THE UNIVERSITY OF CHICAGO

DEVELOPMENT OF SITE-SPECIFIC AND QUANTITATIVE N^6-METHYL ADENOSINE (M^6A) PROFILING METHODS

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF CHEMISTRY

BY

RUIQI GE

CHICAGO, ILLINOIS

AUGUST 2023

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank my advisor, Professor Chuan He, for his instructions and help throughout my Ph.D. years. Prof. He is a pioneering figure in RNA epitranscriptomics. His broad knowledge, insightfulness, diligence, and leadership all deeply impressed me. He always had new ideas and always kept track of projects, despite there were several dozen running in his lab simultaneously. Members in He Lab showed enormous passion in science and brilliance in pursuit of projects, which I deeply admired. Here I was able to learn cutting-edge technology, do world-leading research, and wholeheartedly believe that my endeavors were meaningful. Prof. He was resourceful and supportive. Thanks to his connections, I was able to participate in collaborations across a myriad of fields, which broadened my horizon and improved my communication skills. Even after costly and fruitless attempts in research, Prof. He was still supportive and encouraging. He is also exceptionally understanding to my work schedule. During the last year of Ph.D., I encountered many difficulties in finding post-doctoral positions and preparing for the dissertation. Only with Prof. He's suggestions, recommendations and patience could I overcame these obstacles.

I would also like to thank my committee members, Prof. Weixin Tang and Prof. Mengjie Chen. During our collaboration in projects, both Professors impressed me with their expertise and exceptional diligence. Both professors also helped greatly in my post-doctoral recommendations. I would like to send special thanks to Prof. Tang, with whom I discussed a lot about science and academic career over the past several years. As a young scientist, she also served as a role model for my scientific pursuits. I really appreciate her generosity for the sharing of so many precious suggestions.

I would like to thank He Lab members. Firstly, I would like to thank Dr. Lulu Hu, who is currently a professor in Fudan University in China. We collaborated on the SAC-seq project and various other ideas. I learnt the basics of next-generation sequencing and recombinant protein expression from her. We discussed numerous papers, ideas, experiment designs and career plans throughout my Ph.D. years. All kinds of other things were also discussed, which became my most precious memories of Ph.D. life. She was diligent and determined, dedicated to science. I sincerely wish her success in her career. Next, I would like to thank Dr. Chang Ye, with whom I collaborated on later projects in the past 2-3 years. His profound understanding of next-generation sequencing is fascinating. I would say one can become an expert in this field just by discussing with him. He also showed broad knowledge in various fields of science, oftentimes surprisingly indeed. He was also kind and patient, being very pleasant to work with. I really appreciate our collaborations. I would like to thank Dr. Peng Yong, Dr. Shun Liu, and Dr. Ruitu Lyu, for collaborating with me on data analysis. Their patience and efforts were indispensable for those projects. I would like to thank Dr. Lisheng Zhang and Dr. Qing Dai for various discussions and suggestions. They were both experts in next-generation sequencing, especially in the field of DNA/RNA modifications. Their unreserved sharing of knowledge and experiences was the key to my laying foundations in this field. Dr. Lisheng Zhang was also an extremely prolific graduate student, a role model, a kind, energetic and enthusiastic person. He gave me a lot of encouragement during all these years. I would like to thank Dr. Yu Xiao, for day-to-day discussion of research and non-research topics. Thank her for all those encouragements and kindness. Also thank Dr. Yulan Xiao, Dr. Pingluan Wang, Dr. Tong Wu, Dr. Yawei Gao, Dr. Qiancheng You, Dr. Ji Nie, Yutao Zhao, Yuhao Zhong, Xinran Feng and everyone else for collaborations and discussions.

Thank the University of Chicago. Thank the facility staffs, especially Dr. Pieter Farber, Dr. Lindsay Scarpitta, Dr. Saha Poulami, and Dr. Michael Nuenninghoff. Their efforts were indispensable. Thank my family and girlfriend for continuous support. Thank the city of Chicago for being a really nice place.

ABSTRACT

N^6-methyladenosine (m^6A) is the most abundant mRNA modification in mammalian cells, consisting of 0.1%-0.4% of total adenosine residues. m^6A regulates mRNA stability and translation, pre-mRNA splicing, miRNA biogenesis, lncRNA binding, and many other physiological and pathological processes. While the majority of m^6As occur at a consensus motif of DRm^6ACH (D = A/G/U, R = A/G, H = U/A/C), the presence of such motif does not guarantee methylation. Different copies of the same transcript are not uniformly methylated either. Within a single transcript m^6As are not evenly distributed, showing an enrichment in long internal exons and terminal exons. These characteristics of m^6A deposition calls for sequencing methods that not only pinpoint m^6A sites at base resolution, but also quantitates the abundance of methylation across different RNA copies. We developed m^6A-SAC-seq (m^6A-selective allyl chemical labeling and sequencing) and eTAM-seq (evolved TadA-assisted N^6-methyladenosine sequencing), both being site-specific and quantitative m^6A profiling methods based on next-generation sequencing. Optimizations were made to the shared library construction strategy, which significantly improved the result and reduced the amount of sample requirement. Both methods were able to identify more than 10-thousands of m^6A sites across the transcriptome of HeLa cells, being applicable to various biological systems including cell lines and tissue samples. Efforts had been made to extend both methods to high-throughput single-cell m^6A profiling. Finally, a modified version of eTAM-seq was also applied to the study of transcriptome-wide in-vivo accessibility of RNA.

LIST OF PUBLICATIONS PRESENTED IN THIS THESIS

#1. Hu L*, Liu S*, Peng Y*, **Ge R***, Su R*, Senevirathne C, Harada BT, Dai Q, Wei J, Zhang L, Hao Z, Luo L, Wang H, Wang Y, Luo M, Chen M, Chen J, He C. m^6A RNA modifications are measured at single-base resolution across the mammalian transcriptome. Nat Biotechnol. 2022 Aug;40(8):1210-1219.

#2. **Ge R***, Ye C*, Peng Y, Dai Q, Zhao Y, Liu S, Wang P, Hu L, He C. m^6A-SAC-seq for quantitative whole transcriptome m^6A profiling. Nat Protoc. 2023 Feb;18(2):626-657.

#3. Xiao YL*, Liu S*, **Ge R***, Wu Y, He C, Chen M, Tang W. Transcriptome-wide profiling and quantification of N^6-methyladenosine by enzyme-assisted adenosine deamination. Nat Biotechnol. 2023 Jan 2.

(Asterisks denote co-first authors)

Chapter 1 Introduction

1.1 RNA modifications and m^6A

1.1.1 Brief introduction to RNA modifications

Nucleic acids—DNA and RNA are fundamental to gene expression, which process is known as the central dogma of molecular biology. Apart from the information they encode, nucleic acids are essentially biopolymers assembled from four types of nucleotides: deoxyadenosine (dA), deoxycytidine (dC), deoxyguanosine (dG), deoxythymidine (dT) for DNA, and adenosine (A), cytidine (C), guanosine (G), uridine (U) for RNA. Nucleotides contain certain moieties that could be reacted with, either chemically or biochemically, to yield modified nucleotides. In 1948, 5-methyl-2'-deoxycytidine (5mdC) was discovered in DNA via paper chromatography to be the first naturally occurring modified nucleotide[1]. The first naturally occurring ribonucleotide modification, pseudouridine (Ψ), was discovered two years later[2]. Ever since, more than 170 types of RNA modifications have been identified.[3–8]. Technological advances, especially in the development of next-generation sequencing (NGS) and liquid chromatography-mass spectrometry (LC-MS), promoted a resurgence of the field in recent years, enabling whole transcriptome-wide profiling of RNA modifications including: N^6-methyladenosine (m^6A), N^6, 2′-O-dimethyladenosine (m^6Am), 5-methylcytosine (m^5C), 5-hydroxymethylcytosine (hm^5C), inosine (I), pseudouridine (Ψ), N^1-methyladenosine (m^1A), 2′-O-methylation (Nm), N^4-acetylcytidine (ac^4C), N^7-methylguanosine (m^7G) and dihydrouridine (D)[3].

RNA modifications have been found in almost all kinds of RNA species. Transfer RNAs (tRNAs) are especially rich in modifications. At least 111 types of modifications can be found in

tRNAs[4]. Eukaryotic tRNA molecules each contain 13 modified bases on average[9]. 33 types of modifications can be found on ribosomal RNAs (rRNAs) [8,9], and human rRNA contain 228 modifications of 14 different types[10]. In eukaryotes, mature messenger RNAs (mRNAs) are modified with a 5' cap, which is a m^7G modification appended to the 5' of the first nucleotide via a triphosphate link. Some cap structures contain further modifications of Nm on the first one or two nucleotides, where the 2' OH of the ribose is methylated to 2'-O-methyl (2' OMe)[11,12]. These modifications are essential for mRNA stability maintenance, splicing, polyadenylation, nuclear export, and cap-dependent translation[12]. Apart from Nm, internal mRNA modifications also include m^6A, m^1A, m^5C, hm^5C, Ψ and at least 11 others[4–6]. RNA modifications were also found on long non-coding RNAs (lncRNAs), microRNAs (miRNAs), piwi-interacting RNAs (piRNAs), small nuclear RNAs (snRNAs), small nucleolar RNAs (snoRNAs) and their corresponding precursors[13].

1.1.2 m^6A: distribution, biochemistry, and biological functions

While Ψ is the most abundant RNA modification in general, m^6A is the most abundant internal mRNA modification in eukaryotes [3–8]. First discovered in 1974[14], it has been the most well-studied RNA modification in the past decades[6]. From a chemistry point of view, m^6A is a simple N-monomethylation of the 6-amine of adenosine. When alternatively presented as a DNA modification (N^6-methyldeoxyadenosine, 6mdA), it is the predominant form of DNA methylation in prokaryotes[15]. Some viral RNAs were also discovered to be containing m^6A, including human immunodeficiency virus-1 (HIV-1) genomic RNA[16]. m^6A is highly prevalent in mammalian transcriptome: approximately 0.1-0.4% of the total adenosine are m^6A methylated[17]. This also translates to 0.15-0.6% methylation of total adenosines in poly(A)$^+$ RNA, 3-5 m^6A per transcript,

10-13 m^{6}A sites per gene, or at least one m^{6}A in 25-60% of transcripts[18–21]. m^{6}A is also found in non-coding RNAs. Some lncRNAs are heavily methylated: X-inactive specific transcript (XIST), which is responsible for X-chromosome inactivation during mammalian development, contains at least 78 m^{6}A sites[22]. Metastasis associated lung adenocarcinoma transcript 1 (MALAT1), which is related to tumor metastasis and malignancy, contains at least 31 m^{6}A sites[23]. m^{6}A is also found on primary microRNAs (pri-miRNAs)[24], U6 snRNA[25], snoRNAs, tRNAs and rRNAs[26]. Human rRNAs contain 2 m^{6}A sites, one each on 18S rRNA and 28S rRNA[10,27]. Recent studies revealed that retrotransposons, for example long interspersed elements (LINE) 1, are also heavily enriched in m^{6}A methylations[28–30].

Extensive studies have been conducted in the biochemistry of m^{6}A related proteins, also known as "writers" (methyltransferases), "erasers" (demethylases) and "readers" (m^{6}A binding proteins). The m^{6}A methylation complex was first isolated from HeLa cell as 200 kDa and 875 kDa subcomplexes[31]. The core methyltransferase of the complex, methyltransferase-like 3 (METTL3) was first purified and characterized in 1994[32,33]. Wilms' tumor 1-associating protein (WTAP) was then discovered to be a major adaptor protein for METTL3[34,35]. METTL14, initially discovered via phylogenic analysis in comparison to METTL3 and thought to be an additional methyltransferase, later was proved to be an essential component forming heterodimeric complex with METTL3 yet not enzymatically active when presented alone[36–38]. Vir-like m^{6}A methyltransferase associated (VIRMA, also known as KIAA1429)[39], RNA binding motif protein 15/15B (RBM15/15B)[22], Cbl proto-oncogene like 1 (CBLL1, also known as Hakai), and zinc finger CCCH-type containing 13 (ZC3H13)[40–42] were further identified as key components of the ~1 MDa full complex. For non-coding RNAs, METTL16 was identified as the methyltransferase

for U6 snRNA[43] and MAT2A pre-mRNA[44], METTL5 for 18S rRNA[45], and Zinc finger CCHC-type containing 4 (ZCCHC4) for 28S rRNA[46].

The discovery of in vivo m^6A demethylation activity in fat mass and obesity-associated (FTO) protein[47], a Fe (II)- and 2-oxoglutarate-dependent dioxygenases originally found to demethylate m^3dT/m^3U in vitro [48], inspired the concept of "RNA epigenetics"[49] or "epitranscriptomics"[50,51] that envisioned a dynamic and reversible regulation of RNA modifications in parallel to epigenetic regulation of DNA/histone modifications. m^6Am was also later reported to be a substrate of FTO in vivo[52,53], and alkB homolog 5 (ALKBH5) was identified as a second m^6A demethylase[54]. YT521-B homology (YTH) family of proteins were among the first discovered and the most well studied m^6A binding proteins, in mammals including YTH N^6-methyladenosine RNA binding protein (YTHDF) 1-3 and YTH domain containing proteins (YTHDC) 1,2[22,55,56]. The YTH domain consists of 100–150 amino acid residues, with 4–5 alpha helices surrounding a curved six-stranded beta sheet[57]. YTHDF paralogs are highly conserved cytosolic proteins, sharing about 85% sequence similarity[58]. Other than the conserved C-terminal YTH domain they also contain a N-terminal intrinsically disordered region (IDR) which promotes phase-separation and stress granule formation[59–63]. On the contrary, YTHDC 1&2 bear less sequence resemblance to other YTH family proteins, and YTHDC1 is the only nuclear localized protein of the five[53,56,64]. Other identified m^6A binding proteins outside of the YTH family include insulin-like growth factor 2 mRNA-binding proteins (IGF2BP) 1-3[65], heterogeneous nuclear ribonucleoproteins (HNRNPs)[66–68], eukaryotic initiation factor (eIF) 3[69], fragile X mental retardation protein (FMRP)[70,71], and proline rich coiled-coil 2 A (Prrc2a)[72].

Since its discovery, m^{6}A has been associated with various biological functions, including mRNA stability maintenance[19,62,65,69,70,73–76], mRNA nuclear export[71,77], mRNA translation[62,63,65,7879], pre-mRNA splicing[44,66], miRNA biogenesis[24,80], lncRNA binding[22,81], etc. The most well-established impact of m^{6}A is its negative correlation with mRNA stability and promotion of mRNA turnover. Almost half a century ago, m^{6}A in the cytoplasm was already found to be losing the radioactive label faster than caps (m^{7}G), suggesting a m^{6}A-preferential degradation[82]. Later comparison of mRNA half-life in METTL3 knockdown (KD) cell lines versus normal control reaffirmed this observation[39,73]. Recently, multiple studies of quantitative m^{6}A sequencing were able to categorize all mRNA transcripts by their m^{6}A methylation levels, which again showed strong negative correlation with their half-lives[83–85]. Mechanistically, YTHDF2 was found to promote cytosolic mRNA degradation by recruiting a deadenylation complex, CCR4-NOT, to the m^{6}A methylated transcripts[69]. YTHDF1 was found to promote translation through interaction with translation initiation factors[63], while YTHDF3 was found to promote both translation and degradation possibly through the interaction with YTHDF1 and YTHDF2[62]. Recent studies suggested that YTHDF1 also promotes degradation through interaction with Argonaute RISC catalytic component (AGO) 2 and phase separation[76], which possibly showed a synergistic effect of all three YTHDF paralogs on the acceleration of mRNA turnover[75]. Intriguingly, in acute myeloid leukemia (AML) cells, studies showed that IGF2BPs, which are also m^{6}A binding proteins, increase the stability of mRNA they interact with[65,86]. This indicates that although in general m^{6}A stoichiometry is negatively correlated with mRNA stability, the downstream mechanism can be diverse and complicated.

The regulation of mRNA stability through m^{6}A can be crucial in some physiological and pathological processes. Since the reduction in mRNA half-life effectively reduces the gene product

associated with that transcript, m^{6}A could possibly serve as an additional layer of gene expression regulation downstream to the changes in transcription factors/chromatin states. For example, during maternal-to-zygotic transition, maternal mRNA clearance is required for the zygotic genes to be activated[87]. Such process was found to be m^{6}A dependent, as knockout (KO) of Ythdf2 in mice[88] and zebrafish[89] both resulted in developmental arrest. Mettl3 KO mice, in contrast, were embryonically lethal by embryonic day 5.5 (E5.5), while Mettl3 KO mouse embryonic stem cells (mESCs) were viable yet did not differentiate[90–92]. Similar mechanism also exists in endothelial-to-hematopoietic transition during hematopoietic stem cell (HSC) development in zebrafish[93]. m6A were also found to be important in the development of neuronal cells[94] and reproductive cells[95,96]. Other than developmental controls, a very recent study demonstrated that X-chromosomal genes are also relatively m^{6}A depleted compared to autosomal genes, which resulted in longer half-lives and therefore compensation of gene expression level otherwise lost due to X-chromosome inactivation[74].

1.2 m^{6}A profiling methods

1.1.3 Biophysical profiling of RNA modifications

As previously mentioned, pioneering studies of RNA modifications dated back to several decades ago. At that time, the profiling of RNA modifications was mostly based on their biophysical properties. Common techniques include radioactive isotope labeling, ion-exchange chromatography, affinity chromatography, two-dimensional thin layer chromatography (2D-TLC) and dot blot, in combination with nuclease digestion or selective chemical cleavage[3,8].

Thin layer chromatography (TLC): TLC is a broadly used method for analytic and preparative separation of small molecules mixture. As an improvement over paper chromatography (which identified the first modified base m5C[1]), TLC utilizes a variety of stationary phases including silica gels, aluminum oxide and cellulose to achieve differential retention of small molecules. For nucleotides, cellulose stationary phase is most used. The distinction could be further improved by expanding in two orthogonal directions, termed 2D-TLC. By comparing to the standard retention factor (Rf) reported using three solvent systems[97], the identity of nucleotides could be deduced. Although nucleotides could be visualized by UV absorption, which is a commonly used low-cost visualization method in TLC, radioactive isotopic labeling using [^{32}P], [^{14}C], or [^{3}H] are often preferred due to its femtomol level[97,98] sensitivity. This method was once standard for tRNA modification analysis. When applied to rRNA, the analysis became unsuitable due to much fewer modifications being difficult to distinguish from unmodified nucleotides.

Ion exchange chromatography: Ion exchange chromatography relies on electrostatic interactions between anion/cation residues on the stationary phase and charged small molecules in the mobile phase to achieve differential retention. This method is widely used for biomolecules because proteins, nucleic acids, carbohydrates, and many metabolites are often hydrophilic, polar, and charged. Nucleic acids and nucleotides are negatively charged by phosphate groups, therefore anion exchange resins are suitable for their separation, for example diethylaminoethyl (DEAE)-Sepharose (commercialized name for a type of crosslinked agarose). Visualization of eluate could be achieved by UV absorption or radioactive signals, like the case in TLC.

Some sequence information is retained in these assays. With a combination of nuclease P1, which promiscuously hydrolyzes all phosphodiester bonds; RNase A, which cleaves after pyrimidines; and RNase T1 which cleaves after guanosines, it is possible to deduce the position of the modification. RNase H, which cleaves the RNA strand of DNA: RNA hybrid, could be used with DNA probes to deduce the modification position in a sequence specific manner. However, for RNA species longer than tRNA it would be time consuming to sequence the modification sites solely by digestion and chromatography.

Technological advances in mass spectrometry (MS) provided better tools. In mass spectrometry, molecules are ionized into gas-phase ions and analyzed by their corresponding mass-to-charge ratio (m/z) using one of the many designs of mass analyzers. The advantage of measuring m/z of modifications for their identification lies in its generality and directness. Except for pseudouridine versus uridine, all other RNA modifications form a mass shift compared to their base nucleotides. On the contrary, despite there are modifications like m^1A and m^7G which significantly alters the original electrostatic charge distribution[99], many other modifications like m^6A only result in subtle changes from the base nucleotides which are difficult to observe via other biophysical measurements.

Many ionization methods have been developed for mass spectrometry[100]. The two most used for nucleic acid/nucleotides are:

Electron spray ionization (ESI)[101]: Samples are dissolved in polar solvents and passed through a thin needle applied with a high voltage. This generates a fine mist of tiny, charged droplets at the needle tip. Desolvation gas, which is typically warm nitrogen, is blown towards the charged droplets and reduces the volume of droplets through evaporation of solvents. After the

droplets shrink to a critical size, the repulsion force of concentrated charges will eventually exceed the surface tension and result in an explosion (also known as Coulombic fission) of the droplet, producing charged ions without the solvent into the gas phase. ESI source is inherently suitable for a tandem high performance liquid chromatography (HPLC) separation prior to the MS analysis, for less complexity of ions at any given point and therefore better identification qualitatively and quantitatively. This adds another layer of differentiation for modified nucleotides, for example m^6A is more hydrophobic than adenosine therefore has longer retention time on a reversed-phase (RP) HPLC column.

Later advances applied tandem mass spectrometry (MS/MS) for accurate quantification of modified nucleotides[102]. MS/MS is a technology of adding a fragmentation step to the first ion, then the fragmented secondary ion is analyzed by another mass analyzer (MS2). Similar to the ion source and mass analyzer, there are many designs for MS/MS, with triple quadruple (QQQ) being the most commonly used for RNA modification analysis. A quadruple is a design of mass analyzer which features four parallel poles with voltages between each opposing pairs. Only ions of a certain m/z will reach the detector for a given set of voltages, while the other ions collide with the poles[103]. In QQQ, the first quadruple is used for selection of ions with a predefined m/z (also known as precursor ions). This process is known as multiple reaction monitoring (MRM). If a predefined retention time window is also selected alongside the m/z, the sensitivity and accuracy could be further enhanced, known as dynamic multiple reaction monitoring (DMRM). Then the second quadruple fragments the precursor ion via collision with neutral gas (N_2 or Ar) under a selected voltage, which process is known as collision induced dissociation (CID), generating product ions. The voltage used for CID could be adjusted to fine tune the position of cleaved chemical bonds. For modified nucleotides, this is typically set to cleave the N-glycosidic bond, leaving the

cationized base as the product ion. With a combination of retention time, MS1 spectra and MS2 spectra, each modified nucleotide could be precisely assigned. For absolute quantitation, external standards could be used for calibration calculated with the integration of UV absorption in the LC chromatograph.

Samples used for HPLC-QQQ-MS/MS analysis are typically first digested into nucleotides by nuclease P1, then dephosphorylated using commercialized alkaline phosphatases (optionally also with snake venom phosphodiesterase which also dephosphorylates the cap and 2'3'-cyclic phosphates)[104]. If left unremoved, phosphate group will attract metal cations (Na^+, for example) in the solution forming ion-pairs, which reduces the ionization efficiency in the ESI source, and complicates the spectra by forming several different adduct species.[105–107]. Due to the same reason samples are also desalted. It is worth noting that, protein post translational modifications (PTM) have long been identifiable using ESI-MS/MS at peptide-level, retaining the sequence information by analyzing the product ions in MS2, rather than full digestion to amino acids as what is common for RNA modifications. This is partly due to the cation adduct issue just discussed, also due to lack of an established MS2 database since the interest in RNA modifications only revived recently[8]. Therefore, the MS based sequencing of RNA modifications is currently limited to short oligonucleotides, for example siRNAs[105] and fragmented rRNAs[10], yet hold huge potential.

Matrix assisted laser desorption/ionization (MALDI)[101]: As the name suggests, MALDI utilizes matrix to promote desorption and ionization of analytes that are otherwise difficult to ionize, which are often large biomolecules intended to be analyzed with minimal degradation. Matrix are small organic crystals that have a high absorption at the corresponding laser wavelengths[108]. Again due to the cation adduct issue, oligonucleotides require dedicated matrices

to achieve equal ionization efficiency compared to proteins, namely: 2,4,6-trihydroxyacetophenone (2,4,6-THAP), 2,3,4-trihydroxyacetophenone (2,3,4-THAP), picolinic acid (PA), 3-aminopicolinic acid (3-APA), 6-aza-2-thiothymine (ATT), 5-methoxysalicylic acid (5-MSA), quinaldic acid (QA), pyrazinecarboxylic acid (PCA), 3-hydroxycoumarin (3-HC), and 3,4-diaminobenzophenone (DABP)[109]. The resulting large ion is typically analyzed using time-of-flight (TOF) mass analyzer, which measures the time ions take to travel a certain distance after being accelerated in a predefined electric field. The modification is read out as a mass shift from the control oligonucleotide, which could be quantitated as a modification ratio relative to control by calculating the peak area of shifted peak versus the original peak on the MS spectra. By using a combination of RNases digestion, it is also possible to obtain some sequence information.

1.2.2 Site-specific, low throughput m^{6}A profiling

A shared limitation of all previously described biophysical assays is their inability to precisely locate the modification within a transcript sequence. Although RNase digestion could provide some information, the specificity of RNases A/T1 lacks enough complexity, and the procedure is too labor intensive to allow for parallel analysis of multiple transcripts. That being said, RNase H possesses the potential of site specificity, in that the opposing DNA strand could be artificially synthesized to direct the cleavage site. Based on this idea Liu et al. developed:

Site-specific cleavage and radioactive-labeling followed by ligation-assisted extraction and thin-layer chromatography (SCARLET)[110]: RNase H cleaves RNA only at sites where the 2' hydroxyl of the ribose is not modified.[111,112]. Therefore, by designing a DNA-2' OMe RNA chimeric probe complementary to the sequence of the target RNA, the RNase H cleavage will happen on the phosphodiester bond between the last 2' OMe ribonucleotide and the first

deoxyribonucleotide (2' H). If the last 2' OMe ribonucleotide on the probe is positioned at the very opposite to the assumed modified ribonucleotide on the target, this cleavage will expose the modified ribonucleotide as the first 5' base of the 3' half of the cleaved RNA. The resulting RNA fragments were subsequently dephosphorylated with calf intestine phosphatase (CIP), and rephosphorylated with T4 polynucleotide kinase (PNK) and γ-[^{32}P] ATP. Radioactively labeled fragments were isolated using polyacrylamide gel electrophoresis (PAGE), with optional ligation to increase the length to facilitate purification. Eventually the RNA fragment was trimmed and digested to nucleotides by RNases A/T1 and nuclease P1 and run by 2D-TLC for identification.

This method still relies on biophysical properties for m^{6}A identification, therefore, is relatively labor-intensive. In 2018, Xiao et al. reported a protocol that also requires artificially synthesized complementary probes, yet rely on quantitative polymerase chain reaction (qPCR) for detection:

Single-base elongation- and ligation-based PCR amplification (SELECT)[113]: This method relies on the nick translation ability of Bacillus stearothermophilus (Bst) DNA polymerase. DNA polymerases that possess the proofreading function (biochemically, 3'-5' exonuclease activity) can excise the nucleotides surrounding a nick, then extend the 5' remaining strand in a 5' to 3' direction using the DNA-dependent DNA synthesis activity, effectively "moving" the nick forward and replacing the original nucleotides downstream. The length of the newly synthesized strand is dependent on the processivity of the polymerase used, and frequently requires the aid of DNA ligases to seal the nick and achieve full length. This process is known as nick translation and was once a popular method for radioactive labeling[114]. In SELECT, two DNA probes were designed complementary to the RNA target, with the assumed modified nucleotide skipped. After

annealing, this forms a DNA: RNA heteroduplex with only the modified base exposed (technically not a nick but rather a single nucleotide excision, which nonetheless could be 5' to 3' digested by polymerases). Nick translation was then performed with Bst DNA polymerase and Chlorella virus-1 DNA Ligase (commercialized as SplintR ligase). Due to the existence of m^6A, the nick translation efficiency would be slightly lower than if an unmodified adenosine was present, which could be reflected as higher quantification cycles (C_q) in the subsequent qPCR amplification of the product. This method is facile and quantitative, serving as a standard verification method for some later studies[84,115]. However, it would be unsurprising that for lowly methylated sites, changes in Cq (ΔCq) can be too minimal to be statistically significant.

A shared limitation of these probe-based assays is the requirement of a priori knowledge of the assumption of m^6A presence. Combined with the low-throughput nature, these assays best serve as the orthogonal validation or functional investigation methods for sites identified in de novo high-throughput m^6A profiling.

1.2.3 Antibody based, high throughput m^6A profiling

Next generation sequencing (NGS) has fundamentally changed many disciplines of biology in the past two decades. The most commonly used platforms, developed by Illumina, feature a technology known as "sequencing by synthesis" (SBS)[116,117]. Double stranded (ds) DNA molecules are first denatured into single strands (ss), then annealed to adapters deposited on specialized glass slides (called flow cells). Through a process called "bridge amplification", the ssDNA is first amplified using its annealed adapter as the primer. After denaturation, the amplification product bends over (bridges) and anneals to another adapter complementary to its distant end to initiate another round of amplification. After several rounds the original DNA will

be amplified into localized and flow cell-tethered clones, known as clusters. Then the clusters were amplified using sequencing primers and fluorescent reversible-terminator dNTPs. These dNTPs temporarily terminate the strand extension once incorporated, and their fluorescent tags are excited by lasers. The fluorescence signal is recorded and interpreted as a sequencing result of either one of the four bases, then the fluorescent tag and 3' blocking group are cleaved off, unblocking the growing strand end for another cycle of reaction.

Initially used for genome sequencing, the NGS technology was soon applied to transcriptomes, leading to the technology known as RNA sequencing (RNA-seq)[118,119]. RNA-seq has been the golden standard for gene expression analysis in the past decade and continues to be so. More than 100 different sequencing methods based on RNA-seq has been published[120,121], handling different topics ranging from gene expression, RNA-protein interactions, RNA secondary structures, low-input RNA and RNA modifications. The most commonly used RNA-seq strategy by now is the sequencing of short cDNA fragments. mRNAs are typically enriched to avoid most sequencing output (known as reads) being aligned (known as mapped) to the most abundant rRNA or tRNA region, either by oligo-dT complementary pull-down or rRNA complementary probe-based depletion. The depletion could be achieved either by biotinylated probe pull-down or RNase H based digestion. Oligo-dT based enrichment leaves mostly mature mRNA fractions (or poly(A)$^+$ RNA, by definition of oligo-dT) together with some remainders of rRNA, consisting of 1-5%[122] of the starting material; while rRNA depletion method leaves mRNA, pre-mRNA, and other non-coding RNAs intact, alongside some remainders of rRNA, consisting of approximately 10% of the starting material. Then the RNA is subjected to physical (mechanical shearing or sonication), chemical (heating with divalent metal ions) or enzymatic fragmentation to yield 150-200 bp[120] short fragments. Shorter fragments form more clusters in Illumina sequencers, yet too short

fragments cannot be mapped to unique positions on the genome. Then the RNA is reverse transcribed (RT) and adapter-ligated using various methods. More recent methods also allow for full-length RT before fragmentation of the cDNA using Tn5 transposase[123–125], as opposed to the traditional "fragmentation before RT" design. 20-30 million reads [120] are typically required for each sample to ensure enough coverage to weigh against the inherent error rate of the sequencer.

RNA-seq itself could be viewed as an abridged version of genome sequencing, enriched in transcriptome or exosomes (mRNA-seq). Similarly, modification sequencing could be achieved by enrichment of the modification.

Methylated RNA immunoprecipitation sequencing (MeRIP-seq)[126]/m^{6}A-seq[55]: RNA immunoprecipitation sequencing (RIP-seq)[127] is the tandem assay combining immunoprecipitation (IP) of RNA binding proteins (RBP) with RNA-seq. The RBP binding targets are co-eluted either by protein-RNA binding (native RIP) or chemical fixation with formaldehyde (alternatively known as RIPiT-seq[128], where iT = "in tandem"). Using commercialized m^{6}A-specific antibody, this method was extended to m^{6}A modification by two groups almost simultaneously in 2012. If the enriched RNAs were subjected to qPCR analysis instead of NGS, where the enrichment over the IP input could be calculated form ΔCq and normalized against housekeeping genes (also known as ΔΔCq method[129]), this serves as a low-throughput version of MeRIP-seq known as MeRIP-qPCR.

By adding a fragmentation step prior to the immunoprecipitation, it is possible to narrow down the sequences that contain m^{6}A to a 100-200 bp window. Then the precise location of m^{6}A can be deduced from the m^{6}A consensus motif, DRm^{6}ACH (D = A/G/U, R = A/G, H = U/A/C)[21,26,53,56,64,130–133]. Initially identified as Rm^{6}AC via nuclease digestion and chromatography[134],

the consensus sequence was later identified in detail by sequencing as DRACH for Homo sapiens and Mus musculus, RRACH for Arabidopsis thaliana, and RGAC for Saccharomyces cerevisiae[130]. METTL3/14 complex[37] and YTH domain[73] were both found to bind preferentially to the consensus motif. However, within a MeRIP-seq peak (the 100-200 bp aligned fragments) there could possibly be more than one DRACH motifs. Actually, more recent site-specific studies[83–85,135] are finding a much larger dataset of m^{6}A than previously estimated, which is also supported mechanistically[19,115,136]. In a 2023 work by Uzonyi et al [19], the authors presented an in silico estimation of 3.4 m^{6}A per "peak", which was defined as a consecutive stretch of bases with a predicted methylation signal > 0. This signifies the limitation in the resolution of MeRIP-seq.

m^{6}A individual-nucleotide-resolution cross-linking and immunoprecipitation sequencing (miCLIP-seq)[137]/photo-crosslinking-assisted m^{6}A sequencing (PA-m^{6}A-seq)[138]: As an effort to improve the site-specificity of MeRIP-seq, miCLIP-seq and PA-m^{6}A-seq were developed based on the idea of crosslinking and immunoprecipitation sequencing (CLIP-seq)[139] and photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation (PAR-CLIP)[140]. 254 nm ultraviolet (UV) light is capable of inducing free radical reactions between amino acid residues and RNA bases, forming a covalent bond if the two residues are physically close enough. Compared to native RIP, this crosslink greatly enhances the binding affinity between RBP and its target, enabling stringent wash protocols which improve target specificity. Another advantage is that crosslinked amino acid residues can disrupt Watson-Crick base pairing, inducing mispairing (mutation) or premature truncation (stop) or skipped bases (deletion) during RT. In the case of m^{6}A, such "RT signature" could be used with the DRACH motif to deduce the precise location of the modification. 254 nm UV is considered energy-intensive for biomolecules, inducing undesirable non-specific crosslinks and chemical bond cleavages. By metabolically

incorporating a UV-sensitive modified base, 4-thiol uridine (4sU), CLIP reactions could be performed under 365 nm UV lights. This improvement of specificity and crosslinking efficiency is also known as PAR-CLIP. In the case of m^{6}A, because GGACU is the most prominent motif among DRACHs[130], the likelihood of having a U adjacent to m^{6}A is high.

MeRIP-seq continued to be more popular than miCLIP-seq after the publication of the latter[141]. The 254 nm UV is prone to non-specific crosslinking and random mutagenesis. Metabolic labeling adds the factor of incorporation efficiency into the already existing crosslinking efficiency and library construction efficiency, which further reduces the sensitivity and transcriptome coverage, while increasing the required RNA sample amount. The RNase T1 digestion combined with the crosslink-induced RT stop typically generates shorter reads than standard RNA-seq. Shorter reads are more difficult to map to unique positions in the genome[142], which problem is exacerbated when mutations/stop/deletions are present in the said reads.

1.2.4 Antibody free, high throughput m^{6}A profiling

There are two shared limitations with the antibody-based m^{6}A profiling: 1. Antibody induced non-specificity[141,143–145]. Because antibodies are raised using synthetic moieties (known as haptens) attached to larger proteins for immunostimulation[8], the recognition pattern on RNA strands will inevitably be different to some extent. One early study[145] published just one year after the initial MeRIP/m^{6}A-seq works already identified ~50% of the peaks to be false positives. An interesting feature of MeRIP-seq data is its inclusion of ~25% of non-DRACH sites[146]. In the first paper that reported the m^{6}A consensus motif[134], chromatography combined with RNase digestion and radioactive labeling revealed that apart from 70% GAC and 30% AAC, only "very low amount" of radioactivity remained in other peaks. Later PAR-CLIP study revealed the preferential binding

of METTL3, METTL14 and WTAP to the consensus sequence[36]. Given that later discovered methyltransferases METTL16, METTL5 and ZCCHC4 only methylate a limited range of substrates (among these three, only METTL16 has a physiological mRNA substrate), it is unlikely that non-DRACH sites will be common, at least not in mRNA/poly(A)$^+$ RNA studies. In contrast, recent site-specific quantitative methods[83,85,135] reported much lower non-DRACH ratios. Another rational explanation for MeRIP's overestimation of non-DRACH site lies in the fact that anti-m^{6}A antibody also binds to m^{6}Am. Later studies found that the abundance of m^{6}Am is 2.2-11.4%[147] of that of m^{6}A, and that m^{6}Am resides in its own consensus motif of BCA (B = C/U/G)[147,148].

2. Lack of quantitation. Variations in the pull-down efficiency and binding affinity to the antibody due to motifs and secondary structures[141] disabled accurate quantitation of m^{6}A stoichiometry in antibody-dependent profiling methods. Why is stoichiometry a concern? Again, in the pioneering paper of m^{6}A consensus motif[134], the authors estimated that with the assumption of each base sharing an equal appearance chance, the consensus motif of RACH should occur once per 43 nucleotides (1/2 × 1/4 × 1/4 × 3/4). Although m^{6}A is the most abundant internal mRNA modification, it is still rather scarce compared to this estimation. Based on MeRIP results, approximately 5% of the DRACH motifs are methylated[21]. This could be an underestimation due to multiple m^{6}A sites present within each peak, or an overestimation due to false positives. Within those 5% "peaks", not all transcripts are equally methylated. This fact could be deduced from MeRIP-qPCR, or the quantitative low-throughput assays described in Section 1.2.2. Therefore, the biological information implied by m^{6}A is not just if an adenosine is methylated, but also in what ratio, or stoichiometry.

