Development of direct, quantitative, and base-resolution sequencing methods for DNA and RNA modifications

Haiqi Xu

Ludwig Institute for Cancer Research

Nuffield Department of Medicine

Magdalen College

University of Oxford

This thesis is submitted for the degree of

Doctor of Philosophy in Clinical Medicine

August 2024

Declaration

This thesis is submitted to the Department of Medical Science, University of Oxford, for the degree of Doctor of Philosophy. This thesis is the result of my own work unless otherwise stated. This thesis has not been submitted for any other degree or qualification.

I declare that part of the figures in Chapter 1 were published in Xu, et al. *Isr. J. Chem.* **2024**, *64*, e202400007; most of the figures in Chapter 2 were published in Xu, et al. *J. Am. Chem. Soc.* **2023**, *145*, 7095–7100; most of the figures in Chapter 3 and Chapter 4 were published in Xu, et al. *Nat. Methods* **2024**.

Word count: 31208

Haiqi Xu

Abstract

DNA and RNA are decorated with various chemical modifications. Sequencing these modifications is crucial for us to understand their biological functions. To meet the growing demand for detecting DNA and RNA modifications, it is highly desirable to develop ideal chemical tools and sequencing methods that enable direct, quantitative, and base-resolution detection. In this thesis, I introduce the development of two such methods: Chemical-Assisted Pyridine borane Sequencing plus (CAPS+) for DNA 5-hydroxyethylcytosine (5hmC) sequencing and 2-Bromoacrylamide Assisted Cyclization Sequencing (BACS) for RNA pseudouridine (Ψ) sequencing. CAPS+ enhances the toolbox for sequencing DNA cytosine modifications, while BACS revolutionizes the Ψ detection. BACS has been applied to a series of studies on Ψ and pseudouridine synthases (PUS), revealing their functional importance in various biological processes and significantly deepening our understanding of this long known but enigmatic RNA modification. Overall, the development of CAPS+ and BACS clearly demonstrates the power of chemical tools in biological research. Further advancements in chemistry will continue to facilitate the study of DNA and RNA modifications.

Dedication

This thesis is dedicated to

My parents

And my grandparents

Acknowledgements

Firstly, I would like to thank my supervisor, Prof. Chunxiao Song, for his guidance and support throughout my PhD study. I would also like to thank all current and former lab members for their contributions to my projects, especially Dr. Yibin Liu for discussions in chemistry and Linzhen Kong for help with sequencing data analysis. Additionally, I want to thank Dr. Jingfei Cheng, Dr. Xiufei Chen, Dr. Jinfeng Chen, Dr. Masato Inoue, and Dr. Felix Jackson for their contributions to the applications of CAPS+.

Secondly, I would like to thank my co-supervisors, Prof. Xin Lu and Prof. Skirmantas Kriaucionis, for their continuous support. I would also like to acknowledge all collaborators: Dr. Khatoun Al Moussawi for providing the EBV-infected cell lines; Thomas McMahon for his assistance with UHPLC-MS/MS experiments; Prof. Jane McKeating, Dr. Peter Wing, Dr. James Harris, Dr. Senko Tsukuda, Prof. Alfredo Castello, Dr. Azman Embarc-Buh, and Dr. Guifeng Wei for providing RNA virus samples; Mengjie Li and Dr. Parinaz Mehdipour for generating the knockout cell lines.

Lastly, I wish to give my special thanks to everyone who has supported me throughout my journey in chemistry. My heartfelt gratitude goes to my family — my parents and grandparents — for raising me up and for their unwavering support. A special acknowledgement goes to my grandfather for teaching me the periodic table when I was five years old, which sparked my interest in chemistry. However, I was more absorbed in other natural sciences for a long time. This situation did not change until I entered high school and met my chemistry teacher Mr. Lifeng Wan. During high school, Lifeng showed me the wonders of chemistry, provided me with many books, and encouraged me to participate in the Chinese Chemistry Olympiad, which finally helped me gain admission

to the College of Chemistry and Molecular Engineering at Peking University — the best place to study chemistry in China. Studying at Peking University was a great experience, where I made friends with many talented peers and learned from dedicated professors, including my undergraduate supervisor Prof. Meiping Zhao. Meiping helped me quickly adapt to the research environment and continued to support me as I started my PhD study remotely during the pandemic. Here, I would also like to extend my thanks to those who helped me during that difficult time, especially Dr. Honggang Nie, Shuangshuang Di, and Dr. Jiang Zhou for their assistance with mass spectrometry. Finally, the pandemic went away, and I joined Prof. Chunxiao Song's lab at University of Oxford, where I was introduced to the fascinating field of DNA and RNA modification sequencing. Chunxiao granted me the freedom to explore this exciting area, and without his guidance, my achievements would not have been possible.

Thank you all for your invaluable support during my PhD study.

Table of Contents

1 Introduction

1.1 DNA and RNA modifications

1.1.1 Overview

Nucleic acids are more than just combinations of four canonical nucleotides (A, C, G, T for DNA and A, C, G, U for RNA). They undergo numerous modifications in all cellular organisms across all three domains of life [1] (Figure 1.1). A wide variety of chemical groups can be attached to both the nucleobase and the backbone of nucleic acids, ranging from simple methyl groups to amino acids, cyclopentane derivatives, polyamines, and saccharides. To date, more than 15 and 150 natural modifications have been identified in DNA and RNA, respectively [2,3].

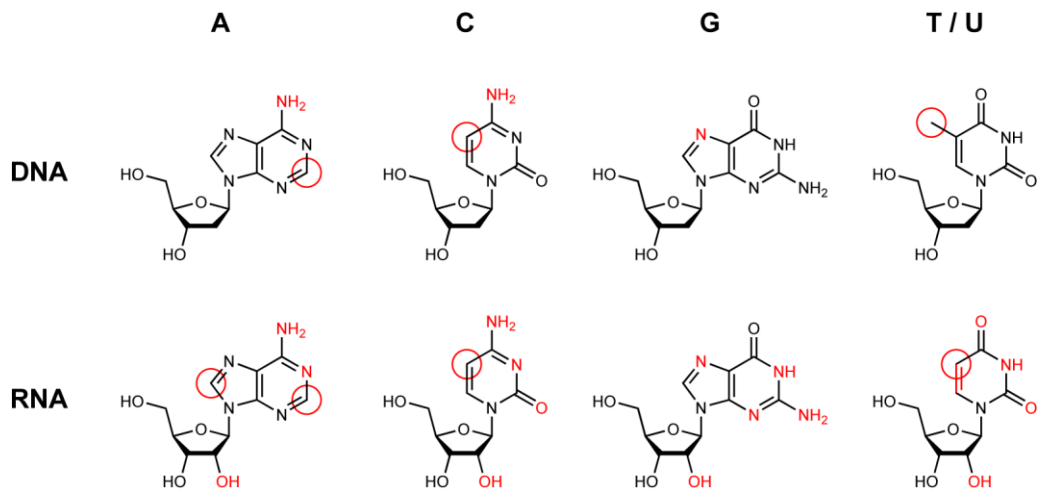


Figure 1.1 | The four canonical DNA and RNA nucleosides. Red color denotes where the modification can occur.

These modifications play crucial roles in many biological processes, including regulation of gene expression, nucleic acids stability, and the distinction between self and non-self. Alterations in DNA and RNA modifications are linked to various human diseases, including inherited genetic disorders and cancer [4-7]. Beyond naturally occurring modifications, artificially synthesized modified nucleobases, nucleosides, or nucleotides (and their analogs) are widely used as anti-tumor and antiviral drugs [8], as well as key components in therapeutic nucleic acids [9]. Notably, the 2023 Nobel Prize in Physiology or Medicine was awarded to Katalin Karikó and Drew Weissman for their contributions to the development of effective messenger RNA (mRNA) vaccines using modified nucleobases. Therefore, understanding the chemical properties and biological functions of DNA and RNA modifications is important not only for studying the mechanisms underlying relevant biological processes, but also for developing novel modified nucleobases, nucleosides, nucleotides, or nucleic acids for diagnostic and therapeutic applications. In this section, I will briefly introduce the history and current state of research on DNA and RNA modifications.

1.1.2 DNA cytosine modifications

The first discovered DNA modification was 5-methylcytosine (5mC) (Figure 1.2). In 1925, Treat B. Johnson and Robert D. Coghill observed 5mC in the DNA of *Mycobacterium tuberculosis* [10]. Although initially disputed by other researchers [11], the existence of 5mC was finally confirmed in calf thymus DNA by Rollin D. Hotchkiss in 1948 [12]. Later, methylation of the fifth carbon of DNA cytosines was found to be a common phenomenon [13,14]. In particular, 5mC was demonstrated to be the most abundant DNA modification in mammalian genomes, constituting 4–5% of all cytosines

[15]. Furthermore, 5mC is predominantly located in the CpG dinucleotides in a symmetrical manner in mammals, with 60–80% of the CpG sites in the human genome methylated to different extents [16].

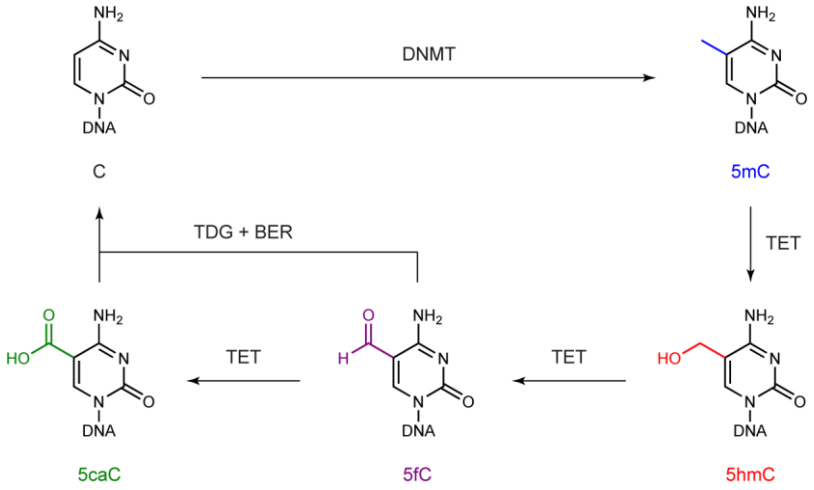


Figure 1.2 | DNA cytosine modifications and the active demethylation pathway.

5mC is generated by DNA methyltransferase (DNMT) enzymes [17] (Figure 1.2). 5mC can undergo demethylation processes; however, it cannot be simply explained by the dilution of 5mC after DNA replication [18,19], which is also known as the passive demethylation pathway [20]. This mechanism was not revealed until 2009–2011. Multiple groundbreaking studies demonstrated that 5mC can be oxidized by ten-eleven translocation (TET) family enzymes to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxycytosine (5caC) sequentially [21-24] (Figure 1.2). Subsequent thymine-DNA glycosylase (TDG) excision of 5fC and 5caC to abasic sites and repair through the base excision repair (BER) pathway complete the demethylation process [24], now known as the active demethylation pathway [25]. Consistent with these findings, 5hmC was found to be the second most abundant DNA modification in

mammalian genomes (0.1–0.7% of all cytosines) [15], while 5fC and 5caC exist at much lower levels (0.0001–0.001% of all cytosines) [23].

1.1.3 DNA cytosine modifications and epigenetics

Epigenetics refers to changes in phenotypes without changes in genotypes [26]. From a modern perspective, epigenetics is linked to a series of underlying mechanisms without altering the genetic information [27,28]. Over the past few decades, DNA cytosine modifications have been recognized as major contributors to epigenetic regulation, alongside histone modifications [29,30]. Indeed, 5mC and 5hmC play important regulatory roles in gene expression [31,32]. Generally, promoter 5mC is considered as a repressive marker of gene expression, while 5hmC in gene bodies is recognized as an active marker. Consistent with their regulatory roles, 5mC and 5hmC are dynamic during cell differentiation and disease development [33,34]. Additionally, aberrant DNA methylation and hydroxymethylation have been observed during tumorigenesis [35-37]. More specifically, global hypomethylation and hypohydroxymethylation but local hypermethylation in promoters are observed in many different cancer types. Hypomethylating agents (HMA), such as 5-azacytidine (azacitidine) and 5-aza-2'-deoxycytidine (decitabine), can inhibit DNMT and have been used as primary treatments for myelodysplastic syndrome (MDS) and acute myeloid leukemia (AML), marking the advent of epigenetic therapies [38-41]. More importantly, 5mC and 5hmC in cell-free DNAs (cfDNAs) are considered as promising biomarkers in liquid biopsies for early cancer diagnostics [42-45]. Therefore, DNA cytosine modifications have garnered extensive attention and have been the primary focus in the study of DNA and RNA modifications for a long time.

1.1.4 RNA modifications and epitranscriptomics

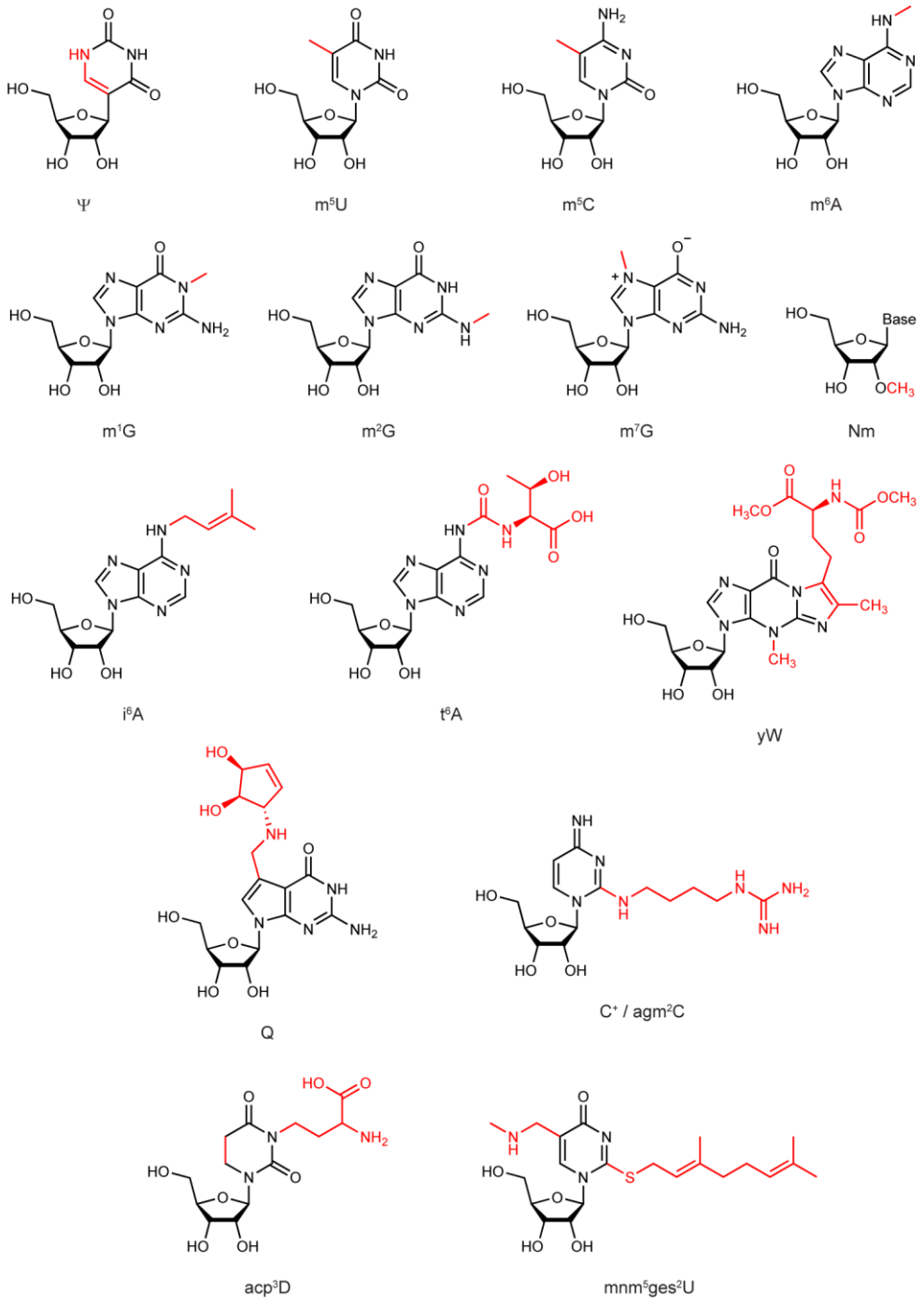


Figure 1.3 | The variety of RNA modifications.

In contrast to the relatively limited repertoire of DNA modifications, RNA modifications are far more diverse (Figure 1.3). The first RNA modification was discovered by Waldo E. Cohn and Elliot Volkin in 1951 [46], which was later confirmed to be pseudouridine (Ψ) [47-49]. Subsequently, several simple types of RNA methylation, including 5-methyluridine (m^5U), 5-methylcytidine (m^5C), N^6-methyladenosine (m^6A), 1-methylguanosine (m^1G), N^2-methylguanosine (m^2G), 7-methylguanosine (m^7G), 2'-*O*-methylation (Nm), and others [50-54], and more structurally complicated modifications, such as N^6-isopentenyladenosine (i^6A) [55], N^6-threonylcarbamoyladenonsine (t^6A) [56], wybutosine (yW) [57], queuosine (Q) [58,59], agmatidine (C^+ or agm^2C) [60,61], 3-(3-amino-3-carboxypropyl)-5,6-dihydrouridine (acp^3D) [62], and 5-methylaminomethyl-2-geranylthiouridine (mnm^5ges^2U) [63,64], were identified in the following decades. The majority of these RNA modifications are found in non-coding RNAs (ncRNAs), such as transfer RNAs (tRNAs) and ribosomal RNAs (rRNAs) [65,66].

Probably due to historical reasons, RNA modifications have conventionally been viewed as stable and playing adaptive rather than regulatory roles in gene expression [67], leading to a long-standing neglect of these modifications. This situation did not change until 2011, with the first identification of an m^6A demethylase called fat mass and obesity-associated protein (FTO) *in vitro* by Chuan He and colleagues [68]. Since m^6A is likely the most prevalent post-transcriptional internal modification in eukaryotic mRNAs, the discovery of its demethylase suggests that m^6A modification could also be reversible and have regulatory roles by affecting the properties of mRNA, similar to cytosine modifications in DNA [69,70]. Despite ongoing controversy over whether FTO is a *bona fide* m^6A demethylase *in vivo* [71,72], this exciting finding opened the field of epitranscriptomics,

which essentially means RNA epigenetics, and has garnered growing interest ever since it was proposed [73-75].

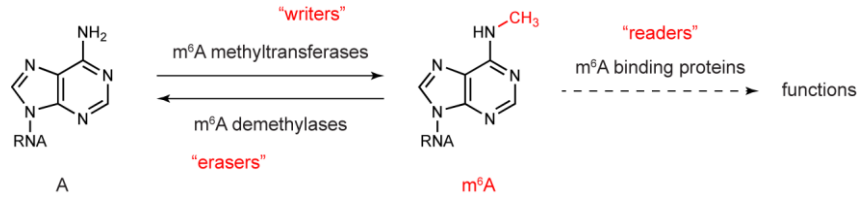


Figure 1.4 | The current view of m^{6}A epitranscriptomics.

In the following years, most efforts have focused on studying m^{6}A. Researchers have characterized the methyltransferases, demethylases, and binding proteins of m^{6}A, also termed writers, erasers, and readers, respectively [76-78]. Based on these discoveries, a comprehensive picture of m^{6}A dynamics has been drawn similarly to DNA cytosine modifications [79,80] (Figure 1.4). Importantly, the first-in-class small molecule inhibitor of m^{6}A methyltransferase METTL3 was developed in 2021 [81], and its Phase I clinical trials for treating AML have begun [82], highlighting the therapeutic potential of inhibiting RNA-modifying enzymes [83,84].

1.1.5 Limitations of the current view of epitranscriptomics

The fast-expanding field of epitranscriptomics has inspired researchers to revisit old modifications in ncRNAs, including Ψ, 1-methyladenosine (m^{1}A), m^{5}C, *N*^{4}-acetylcytidine (ac^{4}C), Nm, m^{7}G, and others [85]. However, only limited progress has been made so far. This failure can largely be attributed to the efforts to replicate the success of m^{6}A research. Except for the writers, researchers have been struggling to define the erasers and readers of these ncRNA modifications [86-88]. Additionally, researchers have

been spending significant effort studying the functions of these modifications in mRNA, despite their lower prevalence [89-91]. These differences between ncRNA modifications and m^6A suggest limitations in applying m^6A models to ncRNA modifications.

Perhaps it is time to delve into the principles of epigenetics and epitranscriptomics. Let's start with well-studied epigenetics. As a chemist, it is always interesting to understand why DNA cytosine modifications and histone modifications were selected as the main regulators of epigenetics through evolution. Although we may never know the answer, we can easily identify one important similarity between them, that is, the reversibility of the modification. Indeed, only reversible modifications can enable both positive and negative regulation, especially for long-lived molecules, while irreversible modifications are more likely to play adaptive roles. Given that it is extremely challenging to monitor any modification site-specifically at the single-molecule level, nearly all the current detection methods detect integrated signals of all molecules, namely the whole ensemble. Therefore, researchers have been using the dynamics of a modification as an important criterion for whether the modification is reversible or further plays regulatory roles. Indeed, this approach has worked well for both DNA methylation and histone modifications. However, the observed dynamics of a modification actually depends on the turnover rate of itself (modification vs demodification) and the turnover rate of the subject biomacromolecule (synthesis vs degradation) (Figure 1.5 and Table 1.1). This theory provides a possible explanation for why the above criterion has been effective for DNA methylation and histone modifications — both DNA and histone proteins are long-lived molecules, which means the dynamics of a modification mainly reflect the turnover of itself. Additionally, histone modifications are generally less stable than DNA

methylation [92], which could also be partially attributed to the relatively shorter lifetime of histone proteins (half-life up to 100–200 d) [93-96] compared to DNA (presumably as long as the cell life), alongside the high turnover rates of histone modifications [97].

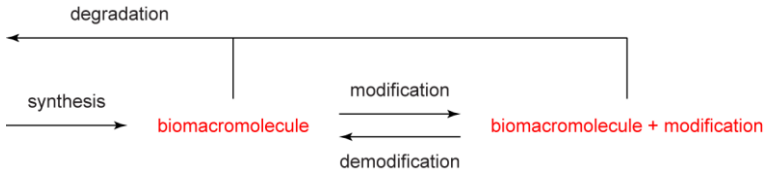


Figure 1.5 | The dynamics of a modification.

Table 1.1 | Estimation of turnover rates of different biomacromolecules and the corresponding modifications.

	Turnover	Half-life		Turnover	Half-life
DNA	Extremely slow	Very long	DNA modification	Fast	~ 1 hour to ?
Histone protein	Slow	~ 100 days	Histone modification	Very fast	~ 1 min to ~ 1 day
RNA	Fast	~ 1 hour to ~ 5 days	RNA modification	? Probably no	?

Now let's return to RNA modifications. Apparently, researchers have been pre-occupied with the similarity between DNA and RNA modifications. Indeed, they share similar or even identical chemical structures of the nucleobases (for example, 5mC vs m^5C), where the majority of modifications occur. However, compared to DNA, RNA has a much shorter lifetime (half-life ranging from only several hours for mRNA to 3–5 d for rRNA) [98-100]. This nature makes it less likely or even not worthwhile for cells to induce reversible modifications on RNA molecules, as they will degrade quickly anyway. Consequently, the dynamics of RNA modifications may stem more from the inherent

dynamics of RNA rather than genuine reversible modifications. Nonetheless, the rapid turnover of RNA molecules allows them to play a broad range of regulatory roles. Indeed, various types of ncRNAs are important regulators of gene expression in eukaryotic cells, including rRNAs and tRNAs in translation, spliceosomal small nuclear RNAs (snRNAs) in mRNA splicing, and microRNAs (miRNAs) in gene silencing. In these cases, whether an RNA modification is reversible or not is less critical — even irreversible RNA modifications can be significant enough to regulate gene expression by fine-tuning the properties of their subject RNA molecules. This irreversible approach might have been selected as an energy-efficient strategy as early as the emergence of cellular organisms [101].

1.1.6 Ψ — a representative of ncRNA modifications

Based on the above discussions, it is necessary to find a representative ncRNA modification as the model to facilitate research. Among all ncRNA modifications, the first discovered one, Ψ, stands out as a prime candidate:

- Ψ is highly abundant in various types of ncRNAs, including rRNAs, tRNAs, and snRNAs [102,103]. Recently, Ψ has also been found in mRNAs, although less prevalent than in ncRNAs [104,105].

- Ψ in present in the RNA of a wide range of cellular organisms across all three domains of life [106]. Given its close association with rRNAs and tRNAs, Ψ modifying enzymes, also known as pseudouridine synthases (PUS), are presumed to be among the most conserved proteins during evolution, similar to ribosomal proteins and aminoacyl-tRNA synthetases [107]. This suggests that Ψ may have been present in

the last universal common ancestor (LUCA), as a vital RNA modification in rRNAs and tRNAs.

- Ψ plays important roles in many biological processes, including translation, mRNA splicing, and innate immune responses [108]. Furthermore, Ψ and PUS enzymes are linked to many diseases including cancer, which is summarized in Table 1.2.

Table 1.2 | PUS enzymes and human diseases.

PUS enzyme	Disease	Possible reasons	Ref.
DKC1 mutation	X-linked dyskeratosis congenita	Telomerase RNA component (TERC) destabilization	[109]
PUS1 mutation	Mitochondrial myopathy and sideroblastic anemia (MLASA)	Mitochondrial tRNA Ψ dysregulation	[110,111]
PUS3 mutation	Intellectual disability	Cytosolic tRNA Ψ dysregulation ?	[112]
PUS7 mutation	Intellectual disability	Cytosolic tRNA Ψ dysregulation ?	[113,114]
PUS7	Acute myeloid leukemia ?	?	[115,116]
PUS7	Glioblastoma ?	?	[117]

Pioneering researchers have made substantial contributions to determining the positions of Ψ, elucidating the catalytic mechanisms of PUS enzymes, and understanding their biological functions using *E. coli* and yeast as model organisms [118]. However, progress in human systems has been limited, possibly due to increased complexity compared to *E. coli* and yeast. Clearly, more work is needed to deepen our understanding of the underlying mechanisms of Ψ modifications and PUS enzymes and provide insights into other ncRNA modifications.

