825–837.

Śledź, Paweł and Martin Jinek (2016). "Structural insights into the molecular mechanism of the m6A writer complex". In: *elife* 5, e18434.

Snell, Daniel M and James MA Turner (2018). "Sex chromosome effects on male–female differences in mammals". In: *Current Biology* 28.22, R1313–R1324.

Splinter, Erik, Elzo de Wit, Elphège P Nora, Petra Klous, Harmen JG van de Werken, et al. (2011). "The inactive X chromosome adopts a unique three-dimensional conformation that is dependent on Xist RNA". In: *Genes & development* 25.13, pp. 1371–1383.

Stoilov, Peter, Ilona Rafalska, and Stefan Stamm (2002). "YTH: a new domain in nuclear proteins". In: *Trends in biochemical sciences* 27.10, pp. 495–497.

Sun, Hui-Lung, Allen C Zhu, Yawei Gao, Hideki Terajima, Qili Fei, et al. (2020). "Stabilization of ERK-phosphorylated METTL3 by USP5 increases m6A methylation". In: *Molecular cell* 80.4, pp. 633–647.

Takagi, Nobuo and Kenji Abe (1990). "Detrimental effects of two active X chromosomes on early mouse development". In: *Development* 109.1, pp. 189–201.

Tegowski, Matthew, Mathieu N Flamand, and Kate D Meyer (2022). "scDART-seq reveals distinct m6A signatures and mRNA methylation heterogeneity in single cells". In: *Molecular Cell* 82.4, pp. 868–878.

Tuck, Martin T (1992). "The formation of internal 6-methyladenine residues in eucaryotic messenger RNA". In: *International journal of biochemistry* 24.3, pp. 379–386.

Tukiainen, Taru, Alexandra-Chloé Villani, Angela Yen, Manuel A Rivas, Jamie L Marshall, et al. (2017). "Landscape of X chromosome inactivation across human tissues". In: *Nature* 550.7675, pp. 244–248.

Uzonyi, Anna, David Dierks, Ronit Nir, Oh Sung Kwon, Ursula Toth, et al. (2023). "Exclusion of m6A from splice-site proximal regions by the exon junction complex dictates m6A topologies and mRNA stability". In: *Molecular Cell*.

Visvanathan, A, V Patil, A Arora, AS Hegde, A Arivazhagan, V Santosh, and K Somasundaram (2018). "Essential role of METTL3-mediated m6A modification in glioma stem-like cells maintenance and radioresistance". In: *Oncogene* 37.4, pp. 522–533.

Wang, Ping, Katelyn A Doxtader, and Yunsun Nam (2016a). "Structural basis for cooperative function of Mettl3 and Mettl14 methyltransferases". In: *Molecular cell* 63.2, pp. 306–317.

Wang, Xiang, Jing Feng, Yuan Xue, Zeyuan Guan, Delin Zhang, et al. (2016b). "Structural basis of N6-adenosine methylation by the METTL3–METTL14 complex". In: *Nature* 534.7608, pp. 575–578.

Wang, Xiao, Zhike Lu, Adrian Gomez, Gary C Hon, Yanan Yue, et al. (2014). "N6-methyladenosine-dependent regulation of messenger RNA stability". In: *Nature* 505.7481, pp. 117–120.

Wang, Xiao, Boxuan Simen Zhao, Ian A Roundtree, Zhike Lu, Dali Han, et al. (2015). "N6-methyladenosine modulates messenger RNA translation efficiency". In: *Cell* 161.6, pp. 1388–1399.

Wang, Ye, Yu Xiao, Shunqing Dong, Qiong Yu, and Guifang Jia (2020a). "Antibody-free enzyme-assisted chemical approach for detection of N6-methyladenosine". In: *Nature chemical biology* 16.8, pp. 896–903.

Wang, Zhong-Yi, Evgeny Leushkin, Angélica Liechti, Svetlana Ovchinnikova, Katharina Mößinger, et al. (2020b). "Transcriptome and translatome co-evolution in mammals". In: *Nature* 588.7839, pp. 642–647.