FTO-assisted m^{6}A selective chemical labeling (m^{6}A-SEAL)[149]**/m^{6}A-label-seq**[150]: These two methods are antibody-free yet not quantitative. m^{6}A-SEAL utilizes FTO demethylation reaction intermediates for enzymo-chemical labeling of m^{6}A. FTO, as a dioxygenase, demethylates the amine group via oxidation, which first transforms the m^{6}A into N^6-hydroxylmethyladenosine (hm^{6}A), then N^6-formyladenosine (f^{6}A). The N-formyl group will be hydrolyzed to N-H, effectively demethylates m^{6}A to A. The authors reacted hm^{6}A with dithiothreitol (DTT) to form N^6-dithiolsitolmethyladenosine, which contains a free thiol that could be functionalized with thiol-reactive linkers (the authors used methanethiosulfonate) and biotin. Then the biotin was used for pull-down enrichment. This method circumvents the antibody specificity issue yet does not provide accurate quantitation due to the use of pull-down enrichment. m^{6}A-label-seq utilizes metabolic labeling of N^6-allyladenosine (a^{6}A). Because METTL3/14, like many other methyltransferases, use S-Adenosyl methionine (SAM) as the methyl donor (also known as cofactor). SAM is metabolically synthesized from methionine. By feeding the cell with chemically synthesized methionine analog, Se-allyl-L-selenohomocysteine (methionine is S-methyl-L-homocysteine), Se-allyl, Se-adenosylselenohomocysteine (allyl-SeAM) will be synthesized in vivo and used by METTL3/14 to transfer the allyl group to physiological m^{6}A sites. This transforms A to a^{6}A in place of m^{6}A, which could be further chemically modified (with iodine) to N^1, N^6-cyclized adenosine[151], which could induce mispairing in RT with low fidelity reverse transcriptases (HIV-1 RTase, for example). This method does not provide quantitation either, like PA-m^{6}A-seq, due to variations in the efficiency of metabolic labeling.

m^{6}A-sensitive RNA-endoribonuclease–facilitated sequencing (m^{6}A-REF-seq)[152]**/MAZTER-seq**[153]: Both methods rely on the property of Escherichia coli RNA endonuclease MazF[154]. This enzyme digests ssRNA at the 5' of ACA sequence yet is blocked by

m^{6}ACA. This interesting feature leads to site specific identification of m^{6}ACA and quantitation by counting the truncated reads versus total reads mapped to a given transcript. Using this method, as much as 17,007 sites were identified from 100 ng of mRNA, which was a great improvement over antibody-based approaches. The limitation of this method is apparent: m^{6}ACA is not prevalent among DRACH motifs. The authors' own estimation was that at most 16-25% of total m^{6}A can be detected. Like MeRIP, this method could also be adapted to qPCR as a low throughput verification method.

Deamination adjacent to RNA modification targets (DART-seq)[155]: DART-seq utilizes the famous RNA C-to-U editor apolipoprotein B-editing enzyme, catalytic polypeptide-1 (APOBEC1), and exploits the fact that in DRACH there is always a C adjacent to m^{6}A. A YTH domain was fused genetically to APOBEC1 and overexpressed in cell lines to direct the deaminase to m^{6}A sites. The limitation of this method is that transfection/transduction is not always available for biological samples, for example frozen or formalin-fixed, paraffin-embedded (FFPE) tissue samples which are predominant in clinical studies. Overexpression of APOBEC1 also might result in off-target editing and unintended changes in cellular state. The strength of this method is that it does not involve any enzymatic or chemical treatment after cell lysis other than the RNA-seq library construction. Therefore, it could be used in tandem with any developments in RNA-seq methods. This was excellently demonstrated in **single-cell (sc)DART-seq**[156], where scRNA-seq technologies of Smart-seq2[125] and commercialized 10x Genomics Next GEM 3'[157] were applied. Using regular amount of input samples (also known as bulk, in contrast to single-cell), DART-seq can identify 12,672 sites with as low as 10 ng of total RNA, which is impressively low even up to date. scDART-seq was able to detect 16,934 sites from 10,352 cells.

Glyoxal and nitrite-mediated deamination of unmethylated adenosines (GLORI)[83]: This method features chemical deamination of A instead of m^{6}A to inosine (I). Glyoxal, a highly reactive dialdehyde, react with the purine-amine of A and G via condensation reactions. If then oxidized by sodium nitrite ($NaNO_2$), the 6-amine of A will be converted to diazonium and subsequently hydrolyzed to hydroxyl, forming inosine (the enol tautomer). G, on the other hand, can be protected with boric acid to avoid oxidation, and completely reversed to G via heating at 95 ºC for 10 min in an acidified formamide-water mixture. The conversion rate of this method is exceptionally high (98-99%), which enables highly accurate quantitation of m^{6}A ratio. Another advantage of the method lies in its low-cost and easily commercially available reagents. The method detected 176,642 sites from 200 ng of mRNA, more than 10 folds higher than what was reported in antibody-based methods. As previously discussed, this magnitude was supported mechanistically[19,115,136]. The sample amount requirement is much higher than contemporary methods, indicating sample loss due to harsh reactions.

m^{6}A-selective allyl chemical labeling and sequencing (m^{6}A-SAC-seq)[84]**/evolved TadA-assisted N^6-methyladenosine sequencing (eTAM-seq)**[85]: These two methods were developed by us and will be described in detail in Chapters 2-4. The optimized m^{6}A-SAC-seq[135] identified 71,547 sites from 50 ng of mRNA or 31,233 sites from 2 ng of mRNA, while eTAM-seq identified 69,834 sites from 50 ng of mRNA. The two had 76% overlap, cross-validating each other.

1.3 Scope of this dissertation

This dissertation is focused on the development of site-specific and quantitative m^{6}A profiling methods.

Chapter 2 discusses m^{6}A-SAC-seq, the antibody-free whole transcriptome profiling of m^{6}A that relies on the m^{6}A-favored methyltransferase activity of Methanocaldococcus jannaschii dimethyltransferase-1 (Mjdim1). Data is presented to reflect the validity of the detected sites. Limitation of the method is also discussed.

Chapter 3 discusses optimization of m^{6}A-SAC-seq, and modification oriented RNA-seq in general. Rationales, detailed methods, and improvements in data are discussed. Comparison with many m^{6}A profiling methods described in Chapter 1 Section 1.2.3 & 1.2.4 will be discussed.

Chapter 4 discusses eTAM-seq, the highly quantitative, non-biased whole transcriptome profiling of m^{6}A that relies on the m^{6}A-repelled deamination activity of tRNA adenosine deaminase (TadA) evolved variant 8.20. Data will be presented to demonstrate validity, and cross-validation with m^{6}A-SAC-seq.

Chapter 5 discusses experimental designs and preliminary results for single-cell m^{6}A profiling. General concerns regarding the methodology are discussed in detail.

Chapter 6 discusses experimental designs and preliminary results for eTAM-seq derived RNA accessibility profiling.

Chapter 7 summarizes the dissertation and discusses future directions in the single-cell m^{6}A sequencing and third-generation sequencing of RNA modifications.

Chapter 2 m^{6}A-SAC-seq

2.1 Introduction

m^{6}A-SAC-seq was developed based on previous discoveries that Dim1/KsgA family dimethyltransferases could convert m^{6}A residues to dimethyl A (m^{6,6}A)[158,159], and that iodine treatment of N^6-allylation generates N^1-N^6 cyclization which could be read out as mismatches during reverse transcription (RT)[151]. In m^{6}A-SAC-seq, allyl-SAM is synthesized as previously reported, and purified with semi-preparative HPLC. The dimethyltransferase that originated from Methanocaldococcus jannaschii, known as MjDim1, is recombinantly expressed and purified. The purified enzyme and cofactor are quality-controlled on synthetic probes using LC-MS to check the conversion ratio of m^{6}A. The library preparation starts by poly(A) selection or ribosomal depletion of the sample RNA. Either approach could be chosen depending on the RNA integrity and desired coverage of transcriptome (mRNA-only, or also including non-coding RNA, respectively). Next, the poly(A) tract is depleted using oligo-dT probes and RNase H, which reduces background labeling and improved mapping ratio. The RNAs is chemically fragmented to <200 bp to facilitate MjDim1 labeling and adapter ligation. After end-repair and clean-up, the fragmented RNA is ligated to a 3' adenylated ssDNA adapter. The adapter is 5' biotinylated to facilitate on-beads purification and labeling. Reverse transcription is performed using human immunodeficiency-1 (HIV-1) reverse transcriptase (RTase), which shows the best balance between high mutation ratio and high readthrough as compared to Moloney murine leukemia virus (MMLV) RTase and related variants. The cDNA is ligated to another ssDNA adapter carrying unique molecular identifiers (UMIs), then PCR amplified with indexed primers. The resulting dsDNA is size selected for >150

bp fragments and sequenced using Illumina NGS platforms (**Figure 2.1a**). The analysis could then be performed using our codes (**Figure 2.1b**).

Although m^{6}A-label-seq[150] uses a similar chemistry, our method has the advantage of 1. In vitro instead of metabolic labeling circumvents the labeling efficiency, therefore enabling accurate quantitation and improves sensitivity. 2. Metabolic labeling cannot be applied to frozen or FFPE tissue samples or some demanding primary cell lines, while In vitro labeling is broadly applicable. 3. Cyclic-allyl-m^{6}A provides more steric hindrance than cyclic-allyl-A, which translates to higher mutation ratio in RT and improves sensitivity. 4. S-derivatives are easier to synthesize and more stable than Se-derivatives. The reason why m^{6}A-label-seq used Se-derivative was lower efficiency caused by a combination of metabolic labeling and less steric hindrance, as stated before.

2.2 Results and discussion

As shown in the schematics, Mjdim1 uses allyl-SAM in place of SAM and transfers the allyl group to m^{6}A, forming N^6-allyl, N^6-methyladenosine (a^{6}m^{6}A). Subsequent iodine treatment converts a^{6}m^{6}A to cyclic-allyl-m^{6}A, which readout as a mutation during RT (**Figure 2.1a**). We synthesized a 12-mer RNA probe with a GGA/m^{6}ACU consensus motif containing 'A' or 'm^{6}A' (underlined) in the middle (**Figure 2.1b,c**). In the presence of allyl-SAM, MjDim1 exhibited an approximately tenfold preference for m^{6}A over A in a model allyl group transfer reaction (**Figure 2.1d,e**), converting m^{6}A into allyl-modified m^{6}A (N^6-allyl, N^6-methyladenosine or a^{6}m^{6}A) and A into allyl-modified A (N^6-allyl-adenosine or a^{6}A). Human immunodeficiency virus 1 (HIV-1) reverse transcriptase (RTase) reads through synthetic oligonucleotides containing the allyl-labeled

and cyclized adducts with negligible RT stops. To our delight, the HIV-1 RTase generated ~10-fold higher mutation rates at the cyclized a^6m^6A sites (true-positive m^6A sites) than cyclized a^6A sites (unmodified A sites) in almost all sequence contexts tested using model oligonucleotides containing NNXNN (X is either cyclized a^6m^6A or a^6A; **Figure 2.1f**). Therefore, m^6A-SAC-seq directly detects m^6A and exhibits high selectivity toward m^6A over A at two steps of the procedure: (1) the allyl transfer from allylic-SAM catalyzed by MjDim1 is ~tenfold more selective for m^6A than A, and (2) the labeled and cyclized m6A adducts generate higher mutation rates than the corresponding adducts formed from unmodified A. Even with the high selectivity of m^6A-SAC-seq (~100-fold), there is a ~200-fold excess of unmodified A that could produce non-specific background noise. To confirm the specificity of m^6A-SAC-seq, we included a control sample in which RNA is treated with the m^6A demethylase FTO before MjDim1 labeling. FTO treatment removed a large portion of the m^6A in isolated mRNAs from HeLa cells, justifying its use to erase m^6A in the transcriptome as a background control.

We identified more than 10,000 high-confidence m^6A sites with stoichiometrical information in each of these three cell lines with two biological replicates (**Figure 2.2a,b**). m^6A stoichiometry was reproducible within biological replicates (**Figure 2.2c**). Consistent with previous observations[55,126], most m^6A sites are enriched around the stop codon and located in the 3′-untranslated region (3′-UTR) and coding DNA sequence (CDS) regions (**Figure 2.2d,e**), with identified m^6A sites in the frequent m^6A-methylated GGACU and AGACU motifs displaying the highest frequency (**Figure 2.2f**). We noticed that the number of identified sites in the less frequent m^6A-methylated DAACH motifs was low. This is most likely caused by the fact that m^6A-SAC-seq yields low mutation rates for certain Am^6AC motifs. To estimate how many DAACH sites m6A-SAC-seq might have missed, we analyzed DRACH motifs under MeRIP-seq peaks and

compared them with m^{6}A-SAC-seq sites. We could observe seven to nine DRACH motifs per peak, on average, from MeRIP-seq data[37,39,55] in three cell lines that we studied (**Figure 2.2g**). We decided to focus on Gm^{6}AC- or Am^{6}AC-enriched peaks to avoid overestimating DRACH sites under MeRIP-seq peaks that are not methylated and not detected by m^{6}A-SAC-seq. We identified ~72% of these peaks, on average, that were enriched with the Gm^{6}AC motif (**Figure 2.2h**), which is consistent with previous reports[18,148,160]. Based on this analysis, m^{6}A-SAC-seq likely missed ~85% of Am^{6}AC motif occurrences (**Figure 2.2i**). However, the Am^{6}AC sites detected by m^{6}A-SAC-seq were significantly enriched in high-signal peaks (**Figure 2.2j**), indicating that m^{6}A-SAC-seq can still uncover highly modified Am^{6}AC sites. As an orthogonal validation, we applied SELECT[113] and MazF digestion[152,161] to several randomly selected sites, including one m^{6}A negative and four m^{6}A positives using SELECT (**Figure 2.3a**) and one m^{6}A negative and five m^{6}A positives using MazF digestion (**Figure 2.3b**).

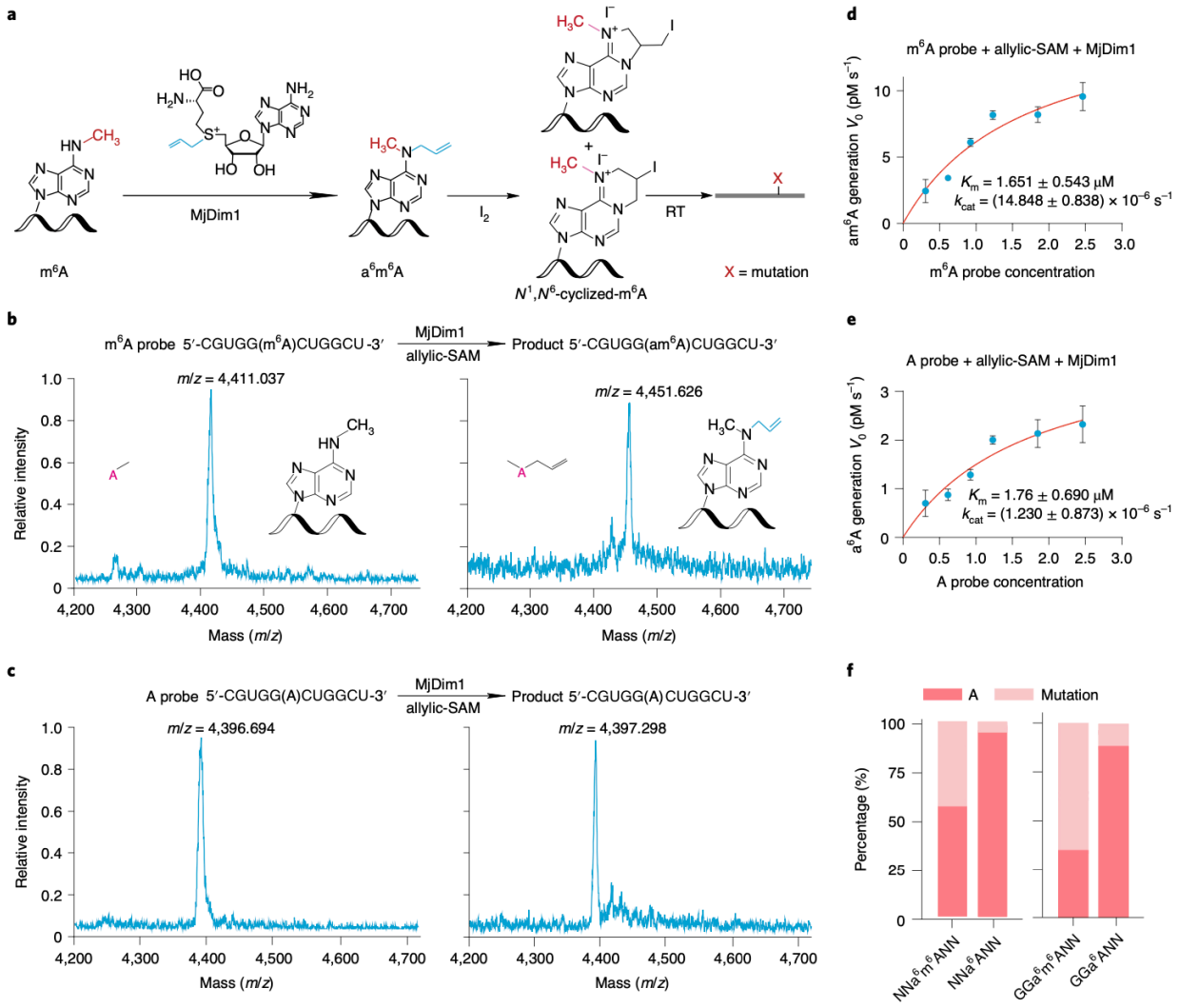


Figure 2.1 m⁶A-SAC-seq strategy and development

a, MjDim1 uses allylic-SAM as a cofactor to label m⁶A to a⁶m⁶A, which undergoes cyclization following I_2 treatment. **b**, An m6A-modified 12-mer RNA probe was treated with MjDim1 and allylic-SAM, followed by matrix-assisted laser desorption ionization (MALDI) characterization. The added molecular weight is that of the allyl group. **c**, An m⁶A-free 12-mer RNA probe was treated with MjDim1 and allylic-SAM, followed by MALDI characterization. No detectable new product appeared. **d**, Michaelis–Menten steady-state kinetics of the MjDim1-catalyzed allyl

transfer to an m^{6}A-containing probe (CGUGGm^{6}ACUGGCU-biotin). Data are represented as mean ± s.e.m. for two biological replicates × two technical replicates. e, Michaelis–Menten steady-state kinetics of the MjDim1-catalyzed allyl transfer to an unmodified control probe. Data are represented as mean ± s.e.m. for two biological replicates × two technical replicates. **f**, Cyclized a^{6}m^{6}A induces higher mutation rates than cyclized a^{6}A in various RNA sequence contexts when using HIV RT. RNA oligonucleotides containing a^{6}A or a^{6}m^{6}A were synthesized by incorporating O^6-phenyl-adenosine phosphoramidite into the designed sequence containing an NNXNN motif (X = a^{6}A or a^{6}m^{6}A).

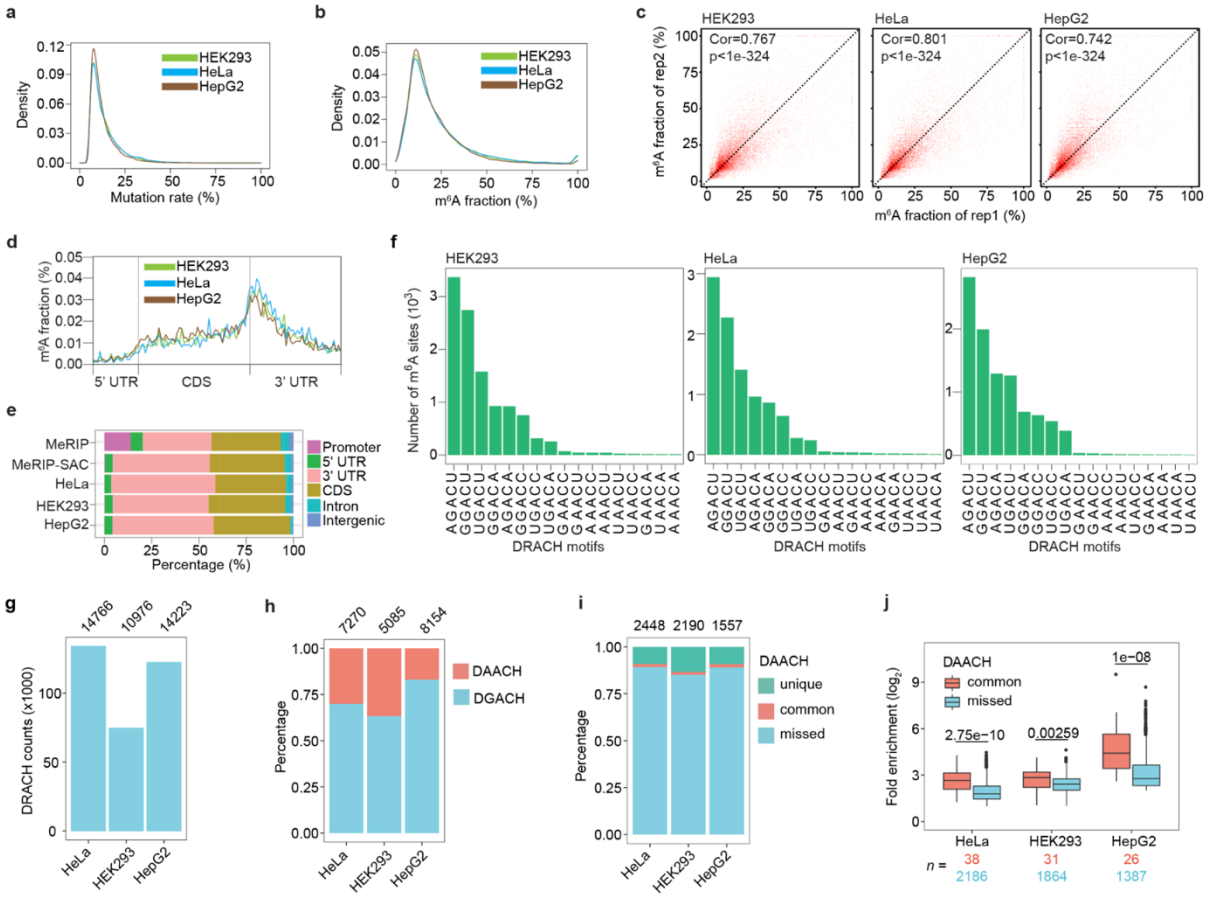
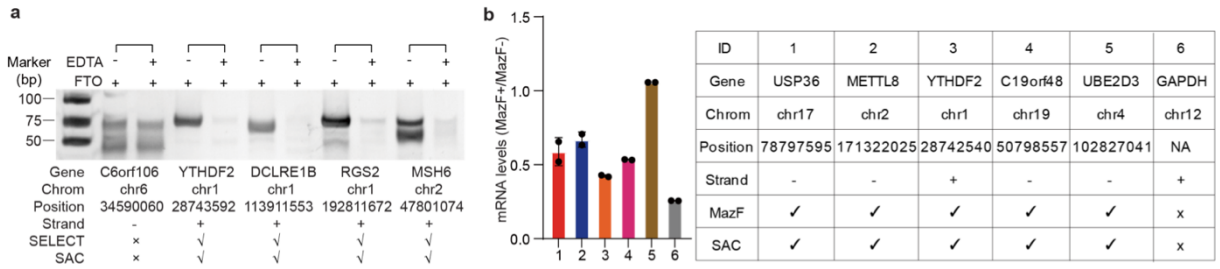


Figure 2.2 m^{6}A-SAC-seq validation

a, Distributions of mutation rates of the identified m^{6}A sites in HeLa, HEK293, and HepG2 RNA samples, respectively. **b**, Distributions of m6A fractions of the identified m6A sites in HeLa, HEK293, and HepG2 RNA samples, calculated using the calibration curves obtained from the spike-in probes. **c**, Scatter plots showing the reproducibility of m^{6}A fractions between two replicates from HeLa, HEK293 and HepG2 cell lines. The Pearson correlation coefficient was used to measure the strength of a linear association between two replicates. The p-value is determined by two-tailed t-test. **d**, A metagene profile depicting m6A fraction along the length of transcripts in HeLa, HEK293, and HepG2 cells, respectively. **e**, The percentage distribution of m^{6}A sites within the 3' UTR, CDS, 5' UTR, and intronic regions. **f**, Distributions of m^{6}A sites in different m6A consensus motifs in transcripts from HeLa, HEK293, and HepG2 cells. **g**, Distribution of counts of DRACH motifs under MeRIP-seq peaks in RNA samples from HeLa1, HEK2932 and HepG23 cells, respectively. Numbers in the top represent counts of total peaks in the corresponding cell line. **h**, Percentages of only DGACH- ($FC_{DGACH/DAACH} > 1.5$) or DAACH- ($FC_{DAACH/DGACH} > 1.5$) enriched peaks with more than one motif but not overlapping with m^{6}A-SAC-seq sites in the three cell lines. Numbers in the top represent counts of total peaks with significant enrichment of DGACH or DAACH motifs in the corresponding cell line. **i**, Percentages of unique, common or missed DAACH motifs from m6A-SAC-seq data compared with MeRIP-seq data in the three cell lines. Numbers in the top represent counts of DAACH motif occurrences (**Figure 2.2, continued**) in the corresponding cell line. **j**, Fold enrichment distribution of MeRIP-seq peaks with common or missed DAACH sites in m^{6}A-SAC-seq. The site number of each group were shown in the plot. P-values were determined using one-tailed Mann-Whitney U-test and then adjusted by BH method. In box plots, lower and upper hinges represent first and third quartiles, the center line represents the median, and whiskers represent ±1.5× the interquartile range.



ID	1	2	3	4	5	6
Gene	USP36	METTL8	YTHDF2	C19orf48	UBE2D3	GAPDH
Chrom	chr17	chr2	chr1	chr19	chr4	chr12
Position	78797595	171322025	28742540	50798557	102827041	NA
Strand	-	-	+	-	-	+
MazF	✓	✓	✓	✓	✓	x
SAC	✓	✓	✓	✓	✓	x

Figure 2.3 Validation of m6A-SAC-seq sites via different methods

a, PAGE showing FTO-assisted SELECT results for validation of selected m^{6}A sites identified by m^{6}A-SAC-seq. m^{6}A marks that are present in the RNA template selectively prohibit DNA polymerase mediated elongation and DNA-ligase-catalyzed nick ligation. FTO erased the m6A and increased the amount of final ligation products. After qPCR-based quantification of FTO$^-$ versus FTO$^+$ treated target templates, PAGE showed stronger product bands for FTO$^+$ treated group containing m6A modification. The experiments of reverse transcription were repeated 3 times independently with similar results and one representative result was displayed. **b**, Validation of five positive sites and one negative site using MazF digestion. MazF was sensitive to m^{6}A modification within RNA ACA motif and could cleave the unmethylated ACA motif, leaving methylated (m^{6}A) CA motifs intact. The RT-qPCR quantitation of MazF$^+$ group versus MazF$^-$ group could reflect the m^{6}A stoichiometry within the ACA motif of the RNA template. Data are presented as mean values +/- SEM, n=2 independent experiments

2.3 Methods

Expression and purification of recombinant MjDim1. The MjDim1 gene was codon optimized and synthesized by Thermo Fisher Scientific and cloned into a pET-His-SUMO vector. T7 Express Competent Escherichia coli (NEB) was transformed with the plasmid and cultured at 37 °C. When the optical density at 600 nm (OD600) reached 1, cells were cooled to 16 °C, IPTG was added to a final concentration of 0.1 mM for inducible expression, and cells were cultured at 16 °C for an additional 18 h. Cells were collected and lysed by EmulsiFlex-C3 (Avestin) in lysis buffer (50 mM Tris-HCl (pH 7.5), 300 mM NaCl). The soluble recombinant protein was purified using a nickel resin column (GE Healthcare) washed with washing buffer (50 mM Tris-HCl (pH 7.5), 300 mM NaCl, 5 mM imidazole). Ulp1 was added in the resin to cleave the SUMO tag on the column at 4 °C for 16 h. The enzyme was eluted in elution buffer (20 mM Tris-HCl (pH 7.5), 150 mM NaCl, 20 mM imidazole) and subjected to anion-exchange chromatography (Source15-Q 10/10, GE Healthcare) on an AKTA Purifier 10 system (GE Healthcare) to getrid of the RNA and DNA bound with the enzyme. Flow-through was collected for the second round of cation-exchange chromatography (Source15-S 10/10, GE Healthcare), and the fractions coming out between conductivity 20 and 35 were collected and concentrated to ~1.6 mM. Glycerol was added in the enzyme to a final concentration of 30% and stored at –80 °C for future use.

Synthesis of allylic-SAM analog. S-Adenosyl-l-homocysteine (20 mg, 0.05 mmol, 1 equiv.) was dissolved in formic and acetic acids (1:1, 2 ml). Allyl bromide (425 µl, 5 mmol, 100 equiv.) and $AgClO_4$ (10.4 mg, 0.05 mmol, 1 equiv.) were added and stirred at ambient temperature (22 °C) for 8 h. The reaction was quenched with 20 ml of 0.01% trifluoroacetic acid (TFA; vol/vol) in water. The aqueous phase was washed three times with diethyl ether (3 × 10 ml) and then passed through a 0.2-µm syringe filter. The crude mixture of allyl-SAM was purified using a preparative reversed-phase high-performance liquid chromatography (RP-HPLC) column (XBridge Prep C18

5 μm OBD 19 × 150 mm). A diastereomeric mixture of the allylic-SAM analog was collected, concentrated and lyophilized. The resultant compounds were redissolved in water containing 0.01% TFA (vol/vol), aliquoted and stored at –80 °C before use.

Biochemical assay for a^6m^6A methyltransferase activity in vitro. The in vitro methyltransferase activity assay was performed in a standard 20-μl reaction mixture containing the following components: 50 ng to 1 μg of RNA probe or 30 ng of mRNA, 10 nmol of fresh recombinant MjDim1 enzymes, 3 mM allylic-SAM, 40 mM HEPES (pH 8.0), 40 mM NH_4Cl, 4 mM $MgCl_2$ and 1 U/μl SUPERase In RNase Inhibitor (Thermo Fisher Scientific). For the RNA probe, the reaction was incubated at 50 °C for 1 h. For mRNA or other biological samples, RNA fragments were ligated with biotin-modified 3′-adapter and bound with Dynabeads MyOne Streptavidin C1 (Thermo Fisher Scientific). The reaction was performed on the solid phase using the following procedure: 50 °C for 1 h and change new reaction system for three rounds to efficiently label m6A sites with an allyl group.

MALDI-time-of-flight (MALDI-TOF) and HPLC. The RNA reaction products were purified using Dynabeads MyOne Streptavidin C1 (Thermo Fisher Scientific) and eluted by heating at 98 °C for 10 min. One microliter of the supernatant was then mixed with an equal amount of MALDI matrix, which was composed of a 9:1 (vol/vol) ratio of 2′,4′,6′-trihydroxy acetophenone (THAP; 10 mg/ml in 50% acetonitrile and water):diammonium citrate (50 mg/ml in water). The mixture was then spotted on a MALDI sample plate, dried under vacuum and analyzed by a Bruker Ultraflextreme MALDI-TOF-TOF mass spectrometer in a reflector, positive mode. The HPLC profiles were acquired using Waters e2695 equipment.

Quantification of am^{6}A in RNA by LC–MS/MS. RNA oligonucleotides or mRNAs were digested into nucleosides, and the amount of am^{6}A was measured by using Agilent 6460 Triple Quad MS–MS with a 1290 UHPLC supplied with a ZORBAX Eclipse XDB-C18 column (UHPLC–QQQ–MS/MS) and calculated based on the standard curve generated by pure standards. For each sample, RNA was digested by using 1 U of nuclease P1 (Wako) in a 25-μl reaction containing 10 mM ammonium acetate at 37 °C for 16 h. Then, 1 μl of FastAP thermosensitive alkaline phosphatase and 3 μl of 10× FastAP buffer (Thermo Scientific) was added, and the reaction was incubated at 37 °C for 2 h. Samples were then filtered using a 0.22-μm filter (Millipore) and injected into LC–MS/MS. The nucleosides were quantified by using the nucleoside-to-base ion mass transitions of 282 to 150 (m^{6}A), 323 to 191 (am^{6}A), 268 to 136 (A) and 284 to 152 (G). Quantification was performed in comparison to the standard curve obtained from pure nucleoside standards run on the same batch of samples. The ratio of m^{6}A to G was calculated based on the calibrated concentrations.

Steady-state kinetics of MjDim1-catalyzed am6A and a6A modifications. MjDim1 (168 μM) was used for the kinetic measurements using both MALDI_ Probe_m6A (CGUGGm^{6}ACUGGCU-biotin) and MALDI_Probe_A (CGUGGACUGGCU-biotin) with allyl-SAM as a cofactor. Considering that allyl-SAM is not an optimal cofactor for the enzyme, we chose relatively long reaction time points as linear intervals: 0, 1, 2.5, 5, 7.5 and 10 min. The reaction products were analyzed by LC–MS/MS. Relative amounts of A, m^{6}A, a^{6}A and ma^{6}A and G were calculated for each measurement according to standard curves. The amounts of the adenosine derivatives are normalized to the amount of G nucleotide. Error bars indicate s.d. for duplicate experiments from two independent assays.

m6A-SAC-seq scheme. m6A-SAC-seq experiments require parallel construction of three libraries: (1) the input library, where the RNA is subjected to standard library construction without any treatment by MjDim1 or I and is used as a reference to call mutations; (2) the experimental library (FTO^{-}), where RNA is treated with the MjDim1 enzyme with cofactor allylic-SAM to convert m^6A sites into allylic-m^6A by MjDim1 and (3) the background noise group (FTO^{+}), where m^6A sites were erased by FTO first, followed by allylic labeling. Both experimental groups and the background noise group were treated with I_2 for the cyclization reaction, followed by reverse transcription with HIV RT (Worthington Biochemical).

RNA poly(A) tail elimination, fragmentation and 3′-adapter ligation. 30 to 100 ng of $poly(A)^{+}$ RNA or $ribo^{-}$ RNA (300 ng to 1 μg of total RNA) were annealed with oligo(dT), digested with RNase H (NEB) and DNase I (NEB) to remove oligo(dT) and purified by RNA Clean & Concentrator kits (Zymo Research). The purified RNA was fragmented by sonication using Bioruptor (Diagenode; 30 cycles of 30 s on/30 s off to obtain ~150-nt fragments), followed by PNK enzyme (NEB) treatment at 37 °C for 30 min to expose the 3′-hydroxyl group. Calibration spike-in mix (0.6%) was added in the reaction and subjected to 3′-adapter ligation with T4 RNA ligase 2, truncated KQ (NEB). The excessive RNA adaptor was digested by adding 1 μl of 5′ Deadenylase (NEB) into the ligation mix followed by incubation at 30°C for 1h. Then, 1μl of RecJf (NEB) was added and incubated at 37°C for 1h. One microliter of RT primer (50 μM) was added with the following parameters: anneal 75°C for 5min, 37°C for 15min and 25°C for 15min.

m^6A site labeling and reverse transcription. Fifteen microliters of dynabeads C1 (Thermo Fisher Scientific) was added to the reaction to purify the 3′-adapter-ligated RNA. The beads were washed, resuspended in 6 μl of waterand denatured at 70 °C for 30 s and cooled on ice.

m6A enzymatic labeling was performed on beads. Two microliters of 10× buffer (400 mM HEPES (pH 8.0), 400 mM NH4Cl, 40 mM $MgCl_2$,), 2 μl of SUPERase In RNase Inhibitor (Thermo Fisher Scientific), 6 μl of allylic-SAM and 4 μl of MjDim1 enzyme (1.6 mM) were added in the reaction and incubated at 50 °C for 1 h. The supernatant was removed, and 4 μl of water, 1 μl of 10× buffer, 1 μl of RNase inhibitor, 2 μl of allylic-SAM and 2 μl of enzyme were added in the reaction and incubated at 50 °C for 20 min (this labeling step was repeated six times). Beads were washed and resuspended in 25 μl of water. One microliter of 125 mM I_2 was added and mixed thoroughly and kept in the dark at room temperature for 1 h, then 1 μl of 40 mM $Na_2S_2SO_3$ was added to quench I_2. Beads were washed and resuspended in 9 μl of water. Then, 2 μl of 10× RT buffer (SuperScript III First-Strand Synthesis SuperMix, Thermo Fisher Scientific), 2 μl of 10 mM dNTP, 2 μl of 25 mM $MgCl_2$, 1.25μl of 0.1M DTT, 2μl of RNaseOUT and 2 μl of HIV RT enzyme (Worthington Biochemical) were added in the tube to perform reverse transcription at 37 °C for 3 h. For input RT, 1 h with 1 μl enzyme was sufficient. Beads were washed and resuspended in 8 μl of water.

cDNA 3′-adapter ligation, library construction, purification and sequencing. One microliter of RNase H buffer and 1 μl of RNase H were added into the resuspended reverse transcription product and placed in a thermocycler (Bio-Rad) at 37 °C for 30 min. Beads were washed and resuspended in 50 μl of water. cDNA was eluted by boiling the beads at 95 °C for 10 min, purified using a DNA Clean & Concentrator kit (Zymo Research) to remove short adapters and eluted into 10 μl of water. Two microliters of 10× T4 RNA ligase buffer, 2 μl of 10 mM ATP, 10 μl of 50% PEG8000, 1 μl of cDNA_3′ adapter (50 μM; Phos-NNNNNGATCGTCGGACTGTAGAACTCTGAAC/3SpC3/) and 1 μl of T4 RNA ligase 1 were added into the eluted cDNA, and the ligation was performed at 25 °C overnight. The reaction was purified using a DNA Clean & Concentrator kit (Zymo Research) and eluted with 21 μl of water.

One microliter of supernatant was used for quantitative real-time PCR (qPCR) testing, and the remaining 15 μl was used for library construction. NEBNext Ultra II Q5 Master Mix and NEBNext adaptors were used for library amplification. Amplified libraries were purified using 0.8× Ampure beads. The purified libraries were sent for next-generation deep sequencing. The libraries were sequenced on an Illumina HiSeq X Ten with paired-end 2 × 150 bp read length.

FTO demethylation. The fragmented RNA materials were ligated to the biotin-modified 3′-adapter (rAppNN NNN ATC ACG AG ATC GGA AGA GCA CAC GTC /iBiodT//3SpC3/) and bound with Dynabeads MyOne Streptavidin C1 (Thermo Fisher Scientific). RNA on beads was denatured at 70 °C for 30 s and quickly put on ice to quench secondary structure formation. Demethylation was performed on the solid phase in a 50-μl reaction system of 50 mM HEPES buffer (pH 7.0), 75 μM $(NH_4)_2Fe(SO_4)_2$, 2 mM L-ascorbic acid, 0.3 mM α-ketoglutarate, 2 U/μl RNase inhibitor and 0.2 nmol of FTO. The demethylated RNA was washed and subjected to m^{6}A-SAC-seq.

m^{6}A-immunoprecipitation (m^{6}A-IP). HeLa poly(A)$^+$ RNA was purified with a Dynabeads mRNA DIRECT Purification kit (Thermo Fisher Scientific) and sonicated to ~150-nt fragments using Bioruptor (Diagenode) with 30 cycles of 30 s on/30 s off. An EpiMark N^6-Methyladenosine Enrichment kit (NEB) was used to enrich m^{6}A-containing RNA fragments. m^{6}A-IP RNA fragments (120 ng) were prepared, and a 0.6% calibration spike-in mix was added. Half of the materials were subjected to m^{6}A-SAC-seq (demethylase$^-$ group), and the other half was treated with FTO followed by m^{6}A-SAC-seq (demethylase$^+$ group).

Quantitation of m6A stoichiometry by standard curve construction using model oligonucleotides. Four pairs of synthetic RNA oligonucleotides containing either m6A or A were

used to examine the efficiency of m^{6}A-SAC-seq in causing misincorporation during reverse transcription in a quantitative manner.