1.2 Chemistry of DNA and RNA modifications

1.2.1 Overview

The discovery of DNA and RNA modifications has expanded the chemical diversity of nucleic acids. The attachment, substitution, or isomerization of functional groups in nucleic acids fundamentally alters their properties. Naturally, chemists have been extensively studying the chemical properties and developing novel chemical reactions for DNA and RNA modifications since their discovery. This variety of chemical tools allows us to understand their biological functions and create new analogs for therapeutic applications. In this section, I will summarize the chemical properties of DNA and RNA modifications and introduce one of the most important chemical reactions in this field — bisulfite-mediated deamination of unmodified cytosine to uracil.

1.2.2 Chemical properties of DNA and RNA modifications

Before delving into the chemical properties of nucleic acid modifications, it is essential to introduce the chemical properties of the canonical nucleotides in DNA and RNA. Compared to the phosphate and sugar backbone, the nucleobases possess most of the reactivities. Since thymine can be considered as 5-methyluracil (5mU), there are four major types of nucleobases in natural nucleic acids: two purine bases (A and G) and two pyrimidine bases (C and U). An overview of the chemical reactivities of these four canonical nucleobases is summarized in Figure 1.6 [119-121]. In general, their reactivities can be divided into nucleophilic and electrophilic reactivities. Due to their intrinsic electron properties, more electron-rich purine nucleobases (A and G) predominantly act

as nucleophiles, while more electron-deficient pyrimidine nucleobases (especially U) mainly serve as electrophiles.

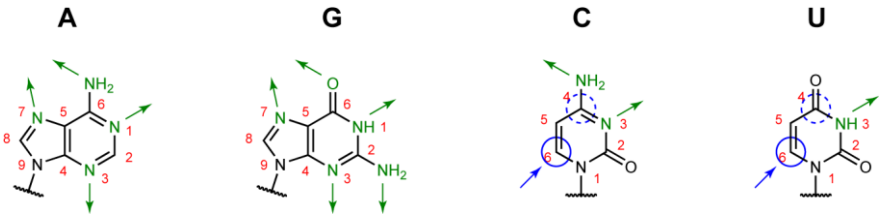
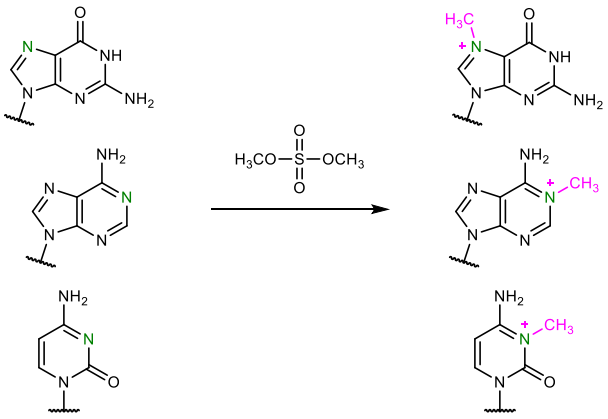
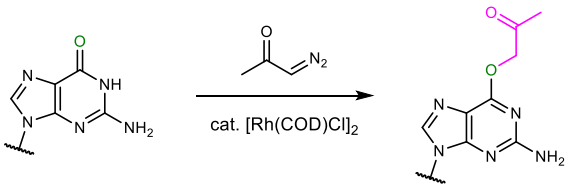
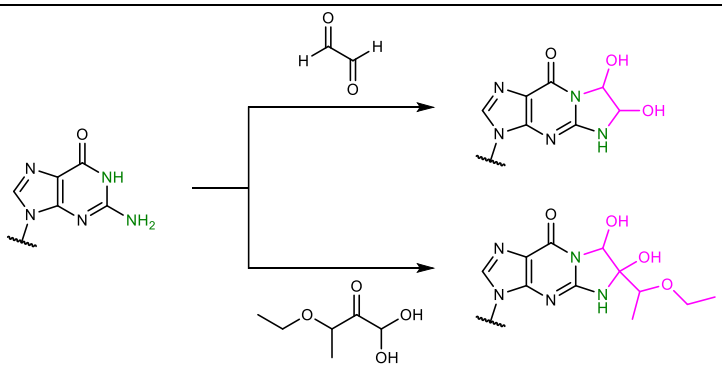


Figure 1.6 | The chemical properties of the four canonical nucleobases. Green and blue color denote the nucleophilic and electrophilic sites, respectively.

More specifically, most of the nitrogen and oxygen atoms in the nucleobases that do not form glycosidic bonds, including N^1, N^3, N^6, N^7 of adenine, N^1, N^2, N^3, O^6, N^7 of guanine, N^3 and N^4 of cytosine, and N^3 of uracil, can behave as nucleophiles, preferring to react with relatively strong electrophiles. These range from alkyl halides, alkylsulfonates, sulfur or nitrogen mustard-type compounds, nitroso compounds, metal carbenes, aldehydes, carbodiimides, to Michael acceptors. The correct positions of two nucleophiles, such as N^1 and N^3 of adenine, N^1 and N^2 of guanine, and N^3 and N^4 of cytosine, enable them to react with bifunctional electrophiles, such as chloroacetaldehyde and glyoxal. In contrast to nucleophilic reactivities, only a few positions in the nucleobases can serve as electrophiles. The main electrophilic centers are at C^6 position of cytosine and uracil, generally acting as Michael acceptors due to the electron-deficient nature and polarized C^5-C^6 double bonds. It is worth mentioning that the C^4 position of cytosine and uracil can also serve as potential electrophiles, particularly after Michael addition of the C^5-C^6 double bond, which removes the limitation of aromaticity. Besides, as heterocycles, nucleobases can undergo various radical-type or photo-induced reactions,

including the formation of 8-oxoguanine (8-oxoG) and thymidine dimers [122-127]. However, these nucleobases are traditionally considered as DNA or RNA damage and therefore are not discussed in detail here. A selected collection of the chemical reactions on the four canonical nucleobases can be found in Table 1.3 and 1.4.

Table 1.3 | Chemical reactions on the four canonical nucleobases.

Chemical reaction	Ref.
	[128]
	[129]
	[130,131]

The modification of nucleobases inevitably changes the chemical properties and reactivities of canonical nucleobases. The exact effect requires a case-specific study. For

instance, the modification at C^5 of cytosine can alter the electron property of the C^5-C^6 double bond. Modifications with electron-withdrawing groups (as in 5fC and 5caC) enhance the electrophilicity, while methyl and hydroxymethyl groups increase the nucleophilicity, thereby enabling 5mC and 5hmC to behave as electron-rich alkenes. The different reactivities between the C^5-C^6 double bonds of thymine and uracil can be explained in the same way. Another common type of modification that can alter electrophilicity is the methylation of saturated nitrogen atoms, finally inducing positively charged nucleobases like m^1A, m^7G, and 3-methylcytidine (m^3C). These modifications generally enhance the electrophilicity of adjacent or conjugated positions and enable a series of unique reactions. Conversely, methylation of amino groups in nucleobases slightly increases their nucleophilicity. However, the methyl group also blocks the dehydration step involved in the formation of diazonium salt. An overview of the above chemical reactions can be found in Table 1.4.

Table 1.4 | Chemical modifications alter the reactivities of nucleobases.

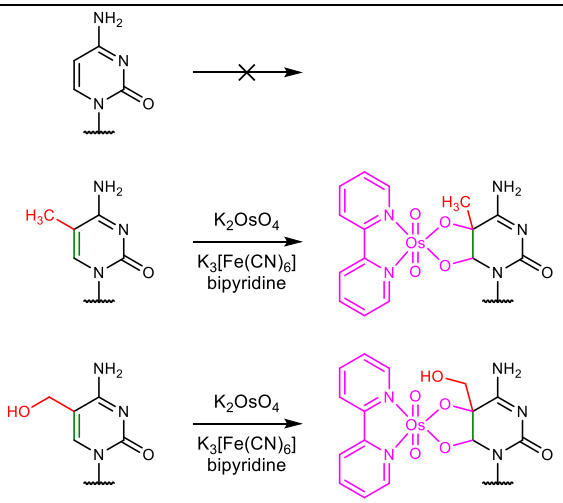
Chemical reaction	Ref.
	[132,133]

Table 1.4 continued | Chemical modifications alter the reactivities of nucleobases.

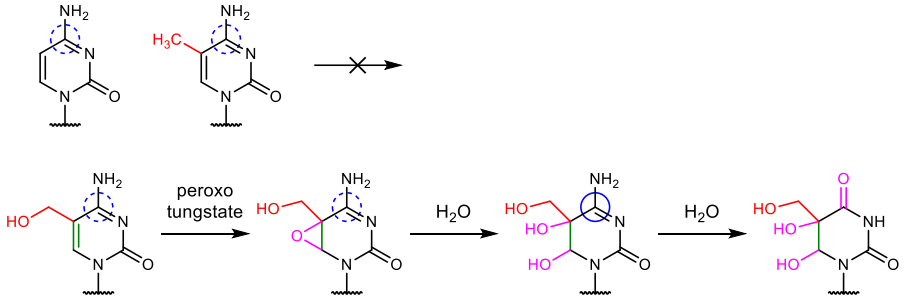
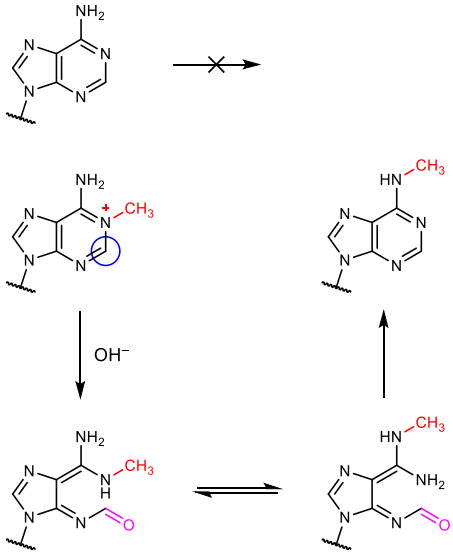
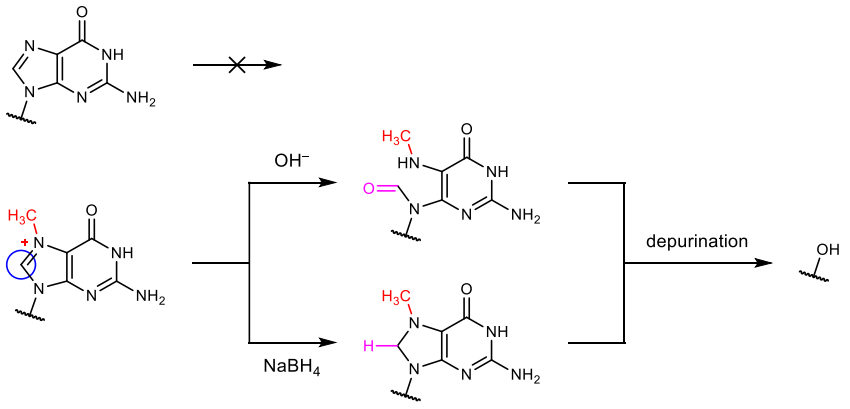
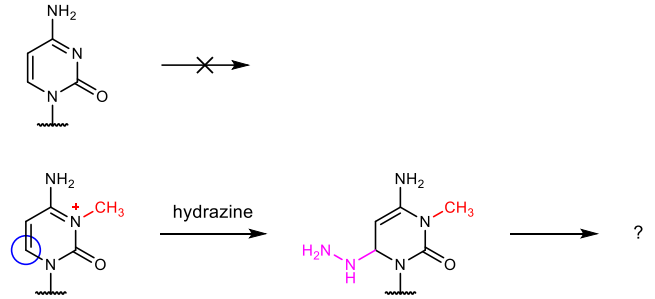
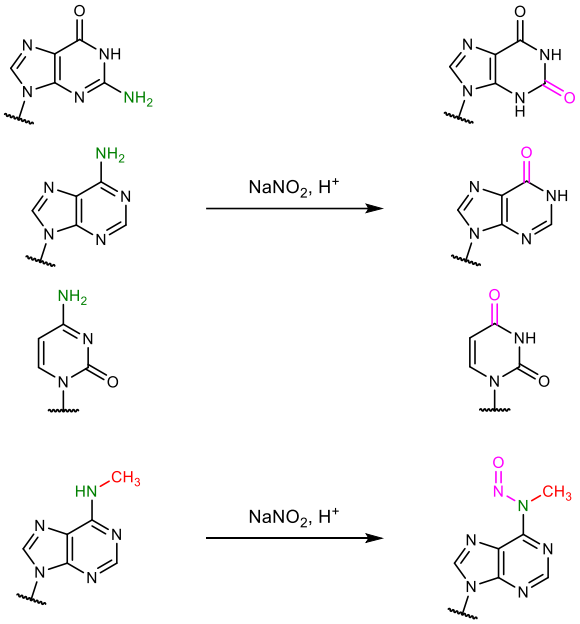
Chemical reaction	Ref.
	[134]
	[135,136]
	[137] [138]

Table 1.4 continued | Chemical modifications alter the reactivities of nucleobases.

Chemical reaction	Ref.
	[139]
	[140,141]

In addition to alteration of the reactivities of canonical nucleobases, DNA and RNA modifications also retain their own reactivity. For example, the hydroxyl group in 5hmC can be readily oxidized, and the formyl group in 5fC or carboxyl group in 5caC can react with hydroxylamines and amines, respectively. These types of reactivities constitute the majority of reactivities of modified nucleobases. Other examples include but are not limited to alkylation of thiolated nucleobases, oxidation of thiolated nucleobases, Michael addition with amide groups in nucleobases, and diazo transfer to aliphatic amines. Also,

correct positioning enables tandem reactions, for example, Friedländer synthesis on 5fC and iodation of alkenes and subsequent intramolecular cyclization on i^6A (similar to iodolactonization). An overview of the above chemical reactions can be found in Table 1.5.

Table 1.5 | The intrinsic reactivities of modified nucleobases.

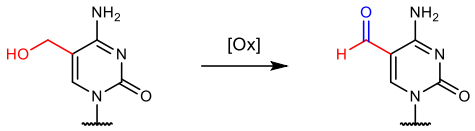
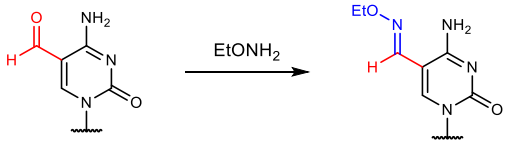
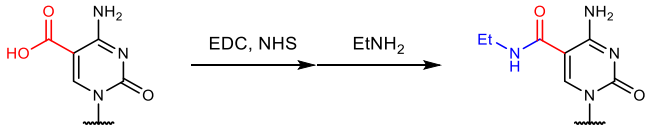
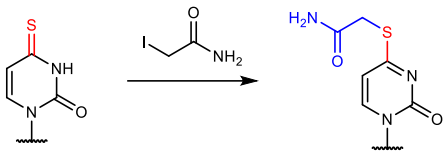
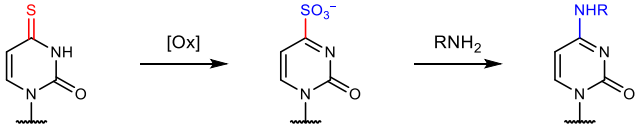
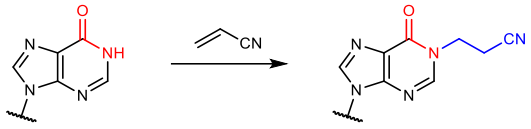
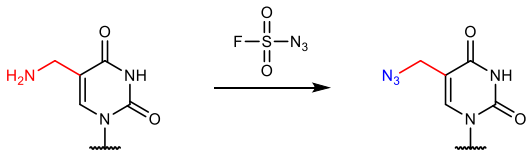
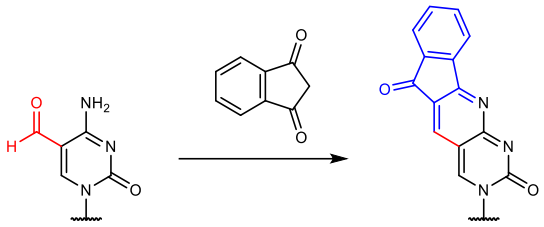
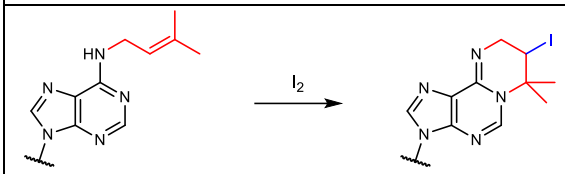
Chemical reaction	Ref.
	[142] [143]
	[144]
	[145]
	[146]
	[147] [148] [149,150]
	[151,152]
	[153]

Table 1.5 continued | The intrinsic reactivities of modified nucleobases.

Chemical reaction	Ref.
	[154]
	[155]

Compared to the variety of chemical reactivities of nucleobases, the chemical reactivities of the phosphate and sugar backbone are relatively simple. Basically, the 2'-OH group in ribose could serve as a nucleophile to react with acylating reagents, or to attack its own 3'-phosphate group. This nucleophilicity is blocked when 2'-OH is methylated. Additionally, the 3'-end of RNA possesses a 1,2-diol structure and can be readily oxidized by periodate, while this reactivity can be blocked by 2'-*O*-methylation. A summary of these chemical reactions can be found in Table 1.6.

Table 1.6 | The reactivities of the phosphate and sugar backbone.

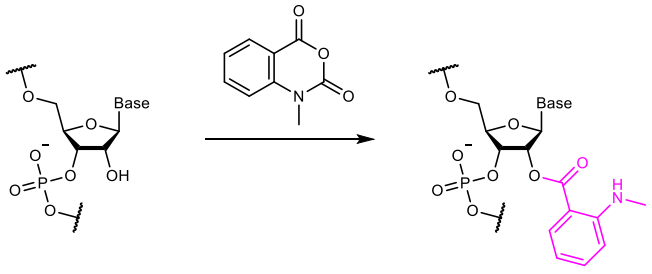
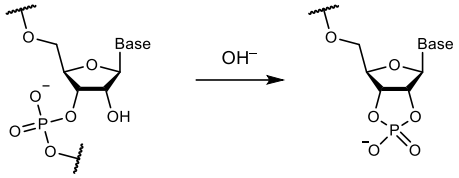
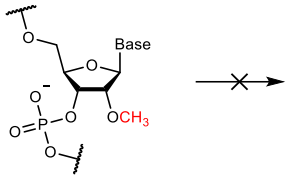
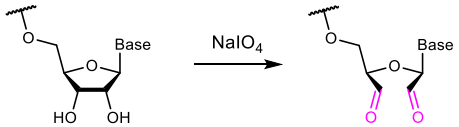
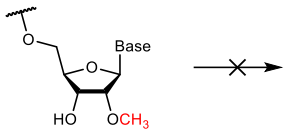
Chemical reaction	Ref.
	[156]

Table 1.6 continued | The reactivities of the phosphate and sugar backbone.

Chemical reaction	Ref.
 	[157]
 	[158]

Beyond these chemical properties, modifications can also alter the base-pairing properties of nucleobases. Generally, whether a modification interrupts the canonical U·A or C·G base pairing or not depends on whether it blocks the hydrogen bond on the Waston-Crick side. For example, m^1A and m^3C directly block the hydrogen bond donor or acceptor, thereby strongly inhibiting base pairing, while m^6A and 5mC have minimal effects (Figure 1.7.a). Some other types of modifications, like agm^2C [159], can also block canonical C·G base pairing by introducing steric hindrance to the Waston-Crick side (Figure 1.7.b). Additionally, a set of RNA modifications at tRNA position 34 can pair with two or more bases, thereby enabling decoding of multiple codons [160-162]. Some of them achieve this through adjusting the acidity or basicity and the tautomerization of

nucleobases by chemical modifications, further impacting hydrogen bond formation, exemplified by 5-methylaminomethyl-2-thiouridine (mnm^5s^2U) [163] (Figure 1.7.c).

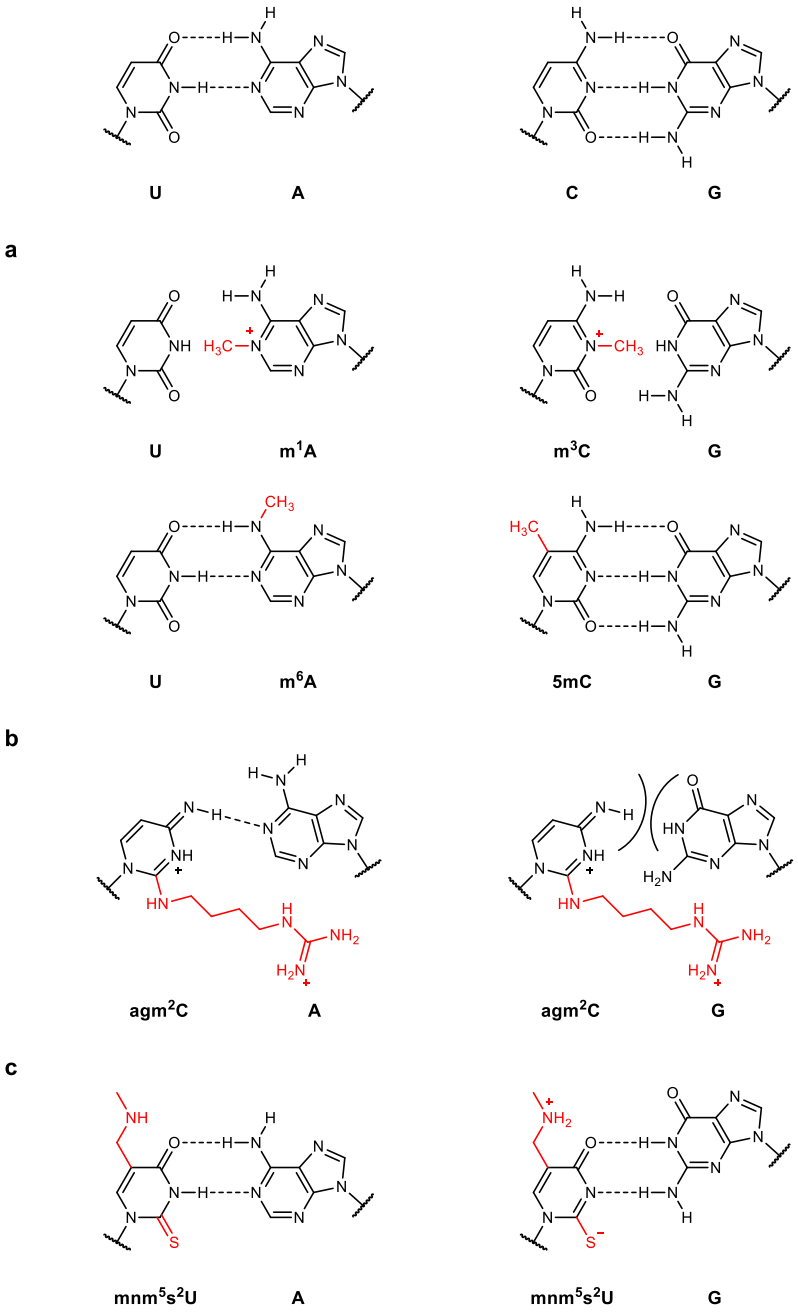


Figure 1.7 | Chemical modifications can alter the base-pairing properties of nucleobases.

Taken together, extensive studies of the chemical properties of DNA and RNA modifications not only enhance our understanding of their unique functions, but also provide various chemical tools to manipulate them.

1.2.3 Bisulfite-mediated deamination of unmodified cytosine to uracil — where all stories began

Among all chemical reactions related to DNA and RNA modifications, bisulfite-mediated deamination is undoubtedly the most famous one (Figure 1.8). This reaction was first independently developed in 1970 by Robert Shapiro group and Hikoya Hayatsu group [164,165]. They discovered that cytosine could be readily converted into uracil using high concentrations of bisulfite treatment under thermal and acidic conditions. Later observations revealed that bisulfite-mediated deamination of 5mC is significantly slower than that of unmodified cytosine [166], making it possible to distinguish them. The reaction mechanism involves an initial Michael addition of a sulfite ion to the cytosine ring, followed by deamination of cytosine, which is defined as the rate-limiting step. Without further workup, the bisulfite reaction will stop at this stage and yield the bisulfite-uracil adduct. To generate uracil as the final product, desulphonation under alkaline conditions is required. As discussed in the previous section, methylation of C^5 significantly reduces the electrophilicity of the C^5-C^6 double bond, impeding the initial Michael addition step. Additionally, detailed mechanistic studies have shown that the optimum pH for bisulfite-mediated deamination of unmodified cytosine is around 5.0 [167,168]. This could be attributed to the balance of two contradictory factors. On the one hand, sulfite ion serves as the actual nucleophile in the reaction, whose concentration increases with higher pH. On the other hand, the protonation of N^3 of cytosine ring occurs

at a lower pH around 4.5, which enhances the electrophilicity of the C^5-C^6 double bond. Despite these insights, some aspects of the mechanism remain unresolved [167,168]. For instance, both research groups noted that very high concentrations of bisulfite induce unexpectedly rapid deamination of unmodified cytosine, suggesting a catalytic role of bisulfite or sulfite ion as a general base. When applied to nucleic acids, bisulfite reaction can only work efficiently on single-stranded DNA or RNA [169], which could be attributed to the combinatory effects of several factors, including the steric hindrance within the double-stranded nucleic acids, the difficulty in protonation and deamination of cytosine when base-paired with guanine, and the electrostatic repulsion of bisulfite and sulfite ion with the negatively charged phosphate backbone. Therefore, alkaline denaturation is required for complete conversion of unmodified cytosines in DNA. Finally, a personal account by Hikoya Hayatsu of his discovery of the bisulfite reaction can be found in ref. [170].

Further research on oxidized cytosine modifications (5hmC, 5fC, and 5caC) revealed novel aspects of bisulfite-mediate deamination (Figure 1.8). The bisulfite reaction with 5hmC is apparently similar to 5mC, in that it does not lead to deamination. However, the hydroxyl group of 5hmC can serve as a leaving group in sulfite Michael addition, finally yielding 5-methylenesulfonate (CMS) product through an apparent $S_{N2'}$ reaction (initial Michael addition with sulfite followed by reverse Michael addition to release water molecule) [171]. This mechanism is further supported by the sulfinate reaction with 5hmC, which will generate CMS derivatives [172]. Notably, glucosylated 5hmC is inert towards bisulfite-mediated deamination [171,173], possibly because glycosyl groups serve as much worse leaving groups compared to the hydroxyl group in 5hmC.