Wei, Cha-Mer, Alan Gershowitz, and Bernard Moss (1975). "Methylated nucleotides block 5 terminus of HeLa cell messenger RNA". In: *Cell* 4.4, pp. 379–386.

Wei, Cha Mer, Alan Gershowitz, and Bernard Moss (1976). "5′-Terminal and internal methylated nucleotide sequences in HeLa cell mRNA". In: *Biochemistry* 15.2, pp. 397–401.

Wutz, Anton and Rudolf Jaenisch (2000). "A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation". In: *Molecular cell* 5.4, pp. 695–705.

Xiao, Wen, Samir Adhikari, Ujwal Dahal, Yu-Sheng Chen, Ya-Juan Hao, et al. (2016). "Nuclear m6A reader YTHDC1 regulates mRNA splicing". In: *Molecular cell* 61.4, pp. 507–519.

Xiao, Yu-Lan, Shun Liu, Ruiqi Ge, Yuan Wu, Chuan He, Mengjie Chen, and Weixin Tang (2023). "Transcriptome-wide profiling and quantification of N 6-methyladenosine by enzyme-assisted adenosine deamination". In: *Nature Biotechnology*, pp. 1–11.

Xiong, Yuanyan, Xiaoshu Chen, Zhidong Chen, Xunzhang Wang, Suhua Shi, et al. (2010). "RNA sequencing shows no dosage compensation of the active X-chromosome". In: *Nature genetics* 42.12, pp. 1043–1047.

Xu, Chao, Xiao Wang, Ke Liu, Ian A Roundtree, Wolfram Tempel, et al. (2014). "Structural basis for selective binding of m6A RNA by the YTHDC1 YTH domain". In: *Nature chemical biology* 10.11, pp. 927–929.

Yang, Fan, Tomas Babak, Jay Shendure, and Christine M Disteche (2010). "Global survey of escape from X inactivation by RNA-sequencing in mouse". In: *Genome research* 20.5, pp. 614–622.

Yang, Xin, Robinson Triboulet, Qi Liu, Erdem Sendinc, and Richard I Gregory (2022). "Exon junction complex shapes the m6A epitranscriptome". In: *Nature Communications* 13.1, pp. 1–12.

Yankova, Eliza, Wesley Blackaby, Mark Albertella, Justyna Rak, Etienne De Braekeleer, et al. (2021). "Small-molecule inhibition of METTL3 as a strategy against myeloid leukaemia". In: *Nature* 593.7860, pp. 597–601.

Yildirim, Eda, Ruslan I Sadreyev, Stefan F Pinter, and Jeannie T Lee (2012). "X-chromosome hyperactivation in mammals via nonlinear relationships between chromatin states and transcription". In: *Nature structural & molecular biology* 19.1, pp. 56–61.

Yin, Shanye, Wenjun Deng, Hancheng Zheng, Zhengguo Zhang, Landian Hu, and Xiangyin Kong (2009). "Evidence that the nonsense-mediated mRNA decay pathway participates in X chromosome dosage compensation in mammals". In: *Biochemical and biophysical research communications* 383.3, pp. 378–382.

Zaccara, Sara and Samie R Jaffrey (2020). "A unified model for the function of YTHDF proteins in regulating m6A-modified mRNA". In: *Cell* 181.7, pp. 1582–1595.

Zheng, Guanqun, John Arne Dahl, Yamei Niu, Peter Fedorcsak, Chun-Min Huang, et al. (2013). "ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility". In: *Molecular cell* 49.1, pp. 18–29.

Zhu, Tingting, Ian A Roundtree, Ping Wang, Xiao Wang, Li Wang, et al. (2014). "Crystal structure of the YTH domain of YTHDF2 reveals mechanism for recognition of N6-methyladenosine". In: *Cell research* 24.12, pp. 1493–1496.

Żylicz, Jan Jakub, Aurélie Bousard, Kristina Žumer, Francois Dossin, Eusra Mohammad, et al. (2019). "The implication of early chromatin changes in X chromosome inactivation". In: *Cell* 176.1-2, pp. 182–197.