Procedures of SELECT and MazF for m6A-SAC-seq validation. The SELCET procedures were described previously[113]. Total RNA was mixed with 40 nM Up Primer, 40 nM Down Primer and 5 μM dNTP in 17 μl of 1× CutSmart buffer (50 mM potassium acetate, 20 mM Tris-acetic acid, 10 mM magnesium acetate, 100 μg/ml BSA, pH 7.9, at 25 °C). The RNA and primers were annealed by incubating the mixture at a temperature gradient: 90 °C for 1 min, 80 °C for 1 min, 70 °C for 1 min, 60 °C for 1 min, 50 °C for 1 min and 40 °C for 6 min. Subsequently, 3 μl of a mixture containing 0.01 U Bst 2.0 DNA polymerase, 0.5 U SplintR ligase and 10 nmol of ATP was added in the former mixture to a final volume of 20 μl. The final reaction mixture was incubated at 40 °C for 20 min, denatured at 80 °C for 20 min and kept at 4 °C. Afterward, qPCR was performed in high-performance real-time PCR with LightCycler Systems (Roche). The 20-μl qPCR reaction was composed of 2× qPCR SYBR Green Master Mix (Roche), 200 nM qPCRF primer, 200 nM qPCRR primer, 2 μl of the final reaction mixture and double-distilled water. qPCR was run using the following conditions: 95 °C for 5 min, 95 °C for 10 s and 60°C for 35s for 40 cycles, 95°C for 15s, 60°C for 1min, 95°C for 15s and hold at 4 °C. PCR products were analyzed by PAGE. The MazF procedures were reported previously. Total RNA was incubated with 2.5 U of MazF (mRNA interferase MazF; Takara Bio, 2415A) in the 20-μl reaction mixture of MazF buffer (40 mM sodium phosphate (pH 7.5) and 0.01% Tween 20) at 37 °C for 30 min. Total RNA with or without MazF treatment was subjected to gene-specific quantitative PCR with reverse transcription (RT–qPCR)

Mutation calling and identification of m^{6}A sites. First, to maximize the read coverage, bam files from input and FTO$^+$ data sets were merged into a single file by cell lines and time points, respectively. Merged bam files and those bam files from FTO$^-$ data sets were split by strands and piled up using the samtools subcommand mpileup. Second, mutation calling was performed using VarScan v2.3 subcommand somatic in two pair-wise comparisons: (1) FTO– versus input and (2) FTO– versus FTO+. For each comparison (X versus Y), mutation sites were kept if (1) the reference position was adenine (A), (2) the P value was <0.1, (3) the mutation frequency was X – Y > 5% (FTO$^-$ versus input) or X – Y > 2% (FTO– versus FTO$^+$), (4) coverage was more than five reads, (5) the 5-mer context was DRACH and (6) only the common sites from two or three biological replicates of either FTO$^-$ or FTO$^+$ were kept. Finally, sites in both types of comparisons (FTO$^-$ versus input and FTO$^-$ versus FTO$^+$) were identified as m^{6}A sites.

Spike-in analysis and calibration curve for m6A stoichiometry estimation. Spike-in sequences were directly extracted from preprocessed m^{6}A-SAC-seq data sets, and observed mutation rates of the target A were calculated by motifs. To correlate observed mutation rates and m6A fractions in each motif to fit the values measured from the spike-in samples, linear regression models were used: y=ax+b, where y is the observed mutation rate, and x is the m6A fraction. We fit spike-in mutation rates and m6A fractions using the above model in R 3.5.1. Thus, the best-fit calibration curve in each DRACH motif was used for the estimation of m6A stoichiometry in m^{6}A-SAC-seq data sets.

Chapter 3 Optimization of m^{6}A-SAC-seq

3.1 Introduction

Further optimizations of the protocol (**Figure 3.1)** were introduced after the initial publication[84]. These include: 1. The elimination of the FTO-treated control. MjDim1 has a minimal background labeling on A residues (**Figure 3.2a**). Considering the intrinsic abundance of A, m^{6}A-depleted controls were previously introduced using FTO-treated RNA samples. However, this not only increases the requirement of RNA sample amount, but also results in under-estimation of m^{6}A sites that could be partially demethylated by FTO. The incomplete demethylation could be explained as inaccessibility of certain secondary structures or sequence preference of FTO. The problem is solved by increasing the stringency of mutation calling, in which the filtering ratio for called m^{6}A sites is set higher than the highest possible mismatch ratio of A residues on spike-in probes in treated samples. We now set the cutoff at 5% with 3 mismatches in 20× coverage. This is sufficient to rule out all false positives from background labeling by MjDim1. An untreated input library is still required using 1/3 of the starting material, which serves as a control for somatic mutations (input mismatch ratio ≥2.5% will not be considered as sites, **Figure 3.2b**) and an RNA-seq library for gene expression and isoform profiling.

2. Introduction of template-switch inhibitor Actinomycin D in the RT reaction. HIV-1 RTase is known to possess terminal transferase activity[162]. The randomly appended sequences to the 3' end of cDNA could anneal with random fragments in the transcriptome, resulting in template-switch RT and apparent "ligation" of two genetically non-adjacent transcripts. This could be solved simply by the addition of Actinomycin D, which according to previous studies, inhibits

template-switch by preventing the formation of annealed intermediates[163]. This optimization improves the mapping ratio to over 80% of the reads mapped to transcriptome (**Figure 3.2c**).

3. Optimization of ligation. The ssDNA ligation efficiency is greatly improved with the addition of DMSO and hexaamminecobalt chloride ($Co(NH_3)_6Cl_3$). DMSO is known to reduce secondary structures and thus improve single-strand ligation efficiency[164]. $Co(NH_3)_6Cl_3$ was reported to improve the dsDNA ligation efficiency of T4 DNA ligase[165], but its application in ssDNA ligation is not well-documented. We found that combining these two additives could improve the ligation efficiency by 2^6 to 2^8 folds, as estimated by product concentration increases after 12 rounds of PCR. This optimization increases the remaining reads after de-duplication to approximately 60% of the raw reads (**Figure 3.2c**).

4. Optimization of size-selection. The improved ligation efficiency mentioned above enables size-selection for longer fragments. This is because shorter fragments intrinsically have higher ligation efficiency due to the lack of secondary structures[166] but contribute less to mutation calling since they cannot be uniquely mapped. Size-selection of longer fragments will improve the mapping ratio and reduce adapter-dimer contamination (**Figure 3.2c**). The purification is introduced before and after the ssDNA ligation, and after the final PCR. The resulting library shows a single peak around 260-300 bp containing an average insert of 150 bp, which is suitable for paired-end sequencing (**Figure 3.2d**).

5. Streamlined protocol. The labeling process of MjDim1 is simplified to two rounds in two hours. Unnecessary purification steps have been omitted. All purification steps are also changed to beads-based for simplicity in multiplexing of samples and potential automation in the future.

3.2 Results and discussion

These optimizations greatly improve the data quality of m^{6}A-SAC-seq, thus generating more detectable sites and higher reproducibility of the sites. Two technical replicates of 50 ng of HeLa poly-A RNA samples exhibit a correlation coefficient of 0.902, enabling the detection of 129,263 raw sites upon initial filtering (input < 2.5%, treated ≥ 5% with 3 mismatches in 20× coverage. **Figure 3.2e**). As low as 2 ng of HeLa poly-A RNA could still generate comparable results. The number of sites identified from 2 ng mRNA libraries is proportional to the effective reads as compared to 50 ng mRNA libraries, indicating that neither sequencing depth is saturated thus more sites could still be discovered with additional reads (**Figure 3.2f**). Despite the differences in sample amount and sequencing depth, the 2 ng mRNA libraries could still re-discover 40,409 sites from the 129,263 sites identified in 50 ng mRNA libraries, showing strong reproducibility (r = 0.942. **Figure 3.2g**). The mutation ratios are then converted into modification fractions based on calibration curves. The sites could be further selected for high modification fractions, retaining 31,233-71,547 sites even if the threshold is set at 20% (**Figure 3.2h**).

The validity of these sites is supported by the strong transcriptome-dependent enrichment of DRACH motifs (**Figure 3.3a-b)**, consistent with previous studies[32,33]. The sites are also enriched in 3'UTR near stop codon region (**Figure 3.3c-d)**, as was previously reported[16,17]. Notably, the sites are enriched upon entering the last exon (**Figure 3.3e)**. This was recently revealed to be an exclusion effect of exon junction complex (EJC)[19,115,136], showing the strong site-specificity of m^{6}A-SAC-seq data and implying additional mechanisms regulating m^{6}A deposition.

Together these results show that: 1. m^{6}A-SAC-seq could identify a large number of m^{6}A sites even with stringent criteria. 2. mRNA samples ranging from 2 ng to 50 ng could be used to

generate data of comparable quality. The data quality could be improved by increasing the sequencing depth, when limited samples or low quality RNAs are to be used. 3. The method does not require metabolic labeling or transfection of the cells, nor does it need intact RNA. This also enables RNA m^{6}A profiling from fresh or frozen tissue samples as well as FFPE samples.

Comparison with various alternative methods is summarized in **Table 3.1**. Commonly used transcriptome wide m^{6}A profiling methods could be categorized using the following characteristics: direct or indirect, site specificity, antibody usage, quantitativeness, and compatibility with frozen or FFPE samples. miCLIP-seq and DART-seq use indirect inference of m^{6}A sites, from crosslink-induced mutation or truncation (typically occurs on U or C[148]) and adjacent C[155], respectively. This would affect the precision of identified sites, especially the characterization of closely deposited m^{6}A clusters. m^{6}A-seq or MeRIP-seq and m^{6}A-SEAL[149] identify m^{6}A as peaks rather than specific sites at a resolution of 100-200 nt. Statistically there are more than one m^{6}A within each MeRIP-seq peak[21], thus again preventing detailed characterization of closely deposited m^{6}A clusters. MeRIP-seq, miCLIP and m^{6}A-label-seq are also antibody-based. The enrichment of m^{6}A-bearing transcripts reduces the starting material for library construction by approximately 100 folds. This would impair the performance of the ensuing RNA-seq, and in turn require more RNA samples before immunoprecipitation. Typically, μg-scale of RNA is required, as compared to ng-scale in m^{6}A-SAC-seq and DART-seq. Antibodies could potentially recognize non-m^{6}A structures and induce false positives, which is suggested by a recent analysis[144]. All of these methods do not provide stoichiometry information and approaches using transfection or metabolic labeling could not be applied to frozen samples, patient samples, FFPE samples and many other primary-cell applications.

m^{6}A-SAC-seq has one of the lowest sample requirements (on-par with DART-seq) and identifies the largest number of confident sites by far. We compared the 71,547 identified sites with >20% mutation fraction from the two 50 ng-scale HeLa mRNA libraries with two independent MeRIP-seq data[37,167], which were also performed on HeLa mRNA (**Figure 3.3f**). 66.9% and 68.4% sites of respective MeRIP-seq peaks overlap with sites revealed from our data. Note that these two MeRIP-seq datasets also only dispplay 71.8% to 74.9% overlap between themselves. As is discussed above, our sites also show similar motif and distribution profile to MeRIP-seq datasets. Together these demonstrate that m^{6}A**-**SAC-seq could reproduce the results from the standard antibody-based m^{6}A profiling, but in a site-specific and quantitative manner.

Compared to other site-specific methodologies, our data covers 35.5% of the sites from one of the miCLIP-seq studies[92]. The previously mentioned two MeRIP-seq datasets covers 49.4% and 71.8% of the miCLIP-seq sites, respectively, while m^{6}A-REF-seq or MAZTER-seq covers only 1.6% of the latter. This could be explained as 1. miCLIP-seq correctly locates the inferred sites within MeRIP-seq peaks yet deviates from the precise location identified by site-specific methods. 2. miCLIP-seq may share same antibody-induced false positives with MeRIP-seq, because it is also antibody based. m^{6}A-REF-seq or MAZTER-seq show a limited overlap of 18.5% with our dataset. This is comparable with the overlapping ratio of the same m^{6}A-REF-seq data with the two previously mentioned MeRIP-seq datasets. This could be explained by the bias towards ACA motif of the MazF-based methods, which only represents 16-25% of all m^{6}A sites[153]. It is worth noting that the m^{6}A-REF-seq or MAZTER-seq results also contain many low modification-fraction sites, which are excluded from the 71,547 m^{6}A-SAC-seq sites.

As is described in previous sections, allyl-SAM corresponding to 400-600 reactions could be prepared in 3 days, while MjDim1 corresponding to 100-150 reactions could be prepared in 6 days. When quality control is performed on probes, the prepared MjDim1 should have over 60% of conversion within an hour of labeling. Up to 48 libraries, corresponding to 24 technical replicates or 12 biological replicates could be prepared within 4 days, excluding the time of sample preparation. The libraries should take no more than 14 cycles of PCR to obtain 3 nM of concentration after purification. The purified libraries should contain a single peak centered at 260-300 bp range with no significant contamination of <150 bp fragments.

In the sample data, we prepared two technical replicates of SAC-seq libraries form 50 ng each of HeLa mRNA. 137 and 142 million paired end reads (respectively, omitted hereafter) were obtained from sequencing. After filtering low-quality reads and adapters, 117 and 120 million reads could be mapped to human genome GRCh38, among which 111 and 114 million were uniquely mapped. After de-duplication, 81 and 84 million reads remained. 129,263 filtered mutations were detected from these reads, using the filtering criteria described in step 99. Modification fractions could be calculated using the standard curves deducted from spike-in data. Finally, 105,320 or 71,547 high confidence sites could be selected using 10% or 20% mutation as threshold, respectively.

The sample data is from two technical replicates of SAC-seq 2 ng of HeLa mRNA, similar data processing was performed on 84 and 87 million raw reads where 71 and 74 million mapped to human genomes with 68 and 71 million being unique. 59,399 filtered sites were identified from 32 and 32 million de-duplicated reads, among which 49,235 sites had mutation fractions over 10% and 31,233 sites over 20%.

The motif enrichment and metagene profile of the identified sites could also be checked for the patterns described in previous sections (**Figure 3.3b-c**). 77.6% of the identified sites follow the consensus DRACH motifs. The sites were distributed along CDS region and 3'UTR and enriched near stop-codon.

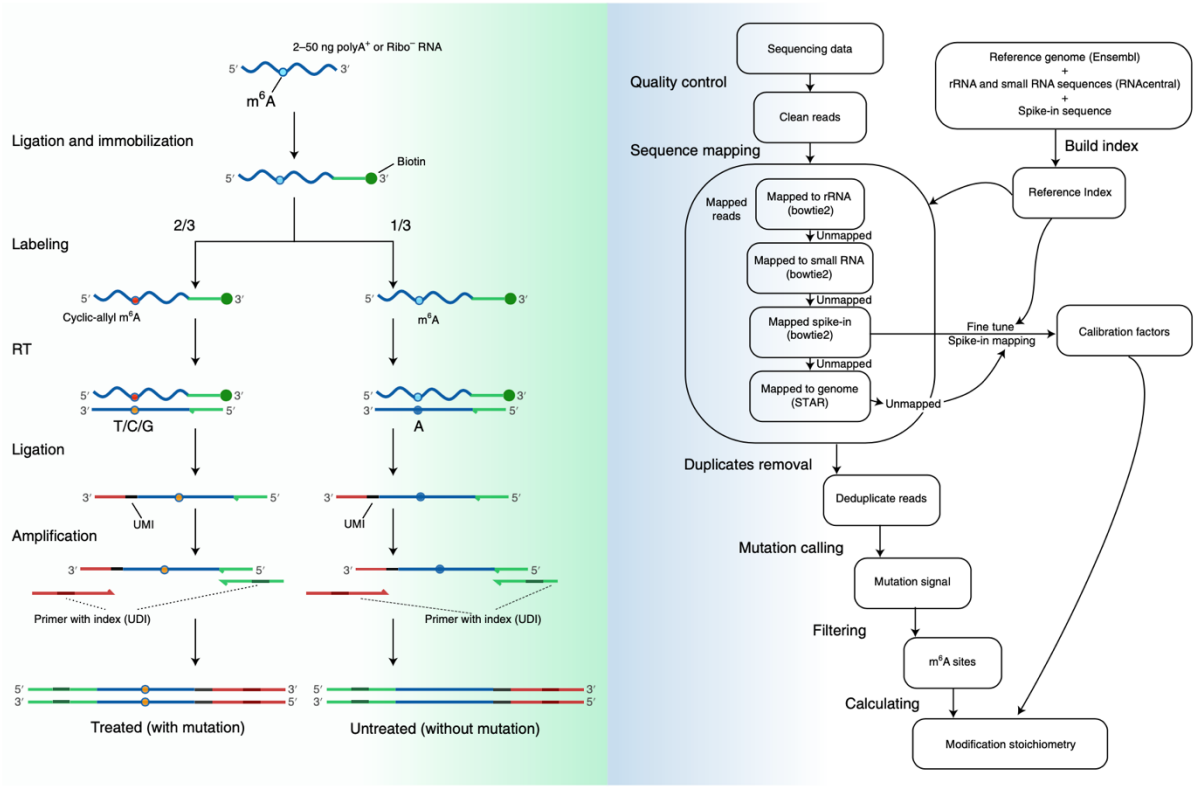


Figure 3.1 Overview of the m⁶A-SAC-seq protocol

a, Schematic representation of library preparation steps. 2-50 ng of poly-A enriched or ribosome RNA-depleted RNAs were fragmented and ligated, then divided in a 2:1 ratio. 2/3 of the starting materials are labeled by MjDim1, while the remaining 1/3 serve as the untreated control. After RT the cyclic allyl m⁶A sites (red dot) are converted to mismatches (orange dot), while unconverted m⁶A sites (cyan dot) in the control group are read as A (blue dot). **b**, Schematic representation of data analysis pipeline.

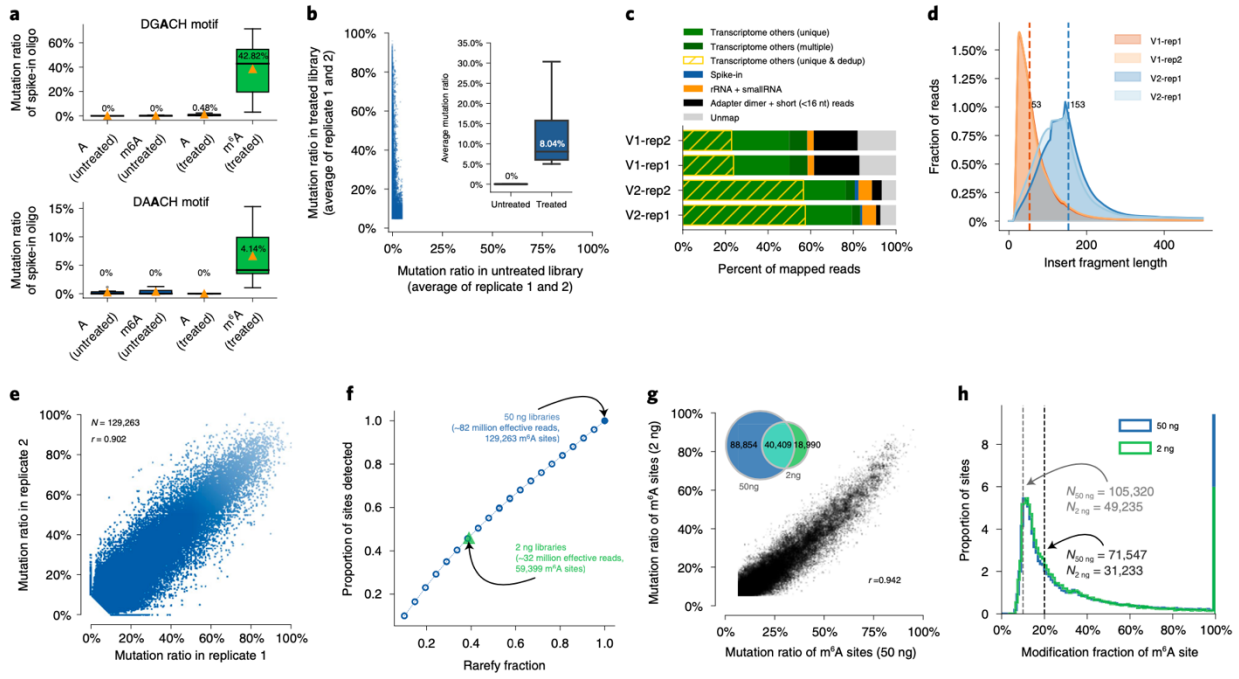


Figure 3.2 m⁶A-SAC-seq could reproducibly identify a large number of m6A sites

a, Mutation ratio of A and m⁶A sites on spike-in probes, treated or untreated. Although AAC motifs show lower mutation ratio after treatment compared to GAC, they also have a much lower background. Median is labelled and marked as black in the box. Orange triangle shows the average mutation ratio. **b**, Average mutation ratio of two HeLa mRNA replicates for treated or untreated libraries. Each dot represents a filtered site. **c**, The proportion of mapped reads. The optimized protocol (V2) shows less adapter dimer (black bar), higher mapping ratio (light and dark green bar) and more unique reads after de-duplication (yellow diagonal shades) compared to the previous protocol (V1). **d**, Fragment length distribution of mapped reads. **e**, Mutations called after initial filtering from two technical replicates. In 20× of coverage, each site has a treated mismatch ratio ≥5%, mutation frequency ≥3, input mismatch ratio <2.5% (or its mutation frequency ≤1). N, the number of sites called. r, Pearson correlation coefficient. **f**, Random subsampling of the data of

libraries prepared from 50 ng mRNA and detected m^{6}A sites using some cutoff. Results for 2 ng libraries match the simulated data. **g**, Correlation of sites called from of the libraries prepared from 50 ng or 2 ng of HeLa mRNA. Inset plot shows the number of overlapping sites. **h**, Density distribution of the mutation fraction of identified sites. Threshold could be placed on either of the two dashed lines, resulting in the corresponding number of filtered sites. Grey dashed line (left): 10%, Black dashed line (right): 20%.

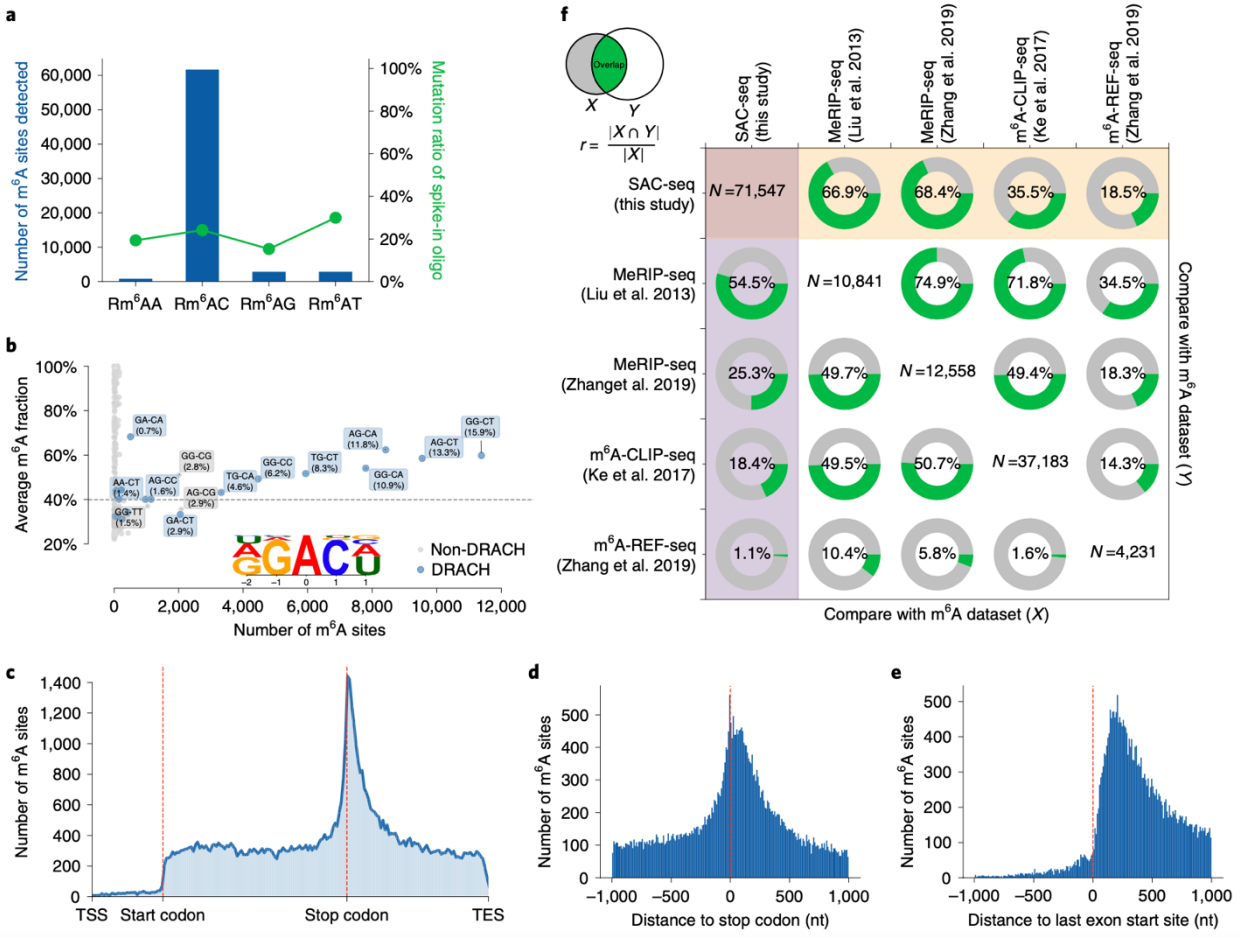


Figure 3.3 Quality control of m^{6}A-SAC-seq datasets

a, Motif enrichment of the identified m^{6}A sites. Note the difference between the internal sites (blue) and spike-in probes (green). The RAC motif has no bias on the spike-in probe but is significantly

enriched on the internal sites. **b**, m^{6}A fraction and site-number distribution of all motifs. **c**, Metagene profile of the identified sites. **d-e**, Distribution of the sites when centered around stop codon or the entry of last exon. **f**, Comparison of the m^{6}A-SAC-seq dataset with results from other methods. Ratio is expressed as the number of overlapping sites between method X and Y, divided by the site number of method X. Diagonal numbers show the number of sites for each method.

Table 3.1 Technical comparisons of commonly used whole transcriptome m^{6}A profiling methods

Method	m^{6}A-seq[16] and MeRIP-seq[8,17,36]	m^{6}A-SEAL[35]	miCLIP-seq[18,37]	DART-seq[19]	m^{6}A-REF-seq[20] and MAZTER-seq[21]	m^{6}A-label-seq[22]	m^{6}A-SAC-seq[23]
Site-specific	No	No	Yes	Yes	Yes	Yes	Yes
Quantitative	No	No	No	No[a]	Yes	No	Yes
Covering all motifs	Yes	Yes	Yes	No	No	Yes	Yes

Directly detecting m^{6}A sites	No	No	No[b]	No[b]	No[c]	Yes	Yes
Frozen sample compatible	Yes	Yes	Yes	No[d]	Yes	No[e]	Yes
Antibody-free	No	Yes	No	Yes	Yes	No	Yes
Starting material	2-400 μg mRNA	5 μg mRNA	20 μg mRNA	10 ng-1 μg total RNA[f]	100 ng mRNA	5 μg total RNA	2-50 ng mRNA
Identified sites	10,841-12,558	8,605	21,494-37,183	12,672	4,231-17,007	2,479	71,547

[a] Can estimate the modification fraction based on C to U mutation ratio, which is indirect and inaccurate. [b] Mutate flanking U (miCLIP-seq) or C (DART-seq) sites, thereby inferring the position of m^{6}A sites. [c] Cut at A sites, thereby inferring the fraction of m^{6}A modification. [d] Need to express the APOBEC1-YTH fusion protein in cells. [e] Need to culture cells in Se-allyl-L-selenohomocysteine medium for 16 h. [f] DART-seq protocol has been applied to single cell m^{6}A profiling in a recent work[42].

3.3 Methods

Materials

Reagents

- S-(5′-Adenosyl)-L-homocysteine ($C_{14}H_{20}N_6O_5S$, Sigma-Aldrich, cat. no. A9384)
- Silver perchlorate ($AgClO_4$, Sigma-Aldrich, cat. no. 674583) **!CAUTION** Silver perchlorate is oxidative and toxic. Handle with appropriate PPE and keep away from fire.
- Allyl bromide (CH_2=$CHCH_2Br$, Sigma-Aldrich, cat. no. A29585) **!CAUTION** Allyl bromide is highly toxic. Handle with appropriate PPE in a fume hood.
- Acetic acid (CH_3CO_2H, Sigma-Aldrich, cat. no. 695092)
- Formic acid (HCOOH, Sigma-Aldrich, cat. no. 695076) **!CAUTION** Formic acid is highly volatile and the vapor is corrosive. Handle with appropriate PPE in a fume hood.
- Trifluoroacetic acid (CF_3CO_2H, Sigma-Aldrich, cat. no. T6508) **!CAUTION** Trifluoroacetic acid is highly volatile and the vapor is highly corrosive. Handle with appropriate PPE in a fume hood.
- Diethyl ether ($(CH_3CH_2)_2O$, Sigma-Aldrich, cat. no. 346136) **!CAUTION** Diethyl ether is highly volatile and the vapor is highly flammable. Handle with appropriate PPE in a fume hood and keep away from fire.
- DNase I, from bovine pancreas (Roche, cat. No. 11284932001)
- ULP1 (previously purified in lab. Commercial SUMO proteases with His-tag will also work)
- Nuclease P1 (New England BioLabs, cat. no. M0660S)
- FastAP Thermosensitive Alkaline Phosphatase (1 U/μL) (Thermo Fisher Scientific, cat. no. EF0651)
- 10X FastAP Buffer (100 mM Tris-HCl (pH 8.0 at 37 °C), 50 mM $MgCl_2$, 1 M KCl, 0.2 % Triton X-100 and 1 mg/mL BSA. Included in Thermo Fisher Scientific, cat. no. EF0651)
- UltraPure™ 1 M Tris-HCI Buffer, pH 7.5 (Thermo Fisher Scientific, cat. no. 15567027)

- UltraPure™ 0.5M EDTA, pH 8.0 (Thermo Fisher Scientific, cat. no. 15575020)
- NaCl (5 M), RNase-free (Thermo Fisher Scientific, cat. no. AM9760G)
- UltraPure™ DEPC-Treated Water (Thermo Fisher Scientific, cat. no. 750023)
- Ethyl alcohol, Pure (CH_3CH_2OH, Sigma-Aldrich, cat. no. E7023) **!CAUTION** Ethanol is flammable. Keep away from fire.
- Methanol (CH_3OH, Sigma-Aldrich, cat. no. 34860) **!CAUTION** Methanol is volatile and the vapor is toxic and flammable. Handle with appropriate PPE and keep away from fire.
- Acetonitrile (CH_3CN, Sigma-Aldrich, cat. no. 34851) **!CAUTION** Acetonitrile is volatile and the vapor is toxic and flammable. Handle with appropriate PPE and keep away from fire.
- Nuclease Decontamination Solution (IDT, cat. no. 11-05-01-01)
- HEPES (Fisher Scientific, cat. no. BP310-100)
- Sodium hydroxide (NaOH, Sigma-Aldrich, cat. no. 221465)
- Ammonium chloride (NH_4Cl, Sigma-Aldrich, cat. no. 1.01142)
- Magnesium chloride ($MgCl_2$, Sigma-Aldrich, cat. no. M4880)
- Zinc chloride ($ZnCl_2$, Sigma-Aldrich, cat. no. Z0152)
- Iodine (I_2, Sigma-Aldrich, cat. no. 376558)
- Potassium iodide (KI, Sigma-Aldrich, cat. no. 207969)
- Hexaamminecobalt(III) chloride ($Co(NH_3)_6Cl_3$, Sigma-Aldrich, cat. no. 481521)
- Sodium thiosulfate ($Na_2S_2O_3$, Sigma-Aldrich, cat. no. 217263)
- Tween™ 20 (Fisher Scientific, cat. no. BP337-100)
- DPBS, no calcium, no magnesium (Thermo Fisher Scientific, cat. no. 14190144)
- SDS, 20% Solution, RNase-free (Thermo Fisher Scientific, cat. no. AM9820)

- Dynabeads™ mRNA DIRECT™ Purification Kit (Thermo Fisher Scientific, cat. no. 61012)
- RiboMinus™ Eukaryote System v2 (Thermo Fisher Scientific, cat. no. A15026)
- RNA Clean & Concentrator-5 (Zymo Research, cat. no. R1013)
- Oligo(dT)18 Primer (Thermo Fisher Scientific, cat. no. SO132)
- RNase H (New England BioLabs, cat. no. M0297L)
- RNase H Reaction Buffer (10x, 500 mM Tris-HCl, pH 8.3, 750 mM KCl, 30 mM $MgCl_2$, 100 mM DTT. included in New England BioLabs, cat. no. M0297L)
- RNaseOUT™ Recombinant Ribonuclease Inhibitor (Thermo Fisher Scientific, cat. no. 10777019)
- Dynabeads™ MyOne™ Silane (Thermo Fisher Scientific, cat. no. 37002D)
- Buffer RLT (220 mL, Qiagen, cat. no. 79216)
- T4 Polynucleotide Kinase (New England BioLabs, cat. no. M0201L)
- T4 Polynucleotide Kinase Reaction Buffer (700 mM Tris-HCl, pH 7.6,100 mM $MgCl_2$, 50 mM DTT. included in New England BioLabs, cat. no. M0201L)
- SUPERase•In™ RNase Inhibitor (Thermo Fisher Scientific, cat. no. AM2696)
- T4 RNA Ligase 2, truncated KQ (New England BioLabs, cat. no. M0373L)
- T4 RNA Ligase Reaction Buffer (500 mM Tris-HCl, pH 7.5, 100 mM $MgCl_2$, 10 mM DTT. included in New England BioLabs, cat. no. M0373L and M0437M)
- PEG 8000 (50% w/v, included in New England BioLabs, cat. no. M0373L and M0437M)
- 5´ Deadenylase (New England BioLabs, cat. no. M0331S)
- $RecJ_f$ (New England BioLabs, cat. no. M0264L)
- Dynabeads™ MyOne™ Streptavidin C1 (Thermo Fisher Scientific, cat. no. 65002)

- AMV Reverse Transcriptase Reaction Buffer (10x, 500 mM Tris-acetate, pH 8.3, 750 mM Potassium acetate, 80 mM Magnesium acetate, 100 mM DTT. Included in New England BioLabs, cat. no. M0277S)
- Reverse Transcriptase, Recombinant HIV (Worthington Biochemical, cat. no. LS05006)
- Actinomycin D ($C_{62}H_{86}N_{12}O_{16}$, Sigma-Aldrich, cat. no. A1410)
- Deoxynucleotide (dNTP) Solution Mix (New England BioLabs, cat. no. N0447L)
- T4 RNA Ligase 1 (ssRNA Ligase), High Concentration (New England BioLabs, cat. no. M0437M)
- Adenosine-5'-Triphosphate (ATP) (100 mM, included in New England BioLabs, cat. no. M0437M)
- DMSO (100%, included in New England BioLabs, cat. no. M0530L)
- NEBNext® Ultra™ II Q5® Master Mix (New England BioLabs, cat. no. M0544L)
- NEBNext® Multiplex Oligos for Illumina® (96 Unique Dual Index Primer Pairs) (New England BioLabs, cat. no. E6440S)
- Qubit™ dsDNA HS Assay Kit (Thermo Fisher Scientific, cat. no. Q32854)
- Qubit™ RNA HS Assay Kit (Thermo Fisher Scientific, cat. no. Q32852)
- KAPA Library Quantification Kits - Complete kit (Universal) (Roche, cat. no. 07960140001)

Equipment

- 50 mL round bottom flask (Synthware)

- 250 mL Erlenmeyer flask (Synthware)
- 100 mL separation funnel (Synthware)
- Fisherbrand™ Octagon Spinbar™ Magnetic Stirring Bars (Fisher Scientific, cat. no.14-513-51)
- IKA 3622001 RET basic MAG Stirring Hot Plate (IKA)
- Rotavapor® R-215 (Büchi)
- Millex-GS Syringe Filter Unit, 0.22 μm (Millipore, cat. no. SLGSV255F)
- VirTis Sentry™ 12SL Freeze Dryer (VirTis)
- Waters Alliance HPLC System (Waters, 176803000)
- Galaksil EP-C18M 10 μm, 10 x 250 mm (Galak Chromatography)
- Avestin EmulsiFlex C3 (Avestin)
- Vivaspin® 6, 10 kDa MWCO Polyethersulfone (Cytiva, cat. no. GE28-9322-96)
- SOURCE™ 15Q 4.6/100 PE (Cytiva, cat. no. GE17-5181-01)
- SOURCE™ 15S 4.6/100 PE (Cytiva, cat. no. GE17-5182-01)
- Ni Sepharose® 6Fast Flow (Cytiva, cat. no. GE17-5318-01)
- ÄKTA pure protein purification system (Cytiva)
- Millex Syringe Filter, Durapore® (PVDF), Non-sterile (Millipore, cat. no. SLGVR04NL)
- 11mm Plastic Crimp/Snap Top Autosampler Vials (Thermo Fisher Scientific, cat. no. C4011-13)
- 11mm Autosampler Snap-It Caps (Thermo Fisher Scientific, cat. no. 60180-676)
- SCIEX Triple Quad™ 6500+ LC-MS/MS Systems (SCIEX)
- Agilent Eclipse XDB-C18 reversed phase HPLC column 4.6 mm ID x 250 mm (5 μm) 80Å (Agilent)

- NanoDrop™ 8000 Spectrophotometer (Thermo Fisher Scientific, cat. no. ND8000)
- 1.5 mL Low Adhesion Microcentrifuge Tubes (USA scientific, cat. No. 1415-2600)
- Tubes and Domed Caps, strips of 8 (Thermo Fisher Scientific, cat. no. AB0266)
- Reach Olympus Premium Barrier Tips (Genesee Scientific, cat. no. 10 µL: 23-401; 200 µL: 23-412; 1000 µL: 23-430)
- ART™ Barrier Speciality Pipette tips (Thermo Fisher Scientific, cat. no. 2149)
- Pipetman L single channel pipette (Gilson, P2L, cat. no. P2L, P10L, P20L, P100L, P200L, P1000L: SKU: FA10001M-10006M, respectively.)
- Fisherbrand™ Elite™ Multichannel Pipettes (Fisher Scientific, cat. no. 1-10 µL: FBE1200010; 30-300 µL: FBE1200030)
- 12 -Tube Magnetic Separation Rack (New England BioLabs, cat. no. S1509S)
- DiaMag 0.2ml - magnetic rack (Diagenode, cat. no. B04000001)
- Qubit™ 2.0 Fluorometer (Thermo Fisher Scientific, cat. no. Q33216)
- Qubit™ Assay Tubes (Thermo Fisher Scientific, cat. no. Q32856)
- 2100 Bioanalyzer Instrument (Agilent, cat. no. G2939BA)
- NovaSeq 6000 System (Illumina)

Reagent Setup

0.1% TFA (v/v)

Prepare 10 mL of 0.1% TFA (v/v) by combing 10 µL of trifluoroacetic acid with 10 mL of deionized water. The buffer can be stored at room temperature for at least 3 months.

5 M Imidazole, pH 8.0

Dissolve 17.02 g of Imidazole in 50 mL of deionized water to produce 5 M solution. Adjust the pH to 8.0 with NaOH. The buffer can be stored at 4 ºC for at least 3 months.

Lysis Buffer

Lysis buffer is composed of 50 mM Tris-HCl, pH 7.5 and 300 mM NaCl. To prepare 1 L of lysis buffer, combine 50 mL of 1 M Tris-HCl, pH 7.5 and 60 mL of 5 M NaCl. Add deionized water until the total volume is 1 L. The buffer can be stored at 4 ºC for at least 3 months.