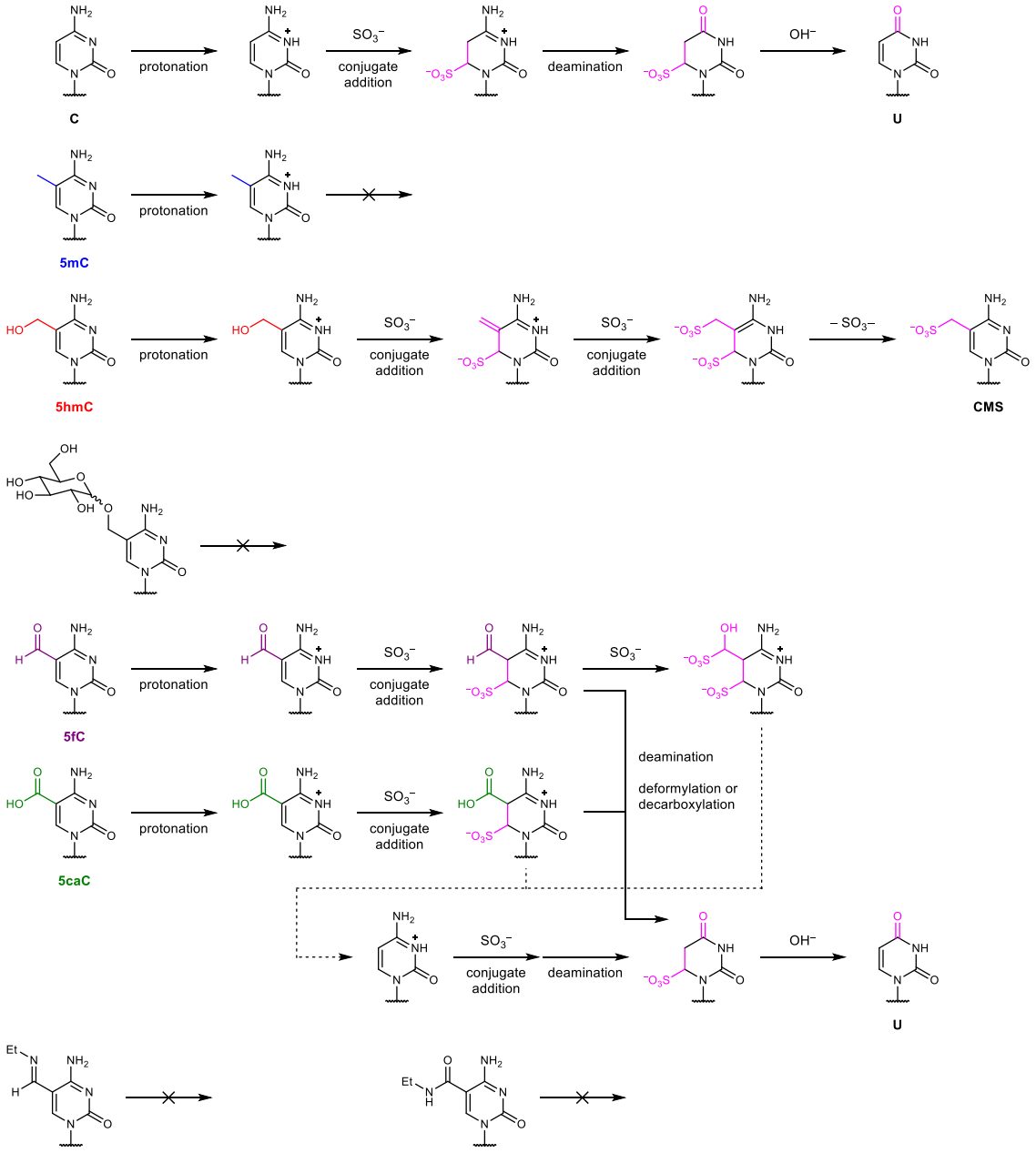


Figure 1.8 | Overview of bisulfite-mediated deamination of unmodified cytosine and cytosine modifications.

In contrast to 5mC and 5hmC, bisulfite treatment will convert 5fC and 5caC into uracil [142,174]. The detailed mechanisms of these two reactions are not available, yet it is believed to undergo similar Michael addition and deamination steps, where the

deformylation or decarboxylation step may happen before or after the deamination. However, it is noted that several studies postulate that the 5fC and 5caC intermediates will also undergo a fragmentation step to release sulfite ion and generate unmodified cytosines, which will be further converted into uracils [142,174]. Interestingly, oximation of 5fC or amidation of 5caC can efficiently block the bisulfite-mediated deamination [144,145], possibly due to the reduction of the electrophilicity and the inability to deformylation or decarboxylation.

With advancements in DNA sequencing and amplification tools, the bisulfite reaction has become one of the most important tools for studying the biological function of 5mC and other DNA cytosine modifications, which finally gives rise to 'bisulfite sequencing'. This revolutionary approach completely changes the study of DNA methylation, marking the inception of epigenetic sequencing.

1.3 Sequencing of DNA and RNA modifications

1.3.1 Overview

DNA and RNA are macromolecules using their unique sequences to store genetic information or realize specific biological functions. Considering the length of nucleic acid (easily up to kb level), deciphering their sequences has long been an almost impossible task until the development of Sanger sequencing and polymerase chain reaction (PCR). Similarly, it is also important to know where the DNA and RNA modifications are, to help us better understand their biological functions and implications. Although several modified bases display non-canonical base pairing properties, most of the nucleic acid modifications behave similarly as the four canonical bases, and therefore could not be

directly detected by sequencing. To overcome this limitation, researchers have utilized the above-mentioned chemical tools to directly convert the DNA and RNA modifications or indirectly transform all unmodified bases into detectable signals, such as mutations, deletions, and truncations, thereby enabling detection through comparing chemical treated samples with untreated controls. Ever since this strategy was first applied in the development of bisulfite sequencing, it has become the major approach to detect DNA and RNA modifications. In this section, I will briefly introduce the history and current aspects of sequencing of DNA and RNA modifications.

1.3.2 Bisulfite sequencing — the current gold standard of DNA methylation sequencing

As mentioned above, a hallmark of sequencing of DNA and RNA modifications is the development of bisulfite sequencing. 22 years after the first discovery of the bisulfite reaction, researchers combined it with PCR and Sanger sequencing to determine the 5mC position in DNA [175]. Upon bisulfite treatment, all the unmodified cytosines will be converted into uracils and finally be read as thymine in PCR, while 5mC remains intact and is still detected as cytosine in sequencing. The whole approach is termed as bisulfite sequencing. In the 21st century, bisulfite sequencing has made tremendous progress, especially when combined with next-generation sequencing (NGS) technologies, which enabled high-throughput sequencing of DNA methylation [176]. Moreover, a series of modified bisulfite sequencing methods have been developed to distinguish different types of DNA cytosine modifications, as summarized in Table 1.7. Nowadays, bisulfite sequencing is considered as the ‘gold standard’ of DNA methylation sequencing and is used in a wide range of biological studies.

Table 1.7 | Bisulfite sequencing and modified bisulfite sequencing for detection of DNA cytosine modifications.

Method	C	5mC	5hmC	5fC	5caC	Ref.
BS-seq	T	C	C	T	T	
oxBS-seq	T	C	T	T	T	[142]
TAB-seq	T	T	C	T	T	[173]
fCAB-seq	T	C	C	C	T	[144]
CAB-seq	T	C	C	T	C	[145]
redBS-seq	T	C	C	C	T	[177]
MAB-seq	C	C	C	T	T	[178]

The success of bisulfite sequencing can be attributed to the following reasons:

1. Base-resolution detection. Bisulfite sequencing induces C-to-T mutations on unmodified cytosines, which achieves 5mC detection at single-base resolution.

2. High selectivity. Bisulfite reaction possesses high conversion rates of unmodified cytosines and low false-positive rates of 5mC, which enables accurate quantification of 5mC modification levels.

3. Easy to implement. Bisulfite reaction does not require special chemical reagents and equipment and can be easily conducted in standard molecular biology lab.

However, there are also two major limitations for bisulfite sequencing:

1. Indirect detection. Bisulfite sequencing converts all unmodified cytosines into uracils. Given that the vast majority (~95%) of cytosines are unmodified, bisulfite sequencing significantly reduces the complexity of sequencing libraries and causes difficulty in subsequent data analysis. Additionally, bisulfite sequencing relies on the complete

conversion of unmodified cytosines, which means incomplete conversion, even in a low frequency, will lead to considerable false-positives and an overestimation of methylation level considering the large amount of unmodified cytosines.

2. Harsh chemical conditions. Bisulfite treatment causes significant degradation of DNA [179], limiting its application to low-input samples, such as cfDNA and single-cell samples.

Therefore, an ideal method which can preserve the advantages and overcome the limitations of bisulfite sequencing is highly desired, to fulfil the growing needs for detection of DNA and RNA modifications.

1.3.3 What makes an ideal sequencing method

Here, I propose three key aspects to realize an ideal sequencing method for DNA and RNA modifications:

1. Absolute single-base resolution, which requires the chemistry to change the base pairing properties of the modification, instead of inducing deletion and truncation signals.

2. Accurate quantification, which requires the chemistry to be highly efficient and selective to achieve high conversion rates and low false-positive rates.

3. Direct detection, which requires the chemistry to convert the modified nucleoside rather than unmodified nucleosides. Compared with indirect detection, direct detection preserves the sequence complexity and therefore has a number of advantages such as higher sequencing quality and mapping rates (Figure 1.9).

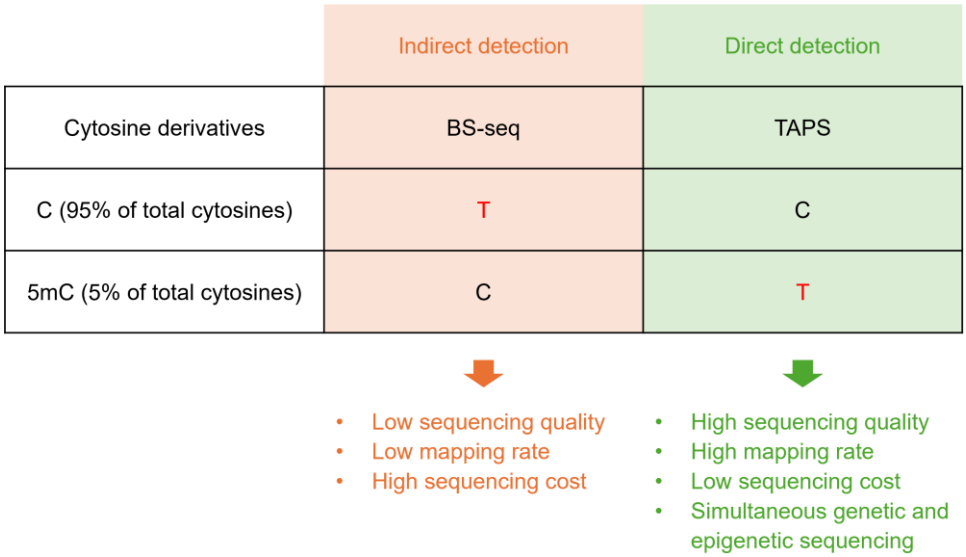


Figure 1.9 | Comparison of indirect and direct sequencing methods for DNA and RNA modification detection.

Several minor aspects are also preferred on top of fulfillment of above three criteria:

1. Mildness, which requires the chemistry not to cause significant DNA and RNA degradation.

2. User-friendly, which requires the chemistry to be simple, straightforward, and easy to implement in standard molecular biology lab. For example, photo-induced chemistry should be avoided due to the inconvenience.

3. Cheap, which requires the chemistry to use common and inexpensive reagents to reduce the cost.

1.3.4 Overview of existing direct, quantitative, and base-resolution sequencing methods for DNA and RNA modifications

Although a number of methods have been developed in the past two decades, only two of them can fully meet the above requirements to serve as ideal sequencing methods for DNA and RNA modifications. Accidentally, both methods utilize borane-mediated reduction of electron-deficient cytosine modifications, with one developed for DNA 5mC and 5hmC sequencing and the other one for RNA ac^4C sequencing (Figure 1.10).

To overcome the limitations of bisulfite sequencing, TET-Assisted Pyridine borane Sequencing (TAPS) was developed by our group in 2019 [180]. Basically, borane compounds, as milder nucleophiles than sulfite ions, were utilized to reduce the 5fC and 5caC into dihydrouracil (DHU), which will be finally read as thymine during PCR amplification. We proposed that borane compounds enabled a conjugate reduction of the C^5-C^6 double bond in the cytosine ring and thus could lead to subsequent deamination and deformylation or decarboxylation process. It is noted that the order of deamination and deformylation or decarboxylation has not been specified and requires more follow-up research. Nevertheless, the borane reduction displayed huge potential to replace bisulfite reaction in DNA methylation sequencing. Indeed, TAPS displayed high conversion rates on 5mC and low false-positive rates on unmodified cytosine, further demonstrating the efficiency and selectivity of borane reduction. Since TAPS does not alter the majority of cytosine bases, it showed higher mapping rates, more even sequencing coverage, and higher base quality compared to bisulfite sequencing. Due to its mild nature, TAPS can also preserve long DNA fragments [181,182] and low-input DNA materials, such as cfDNAs [183]. One potential drawback of TAPS is that it cannot

distinguish between 5mC and 5hmC. This issue was resolved by introducing the β-glucosyltransferase (βGT) blocking step to generate β-glucosyl-5-hydroxymethylcytosine (5gmC) before TET oxidation, which was termed as TAPSβ for 5mC-specfic sequencing [184]. To realize 5hmC-specific sequencing, borane reduction can also be combined with selective oxidation of 5hmC into 5fC. This approach is also known as Chemical-Assisted Pyridine borane Sequencing (CAPS), yet its performances are not satisfactory, which will be discussed in Chapter 2. These results demonstrate that TAPS could replace bisulfite sequencing as the gold standard in DNA methylation sequencing and have widespread applications in both research and clinical settings, such as cfDNA sequencing.

One year later than the development of TAPS, another sequencing method, called ac^{4}C-seq, was developed independently for mapping RNA ac^{4}C modifications [185]. Interestingly, ac^{4}C-seq also uses the borane compounds to reduce ac^{4}C into the tetrahydro ac^{4}C, which will be read as thymine in PCR as well. The proposed mechanism shares many similarities with that of borane reduction of 5fC and 5caC, especially the key initial conjugate reduction of the C^5-C^6 double bond of the cytosine ring. The ac^{4}C-seq was applied to various types of RNA from hyperthermophilic archaea, yeast, and human, which significantly deepens our understanding of ac^{4}C deposition in the transcriptome.

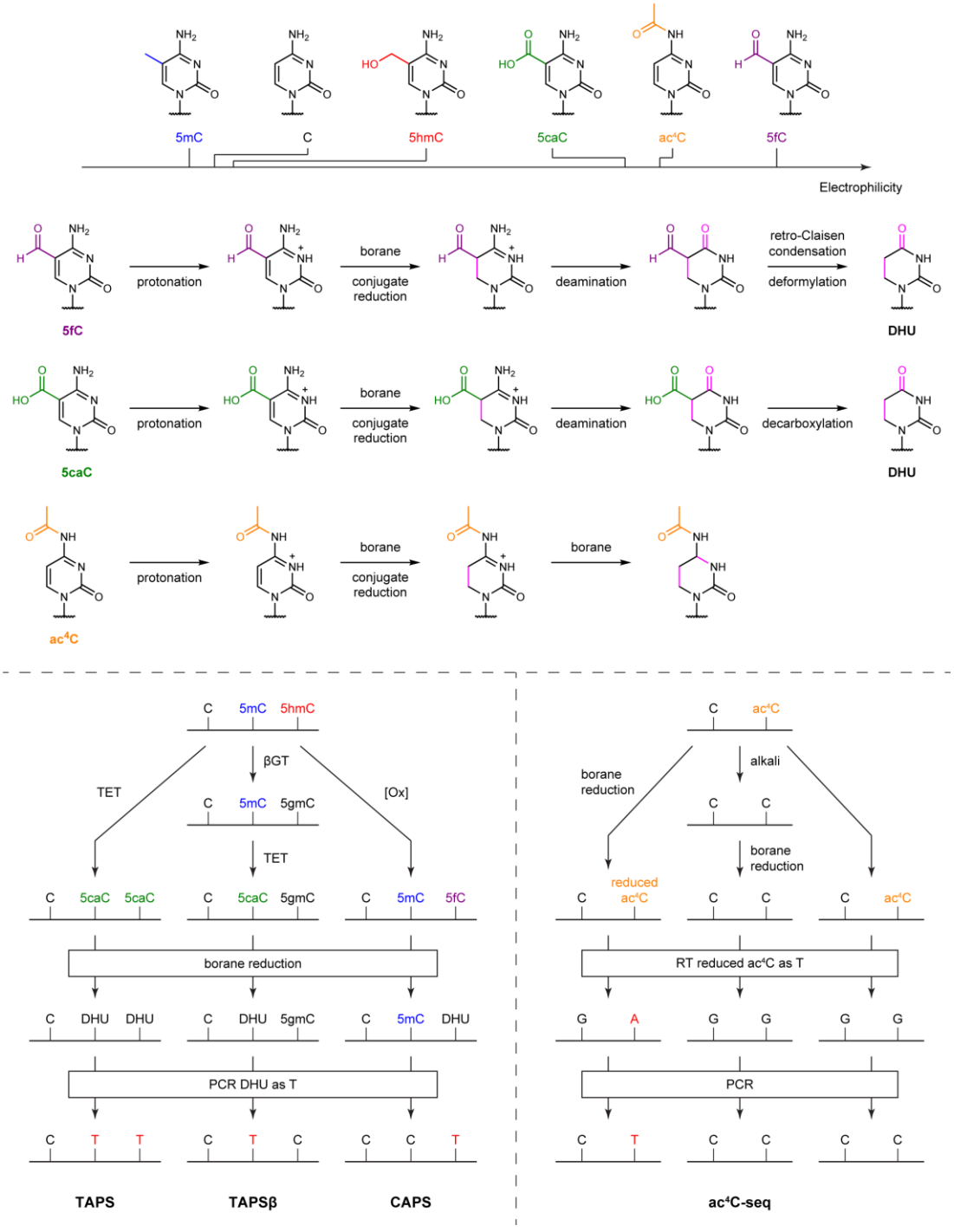


Figure 1.10 | Overview of TAPS and ac^{4}C-seq.

1.4 Research objectives

Given the scarcity of ideal sequencing methods for DNA and RNA modifications, my primary objective is to develop novel chemical tools suitable for direct, quantitative, and base-solution sequencing of DNA and RNA modifications. In this thesis, I describe two novel sequencing methods: CAPS+ for DNA 5hmC sequencing (Chapter 2) and BACS for RNA Ψ sequencing (Chapter 3). Additionally, I applied BACS to studying the biological functions of Ψ modification and PUS enzymes (Chapter 4).

Chapter 2 was published as Xu, et al. *J. Am. Chem. Soc.* **2023**, *145*, 7095–7100. Chapter 3 and part of Chapter 4 were published as Xu, et al. *Nat. Methods* **2024**.

2 Development of CAPS+ for 5hmC sequencing

2.1 Introduction

5hmC is the oxidized product of 5mC and the second most abundant modification in mammalian genomes [32]. As mentioned in Chapter 1, traditional bisulfite sequencing cannot distinguish 5hmC from 5mC. To map genome-wide distribution of 5hmC, the canonical strategy is to oxidize it into 5fC, which was first demonstrated in oxidative bisulfite sequencing (oxBS-seq) using potassium perruthenate ($KRuO_4$) as the oxidant [142]. However, $KRuO_4$ has several limitations. Firstly, it only works on single-stranded DNA due to its negatively charged active species perruthenate (RuO_4^-), thereby requiring an alkaline denaturation step in advance. Secondly, $KRuO_4$ causes significant DNA damage, possibly due to its harsh reactivity. Lastly, $KRuO_4$ has been found to slowly decompose in aqueous solutions, which requires stringent preparation and storage conditions [186]. Although later modification using K_2RuO_4 in chemical-assisted C-to-T conversion of 5hmC sequencing (hmC-CATCH) improved the DNA recovery, the other limitations could not be solved [187]. Furthermore, we found that ruthenium-based oxidation was not fully compatible with the borane reduction, as shown in our original CAPS approach [184]. Therefore, it is necessary to develop a ruthenium-free 5hmC oxidation reaction to improve CAPS performance.

2.2 TEMPO-derived oxoammonium salt oxidation of 5hmC to 5fC

The weaknesses of ruthenium-based oxidants can be attributed to the following two aspects. On the one hand, negatively charged active species (perruthenate RuO_4^- and

ruthenate RuO_4^{2-}) cannot approach double-stranded DNA easily. On the other hand, it has been shown that most of the transition metal-based oxidants would cause significant damage to DNA, due to their unspecific reactivity [188,189]. We therefore aimed to discover a novel 5hmC oxidation using an organic oxidant which is neutral or ideally possesses a positive charge. We noted that two research groups applied nitroxyl radical compounds to catalytic 5hmC oxidation in different strategies and showed some preliminary results in 5hmC sequencing. The first strategy utilized 2-azaadamantane-*N*-oxyl (AZADO) generated *in situ* as the catalyst and bis(acetoxy)iodobenzene (BAIB) as the terminal oxidant, which was also known as a modified approach for Anelli oxidation [190,191]. However, the overall system is not fully compatible with aqueous solution. Later on, another strategy using Stahl approach was developed for 5hmC oxidation, using water soluble copper(II) salt as the catalyst and 2,2,6,6-tetramethylpiperidine-1-oxyl (TEMPO) as the terminal oxidant [143]. Nevertheless, both strategies suffer from low efficiency and selectivity on DNA, due to the complexity of the reaction system. Moreover, the terminal oxidant BAIB in the first approach could directly oxidize DNA rather than the catalyst. Similarly, copper catalyst would also result in DNA degradation due to its intrinsic Lewis acidity and redox activity.

Based on these findings, we envisioned that oxoammonium salts, the oxidative form of nitroxyl radicals which was first synthesized in 1960s, would serve as a better solution to efficient and selective oxidation of 5hmC [192-195]. The oxoammonium salts display several advantages, such as high selectivity, mild reactivity, good solubility and stability in aqueous solutions, and low costs [196,197]. Although oxoammonium salt oxidation has been considered as an old version of modern catalytic TEMPO oxidation, it exhibited

higher comparability with DNA molecules, since it is simple, stoichiometric, and easy to clean up [198]. In contrast to the catalytic 5hmC oxidation, this approach is free of terminal oxidants or transition metal catalysts, thus making it ideal to oxidize 5hmC specifically without causing damage to DNA. More importantly, the oxidative species is the positively charged oxoammonium cation, which can easily approach DNA duplex and overcome the limitations of ruthenium-based oxidants.

To test the feasibility of oxoammonium salt oxidation of 5hmC, we chose ACT^+ BF_4^- (also known as Bobbitt's salt), a kind of TEMPO-derived oxoammonium salt as the oxidant, since it is stable, water-soluble, commercially available, inexpensive, and environmentally benign [196,197] (Figure 2.1.a).

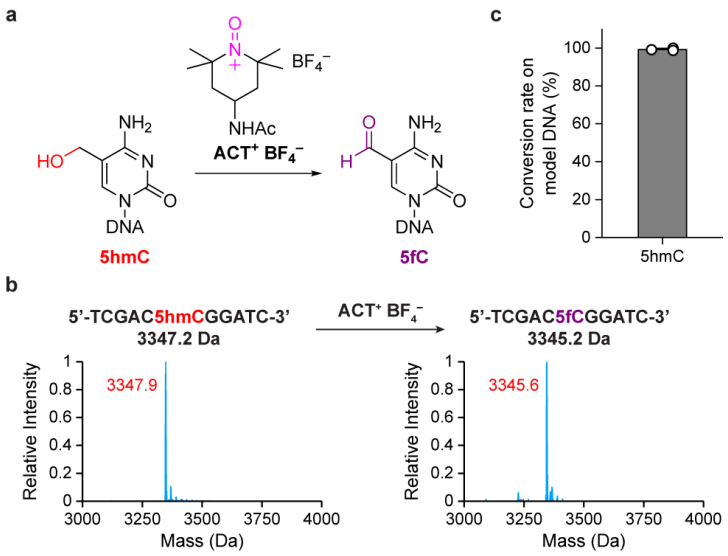


Figure 2.1 | ACT^+ BF_4^- oxidation of 5hmC to 5fC. a. Schematic overview of ACT^+ BF_4^- oxidation. **b.** ESI-MS characterization of an 11mer 5hmC-containing oligonucleotide treated with ACT^+ BF_4^-. Calculated mass is shown in black. Deconvoluted mass is shown in red. **c.** Conversion rate of ACT^+ BF_4^- oxidation of 5hmC on 79 bp model DNA quantified by UHPLC-MS/MS. Data are shown as mean ± s.d. of three independent experiments (n = 3).