Ni Wash Buffer

Prepare 500 mL of Ni wash buffer by adding 0.5 mL of 5 M imidazole, pH 8.0 into 500 mL Lysis Buffer. Prepare freshly before use. The final concentration of imidazole is 5 mM.

Ni Elution Buffer

Prepare 100 mL of Ni wash buffer by adding 0.4 mL of 5 M imidazole, pH 8.0 into 100 mL Lysis Buffer. Prepare freshly before use. The final concentration of imidazole is 20 mM.

Buffer A

Buffer A is 20 mM Tris-HCl, pH 7.5. Prepare 1 L of buffer A by adding 20 mL of 1 M Tris-HCl, pH 7.5 into 1 L of deionized water.

Buffer B

Buffer B is composed of 20 mM Tris-HCl, pH 7.5 and 1 M NaCl. Prepare 1 L of buffer A by dissolving 58.44 g NaCl in 1 L of buffer A.

1 M NH_4Ac, PH 5.3

Dissolve 3.85 g of NH_4Ac in 50 mL of deionized water to produce 1 M solution. Adjust the pH to 5.3 with HCl. The buffer can be stored at room temperature for at least 3 months.

▲ **CRITICAL** The following reagents setup should be performed under RNase-free conditions. Clean the gloves and working area with 70% EtOH and Nuclease Decontamination Solution. Use RNase free water when indicated.

1 M HEPES, PH 8.0

Dissolve 11.92 g of HEPES in 50 mL of RNase-free water to produce 1 M solution. Adjust the pH to 8.0 with NaOH. The buffer can be stored at room temperature for at least 3 months.

1 M NH_4Cl

Dissolve 2.67 g of NH_4Cl in 50 mL of RNase-free water to produce 1 M solution. The solution can be stored at room temperature for at least 3 months.

2 M $MgCl_2$

Dissolve 9.52 g of $MgCl_2$ in 50 mL of RNase-free water to produce 2 M solution. The solution can be stored at room temperature for at least 3 months.

1 M $ZnCl_2$

Dissolve 6.81 g of $ZnCl_2$ in 50 mL of RNase-free water to produce 1 M solution. The solution can be stored at room temperature for at least 3 months.

1 M KI

Dissolve 8.30 g of KI in 50 mL of RNase-free water to produce 1 M solution. The solution can be stored at room temperature for at least 3 months.

125 mM I_2 (200 mM KI solution)

To prepare 1mL of 125 mM I_2 (200 mM KI solution), dissolve 31.73 mg of I_2 in 200 μL of 1 M KI solution and 800 μL of RNase-free water. Protect the solution from light by wrapping the tube with aluminum foil. The solution can be stored at –20 ºC for at least a year. ▲ **CRITICAL** Light sensitive. Keep in dark.

200 mM $Na_2S_2O_3$

Dissolve 31.62 mg of $Na_2S_2O_3$ in 1 mL of RNase-free water to produce 200 mM solution. The solution can be stored at –20 ºC for at least a year.

40 mM $Co(NH_3)_6Cl_3$

Dissolve 10.70 mg of $Co(NH_3)_6Cl_3$ in 1 mL of RNase-free water to produce 40 mM solution. The solution can be stored at –20 ºC for at least a year.

0.1% PBST (v/v)

Prepare 50 mL of 0.1% PBST (v/v) by combing 50 µL of Tween 20 with 50 mL of DPBS. The buffer can be stored at room temperature for at least 3 months.

2x Binding/Wash Buffer

2x Binding/Wash buffer is composed of 10 mM Tris-HCl, pH 7.5, 1 mM EDTA and 2 M NaCl. To prepare 50 mL of 2x Binding/Wash buffer, combine 500 µL of 1 M Tris-HCl, pH 7.5, 100 µL of 0.5 M EDTA, and 20 mL of 5M NaCl. Add RNase-free water until the total volume is 50 mL. The buffer can be stored at room temperature for at least 3 months.

1x Binding/Wash Buffer

Dilute 2x Binding/Wash buffer with an equal volume of RNase-free water.

10x MjDim1 Reaction Buffer

10x MjDim1 reaction buffer is composed of 400 mM HEPES, pH 8, 400 mM NH_4Cl and 40 mM $MgCl_2$. To prepare 50 mL of 10x MjDim1 reaction buffer, combine 20 mL of 1 M HEPES, pH 8, 20 mL of 1 M NH_4Cl, and 1 mL of 2M $MgCl_2$. Add RNase-free water until the total volume is 50 mL. The buffer can be aliquoted into 1 mL each and stored at -20 ºC for at least a year.

Zn Fragmentation Buffer

Zn fragmentation buffer is composed of 100 mM Tris-HCl, pH 7.5 and 100 mM $ZnCl_2$. To prepare 1mL of Zn fragmentation buffer, combine 100 µL of 1 M Tris-HCl, pH 7.5, 100 µL of 1 M $ZnCl_2$ and 800 µL of RNase-free water. The buffer can be stored at –20 ºC for at least a year.

0.1 M EDTA

Dilute 0.5 M EDTA with 4 times volume of RNase-free water.

31.25 μM Actinomycin D

Prepare 0.8 mL of 3 mM Actinomycin D stock solution by dissolving 3 mg of Actinomycin D in 800 μL of DMSO. Measure the exact concentration by diluting 1 μL of the stock solution in 100 μL of methanol. Measure the absorbance at 443 nm using a NanoDrop. The extinction coefficient is 24.4 mM^{-1}. Store the stock solution at –80 ºC. The stock solution is stable at –80 ºC for at least a year. Mix 1 μL of the stock with 23 μL of DMSO and 72 μL of RNase-free water to produce a 31.25 μM Actinomycin D solution in 25% DMSO. Aliquot and store the diluted solution at –80 ºC. Thaw each aliquot only once. ▲ **CRITICAL** Actinomycin D is highly light sensitive. Store the solution in opaque 1.5 mL tubes. Diluted Actinomycin D might adsorb to plastic. Avoid repeated freeze-saw cycles.

MjDim1 QC probe

(rCrGrUrGrG/iN6Me-rA/rCrUrGrGrCrU/3Bio/)

This RNA oligo-ribonucleotide contains a N^6-methyladenosine modification (m^6A, code /iN6Me-rA/ in IDT) and a 3' biotinylation modification (code /3Bio/ in IDT's catalog). The probe should be ordered with RNase-free HPLC purification. Dissolve the oligonucleotide in RNase-free water to produce a 100 μM stock solution. The dissolved probe is stable at –80 ºC for at least a year.

3' Adaptor (/5rApp/AGATCGGAAGAGCGTCGTG/3Bio/)

This ssDNA oligonucleotide contains a 5' adenylation modification (code /5rApp/ in IDT's catalog) and a 3' biotinylation modification (code /3Bio/ in IDT's catalog). The probe should be ordered

with RNase-free HPLC purification. Dissolve the oligonucleotide in RNase-free water to produce a 100 μM stock solution. Dilute and aliquot the stock solution into 20 μL x 20 μM format. Store the aliquots and remaining stock solution at –80 ºC. The dissolved adaptor is stable at –80 ºC for at least a year.

RT Primer (ACACGACGCTCTTCCGATCT)

This ssDNA oligonucleotide should be ordered with RNase-free HPLC purification. Dissolve the oligonucleotide in RNase-free water to produce a 100 μM stock solution. Dilute and aliquot the stock solution into 50 μL x 2 μM format. Store the aliquots and remaining stock solution at –20 ºC. The dissolved primer is stable at –20 ºC for at least a year.

cDNA Adaptor (/5Phos/NNNNNNAGATCGGAAGAGCACACGTCTG/3SpC3/)

This ssDNA oligonucleotide contains a 5' phosphorylation modification (code /5Phos/ in IDT's catalog) and a 3' C3 spacer modification (code /3SpC3/ in IDT's catalog). At the beginning of 5' end the probe contains 6 random nucleotides to serve as the UMI. Order standard mixed bases with an equal ratio of A/T/C/G at each N position. The probe can be ordered with standard desalting purification. Dissolve the oligonucleotide in RNase-free water to produce a 100 μM stock solution. Dilute and aliquot the stock solution into 20 μL x 25 μM format. Store the aliquots and remaining stock solution at –20 ºC. The dissolved adaptor is stable at –20 ºC for at least a year.

Spike-in Probes and Spike-in Mix

(Probe 1:

rUrArUrCrUrGrUrCrUrCrGrArCrGrUrNrNrArNrNrGrGrCrCrUrUrUrGrCrArArCrU

rArGrArArUrUrArCrArCrCrArUrArArUrUrGrCrU

Probe 2:

rUrArUrCrUrGrUrCrUrCrGrArCrGrUrNrNrArNrNrGrGrCrArUrUrCrArArGrCrCrU

rArGrArArUrUrArCrArCrCrArUrArArUrUrGrCrU

Probe 3:

rUrArUrCrUrGrUrCrUrCrGrArCrGrUrNrNrArNrNrGrGrCrGrArGrGrUrGrArUrCrU

rArGrArArUrUrArCrArCrCrArUrArArUrUrGrCrU

Probe 4:

rUrArUrCrUrGrUrCrUrCrGrArCrGrUrNrNrArNrNrGrGrCrUrUrCrArArCrArArCrU

rArGrArArUrUrArCrArCrCrArUrArArUrUrGrCrU

Probe 5:

rUrArUrCrUrGrUrCrUrCrGrArCrGrUrNrN/iN6Me-

rA/rNrNrGrGrCrArUrUrCrArArGrCrCrUrArGrArArUrUrArCrArCrCrArUrArArUrU

rGrCrU

Probe 6:

rUrArUrCrUrGrUrCrUrCrGrArCrGrUrNrN/iN6Me-rA/rNrNrGrGrCrGrArGrGrUrGrArUrCrUrArGrArArUrUrArCrArCrCrArUrArArUrUrGrCrU

Probe 7:

rUrArUrCrUrGrUrCrUrCrGrArCrGrUrNrN/iN6Me-rA/rNrNrGrGrCrUrUrCrArArCrArArCrUrArGrArArUrUrArCrArCrCrArUrArArUrUrGrCrU

Probe 8:

rUrArUrCrUrGrUrCrUrCrGrArCrGrUrNrN/iN6Me-rA/rNrNrGrGrCrGrArUrGrGrUrUrUrCrUrArGrArArUrUrArCrArCrCrArUrArArUrUrGrCrU)

Order 8 oligo-ribonucleotide probes with the sequence listed above. Probes 5-8 contain a N^6-methyladenosine modification (m^6A, code /iN6Me-rA/ in IDT). All probes contain random ribonucleotides designated as rN, which should be ordered as standard mixed bases with an equal ratio of rA/rU/rC/rG. The probes should be ordered with RNase-free HPLC purification. Dissolve each oligo-ribonucleotide in RNase-free water to produce a 10 ng/μL stock solution. Prepare the spike-in mix by combining the stock solutions in the following ratio.

Probe #	1	2	3	4	5	6	7	8
Normalized ratio	0.20	0.15	0.10	0.05	0.05	0.10	0.15	0.20

Volumes of 10 ng/μL stock in 1 mL of 10 ng/μL mix (μL)	200	150	100	50	50	100	150	200
Volumes of 10 ng/μL stock in 1 mL of 1 ng/μL mix (μL)	20	15	10	5	5	10	15	20

Prepare 10 ng/μL mix and dilute 10 times in RNase-free water, or directly prepare 1 ng/μL mix using the above form. Aliquot into 50 μL x 1 ng/μL formula. Store the aliquots at –80 ºC. The aliquoted spike-in mix is stable at -80 ºC for at least a year.

Procedure

Synthesis of Allyl-SAM

((2S)-2-Amino-4-[(RS)-{[(2S,3S,4R,5R)-5-(4-amino-9H-purin-9-yl)-3,4-dihydroxyoxolan-2-yl]methyl}(prop-2-enyl)sulfaniumyl]butanoate) ●Timing 3 d

1. In a 50 mL round bottom flask weigh and add 100 mg of S-(5′-Adenosyl)-L-homocysteine. Add 1 mL of acetic acid and 1 mL of formic acid. **!CAUTION** Perform steps 1-7 in a fume hood with appropriate PPE.
2. Weigh and add 52.0 mg of $AgClO_4$. Add 2.125 mL of allyl bromide. Add a stir bar and stir at room temperature for 8 h. **!CAUTION** Ally bromide is very toxic and should be handled with exceptional caution.

3. Quench the reaction by adding 20 mL of 0.1% TFA.
4. Transfer the quenched reaction mixture excluding the stir bar to a 100 mL separation funnel. Wash the reaction mixture with 10 mL of diethyl ether. Keep the lower phase after separation.
5. Repeat the wash two more times, using 10 mL of diethyl ether each time.
6. Transfer the washed aqueous phase to a 50 mL syringe with the nozzle attached to a 0.22 μm filter. Carefully attach the plunger. Pass the solution through the filter by applying steady pressure to the plunger.
7. Transfer the solution into a 50 mL round bottom flask. Evaporate any residual ether using a rotatory evaporator. Keep the solution at 4 ºC by submerging the flask in the ice-water bath. ▲ **CRITICAL** Using a heated water bath will cause product decomposition.
8. Add 0.1% of the filtrate's volume of TFA (add 20 μL of TFA to 20 mL of filtrate, for example). Aliquot the acidified solution into 1.5 mL tubes. Snap-freeze and lyophilize the solution using a VirTis Sentry Lyophilizer. ▲ **CRITICAL** It is recommended to lyophilize immediately after synthesis.
9. Purify the allyl-SAM by reconstituting the lyophilized crude in 0.1% TFA. Purify the solution on a Waters Alliance HPLC system with a Galaksil EP-C18M 10 μm, 10 x 250 mm reversed phase HPLC column. Inject ~1 μmol in 100 μL each run. Elute the allyl-SAM using the following LC program (retention time = 6-8 min).

Time (min)	A (%)	B (%)	Flow (mL/min)	Max Pressure Limit (Bar)
0.00	98.0	2.0	3.000	600.00

10.00	98.0	2.0	3.000	600.00
10.50	0.0	100.0	3.000	600.00
17.00	0.0	100.0	3.000	600.00
17.50	98.0	2.0	3.000	600.00
25.00	98.0	2.0	3.000	600.00

Where Phase A = 0.01% TFA in deionized water; Phase B = 0.01% TFA in acetonitrile. **?TROUBLESHOOTING**

10. Determine the concentration of purified allyl-SAM by measuring A260 on NanoDrop. The extinction coefficient is 15.4 mM^{-1}.
11. Aliquot the eluate into 1.5 mL tubes each containing 968 nmol of ally-SAM. One aliquot is sufficient for 4 reactions plus 10% extra volume. Lyophilize the eluate. Store the compound at –80 ºC. The lyophilized allyl-SAM is stable at –80 ºC for at least a year. ▲**CRITICAL** The lyophilized allyl-SAM is frozen in acidified conditions. Do not neutralize before use (steps 31 and 64).

Expression and purification of MjDim1 ●Timing 6 d

12. MjDim1 gene was previously codon-optimized and cloned into a pET28a-SUMO vector. Thaw one vial of competent BL21(DE3) cells on ice for 5 min. Add 1 μL of 50 ng/μL pET28a-SUMO-MjDim1 plasmid. Mix by gentle shaking and incubate on ice for 30 min. Then heat-shock the cells at 42 ºC for 30 s. Culture the cells in 1 mL of SOC at 37 ºC for 1 hr before spreading 50 μL of the culture on an agar plate supplemented with 50 mg/mL Kanamycin. Culture the plate overnight at 37 ºC. Pick a single colony and incubate

overnight in 8 mL of LB broth supplemented with 50 mg/mL Kanamycin. Collect 1 mL of overnight culture and mix with an equal volume of glycerol. Store at -80 ºC. The glycerol stock can be stored for at least a year if caution is taken to minimize freeze-saw cycles.

13. Day 1. Dissolve 31 g of 2YT media in 1 L of deionized water. Prepare 6 L and autoclave the media. Typically, 6 L of culture will produce enough enzymes for 150 SAC-seq reactions.
14. Stuck a pipette tip into the glycerol stock without sawing the stock. Inoculate 60 mL of sterile LB media with the tip. Add 60 μL of 50 mg/mL kanamycin. Culture the cells at 37 ºC overnight (14-18 h) with 220 rpm shaking.
15. Day 2. Expand the starter culture by diluting 1/100 into the 6 L 2YT media. Per 1 L of media add 10 mL of starter culture and 1 mL of 50 mg/mL kanamycin. Culture the cells at 37 ºC 4-5 h with 220 rpm shaking. Then cool the incubator to 16 ºC (with the cells inside the incubator). Per 1 L of media add 0.1 mL of 1 M IPTG. Culture the cells at 37 ºC overnight (14-18 h) with 220 rpm shaking.
16. Prepare a nickel column by washing 20-25 mL of Ni-NTA resin in 150 mL of H_2O. Wash 3 times using 50 mL each time. Then wash in 50 mL H_2O containing 100 μL of 5 M imidazole, pH 8.0. Then wash in 150 mL of lysis buffer. Wash 3 times using 50 mL each time. Store at 4 ºC overnight.
17. Day 3. Collect the cells by centrifuging at 6,000 rpm for 10 min.
18. Resuspend the cells in 300 mL of lysis buffer (50 mL/1 L culture). Add 100 μL of 10 mg/mL DNase I. Keep the resuspension on ice.

19. Rinse the interior of EmulsiFlex C3 with deionized water. Homogenize the cells by passing the resuspension through the machine while applying 100 bars of pressure. Lyse the cells at 1000-1500 bars. Collect the lysate. Keep the lysate on ice.
20. Centrifugate the lysate at 14,000 x g for 20 min. Load the supernatant onto the prepared nickel column. After the loaded volume completely passes through, wash the column with 300 mL of Ni wash buffer.
21. Add 15 mL of 0.5 mg/mL ULP1 to the column. Let 1-2 mL of the buffer pass through the column. Then attach the cap to the outlet nozzle and put the column in an upright position. Incubate at 4 ºC overnight.
22. Prepare and balance a SOURCE 15Q-column and a 15S-column by washing both sequentially in 60 mL of 300 mM NaOH, 60 mL of H_2O, 60 mL of buffer B, and 60 mL of buffer A. Store at 4 ºC overnight.
23. Day 4. Elute the SUMO-cleaved protein from the nickel column with 80 mL of Ni elution buffer.
24. Add 20 μg of HeLa total RNA into the eluate. Mix well and incubate at 37 ºC for 30 min. This step removes the endogenous SAM-cofactors and facilitates later binding with allyl-SAM. Adding any type of RNA will work.
25. Load the eluate with RNA onto the Q-column. Collect the flowthrough while loading. Then apply 100% Buffer A and collect another 30 mL. If using AKTA Pure FPLC system, a significant reduction in UV-280 nm signal should be observed around this time point.

    ▲ **CRITICAL** The enzyme is not retained on column. Do not discard the flowthrough.
26. Load the eluate onto S-column. Balance the column with 60 mL of buffer A. Then apply a gradient of 50% buffer B in 20 min. If using AKTA Pure FPLC system, a significant peak

in UV-280 nm signal should be observed peaking at 35 mS/cm conductivity, and it should be the only peak.

27. Concentrate the protein with 10 kD MWCO spin filters to 2 mL. Spin at 7,000 x g for 20 min each time. Wash the concentrate with buffer A 3 times using 2 mL each time. Concentrate again to 2 mL and add 860 µL of sterile glycerol (final concentration 30%, v/v). Dilute 10 times before measuring A280 on NanoDrop. The concentration should be 1.2-2 mM (Extinction coefficient 23.38 mM^{-1}). ▲ **CRITICAL** It is recommended to concentrate and store the enzyme immediately. Otherwise, keep the solution at 4 ºC for no more than one night. **?TROUBLESHOOTING**
28. Day 5. Measure the reactivity of purified MjDim1 by performing the labeling reaction on 50 ng of MjDim1 QC probe. Wash 40 µL of Dynabeads MyOne Streptavidin C1 beads in 40 µL of 0.1 M NaOH. Wash again in 40 µL of 10 mM Tris-HCl, pH 7.5. Resuspend the beads in 40 µL of 1x Binding/Wash buffer. Add 2 µL of 50 ng/µL MjDim1 QC probe. Mix well and incubate at room temperature for 15 min.
29. Place the beads on a magnetic rack. Decant the beads. Wash the beads once in 50 µL of 1x Binding/Wash buffer. Then twice in 50 µL of 10 mM Tris-HCl, pH 7.5 (using 50 µL each time).
30. Resuspend the beads in 12 µL of RNase-free H_2O. Divide into two halves. Use one half for labeling and the other as a control. Keep the control on ice.
31. Right before starting the next step, reconstitute 968 nmol of lyophilized allyl-SAM in 48.4 µL of 0.1% TFA. The resulting concentration is 20 mM. 11 µL of dissolved allyl-SAM is necessary for one reaction. Prepare enough amount for all reactions plus 10% of the volume. The remaining dissolved allyl-SAM should be stored at –80 ºC without proceeding to the

next step. ▲ **CRITICAL** Neutralize right before usage. Do not re-freeze neutralized allyl-SAM.

Component	Volume (μL)
Allyl-SAM (20 mM in 0.1% TFA)	3 μL
RNase-free H_2O	3 μL
1 M Tris-HCl, pH 8.3	0.6 μL
Total	6.6 μL

32. Add the following components to the beads (resuspended in 6 μL of RNase-free H_2O).

Component	Volume (μL)	Final concentration in 40 μL reaction
Neutralized Allyl SAM (~9 mM)	6 μL	2.7 mM
SUPERase•In RNase Inhibitor (20 U/μL)	2 μL	2 U/μL
10x MjDim1 Reaction Buffer	2 μL	1x
Purified MjDim1 (1.2-2 mM)	4 μL	0.4 mM
Total	14 μL	

33. Mix well and incubate at 50 ºC for 1 h.

34. Wash the beads sequentially with 50 μL of 0.1% PBST (v/v), 50 μL of 1x Binding/Wash buffer and twice with 50 μL of 10 mM Tris-HCl, pH 7.5 (using 50 μL each time). Decant and resuspend in 6 μL of RNase-free H_2O.

35. For both the labeled probe and the control (resuspended in 6 μL of RNase-free H_2O), add the following components.

Component	Volume per sample (μL)	Final concentration in 21.3 μL reaction
RNase-free H_2O	14 μL	-
1 M NH_4Ac, pH 5.3	0.3 μL	14 mM
Nuclease P1 (100 U/μL)	1 μL	4.7 U/μL
Total	15.3 μL	

36. Mix well and incubate at 37 ºC overnight, then add the following components.

Component	Volume per sample (μL)	Final concentration in 25.3 μL reaction
10x Fast AP Reaction Buffer	3 μL	1.18x
Fast AP (1 U/μL)	1 μL	0.04U/μL
Total	4 μL	

37. Mix well and incubate at 37 ºC for 2 h.

38. Load all the supernatant onto a 0.22 μm PVDF spin filter. Centrifuge at 10,000 x g at room temperature for 1 min. Transfer the filtrate into 250 μL 11-mm cap LC vials. Measure the 2 samples using a SCIEX Triple Quad LC-MS/MS System with Agilent Eclipse XDB-C18 reversed phase HPLC column. Use the following LC program.

Time (min)	A (%)	B (%)	Flow (mL/min)	Max Pressure Limit (Bar)
0.00	98.00	2.00	0.500	600.00
3.00	82.00	18.00	0.500	600.00
4.00	50.00	50.00	0.500	600.00
5.00	10.00	90.00	0.500	600.00
6.00	10.00	90.00	0.500	600.00
6.10	2.00	98.00	0.500	600.00
7.00	2.00	98.00	0.500	600.00
7.10	98.00	2.00	0.500	600.00
9.00	98.00	2.00	0.500	600.00

Where Phase A = 0.1% formic acid in water, Phase B = 0.1% formic acid in MeOH.

Set the scan type to MRM. Polarity = Positive. Duration = 9 min, Cycle = 1270 x 0.425 s per cycle.

Use the following MS1/MS2 parameters:

Q1 Mass (Da)	Q3 Mass (Da)	Time (ms)	ID	CE (V)
268.000	136.000	100.0	A	47.000
282.101	150.100	100.0	m^6A	25.000
284.004	152.100	50.0	G	17.000
244.936	113.000	50.0	C	19.000
244.037	112.000	100.0	U	17.000

39. Calculate the conversion rate as follows:

$$\text{Conversion \%} = \left(1 - \frac{S_{m^6A,\,treated} / S_{G,\,treated}}{S_{m^6A,\,control} / S_{G,\,control}}\right) \times 100\%$$

Where S stands for MS2 peak area. The conversion should not be lower than 60%. **?TROUBLESHOOTING**

Preparation of RNA samples for SAC-seq ●Timing 1 d

40. Refer to the following Box 1 for appropriate sample treatment before library preparation.

▲ **CRITICAL** steps 40-77 should be performed under RNase-free conditions. Clean the gloves and working area with 70% EtOH and Nuclease Decontamination Solution. Use RNase free water when indicated.

Box 1. Sample preparation for cell line, fresh-frozen and FFPE tissue samples
It is necessary to reduce ribosomal RNA fractions in RNA samples to maximize useful reads in sequencing data. Whenever possible, tRNA should also be reduced. We recommend the following treatments before starting step 41. 1. For cell line and fresh-frozen tissue samples: Purify total RNA with Trizol or spin-column based extraction kits. Use tissue homogenizer when necessary. i. If only mRNA m^{6}A modification is desired, extract mRNA with Dynabeads mRNA DIRECT Kit. It is necessary to perform the "Elimination of rRNA contamination" step in kit's manual, which essentially purified the mRNA twice.

ii. If whole transcriptome m^6A modification (e.g., lncRNA) is desired, deplete ribosomal RNA with RiboMinus Eukaryote System v2. Then purify the longer (>200 nt) RNA using RNA Clean & Concentrator-5.

Follow the manufacturer's instructions for detailed protocols. Measure concentration with Qubit RNA HS kit. The expected yield from total RNA is 0.5-1.5%.

2. For FFPE samples and limited supply of other types of samples: Purify total RNA with Trizol or spin-column based extraction kits. Use tissue homogenizer when necessary. Deplete ribosomal RNA with RiboMinus Eukaryote System v2. Then collect all RNA with spin-column based, bead based, or ethanol precipitation RNA purification methods. Measure concentration with Qubit RNA HS kit. The expected yield from total RNA is 5-15%.

▲ **CRITICAL** When using isolated nucleus as starting material, it is recommended to include a DNase treatment step in total RNA extraction. Otherwise, DNase treatment is not considered necessary for general purpose SAC-seq. **?TROUBLESHOOTING**

Poly A removal ●Timing 45 min

41. In 0.2 mL PCR tubes, assemble the following components.

Component	Volume per sample (μL)	Final concentration in 20 μL reaction
2-50 ng mRNA or ribo-depleted RNA	15 μL	-

Oligo(dT)18 Primer (100 µM)	1 µL	5 µM
Total	16 µL	

42. Incubate at 70 °C for 2 min. Hold on 4 °C. Then add the following components.

Component	Volume per sample (µL)	Final concentration in 20 µL reaction
10x RNase H Reaction Buffer	2 µL	1x
RNase H (5 U/µL)	1 µL	0.25 U
RNaseOUT (100 mM)	1 µL	5 mM
Total	(16 µL) + 4 µL	

43. Mix well and incubate at 37 °C for 30 min.

Fragmentation ●Timing 15 min

44. To the original tube add the following components.

Component	Volume per sample (µL)	Final concentration in 30 µL reaction
RNase-free H_2O	7 µL	-
10x Zinc Fragmentation Buffer	3 µL	1x
Total	(20 µL) + 10 µL	

The buffer will turn opaque due to the presence of DTT in the previous step precipitating some of the zinc ion. However, this does not negatively affect fragmentation.

45. Mix well and incubate at 70 °C for exactly 5 min. Hold on 4 °C.

▲ **CRITICAL** It is recommended to check fragment size using Bioanalyzer 2000. Quench the aliquot for a run on bioanalyzer with 1 μL of 0.5 M EDTA. Refer to Box 2.

?TROUBLESHOOTING

Box 2. Fragment size for SAC-seq
The RNA size should be between 50-500 nt, peaking around 150-200 bp, with an RIN of ~2.4. It is recommended to perform an optimization run for FFPE samples, using 6 replicates to fragment for 0, 1, 2, 3, 4, 5 minutes, respectively. Check the resulting sizes and determine the most appropriate time. Typically, the FFPE samples with higher DV200 (percentage >200 nt) should be fragmented for 3-5 min, while low DV200 samples could omit the fragmentation step.

End-repair ●Timing 90 min

46. To the original tube add the following components.

Component	Volume per sample (μL)	Final concentration in 50 μL reaction
0.1 M EDTA	3 μL	6 mM

10x T4 Polynucleotide Kinase Reaction Buffer	5 μL	1x
T4 Polynucleotide Kinase (10 U/μL)	5 μL	1 U /μL
SUPERase•In RNase Inhibitor (20 U/μL)	2.5 μL	1 U /μL
RNase-free H_2O	4.5 μL	-
Total	(30 μL) + 20 μL	

47. Mix well and incubate at 37 °C for 1 h.
48. Silane beads clean-up (steps 48-52). Wash 20 μL of beads in 175 μL of buffer RLT. Then resuspend them in 175 μL of buffer RLT.
49. Mix the beads with 50 μL of reaction mixture and 300 μL of EtOH. Incubate at room temperature for 5 min.
50. Separate the beads on magnetic racks (~1 min). Wash the beads with 200 μL of 70% EtOH by resuspension. Repeat the wash for once. ▲ **CRITICAL** Use freshly prepared 70% EtOH.
51. Briefly spin down the beads. Decant completely on magnetic racks. Air dry with lids open for 3 min. ▲ **CRITICAL** Over-drying might result in sample loss.
52. Resuspend the beads in 11 μL of RNase-free H_2O. Incubate at room temperature for 5 min. Separate the beads on magnetic racks and transfer 10 μL of the eluate to a new tube.

RNA adaptor ligation ●Timing 16 h

53. Assemble the following reaction.

Component	Volume per sample (μL)	Final concentration in 25 μL reaction
End-repaired RNA	10 μL	-
3' Adaptor (20 μM)	1 μL	0.8 μM
Spike-in Mix (1 ng/μL)	1 μL	2% of the input RNA
Total	12 μL	

If using less than 50 ng input RNA, it is recommended to dilute the spike-in mix proportionally.

54. Incubate at 70 °C for 2 min. Then add the following components in the listed order.

▲ **CRITICAL STEP** Add the ligase separately. Do not mix the ligase in 13 μL of master-mixes.

Component	Volume per sample (μL)	Final concentration in 25 μL reaction
10x T4 RNA Ligase Reaction Buffer	2.5 μL	1x
SUPERase•In RNase Inhibitor (20 U/μL)	1 μL	0.8 U/μL
50% PEG 8000	7.5 μL	15% (w/v)
T4 RNA Ligase 2, truncated KQ (200 U/μL)	2 μL	16 U/μL
Total	(12 μL) + 13 μL	

55. Mix well and incubate at 25 ºC for 2 h, then at 16 ºC for 12 h. Hold on 4 ºC.

56. Digest excess adaptors by adding:

Component	Volume per sample (μL)	Final concentration in 51 μL reaction
RNase-free H_2O	25 μL	-
5' Deadenylase (50 U/μL)	2 μL	~2 U/μL
Total	(25 μL) + 27 μL	

57. Mix well and incubate at 30 ºC for 1 h. Then add:

Component	Volume per sample (μL)	Final concentration in 52 μL reaction
$RecJ_f$ (30 U/μL)	1 μL	~0.6 U/μL
Total	(52 μL) +1 μL	

58. Mix well and incubate at 37 ºC for 1h.

■ **PAUSE POINT** Do not proceed to step 59 if not performing steps 59-73 within the same day. Instead, the ligated RNA could be stored at –80 ºC for at least a week.

59. Dynabeads MyOne Streptavidin C1 clean up (steps 59-63). Wash 40 μL of beads in 40 μL of 0.1 M NaOH. Wash again in 40 μL of 10 mM Tris-HCl, pH 7.5.
60. Resuspend the beads in 50 μL of 2x Binding/Wash buffer. Mix the beads with 50 μL of ligation mixture. Incubate at room temperature for 15 min.
61. Separate the beads on magnetic racks. Wash the beads in 50 μL of 0.1% PBST (v/v) and 50 μL of 1x Binding/Wash buffer. Then wash twice in 50 μL of 10 mM Tris-HCl, pH 7.5 (using 50 μL each time).

62. Resuspend the beads in 36 μL of RNase-free H_2O. Mix well and transfer 12 μL of the beads to a new PCR tube. Label these new tubes as the RNA-seq input of the corresponding SAC-seq experiment. Store the input in a -80 °C freezer.
63. Place the original tubes on magnetic racks. Remove 12 μL of RNase-free H_2O. Resuspend the beads in the remaining volume of RNase-free H_2O. Incubate at 70 °C for 2 min. Hold on 4 °C.

MjDim1 labeling ●Timing 2.5 h

64. Right before starting the next step, reconstitute 968 nmol of lyophilized allyl-SAM in 48.4 μL of 0.1% TFA. The resulting concentration is 20 mM. 11 μL of dissolved allyl-SAM is necessary for one reaction. Prepare enough amount for all reactions plus 10% of the volume. The remaining dissolved allyl-SAM should be stored at –80 ºC without proceeding to the next step. ▲ **CRITICAL** Neutralize right before usage. Do not re-freeze neutralized allyl-SAM.
65. Prepare neutralized allyl SAM by mixing the following components. The listed volume is necessary for one reaction. Prepare enough amount for all reactions plus 10% of the volume.

Component	Volume per sample (μL)	Volume per 24 samples (μL)
Allyl-SAM (20 mM in 0.1% TFA)	11 μL	264 μL
RNase-free H_2O	11 μL	264 μL
1 M Tris-HCl, pH 8.3	2.2 μL	52.8 μL
Total	24.2 μL	580.8 μL

66. Add the following components to the beads (resuspended in 12 μL of RNase-free H_2O).

Component	Volume per sample (μL)	Final concentration in 40 μL reaction
Neutralized Allyl SAM (~9 mM)	12 μL	2.7 mM
SUPERase•In RNase Inhibitor (20 U/μL)	4 μL	2 U/μL
10x MjDim1 Reaction Buffer	4 μL	1x
MjDim1 (1.2-2 mM)	8 μL	0.4 mM
Total	(12 μL) + 28 μL	

67. Mix well and incubate at 50 ºC for 1 h.
68. Decant the beads. Resuspend the beads with 12 μL of RNase-free H_2O and repeat the labeling reaction by adding the same components. Mix well and incubate at 50 ºC for 1 h.
69. Wash the beads sequentially with 50 μL of 0.1% PBST (v/v), 50 μL of 1x Binding/Wash buffer and twice with 50 μL of 10 mM Tris-HCl, pH 7.5 (using 50 μL each time). Then decant completely.

Iodine labeling ●Timing 90 min

70. Resuspend the beads in 25 μL of RNase-free H_2O. Add 1 μL of 125 mM I_2 (200 mM KI solution).
71. Mix well and incubate at room temperature for 1h. ▲**CRITICAL** The KI solution is light sensitive. Keep in dark.
72. Add 1 μL of 200 mM $Na_2S_2O_3$ for quenching, mix well.

73. Wash the beads sequentially with 50 μL of 1x Binding/Wash buffer and twice with 50 μL of 10 mM Tris-HCl, pH 7.5 (using 50 μL each time). Then decant completely. Resuspend the beads in 12 μL of RNase-free H_2O. Store in the -80 °C freezer with corresponding inputs.

■ **PAUSE POINT** Both the labeled and the input RNA could be stored at -80 °C for at least a week.

Reverse transcription ●Timing 2.5 h

74. Thaw the frozen PCR tubes for SAC-seq and inputs. Add the following components.

Component	Volume per sample (μL)	Final concentration in 25 μL reaction
2 μM RT Primer	1 μL	0.08 μM
dNTP Solution Mix (10 mM)	2.5 μL	1 mM
Total	3.5 μL	

75. Incubate at 70°C for 2 min. Hold on 4 °C.

76. Add the following components.

Component	Volume per sample (μL)	Final concentration in 25 μL reaction
10x AMV Reverse Transcriptase Reaction Buffer	2.5 μL	1x

RNaseOUT (100 mM)	1 μL	4 mM
Reverse Transcriptase, Recombinant HIV (50 U/μL)	2 μL	4 U/μL
Actinomycin D (31.25 μM)	4 μL	5 μM
Total	9.5 μL	

77. Mix well and incubate at 37°C for 2 h.

78. Add the following components.

Component	Volume per sample (μL)	Final concentration in 50.5 μL reaction
RNase-free H_2O	25 μL	-
10% SDS	0.5 μL	~0.1% (w/v)
Total	25.5 μL	

79. Mix well and incubate at 95°C for 10 min. Hold on 4 °C.

80. Separate the beads on magnetic racks. Transfer the supernatant to a new tube.

81. Silane beads clean-up (steps 81-85). Wash 10 μL of beads in 150 μL of buffer RLT. Then resuspend them in 150 μL of buffer RLT.

82. Mix the beads with 50 μL of reaction mixture and 75 μL of EtOH. Incubate at room temperature for 5 min.

83. Separate the beads on magnetic racks (~1 min). Wash the beads with 200 μL of 70% EtOH by resuspension. Repeat the wash for once. ▲ **CRITICAL** Use freshly prepared 70% EtOH.

84. Briefly spin down the beads. Decant completely on magnetic racks. Air dry with lids open for 3 min. ▲ **CRITICAL** Over-drying might result in sample loss.

85. Resuspend the beads in 12.5 μL of RNase-free H_2O. Incubate at room temperature for 5 min. Separate the beads on magnetic racks and transfer 11.5 μL of the eluate to a new tube.

■ **PAUSE POINT** cDNA could be stored at -20 °C for at least a month.

cDNA adaptor ligation ●Timing 16 h

86. Add 2 μL of 25 μM cDNA adaptor to the eluate. Incubate at 70 °C for 2 min. Hold on 4 °C. Then add the following components.

Component	Volume per sample (μL)	Final concentration in 50 μL reaction
10x T4 RNA Ligase Reaction Buffer	5 μL	1x
50% PEG 8000	25 μL	25% (w/v)
ATP (100 mM)	0.5 μL	1 mM
DMSO	3.75 μL	7.5% (v/v)
$Co(NH_3)_6Cl_3$ (40 mM)	1.25 μL	1 mM
Total	35.5 μL	

87. Mix well by pipetting up and down at least 10 times with the pipette set at 50 μL. This is critical for successful ligation. Then add the ligase. ▲ **CRITICAL STEP** Add the ligase

separately. Do not mix the ligase in 36.5 μL of master-mixes. It is recommended to mix an extra time before the addition of the ligase.

Component	Volume per sample (μL)	Final concentration in 50 μL reaction
T4 RNA ligase 1, high conc. (30 U/μL)	1 μL	0.6 U/μL

88. Mix well again by pipetting up and down at least 10 times with the pipette set at 50 μL. Incubate at 25 °C for at least 14 h.
89. Silane beads clean-up (steps 89-93). Wash 10 μL of beads in 150 μL of buffer RLT. Then resuspend them in 150 μL of buffer RLT.
90. Mix the beads with 50 μL of reaction mixture and 75 μL of EtOH. Incubate at room temperature for 5 min.
91. Separate the beads on magnetic racks (~1 min). Wash the beads with 200 μL of 70% EtOH by resuspension. Repeat the wash for once. ▲ **CRITICAL** Use freshly prepared 70% EtOH.
92. Briefly spin down the beads. Decant completely on magnetic racks. Air dry with lids open for 3 min. ▲ **CRITICAL** Over-drying might result in sample loss.
93. Resuspend the beads in 12.5 μL of RNase-free H_2O. Incubate at room temperature for 5 min. Separate the beads on magnetic racks and transfer 11.5 μL of the eluate to a new tube.

the ■ **PAUSE POINT** ligated cDNA could be stored at -20 °C for at least a month.

PCR amplification ●Timing 2 h

94. Assemble the PCR reaction by mixing the following components in the PCR tube containing cDNA.

Component	Volume per sample (μL)	Final concentration in 50 μL reaction
NEBNext Ultra II Q5 Master Mix (2x)	25 μL	1x
NEBNext Unique Dual Index Primer for Illumina (10 μM)	10 μL	2 μM
cDNA	15 μL	
Total	50 μL	

▲ **CRITICAL** Any primers following standard TruSeq design could be used. Check the adaptors' sequences to ensure compatibility.

95. Mix well and start the following PCR program.

Cycle number	Denature	Anneal & Extend
1	98 ºC 30 s	
2-11(10 cycles)	98 ºC 10 s	65 ºC 75 s
12		65 ºC 5 min

The cycle number should be adjusted based on input RNA amount. Typically:

Input RNA (ng)	PCR Cycles
2-10	13-14

10-25	11-12
25-50	9-10

■ **PAUSE POINT** Amplified libraries could be left in the PCR block at 4 ºC overnight or stored at -20 °C for at least a month.

AMPure XP beads clean up ●Timing 30 min

96. Equilibrate the AMPure XP beads to room temperature for 30 min. Add 40 μL (0.8x) beads to each PCR mixture. Mix well by pipetting up and down at least 15 times with the pipette set at 90 μL. Incubate at room temperature for 10 min.
97. Place the beads on a magnetic rack for 5min. Decant the beads.
98. Add 200 μL freshly prepared 70% EtOH to the beads without mixing them. Incubate at room temperature for 30s. Decant the beads. Repeat the wash. ▲**CRITICAL** Use freshly prepared 70% EtOH.
99. Briefly spin down the beads and decant completely. ▲**CRITICAL** Remaining droplets might introduce primer dimers in the purified library.
100. Air dry for 2 min. ▲**CRITICAL** Over-drying might result in sample loss.
101. Add 18 μL of RNase-free H_2O. Mix well and incubate at room temperature for 10 min. Place the beads on a magnetic rack for 5 min.
102. Transfer 15 μL of the eluate to a new 1.5 mL tube. Use 1 μL from the remainder for concentration measurement.

■ **PAUSE POINT** Purified libraries could be stored at -20 °C for at least a month.

Quality control and sequencing ●Timing 1-2 d

103. Measure concentration with Qubit dsDNA HS Assay Kit. Follow the manufacturer's directions. The final concentration should be 2-10 ng/μL. **?TROUBLESHOOTING**

104. Check the size distribution using Agilent 2100 Bioanalyzer. Any peak <150 bp should be minimal. The major peak should peak around 260-300 bp with a concentration of at least 3 nM. **?TROUBLESHOOTING**

105. Use KAPA Library Quantification Kits to determine the precise concentration of libraries. Follow the manufacturer's instructions. Pool the samples with the desired ratio based on qPCR results and Box 3. Perform sequencing on a Novaseq 6000 system with PE150 mode. **?TROUBLESHOOTING**

Box 3. Sequencing depths for SAC-seq
Typically, mRNA samples will require 50 M paired-end reads while size-selected (>200 bp) Ribo-depleted RNA samples will require 100-150 M paired-end reads. For non-size-selected Ribo-depleted RNA samples (e.g., FFPE), 150-200 M paired-end reads are necessary.

Quality control of raw sequencing data ●Timing 10 min

106. ▲ **CRITICAL** Replace the parameters within the curly brackets from steps 90-103 with customized input files and computation resources.

Use the cutadapt tool (version 3.5) with customize settings to trim adapter and low-quality bases. Through this command, 6 bp of UMI and 5 bp of random RT tail will be removed from the sequence and appended into the reads name for downstream PCR duplicates removal. Meanwhile, low quality bases with a Phred quality of less than 20 will be trimmed, and reads with a final length of less than 15 bp will be discarded.

```
$ cutadapt -j {threads} -U 11 --rename='{id}_{r2.cut_prefix} {comment}' \
    --max-n=0 -e 0.15 -q 20 --nextseq-trim=20 -O 6 --pair-filter=both \
    -a AGATCGGAAGAGCACACGTCTG -A AGATCGGAAGAGCGTCGTGT \
    -o {output.inter_1} -p {output.inter_2} \
    {input.fq_1} {input.fq_2} >{output.report1}
$ cutadapt -j {threads} -m 15 -u -11 -n 5 -O 12 \
    -g ACACGACGCTCTTCCGATCT -a AGATCGGAAGAGCGTCGTGT \
    -G ACACGACGCTCTTCCGATCT -A AGATCGGAAGAGCGTCGTGT \
    -o {output.trimmed_1} -p {output.trimmed_2} \
    {output.inter_1} {output.inter_2} >{output.report2}
```

107. Run the falco tool (version 0.3.0) to check the quality of reads after adapter trimming.

```
$ falco -o {output_dir} {input_fq}
```

Sequence mapping and mutation calling ●Timing 4h

108. Download the reference sequence (.fa) and annotation (.gtf) file from public database. Download the human genome (GRCh38) form ENSEMBL database (http://ftp.ensembl.org/pub/current_fasta/homo_sapiens/dna/) and download the corresponding gene annotation from GENCODE database (https://www.gencodegenes.org/human/). Fetch and rRNA, tRNA, snoRNA and miRNA sequences from RNAcentral database and coalesce identical sequences into one.

109. Build bowtie2 (version 2.4.5) index for ribosomal RNA sequences (rRNA), small RNA sequences (tRNA + snoRNA + miRNA) and spike-in RNA sequences respectively.

```
$ bowtie2-build --threads {threads} {input.fa} {output.bt2_index}
```

110. Build STAR (version 2.7.10a) index for the reference genome.

```
$ STAR --runThreadN {threads} \
    --runMode genomeGenerate \
    --limitGenomeGenerateRAM=55000000000 \
    --sjdbGTFfile {input.gtf} \
    --genomeFastaFiles {input.fa}
    --genomeDir {output.star_index}
```

111. The trimmed reads are mapped to the spike-in RNA, rRNA and small RNA by the bowtie2 tool. Mapping output are filtered and sorted by the samtools (version 1.14-13). Firstly, trimmed reads are mapped to spike-in RNA sequences, and unmapped reads will be mapped to ribosomal RNA (rRNA) afterward. Similarly, unmapped reads from rRNA mapping step are mapped to the small RNA.

```
$ bowtie2 -p {threads} --no-unal --end-to-end -L 16 -N 1 --mp 5 \
    --un-conc {params.un} -x {params. bt2_index} -1 {input.r1} -2 {input.r2} | \
    samtools sort -@ {threads} --input-fmt-option 'filter=[NM]<=10' -O BAM -o {output.bam}
```

112. Unmapped reads from the output of bowtie2 aligner are mapped to the reference genome by STAR aligner.

```
$ STAR --runThreadN {threads}
        --genomeDir {params.star_index} \
                --readFilesIn {input.r1} {input.r2}
\
        --alignEndsType Local \
                --outFilterMatchNminOverLread 0.66
\
```

```
--outFilterMatchNmin 15 \

--outFilterMismatchNmax 5 \

--outFilterMismatchNoverLmax 0.2 \

--outFilterMultimapNmax 50 \

--outSAMmultNmax -1 \

--outReadsUnmapped Fastx \

--outSAMattrRGline ID:SACseq SM:sample LB:RNA PL:Illumina PU:SE \

--outSAMattributes NH HI AS nM NM MD jM jI MC \

--limitBAMsortRAM 8000000000 \

--outSAMtype BAM SortedByCoordinate \

--outFileNamePrefix {params.output_pre}
```

113. After mapping to the reference genome, reads with identical UMI and mapped to the same location are treated as PCR duplicates and dropped through the umicollapse tool (version 88d7949, https://github.com/Daniel-Liu-c0deb0t/UMICollapse).

```
$ umicollapse bam --two-pass -i {input.bam} -o {output.bam}
```

114. Detect mutation sites on the transcriptome. The mapping results are filtered by masking flags (UNMAP, SECONDARY, QCFAIL, DUP), so unmapped, secondary, low

quality, or duplicate reads are not used for downstream analysis. Then, the bam file is split into two parts, and forward and reverse reads will be analyzed respectively. The mutations of each site among all the samples are detected simultaneously with a customized tool, cpup (version 0.0.1, https://github.com/y9c/cpup). Sites with more than one (≥ 2) mutation reads in any library are reported in the output file. To improve the speed of mpileup tool, the genomic region can be cut into bins, and run in parallel.

```
$ (
  samtools mpileup --input-fmt-option 'filter=(flag & 99 == 99 || flag & 147 == 147)' -d 0 -Q 10 --reverse-del -f {params.ref} -l {input.bedT} {input.bam} |  cpup -H -S -f mut:2 | sed 's/\\t/\\t-\\t/3'
  samtools mpileup --input-fmt-option 'filter=(flag & 83 == 83 || flag & 163 == 163)' -d 0 -Q 10 --reverse-del -f {params.ref} -l {input.bedA} {input.bam} | cpup -H -S -f mut:2 | sed 's/\\t/\\t+\\t/3'
  ) >{output}
```

Site-by-site m^{6}A level quantification quality control ●Timing 20min

115. Filter sites by the mutation ratio. The mutation ratios in the un-treated libraries are used for masking SNP sites, and mutation ratios in the treated libraries are used for selected putative m^{6}A sites. List of criteria are applied to each site.

- Average mutation ratio of the input libraries should be < 2.5%, or the number of mutations is≤ 1.
- Average mutation ratio of the treated libraries should be ≥ 5%.

- There should be ≥ 3 mutations in the treated libraries.
- The minimum sequencing depth of all the libraries should be ≥ 20.

116. Gene/transcript name, relative position to transcript and motif sequence at RNA level can be annotated by variant-effect command in Python package, variant (version 0.0.16, https://github.com/y9c/variant).

```
$ variant-effect -i {input.sites_tsv} -r human --rna
```

117. Fine-tune the result of spike-in sequences alignment. There is only one base difference among two similar spike-in sequences, so bowtie2 aligner which is based on Burrows-Wheeler Transform algorithm cannot completely distinguish each sequence correctly. The BLAST aligner is used for fine-tuning the sequence alignment. Reads that mapped to spike-in RNA or unmapped to the genome are combined, and alignment to a degenerate reference spike-in sequence in which all m^6A motifs (5bp) are masked by N.

```
$ blastn -num_threads {threads} -max_target_seqs 1 -db {params.spike-in_with_N} \
  -query {input.fa} -outfmt 5 > {output.xml}
```

118. Fitting calibration curves using spike-in RNA data. Reads that mapped into the spike-in sequence are assigned to different motifs based on the flanking NN-NN sequence. The number of A site (mutation) and non-A sites (non-mutation) are counted, and the mutation ratio of each motif at 0%, 25%, 50%, 75% and 100% modification ratio can be calculated. A linear function is used for fitting the scaling factor (slope), which can be used for converting the mutation ratio into m^6A faction.

119. Estimate modification fraction of m^6A sites. The mutation ratio of each site is transformed by the scaling factor of the corresponding motif. Sites with a fraction greater than 20% can be regarded as a high modified one and will be used for functional analysis.

Troubleshooting

Troubleshooting advice can be found in **Table 3.2.**

Timing

Preparation of allyl-SAM and MjDim1

Day 1-3 preparation of allyl-SAM, steps 1-11

Day 4-9 preparation of MjDim1, steps 12-39

Performing SAC-seq

Day 1 Preparation of RNA samples. Step 40, Time depends on sample type

Day 2 Until the first ligation, steps 41-55. 3h, hands-on time ~1h

Day 3 Until iodine labeling, steps 56-73. 6h, hands-on time ~1h.

Day 4 Until the second ligation, steps 74-88. 3h, hands-on time ~1h.

Day 5 Finish the library preparation, steps 89-102. 3h, hands-on time ~1h

NGS sequencing

Step 103-105. Time depends. Typically, 1-2 days if the instrument is available.

Data analysis

Step 106-119. Data processing, 4.5 h.

Table 3.2 Troubleshooting table

Step	Problem	Possible reason	Solution
9	Multiple peaks shown in HPLC Cannot identify the desired product Low yield of desired product	Crude Allyl-SAM degraded	Lyophilize immediately after synthesis Determine the correct product peak via LC-MS
27	No peak shown in FPLC Low yield of enzyme	Discarded the flowthrough in step 25 Incorrect inoculation, proliferation or induction	Remember to keep the flowthrough Check turbidity via OD600
39	Low conversion	Enzyme activity impaired during purification	Keep the enzyme on ice whenever possible Use the appropriate time and speed when concentrating the protein, as is indicated in step 27

		LC-MS data variability	Perform technical replicates
40	Low yield after Box 1, option 1 Too high yield after Box 1, option 2	The total RNA is more fragmented than expected	Check total RNA quality with Bioanalyzer 2100, intact RNA should have a RIN of 10 Use RNase free reagents and consumables Snap-freeze cell line and tissue samples
Table 3.2. Troubleshooting table (continued)			
45	Fragment too long or short	Using non-intact starting RNA materials, e.g., FFPE	Perform pilot run as indicated in Box 2
103	Library concentration > 30 ng/μL Library concentration <	Over-amplification Under-amplification	Reduce cycle number for next library preparation

	1 ng/μL		Add another 4-6 cycles of PCR to the library using the same index primers or P5/P7 primers. Repeat steps 96-102 after amplification.
104	Library contains significant peaks under 150 bp	Incomplete removal of adaptor dimer, primer dimer, or primers	Repeat steps 96-102. Pay attention to EtOH percentage and complete drainage before drying
	Library size < 200 bp	Over-fragmentation	Perform pilot run as indicated in Box 2
	Library contains a secondary peak > 500 bp	Over-amplification	Reduce cycle number for next library preparation. Remove large fragments with AMPure XP beads, according to the

			manufacturer's instructions

Table 3.2. Troubleshooting table (continued)

105	Many reads mapped to E. coli or mycoplasma	Contamination	Use sterile reagents and consumables Check the starting biological samples

	Many reads mapped to rRNA or tRNA	Incomplete removal of rRNA or tRNA	If possible, follow the instructions in Box 1. Otherwise (e.g., FFPE), increase sequencing depths
	Reads contain high level of duplication	Over-amplification	Reduce cycle number for next library preparation
	Low mutation rate detected on spike-in probes	Low quality MjDim1 or allyl-SAM used	Ensure step 28-39 is performed Avoid freeze-saw cycles

Chapter 4 eTAM-seq

4.1 Introduction

eTAM-seq is inspired by bisulfite sequencing[168], which profiles m^5C in DNA via C-to-U conversion through chemical deamination. C is deaminated while m^5C retains, which could be read out as unmutated sites in NGS sequencing, in contrast to the modification-to-mutation paradigm demonstrated by m^6A-SAC-seq. The modification ratio can be quantified as 1-mismatch%. Cytidine deaminase derived base editor APOBEC1 has already been employed for m^6A-profiling in DART-seq[155]. Although the site-specificity of identified m^6A could be ensured by the adjacency of C in the DRACH motif, the accuracy of quantitation can be compromised due to the indirectness of the measurement. Therefore, the deamination of A to inosine (I), which base pairs with dA/dC/dG (preferentially dC)[169], while leaving the m^6A intact is highly desirable for m^6A profiling[169]. Escherichia coli tRNA adenine deaminase (TadA) converts A to I in the single-stranded anticodon loop of $tRNA^{Arg}$, making it a promising candidate for the ssDNA/RNA A-to-G base editor. Using directed evolution[170], TadA has been engineered to create several generations of base editors, with TadA8.x series showing 98-99% conversion[171–173]. This could be exploited for m^6A profiling if appropriate reaction conditions can be found.

Compared to GLORI[83] (see Chapter 1 Section 1.2.4 for a detailed introduction), its chemical deamination protocol saves the effort to express deaminases. However, our method requires much less starting material (several ng versus 200 ng mRNA) due to milder treatment, which is a highly valued aspect when investigating clinical samples, pre-implantation embryos and other RNA samples with limited availability.

4.2 Results and discussion

As shown in the schematic (**Figure 4.1a**), an ideal deaminase would achieve global deamination of A but not m^{6}A—all unmethylated A is converted into I; I base pairs with C and is read as G by reverse transcriptases. Persistent A corresponds to m^{6}A. When we placed a single A or m^{6}A in different sequence contexts, TadA8.20 deaminated A close to completion without acting on m^{6}A (**Figure 4.1b–e**). TadA8.20 preserved RNA integrity, as RNA pre- and post-enzymatic treatment produced the same amount of complementary DNA (cDNA) during reverse transcription. Deamination efficiency is critical for faithful detection of m^{6}A. We screened a series of assay conditions and found that temperature elevation from 37 °C to 44 °C or 53 °C markedly improved A-to-I conversion rates, especially in regions resistant to deamination at 37 °C.

We incorporated TadA8.20 into an RNA-sequencing (RNA-seq) workflow and developed eTAM-seq. We first assessed the efficiency and context dependence of the enzyme using synthetic RNA probes with A/m^{6}A flanked by two Ns (N = A, C, G or U). TadA8.20 attained a global A-to-I conversion rate of 99%, close to the efficiency offered by bisulfite treatment of C in RNA[174], and rejected m^{6}A completely (**Figure 4.1f**). Importantly, TadA8.20 efficiently deaminates A in all DRACH sequences (D = A, G or U; R = A or G; H = A, C or U), the consensus motif hosting m^{6}A modifications in eukaryotes.

We next investigated whether global A deamination enforced by TadA8.20 enables quantitative detection of m^{6}A. We synthesized additional RNA probes containing 25%, 50% and 75% m^{6}A at the N-flanked position (NNA/m^{6}ANN) and labeled them with unique molecular identi- fiers (UMIs). After incubation with TadA8.20, we detected 1.69 ± 0.03%, 26.1 ± 0.6%, 46.5 ± 0.3%, 73.9 ± 0.5% and 98.7 ± 0.1% A in probes hosting 0%, 25%, 50%, 75% and 100%

m^{6}A, respectively (**Figure 4.1g**). As the persistent A ratio correlates linearly with the extent of m^{6}A in synthetic RNA probes (r2 = 1.00; **Figure 4.1g**), we conclude that eTAM-seq quantifies m^{6}A at the site of interest. Collectively, we demonstrate that TadA8.20 is mild, robust, selective and insensitive to sequence contexts, paving the way for quantitative and base-resolution detection of m^{6}A in biological samples.

We applied eTAM-seq to 50 ng of mRNA extracted from HeLa cells. To reduce secondary structures and to facilitate downstream sequencing, we processed HeLa mRNA into ~150-nt fragments before incubation with TadA8.20. Capillary gel electrophoresis of RNA incubated with TadA8.20 for 3 h indicated that the size distribution of RNA remained unchanged with no noticeable sample loss (**Figure 4.1h**). Fragmented mRNA treated with or without TadA8.20 behaved similarly during library construction and RNA-seq, suggesting that the increased GC content[175], a consequence of global A deamination, poses minimal impact on cDNA synthesis, amplification and sequencing.

After adapter removal, reads were mapped to the transcriptome, allowing both A and G to be matched to genomic A sites. Given the reduced complexity of the transcriptome post-TadA8.20 treatment, we took conservative measures to ensure mapping accuracy: we accepted only reads ≥40 nt and discarded those that could be mapped to more than one genomic locus. Importantly, mRNA abundances reported by eTAM-seq are consistent with a published RNA-seq dataset[36] (Pearson's r = 0.84–0.85), indicating that eTAM-seq sustains gene expression information captured by canonical RNA-seq. The conversion rate of a given A in the HeLa transcriptome is highly reproducible across biological replicates (Pearson's r = 0.99; **Figure 4.1i**), suggesting that deamination efficiency is governed by intrinsic properties of RNA rather than random factors

introduced during sample preparation. Two well-characterized m^{6}A sites in human rRNA[10], at position 1832 in 18S rRNA and position 4220 in 28S rRNA, are cleanly detected by eTAM-seq (**Figure 4.1j**).

We next profiled m^{6}A in the HeLa transcriptome. To assess reproducibility of eTAM-seq, we carried out three biological replicates: replicate 1 (HeLa-1 and HeLa-IVT1), replicate 2 (HeLa-2 and HeLa-IVT2) and replicate 3 (HeLa-3 and HeLa-IVT3). We processed the three replicates separately and called out m^{6}A sites with exposed methylation levels ≥10% (methylation level × site accessibility ≥ 10%). We chose this cutoff to remove sites of extremely low methylation and accessibility. We identified 18,712, 19,439 and 15,159 m^{6}A sites from replicates 1, 2 and 3, respectively, 16,376 (88%), 16,600 (85%) and 13,151 (87%) of which were found in DRACH motifs (**Figure 4.2a**). As DRACH motifs were reported to host ~70% of m^{6}A sites[148], and only ~7% of all A sites, in mammalian transcriptomes, the observed hit distribution among DRACH and non-DRACH sequences supports the robustness of eTAM-seq. Of the identified hits, only 2,607 (14%), 2,719 (14%) and 661 (4%) are unique to individual replicates (**Figure 4.2a**). Sites common to the three replicates show highly consistent methylation levels (Pearson's r = 0.96; **Figure 4.2b**), confirming the reproducibility of eTAM-seq.

Next, we sequenced replicate 1 deeper to uncover the m^{6}A landscape in HeLa cells. We detected 80,941 m^{6}A sites in HeLa mRNA, of which 12,454 sites (15%) showed methylation levels >90%. Methylated sites are enriched in DRACH sequences (**Figure 4.2c**), highlighting the strong motif preference of the m^{6}A writer complex. m^{6}A constitutes 0.41% of all A subject to evaluation (**Figure 4.2d**). Given that the m^{6}A fraction in the HeLa transcriptome was determined

to be 0.2–0.6% by liquid chromatography–mass spectrometry[169] (0.55% in this work), we conclude that eTAM-seq captures most of the m^{6}A sites in the input RNA.

We extracted m^{6}A sites with exposed methylation levels ≥10% (69,834) for further analysis. Of these m^{6}A sites, 34,049 (49%) overlap with a published MeRIP-seq dataset[37] and account for 8,398 (52%) of peaks detected by MeRIP-seq. When we consider sites with ≥200 counts and ≥90% methylation, 89% of eTAM-seq hits are codiscovered by MeRIP-seq (**Figure 4.2e**). One MeRIP-seq peak covers 4.1 m^{6}A sites on average. The m^{6}A sites are enriched around stop codons with distribution across 5′-untranslated regions (UTRs), coding sequences and 3′-UTRs (**Figure 4.2f**), consistent with the transcriptome-wide m^{6}A distribution reported by orthogonal detection methods[55,126]. A clear DRACH motif emerges in sequences surrounding m^{6}A (**Figure 4.2g**). Further sequence context dissec- tion of these m^{6}A sites reveals 14.7% from GGACU, 11.8% from GAACU and 10.5% from AGACU, a distribution fully corroborated by miCLIP[148] (**Figure 4.2g**). We also overlapped eTAM-seq hits with m^{6}A-SAC-seq[84,135], a recently developed base-resolution m^{6}A detec- tion method. As m^{6}A-SAC-seq is more sensitive to GA sequences, we limited our comparison to DGACH hits. eTAM-seq detects 36,993 m^{6}A sites in DGACH, 28,067 (76%) of which are codiscovered by m^{6}A-SAC-seq (28,067 of 36,737, 76%; **Figure 4.2h**).

We next evaluated the quantitative feature of eTAM-seq using five representative sites in HeLa RNA—MALAT1_2515, _2577 and _2611 and TPT1_687 and _703—the methylation levels of which have been previously determined as 61%, 80%, 38%, 15% and 1% by site-specific cleavage and radioactive-labeling followed by ligation-assisted extraction and thin-layer chromatography (SCARLET)[110]. eTAM-seq reported m^{6}A fractions of 58%, 81% and 51% for MALAT1_2515, _2577 and _2611 (**Figure 4.2i**), respectively, in line with SCARLET results.

However, we observed much higher modification levels than previously reported for TPT1_687 and _703—42% and 77% (**Figure 4.2i**). As results from eTAM-seq (HeLa/IVT), eTAM-seq (HeLa/FTO) and m^{6}A-SAC-seq corroborate each other, we find these quantification results reliable. Discrepancies between our study and the SCARLET study may be explained by (1) differences in HeLa cells and (2) inefficient/off-target probe annealing in SCARLET.

Last, we extracted eTAM-seq signals from 18 additional loci and overlaid them with a published MeRIP-seq dataset[37] (**Figure 4.2i**). In all 20 cases, including MALAT1 and TPT1, eTAM-seq peaks were found in or close to MeRIP-seq peaks. Of the 20 transcripts inspected, 19 host multiple m^{6}A sites with several transcripts heavily methylated, such as ZBED5, MYC and CXCR4, which bear 31, 22 and 23 m^{6}A sites, respectively. These results, taken together, confirm that eTAM-seq is robust and reliable in capturing and quantifying m^{6}A sites in the whole transcriptome.

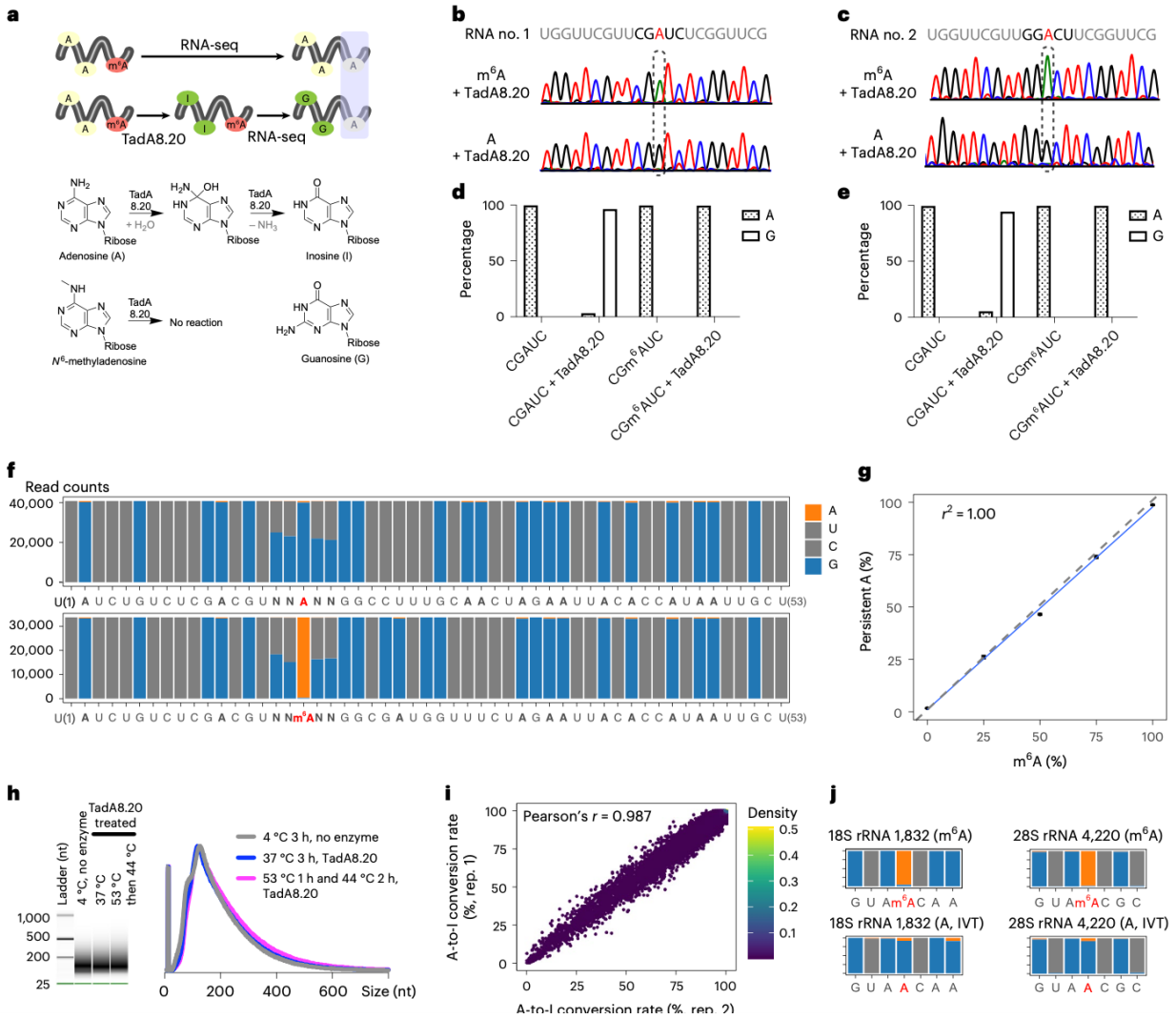


Figure 4.1 Global A deamination by TadA8.20

a, Proposed m⁶A detection scheme. TadA8.20 selectively converts A into I, without acting on m⁶A. I is recognized as G by reverse transcriptases. Persistent A post-TadA8.20 treatment corresponds to m⁶A. **b**,**c**, In vitro deamination of RNA probes hosting A or m⁶A in 'CGAUC' (b) and 'GGACU' (c) motifs by TadA8.20. Unmethylated and methylated RNA sequences were prepared through in vitro transcription using ATP and N6-methyl-ATP as starting materials, respectively. Treated RNA was reverse transcribed, amplified and subjected to Sanger sequencing. **d**,**e**, TadA8.20-

catalyzed A-to-I conversion rates in 'CGAUC' (d) and 'GGACU' (e) probes quantified by next-generation sequencing (NGS). **f**, Deamination of synthetic A/m^{6}A RNA probes by TadA8.20. The 53-nt RNA probes hosting NNANN and NNm^{6}ANN motifs were treated byTadA8.20. Deaminated RNA underwent RT and NGS. **g**, Correlation of persistent A signals captured by eTAM-seq and m^{6}A contents in RNA probes. **h**, Capillary gel electrophoresis analysis of fragmented HeLa mRNA treated with or without TadA8.20 at different temperatures for 3 h. RNA size distribution is plotted on the right. For eTAM-seq, RNA is incubated with TadA8.20 at 53 °C for 1 h followed by a 2-h treatment at 44 °C. Experiments were repeated independently with similar results. **i**,

(**Figure 4.1, continued**) Transcriptome-wide A-to-I conversion rates in two independent replicates (rep.). Of A sites with ≥100 counts, 10% were randomly sampled to make the scatter plot. Pearson's r was calculated for all A sites with ≥100 counts. j, Two m^{6}A sites in human rRNA. Positions 1829–1835 of 18S rRNA and positions 4217–4223 of 28S rRNA are plotted.

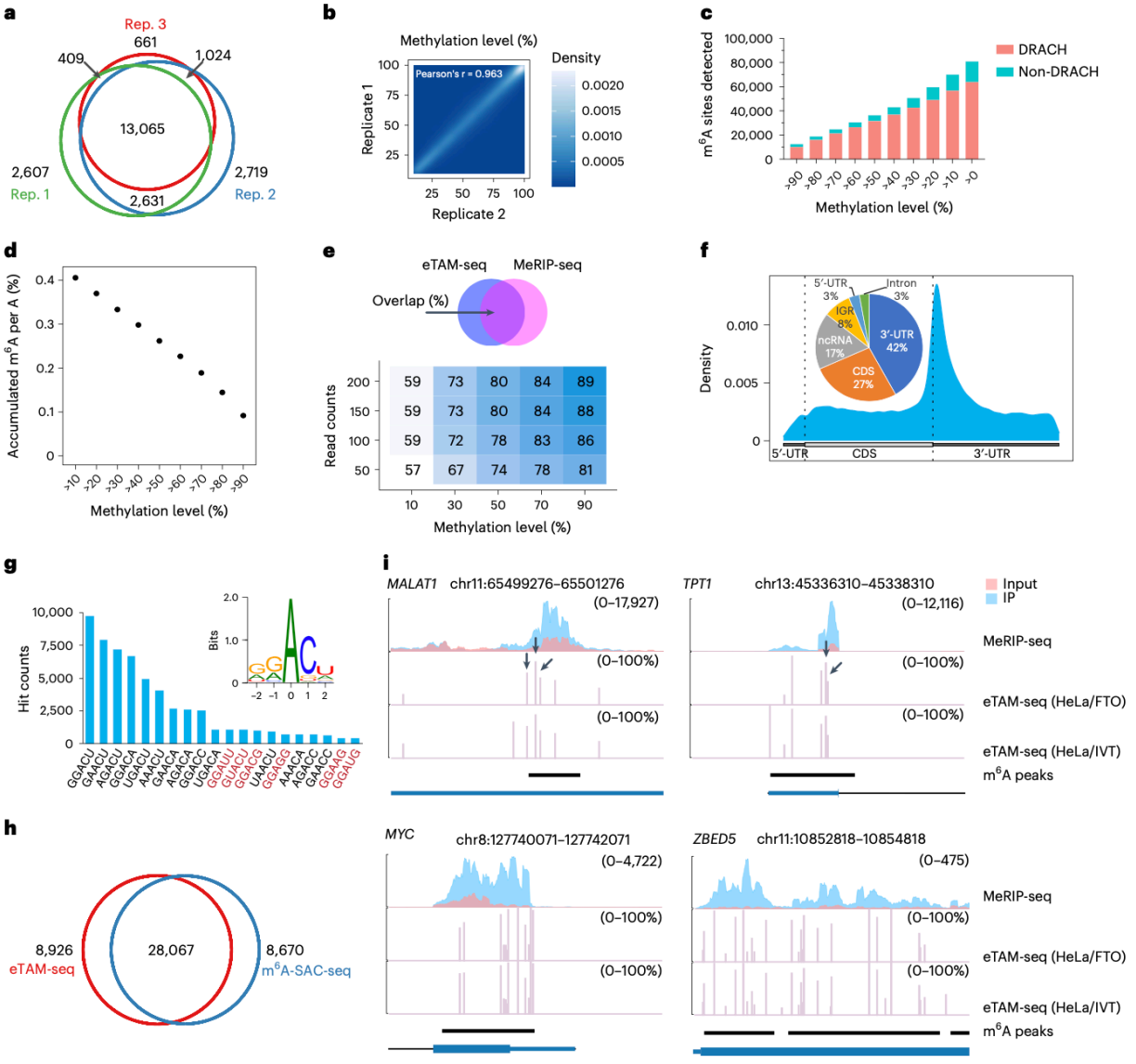


Figure 4.2 Transcriptome-wide m6A profiling in HeLa cells by eTAM-seq

a, Overlap analysis of m^{6}A sites identified in three biological replicates (Rep.) of eTAM-seq (HeLa/IVT). **b**, Methylation levels of common sites detected in eTAM-seq (HeLa/IVT1) and eTAM-seq (HeLa/IVT2). **c**, Hit distributions in DRACH and non- DRACH sequences at different methylation levels. **d**, Cumulative m^{6}A signals from highly methylated sites to lowly methylated sites (right to left). eTAM- seq-captured m^{6}A constitutes 0.41% of all A subject to evaluation in

the HeLa transcriptome. **e**, Overlap analysis of m^{6}A sites identified by eTAM-seq and peak clusters generated via MeRIP-seq. The overlap between eTAM-seq and MeRIP- seq increases with higher (**Figure 4.2, continued**) read depth and methylation levels. **f**, Metagene plot of transcriptome-wide distribution of m^{6}A. m^{6}A distributions across different RNA regions are provided in the inserted pie chart. CDS, coding sequences; IGR, intergenic region; ncRNA, noncoding RNA. **g**, Major sequence motifs hosting m^{6}A. DRACH motifs are in black and non-DRACH motifs in red. The consensus sequence hosting m^{6}A is inserted. **h**, m6A sites codiscovered by eTAM-seq and m^{6}A-SAC-seq. Hits in DGACU captured by both methods are subject to overlap analysis. The m^{6}A- SAC-seq dataset is under GEO accession no. GSE198246 (ref. 40). **i**, m^{6}A positions and fractions in MALAT1, TPT1, MYC and ZBED5. eTAM-seq signals are plotted as methylation levels (%) alongside MeRIP-seq peaks in normalized read coverage. Note that eTAM-seq (HeLa/IVT) has slightly higher coverage than eTAM-seq (HeLa/ FTO) and may therefore capture more m6A sites. MALAT1_2515, _2577 and _2611 and TPT1_687 and _703 are indicated by arrows. The coding sequence for TPT1 is on the minus strand of the genome.

4.3 Methods

Cell culture. HeLa cells and mESCs were purchased from American Type Culture Collection (ATCC). HeLa cells were grown in Dulbecco's modified Eagle's medium (DMEM; Gibco, catalog no. 11965092) supplemented with 10% fetal bovine serum (FBS; Gibco) and 1% 100× penicillin–streptomycin (Gibco). WT, control KO and Mettl3 conditional KO (cKO) mESCs were maintained in DMEM (Invitrogen) supplemented with 15% FBS (Gibco), 1% nucleosides (100×; Millipore), 1 mM l-glutamine (Gibco), 1% nonessential amino acids (Gibco), 0.1 mM 2-

mercaptoethanol (Sigma-Aldrich), 1,000 U/ml of leukemia inhibitory factor (Millipore), 3 μM CHIR99021 (STEMCELL Technologies) and 1 μM PD0325901 (STEMCELL Technologies). All cells were cultured at 37 °C under 5.0% CO2. Mettl3 cKO mESC lines were generated following previously reported methods[176]. Briefly, mESCs derived from Met- tl3flox/flox mouse blastocysts were transfected with 200 ng of PB-CAG-puromycin-P2A-CreERT2 and 100 ng of PBase by electropo- ration. After 24 h, electroporated cells were treated with 1 μg/ml of puromycin to generate stable Mettl3flox/flox;CreERT2 mES clones. To induce deletion, Mettl3flox/flox;CreERT2 ESCs were treated with 1 μg/ml of 4-hydroxytamoxifen (Sigma-Aldrich). These Mettl3 KO cells were cultured for 48 h before harvesting. Untreated Mettl3flox/flox;CreERT2 ESC cells were used as ctrl mESCs.

Western blotting. Cells were lysed in radioimmunoprecipitation lysis buffer (Pierce) supplemented with complete protease inhibitor cocktail (Takara). Lysates were boiled at 95 °C in NuPAGE LDS loading buffer (Invitrogen) for 10 min and then stored at –80 °C for use in the next step. A total of 30 μg of protein per sample was loaded on to 4–12% NuPAGE Bis–Tris gel and transferred to poly(vinylidene) membranes (Life Technologies). Membranes were blocked in 5% milk phosphate-buffered saline–Tween (PBS-T) for 30 min at room temperature, incubated in 1:1,000 (v:v) dilution of anti-METTL3 antibody (abcam, catalog no. ab195352) at 4 °C overnight, washed and incubated in 1:5,000 (v:v) dilution of goat anti-rabbit immunoglobuin G–horseradish peroxidase (HRP; abcam, catalog no. ab6721) for 1 h at rooom temperature. The membrane region lower than 50 kDa was used as a loading control and directly washed and incubated in 1:1,000 (v:v) dilution of anti-glyceraldehyde 3-phosphate dehydrogenase monoclonal antibody–HRP (CST, catalog no. 3683) for 1 h at room temperature. Protein bands were detected using

SuperSignal West Dura Extended Duration Substrate kit (Thermo Fisher Scientific) and FluroChem R (Proteinsimple).