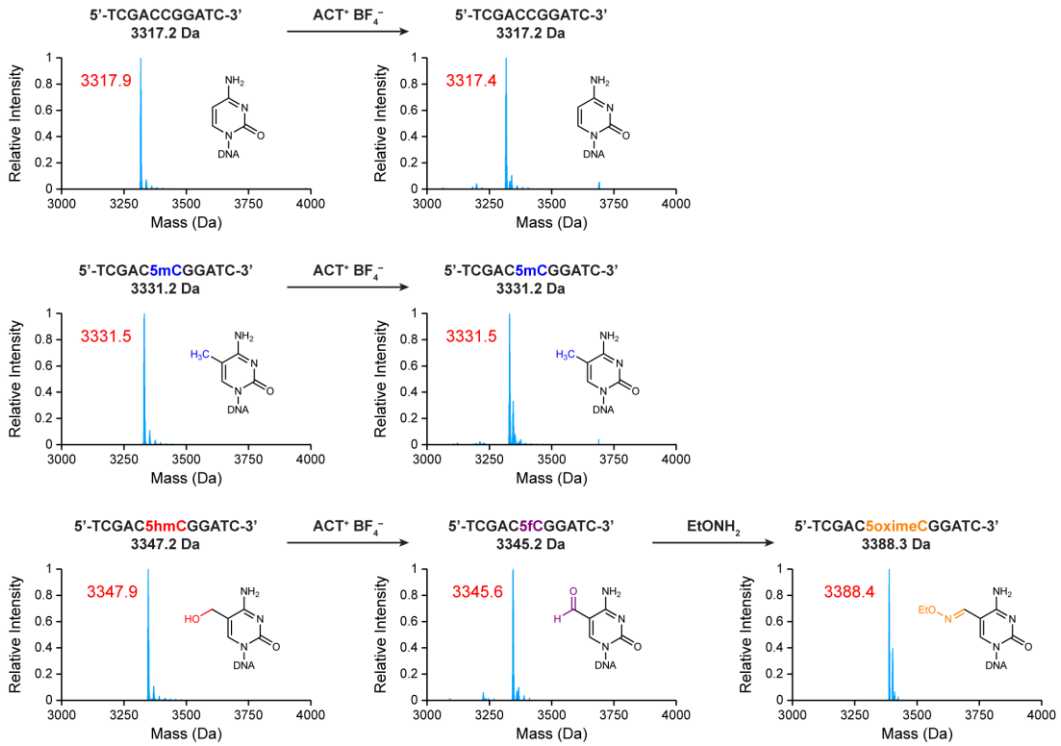


Figure 2.2 | ESI-MS characterization of 11mer oligonucleotide treated with ACT$^+$ BF$_4^-$. Hydroxylamine chemistry was applied to better distinguish the newly generated 5fC oligonucleotide from the 5hmC oligonucleotide [144]. Calculated mass is shown in black. Deconvoluted mass is shown in red. All experiments were performed once.

We first performed the oxidation on an 11mer oligonucleotide that contains one 5hmC site. After ACT$^+$ BF$_4^-$ oxidation, we found that the starting material was transformed into 5fC efficiently, which was confirmed by a 2 Da reduction in mass, detected by electrospray ionization-mass spectrometry (ESI-MS) (Figure 2.1.b). In comparison, an unmodified or 5mC-labeled oligonucleotide was not oxidized under the same condition (Figure 2.2). Next, we applied the ACT$^+$ BF$_4^-$ oxidation to 79 bp double-stranded 5hmC-containing DNA and used ultra-high-performance liquid chromatography-tandem mass spectrometry (UHPLC-MS/MS) to monitor the conversion rate of 5hmC after oxidation. A 99.2% conversion was achieved under the nondenaturing condition (Figure 2.1.c). We

hypothesized that the mechanism of ACT^{+} BF_4^{-} oxidation involved a direct intermolecular hydride transfer step similar to previous predictions [199,200] (Figure 2.3).

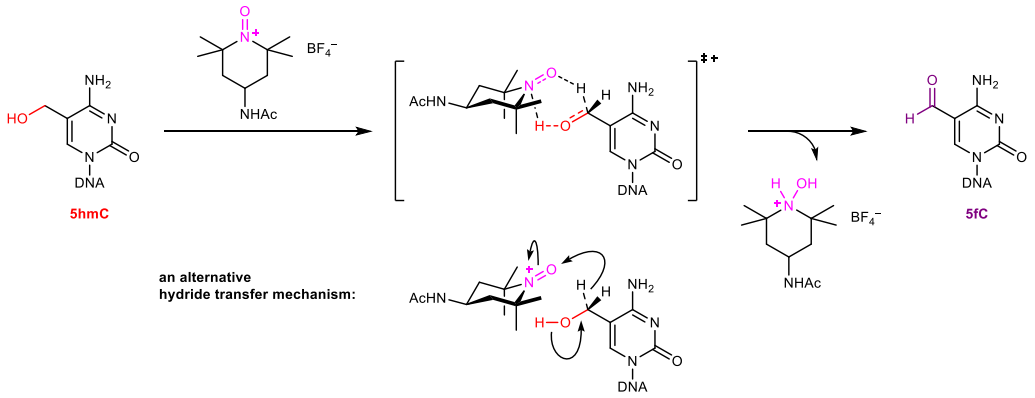


Figure 2.3 | Proposed mechanisms of ACT^{+} BF_4^{-} oxidation of 5hmC into 5fC. A direct intermolecular hydride transfer step is involved.

2.3 Pinnick oxidation of 5fC to 5caC

During the development of CAPS, we found that 5fC is a suboptimal substrate compared to 5caC in borane reduction [184]. Interestingly, this trend is opposite to their electrophilicity order. This phenomenon could be attributed to the softer nature of C^5-C^6 double bond and stronger basicity of N^3 in 5caC [201,202]. Consequently, we aimed to further oxidize 5fC into 5caC to improve the performance of CAPS. It is interesting that 5fC is much more inert than 5hmC towards oxidation, since all the aforementioned 5hmC oxidation reactions will stop in the 5fC stage, without overoxidation to 5caC [142,143,190,191]. This phenomenon could be explained by the low equilibrium constant of 5fC to form the key *gem*-diol intermediate in aqueous solutions under ambient conditions ($K_w < 4.5\times10^{-4}$) [203] (Figure 2.4). Inspired by this finding, we aimed to focus on aldehyde oxidation through a unique mechanism without the formation of *gem*-diol

intermediates. Finally, we envisioned that the Pinnick oxidation using sodium chlorite ($NaClO_2$) would be most suitable for this purpose, since it can work on sterically hindered aldehydes and is known to tolerate a wide range of functional groups, including alcohols, electron-deficient aromatic amines, and heterocycles [204-207] (Figure 2.5).

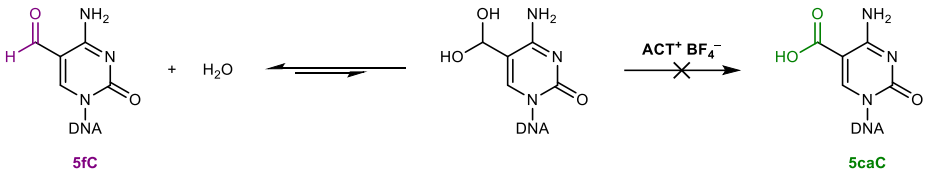


Figure 2.4 | Schematic explanation of the resistance of 5fC towards ACT^+ BF_4^- oxidation.

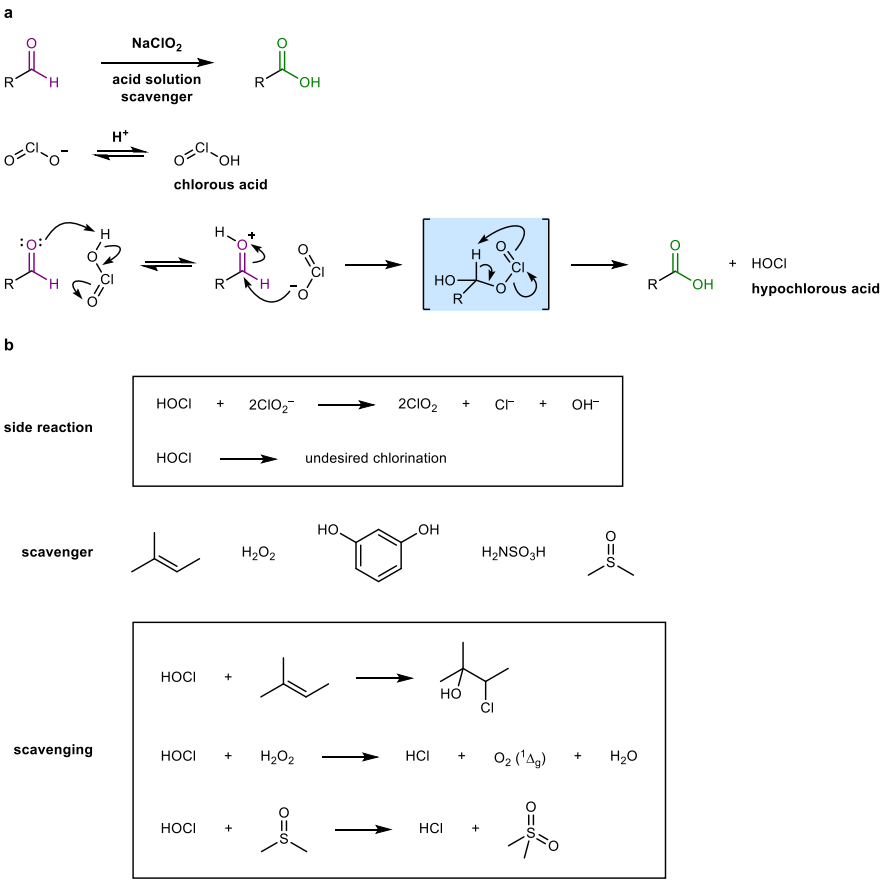


Figure 2.5 | Mechanistic overview of Pinnick oxidation. a. Proposed mechanisms of Pinnick oxidation. **b.** Common scavengers in Pinnick oxidation and their scavenging reactions.

To test the Pinnick oxidation, we started with an 11mer oligonucleotide that contains one 5fC modification and monitored the reaction by matrix-assisted laser desorption/ionization mass spectrometry (MALDI). We chose 2-methyl-2-butene as the scavenger (Figure 2.6.a), and after Pinnick oxidation, we found that 5fC was completely converted into 5caC, which was confirmed by a 16 Da increase in mass (Figure 2.6.b). In comparison, unmodified cytosine, 5mC-labeled, or 5hmC-labeled oligonucleotides were not oxidized (Figure 2.7). Moreover, nearly quantitative conversion was observed on 99 bp 5fC-contaning double-stranded DNA under the optimized reaction condition (Figure 2.6.c). To the best of our knowledge, this is the first successful selective chemical oxidation of 5fC to 5caC on DNA.

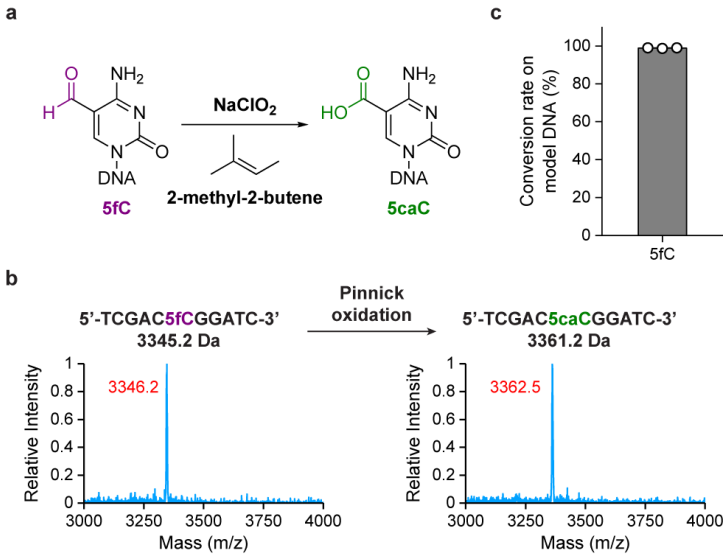


Figure 2.6 | Pinnick oxidation of 5fC to 5caC. a. Schematic overview of Pinnick oxidation using $NaClO_2$ as the oxidant and 2-methyl-2-butene as the scavenger. **b.** MALDI characterization of Pinnick oxidation of 11mer 5fC-containing oligonucleotide. Calculated mass is shown in black.

Observed mass is shown in red. **c.** Conversion rate of Pinnick oxidation of 5fC on 99 bp model DNA quantified by UHPLC-MS/MS. Data are shown as mean ± s.d. of three independent experiments (n = 3).

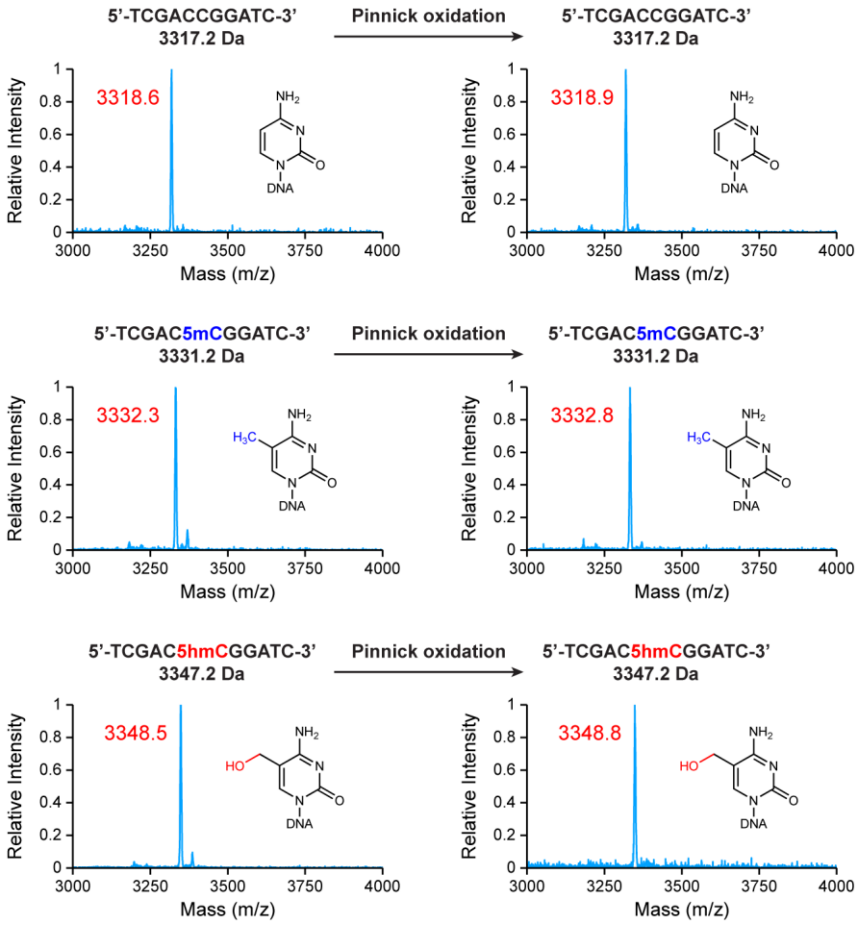


Figure 2.7 | MALDI characterization of Pinnick oxidation of 11mer oligonucleotide. Calculated mass is shown in black. Observed mass is shown in red. All experiments were performed once.

2.4 CAPS+ for improved 5hmC sequencing

Previously, by combining K_2RuO_4 oxidation and borane reduction, we have developed CAPS for 5hmC sequencing [184] (Figure 2.8). However, it has a relatively low

conversion rate and a high false-positive rate, possibly due to the following reasons. On the one hand, the denaturing ruthenate oxidation may not be compatible with the nondenaturing borane reduction. On the other hand, 5fC could be a worse substrate for borane reduction than 5caC according to the genome-wide sequencing results of CAPS and related methods (~95% conversion rate of 5caC, 75−85% conversion rate of 5fC) [184], although they showed similar reactivity on short model DNA [180]. Therefore, we combined nondenaturing ACT^{+} BF_4^{-} oxidation of 5hmC and Pinnick oxidation of 5fC with borane reduction to develop Chemical-Assisted Pyridine borane Sequencing plus (CAPS+) for improved base-resolution sequencing of 5hmC (Figure 2.8).

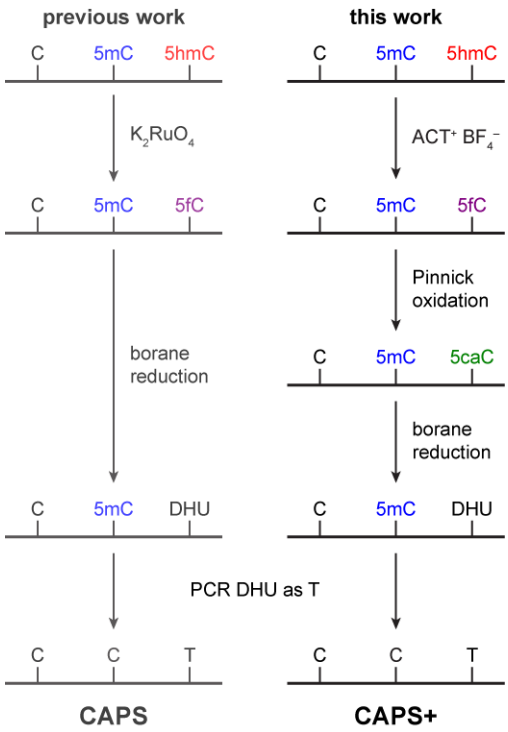


Figure 2.8 | Comparison between CAPS and CAPS+.

We first applied CAPS+ to fragmented mESC genomic DNA (gDNA) and compared the efficiency of two new oxidation reactions with the K_2RuO_4 oxidation previously used in

CAPS using UHPLC-MS/MS (Figure 2.9). In gDNA, ACT^+ BF_4^- oxidation showed higher efficiency than K_2RuO_4 oxidation, achieving 99.0% oxidation of 5hmC to 5fC, compared to 95.5% in K_2RuO_4 oxidation. Interestingly, we observed a slight increase of 8-oxoG (one of the most common mutagenic oxidative damages [208]) in gDNA after K_2RuO_4 oxidation (Figure 2.10). In contrast, ACT^+ BF_4^- and Pinnick oxidation did not lead to 8-oxoG formation, which further demonstrated their superior selectivity and mildness.

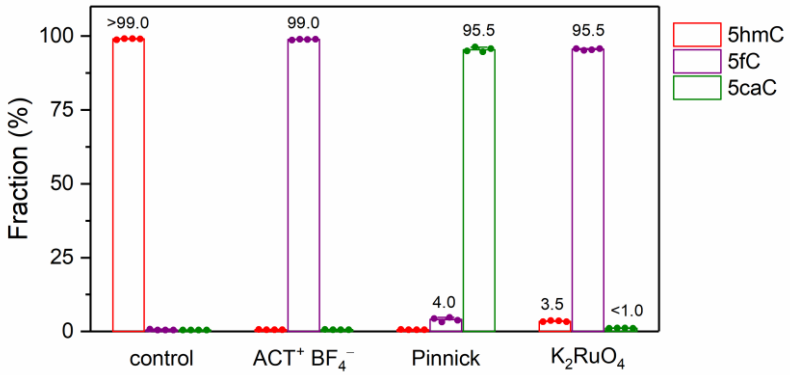


Figure 2.9 | Comparisons of efficiency of ACT^+ BF_4^- oxidation, Pinnick oxidation (both used in CAPS+), and ruthenium-based oxidation (used in CAPS) on 300–500 bp fragmented mESC gDNA. Data are shown as mean ± s.d. of four independent experiments (n = 4).

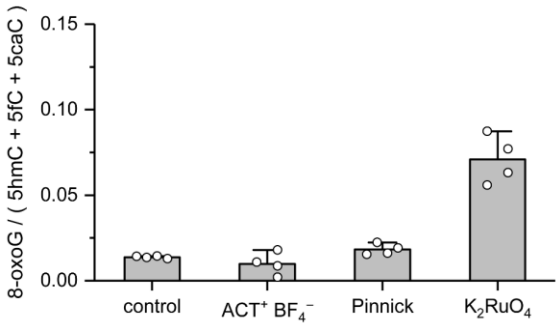


Figure 2.10 | Comparisons of 8-oxoG formation after ACT^+ BF_4^- oxidation, Pinnick oxidation (both used in CAPS+), and ruthenium-based oxidation (used in CAPS). 8-oxoG level is normalized to the sum of oxidized cytosine modification level (5hmC + 5fC + 5caC). Data are shown as mean ± s.d. of four independent experiments (n = 4).

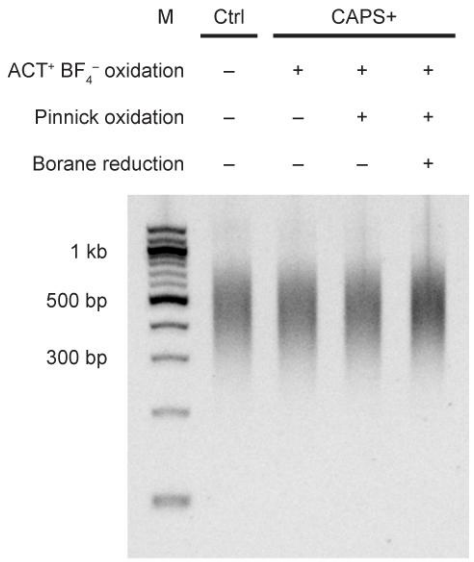


Figure 2.11 | No detectable degradation of ligated DNA was observed on 2% agarose gel after CAPS+. Experiment was performed once.

As expected, CAPS+ did not cause notable DNA degradation after each step of chemical treatment (Figure 2.11). Based on the analysis of spike-in controls (including synthetic 5hmC spike-in, CpG-methylated lambda DNA, and 2 kb unmodified spike-in), we found that CAPS+ displayed a higher conversion rate on 5hmC (94.5%) and lower false-positive rates on 5mC (0.15%) and unmodified cytosines (0.17%) than CAPS (83.1% for 5hmC, 0.38% for 5mC, and 0.72% for unmodified cytosines) [184] (Figure 2.12). The 5hmC-to-T conversion rate of CAPS+ was comparable with the 5hmC protection rates of APOBEC-coupled epigenetic sequencing (ACE-seq) (98.5%) [209] and TET-assisted

bisulfite sequencing (TAB-seq) (92.0%) [173], while the 5mC and unmodified cytosine false-positive rates of CAPS+ were similar to or lower than the reported nonconversion rates in ACE-seq (0.5% and 0.1%, respectively) [209] and TAB-seq (2.2% and 0.4%, respectively) [173].

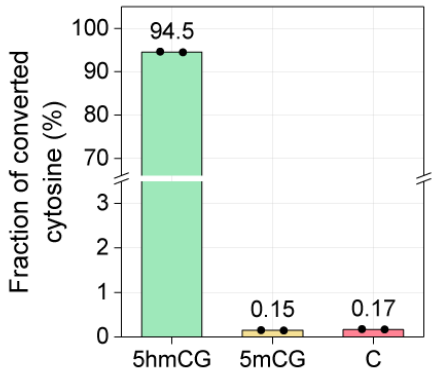


Figure 2.12 | Conversion rate for synthetic spike-in (5hmCG) and false-positive rates for CpG-methylated lambda DNA (5mCG) and 2 kb unmodified spike-in control in CAPS+. Data are shown as means of two technical replicates.

Next, we examined the 5hmC raw signals from two technical replicates of CAPS+, which revealed a strong correlation between them (Pearson's $r = 0.85$) (Figure 2.13). We then merged reads from two replicates and found that CAPS+ showed good correlation with CAPS (Pearson's $r = 0.87$) (Figure 2.14) and other published data sets (Pearson's $r = 0.64$ with ACE-seq and 0.79 with TAB-seq) (Figure 2.14). We detected a mean 5hmCG level of 2.3% with CAPS+, compared to 2.7% detected with CAPS [184], which could be attributed to the low false-positive rates of CAPS+. Additionally, CAPS+ maintained high sequencing quality of CAPS and therefore exceeded ACE-seq and TAB-seq in mapping rate (91.1% for CAPS+, 90.7% for CAPS [184], 71.8% for ACE-seq [209], and 53.4% for TAB-seq [173]), base quality (Figure 2.15), and sequencing coverage uniformity

(Figure 2.16). Comparison of results from different methods demonstrated good consistency between CAPS+ and other approaches (Figure 2.17). We also analyzed the genomic distribution and enrichment profile of 5hmC sites, both of which were in agreement with previous reports [173,184,209] (Figure 2.18).

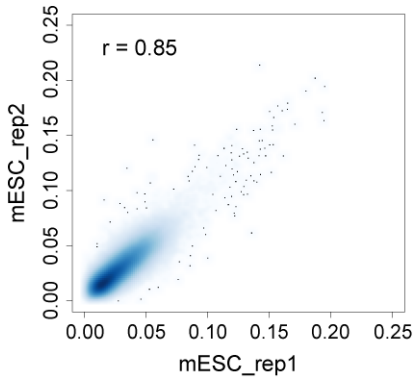


Figure 2.13 | Correlation density plot between two technical replicates of CAPS+ in 10 kb bins. The color scale represents density.

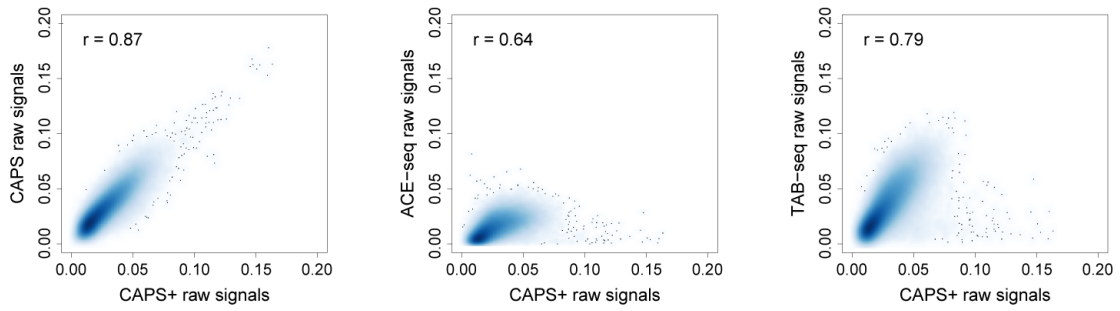


Figure 2.14 | Correlation density plot between CAPS+, CAPS, ACE-seq, and TAB-seq in 10 kb bins. The color scale represents density.

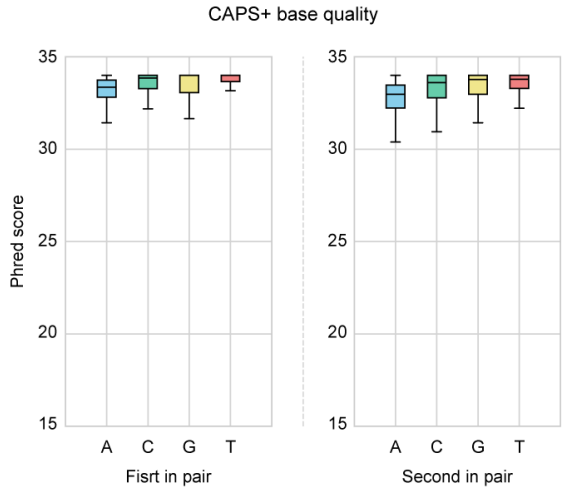


Figure 2.15 | Sequencing base quality of CAPS+. Reads from two replicates of CAPS+ were merged. Nucleotide is denoted by color. Boxplots visualize all sequencing reads showing medians, upper and lower fourth quantiles and non-outlier extreme values (n = 10 million random sequencing reads).