Poly(A)$^+$ RNA extraction. Cells were cultured to 70–80% confluency, rinsed with 1× PBS (Gibco) and lysed by the direct addition of TRIzol reagent (Invitrogen). Total RNA was then collected following the manufacturer's protocol. Poly(A)$^+$ RNA was extracted from purified total RNA using Dynabeads mRNA DIRECT Purification Kits (Invitrogen).

Overexpression and purification of recombinant TadA8.20 protein. WT TadA and TadA8.20 fused to an amino-terminal, hexahistidine-tagged, maltose-binding protein (6×His-MBP) were cloned into a pET28a vector. A TEV (tobacco etch virus) protease cleav- age site (ENLYFQ|G) was installed between MBP and TadA variants. Expression plasmids will be deposited to Addgene. BL21 Rosetta 2 (DE3) competent cells were transformed with the recombinant plasmids and grown on Luria broth (LB) agar plates sup- plemented with 50 μg ml−1 of kanamycin and 25 μg ml−1 of chloram- phenicol. Successfully transformed bacteria were always cultured in the presence of 50 μg ml−1 of kanamycin and 25 μg ml−1 of chloramphenicol unless otherwise noted. Single colonies were inoculated into fresh LB medium and grown in an incubator shaker (37 °C, 220 r.p.m.) for 12–18 h. A 10-ml saturated start culture was used to inoculate 1 l of fresh medium. Bacteria were grown at 37 °C until the absorbance at 600 nm (OD600) reached 0.5. The culture was cooled down immediately to 4 °C and induced with 0.1 mM isopropyl β-d-1-thiogalactopyranoside (IPTG). Bacteria were cultured at 16 °C for an additional 20 h before pelleting by centrifugation at 4,000 g. Bacterial pellets were lysed by sonication in buffer A (50 mM Tris, 500 mM NaCl, 10 mM 2-mercaptoethanol and 10% (v:v) glycerol, pH 7.5). Lysed bacteria were clarified by centrifugation at 4 °C and 23,000 g. The supernatant was loaded

on to a Ni-NTA Superflow Cartridge (QIAGEN, catalog no. 30761), washed with 30 ml of buffer A supplemented with 50 mM imidazole and eluted with a gradient of imidazole from 50 mM to 500 mM in buffer A. The eluted protein was incubated with TEV protease and dialyzed in buffer A at 4 °C overnight. The protein mixture was reloaded on to a Ni-NTA Superflow Cartridge, washed with buffer B (50 mM Tris, 1 M NaCl, 10 mM 2-mercaptoethanol and 10% (v:v) glycerol, pH 8.0) and eluted by buffer B supplemented with 50 mM imidazole. Finally, MBP-free TadA8.20 was purified by size-exclusion chromatography (Enrich SEC 650 10 × 300 mm2 column; BioRad, catalog no. 7801650) and concentrated to approximately 4 mg ml−1. The column was bal- anced and eluted with buffer C (50 mM Tris, 200 mM NaCl, 10 mM 2-mercaptoethanol and 10% (v:v) glycerol, pH 7.5).

Preparation of A- and m^{6}A-bearing E. coli tRNA (Arg2, CGT) and RNA probes. DsDNA templates carrying T7 promoter were prepared by primer extension with two single-stranded DNA oligos. Unmethylated and methylated E. coli tRNA (Arg2, CGT), RNA no. 1 and RNA no. 2, were synthesized by IVT using T7 RNA polymerase. ATP and N6-methyl-ATP (TriLink, catalog no. N-1013) were supplied in the presence of UTP, CTP and GTP to synthesize unmethylated and methylated RNA, respectively. RNA was purified by E.Z.N.A Micro RNA kits (Omega Bio-Tek, catalog no. R7034) and quantified by NanoDrop One (Thermo Fisher Scientific).

Preparation of IVT transcriptomes. Modification-free control RNAs were prepared from HeLa cell and mESC total mRNA based on previously published protocols[125,144] with minor modifications. Oligo-dT(30)VN primer (TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN, 100 pmol) was annealed to 100 ng of purified poly(A)$^+$ RNA at 65 °C for 5 min. RNA was reverse transcribed in 20 μl of 1× RT buffer (Thermo Fisher Scientific, catalog no. EP0753, containing 50

mM Tris-HCl, pH 8.3, 75 mM KCl, 3 mM $MgCl_2$ and 10 mM dithiothreitol (DTT)) in the presence of 40 pmol of 5Bio-T7-TSO (/5Biosg/ ACTCTAATACGACTCACTATAGGGAGAGGGCrGrGrG), 1 mM of GTP, 5% (w:v) polyethyleneglycol (PEG) 8000, 0.5 mM of each dNTP, 5 mM RNaseOUT (Invitrogen, catalog no. 10777019) and 200 U of Maxima H^- Reverse Transcriptase (Thermo Fisher Scientific, catalog no. EP0753) under the following conditions: 42 °C for 90 min, 10 cycles of (50 °C for 2 min + 42 °C for 2 min) and 85 °C for 5 min. To 20 µl of RT reaction, 10 µl of RNase H (New England Biolabs (NEB), catalog no. M0297L), 70 µl of RNase-free H2O and 100 µl of Ultra II Q5 Master Mix (NEB, catalog no. M0544X) were added to make the second-strand synthesis mixture, which was incubated under the following conditions: 37 °C for 15 min, 95 °C for 1 min and 65 °C for 10 min. The reaction was purified with 160 µl (0.8×, v:v) of AMPure XP beads (Beckman Coulter, catalog no. A63882) following the manufacturer's directions. The purified and concentrated dsDNA was in vitro transcribed in 1× T7 Reaction Buffer (NEB, catalog no. E2040S, containing 40 mM Tris-HCl, 6 mM $MgCl_2$, 1 mM DTT and 2 mM spermidine, pH 7.9) with 10 mM of each NTP and 2 µl of T7 RNA Polymerase Mix (NEB, catalog no. E2040S) in 20-µl volume at 37 °C overnight. The IVT mixture was further treated with TURBO DNase (Invitrogen, catalog no. AM2238) and purified by acid–phenol chloroform (Invitrogen, catalog no. AM9722) extraction and ethanol precipitation to yield 2.5–10 µg of IVT RNA.

In vitro deamination of RNA probes by TadA8.20. All reactions were carried out in a deamination buffer (50 mM Tris, 25 mM KCl, 2.5 mM MgCl2, 2 mM DTT and 10% (v:v) glycerol, pH 7.5) in the presence of 10 U of SUPERase•In RNase Inhibitor (Thermo Fisher Scientific, catalog no. AM2694). RNA was always preheated to 95 °C for 3 min and immediately cooled down before use. To assay deaminase activity on the natural substrate E. coli tRNA, 200 ng of

RNA and 100 nM WT TadA or TadA8.20 were incubated at 37 °C for 1 h. For a typical deamination assay on RNA probes, 10 ng of RNA was incubated with 10 μM TadA8.20 in 20 μl of deamination buffer at 37 °C for 3 h. All reactions were quenched by incubating at 95 °C for 10 min. Temperature and pH were adjusted to identify the optimal condition for in vitro deamination.

RT–PCR. To convert RNA into cDNA for sequencing purposes, a 2-μl deamina- tion reaction was aliquoted, to which 0.5 μl of 50 μM RT primer was supplied. Primer annealing was enabled by heating up the mixture to 95 °C for 3 min, cooling down at a ramping rate of 2 °C/s and incubat- ing at 25 °C for 2 min. To the reaction, 0.5 μl of GoScript reverse tran- scriptase (Promega, catalog no. A5003) was added together with 2 μl of 5× GoScript RT buffer, 1 μl of 25 mM MgCl2, 0.5 μl of 10 mM dNTPs and 3.5 μl of nuclease-free H2O. The RT reaction was incubated at 42 °C for 1 h and then quenched at 65 °C for 20 min. To a 20-μl PCR or quantitative (q)PCR reaction (EvaGreen qPCR Master Mix, Biotium, catalog no. 31041), 0.1 μl of the RT reaction was supplied as the template. A typical PCR program includes initiation at 95 °C for 3 min, 30 cycles of amplification (denaturing at 95 °C for 10 s, annealing at 60 °C for 10 s followed by extension at 72 °C for 20 s) and final extension at 72 °C for 5 min. The qPCR reactions were performed on a CFX96 Real-Time PCR System (BioRad).

Overexpression and purification of recombinant FTO. The human FTO gene was cloned into a pET28a vector and transformed into BL21(DE3) cells (NEB). Successfully transformed bacteria were cultured at 37 °C in 2× YT broth (Teknova) to an OD of 0.8–1.0. The culture was cooled to 16 °C and supplemented with 0.1 mM IPTG (Sigma-Aldrich), 10 μM $ZnSO_4$ (Sigma-Aldrich) and 2 μM $(NH_4)_2Fe(SO_4)_2$ (Sigma-Aldrich). Bacteria were cultured overnight at 16 °C after induction. Bacteria were collected via centrifugation and lysed in buffer D (300 mM

NaCl (Thermo Fisher Scientific), 50 mM imidazole (Thermo Fisher Scientific) and 50 mM of Na_2HPO_4 (Sigma-Aldrich), pH 8.0). The lysate was clarified by centrifugation and loaded on to a nickel column (Ni Sepharose 6 FF, Cytiva), washed with buffer E (150 mM NaCl, 25 mM imidazole and 10 mM Tris-HCl (Invitrogen), pH 7.5) and eluted with buffer F (150 mM NaCl, 250 mM imidazole and 10 mM Tris-HCl, pH 7.5). The eluate was loaded on to an anion-exchange column (SOURCE 15Q, Cytiva) and fractionated with 0–50% buffer G (1.5 M NaCl, 20 mM Tris-HCl, pH 7.5) over 30 min. The resulting protein was concentrated and buffer exchanged using a 10-kDa MWCO filter (Cytiva, catalog no. 28932296) before being flash frozen in 30% glycerol for future use

Library preparation for eTAM-seq. Poly(A)$^+$ RNA purified from HeLa cells or mESCs (50 ng) and IVT control RNA (25 ng) were depleted of poly(A) tails, end-repaired and ligated to 3′-adapters following a previously published protocol[135]. Briefly, RNA was annealed to 100 pmol of oligo-dT (Thermo FIsher Scientific, catalog no. SO132) and digested by 5 U of RNase H (NEB, catalog no. M0297L) at 37 °C for 30 min. Without purification, the RNA was fragmented in 1× zinc fragmentation buffer (10 mM $ZnCl_2$ and 10 mM Tris-HCl, pH 7.5) at 70 °C for 5 min. The fragmentation reaction was quenched by the addition of 10 mM EDTA, then treated with 50 U of T4 PNK (NEB, catalog no. M0201L) at 37 °C for 1 h. The end-repaired RNA was purified by RNA Clean & Concentrator (Zymo) kits, mixed with 2% (w:w) spike-in probes and ligated to 20 pmol of 3′-adapter (/5rApp/AGATCGGAAGAGCGTCGTG/3Bio/) using 400 U of T4 RNA ligase 2, truncated KQ (NEB, catalog no. M0373L) at 25 °C for 2 h, then 16 °C overnight. Excess adapters were first digested by 5′-deadenylase (NEB, catalog no. M0331S) at 30 °C for 1 h, then by RecJf (NEB, catalog no. M0264L) at 37 °C for 1 h. The ligated poly(A)$^+$ RNA was divided into two halves and designated to the FTO$^-$ and FTO$^+$ groups. The FTO$^-$, FTO$^+$ and IVT groups were

immobilized on Dynabeads MyOne Streptavidin C1 (Invitrogen, catalog no. 65002). The FTO^+ group was demethylated by incubating with 200 pmol of FTO in 1× FTO reaction buffer (2 mM sodium ascorbate (Sigma-Aldrich), 65 μM ammonium iron(II) sulfate (Sigma-Aldrich), 0.3 mM α-ketoglutarate (Sigma-Aldrich), 0.1 mg ml−1 of bovine serum albumin (NEB) and 50 mM Hepes-KOH, pH 7.0) supplemented with 10% (v:v) SUPERase•In RNase Inhibitor (Invitrogen, catalog no. AM2696) at 37 °C for 1 h. The beads were washed by resuspension in 0.1% PBS-T (1× PBS (Gibco) supplemented with 0.1% (v:v) Tween-20 (Sigma-Aldrich)), once in binding/wash buffer (1 M NaCl, 0.5 mM EDTA, 5 mM Tris-HCl, pH 7.5) and twice in 10 mM Tris-HCl, pH 7.5, consecutively. The three RNA samples were then deaminated on beads with 200 pmol of TadA8.20 in the deamination buffer (50 mM Tris, 25 mM KCl, 2.5 mM $MgCl_2$, 2 mM DTT and 10% (v:v) glycerol, pH 7.5) supple- mented with 10% (v:v) SUPERase•In RNase Inhibitor at 53 °C for 1 h. This reaction was repeated twice at 44 °C for 1 h each by draining the supernatant on a magnetic rack and resuspending the beads in fresh reaction mixtures, lasting 3 h in total. The beads were washed sequentially by resuspension in 0.1% PBS-T (v:v), once in binding/wash buffer and twice in 10 mM Tris-HCl, pH 7.5. RNA was annealed to 2 pmol of RT primer (AGACGTGTGCTCTTCCGATCT) at 70 °C for 2 min, then reverse transcribed in 1× RT buffer (Thermo Fisher Scientific, catalog no. EP0753, containing 50 mM Tris-HCl, pH 8.3, 75 mM KCl, 3 mM MgCl2 and 10 mM DTT) with 5 mM RNaseOUT (Invitrogen, catalog no. 10777019) and 200 U of Maxima H− Reverse Transcriptase (Thermo Fisher Scientific, caatalog no. EP0753) at 50 °C for 1 h. The cDNA was released by boiling the beads in 0.5% (v:v) sodium dodecylsulfate for 10 min. The eluate was purified by DNA Clean & Concentrator (Zymo) kits. Purified cDNA was ligated to cDNA Adapter (/5Phos/NNNNNN AGATCGGAAGAGCACACGTCTG/3SpC3/) using 30 U of T4 RNA ligase (NEB) at 25 °C overnight, following a pre- viously published

protocol14. The reaction was purified again by DNA Clean & Concentrator kits and then PCR amplified with NEBNext Ultra II Q5 Master Mix (NEB, catalog no. M0544X) and NEBNext Unique Dual Index Primer for Illumina (NEB, catalog no. E6440S) following the manufacturer's directions. Typically, 10–11 cycles of PCR were carried out to generate enough DNA. The resulting library was purified by AMPure XP beads (Beckman Coulter, catalog no. A63882) following the manufacturer's directions and submitted for NGS.

Site-specific amplification and barcoding of HeLa mRNA and IVT samples. Purified HeLa poly(A)$^+$ RNA (300 ng) was depleted of poly(A) tails, end-repaired, ligated to 3′-adapters and immobilized on Dynabeads MyOne Streptavidin C1 (Invitrogen) as described in 'Library preparation for eTAM-seq'. RNA was deaminated and reverse transcribed following a similar protocol for NGS library construction. Specifically, RNA samples were deaminated on beads with 200 pmol of TadA8.20 in the deam- ination buffer (50 mM Tris, 25 mM KCl, 2.5 mM MgCl2, 2 mM DTT and 10 % (v:v) glycerol, pH 7.5) supplemented with 10% (v:v) SUPERase•In RNase Inhibitor at 37 °C for 1 h. This reaction was repeated twice by draining the supernatant on a magnetic rack and resuspending the beads in fresh reaction mixtures, lasting 3 h in total. The beads were washed by resuspension in 0.1% PBS-T (v:v), once in binding/wash buffer and twice in 10 mM Tris-HCl, pH 7.5, consecutively. The cDNA was eluted by boiling the beads in DNase-free water (Invitrogen) and transferred into new tubes immediately. Sites of interest were PCR amplified from the eluted cDNA using transcript-specific primers. To demonstrate m^{6}A quantification with limited input, the same protocol was applied to 50 ng, 5 ng or 500 pg of HeLa total RNA. The resulting cDNA was split into two halves to amply ACTB and EIF2A fragments. Both rounds of PCR were set up using EvaGreen qPCR Master Mix. The first-round PCR was carried out at a 20-μl scale with cDNA gener- ated from different amounts of mRNA or total RNA,

and 0.5 μM forward and reverse primers. Primers were designed to recognize sequences post-deamination, leaving out 10- to 20-nt sequences surrounding the target m^{6}A sites. PCR reactions were analyzed by agarose gel elec- trophoresis. Then, 1 μl of PCR products was subjected to enzymatic cleanup (exonuclease I and shrimp alkaline phosphatase; NEB, catalog nos. M0293 and M0371) and Sanger sequencing. Gel purification was performed for PCR reactions that did not yield bright and single bands. The second-round PCR was carried out only for small-scale amplicon deep sequencing. To a 20-μl reaction, 1 μl of the first-round PCR reaction was supplied as the template together with 0.5 μM Illumina P7 and P5 index primers. Barcoded PCR products were pooled, gel purified using QIAquick Gel Extraction Kit (QIAGEN, catalog no. 28706) and sequenced on an Illumina MiSeq instrument.

Data preprocessing of eTAM-seq. Adapters were removed from raw eTAM-seq data by Cutadapt v.1.18. The 6-nt random barcodes at the 5′-end of R2 were extracted by the subcommand extract from UMI-tools56 v.1.1.1. R2 reads longer than 39 nt were used for further analysis. For HeLa samples, reads were first mapped to human rRNA sequences using HISAT-3N57 v.2.2.1 (--time --base-change A,G --no-spliced-alignment --no-softclip --norc --no-unal --rna-strandness F). The remaining non-rRNA reads were mapped to the human genome (hg38) and the GENCODE v.27 gene annotation (--time --base-change A,G --repeat --mit 1000 --bowtie2-dp 0 --no-unal --rna-strandness F). For mESC samples, reads were first mapped to mouse rRNA sequences using HISAT-3N v.2.2.1, with the remaining non-rRNA reads mapped to the mouse genome (mm10) and the GENCODE v.M25 gene annotation. We used the same HSAT-3N parameters for HeLa cell and mESC data. Only uniquely mapped reads were kept and deduplicated by the subcommand dedup from UMI-tools. To apply the statistical models for m^{6}A detection and quantification, we used high-quality datasets in which RNA fragments poorly processed by TadA8.20 (>50%

unconverted A) were eliminated. Customized scripts were developed to count converted and unconverted As across the whole genome.

Detection and quantification of m^{6}A from eTAM-seq data Detection with IVT controls. We first estimated sample-specific conversion rates for mRNA and IVT samples using most of the A sites, based on the hypothesis that most A sites are unmethylated and accessible to TadA8.20. We plugged in the sample-specific conversion rate to normalize the A and G counts observed at individual A sites in the mRNA sample and calculated their apparent methylation levels using a maximum likelihood estimator based on a binomial model. Meanwhile, we estimated deaminase accessibility for each A site using adjusted A and G counts reported by the IVT sample. We fitted a linear model between site accessibility and total counts of A, G and A + G observed at individual A sites to shrink accessibility estimates and reduce esti- mation bias. The model was trained using 2,000 randomly sampled sites with 10-fold crossvalidation before being applied to predict site accessibility. We adjusted the apparent methylation levels calculated from the mRNA data with site accessibility and obtained the final esti- mated methylation levels (true methylation levels). Methylation sites are defined as A sites that: (1) have ≥10 read counts in both mRNA and IVT samples; (2) pass Fisher tests with a false discovery rate <0.05; and (3) show exposed methylation levels (estimated methylation level × site accessibility) ≥ 10%.

Detection with FTO controls. We first estimated sample-specific conversion rates similar to 'detection with IVT controls'. As FTO partially demethylated m^{6}A, we further estimated an upper boundary for FTO efficiency in FTO-treated samples. Assuming a binomial model for both mRNA and FTO-treated mRNA, we jointly estimated site acces- sibility and methylation levels

using a maximum likelihood approach. Methylation sites are defined with the same criteria as described in Detection with IVT controls.

Chapter 5 Single-cell m^{6}A sequencing

5.1 Introduction

Single-cell RNA sequencing (scRNA-seq) has been the most impactful development in RNA-seq in the past decade, if not in genomics as a whole. There are two major motives for pursuing single-cell level genomics/transcriptomics: 1. Limitations in sample availability. The very first scRNA-seq paper by Tang et al. in 2009[177] was performed on only one mouse 4-cell stage blastomere. Pre-implantation embryos remain to be precious sample up to this date. Extremely low-input detection can also be used for detecting rare precancerous cells/cell-free nucleic acids in body fluids for early cancer diagnosis[178]. 2. Individual cell identity may be hidden under bulk information. Although cells have been categorized by morphology and surface markers for a long time, a decade of progress in scRNA-seq has already revealed new cell types in brain[179–181], blood & immune cells[182,183], and various other systems[184]. This is because cell identity is essentially a reflection of differential gene expression (DGE), while morphological change and membrane protein expression is a consequence. Bulk assays which typically collect samples based on morphology and surface markers, will average out any hidden heterogeneity. Heterogeneity is also a key feature of tumor cells due to their genomic instability[185].

These aspects may also hold true for RNA modifications. The m^{6}A stoichiometry measured by the bulk profiling methods we discussed in Chapters 1-4 could be a reflection of uniform distribution within each cell, but also could be a statistical average of differential distribution patterns across the cell population. This could be answered by the development of single-cell or low-input m^{6}A profiling methods. Up to this date, two methods dedicated to this aim has been published:

scm^{6}A-seq[186]: As the name suggests, this method is based on MeRIP-seq/m^{6}A-seq, the antibody-based m^{6}A profiling. Mouse oocytes and blastomeres were collected individually in 96-well plates, lysed, RNA-fragmented, and ligated to barcoded 3' adaptors. This step appends a hexamer sequence (cell barcode) unique to each cell covalently to all RNA fragments. Then the barcoded RNAs from all cells were mixed and immunoprecipitated by the anti m^{6}A-antibody. Because mouse rRNA is also m^{6}A methylated and consists of more than 85%[187] of total RNA, the reads will be mapped almost exclusively to known rRNA sites unless the rRNA is depleted. Therefore, the authors incorporated T7 promoter sequences via PCR amplification after RT and performed in vitro transcription using T7 RNA polymerase[188]. This regenerated the RNA (amplified RNA, also known as aRNA) from cDNA which was then subjected to commercialized RNase H/complementary probe-based rRNA depletion. The remaining uncleaved RNAs were reverse transcribed again and PCR amplified for sequencing on Illumina platforms. The authors found better classification of oocytes and blastomeres from different cell stage when scRNA-seq data was overlayed with scm^{6}A-seq and recapitulated the asynchronous development of the two 2-cell blastomeres by changes in their m^{6}A methylation profiles.

scDART-seq[156]: Already briefly introduced in Chapter 1 Section 1.2.4, scDART-seq features the combination of DART-seq and more advanced scRNA-seq techniques, namely plate-based Smart-seq2[125] and droplet-based 10x Genomics Next GEM 3'[157]. Smart-seq was first invented in 2012[123] and has been continuously optimized to yield Smart-seq2[125] and 3[189] in later years. Although still plate-based and therefore limited to scales of several thousand cells, compared to ligation-based RNA-seq (broadly known as TruSeq, a commercialized name used by Illumina), Smart-seq features low-input full-length mRNA-seq by exploiting the terminal deoxynucleotidyl transferase (TdT) activity of Moloney murine leukemia virus (MMLV, or MoMLV/M-MLV)

RTase. This enzyme adds consecutive 3 dCs to the 3' end of its cDNA product. If adding a probe (known as template switch oligo, TSO) with consecutive 3 rGs (or rGrG+G, where +G = 2'-O,4'-C-methyleneguanosine, also known as locked nucleic acid or LNA[190]) to anneal at the cDNA 3' end, MMLV RTase will extend the cDNA by the sequence of TSO using its DNA-dependent DNA polymerase activity. This process, known as template-switch (TS), effectively appends any desired sequence to cDNA 3' end without ligation. This is preferred over single stranded ligation because ligation is prone to sequence/secondary structure bias at the joining ends[166], thus inefficient for longer nucleic acid substrates like full length mRNA, due to the presence of probabilistically more secondary structure in longer sequences. In addition, lower concentrations in low-input ligations reduce the collision chance of second-order reaction kinetics, further limiting the ligation efficiency[191].

Initially developed in 2015 as inDrop[192] and Drop-seq[193], droplet-based scRNA-seq represents another great advance in single cell technologies. Cells were individually encapsulated in water-in-oil (W/O, oil = perfluoro-1-octanol, PFO) emulsion droplets, which contain barcoded RT primers. In Drop-seq, mRNAs are mixed while annealed to the primer and RT reaction is done in tubes. In the more advanced, currently most popular commercialized protocol of 10x Genomics Next GEM 3', RT reaction is performed in-droplets to avoid barcode swapping. RT reactions used in both methods are based on the template-switching RT used in Smart-seq, but only the 3' of transcripts are retained for enrichment of barcodes. Droplet-based scRNA-seq harbors over 10s of thousands of cells in each run, and developments in commercialized engineered MMLV RTase mutants are pushing the number even higher[194].

Using these scRNA-seq techniques, scDART-seq found sheer heterogeneity in HeLa cells' mRNA methylation profiles, where more than 80% of the sites only appear in 20% of the cells. To quote from the paper[156], "While most mRNAs contain many total m^{6}A sites across all cells in the population, mRNAs are methylated at an average of only 1–3 sites in each cell, and individual sites are methylated in only 4.5% of cells on average". These findings await cross-validation from other site-specific single cell m^{6}A sequencing methods. And as previously discussed, DART-seq is inherently unsuitable for clinical frozen/FFPE tissue samples due to the overexpression of recombinant proteins. Therefore, we seek to develop a single-cell/low-input version of m^{6}A-SAC-seq/eTAM-seq.

5.2 Results and discussion

We first seek to apply the plate-based small RNA scRNA-seq protocols[195,196] to m^{6}A-SAC-seq. The advantage of this design is that small RNAs were also barcoded via ligation, which best accommodate to the bulk m^{6}A-SAC-seq protocol. However, small RNAs are inherently easier to ligate due to the short sequence being relatively free of secondary structures. If longer RNAs are kept intact, the ligation reaction is virtually selective for small RNAs (except for the 5.8S rRNA, which could be blocked by adding a complementary probe. See methods section). Therefore, for an adapted design that featured in-plate cell lysis, fragmentation, end-repair and ssRNA ligation (**Figure 5.1a**) with the aim to cover mRNA, the center problem would be rRNA depletion. Commercialized complementary probe-based rRNA depletion was not ideal in this application: Biotinylated probe pull-down based approach typically covered less rRNA sequence, because the pull-down of any region in an intact rRNA was sufficient to pull down the entire molecule. However, this was not true for fragmented rRNA. RNase H digestion based approach, on the other

hand, left the already ligated 3' remnant of the rRNA fragments in the library even after digestion, because the enrichment and pooling of ligated RNAs was achieved by the pull-down of biotinylated 3' adaptors. Therefore, the depletion efficiency of both methods was expected to be limited. Regarding the in vitro transcription (IVT)-RNase H based depletion method devised by scm^{6}A-seq[186], firstly it was published after our investigations; secondly, repeated conversion from RNA to cDNA to aRNA to cDNA again, was not only tedious but also would result in sample loss, which translated to more biased library profile after PCR amplification and thus less transcriptome coverage. We reasoned that the best approach was to not deplete rRNA itself, but rather deplete the dsDNA library with sequences corresponding to rRNA (termed rDNA) after PCR amplification. This had two advantages: 1. During the enzymatic and chemical treatment steps of m^{6}A-SAC-seq, the rRNA remained as a carrier nucleic acid which would reduce non-specific sample loss during these processes (for example, adsorption to plasticwares/digestion by environmental RNase). 2. Depletion happened after PCR step, which allowed for application of established protocols designed for bulk DNA/RNAs. Indeed, CRISPR-Cas9 system has already been employed for the purpose of removing undesired sequences in bulk RNA-seq libraries[197,198]. In this method, termed depletion of abundant sequences by hybridization (DASH), guide RNA (gRNA) sequence was designed computationally to cover undesired sequences which led to cleavage by Streptococcus pyogenes Cas9 nuclease. We optimized the gRNA design to increase the rRNA sequence coverage, which resulted in up to 75% of rRNA removal in four representative loci by qPCR assessment (**Figure 5.1b**). Then we subjected the resulting RNA to sequencing and mapped to the transcriptome. We used a commercialized RNA-seq kit component (ZapR v2 from Takara SMARTer Stranded Total RNA-Seq Kit v2)[199], which was used for rDNA depletion yet did not

explicitly state its mechanism of action, as a control. The results showed that our optimized sequences achieved better rRNA depletion (**Figure 5.1c**) and mapping (**Figure 5.1d**).

Despite showing 50-75% depletion in four representative loci, sequencing result showed that globally 60% of rRNA remained (**Figure 5.2a**). This was due to the presence of several Cas9-resistent peaks (**Figure 5.1d**) across the rRNA sequences. We therefore reinforced the gRNA sequences targeting these regions by adding another set (set #2, see methods and appendix, **Table S2**) of gRNAs. We screened through the mixing ratio between set #1 and 2, and the reaction time, purification conditions (**Table S2**) for optimization. One interesting finding is that the addition of proteinase K after reaction significantly improved library recovery, possibly due to Cas9 binding to dsDNA interfering with PCR. Using the optimized condition, we performed low-input m^{6}A-SAC-seq on three biological replicates of mouse meiosis II (MII) stage oocytes, each corresponds to approximately 5 ng of total RNA. The result showed optimal mapping ratio and rRNA depletion (**Figure 5.2b**). However, the duplication was relatively high. This was most likely a result of inherent ligation bias of T4 RNA ligases. Such bias was magnified when limited amount of sample was used. 3,102, 4,910, and 11,923 high confident m^{6}A sites were identified in three biological replicates, respectively (**Figure 5.2c**). The motif of these sites also fell in the DRACH consensus sequences (**Figure 5.2d**). However, the overlap sites between three replicates were low, accounting for only <1% of all sites identified. This could be explained by the high duplication ratio, where biased ligation resulted in stochastic capturing of transcripts other than a limited set of most enriched (the overlapping 121 sites) ones.

We reasoned that the limited performance of ligation-based approach could be improved by circumventing the ligation. Indeed, many m^{6}A sites were lowly methylated. When multiplied by

the conversion ratio of profiling methods, it was unlikely that those mutations were still observable under the several-thousand-cell scale of plate-based protocols. Therefore, we seek to develop a protocol based on droplet-based scRNA-seq. The droplet-based methods typically perform RT either before pooling or right after, because the barcoding step that must happen before pooling was achieved via RT with barcoded primers. However, this contradicted with the need to perform enzymatic/chemical labeling before RT and after barcoding, which was required for modification readout. We reasoned that this dilemma could be solved by transferring the barcoding information on the RT primer to the RNA strand, via a DNA-dependent-DNA synthesis reaction which used the RNA as the primer for extension. This biochemical reaction could be achieved by E. coli DNA polymerase large subunit (also known as Klneow fragment). The 3' overhang of poly(A) tails should also be trimmed before primer extension, which could be achieved by ssRNA exonucleases. After screening we found E. coli RNase I and Exonuclease T to be both capable of this function. The biotinylated uridine derivative was also incorporated via the addition of biotin-11-dUTP in the primer extension step, which eased the downstream purification and enzyme-chemical labeling of m^{6}A-SAC-seq (**Figure 5.3a**). Another advantage of this design was that the barcoding step was readily accommodated into the commercialized 10x Genomics Next GEM 3' protocol. We also designed a different approach based on Drop-seq[193], where mRNAs were annealed to barcoded primers and pooled without RT. In Drop-seq, the RT reaction was performed right after. The paper reasoned that barcode swapping was unlikely if large volume was used, due to the annealing being a second order reaction dependent on the product of the concentrations which was inverse-square proportional to the volume. In order to ensure barcode fidelity even after enzymatic treatment at 50 ºC, we removed all the denaturing steps (**Figure 5.3b**).

The preliminary results showed perfect mapping of untreated mRNA reads to the transcriptome, yet the conversion rate in the treated group remained low even after efforts of optimization. This could be a result of using MMLV-based RTases which have higher fidelity than HIV-1 RT. However, such difference showed in probe reads were still not explanatory for the discrepancies in transcriptome reads. We proposed that this could be a result of either EJC dependent m^6A depletion in shorter reads (which mapped exclusively to 3' UTR), or RNA secondary structure induced labeling efficiency decrease in longer reads due to the preservation of full RNA sequences from 3' end. More optimizations in the fragment length, labeling conditions, or the adoption of a completely different methodology (see Chapter 7) were among the future directions.

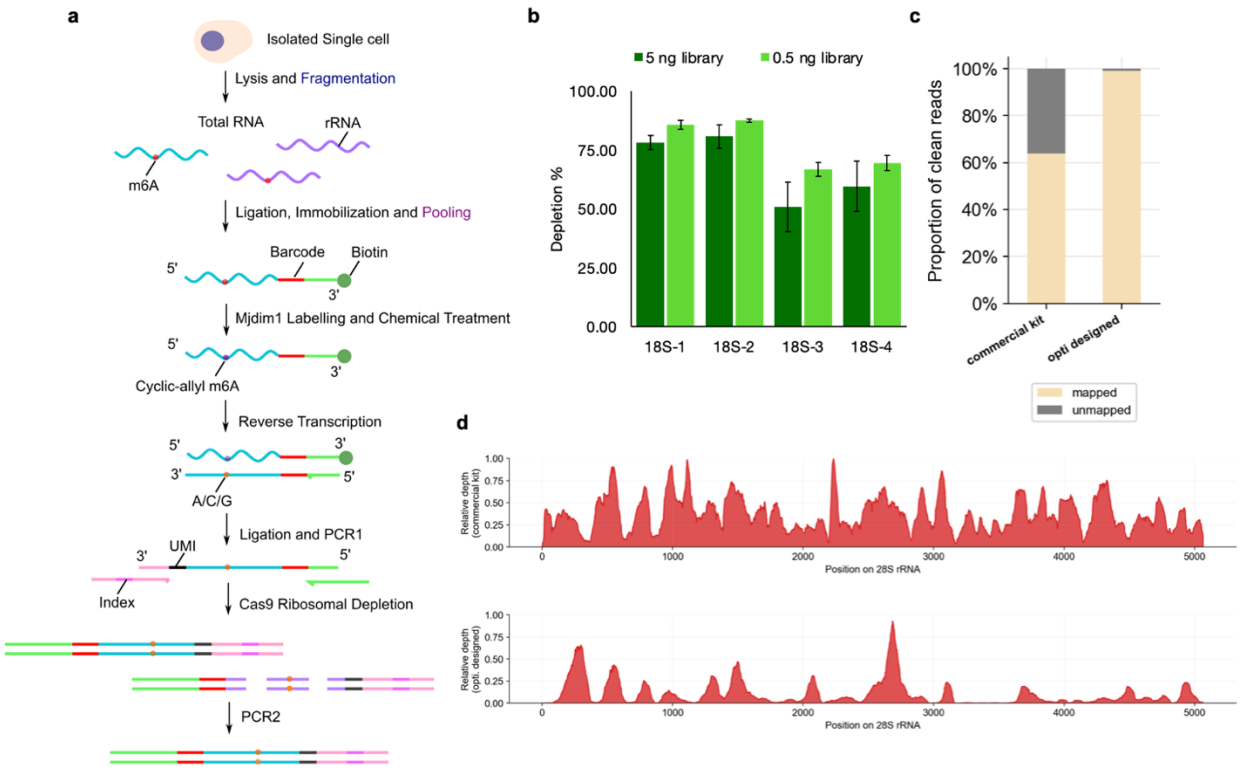


Figure 5.1 Plate-based low-input m⁶A-SAC-seq

a, Schematic representation of the plate-based design. Cells were sorted individually into wells of 96-well plates. Barcoded adaptors unique to each well were ligated to RNAs after cell lysis, fragmentation, and end repair, all done in plates. Then the barcoded RNAs were pooled and bulk m⁶A-SAC-seq was performed. rRNA-corresponding DNA libraries (rDNA) were cleared in the "Cas9 Ribosomal Depletion" step. **b**, qPCR assessment of rDNA depletion efficiency. Primers corresponding to "18S1-4" can be found in Appendix, Table S2. **c**, Mapping ratio comparison of optimized gRNA sequence versus commercialized (Takara SMARTer pico v2). **d**, Peak distribution of mapped reads on 28S rRNA using optimized gRNA sequence versus commercialized. Upper panel: Takara SMARTer pico v2. Lower panel: optimized.

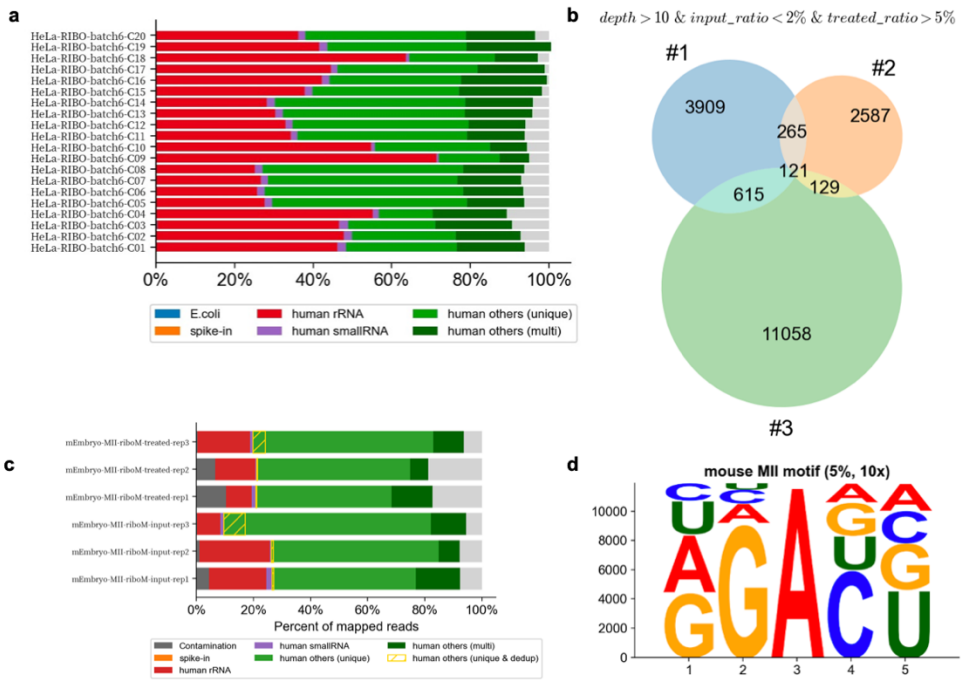


Figure 5.2 Optimization of Cas9-based rDNA depletion

a, Distribution of mapped reads in conditions screening. Details of the screened conditions can be found in Appendix, Table S1. Several conditions showed less than 10% unmapped (grey) and 25% mapped to rRNA (red). **b**, Venn diagram representing overlap of sites between three biological replicates. **c**, Distribution of mapped reads in mouse meiosis II (MII) oocytes. Yellow checkered region represents uniquely mapped deduplicated reads. **d**, Motif profile of the identified sites.