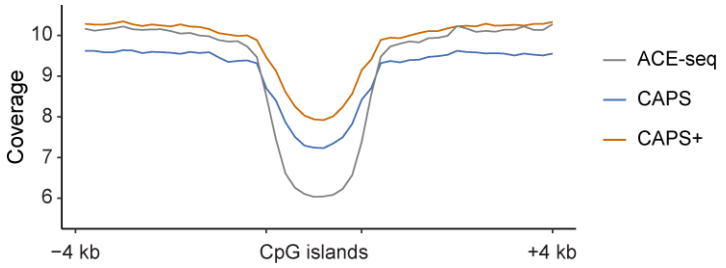


Figure 2.16 | Average sequencing coverage of CAPS+, CAPS, and ACE-seq at all CpG islands (CGI) and 4 kb flanking regions. Reads from two replicates of CAPS+ were merged and subsampled to the similar coverage of ACE-seq.

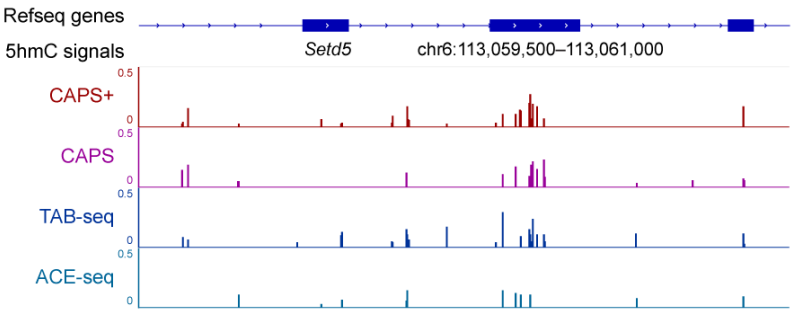


Figure 2.17 | Example of genome browser view demonstrating that CAPS+ detected consistent 5hmC sites with CAPS, TAB-seq, and ACE-seq.

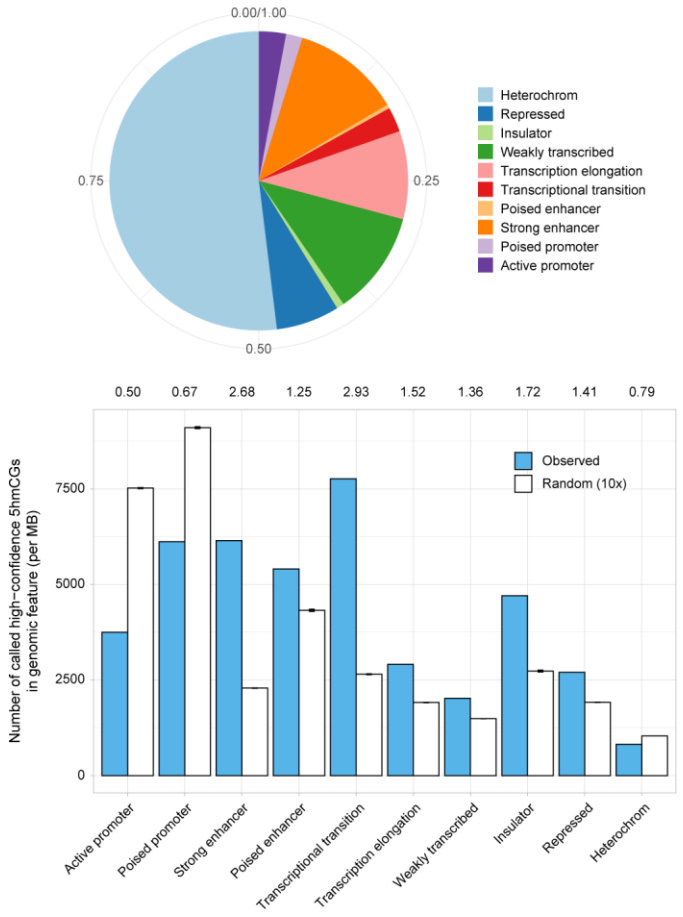


Figure 2.18 | Pie chart shows the overlap of called 5hmCGs with putative genomic regulatory elements and the relative enrichment of 5hmCGs (blue) and random sites (white)

at genomic regulatory elements in mESC. 'Random' consists of ten random samplings. The mean is shown as the bar height and the error bars denote standard deviation (n = 10 random sampling events). The ratios between observed and random are shown at the top.

To further demonstrate the utility of CAPS+ in clinical samples, we applied it to generate base-resolution maps of hydroxymethylome from human normal brain and glioblastoma gDNA. In normal brain, we detected a high 5hmC level as anticipated (mean 18.5% on CpG sites) (Figure 2.19). We then compared 5hmC raw signals from CAPS+ with those from TAB-seq in adult prefrontal cortex [210] and found a good correlation between them (Pearson's r = 0.89) (Figure 2.20). In stark contrast to the normal brain, we observed a sharp decrease of global 5hmCG level to 2.5% in glioblastoma, which was in accordance with previous studies [211] (Figure 2.19).

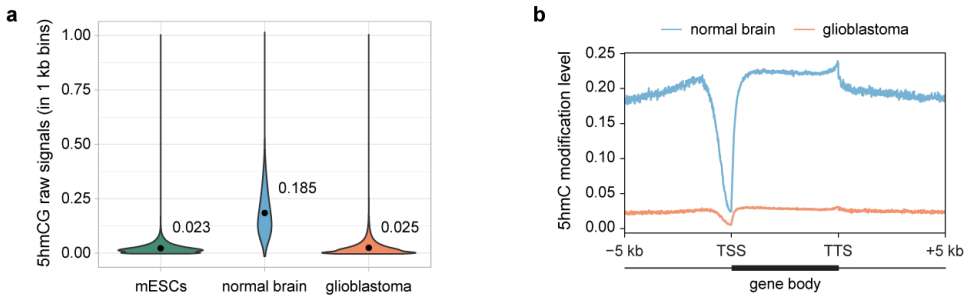


Figure 2.19 | 5hmC profile of human normal brain and glioblastoma. a. Violin plots comparing 5hmCG raw signals in 1 kb bins between mouse embryonic stem cells (mESCs) (green), normal brain (blue), and glioblastoma (orange), with mean values listed above each plot. **b.** Metagene profiles of 5hmC in normal brain (blue) and glioblastoma (orange). TSS and TTS indicate the transcription start site and the transcription termination site, respectively.

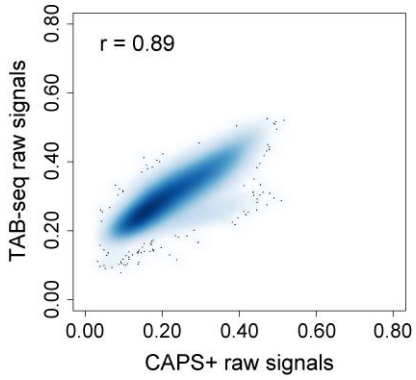


Figure 2.20 | Correlation density plot between CAPS+ signals in human normal brain and TAB-seq signals in adult prefrontal cortex in 10 kb bins. The color scale represents density.

2.5 Summary

In this chapter, I introduced the development of two novel chemical reactions for selective oxidation of cytosine modifications: ACT^+ BF_4^- oxidation of 5hmC into 5fC and Pinnick oxidation of 5fC into 5caC. Both reactions are highly efficient and nondestructive. In addition, they are fully compatible with double-stranded DNA and are milder than ruthenium-based oxidation with less oxidative damage. These reagents are also commercially available and easy to use without the need for special preparation and storage as for the ruthenium compound [142,187]. We expect ACT^+ BF_4^- oxidation to replace ruthenium as the new standard in 5hmC oxidation, and Pinnick oxidation, by converting 5fC to 5caC, to further improve its conversion in borane reduction and bisulfite reaction.

We further combined them with borane reduction to establish CAPS+, an updated version of CAPS, for base-resolution sequencing of 5hmC modifications. CAPS+ showed higher conversion rates of 5hmC and lower false-positive rates of 5mC and unmodified cytosine

than CAPS. CAPS+ can work on double-stranded DNA without denaturation and also inherited several advantages of TAPS, including high mapping rates, even genome coverage, and high base quality. Therefore, we expect CAPS+ to be widely used for 5hmC sequencing. For instance, CAPS+ have been used together with TAPSβ to unravel the dynamics of 5hmC and 5mC during human germline development and reveal the role of DMRT1 in germline commitment [212].

3 Development of BACS for Ψ sequencing

3.1 Introduction

Ψ, as the C-C glycoside isomer of uridine, is the most abundant post-transcriptional modification in cellular RNA [213]. It is not only prevalent in nearly all kinds of ncRNAs, including rRNAs, snRNAs, and tRNAs [102,103], but has also been known to be present in messenger RNA (mRNA) [104,105]. Ψ plays an important role in splicing, translation, RNA stability, and RNA-protein interactions [108]. In eukaryotes, Ψ is installed by various PUS enzymes [214,215], which have been shown to associate with many diseases including cancer (see Table 1.2). Therefore, establishing an accurate and sensitive method to detect Ψ is highly desirable.

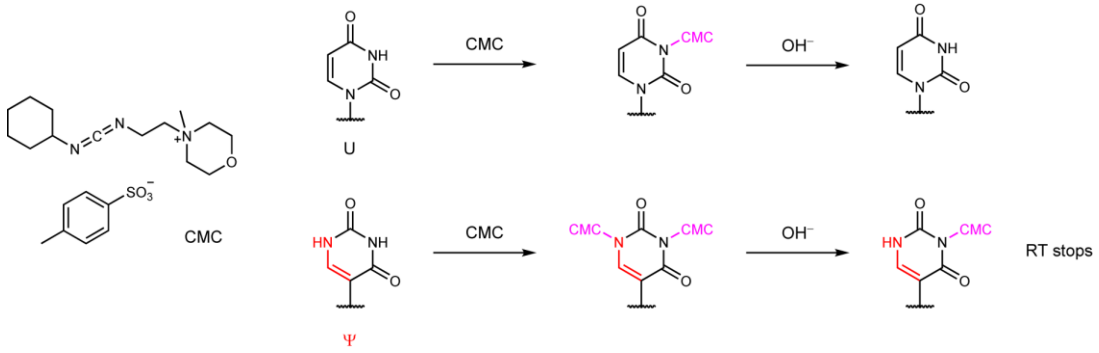


Figure 3.1 | Overview of CMC chemistry for Ψ detection.

Traditionally, the detection of Ψ has been reliant on the *N*-cyclohexyl-*N*'-(2-morpholinoethyl)carbodiimide methyl-*p*-toluenesulfonate (CMC) chemistry [216] (Figure 3.1). CMC can react with amide or imide functional group in nucleobases (for example, amide for guanosine and imide for uridine) [217,218], while it can form a more

stable adduct with N^3 of Ψ, therefore enabling discrimination of Ψ with uridine through subsequent alkaline treatment (pH 10.4) [219]. Since N^3-CMC adduct of Ψ strongly interferes with base pairing and would lead to reverse transcription (RT) truncations, CMC chemistry has been widely applied to transcriptional-wide mapping of Ψ, as shown in Pseudo-seq [104], Ψ-seq [105], PSI-seq [220], and CeU-seq [221]. However, CMC-based methods have low labeling efficiency and selectivity for Ψ, making it intrinsically difficult to distinguish between true Ψ signals and background noise emerging from other bases and RNA secondary structures [222,223]. Consequently, CMC-based methods lack stoichiometry information of Ψ.

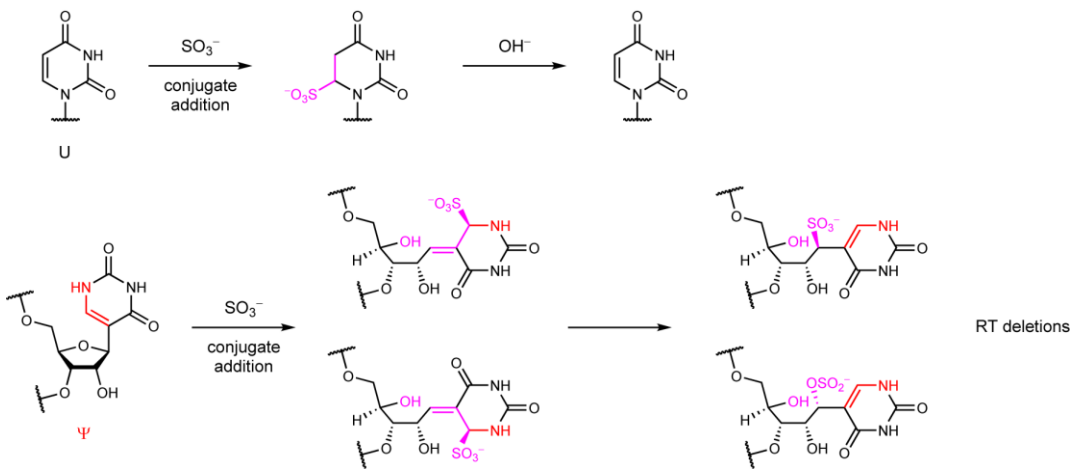


Figure 3.2 | Overview of bisulfite chemistry for Ψ detection.

Recently, RBS-seq reexamined the bisulfite (BS)-mediated conversion of Ψ [224,225], and surprisingly found that the Ψ-BS adduct could lead to deletion signatures during RT, thus providing a novel way to detect Ψ (Figure 3.2) [226,227]. Since unmodified cytosine would also be deaminated to uridine in conventional bisulfite reaction [164,165], BID-seq and similarly designed PRAISE further optimized bisulfite treatment at near neutral

pH to eliminate most of the side reaction on unmodified cytosine, enabling quantitative detection of Ψ [228,229]. Although BID-seq substantially improved this approach, it still suffered from low deletion rates in certain sequence contexts. Crucially, due to the deletion signature, BS-based methods cannot determine the exact position of Ψ in consecutive uridine sequences or consecutive Ψ sites and struggle to detect densely modified Ψ sites. Although local realignment can partially solve this issue, it can generate artefacts including overestimation of Ψ modification level and misidentification of sequence context [230]. Therefore, it is necessary to develop novel chemistry to improve the Ψ sequencing.

3.2 Selective labeling of Ψ using 2-bromoacrylamide-mediated cyclization

Given that the limitations of CMC- and BS-based methods originate from the chemistry and readout of sequencing, we aimed to develop a novel reaction which can induce mutation signals for Ψ. We noted that Ψ possesses a highly reactive N^1 towards various kinds of Michael addition receptors, including acrylonitrile [151,152,231-233], acrylamide [234], and other acrylic compounds [235] (Figure 3.3). Such reactions have been widely applied to distinguish Ψ from uridine through mass spectrometry [236]. However, these products cannot induce mutations of Ψ during reverse transcription. Further inspired by the mutation profile of O^4-methyl thymidine [237], we envisioned that installing an α-leaving group (like halogen) to the acrylic compounds would lead to an intramolecular cyclization, which could result in a Ψ-to-C mutation (Figure 3.3 and 3.4.a).

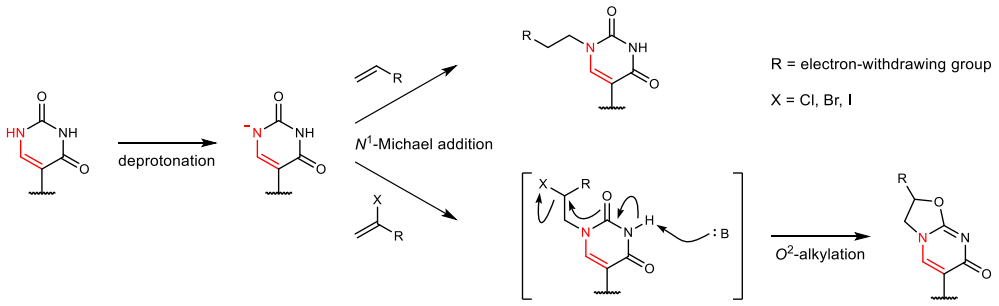


Figure 3.3 | Design of BACS.

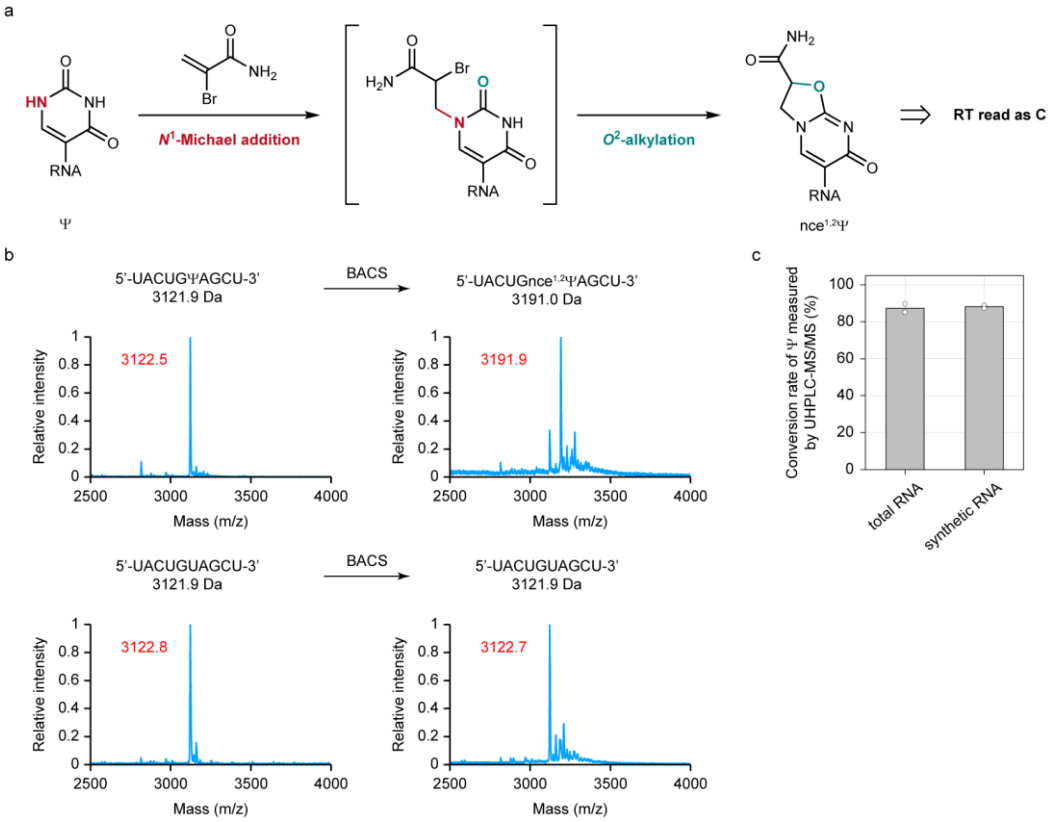


Figure 3.4 | Overview of BACS for Ψ detection. a. Schematic overview of BACS reaction. **b.** MALDI characterization of BACS labeling of a 10mer Ψ-containing RNA oligonucleotide. Calculated mass is shown in black. Observed mass is shown in red. Experiment was performed once. **c.** Conversion rates of Ψ in HeLa total RNA and 1.8-kb 10% Ψ-modified synthetic RNA upon BACS treatment, quantified by UHPLC-MS/MS. Data are presented as means of two independent experiments.

As expected, we treated the Ψ-labeled RNA with 2-bromoacrylamide and yielded the intramolecular cyclized product carbamido-1,O^2-ethano Ψ (nce^{1,2}Ψ) (Figure 3.4.b). This reaction was further confirmed by ultra-high-performance liquid chromatography-tandem mass spectrometry (UHPLC-MS/MS) (Figure 3.4.c). To validate the mutation signature of nce^{1,2}Ψ, we applied this chemistry to a 72mer model RNA which contains two Ψ sites. Through RT and next-generation sequencing, we obtained 86.6% U-to-C mutation rates on these two sites, while U-to-R (R = A or G) mutation rates were lower than 1% (Figure 3.5.a). Therefore, we confirm that the U-to-C mutation rate can serve as the conversion rate identified from sequencing. We named this novel approach as 2-Bromoacrylamide Assisted Cyclization Sequencing (BACS) for absolute quantitative and base-resolution sequencing of Ψ.

To understand the sequence preference of BACS, we built libraries with synthetic 30mer RNA spike-in containing NNΨNN and NNUNN (N = A, C, G or U), respectively (Figure 3.5.b). After BACS, we observed an 87.6% conversion rate of Ψ and a 0.75% false-positive rate of uridine when accumulating all motifs. Among all the 256 motifs, 230 showed a conversion rate higher than 85% and 254 displayed a conversion rate higher than 80%, suggesting the high efficiency of BACS chemistry. We observed a low false-positive rate (<1%) in most motifs (213 out of 256 motifs). Certain motifs, especially those with one or more cytidines 5'- or 3'-flanking to the uridine site (for example, GCUCC and ACUCC), displayed slightly higher false-positive rates (3–4%). Nevertheless, BACS clearly showed higher conversion rates than BID-seq both in general and in specific motifs [228]. By mixing NNΨNN and NNUNN spike-in in different ratios,

we generated excellent linear calibration curves for accurate quantification of Ψ modification levels (R^2 = 1.00, Figure 3.5.c).

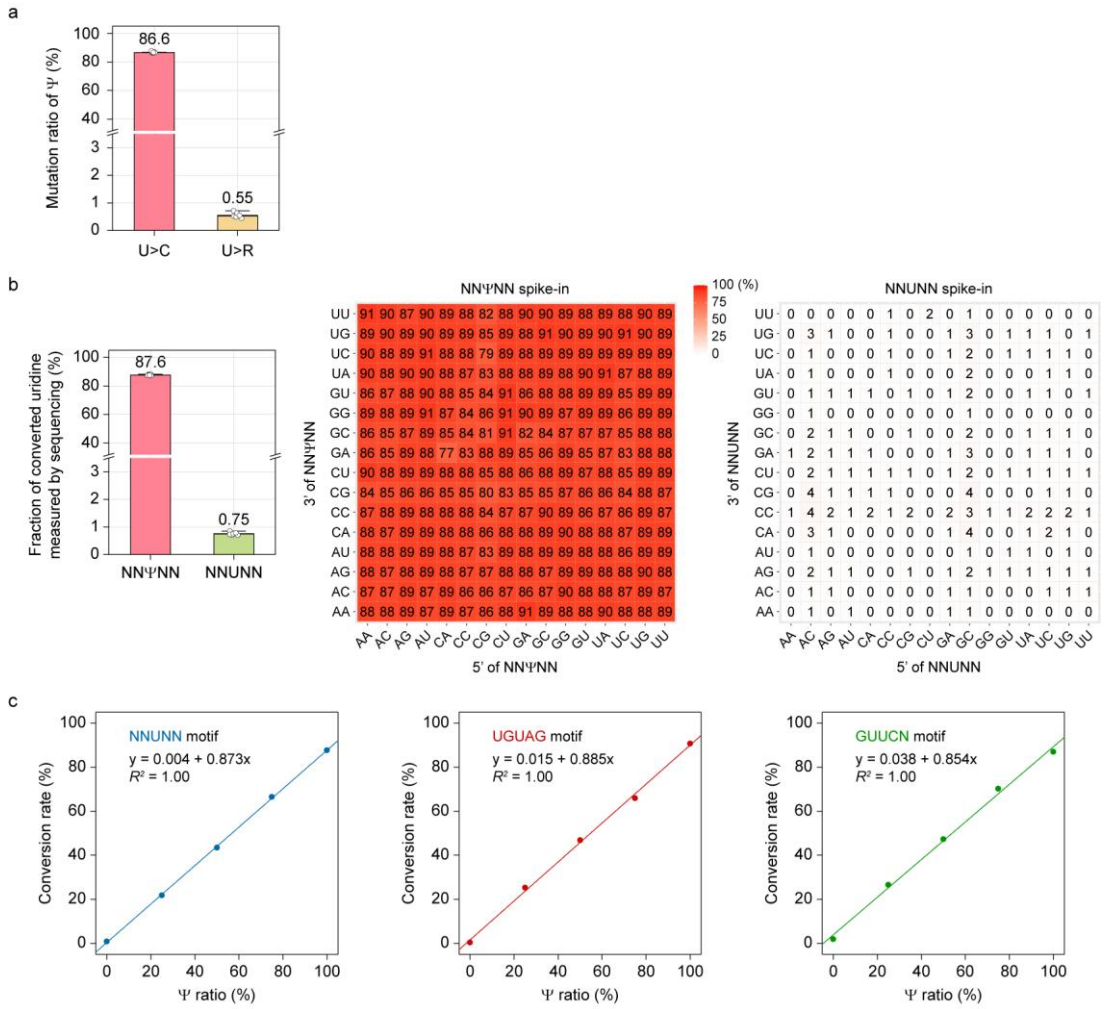


Figure 3.5 | Sequencing performances of BACS. a. Mutation ratios of Ψ sites in 72mer model RNA after BACS treatment. Data are shown as means ± s.d. of six independent experiments (n = 6). **b.** Cumulative (left) and motif-dependent (middle and right) results of BACS conversion rates and false-positive rates on synthetic 30mer NNΨNN and NNUNN spike-in. Data are shown as means ± s.d. of six independent experiments (n = 6). **c.** BACS calibration curve for quantification of Ψ stoichiometry in NNUNN (blue), UGUAG (red), and GUUCN (green) motif. Experiment was performed once.