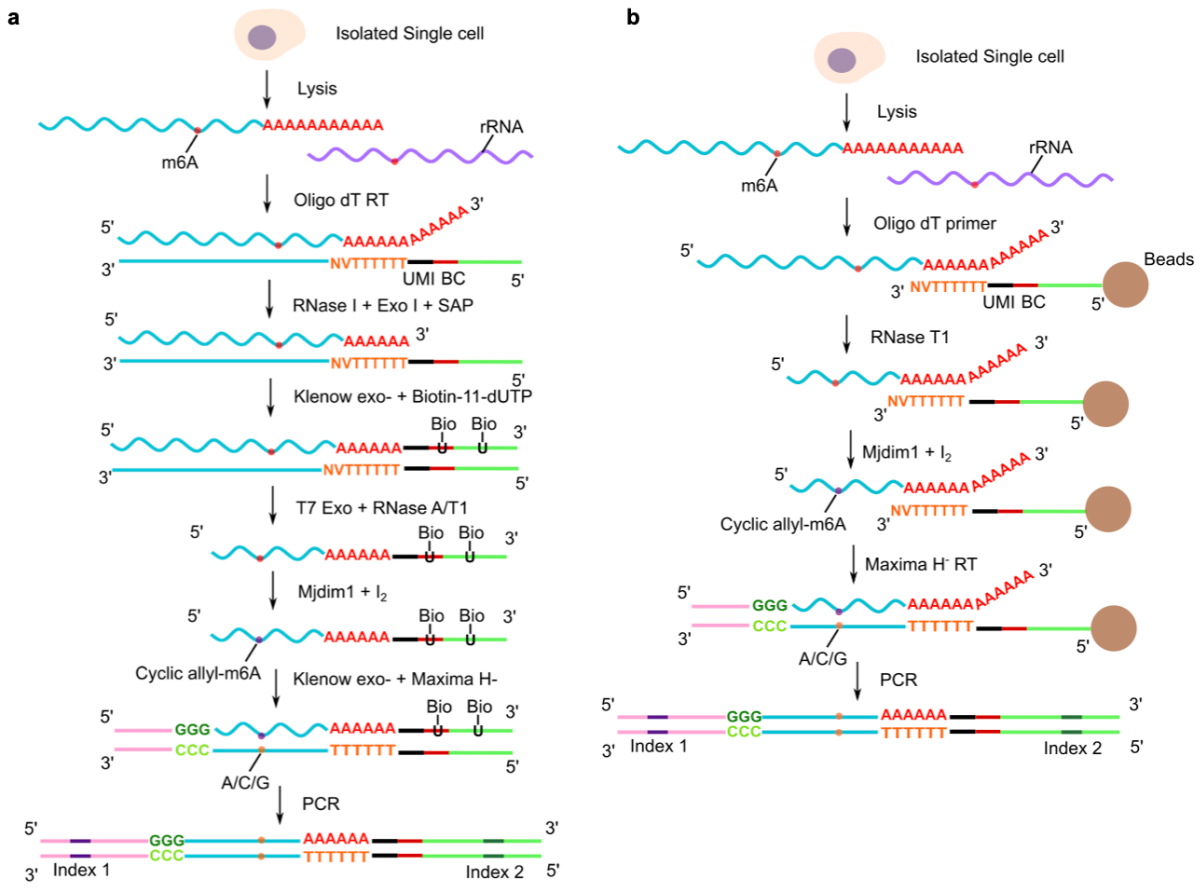


Figure 5.3 Droplet-based m^{6}A-SAC-seq.

a, Schematic representation of "Barcode-transfer" design (see methods). Cells were lysed and reverse transcribed with barcoded oligo-dT primers in droplets. Then the DNA: RNA hybrids were pooled and the 3' overhang of poly A was digested by RNase I. The remaining RNA strand was extended by the sequence on the barcoded primer, which transfers the barcode. Bulk m^{6}A-SAC-seq was performed subsequently with the modification of template switch RT. UMI (black), unique

molecular identifier. BC (crimson), cell barcodes. Bio, biotin. **b**, Schematic representation of "Direct-annealing" design (see methods). Cells were lysed and annealed to barcoded oligo-dT primers. Then the RNA was pooled without denaturation to avoid barcode swapping. m^6A-SAC-seq labeling protocols and template-switch RT was performed on-beads without any denaturation. Beads (brown) represent streptavidin conjugated paramagnetic beads.

5.3 Methods

Preparation of gRNA for Cas9 rDNA depletion: gRNA templates were synthesized by Twist Biosciences. 1 μL of 20 pg/μL template (set #1 or #2) were mixed with 25 μL of 2x KAPA HiFi HotStart ReadyMix (Roche), 1.5 μL of gRNA Primer F (10 μM), 1.5 μL of gRNA Primer R (10 μM), and 21 μL of RNase-free H_2O. The template was PCR amplified at 95 ºC for 3 min, 25 cycles of [98 ºC 20 s, 65 ºC 15 s, 72 ºC 15 s], then incubated at 72 ºC 5 min. The dsDNA was purified with DNA Clean & Concentrator 5 (Zymo). Measure the concentration of the purified dsDNA. The T7 in vitro transcription (IVT) reaction was assembled as follows: 500 ng of dsDNA template, set #1 or #2, dissolved in 10 μL of nuclease-free water; 1.5 μL of 10x T7 Reaction Buffer (NEB), 1.5 μL of 100 mM ATP, 1.5 μL of 100 mM UTP, 1.5 μL of 100 mM ATP, 1.5 μL of 100 mM CTP, 1.5 μL of 100 mM GTP, 1.5 μL of 100 mM ATP, T7 RNA Polymerase Mix (NEB, E2040S), 1 μL of SUPERase•In RNase Inhibitor (20 U/μL, Invitrogen). The reaction was carried out overnight at 37 ºC. Then the reaction was mixed with 5 μL of TURBO DNase (2 U/μL,

Ambion), 5 µL of 10× TURBO DNase Reaction Buffer, 2 µL of SUPERase•In RNase Inhibitor (20 U/µL), and 18 µL of RNase-free H_2O. The reaction was incubated at 37 ºC for 30 min. Then the reaction was diluted with 150 µL of RNase-free water and extracted with 200 µL of acid phenol-chloroform (Ambion). The aqueous phase was ethanol precipitated overnight. gRNA pellets were collected by centrifugation at 14,000 xg and washed with 70% ethanol, then the concentration was measured with Qubit RNA High Sensitivity kit (Invitrogen). gRNA from set #1 and set #2 were mixed in a molar ratio of 1:2. The final concentration was adjusted to 640 ng/µL.

Cas9 rDNA depletion: 250 µg of spCas9 (Sigma) was reconstituted with 390 µL of reconstitution buffer to a final concentration of 4 µM. The dsDNA library to be digested should be dissolved in 10 µL of nuclease-free water at a concentration of 1-3 nM (~5 ng, for 300 bp libraries). 5 µL of gRNA (640 ng/µL, or 20 µM) were mixed with 12.5 µL of spCas9 (4 µM), 5 µL of 10x Cas9 Buffer (500 mM Tris HCl, pH 8.0, 1 M NaCl, 100 mM $MgCl_2$, 10 mM DTT), 7.5 µL of nuclease-free water and incubated at 37 ºC for 5 min. Then the pre-incubated Cas9 RNP was mixed with 10 µL of library and incubated at 37 ºC for 1 h. The digested library was purified with DNA Clean & Concentrator 5 (Zymo) and eluted in 10 µL of nuclease-free water. Then the digestion was performed again with equal amount of gRNA and Cas9, following exactly the same protocol. After the reaction was completed, 1 µL of proteinase K (20 mg/mL, Ambion) was added and the reaction was incubated at 50 ºC for 15 min. The rDNA depleted library was purified with 0.8x (v/v) AMPure XP beads (Beckman) according to the manufacturer's instructions and eluted in 15 µL of nuclease-free water. Mix the library with 25 µL of 2x KAPA HiFi HotStart ReadyMix (Roche), 10 µL of P5/P7 Primer Mix (10 µM each, Roche) and PCR amplified at 98 ºC for 45 s, then 12 cycles of [98 ºC 15 s, 60 ºC 30 s, 72 ºC 30 s] with a final incubation at 72 ºC 1 min. The amplified

library was purified with 0.8x (v/v) AMPure XP beads (Beckman) according to the manufacturer's instructions and sequenced on Illumina Novaseq 6000 platforms.

Plate-based SAC-seq: Cells were sorted into 96 well-plates following standard protocols using FACS Aria Fusion (BD). Prior to sorting, 3 μl of lysis buffer (0.13% Triton-X-100, 2 U/μL SUPERase-In RNase inhibitor, Invitrogen) was aliquoted to each well. Plates with sorted cells can be stored in −80 °C freezers before use. After thawing on ice, 1 μl of 5.8S rRNA masking oligo (ATCGGCAAGCGACGCTCAGACAGGCGTAGCCCCGGGAGGAACCCGGGGCCGCAAG TGCGTTCGAAGTGTCGATGAT/3SpC3/) and 0.5 μl of 10× zinc fragmentation buffer (see Chapter 3, same applies for other buffers) were added and the whole plate was incubated at 72 °C for 20 min. Then 5.5 μl of end repair mix was added consisting of 0.6 μl of 100 mM EDTA, 1 μl of T4 PNK buffer (NEB), 1 μl of T4 PNK (NEB), 0.4 μl of SUPERase-In RNase inhibitor (Invitrogen), 2.5 μl of RNase-free water. The reaction was incubated at 37 ºC for 1 h. Ligation was performed according to instructions in Chapter 3, with the modification of 96 barcoded adapters and spike-in mix both diluted by 100 folds, and the volume of T4 RNA Ligase 2, truncated KQ (200 U/μL, NEB) was halved (to 1 μl). After overnight ligation, all reactions (96 × 25 μL) were combined and transferred to a single 15 mL centrifuge tube. An equal volume of RNase-free water was used to transfer remaining liquids in wells. The combination of liquids (96 × 50 μL, approximately 5 mL) was then incubated with pre-washed Dynabeads C1 (Invitrogen), with the remaining steps of m^{6}A-SAC-seq performed according to the instructions in Chapter 3.

Barcode-transfer SAC-seq: <50 ng of HeLa total RNA was resuspended in 12 uL of RNase-free H_2O. Then, 1 μL of 100 μM 28 s5-dT30VN primer (CTACACGACGCTCTTCCGATCTAAACCCAAGAAACACTNNNNNNNNNNN

TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN) and 1 μL of 10 mM dNTP mix (NEB) were added. The mixture was denatured at 65 ºC for 5 min. Next, 5 μL of Maxima H- RT buffer (Invitrogen), 1 uL of Maxima H- reverse transcriptase (200 U/μL, Invitrogen), and 1 uL of RNaseOUT (Invitrogen) were added. The mixture was incubated at 50 ºC for 30 min. After the reaction was complete, 20 μL of 0.2% SDS was added and the mixture was incubated at room temperature for 5 min. The reaction was purified using 0.8x AmpureXP beads, and eluted in 20 μL of RNase-free H_2O. Then, 9 uL of ExoSAP-IT (Applied Biosystems) and 1 uL of RNase I (NEB) were added. The mixture was incubated at 37 ºC for 30 min, followed by denaturation at 80 ºC for 15 min. The biotin extension mixture was added, consisting of 4 μL of RNase-free H_2O, 3 μL of NEBuffer2, 2 uL of 0.67 mM dNTP* (with 50% of dTTP substituted by biotin-11-dUTP, Thermo), and 1 μL of Klenow exo- (NEB). The mixture was incubated at 37 ºC for 30 min. The cDNA digestion mixture was added, which included 5 μL of RNase-free H_2O, 2 μL of NEBuffer 4, 2 μL of SUPERase-In RNase inhibitor (Invitrogen), and 1 μL of T7 Exonuclease (NEB). The mixture was incubated at 25 ºC for 30 min. Finally, the sample was purified using Dynabeads C1 (Invitrogen) with the remaining steps of m^6A-SAC-seq performed according to the instructions in Chapter 3.

Direct-annealing SAC-seq: 20 μL of Dynabeads C1 (Invitrogen) were aliquoted and washed twice in 1x B/W buffer (see Chapter 3, same applies to other buffers). The beads were resuspended in 20 μL of 1x B/W buffer and mixed with 1 uL of 5Bio-28 s5-dT30VN primer (100 uM, /5Biosg/CTACACGACGCTCTTCCGATCTAAACCCAAGAAACACT NNNNNNNNNNNNTTTTTTTTTTTTTTTTTTTTTTTTTTVN). The mixture was incubated at room temperature for 10 minutes with end to end rotation. After a wash in 1x B/W buffer, the beads were resuspended in 20 μL of 1x B/W buffer and mixed with <50 ng of HeLa

total RNA. The mixture was incubated at room temperature for 10 minutes with end to end rotation. The beads were washed once in 1x B/W buffer and then washed twice in 10 mM Tris-HCl (pH 7.5). Next, the beads were resuspended in 50 μL of 50 mM Tris-HCl (pH 7.5), 2 mM EDTA. 1 μL of 1 U/uL RNase T1 (Ambion) was added, and the mixture was incubated at room temperature for 15 minutes. After a wash in 0.1% PBST, followed by a wash in 1x B/W buffer, and two washes in 10 mM Tris-HCl (pH 7.5), the beads were resuspended in SAC-seq reaction mix: 12 μL of RNase-free H_2O, 12 μL of Allyl-SAM (9.68 mM), 4 μL of 10x Mjdim1 buffer, 4 μL of SUPERase-In RNase inhibitor (Invitrogen), and 8 μL of Mjdim1 (1 mM). The mixture was incubated at 50 ºC for 1 hour. The above step was repeated by decanting the previous reaction mixture on magnetic racks and adding fresh mix. Following a wash in 0.1% PBST, a wash in 1x B/W buffer, and two washes in 10 mM Tris-HCl (pH 7.5), the beads were resuspended in 25 μL of RNase-free H2O. One microliter of 125 mM I_2 (200 mM KI solution) was added, and the mixture was incubated at room temperature for 1 hour in the dark. After a wash in 1x B/W buffer and two washes in 10 mM Tris-HCl (pH 7.5), the beads were resuspended in 8 μL of RNase-free H_2O. 2 μL of 50% PEG8000 and 1 μL of 10 mM dNTP were added. The mixture was incubated at 65 ºC for 5 minutes. The following RT reaction mix was added: 4 μL of 5x Maxima H- RT buffer (Thermo), 2 μL of 10 mM GTP, 1 μL of 40 uM 5Spacer-s7-TSO (/5SpC3/CAGACGTGTGCTCTTCCGATCTNNNNNNrGrGrG), 1 μL of RNaseOUT (Invitrogen), and 1 μL of Maxima H- reverse transcriptase (Invitrogen). The mixture was incubated using the following program: 37 ºC for 30 minutes, 42 ºC for 90 minutes, and 10 cycles of [50 ºC for 2 minutes, 42 ºC for 2 minutes]. After mixing with 20 μL of 0.2% SDS, the mixture was incubated at 95 ºC for 10 minutes. The supernatant was purified and concentrated with Dynabeads

Silane (Invitrogen). Finally, PCR amplification, Ampure XP purification, and agarose gel purification were performed according to the instructions in Chapter 3.

Chapter 6 RNA accessibility assay

6.1 Introduction

RNAs are macromolecules that can fold into complex secondary and tertiary structures. Highly structured regions are typically inaccessible to RBPs and enzymes. There has been a myriad of publications probing RNA higher order structures in vitro[200,201], in vivo[202–204], and in silico[205–207], even with machine learning[208]. In a broader sense, RNA accessibility can also be affected by RBP binding, which could be studied by CLIP-seq/PAR-CLIP (see Chapter 1 Section 1.2.3); or by RNA-RNA/RNA-DNA interactions, which has also been probed respectively[209,210]. Under stress conditions, translation-stalled mRNAs together with translation pre-initiation complex and many other RBPs can be translocated into phase-separated membraneless organelles called stress granules (SGs). The formation of stress granules is thought to be a combined effect of RNA secondary structures, RNA-protein interactions, and RNA-RNA interactions[211]. Although the RNA components of SGs can be profiled through core protein IP[212], ultracentrifugation[213], or proximity labeling/editing[214,215], there lacks an all-encompassing method that profiles RNA accessibility change caused by all kinds of interactions.

RNA accessibility can be functionally relevant. The main topic of this dissertation, m^6A, was recently implied to be deposited in an accessibility-dependent manner[19,115,136]. As discussed in Chapters 1-4, not all DRACH motifs are methylated. The methylation on mature mRNA is initially described to be concentrated in long internal exons and 3' UTR near stop codons[55,126], but later assays revealed that it should be rather described as in long internal exons and near the intron-exon junction (start) of the terminal exon[216]. This has been mysterious because 1. METTL3/14 is nuclear localized 2. Several studies found that introns are relatively m^6A-depleted[92,217,218], despite the

opposite was also reported[219]. Recent studies[19,115,136] revealed that instead of being selectively deposited in long internal and terminal exons, m^{6}A deposition is rather "suppressed" at anywhere in proximity to exon-junctions in an exon-junction complex (EJC) core component eIF4A3 dependent manner. Because EJC bind 20-24 nt upstream of exon junctions, long internal exons and terminal exons will be free of EJC. This is suggested to be achieved via physical exclusion of RNA accessibility by EJC from METTL3/14, given that the methylation complex is also large (~ 1 MDa).

6.2 Results and discussion

We reasoned that TadA8.20 used in eTAM-seq was a good candidate for RNA accessibility profiling, because 1. It converted A to I in vitro at high efficiency; 2. The conversion was apparently hindered by secondary structures under certain reaction conditions (see Chapter 4). 3. The resulting conversion, inosine, could be readout in RT without further enzymatic/chemical treatment, which simplified the procedure. The remaining questions would be: 1. Could TadA8.20 maintain high conversion rate in vivo? Given that the amount of RNA substrate was several magnitudes higher (50 ng of mRNA was used in Chapter 4, whereas in 1 million cell there was ~20 μg of total RNA), the turnover rate of the enzyme could be challenged. 2. Could TadA8.20 permeabilize the plasma membrane/nuclear membrane while keeping the transcriptome relatively intact? An outflux of cytosolic mRNA through the permeabilized cell membrane would significantly limit the transcriptome coverage. 3. Would the deamination be blocked by RBP coating and stress granule/P-body phase-separation?

Inspired by the low-input epigenetic profiling method, CUT&RUN[220]/CUT&TAG[221,222], we employed Concanavalin A (ConA) conjugated paramagnetic beads for cell/nuclei immobilization.

ConA is a carbohydrate-binding protein (also known as lectin) which binds to the carbohydrate chains (also known as glycans) on the plasma membrane/nuclear membrane-embedded glycoproteins at high affinity. Using ConA, buffer exchange could be achieved by magnetic rack separation instead of centrifugation, which significantly helped with the maintenance of cell structure and therefore the relative completeness of the cytosolic RNA contents, after the cell was permeabilized. The permeabilization was implemented with digitonin, which allowed the passage of antibody-like fusion proteins in CUT&TAG. Because antibody was much larger in size compared to TadA8.20, we reasoned that this would equally allow for the passage of the latter. After treatment, cells were lysed with Triton X-100, a less stringent detergent which maintains the nuclear membrane. The released nuclei were simultaneously captured by ConA, allowing for facile compartmentalization. The RNA was then extracted with Trizol and treated with subsequent clearance of DNA and rRNA (**Figure 6.1a**).

We digested the extracted RNA with nuclease P1 and alkaline phosphatase and quantitated the resulting nucleotides with HPLC-QQQ-MS/MS (see Chapters 1-4 for detailed description and protocols). Results showed that the nuclear RNA could be accessed by TadA8.20 while the whole cell was immobilized on ConA, achieving as high as 25% global conversion (**Figure 6.1b**). The background inosine signal was less than 5%. Then the same RNA was reverse transcribed with a mixture of random primers and oligo-dT primers and PCR amplified for NGS sequencing. Specifically, random V_6 (V = A/C/G) hexamers were used alongside with the conventional N_6 hexamers for better annealing with our inosine-rich transcriptome. The secondary structure probed on rRNAs and tRNAs using this method well corresponded with the previous reports.

Then we employed this method on stress-induced versus control human HeLa and HepG2 cell lines. The heat shock stress and oxidative stress were implemented according to previous publications[223–225], which both were reported to induce stress granules formation. We observed a reduction of accessibility after oxidative stress (**Figure 6.1c)**, whose causal relationship to the stress granule formation required further investigation. The heat shock stress, on the other hand, resulted in much less RNA which did not generate high quality data for statistically analysis. More replicates would be done in the future. We also applied the method on eIF4A3 KD HeLa cells versus control, which identified accessibility reduction 20 bp around the exon junctions only in nucleus. The causal relationship of this effect with EJC binding still required further investigation.

In summary, these preliminary findings established a protocol using TadA8.20 for RNA accessibility profiling in vivo, which could potentially probe dynamic changes in RNA accessibility regardless of the cause being conformational change, RNA-protein interaction, RNA-RNA interaction, or phase-separation. This has the potential to be a useful tool in the study of RNA behavior in response to stress conditions.

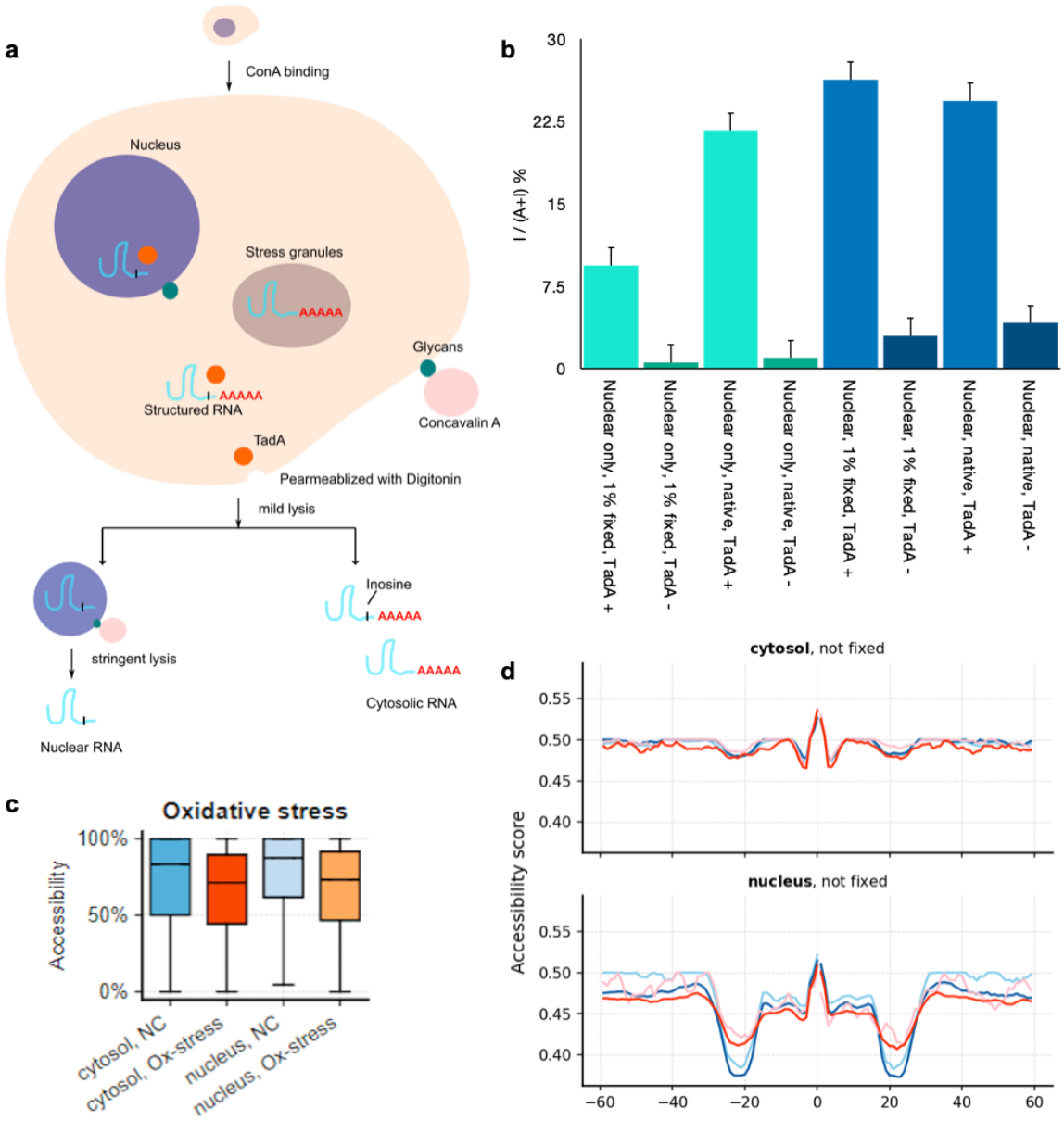


Figure 6.1 TadA RNA accessibility assay

a, Schematic representation of the assay. Cells were immobilized via the interaction between Concanavalin A (ConA, pink) and surface glycans (dark green). TadA8.20 (orange) permeabilized through the cell membrane with the assist from digitonin, and converted A at RNA accessible regions to I. The cells were then lysed and compartmentalized for separate RNA extraction in

nuclei and cytosol. **b**, Global A-to-I conversion measured by LC-MS. In "nuclear only" (cyan), nuclei were isolated prior to TadA8.20 treatment. In "nuclear" (blue), nuclei were isolated after TadA8.20 treatment. $TadA^+$, treated. $TadA^-$, negative control. **c**, Global RNA accessibility in stress-induced cells. NC, negative control. Ox-stress, oxidative stress induced by 1 h treatment of $NaAsO_2$. In box plots, lower and upper hinges represent first and third quartiles, the center line (**Figure 6.1, continued**) represents the median, and whiskers represent ±1.5× the interquartile range. **d**, accessibility reduction near exon junctions, possibly induced by EJC complex.

6.3 Methods

Cell culture and induction of stress conditions: HeLa and HepG2 cell lines were grown in Dulbecco's modified Eagle's medium (DMEM; Gibco, catalog no. 11965092) supplemented with 10% fetal bovine serum (FBS; Gibco) and 1% 100× penicillin–streptomycin (Gibco). Heat shock stress was induced by replacing the cell culture media with fresh media and incubate the cells at 42 ºC for the intended time (1 h or 2 h) before harvesting. Oxidative stress was induced by replacing the cell culture media with fresh media supplemented with fresh media supplemented with 0.5 mM sodium arsenite and incubate the cells at 37 ºC for 1 h before harvesting. Recovery was performed by replacing the cell culture media with fresh media and incubate the stress-induced cells at 37 ºC for 1h before harvesting.

Fixation and cell harvesting: 10 million Hela or HepG2 cells were grown and stress induced as indicated. Before harvesting, cells were washed once with 1× PBS (Gibco). Cells were then harvested and counted. If native cells were intended, 1 million cells were aliquoted per sample

into a 1.5 mL centrifuge tube. If fixed cells were intended, 10 million cells were resuspended in 15 mL of 1× PBS. 1 mL of 16% formaldehyde (Sigma) was mixed with the PBS for a final concentration of 1%. The cells were then incubated at room temperature for 10 mins with end to end rotation. After the incubation was complete, mix the reaction with 3 mL of 2.5 M glycine (aq., Sigma). The cells were then incubated at room temperature for 10 min with end to end rotation. All cells were washed in 1× PBS and spined down at 800 xg centrifugation. Supernatants were aspired and cell pellets were fresh-frozen for future use.

TadA RNA accessibility assay: 20 μL of Concanavalin A (ConA) paramagnetic beads (EpiCypher) were aliquoted and prepared by washing the beads three times, each time in 60 μL of ConA binding buffer (1× PBS, 1 mM $CaCl_2$, 1 mM $MgCl_2$, 1 mM $MnCl_2$). Supernatant was removed by separating the beads on magnetic racks. Then the beads were resuspended in 20 μL of ConA binding buffer. Frozen cell pellets were allowed to defrost on ice. Cell pellets were resuspended in 300 μL of ice-cold 1× PBS supplemented with 1% (v/v) protease inhibitor cocktail (Roche) and 0.2% (v/v) SUPERase In RNase Inhibitor (Invitrogen). The suspension of cells was then mixed with the prepared ConA beads and incubated at room temperature for 10 min with end to end rotation. The beads were then washed with 300 μL of ice-cold wash buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM spermidine, 0.1% BSA) supplemented with 1% (v/v) protease inhibitor cocktail (Roche) and 0.2% (v/v) SUPERase In RNase Inhibitor (Invitrogen). Supernatant was removed by separating the beads on magnetic racks. Then resuspend the cells in 50 μL of TadA labeling mix: 5 μL of TadA8.20 (100 μM), 10 μL of 5× Deamination buffer (250 mM Tris HCl pH 7.5, 125 mM KCl, 12.5 mM $MgCl_2$, 50% Glycerol), 5 μL of SUPERase In RNase Inhibitor, 1 μL of protease inhibitor cocktail (Roche), 1 μL of 2.5% (w/w) digitonin (Sigma) and 28 μL of RNase-free water. The reaction was incubated at 37 ºC for 3 h with end to end rotation or shaking

at 1200 rpm. After the incubation was complete, lyse the cell by directly mixing the reaction with 300 µL of ice-cold NE1 buffer (20 mM HEPES pH 8, 10 mM KCl, 1 mM $MgCl_2$, 0.1% Triton X-100, 20% Glycerol) supplemented with 1% (v/v) protease inhibitor cocktail (Roche) and 1% (v/v) SUPERase In RNase Inhibitor (Invitrogen). Mix the suspension with 20 µL of ConA binding buffer. The suspension was incubated at room temperature with end to end rotation for 10 min. Separate the beads and the supernatant on a magnetic rack. The beads were collected as the nuclear fraction while the supernatant was collected as the cytosolic fraction. The beads were washed with 300 µL of ice-cold wash buffer supplemented with 1% (v/v) protease inhibitor cocktail (Roche) and 0.2% (v/v) SUPERase In RNase Inhibitor (Invitrogen). Resuspend the beads in 120 µL of H_2O, 150 µL of 2× Clear Sample Buffer (200 mM Tris HCl pH 7, 8% SDS, 20 mM EDTA), 30 µL of 1 M DTT. Mix the cytosolic fraction with 375 µL of 2× Clear Sample Buffer, 75 µL of 1 M DTT. If fixed cells were used, incubate both fractions at 75 ºC for 40 min with shaking at 1200 rpm. Otherwise incubate at room temperature for 10 min with end to end rotation. Both fractions were cleared by mixing the nuclear fraction with 1 mL of TRIzol (Invitrogen) and cytosolic fraction with 2 mL of TRIzol. The RNA was extracted according to the manufacturer's instructions (isopropanol precipitation). The nuclear-extracted RNA was treated with 5 µL of 2U/µL TURBO Dnase (Ambion) in 50 µL of 1× TURBO Dnase buffer. Then the reaction was purified with acid-phenol chloroform (Ambion) extraction and ethanol precipitation. Both fractions of RNA were rRNA-depleted with RiboMinu Eukaryote System v2 (Invitrogen) according to manufacturer's instructions. The rRNA-depleted RNA was purified and concentrated with RNA Clean & Concentrator (Zymo) according to manufacturer's instructions. Measure the concentration of RNA. The RT reaction were assembled as follows: 100 ng of RNA dissolved in 7 µL of RNase-free water, 1 µL of 100 µM s5 primer mix (33 µM s5-dT30VN:

ACACTCTTTCCCTACACGACGCTCTTCCGATCTAGACTCNNNNNNTTTTTTTTTTTTTT TTTTTTTTTTTTTTTTVN, 33 μM s5-V6: ACACGACGCTCTTCCGATCTVVVVVV, 33 μM s5-N6: ACACGACGCTCTTCCGATCTNNNNNN), 2 μL of 50% PEG 8000, 1 μL of 10 mM dNTP mix (NEB). Incubate this mixture at 65 ºC for 5 min, then continue to add: 1 μL of 40 μM 5Spacer-s7-TSO (/5SpC3/CAGACGTGTGCTCTTCCGATCT NNNNNN rGrGrG), 1 μL of 5× RT buffer (Thermo), 2 μL of 10 mM GTP, 1 μL of RNaseOUT (Invitrogen), 1 μL of 200 U/μL Maxima H^- reverse transcriptase (Thermo). The reaction was incubated at 42 ºC for 90 min, then 10 cycles of [50 ºC 2 min, 42 ºC 2 min], then at 85 ºC for 5 min. Purify the reaction with 16 μL of AMPure XP beads (Beckman) according to manufacturer's instructions. Amplify the cDNA with NEBNext Ultra II Q5 Master Mix (NEB, catalog no. M0544X) and NEBNext Unique Dual Index Primer for Illumina (NEB, catalog no. E6440S) following the manufacturer's directions with 12 cycles of PCR. The PCR product was further purified by gel excision of fragments >300 bp using 3% low-melting agarose (Fisher, BP165-25) gel. The purified library was sequenced with Illumina Novaseq 6000 platform.

Chapter 7 Summary and perspectives

In summary, we developed two m^{6}A site specific and quantitative profiling methods, m^{6}A-SAC-seq and eTAM-seq. The protocols were optimized for better adaptation of TruSeq for the application of modification profiling, which has the potential to be extended to modifications other than m^{6}A. Preliminary studies have been conducted in low-input and single-cell application of m^{6}A-SAC-seq and eTAM-seq, which still awaits further optimization. Lastly eTAM-seq was repurposed for probing of RNA accessibility and achieved promising preliminary results.

One possible future direction for single cell m^{6}A sequencing is the application of combinatorial indexing techniques developed in recent years. By randomly attaching several rounds of barcodes to evenly distributed cells, if the total number of combinations greatly exceeds the cell number, each cell will almost certainly receive a unique barcode. This idea was first implemented by **single-cell combinatorial indexing RNA sequencing (sci-RNA-seq)**[226], where RT primer barcodes were combined with PCR primer barcodes, yielding 96 × 96 = 9216 combinations, already greatly exceeding the capacity of plate-based scRNA-seq assays. The strength of this design was fully demonstrated in **split-pool ligation-based transcriptome sequencing (SPLiT-seq)**[227], where RT primers were extended via splint ligation (a high efficiency ligation equivalent to nick sealing) for 3 rounds, then combined with PCR primer barcodes. This would yield 96 × 96× 96 × 24 = 21,233,664 combinations, which could unique-label 1 million cells with high confidence, surpassing the upper limit of droplet-based scRNA-seq. For modification sequencing, this means in-tube labeling at the very beginning of the assay without the need to either accommodate the reaction for microfluidic devices, or to achieve barcoding prior to labeling. That said, full-length RNA present in fixed cells might be a challenging scenario for

enzymatic labeling, which requires much optimization or innovations in labeling chemistry, for example mild chemical labeling with minimal sample loss.

Although this dissertation is exclusively focused on NGS sequencing, third generation sequencing techniques are taking on momentum as we write. Oxford Nanopore Technologies (ONT) recently enabled direct RNA sequencing (DRS)[228] which features full-length RNA sequencing without RT. Nanopore sequencing data inherently contains modification information, because the chemical differences of the modified bases will elicit shift in current signal and in the time the nucleic acid sequence resides inside the pore (also known as dwell time)[229]. This is preferred over the current NGS based methods because 1. It envisions a possibility of profiling all kinds of modifications together, in one single run with one sequencing technique; 2. It provides in cis regulation information of modifications present within one single transcript molecule. Although currently still hindered by relatively high error rate compared to NGS, computational methods[230,231] are being invented to mitigate the issue. There is no doubt that more advances in this field will soon emerge.

LIST OF REFERENCES

1. Hotchkiss, R. D. THE QUANTITATIVE SEPARATION OF PURINES, PYRIMIDINES, AND NUCLEOSIDES BY PAPER CHROMATOGRAPHY. Journal of Biological Chemistry **175**, 315–332 (1948).

2. COHN, W. E. & VOLKIN, E. Nucleoside-5′-Phosphates from Ribonucleic Acid. Nature **167**, 483–484 (1951).

3. Zhang, Y., Lu, L. & Li, X. Detection technologies for RNA modifications. Experimental and Molecular Medicine vol. 54 1601–1616 Preprint at https://doi.org/10.1038/s12276-022-00821-0 (2022).

4. McCown, P. J. et al. Naturally occurring modified ribonucleosides. Wiley Interdisciplinary Reviews: RNA vol. 11 Preprint at https://doi.org/10.1002/wrna.1595 (2020).

5. Boccaletto, P. et al. MODOMICS: A database of RNA modification pathways. 2021 update. Nucleic Acids Res **50**, D231–D235 (2022).

6. Arzumanian, V. A., Dolgalev, G. V., Kurbatov, I. Y., Kiseleva, O. I. & Poverennaya, E. V. Epitranscriptome: Review of Top 25 Most-Studied RNA Modifications. International Journal of Molecular Sciences vol. 23 Preprint at https://doi.org/10.3390/ijms232213851 (2022).

7. Nachtergaele, S. & He, C. Chemical Modifications in the Life of an mRNA Transcript. (2018) doi:10.1146/annurev-genet-120417.

8. Helm, M. & Motorin, Y. Detecting RNA modifications in the epitranscriptome: Predict and validate. Nature Reviews Genetics vol. 18 275–291 Preprint at https://doi.org/10.1038/nrg.2016.169 (2017).

9. Roundtree, I. A., Evans, M. E., Pan, T. & He, C. Dynamic RNA Modifications in Gene Expression Regulation. Cell vol. 169 1187–1200 Preprint at https://doi.org/10.1016/j.cell.2017.05.045 (2017).

10. Taoka, M. et al. Landscape of the complete RNA chemical modifications in the human 80S ribosome. Nucleic Acids Res **46**, 9289–9298 (2018).

11. Ramanathan, A., Robb, G. B. & Chan, S. H. mRNA capping: Biological functions and applications. Nucleic Acids Research vol. 44 7511–7526 Preprint at https://doi.org/10.1093/nar/gkw551 (2016).

12. Merrick, W. C. Cap-dependent and cap-independent translation in eukaryotic systems. Gene vol. 332 1–11 Preprint at https://doi.org/10.1016/j.gene.2004.02.051 (2004).

13. Zhang, X., Cozen, A. E., Liu, Y., Chen, Q. & Lowe, T. M. Small RNA Modifications: Integral to Function and Disease. Trends in Molecular Medicine vol. 22 1025–1034 Preprint at https://doi.org/10.1016/j.molmed.2016.10.009 (2016).

14. Perry, R. P. & Kelley, D. E. Existence of methylated messenger RNA in mouse L cells. Cell **1**, 37–42 (1974).

15. O'Brown, Z. K. & Greer, E. L. N6-methyladenine: A conserved and dynamic DNA mark. Adv Exp Med Biol **945**, 213–246 (2016).

16. Chen, S. et al. N6-methyladenosine modification of HIV-1 RNA suppresses type-I interferon induction in differentiated monocytic cells and primary macrophages. PLoS Pathog **17**, (2021).