3.3 Validation of BACS on human rRNA

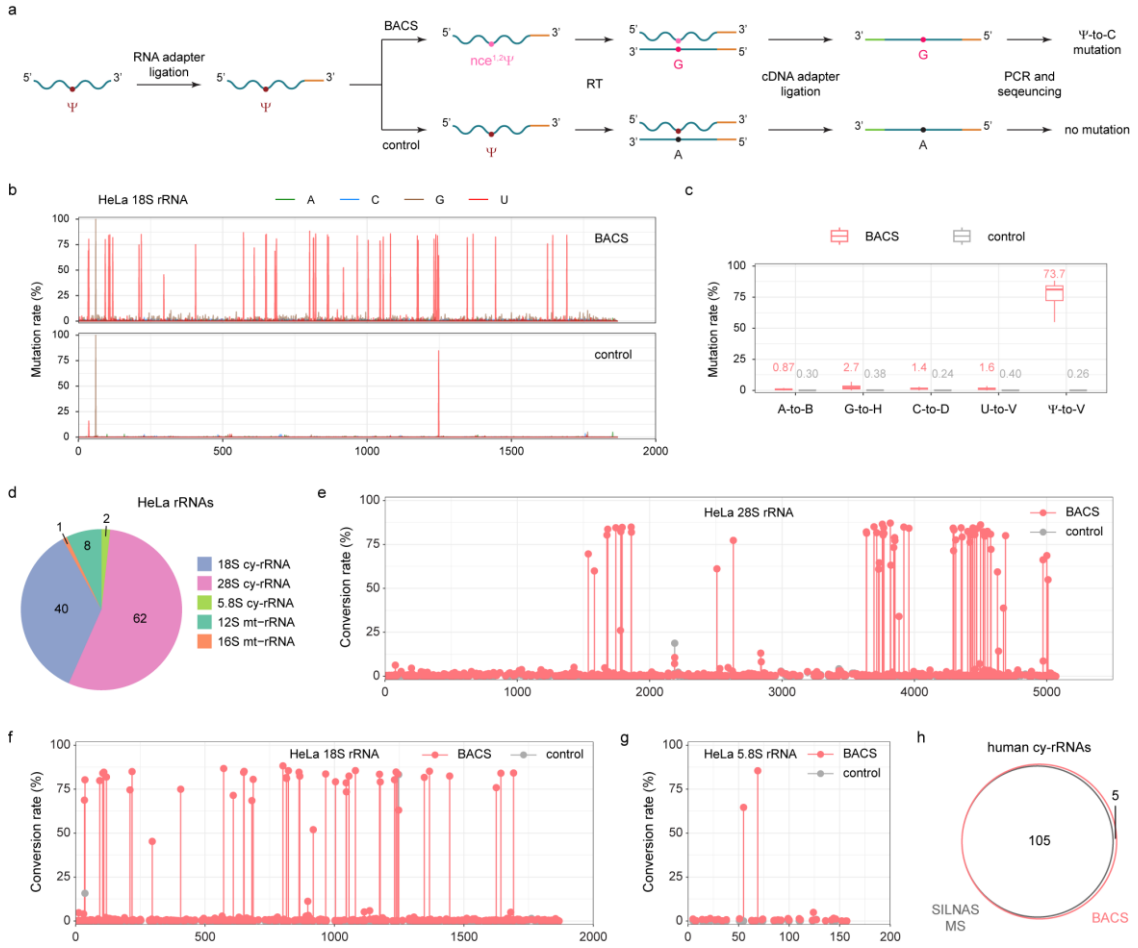


Figure 3.6 | BACS detected known Ψ sites in human rRNA. a. Flowchart of BACS library construction. **b.** BACS and control mutation rates of all bases in HeLa 18S rRNA. **c.** Mutation rates of Ψ and other bases in HeLa cy-rRNAs. For boxplots, boxes represent the 25th to 75th percentiles with a line at the median; whiskers correspond to 1.5 times the interquartile range; mean values are listed above each plot (A-to-B, n = 1268; G-to-H, n = 2459; C-to-D, n = 2272; U-to-V, n = 1113; Ψ-to-V, n = 105). B = C, G or U; H = A, C or U; D = A, G or U; V = A, C or G. **d.** Numbers of Ψ sites identified in HeLa cy-rRNAs and mt-rRNAs. **e–g.** Conversion rates of BACS (pink) and control (grey) samples in HeLa 28S rRNA **(e)**, 18S rRNA **(f)**, and 5.8S rRNA **(g)**, respectively. Data are presented as means of two independent experiments. **h.** Venn diagram illustrating the overlap of Ψ sites detected in human cy-rRNAs between BACS and SILNAS MS.

To evaluate the performance of BACS, we applied it to cytosolic rRNA (cy-rRNA) from cervical and nasopharyngeal cancer cell lines HeLa and C666-1, respectively (Figure 3.6.a). Cy-rRNA is known to possess a series of highly conserved Ψ sites [238]. As expected, we observed high mutation rates only on Ψ, while other bases have minimum mutation rates (Figure 3.6.b,c). Using a 5% modification level cutoff, we detected 2, 40, and 62 Ψ sites in HeLa 5.8S, 18S, and 28S rRNAs, respectively (Figure 3.6.d). Most of the sites displayed a high level of Ψ modification (>80%), consistent with reports that Ψ sites are highly modified in human cy-rRNA [239] (Figure 3.6.e–g and Figure 3.7.a–d). We also examined the raw signals of BACS, which revealed a strong correlation between two biological replicates (Pearson's r = 1.00, Figure 3.7.e). Compared with the reported SILNAS mass spectrometry (SILNAS MS) data, 103 of 105 known Ψ sites in human cy-rRNA (including one 2'-*O*-methylpseudouridine, Ψm site in 28S rRNA) were identified with high confidence [239] (Figure 3.6.e–g and Figure 3.7.a–d). However, Ψ1136 in 18S rRNA was not detected, possibly due to its low modification level in HeLa cells (4.5% by BACS, Figure 3.7.a,b). Interestingly, we found a 16% U-to-C mutation rate of the known 18S rRNA Ψ36 site in control libraries, although the mutation rate increased to 80% after BACS treatment (Figure 3.6.f and Figure 3.7.f). Similar results were obtained from BID-seq control libraries [228], suggesting the presence of an uncharacterized single nucleotide polymorphism (SNP) site in HeLa cells. It is noteworthy that these two sites could be readily detected in C666-1 cell line (Figure 3.7.g). Therefore, BACS could detect all the known Ψ sites in human cy-rRNAs (Figure 3.6.h). In addition, we detected a new Ψ4938 site in 28S rRNA from HeLa and C666-1 cells, located adjacent to the previously known Ψ4937 site (Figure 3.7.c,d). The presence of Ψ4938 was supported by two public databases, both of which predicted that small nucleolar RNA (snoRNA) SNORA17B

would be responsible for catalyzing this modification [240,241]. BACS therefore provided further confirmation of the existence of the Ψ4938 site. We further identified four novel Ψ sites (Ψ31 / Ψ890 / Ψ899 in 18S rRNA and Ψ1674 in 28S rRNA) in cy-rRNAs from C666-1 cells (Figure 3.7.g). It is important to note that while some Ψ or uridine modifications can induce intrinsic mutation signals (such as U-to-C mutation for $m^1acp^3Ψ1248$ in 18S rRNA and U-to-A mutation for m^3U4500 in 28S rRNA) [242], these can be filtered out by comparing results of BACS libraries with control libraries (Figure 3.7.f). In addition to cy-rRNA, we also applied BACS to mitochondrial rRNA (mt-rRNA) and detected 8 and 1 Ψ sites in 12S and 16S rRNAs, respectively (Figure 3.6.d). Among them, 4 sites have also been detected by Pseudo-seq [104]. In general, the modification level of Ψ sites in mt-rRNA was significantly lower than their cytosolic counterparts (Figure 3.7.h).

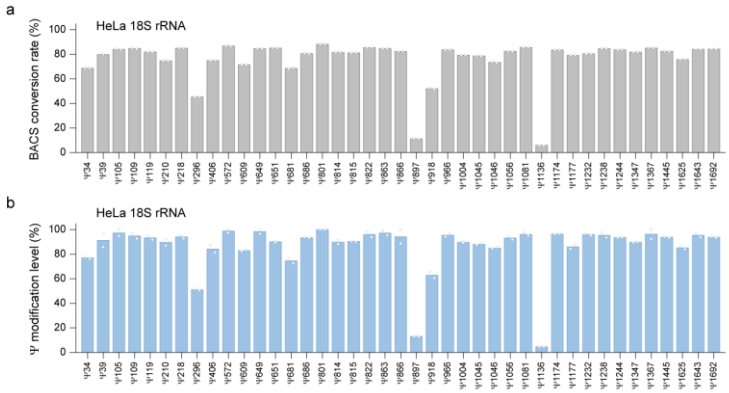


Figure 3.7 | BACS validated known Ψ sites in human rRNA. a. BACS conversion rates of Ψ sites identified in HeLa 18S rRNA. Data are presented as means of two independent experiments. **b.** Modification levels of Ψ sites detected in HeLa 18S rRNA. Data are presented as means of two independent experiments. Ψ modification levels were calculated using the calibration curves.

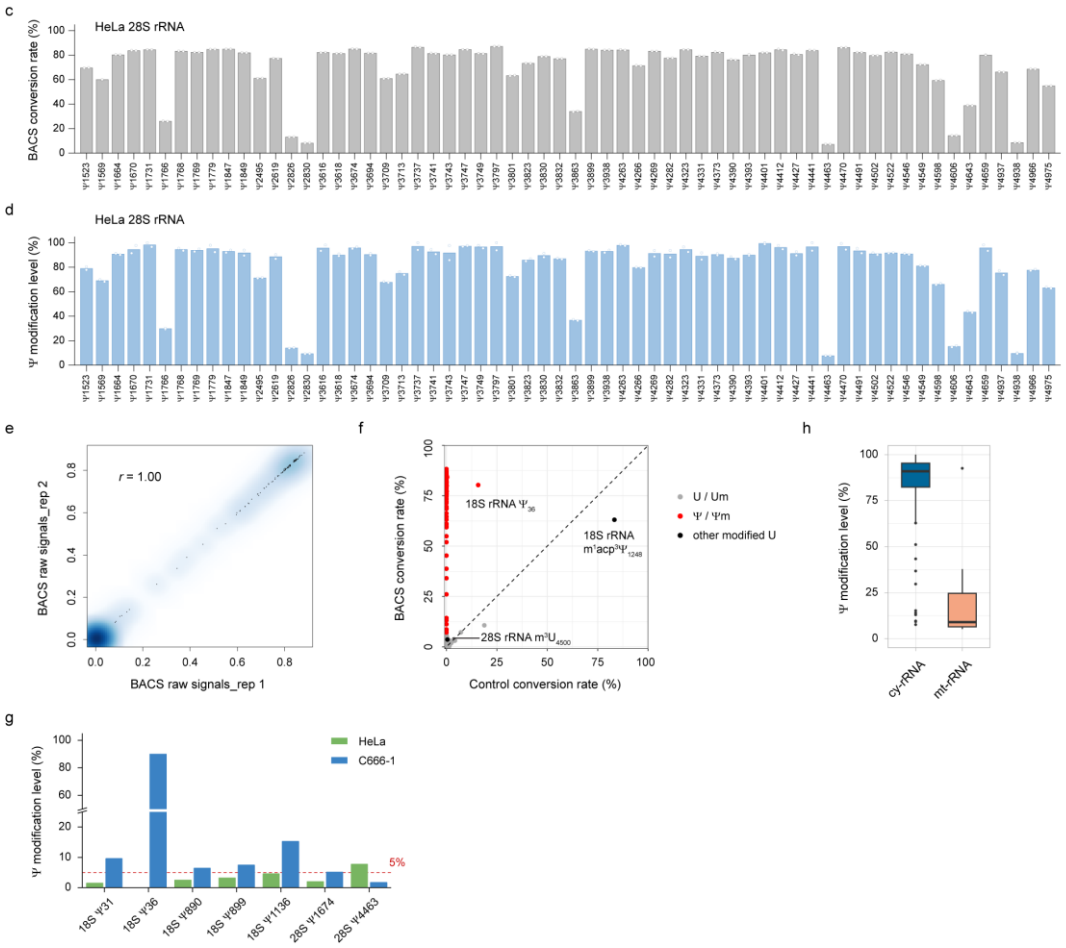


Figure 3.7 | BACS validated known Ψ sites in human rRNA. c. BACS conversion rates of Ψ sites identified in HeLa 28S rRNA. Data are presented as means of two independent experiments. **d.** Modification levels of Ψ sites detected in HeLa 28S rRNA. Data are presented as means of two independent experiments. Ψ modification levels were calculated using the calibration curves. **e.** Correlation density plot between two biological replicates of BACS. The color scale represents density. **f.** Comparison of the conversion rates in HeLa cy-rRNAs between BACS and control samples. **g.** Comparison of the modification levels of selected Ψ sites in cy-rRNAs between HeLa (green) and C666-1 (blue) cell lines. * Ψ36 was found to overlap with a SNP site in HeLa cells. Consequently, its modification level was inferred by comparing the mutation rates between BACS and control libraries. **h.** Comparison of the modification levels of Ψ sites in HeLa cy-rRNAs and mt-rRNAs. For boxplots, boxes represent the 25th to 75th percentiles with a line at the median; whiskers correspond to 1.5 times the interquartile range; dots indicate outliers (cy-rRNA, $n = 104$; mt-rRNA, $n = 9$).

As expected, BACS clearly outperformed BS-based methods in the following aspects [228,229]. Firstly, the U-to-C mutation signature enabled BACS to determine the exact position and number of Ψ sites in consecutive uridine sequences (adjacent to one or more uridines, for instance, Ψ801 / Ψ814 / Ψ815 in 18S rRNA, Ψ1847 / Ψ1849 in 28S rRNA, and Ψ4323 / Ψ4331 in 28S rRNA), which remains challenging for BS-based methods (Figure 3.8.a–c). The improved bioinformatics pipeline of BID-seq with realignment analysis (BID-pipe [230]) could not resolve ambiguity in two consecutive uridine sequences and may introduce artefacts in some cases (for example, Ψ681 in 18S rRNA, Ψ1045 / Ψ1046 in 18S rRNA, and Ψ4549 in 28S rRNA) (Figure 3.8.d–f). These results demonstrate that the deletion signals induced by BS-based methods lack precision in consecutive uridine sequences and these issues are difficult to resolve by subsequent analysis. Secondly, more even conversion rates of Ψ sites were obtained over densely modified regions of rRNA using BACS compared with BS-based methods, indicating that BACS datasets would not be influenced by the density of pseudouridylation (Figure 3.8.g,h). In particular, BACS successfully detected extremely dense Ψ sites in a narrow region (for example, Ψ3737 / Ψ3741 / Ψ3743 / Ψ3747 / Ψ3749 in 28S rRNA and Ψ4263 / Ψ4266 / Ψ4269 in 28S rRNA). Thirdly, when compared to SILNAS MS, BACS provided greatly improved accuracy in quantifying Ψ modification levels than BS-based methods (r = 0.90 for BACS vs r = 0.37 for BID-seq and r = 0.48 for PRAISE) (Figure 3.8.i–k). These findings strongly indicate that quantifying Ψ using deletion signals, as done by BS-based methods, introduces inaccuracies into the analysis. Finally, we found that BACS could achieve a higher conversion rate on 28S rRNA Ψm3797 site (85%) than BS-based methods (10–20%), because BACS solely relied on the availability of N^1 position of Ψ (Figure 3.8.h).

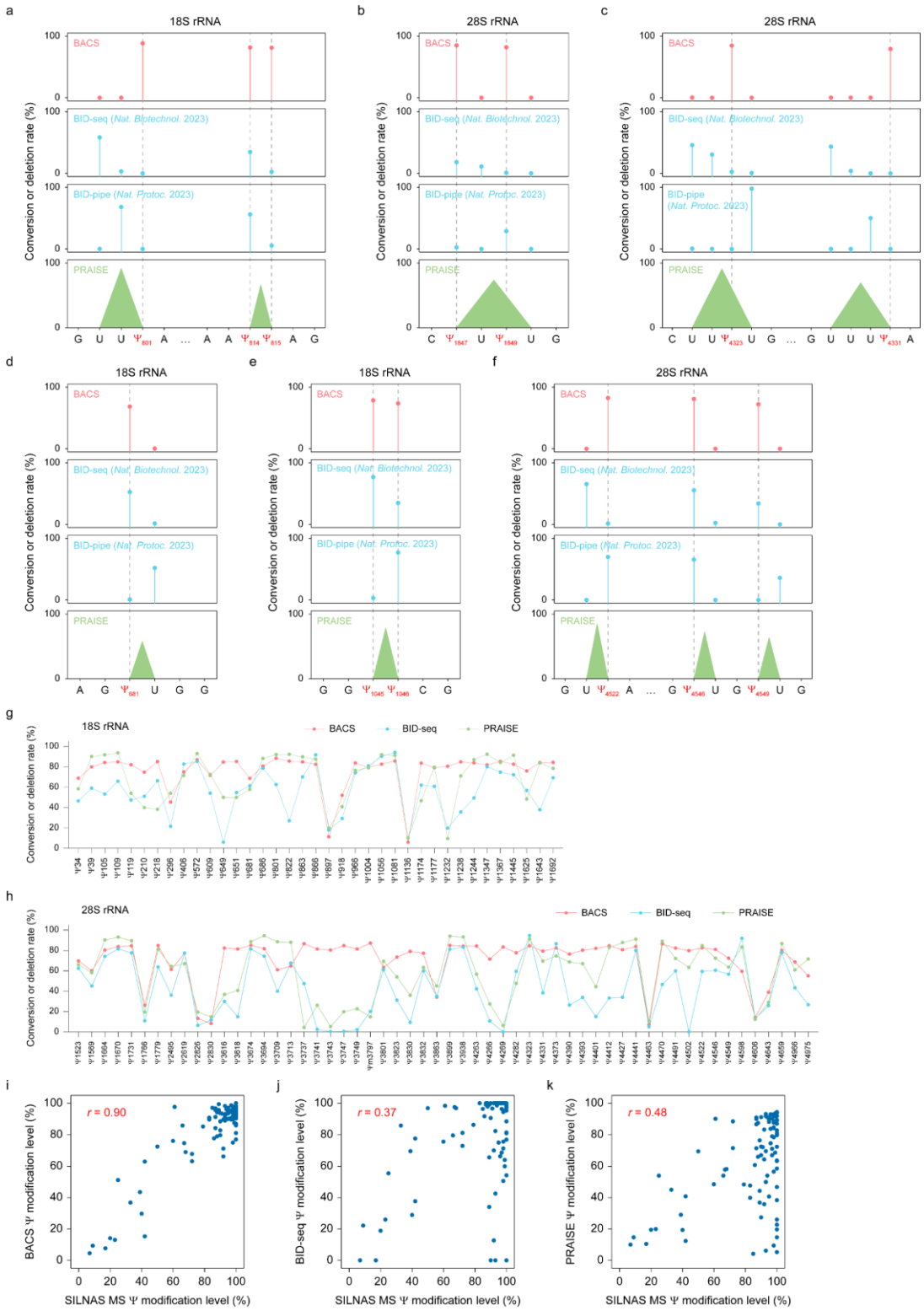


Figure 3.8 | Comparison of BACS and BS-based methods for Ψ detection in human cy-rRNAs. a–f. Examples of BACS, BID-seq, BID-pipe, and PRAISE results in selected consecutive

uridine regions of 18S and 28S rRNA. Due to their deletion signals, BS-based methods cannot determine the exact position or identify the number of Ψ within consecutive uridine contexts. By default, the aligner will position the deletion at the 5'-most uridine, as shown in the BID-seq panel. After realignment, this issue cannot be fully resolved, as shown in the BID-pipe panel. PRAISE considers consecutive uridines as a whole for Ψ calling, resulting in broad peak signals. **g,h.** Comparison of the conversion rates of BACS (pink) with the deletion rates of BID-seq (blue) and PRAISE (green) for selected Ψ sites in 18S rRNA **(g)** and 28S rRNA **(h)**. Because BID-seq and PRAISE cannot quantify multiple Ψ sites (≥ 2) located in the same consecutive uridine context, these sites were excluded. **i–k.** Correlation of cy-rRNA Ψ modification levels reported by SILNAS MS with those reported by BACS **(i)**, BID-seq **(j)**, and PRAISE **(k)**.

3.4 BACS identified highly conserved Ψ sites in human spliceosomal snRNAs

After validating BACS on rRNA, we applied it to spliceosomal snRNAs from HeLa, C666-1, Raji, and Elijah cells, which contain multiple consecutive Ψ sites [243]. We first focused on major spliceosomal snRNA species and detected 2, 14, 3, 4, and 4 Ψ sites in U1, U2, U4, U5, and U6 snRNAs from HeLa cells, respectively, which is highly consistent with the latest SILNAS MS results [244] (Figure 3.9.a). Only Ψ59 in U4 snRNA was not detected by BACS, since it is likely to be lowly modified in HeLa cells. However, this position was modified to higher levels in U4 snRNA from C666-1 and Elijah cells (Figure 3.9.d). Consequently, we can detect all known Ψ sites in human major spliceosomal snRNAs (Figure 3.9.b). It is noteworthy that BACS successfully mapped all 14 Ψ sites in human U2 snRNA, which has not been realized by any other high-throughput sequencing methods, further highlighting the superiority of BACS in detecting dense and consecutive Ψ sites (Figure 3.9.c).

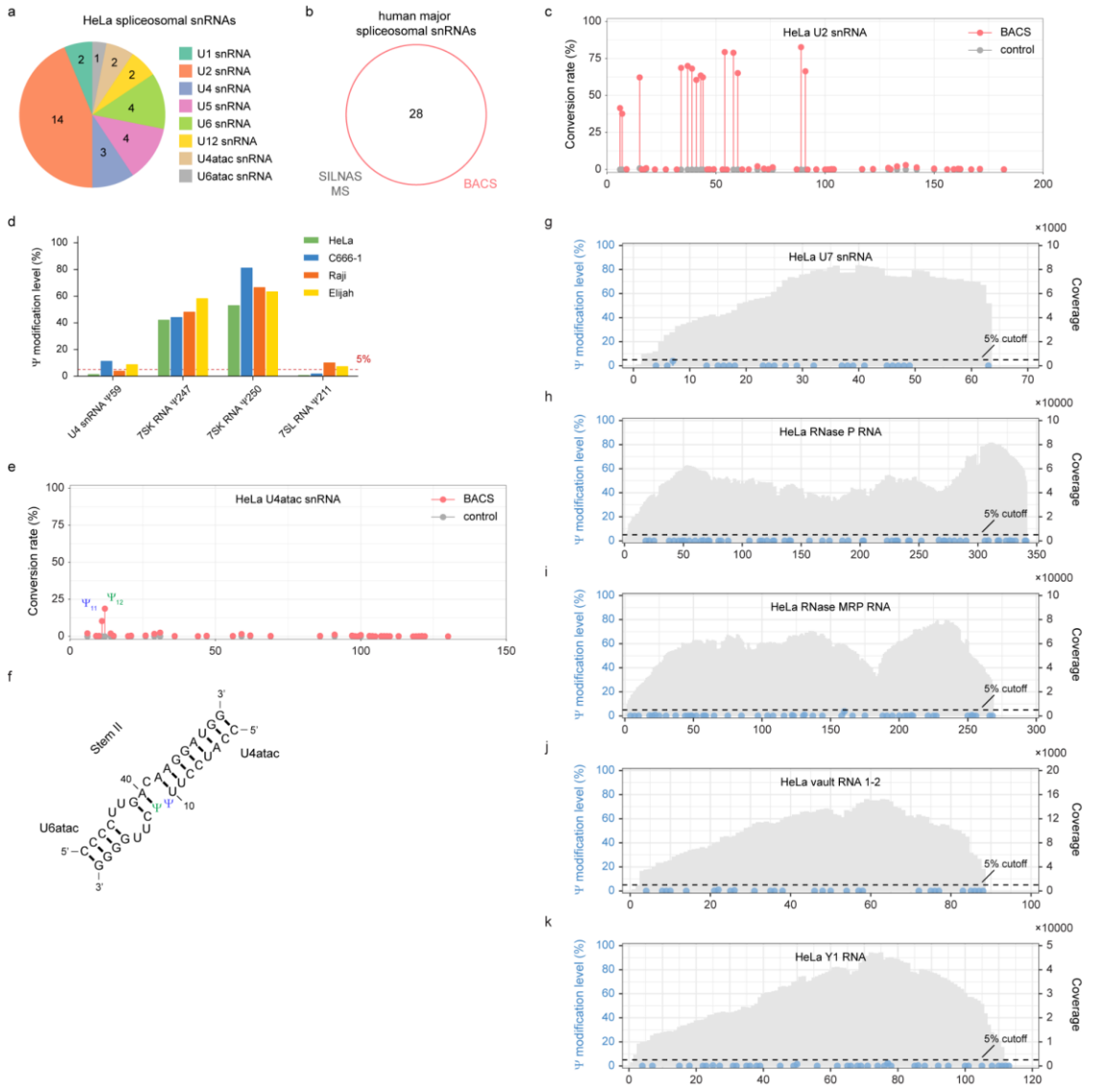


Figure 3.9 | BACS identified conserved and novel Ψ sites in human spliceosomal snRNA. a. Numbers of Ψ sites identified in HeLa spliceosomal snRNAs. **b.** Venn diagram illustrating the overlap of Ψ sites detected in human spliceosomal snRNAs between BACS and SILNAS MS. **c.** Conversion rates of BACS (pink) and control (grey) samples in HeLa U2 snRNA. Data are presented as means of two independent experiments. **d.** Comparison of the modification levels of selected Ψ sites in U4 snRNA, 7SK RNA, and 7SL RNA across HeLa (green), C666-1 (blue), Raji (orange), and Elijah (yellow) cell lines. **e.** Conversion rates of BACS (pink) and control (grey) samples in U4atac snRNA, showing the novel (blue) and known (green) Ψ site. Data are presented as means of two independent experiments. **f.** Base pairing interactions between U4atac and U6atac snRNAs in stem II region. Blue and green color denote the novel Ψ11 and known Ψ12 site,

respectively. **g–k.** Ψ modification levels in U7 snRNA **(g)**, RNase P RNA **(h)**, RNase MRP RNA **(i)**, vault RNA **(j)**, and Y RNA **(k)**. The sequencing coverage of each RNA was plotted in grey. Only one example was shown for vault RNA and Y RNA, respectively.

Unlike the snRNA components of human major spliceosome, the Ψ profile of minor spliceosomal snRNA species has only been revealed using CMC-based primer extension assay, mainly due to their low abundance [245,246]. However, given that CMC-based methods may suffer from partial labeling efficiency and 'stuttering' phenomenon [216], we believed that BACS could be a preferred method to study pseudouridylation in these snRNA species. Indeed, we consistently detected 2, 2, and 1 Ψ sites in U12, U4atac, and U6atac snRNAs from HeLa cells, respectively, while no Ψ site was detected in U11 snRNA (Figure 3.9.a). Notably, we confirmed that there are two consecutive Ψ sites (Ψ11 / Ψ12) rather than one Ψ12 site in U4atac snRNA, providing new insights into its interaction with U6atac snRNA [245,247] (Figure 3.9.e,f). Additionally, we detected highly conserved Ψ247 and Ψ250 sites in 7SK RNA [248] and differentially modified Ψ211 site in 7SL RNA [244] (Figure 3.9.d). Our analysis also showed that there were no high-confidence Ψ sites in U7 snRNA, RNase P RNA, RNase MRP RNA, vault RNA, and Y RNA (Figure 3.9.g–k).