17. Fu, Y. & He, C. Nucleic acid modifications with epigenetic significance. Current Opinion in Chemical Biology vol. 16 516–524 Preprint at https://doi.org/10.1016/j.cbpa.2012.10.002 (2012).

18. Roundtree, I. A., Evans, M. E., Pan, T. & He, C. Dynamic RNA Modifications in Gene Expression Regulation. Cell vol. 169 1187–1200 Preprint at https://doi.org/10.1016/j.cell.2017.05.045 (2017).

19. Uzonyi, A. et al. Exclusion of m6A from splice-site proximal regions by the exon junction complex dictates m6A topologies and mRNA stability. Mol Cell **83**, 237-251.e7 (2023).

20. Liu, C. et al. Absolute quantification of single-base m6A methylation in the mammalian transcriptome using GLORI. Nat Biotechnol (2022) doi:10.1038/s41587-022-01487-9.

21. He, P. C. & He, C. m 6 A RNA methylation: from mechanisms to therapeutic potential . EMBO J **40**, (2021).

22. Patil, D. P. et al. M6 A RNA methylation promotes XIST-mediated transcriptional repression. Nature **537**, 369–373 (2016).

23. Wang, X. et al. N6-methyladenosine modification of MALAT1 promotes metastasis via reshaping nuclear speckles. Dev Cell **56**, 702-715.e8 (2021).

24. Alarcón, C. R., Lee, H., Goodarzi, H., Halberg, N. & Tavazoie, S. F. N6-methyladenosine marks primary microRNAs for processing. Nature **519**, 482–485 (2015).

25. Pendleton, K. E. et al. The U6 snRNA m6A Methyltransferase METTL16 Regulates SAM Synthetase Intron Retention. Cell **169**, 824-835.e14 (2017).

26. Huang, W. et al. N6-methyladenosine methyltransferases: functions, regulation, and clinical potential. Journal of Hematology and Oncology vol. 14 Preprint at https://doi.org/10.1186/s13045-021-01129-8 (2021).

27. Maden, B. E. H. Identification of the locations of the methyl groups in 18 S ribosomal RNA from Xenopus laevis and man. J Mol Biol **189**, 681–699 (1986).

28. Liu, J. et al. The RNA m6A reader YTHDC1 silences retrotransposons and guards ES cell identity. Nature **591**, 322–326 (2021).

29. Chelmicki, T. et al. m6A RNA methylation regulates the fate of endogenous retroviruses. Nature **591**, 312–316 (2021).

30. Hwang, S.-Y. et al. L1 retrotransposons exploit RNA m6A modification as an evolutionary driving force. Nat Commun **12**, 880 (2021).

31. Narayan, P. & Rottman, F. M. An in Vitro System for Accurate Methylation of Internal Adenosine Residues in Messenger RNA. Science (1979) **242**, 1159–1162 (1988).

32. Bokar, J. A., Rath-Shambaugh, M. E., Ludwiczak, R., Narayan, P. & Rottman, F. Characterization and partial purification of mRNA N6-adenosine methyltransferase from HeLa cell nuclei. Internal mRNA methylation requires a multisubunit complex. Journal of Biological Chemistry **269**, 17697–17704 (1994).

33. Bokar, J. A., Shambaugh, M. E., Polayes, D., Matera, A. G. & Rottman, F. M. Purification and cDNA cloning of the AdoMet-binding subunit of the human mRNA (N6-adenosine)-methyltransferase. RNA **3**, 1233–47 (1997).

34. Agarwala, S. D., Blitzblau, H. G., Hochwagen, A. & Fink, G. R. RNA methylation by the MIS complex regulates a cell fate decision in yeast. PLoS Genet **8**, (2012).

35. Zhong, S. et al. MTA is an Arabidopsis messenger RNA adenosine methylase and interacts with a homolog of a sex-specific splicing factor. Plant Cell **20**, 1278–1288 (2008).

36. Wang, X. et al. Structural basis of N6-adenosine methylation by the METTL3-METTL14 complex. Nature **534**, 575–578 (2016).

37. Liu, J. et al. A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. Nat Chem Biol **10**, 93–95 (2014).

38. Wang, P., Doxtader, K. A. & Nam, Y. Structural Basis for Cooperative Function of Mettl3 and Mettl14 Methyltransferases. Mol Cell **63**, 306–317 (2016).

39. Schwartz, S. et al. Perturbation of m6A writers reveals two distinct classes of mRNA methylation at internal and 5' sites. Cell Rep **8**, 284–296 (2014).

40. Wen, J. et al. Zc3h13 Regulates Nuclear RNA m6A Methylation and Mouse Embryonic Stem Cell Self-Renewal. Mol Cell **69**, 1028-1038.e6 (2018).

41. Wan, C. et al. Panorama of ancient metazoan macromolecular complexes. Nature **525**, 339–344 (2015).

42. Liu, J. et al. VIRMA mediates preferential m6A mRNA methylation in 3′UTR and near stop codon and associates with alternative polyadenylation. Cell Discov **4**, (2018).

43. Doxtader, K. A. et al. Structural Basis for Regulation of METTL16, an S-Adenosylmethionine Homeostasis Factor. Mol Cell **71**, 1001-1011.e4 (2018).

44. Pendleton, K. E. et al. The U6 snRNA m 6 A Methyltransferase METTL16 Regulates SAM Synthetase Intron Retention. Cell **169**, 824-835.e14 (2017).

45. van Tran, N. et al. The human 18S rRNA m6A methyltransferase METTL5 is stabilized by TRMT112. Nucleic Acids Res **47**, 7719–7733 (2019).

46. Ma, H. et al. N 6-Methyladenosine methyltransferase ZCCHC4 mediates ribosomal RNA methylation. Nat Chem Biol **15**, 88–94 (2019).

47. Jia, G. et al. N6-Methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. Nat Chem Biol **7**, 885–887 (2011).

48. Jia, G. et al. Oxidative demethylation of 3-methylthymine and 3-methyluracil in single-stranded DNA and RNA by mouse and human FTO. FEBS Lett **582**, 3313–3319 (2008).

49. He, C. Grand Challenge Commentary: RNA epigenetics? Nat Chem Biol **6**, 863–865 (2010).

50. Saletore, Y. et al. The birth of the Epitranscriptome: deciphering the function of RNA modifications. http://genomebiology.com/2012/13/10/175 (2012).

51. Wiener, D. & Schwartz, S. The epitranscriptome beyond m6A. Nature Reviews Genetics vol. 22 119–131 Preprint at https://doi.org/10.1038/s41576-020-00295-8 (2021).

52. Mauer, J. et al. Reversible methylation of m6 Am in the 5′ cap controls mRNA stability. Nature **541**, 371–375 (2017).

53. Meyer, K. D. & Jaffrey, S. R. Annual Review of Cell and Developmental Biology Rethinking m 6 A Readers, Writers, and Erasers. (2017) doi:10.1146/annurev-cellbio-100616.

54. Zheng, G. et al. ALKBH5 Is a Mammalian RNA Demethylase that Impacts RNA Metabolism and Mouse Fertility. Mol Cell **49**, 18–29 (2013).

55. Dominissini, D. et al. Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. Nature **485**, 201–206 (2012).

56. Zhen, D. et al. m6A Reader: Epitranscriptome Target Prediction and Functional Characterization of N6-Methyladenosine (m6A) Readers. Front Cell Dev Biol **8**, (2020).

57. Zhang, Z. et al. The YTH Domain Is a Novel RNA Binding Domain. Journal of Biological Chemistry **285**, 14701–14710 (2010).

58. Hazra, D., Chapat, C. & Graille, M. m6A mRNA Destiny: Chained to the rhYTHm by the YTH-Containing Proteins. Genes (Basel) **10**, 49 (2019).

59. Fu, Y. & Zhuang, X. m6A-binding YTHDF proteins promote stress granule formation. Nat Chem Biol **16**, 955–963 (2020).

60. Fang, Q. et al. YTHDF1 phase separation triggers the fate transition of spermatogonial stem cells by activating the IκB-NF-κB-CCND1 axis. Cell Rep **42**, (2023).

61. Gao, Y. et al. Multivalent m6A motifs promote phase separation of YTHDF proteins. Cell Research vol. 29 767–769 Preprint at https://doi.org/10.1038/s41422-019-0210-3 (2019).

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

63. Wang, X. et al. N6-methyladenosine modulates messenger RNA translation efficiency. Cell **161**, 1388–1399 (2015).

64. Patil, D. P., Pickering, B. F. & Jaffrey, S. R. Reading m6A in the Transcriptome: m6A-Binding Proteins. Trends in Cell Biology vol. 28 113–127 Preprint at https://doi.org/10.1016/j.tcb.2017.10.001 (2018).

65. Huang, H. et al. Recognition of RNA N6-methyladenosine by IGF2BP proteins enhances mRNA stability and translation. Nat Cell Biol **20**, 285–295 (2018).

66. Zhou, K. I. et al. Regulation of Co-transcriptional Pre-mRNA Splicing by m6A through the Low-Complexity Protein hnRNPG. Mol Cell **76**, 70-81.e9 (2019).

67. Wu, B. et al. Molecular basis for the specific and multivariant recognitions of RNA substrates by human hnRNP A2/B1. Nat Commun **9**, 420 (2018).

68. Alarcón, C. R. et al. HNRNPA2B1 Is a Mediator of m6A-Dependent Nuclear RNA Processing Events. Cell **162**, 1299–1308 (2015).

69. Du, H. et al. YTHDF2 destabilizes m6A-containing RNA through direct recruitment of the CCR4–NOT deadenylase complex. Nat Commun **7**, 12626 (2016).

70. Zhang, F. et al. Fragile X mental retardation protein modulates the stability of its m6A-marked messenger RNA targets. Hum Mol Genet (2018) doi:10.1093/hmg/ddy292.

71. Edens, B. M. et al. FMRP Modulates Neural Differentiation through m6A-Dependent mRNA Nuclear Export. Cell Rep **28**, 845-854.e5 (2019).

72. Wu, R. et al. A novel m6A reader Prrc2a controls oligodendroglial specification and myelination. Cell Res **29**, 23–41 (2019).

73. Wang, X. et al. N 6-methyladenosine-dependent regulation of messenger RNA stability. Nature **505**, 117–120 (2014).

74. Rücklé, C. et al. RNA stability controlled by m6A methylation contributes to X-to-autosome dosage compensation in mammals. Nat Struct Mol Biol (2023) doi:10.1038/s41594-023-00997-7.

75. Zaccara, S. & Jaffrey, S. R. A Unified Model for the Function of YTHDF Proteins in Regulating m6A-Modified mRNA. Cell **181**, 1582-1595.e18 (2020).

76. Li, J. et al. YTHDF1 promotes mRNA degradation via YTHDF1-AGO2 interaction and phase separation. Cell Prolif **55**, (2022).

77. Roundtree, I. A. et al. YTHDC1 mediates nuclear export of N6-methyladenosine methylated mRNAs. Elife **6**, (2017).

78. Meyer, K. D. et al. 5′ UTR m6A Promotes Cap-Independent Translation. Cell **163**, 999–1010 (2015).

79. Choe, J. et al. mRNA circularization by METTL3–eIF3h enhances translation and promotes oncogenesis. Nature **561**, 556–560 (2018).

80. Gong, Y. et al. METTL3-mediated m6A modification promotes processing and maturation of pri-miRNA-19a to facilitate nasopharyngeal carcinoma cell proliferation and invasion. Physiol Genomics **54**, 337–349 (2022).

81. Zhou, K. I. et al. N6-Methyladenosine Modification in a Long Noncoding RNA Hairpin Predisposes Its Conformation to Protein Binding. J Mol Biol **428**, 822–833 (2016).

82. Sommer, S., Lavi, U. & Darnell, J. E. The Absolute Frequency of Labeled IV-6-methyladenosine in HeLa Cell Messenger RNA Decreases with Label Time. J. Mol. Biol vol. 124 (1978).

83. Liu, C. et al. Absolute quantification of single-base m6A methylation in the mammalian transcriptome using GLORI. Nat Biotechnol (2022) doi:10.1038/s41587-022-01487-9.

84. Hu, L. et al. m6A RNA modifications are measured at single-base resolution across the mammalian transcriptome. Nat Biotechnol **40**, 1210–1219 (2022).

85. Xiao, Y. L. et al. Transcriptome-wide profiling and quantification of N 6-methyladenosine by enzyme-assisted adenosine deamination. Nat Biotechnol (2023) doi:10.1038/s41587-022-01587-6.

86. Zhang, N. et al. The m6A reader IGF2BP3 promotes acute myeloid leukemia progression by enhancing RCC2 stability. Exp Mol Med **54**, 194–205 (2022).

87. Schulz, K. N. & Harrison, M. M. Mechanisms regulating zygotic genome activation. Nature Reviews Genetics vol. 20 221–234 Preprint at https://doi.org/10.1038/s41576-018-0087-x (2019).

88. Ivanova, I. et al. The RNA m6A Reader YTHDF2 Is Essential for the Post-transcriptional Regulation of the Maternal Transcriptome and Oocyte Competence. Mol Cell **67**, 1059-1067.e4 (2017).

89. Zhao, B. S. et al. M6 A-dependent maternal mRNA clearance facilitates zebrafish maternal-to-zygotic transition. Nature **542**, 475–478 (2017).

90. Geula, S. et al. m ^{6} A mRNA methylation facilitates resolution of naïve pluripotency toward differentiation. Science (1979) **347**, 1002–1006 (2015).

91. Batista, P. J. et al. M6A RNA modification controls cell fate transition in mammalian embryonic stem cells. Cell Stem Cell **15**, 707–719 (2014).

92. 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. (2017) doi:10.1101/gad.301036.

93. Zhang, C. et al. M6A modulates haematopoietic stem and progenitor cell specification. Nature **549**, 273–276 (2017).

94. Yoon, K.-J. et al. Temporal Control of Mammalian Cortical Neurogenesis by m6A Methylation. Cell **171**, 877-889.e17 (2017).

95. Xia, H. et al. Mettl3 Mutation Disrupts Gamete Maturation and Reduces Fertility in Zebrafish. Genetics **208**, 729–743 (2018).

96. Qi, S.-T. et al. N6-Methyladenosine Sequencing Highlights the Involvement of mRNA Methylation in Oocyte Meiotic Maturation and Embryo Development by Regulating Translation in Xenopus laevis. Journal of Biological Chemistry **291**, 23020–23026 (2016).

97. Grosjean, H., Droogmans, L., Roovers, M. & Keith, G. Detection of Enzymatic Activity of Transfer RNA Modification Enzymes Using Radiolabeled tRNA Substrates. in 55–101 (2007). doi:10.1016/S0076-6879(07)25003-7.

98. Kellner, S., Burhenne, J. & Helm, M. Detection of RNA modifications. RNA Biology vol. 7 237–247 Preprint at https://doi.org/10.4161/rna.7.2.11468 (2010).

99. Zhang, C. & Jia, G. Reversible RNA Modification N 1 -methyladenosine (m 1 A) in mRNA and tRNA. Genomics Proteomics Bioinformatics **16**, 155–161 (2018).

100. El-Aneed, A., Cohen, A. & Banoub, J. Mass Spectrometry, Review of the Basics: Electrospray, MALDI, and Commonly Used Mass Analyzers. Appl Spectrosc Rev **44**, 210–230 (2009).

101. Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. & Whitehouse, C. M. Electrospray Ionization for Mass Spectrometry of Large Biomolecules. Science (1979) **246**, 64–71 (1989).

102. Thüring, K., Schmid, K., Keller, P. & Helm, M. Analysis of RNA modifications by liquid chromatography–tandem mass spectrometry. Methods **107**, 48–56 (2016).

103. Yost, R. A. & Enke, C. G. Selected ion fragmentation with a tandem quadrupole mass spectrometer. J Am Chem Soc **100**, 2274–2275 (1978).

104. Crain, P. F. [42] Preparation and enzymatic hydrolysis of DNA and RNA for mass spectrometry. in 782–790 (1990). doi:10.1016/0076-6879(90)93450-Y.

105. Shah, S. & Friedman, S. H. An ESI-MS method for characterization of native and modified oligonucleotides used for RNA interference and other biological applications. Nat Protoc **3**, 351–356 (2008).

106. Boschenok, J. & Sheil, M. M. Electrospray tandem mass spectrometry of nucleotides. Rapid Communications in Mass Spectrometry **10**, 144–149 (1996).

107. Wetzel, C. & Limbach, P. A. Mass spectrometry of modified RNAs: Recent developments. Analyst vol. 141 16–23 Preprint at https://doi.org/10.1039/c5an01797a (2016).

108. Mädler, S., Erba, E. B. & Zenobi, R. MALDI-ToF Mass Spectrometry for Studying Noncovalent Complexes of Biomolecules. in 1–36 (2012). doi:10.1007/128_2011_311.

109. Gao, X., Tan, B.-H., Sugrue, R. J. & Tang, K. MALDI Mass Spectrometry for Nucleic Acid Analysis. in (eds. Cai, Z. & Liu, S.) vol. 331 55–77 (Springer Berlin Heidelberg, 2012).

110. Liu, N. et al. Probing N6-methyladenosine RNA modification status at single nucleotide resolution in mRNA and long noncoding RNA. RNA **19**, 1848–1856 (2013).

111. Yu, Y. T., Shu, M. D. & Steitz, J. A. A new method for detecting sites of 2'-O-methylation in RNA molecules. RNA **3**, 324–31 (1997).

112. Zhao, X. & Yu, Y. T. Detection and quantitation of RNA base modifications. RNA **10**, 996–1002 (2004).

113. Xiao, Y. et al. An Elongation- and Ligation-Based qPCR Amplification Method for the Radiolabeling-Free Detection of Locus-Specific N 6 -Methyladenosine Modification . Angewandte Chemie **130**, 16227–16232 (2018).

114. Rigby, P. W. J., Dieckmann, M., Rhodes, C. & Berg, P. Labeling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase I. J Mol Biol **113**, 237–251 (1977).

115. He, P. C. et al. Exon architecture controls mRNA m6A suppression and gene expression. Science (1979) **379**, 677–682 (2023).

116. Slatko, B. E., Gardner, A. F. & Ausubel, F. M. Overview of Next-Generation Sequencing Technologies. Curr Protoc Mol Biol **122**, (2018).

117. Illumina.inc. An introduction to Next-Generation Sequencing Technology. www.illumina.com/technology/next-generation-sequencing.html.

118. Emrich, S. J., Barbazuk, W. B., Li, L. & Schnable, P. S. Gene discovery and annotation using LCM-454 transcriptome sequencing. Genome Res **17**, 69–73 (2007).

119. Lister, R. et al. Highly Integrated Single-Base Resolution Maps of the Epigenome in Arabidopsis. Cell **133**, 523–536 (2008).

120. Stark, R., Grzelak, M. & Hadfield, J. RNA sequencing: the teenage years. Nature Reviews Genetics vol. 20 631–656 Preprint at https://doi.org/10.1038/s41576-019-0150-2 (2019).

121. Illumina.inc. For all you seq... https://www.illumina.com/content/dam/illumina-marketing/documents/applications/ngs-library-prep/ForAllYouSeqMethods.pdf.

122. ThermoFisher Scientific. User Guide: Dynabeads mRNA DIRECT Kit. https://www.thermofisher.com/document-connect/document-connect.html?url=https://assets.thermofisher.com/TFS-Assets%2FLSG%2Fmanuals%2Fdynabeads_mRNA_direct_man.pdf.

123. Ramsköld, D. et al. Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. Nat Biotechnol **30**, 777–782 (2012).

124. Syed, F. Application of Nextera™ technology to RNA-seq library preparation. Nat Methods **7**, 2–3 (2010).

125. Picelli, S. et al. Full-length RNA-seq from single cells using Smart-seq2. Nat Protoc **9**, 171–181 (2014).

126. Meyer, K. D. et al. Comprehensive analysis of mRNA methylation reveals enrichment in 3′ UTRs and near stop codons. Cell **149**, 1635–1646 (2012).

127. Zhao, J. et al. Genome-wide Identification of Polycomb-Associated RNAs by RIP-seq. Mol Cell **40**, 939–953 (2010).

128. Singh, G., Ricci, E. P. & Moore, M. J. RIPiT-Seq: A high-throughput approach for footprinting RNA:protein complexes. Methods **65**, 320–332 (2014).

129. Livak, K. J. & Schmittgen, T. D. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCT Method. Methods **25**, 402–408 (2001).

130. Wang, K., Peng, J. & Yi, C. The m6A Consensus Motif Provides a Paradigm of Epitranscriptomic Studies. Biochemistry vol. 60 3410–3412 Preprint at https://doi.org/10.1021/acs.biochem.1c00254 (2021).

131. Wiener, D. & Schwartz, S. The epitranscriptome beyond m6A. Nature Reviews Genetics vol. 22 119–131 Preprint at https://doi.org/10.1038/s41576-020-00295-8 (2021).

132. Jiang, X. et al. The role of m6A modification in the biological functions and diseases. Signal Transduction and Targeted Therapy vol. 6 Preprint at https://doi.org/10.1038/s41392-020-00450-x (2021).

133. Nachtergaele, S. & He, C. Chemical Modifications in the Life of an mRNA Transcript. (2018) doi:10.1146/annurev-genet-120417.

134. Wei, C.-M. & Moss, B. Nucleotide sequences at the N ^{6} -methyladenosine sites of HeLa cell messenger ribonucleic acid. Biochemistry **16**, 1672–1676 (1977).

135. Ge, R. et al. m6A-SAC-seq for quantitative whole transcriptome m6A profiling. Nat Protoc (2022) doi:10.1038/s41596-022-00765-9.

136. Yang, X., Triboulet, R., Liu, Q., Sendinc, E. & Gregory, R. I. Exon junction complex shapes the m6A epitranscriptome. Nat Commun **13**, (2022).

137. Grozhik, A. V., Linder, B., Olarerin-George, A. O. & Jaffrey, S. R. Mapping m6A at Individual-Nucleotide Resolution Using Crosslinking and Immunoprecipitation (miCLIP). in 55–78 (2017). doi:10.1007/978-1-4939-6807-7_5.

138. Chen, K. et al. High-Resolution N ^{6} -Methyladenosine (m ^{6} A) Map Using Photo-Crosslinking-Assisted m ^{6} A Sequencing. Angewandte Chemie International Edition **54**, 1587–1590 (2015).

139. Ule, J. et al. CLIP Identifies Nova-Regulated RNA Networks in the Brain. Science (1979) **302**, 1212–1215 (2003).

140. Hafner, M. et al. Transcriptome-wide Identification of RNA-Binding Protein and MicroRNA Target Sites by PAR-CLIP. Cell **141**, 129–141 (2010).

141. McIntyre, A. B. R. et al. Limits in the detection of m6A changes using MeRIP/m6A-seq. Sci Rep **10**, (2020).

142. Lee, H. & Schatz, M. C. Genomic dark matter: the reliability of short read mapping illustrated by the genome mappability score. Bioinformatics **28**, 2097–2105 (2012).

143. Helm, M., Lyko, F. & Motorin, Y. Limited antibody specificity compromises epitranscriptomic analyses. Nature Communications vol. 10 Preprint at https://doi.org/10.1038/s41467-019-13684-3 (2019).

144. Zhang, Z. et al. Systematic calibration of epitranscriptomic maps using a synthetic modification-free RNA library. Nat Methods **18**, 1213–1222 (2021).

145. Schwartz, S. et al. High-resolution mapping reveals a conserved, widespread, dynamic mRNA methylation program in yeast meiosis. Cell **155**, (2013).

146. Liu, S., Zhu, A., He, C. & Chen, M. REPIC: a database for exploring the N6-methyladenosine methylome. Genome Biol **21**, 100 (2020).

147. Liu, J. et al. Landscape and Regulation of m6A and m6Am Methylome across Human and Mouse Tissues. Mol Cell **77**, 426-440.e6 (2020).

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

149. Wang, Y., Xiao, Y., Dong, S., Yu, Q. & Jia, G. Antibody-free enzyme-assisted chemical approach for detection of N6-methyladenosine. Nat Chem Biol **16**, 896–903 (2020).

150. Shu, X. et al. A metabolic labeling method detects m6A transcriptome-wide at single base resolution. Nat Chem Biol **16**, 887–895 (2020).

151. Shu, X. et al. N ^{6} -Allyladenosine: A New Small Molecule for RNA Labeling Identified by Mutation Assay. J Am Chem Soc **139**, 17213–17216 (2017).

152. Zhang, Z. et al. Single-base mapping of m ^{6} A by an antibody-independent method. Sci Adv **5**, (2019).

153. Garcia-Campos, M. A. et al. Deciphering the "m6A Code" via Antibody-Independent Quantitative Profiling. Cell **178**, 731-747.e16 (2019).

154. Imanishi, M., Tsuji, S., Suda, A. & Futaki, S. Detection of N ^{6} -methyladenosine based on the methyl-sensitivity of MazF RNA endonuclease. Chemical Communications **53**, 12930–12933 (2017).

155. Meyer, K. D. DART-seq: an antibody-free method for global m6A detection. Nat Methods **16**, 1275–1280 (2019).

156. Tegowski, M., Flamand, M. N. & Meyer, K. D. scDART-seq reveals distinct m6A signatures and mRNA methylation heterogeneity in single cells. Mol Cell **82**, 868-878.e10 (2022).

157. Zheng, G. X. Y. et al. Massively parallel digital transcriptional profiling of single cells. Nat Commun **8**, 14049 (2017).

158. O'Farrell, H. C., Musayev, F. N., Scarsdale, J. N. & Rife, J. P. Binding of Adenosine-Based Ligands to the MjDim1 rRNA Methyltransferase: Implications for Reaction Mechanism and Drug Design. Biochemistry **49**, 2697–2704 (2010).

159. O'Farrell, H. C., Pulicherla, N., Desai, P. M. & Rife, J. P. Recognition of a complex substrate by the KsgA/Dim1 family of enzymes has been conserved throughout evolution. RNA **12**, 725–733 (2006).

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

161. Garcia-Campos, M. A. et al. Deciphering the "m6A Code" via Antibody-Independent Quantitative Profiling. Cell **178**, 731-747.e16 (2019).

162. Golinelli, M.-P. & Hughes, S. H. Nontemplated Nucleotide Addition by HIV-1 Reverse Transcriptase. Biochemistry **41**, 5894–5906 (2002).

163. Davis, W. R., Gabbara, S., Hupe, D. & Peliska, J. A. Actinomycin D Inhibition of DNA Strand Transfer Reactions Catalyzed by HIV-1 Reverse Transcriptase and Nucleocapsid Protein. Biochemistry **37**, 14213–14221 (1998).

164. Song, Y., Liu, K. J. & Wang, T.-H. Elimination of Ligation Dependent Artifacts in T4 RNA Ligase to Achieve High Efficiency and Low Bias MicroRNA Capture. PLoS One **9**, e94619 (2014).

165. Rusche, J. R. & Howard-Flanders, P. Hexamine cobalt chloride promotes intermolecular ligation of blunt end DNA fragments by T4 DNA ligase. Nucleic Acids Res **13**, 1997–2008 (1985).

166. Zhuang, F., Fuchs, R. T., Sun, Z., Zheng, Y. & Robb, G. B. Structural bias in T4 RNA ligase-mediated 3′-adapter ligation. Nucleic Acids Res **40**, e54–e54 (2012).

167. Zhang, L.-S. et al. Transcriptome-wide Mapping of Internal N7-Methylguanosine Methylome in Mammalian mRNA. Mol Cell **74**, 1304-1316.e8 (2019).

168. Frommer, M. et al. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. Proceedings of the National Academy of Sciences **89**, 1827–1831 (1992).

169. Wright, D. J., Force, C. R. & Znosko, B. M. Stability of RNA duplexes containing inosine·cytosine pairs. Nucleic Acids Res **46**, 12099–12108 (2018).

170. Packer, M. S. & Liu, D. R. Methods for the directed evolution of proteins. Nat Rev Genet **16**, 379–394 (2015).

171. Grünewald, J. et al. Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors. Nature **569**, 433–437 (2019).

172. Gaudelli, N. M. et al. Directed evolution of adenine base editors with increased activity and therapeutic application. Nat Biotechnol **38**, 892–900 (2020).

173. Gaudelli, N. M. et al. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. Nature **551**, 464–471 (2017).

174. Kint, S., De Spiegelaere, W., De Kesel, J., Vandekerckhove, L. & Van Criekinge, W. Evaluation of bisulfite kits for DNA methylation profiling in terms of DNA fragmentation and DNA recovery using digital PCR. PLoS One **13**, e0199091 (2018).

175. Benjamini, Y. & Speed, T. P. Summarizing and correcting the GC content bias in high-throughput sequencing. Nucleic Acids Res **40**, e72–e72 (2012).

176. Liu, J. et al. N ^{6} -methyladenosine of chromosome-associated regulatory RNA regulates chromatin state and transcription. Science (1979) **367**, 580–586 (2020).

177. Tang, F. et al. mRNA-Seq whole-transcriptome analysis of a single cell. Nat Methods **6**, 377–382 (2009).

178. Baslan, T. & Hicks, J. Unravelling biology and shifting paradigms in cancer with single-cell sequencing. Nat Rev Cancer **17**, 557–569 (2017).

179. Usoskin, D. et al. Unbiased classification of sensory neuron types by large-scale single-cell RNA sequencing. Nat Neurosci **18**, 145–153 (2015).

180. Ofengeim, D., Giagtzoglou, N., Huh, D., Zou, C. & Yuan, J. Single-Cell RNA Sequencing: Unraveling the Brain One Cell at a Time. Trends Mol Med **23**, 563–576 (2017).

181. Zeisel, A. et al. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. Science (1979) **347**, 1138–1142 (2015).

182. Villani, A.-C. et al. Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. Science (1979) **356**, (2017).

183. Papalexi, E. & Satija, R. Single-cell RNA sequencing to explore immune cell heterogeneity. Nat Rev Immunol **18**, 35–45 (2018).

184. Zhang, S., Li, X., Lin, J., Lin, Q. & Wong, K.-C. Review of single-cell RNA-seq data clustering for cell-type identification and characterization. RNA **29**, 517–530 (2023).

185. Dagogo-Jack, I. & Shaw, A. T. Tumour heterogeneity and resistance to cancer therapies. Nat Rev Clin Oncol **15**, 81–94 (2018).

186. Yao, H. et al. scm6A-seq reveals single-cell landscapes of the dynamic m6A during oocyte maturation and early embryonic development. Nat Commun **14**, 315 (2023).

187. Kraus, A. J., Brink, B. G. & Siegel, T. N. Efficient and specific oligo-based depletion of rRNA. Sci Rep **9**, 12281 (2019).

188. Chamberlin, M. & Ring, J. Characterization of T7-specific ribonucleic acid polymerase. 1. General properties of the enzymatic reaction and the template specificity of the enzyme. J Biol Chem **248**, 2235–44 (1973).

189. Hagemann-Jensen, M. et al. Single-cell RNA counting at allele and isoform resolution using Smart-seq3. Nat Biotechnol **38**, 708–714 (2020).

190. Gr??nweller, A. & Hartmann, R. K. Locked Nucleic Acid Oligonucleotides. BioDrugs **21**, 235–243 (2007).

191. Tessier, D. C., Brousseau, R. & Vernet, T. Ligation of single-stranded oligodeoxyribonucleotides by T4 RNA ligase. Anal Biochem **158**, 171–178 (1986).

192. Klein, A. M. et al. Droplet Barcoding for Single-Cell Transcriptomics Applied to Embryonic Stem Cells. Cell **161**, 1187–1201 (2015).

193. Macosko, E. Z. et al. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell **161**, 1202–1214 (2015).

194. Zucha, D., Androvic, P., Kubista, M. & Valihrach, L. Performance Comparison of Reverse Transcriptases for Single-Cell Studies. Clin Chem **66**, 217–228 (2020).

195. Faridani, O. R. et al. Single-cell sequencing of the small-RNA transcriptome. Nat Biotechnol **34**, 1264–1266 (2016).

196. Hagemann-Jensen, M., Abdullayev, I., Sandberg, R. & Faridani, O. R. Small-seq for single-cell small-RNA sequencing. Nat Protoc **13**, 2407–2424 (2018).

197. Prezza, G. et al. Improved bacterial RNA-seq by Cas9-based depletion of ribosomal RNA reads. RNA **26**, 1069–1078 (2020).

198. Gu, W. et al. Depletion of Abundant Sequences by Hybridization (DASH): using Cas9 to remove unwanted high-abundance species in sequencing libraries and molecular counting applications. Genome Biol **17**, 41 (2016).

199. Takara bio USA. TECH NOTE: SMARTer Stranded Total RNA-Seq Kit v2 - Pico Input Mammalian: improved ease of use and sequencing performance for whole transcriptome analysis of high-quality or degraded samples. https://www.takarabio.com/learning-centers/next-generation-sequencing/technical-notes/rna-seq/stranded-libraries-from-picogram-input-total-rna-(v2).

200. Kertesz, M. et al. Genome-wide measurement of RNA secondary structure in yeast. Nature **467**, 103–107 (2010).

201. Wan, Y. et al. Landscape and variation of RNA secondary structure across the human transcriptome. Nature **505**, 706–709 (2014).

202. Ding, Y. et al. In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. Nature **505**, 696–700 (2014).

203. Rouskin, S., Zubradt, M., Washietl, S., Kellis, M. & Weissman, J. S. Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. Nature **505**, 701–705 (2014).

204. Spitale, R. C. et al. Structural imprints in vivo decode RNA regulatory mechanisms. Nature **519**, 486–490 (2015).

205. Watkins, A. M., Rangan, R. & Das, R. FARFAR2: Improved De Novo Rosetta Prediction of Complex Global RNA Folds. Structure **28**, 963-976.e6 (2020).

206. Boniecki, M. J. et al. SimRNA: a coarse-grained method for RNA folding simulations and 3D structure prediction. Nucleic Acids Res **44**, e63–e63 (2016).

207. Reuter, J. S. & Mathews, D. H. RNAstructure: software for RNA secondary structure prediction and analysis. BMC Bioinformatics **11**, 129 (2010).

208. Townshend, R. J. L. et al. Geometric deep learning of RNA structure. Science (1979) **373**, 1047–1051 (2021).

209. Cai, Z. et al. RIC-seq for global in situ profiling of RNA–RNA spatial interactions. Nature **582**, 432–437 (2020).

210. Xiao, Q. et al. The landscape of promoter-centred RNA–DNA interactions in rice. Nat Plants **8**, 157–170 (2022).

211. Campos-Melo, D., Hawley, Z. C. E., Droppelmann, C. A. & Strong, M. J. The Integral Role of RNA in Stress Granule Formation and Function. Front Cell Dev Biol **9**, (2021).

212. Khong, A. et al. The Stress Granule Transcriptome Reveals Principles of mRNA Accumulation in Stress Granules. Mol Cell **68**, 808-820.e5 (2017).

213. Namkoong, S., Ho, A., Woo, Y. M., Kwak, H. & Lee, J. H. Systematic Characterization of Stress-Induced RNA Granulation. Mol Cell **70**, 175-187.e8 (2018).

214. Padrón, A., Iwasaki, S. & Ingolia, N. T. Proximity RNA Labeling by APEX-Seq Reveals the Organization of Translation Initiation Complexes and Repressive RNA Granules. Mol Cell **75**, 875-887.e5 (2019).

215. van Leeuwen, W. et al. Identification of the stress granule transcriptome via RNA-editing in single cells and in vivo. Cell Reports Methods **2**, 100235 (2022).

216. Ke, S. et al. A majority of m ⁶ A residues are in the last exons, allowing the potential for 3′ UTR regulation. Genes Dev **29**, 2037–2053 (2015).

217. Chen-Kiang, S., Nevins, J. R. & Darnell, J. E. N-6-methyl-adenosine in adenovirus type 2 nuclear RNA is conserved in the formation of messenger RNA. J Mol Biol **135**, 733–752 (1979).

218. Wei, G. et al. Acute depletion of METTL3 implicates N ⁶ -methyladenosine in alternative intron/exon inclusion in the nascent transcriptome. Genome Res **31**, 1395–1408 (2021).

219. Louloupi, A., Ntini, E., Conrad, T. & Ørom, U. A. V. Transient N-6-Methyladenosine Transcriptome Sequencing Reveals a Regulatory Role of m6A in Splicing Efficiency. Cell Rep **23**, 3429–3437 (2018).

220. Skene, P. J. & Henikoff, S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. Elife **6**, (2017).

221. Kaya-Okur, H. S. et al. CUT&Tag for efficient epigenomic profiling of small samples and single cells. Nat Commun **10**, 1930 (2019).

222. Kaya-Okur, H. S., Janssens, D. H., Henikoff, J. G., Ahmad, K. & Henikoff, S. Efficient low-cost chromatin profiling with CUT&Tag. Nat Protoc **15**, 3264–3283 (2020).

223. Markmiller, S. et al. Context-Dependent and Disease-Specific Diversity in Protein Interactions within Stress Granules. Cell **172**, 590-604.e13 (2018).

224. Wheeler, J. R., Matheny, T., Jain, S., Abrisch, R. & Parker, R. Distinct stages in stress granule assembly and disassembly. Elife **5**, (2016).

225. Maxwell, B. A. et al. Ubiquitination is essential for recovery of cellular activities after heat shock. Science (1979) **372**, (2021).

226. Cao, J. et al. Comprehensive single-cell transcriptional profiling of a multicellular organism. https://www.science.org.

227. Rosenberg, A. B. et al. Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding.

228. Garalde, D. R. et al. Highly parallel direct RNA sequencing on an array of nanopores. Nat Methods **15**, 201–206 (2018).

229. Anreiter, I., Mir, Q., Simpson, J. T., Janga, S. C. & Soller, M. New Twists in Detecting mRNA Modification Dynamics. Trends Biotechnol **39**, 72–89 (2021).

230. Hendra, C. et al. Detection of m6A from direct RNA sequencing using a multiple instance learning framework. Nat Methods **19**, 1590–1598 (2022).

231. Leger, A. et al. RNA modifications detection by comparative Nanopore direct RNA sequencing. Nat Commun **12**, (2021).

Appendix: Optimization Conditions Used in Chapter 4

Optimization conditions used in Chapter 4 (**Tables S1-2**).

Table S1 Cas9-rDNA depletion conditions

#	gRNA ratio	gRNA amount	Purification	Proteinase K
5	1:2	1.28 μg	+	+
6	1:4	1.28 μg	+	+
7	1:2	6.4 μg	+	+
8	1:4	6.4 μg	+	+
9	1:2	0.64 μg	-	+
10	1:4	0.64 μg	-	+
11	1:2	3.2 μg	-	+
12	1:4	3.2 μg	-	+
13	1:2	1.28 μg	+	-
14	1:4	1.28 μg	+	-
15	1:2	6.4 μg	+	-
16	1:4	6.4 μg	+	-
17	1:2	0.64 μg	-	-
18	1:4	0.64 μg	-	-

19	1:2	3.2 µg	-	-
20	1:4	3.2 µg	-	-

Table S2 rRNA targeting primers used for depletion efficiency assessment

1F	AGAAACGGCTACCACATCCA
1R	GTCGGGAGTGGGTAATTTGC
2F	TACCACATCCAAGGAAGGCA
2R	TTTTCGTCACTACCTCCCCG
3F	AAGTCCCTGCCCTTTGTACA
3R	AGGGCCTCACTAAACCATCC
4F	GCGTTGATTAAGTCCCTGCC
4R	GATCCGAGGGCCTCACTAAA