3.5 BACS revealed the Ψ profile of human snoRNA

The Ψ profile of yeast snoRNA has been revealed through Pseudo-seq [104] and Ψ-seq [105], yet it remains relatively unexplored in human snoRNA. Ψ-seq [105] and BID-seq [228] only detected 11 and 39 Ψ sites in human snoRNA, respectively. In contrast, using BACS, we detected 304 Ψ sites in snoRNA from HeLa cells (Figure 3.10.a).

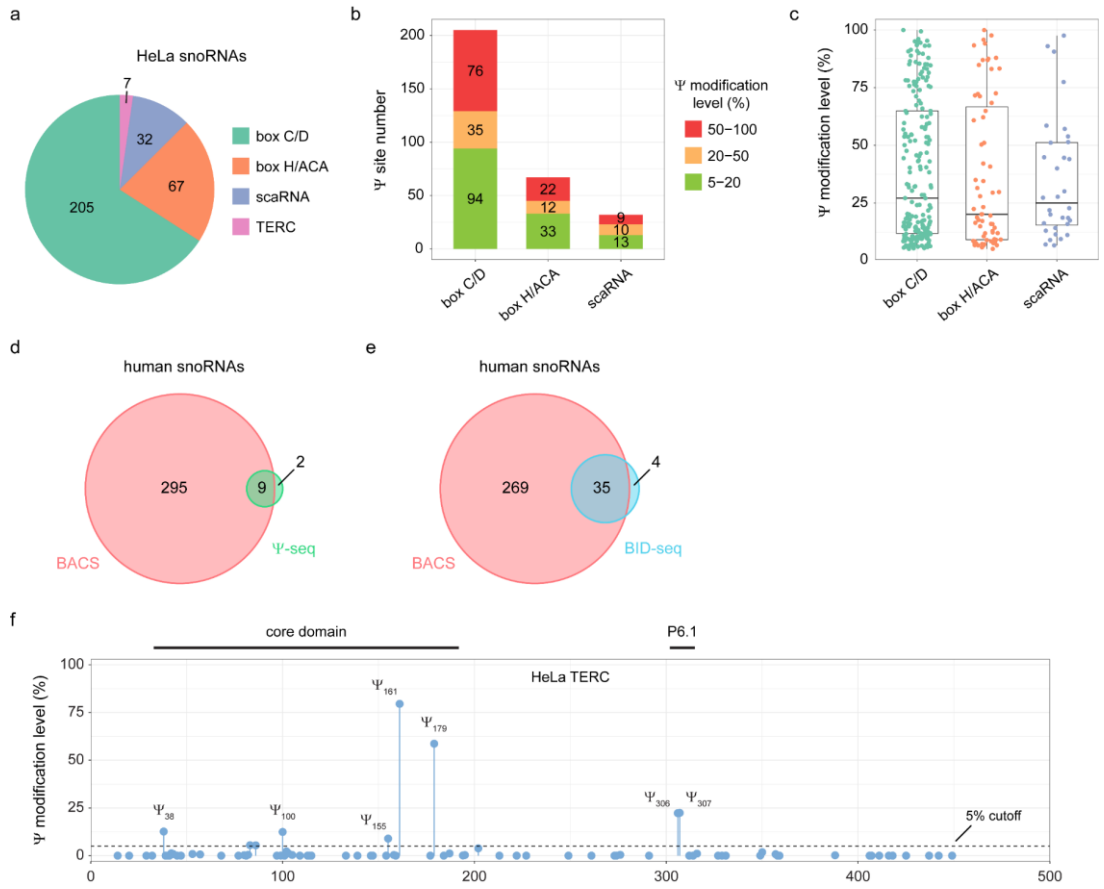


Figure 3.10 | BACS detected abundant Ψ sites in human snoRNAs. a. Numbers of Ψ sites identified in HeLa snoRNAs and TERC. **b.** Numbers of Ψ sites with high (50–100%, red), medium (20–50%, yellow), and low (5–20%, green) modification levels identified in HeLa box C/D snoRNAs, box H/ACA snoRNAs, and scaRNAs. **c.** Modification level distributions of Ψ sites in HeLa box C/D snoRNAs, box H/ACA snoRNAs, and scaRNAs. Boxplots visualize all Ψ sites in each class of snoRNAs; boxes represent the 25th to 75th percentiles with a line at the median; whiskers correspond to 1.5 times the interquartile range (box C/D, $n = 205$; box H/ACA, $n = 67$; scaRNA, $n = 32$). **d.** Venn diagram illustrating the overlap of Ψ sites detected in human snoRNAs between BACS and Ψ-seq. **e.** Venn diagram illustrating the overlap of Ψ sites detected in human snoRNAs between BACS and BID-seq. **f.** Ψ modification levels in HeLa TERC, with each identified Ψ site labeled accordingly. Data are presented as means of two independent experiments.

Further analysis revealed the presence of 205, 67, and 32 Ψ sites in box C/D snoRNAs, box H/ACA snoRNAs, and small Cajal body-specific RNAs (scaRNAs), respectively. Remarkably, all three types of snoRNAs exhibited a substantial number of highly modified Ψ sites (Figure 3.10.b,c). SnoRNA Ψ sites detected through BACS largely covered those previously identified by Ψ-seq [105] and BID-seq [228], demonstrating increased sensitivity of BACS (Figure 3.10.d,e). Furthermore, we observed that Ψ sites in box C/D snoRNAs displayed enrichment in the 5'-upstream regions of box D' and the 3'-downstream regions of box C', while Ψ sites in box H/ACA snoRNAs were enriched in the 5'-upstream regions of box H and ACA (Figure 3.11.a,b). These patterns implied a potential role for Ψ in mediating interactions between snoRNAs and their targets. Indeed, a subset of Ψ sites identified in box C/D and box H/ACA snoRNAs were also located in the predicted guide regions, which was in accordance with the Ψ-seq results [105] (Figure 3.11.c,d).

In addition, human telomerase RNA component (TERC) shares similar characteristics with snoRNAs, as it contains a conserved box H/ACA scaRNA domain at the 3'-end [249]. Upon BACS treatment, we detected 7 Ψ sites in TERC from HeLa cells, 4 of which were putative Ψ sites previously discovered through CMC-based primer extension approach [250] (Figure 3.10.a,f). In particular, all 3 novel Ψ sites (Ψ38 / Ψ100 / Ψ155), together with the known Ψ161 and Ψ179 sites, were found in the core domain of TERC. This observation suggested the potential involvement of Ψ in stabilizing the TERC structure, similar to the scenario that has been demonstrated for Ψ306 and Ψ307 within the P6.1 loop of TERC [251].

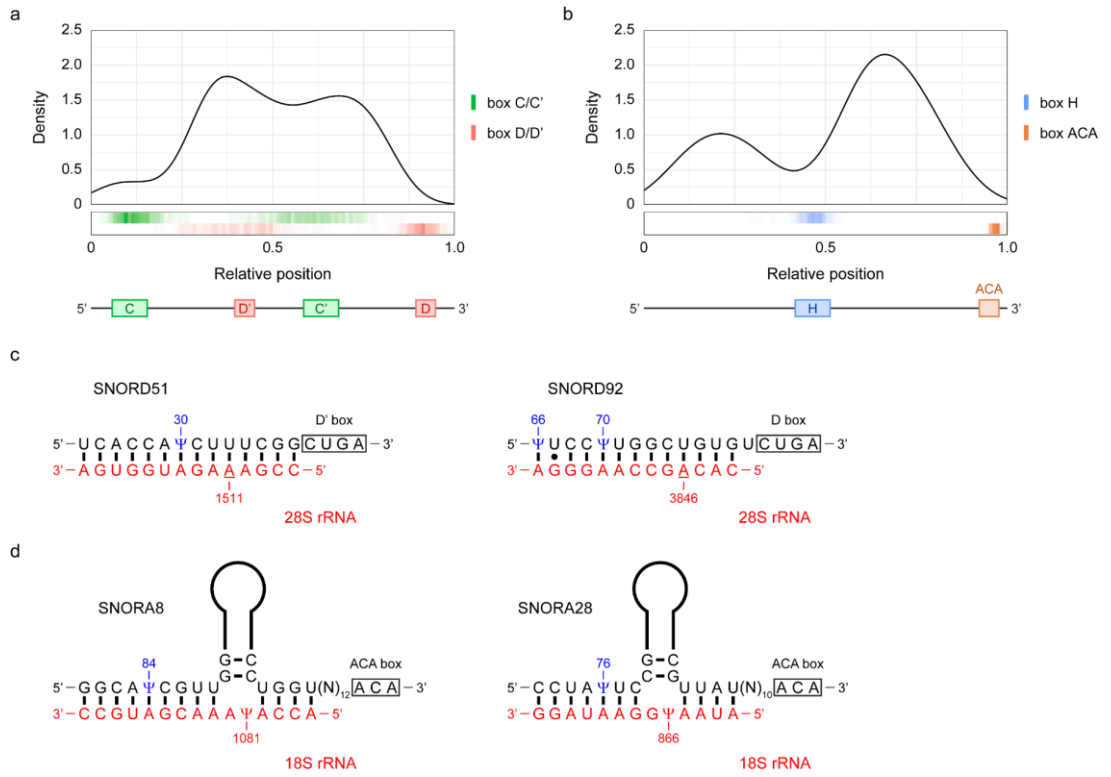


Figure 3.11 | Potential involvement of Ψ in regulating the guiding activity of human snoRNAs. a. Metagene profile of Ψ sites in box C/D snoRNAs. Red and green color denote the box C/C' and D/D', respectively. **b.** Metagene profile of Ψ sites in box H/ACA snoRNAs. Orange and blue color denote the box H and ACA, respectively. **c,d.** Potential base pairing interactions between snoRNAs (black) and their targets in rRNA (red): **c.** box C/D snoRNAs and **d.** box H/ACA snoRNAs. Identified snoRNA Ψ sites are highlighted in blue. 2'-*O*-methylation targets in rRNA are underlined. Structures are adapted from snoRNA Atlas [241].

3.6 A comprehensive Ψ map of human tRNA

Ψ is one of the most fundamental and prevalent modifications in human tRNA [65,252]. However, given that most of tRNA species are extensively modified and highly structured, quantitative profiling of Ψ in tRNA remains challenging by CMC- or BS-based methods [229,253]. We anticipated that BACS would offer a better solution to this problem, since mutation signals induced by BACS would not be influenced by RT blocks or other

intrinsic mismatches. Indeed, while CMC- and BS-based methods failed to map human cytosolic tRNAs (cy-tRNAs), we used BACS to generate the first quantitative Ψ map of cy-tRNAs from HeLa cells with 609 high-confidence Ψ sites (Figure 3.13.a). The number of Ψ sites identified per cy-tRNA varied among different isotypes (Figure 3.12.a).

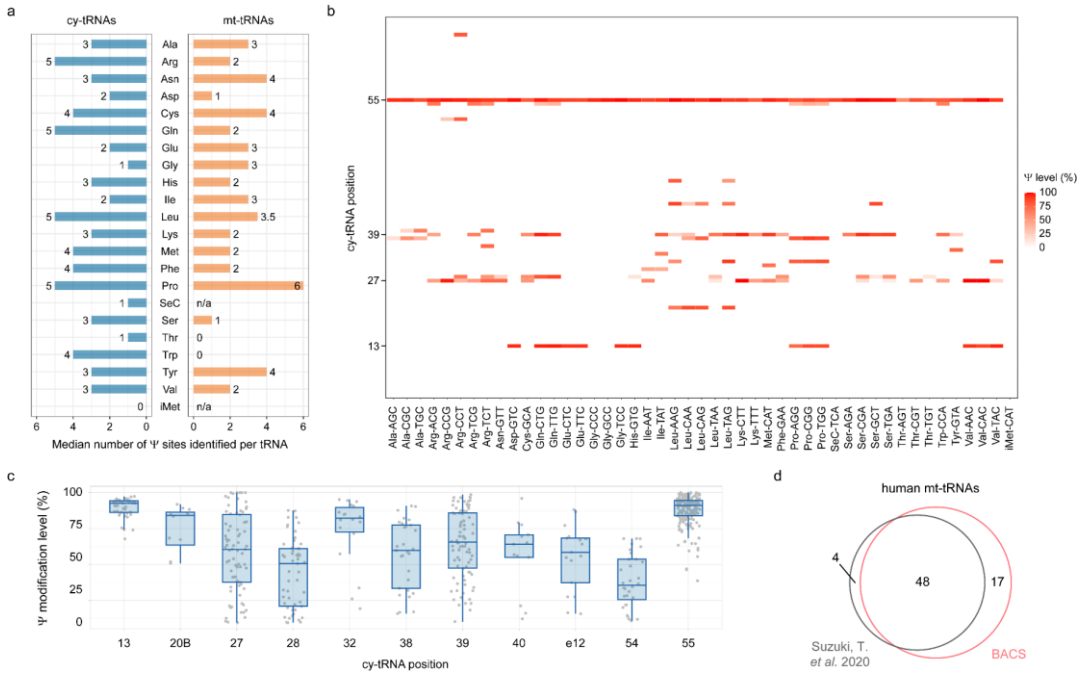


Figure 3.12 | BACS unveiled the comprehensive Ψ profile of human tRNA. a. Median numbers of Ψ sites identified per tRNA in each cy-tRNA (left) and mt-tRNA (right) isotype from HeLa cells. n/a, not applicable. **b.** Heatmap showing the modification levels of high-confidence Ψ sites in HeLa cy-tRNAs. Only one representative tRNA isodecoder was presented for each isoacceptor family. **c.** Comparison of the modification levels of Ψ sites at selected positions of HeLa cy-tRNAs. Boxplots visualize all Ψ sites at each position; boxes represent the 25th to 75th percentiles with a line at the median; whiskers correspond to 1.5 times the interquartile range (tRNA position: 13, n = 43; 20B, n = 12; 27, n = 81; 28, n = 53; 32, n = 18; 38, n = 29; 39, n = 86; 40, n = 13; e12, n = 17; 54, n = 32; 55, n = 185). **d.** Venn diagram illustrating the overlap of Ψ sites in human mt-tRNAs reported by BACS and a previously published dataset [254].

In cy-tRNAs, Ψ sites were predominantly located at highly conserved positions, including position 13, 27–28, 38–40, and 55, while Ψ at other positions were limited to specific types of cy-tRNAs (Figure 3.12.b). An integrated view of the Ψ profile of human cy-tRNAs was then summarized based on the canonical tRNA numbering system [255] (Figure 3.13.b). Subsequently, we compared the Ψ modification level at each tRNA position (Figure 3.12.c). Notably, position 55 emerged as the most frequently and highly modified Ψ site in cy-tRNAs. Moreover, position 13 also displayed a high level of Ψ modification. In contrast, the modification levels of position 27–28 and position 38–40 showed considerable variations.

Using BACS, we also detected 54 Ψ sites in HeLa mitochondrial tRNAs (mt-tRNAs) (Figure 3.12.a and 3.13.a). Applying BACS to other human cell lines, we observed significant differential Ψ modification on mt-tRNAs. For example, Ψ55 in mt-tRNA^{Met} was not characterized as high-confidence sites due to its low modification level in HeLa cells (3.4%), while it was readily detected in C666-1, Raji, and Elijah cell lines (Figure 3.13.c,d). In addition, C666-1 cell lines displayed an elevated Ψ level at position 66–68 compared to HeLa, Raji, and Elijah cells (Figure 3.13.d). We finally mapped a total of 65 Ψ sites in mt-tRNAs by merging the results of HeLa, C666-1, Raji, and Elijah cells, which was highly consistent with the published dataset [254] (Figure 3.12.d). Most of the BACS-only Ψ sites were located near the 3'-end of mt-tRNA, which could not be identified using CMC-based methods. 4 Ψ sites in ref. [254] were not detected by BACS, including mt-tRNA^{Gln} Ψ33 / Ψ40, mt-tRNA^{His} Ψ35, and mt-tRNA^{Pro} Ψ67. Overall, human mt-tRNAs were pseudouridylated to a lesser extent compared with cy-tRNAs (Figure 3.12.b,c and 3.13.c,d).

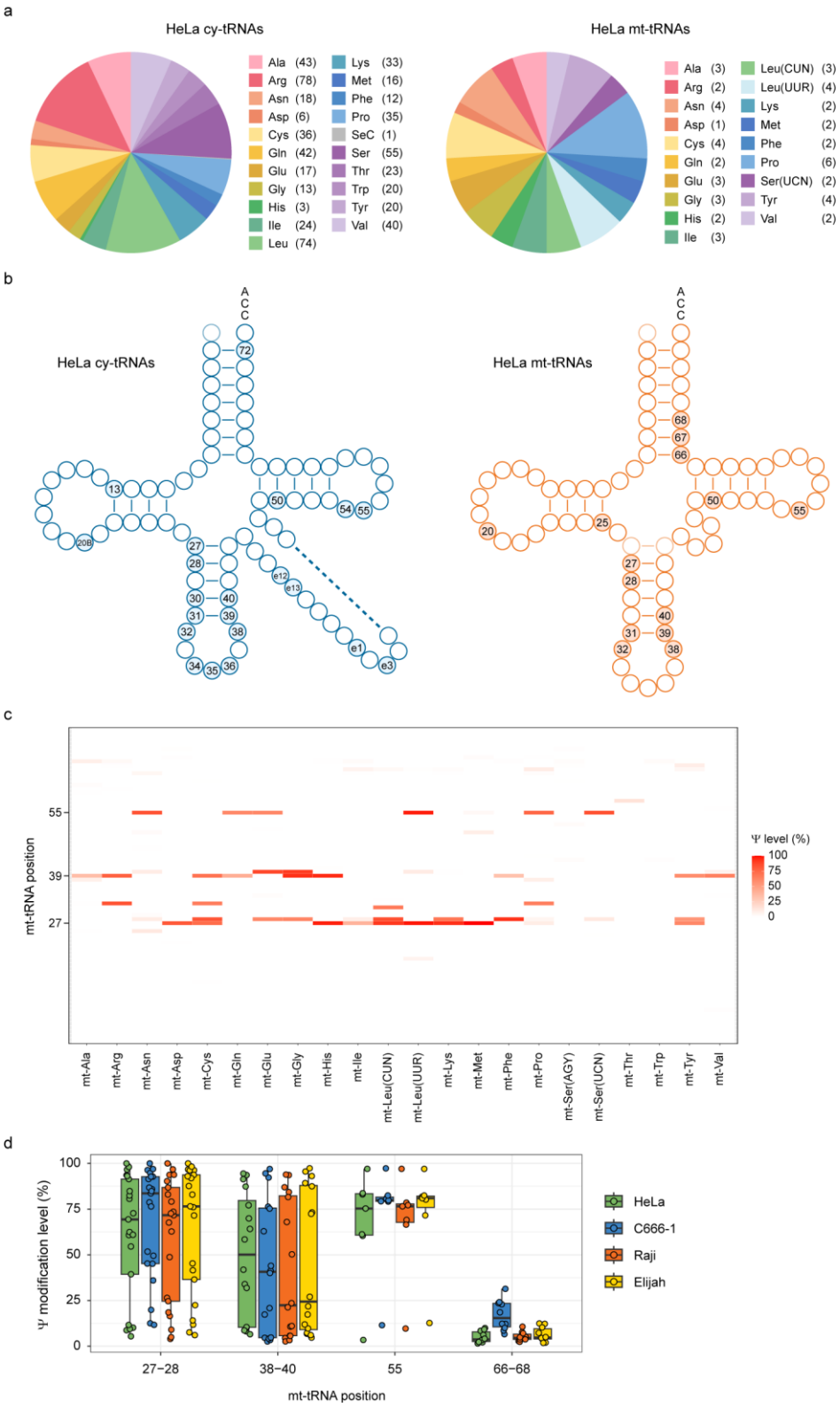

a
HeLa cy-tRNAs
Ala (43)
Arg (78)
Asn (18)
Asp (6)
Cys (36)
Gln (42)
Glu (17)
Gly (13)
His (3)
Ile (24)
Leu (74)
Lys (33)
Met (16)
Phe (12)
Pro (35)
SeC (1)
Ser (55)
Thr (23)
Trp (20)
Tyr (20)
Val (40)
HeLa mt-tRNAs
Ala (3)
Arg (2)
Asn (4)
Asp (1)
Cys (4)
Gln (2)
Glu (3)
Gly (3)
His (2)
Ile (3)
Leu(CUN) (3)
Leu(UUR) (4)
Lys (2)
Met (2)
Phe (2)
Pro (6)
Ser(UCN) (2)
Tyr (4)
Val (2)
b
HeLa cy-tRNAs
HeLa mt-tRNAs
c
mt-tRNA position
Ψ level (%)
mt-Ala
mt-Arg
mt-Asn
mt-Asp
mt-Cys
mt-Gln
mt-Glu
mt-Gly
mt-His
mt-Ile
mt-Leu(CUN)
mt-Leu(UUR)
mt-Lys
mt-Met
mt-Phe
mt-Pro
mt-Ser(AGY)
mt-Ser(UCN)
mt-Thr
mt-Trp
mt-Tyr
mt-Val
d
Ψ modification level (%)
mt-tRNA position
27–28
38–40
55
66–68
HeLa
C666-1
Raji
Elijah

Figure 3.13 | BACS revealed a comprehensive map of Ψ in human tRNA. a. Distributions of Ψ sites identified in each cy-tRNA (left) and mt-tRNA (right) isotype from HeLa cells. **b.** Integrated view of the Ψ profiles of HeLa cy-tRNAs (left) and mt-tRNAs (right). **c.** Heatmap of the Ψ modification levels in HeLa mt-tRNAs. **d.** Comparison of the modification levels of Ψ sites at selected positions of mt-tRNAs across HeLa (green), C666-1 (blue), Raji (orange), and Elijah (yellow) cell lines. Boxplots visualize all Ψ sites at each position; boxes represent the 25th to 75th percentiles with a line at the median; whiskers correspond to 1.5 times the interquartile range (tRNA position: 27–28, n = 21; 38–40, n = 16; 55, n = 7; 66–68, n = 10).

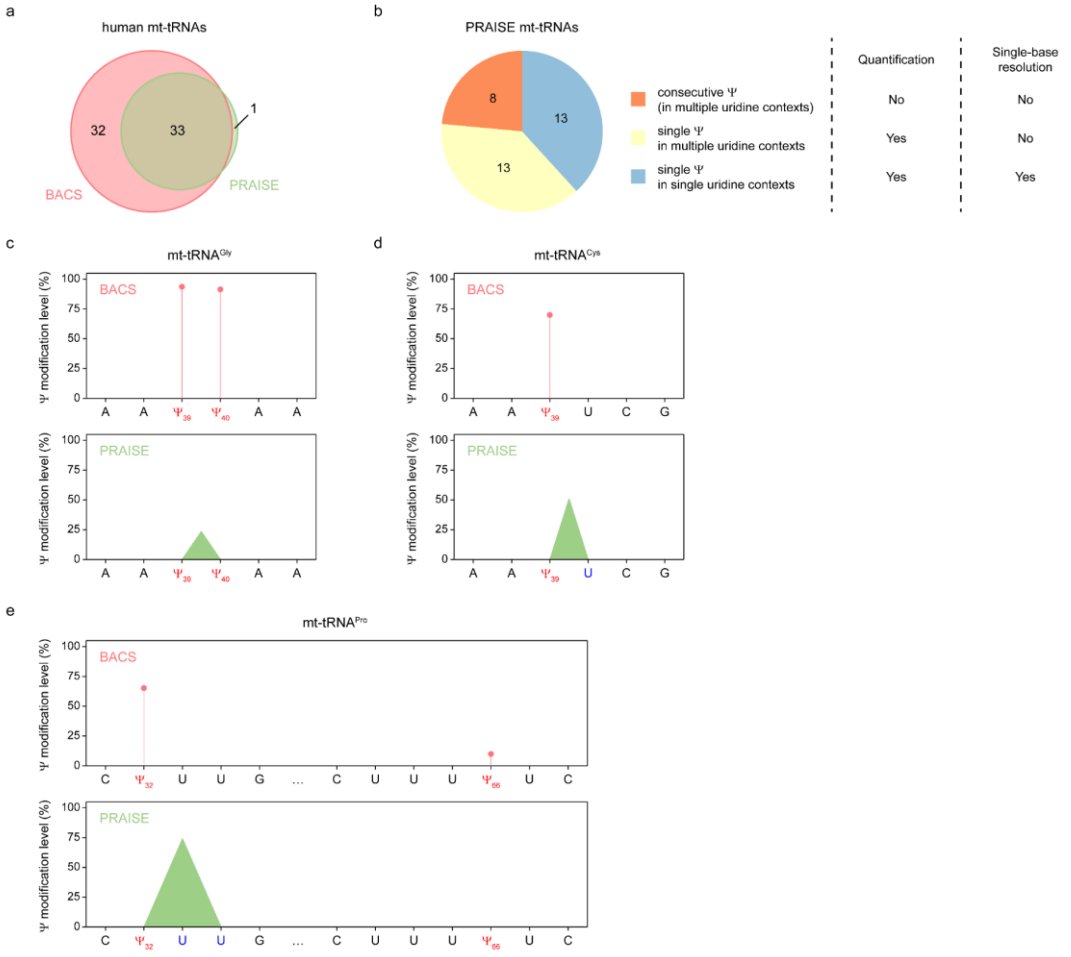


Figure 3.14 | Comparison of BACS and PRAISE for Ψ detection in human mt-tRNAs. a. Venn diagram illustrating the overlap of Ψ sites detected in human mt-tRNAs between BACS and PRAISE. **b.** Distribution of mt-tRNA Ψ sites identified by PRAISE. Only single Ψ site in single uridine contexts can be quantitatively identified by PRAISE at single-base resolution. **c–e.**

Examples of BACS and PRAISE results in selected consecutive uridine regions of human mt-tRNAs. Due to their deletion signals, BS-based methods cannot determine the exact position or identify the number of Ψ within consecutive uridine contexts. PRAISE considers consecutive uridines as a whole for Ψ calling, resulting in broad peak signals.

In contrast, PRAISE detected only 34 Ψ sites in mt-tRNAs from HEK293T cells [229] (Figure 3.14.a). PRAISE encountered challenges in quantifying consecutive Ψ sites (8 out of 34, 23.5%) and determining the precise position for single Ψ site within multiple uridine contexts (13 out of 34, 38.2%) (Figure 3.14.b). The signal produced by PRAISE for these Ψ sites manifested as an ambiguous peak, lacking accurate quantification and single-base resolution (Figure 3.14.c–e). As a result, PRAISE achieved quantitative and single-base resolution detection for only 13 Ψ sites in mt-tRNAs, revealing significant limitations compared to BACS.

3.7 Profiling and quantification of Ψ in HeLa mRNA

After successfully applying BACS to various types of ncRNAs, we extended its usage to map and quantify Ψ modifications in HeLa mRNA. Given that the low stoichiometry of Ψ modification in polyA-tailed RNA, we applied *in vitro* transcribed polyA-tailed RNA (IVT RNA) from HeLa cells as a modification-free control to help with Ψ calling [256] (Figure 3.16.a). We detected a total of 1335 Ψ sites in HeLa polyA-tailed RNA (Figure 3.15.a). The majority of these Ψ sites exhibited low modification levels (<20%), while only a limited number of Ψ sites displayed high levels of modification (>50%) (Figure 3.15.a,b). In contrast to the aforementioned ncRNAs, the Ψ modification level in polyA-tailed RNA was significantly lower (Figure 3.16.b). Among the 1335 Ψ sites, 1294 and

41 were located in mRNA and ncRNA (excluding rRNA, snRNA, snoRNA, and tRNA), respectively (Figure 3.15.c).

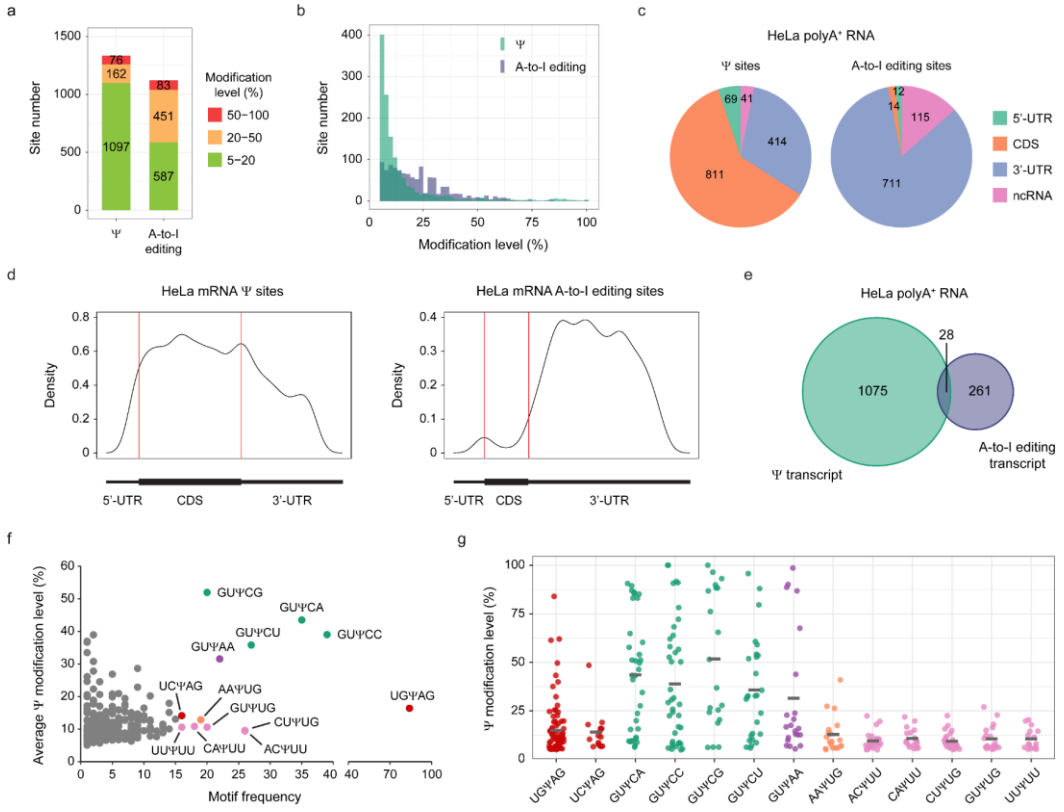


Figure 3.15 | Simultaneous characterization of Ψ and A-to-I editing sites in HeLa mRNA. a. Numbers of Ψ and A-to-I editing sites with high (50–100%, red), medium (20–50%, yellow), and low (5–20%, green) modification levels identified in HeLa polyA-tailed RNA. **b.** Modification level distribution of Ψ and A-to-I editing sites in HeLa polyA-tailed RNA. **c.** Distribution of Ψ and A-to-I editing sites within different features of HeLa mRNA and ncRNA. **d.** Metagene profile of Ψ and A-to-I editing sites in HeLa mRNA. **e.** Venn diagram illustrating the overlap of transcripts possessing Ψ and A-to-I editing sites. **f.** Motif frequency of Ψ sites in HeLa mRNA. **g.** Modification level distributions of HeLa mRNA Ψ sites within selected motifs, with means indicated in each plot by a horizontal line. Motif: UGΨAG, *n* = 84; UCΨAG, *n* = 16; GUΨCA, *n* = 35; GUΨCC, *n* = 39; GUΨCG, *n* = 20; GUΨCU, *n* = 27; GUΨAA, *n* = 22; AAΨUG, *n* = 19; ACΨUU, *n* = 26; CAΨUU, *n* = 18; CUΨUG, *n* = 26; GUΨUG, *n* = 20; UUΨUU, *n* = 16.

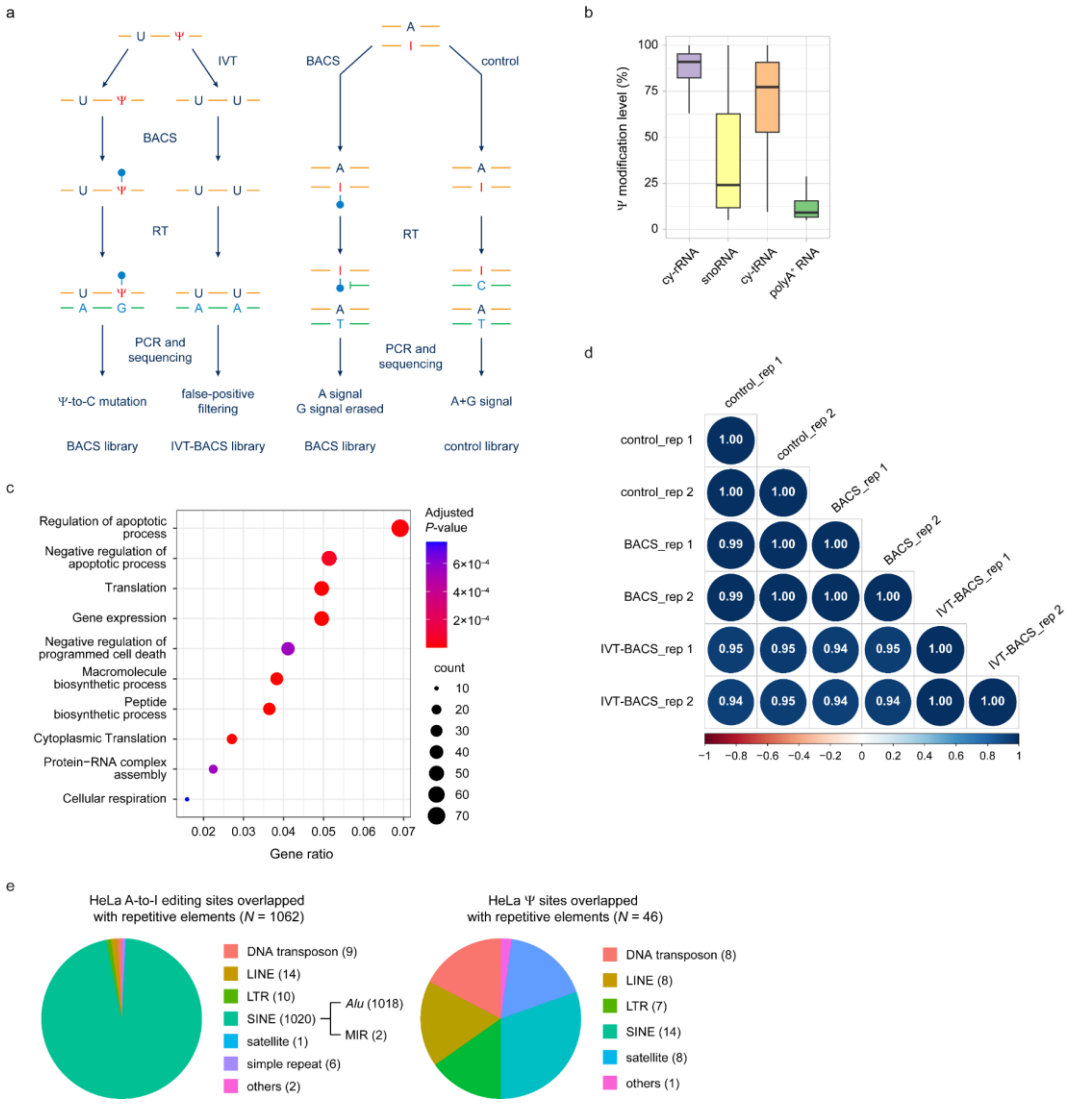


Figure 3.16 | Analysis of BACS libraries for polyA-tailed RNA. a. Schematic overview of Ψ and A-to-I editing site identification in polyA-tailed RNA. Ψ site calling is based on BACS and IVT-BACS libraries. A-to-I editing site calling is based on BACS and control libraries. **b.** Comparison of the Ψ modification levels in different RNA species. For boxplots, boxes represent the 25th to 75th percentiles with a line at the median; whiskers correspond to 1.5 times the interquartile range (cy-rRNA, n = 104; snoRNA, n = 304; cy-tRNA, n = 609; polyA-tailed RNA, n = 1335). **c.** Gene ontology enrichment analysis (biological process) for HeLa mRNA Ψ sites. **d.** Correlation of HeLa RNA expression levels between BACS, control, and IVT-BACS libraries. Pearson's r values are shown. **e.** Distribution of HeLa A-to-I editing and Ψ sites overlapped with repetitive elements.

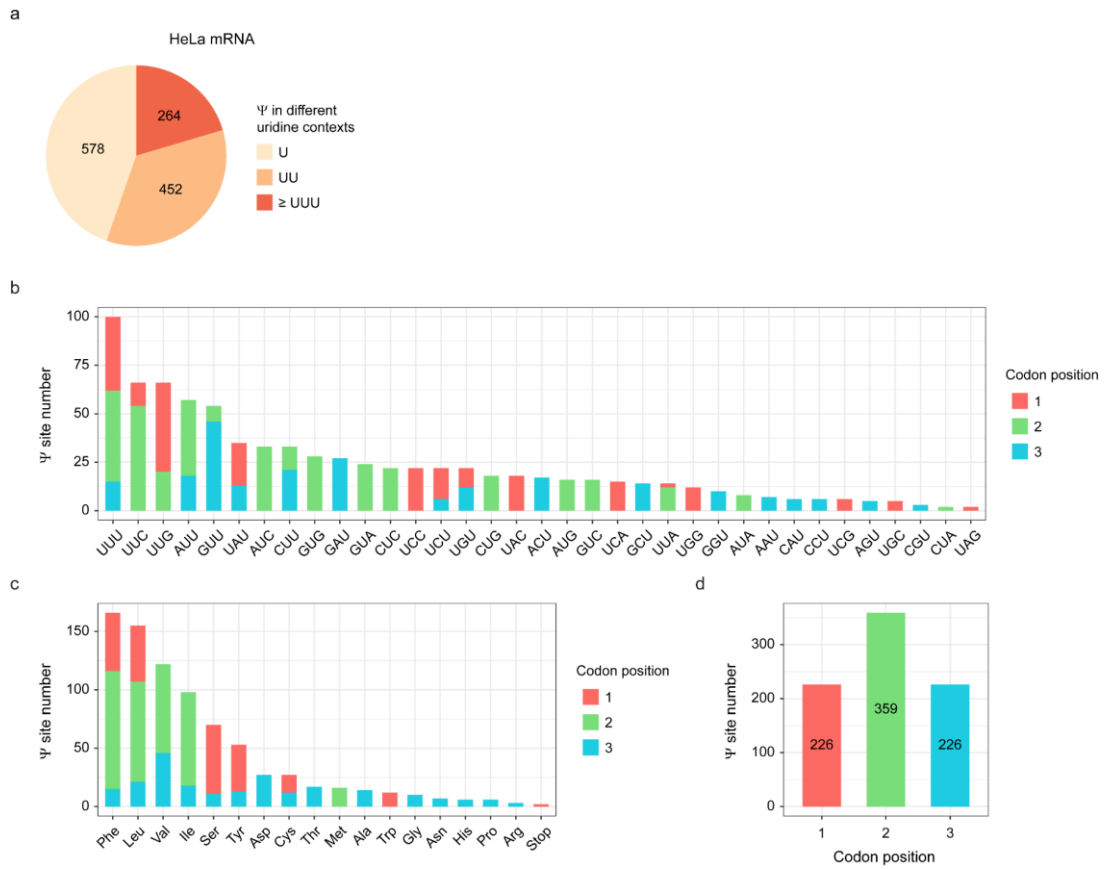


Figure 3.17 | Sequence context and codon preference of Ψ in HeLa mRNA. a. Distribution of mRNA Ψ sites within single and consecutive uridine contexts. **b,c.** Numbers of mRNA Ψ sites located in different codons **(b)** and codons encoding different amino acids **(c)**. Red, green, and blue color denote the first, second, and third codon position, respectively. **d.** Numbers of mRNA Ψ sites located in different codon positions.

Within mRNA, Ψ was enriched in the coding sequence (CDS) and 3'-untranslated region (3'-UTR), while it was relatively depleted in the 5'-untranslated region (5'-UTR), consistent with previous findings [221,228,229] (Figure 3.15.c,d). The 1335 Ψ sites were located across 1103 polyA-tailed RNA transcripts, with the majority carrying only one Ψ site (Figure 3.15.e). Gene ontology (GO) analysis revealed that Ψ-modified mRNA was enriched in functions such as translation and regulation of apoptotic process (Figure

3.16.c). Importantly, BACS could simultaneously provide mRNA expression levels while mapping Ψ, which showed strong correlation with control libraries (Pearson's r = 0.99–1.00), suggesting minimal RNA degradation induced by BACS (Figure 3.16.d).

Next, we analyzed the sequence contexts of Ψ in HeLa mRNA. First, our analysis indicated that the majority of Ψ sites (55.3%) were located in consecutive uridine sequences (Figure 3.17.a). These positions could not be precisely determined through BS-based methods [228,229], further highlighting the advantage of BACS. Benefiting from the high-resolution signals of BACS, we found that Ψ was predominantly enriched in USΨAG (S = C or G) and GUΨCN (N = A, C, G or U) motifs, corresponding to the previously identified PUS7 and TRUB1 motif, respectively [257] (Figure 3.15.f). In addition, we also observed that Ψ tends to be enriched in those motifs containing multiple consecutive uridines, such as CUΨUG, ACΨUU, and even UUΨUU. The stoichiometry of Ψ within these motifs was also compared, demonstrating that GUΨCN exhibited a relatively high modification level (Figure 3.15.g). Furthermore, we analyzed the codon preference of Ψ in mRNA. As expected, Ψ was enriched in those codons containing consecutive uridines, such as UUY (Y = C or U), UUG, AUU, and GUU, which encoded phenylalanine (Phe), leucine (Leu), isoleucine (Ile), and valine (Val), respectively (Figure 3.17.b,c). Within codons, Ψ was mainly located in the second position (Figure 3.17.d). Moreover, we observed one Ψ site positioned in the start codon (AUG), while two sites were found in the stop codon (UAG), which may promote stop codon readthrough according to previous research [228,258].

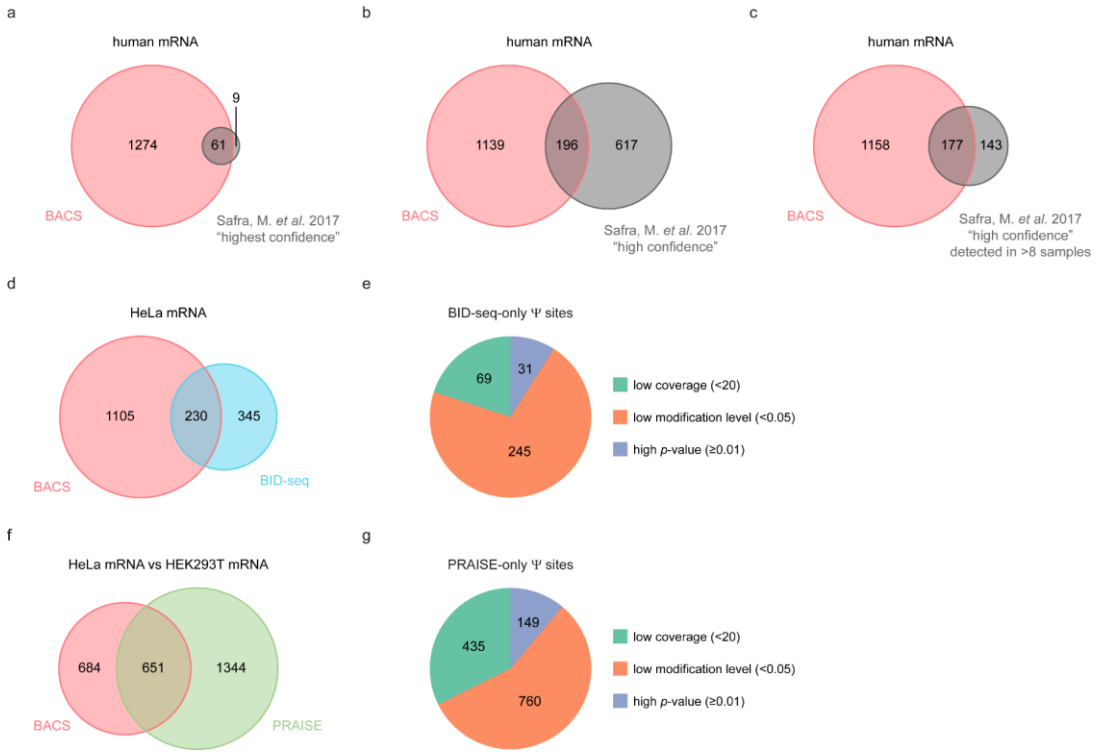


Figure 3.18 | Comparison of mRNA Ψ sites identified by BACS with published datasets. a. Venn diagram illustrating the overlap of mRNA Ψ sites between BACS and the "highest confidence" list in a consolidated CMC-based dataset [257]. **b.** Venn diagram illustrating the overlap of mRNA Ψ sites between BACS and the "high confidence" list in a consolidated CMC-based dataset [257]. **c.** Venn diagram illustrating the overlap of mRNA Ψ sites between BACS and the "high confidence" list in a consolidated CMC-based dataset [257]. Only Ψ sites consistently detected across more than 8 samples in the "high confidence" list were considered. **d.** Venn diagram illustrating the overlap of mRNA Ψ sites between BACS and BID-seq. **e.** Distribution of BID-seq-only Ψ sites in BACS dataset. **f.** Venn diagram illustrating the overlap of mRNA Ψ sites between BACS and PRAISE. **g.** Distribution of PRAISE-only Ψ sites in BACS dataset.

To further evaluate the performance of BACS, we compared our identified mRNA Ψ sites with published datasets. We first compared BACS with a recent dataset which consolidated three CMC-based methods [257]. Remarkably, BACS accurately identified 61 out of 70 Ψ sites (87.1%) listed in the "highest confidence" category (Figure 3.18.a).

However, a strong overlap between BACS and the “high confidence” list was achieved only when considering Ψ sites consistently detected across multiple samples (>8) (177 out of 320 Ψ sites, 55.3%), indicating the considerable variance between different CMC-based datasets (Figure 3.18.b,c). We further compared BACS with two recently developed BS-based methods. Compared to CMC-based approaches, BACS demonstrated a better overlap with BID-seq results, as expected [228] (230 out of 575 Ψ sites, 40.0%, Figure 3.18.d). Most of the sites exclusive to the BID-seq dataset displayed low modification levels in our BACS libraries, possibly due to the inaccurate quantification of BID-seq (Figure 3.18.e). When compared with PRAISE [229], 651 of 1995 Ψ sites (32.6%) showed an overlap with BACS results (Figure 3.18.f). Similarly, the majority of PRAISE-only Ψ sites were lowly modified in our dataset (Figure 3.18.g). Regarding the seven Ψ sites identified in mitochondrial mRNAs (mt-mRNAs) by BACS, four, two, and four of them have also been detected by Pseudo-seq [104], BID-seq [228], and PRAISE [229], respectively. It is important to note that the degree of overlap between different methods may be influenced by differences in sequencing depth and the bioinformatics pipelines used for analysis (for example, mapping to the genome or directly to the transcriptome). Therefore, there is a clear need for developing a standardized pipeline for analyzing RNA modifications.

3.8 Simultaneous profiling of A-to-I editing sites and m^1A in HeLa RNA

In addition to Ψ mapping, BACS enabled the simultaneous detection of A-to-I editing sites. After BACS, the A-to-G mutation signals induced by A-to-I editing would be erased in a similar way as shown in ICE-seq, which could distinguish the true editing sites from potential backgrounds [259] (Figure 3.16.a). We identified 1121 A-to-I editing sites in HeLa polyA-tailed RNA, with a mean modification level of 20% (Figure 3.15.a,b). In stark contrast to Ψ, the majority of A-to-I editing sites were resident in the *Alu* elements (Figure 3.16.e). We further annotated 737 and 115 A-to-I editing sites to mRNA and ncRNAs, respectively (Figure 3.15.c). Within mRNA, the A-to-I editing sites were predominantly enriched in 3'-UTR, consistent with previous findings [260] (Figure 3.15.d). Interestingly, mRNA transcripts that carry A-to-I editing sites did not overlap with those possessing Ψ, suggesting distinct roles of A-to-I editing and pseudouridylation in mRNA processing (Figure 3.15.e).

Similar to RBS-seq [226], BACS induces Dimroth rearrangement of m^1A to m^6A and could potentially detect m^1A together with Ψ [261] (Figure 3.19.a). As expected, we observed a significant reduction of m^1A mutation signals at tRNA position 58 (for cy-tRNAs) and 9 (for mt-tRNAs) after BACS treatment, which was comparable to the efficiency of demethylase [262] (Figure 3.19.b,c). These results highlight the value of BACS to detect multiple modifications simultaneously.

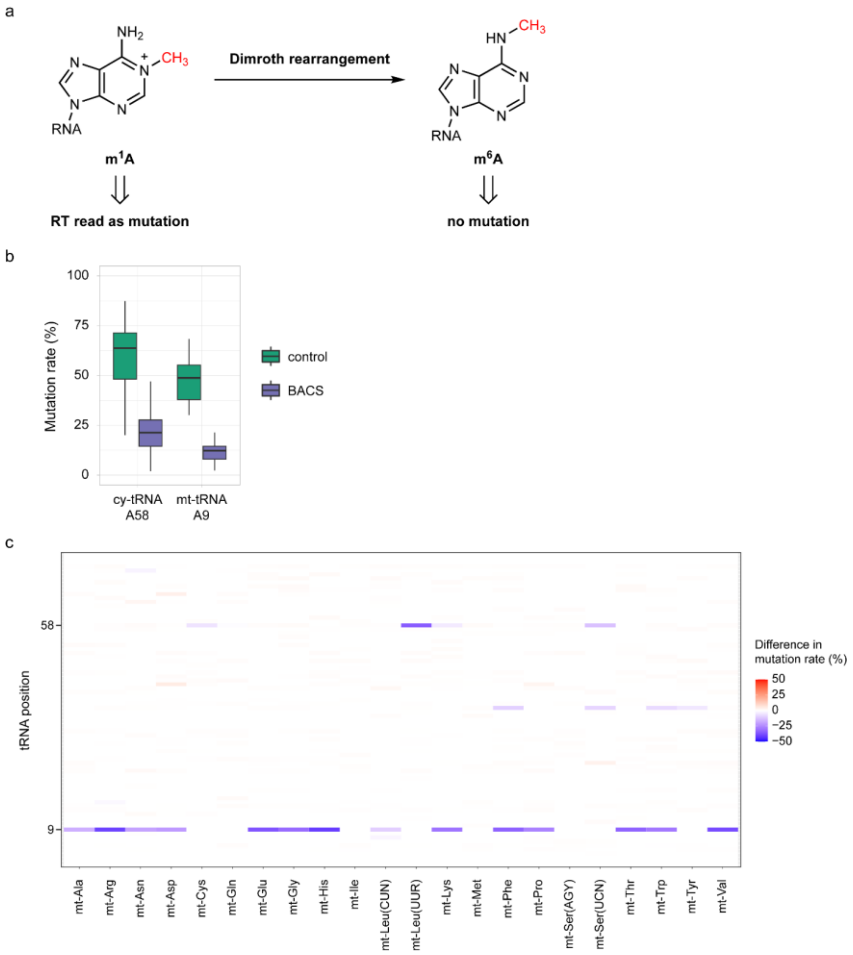


Figure 3.19 | BACS enabled simultaneous detection of m^{1}A with Ψ. a. Schematic overview of Dimroth rearrangement of m^{1}A to m^{6}A. **b.** Comparison of the mutation rates of known m^{1}A sites in control (green) and BACS (purple) samples. For boxplots, boxes represent the 25th to 75th percentiles with a line at the median; whiskers correspond to 1.5 times the interquartile range (cy-tRNA A58, n = 170; mt-tRNA A9, n = 14). **c.** Heatmap showing the changes in mutation rates of all adenosine sites in human mt-tRNAs upon BACS treatment. Red and blue color indicate an increase and decrease of mutation rates, respectively.

3.9 Application of BACS to low-input and single-cell samples

After applying BACS to various types of human RNA, we further applied BACS to low-input samples. As expected, BACS did not induce significant degradation of RNA. With minor modifications, BACS has been successfully applied to RNA samples down to 50 pg, which is the same level as single cell (Figure 3.20).

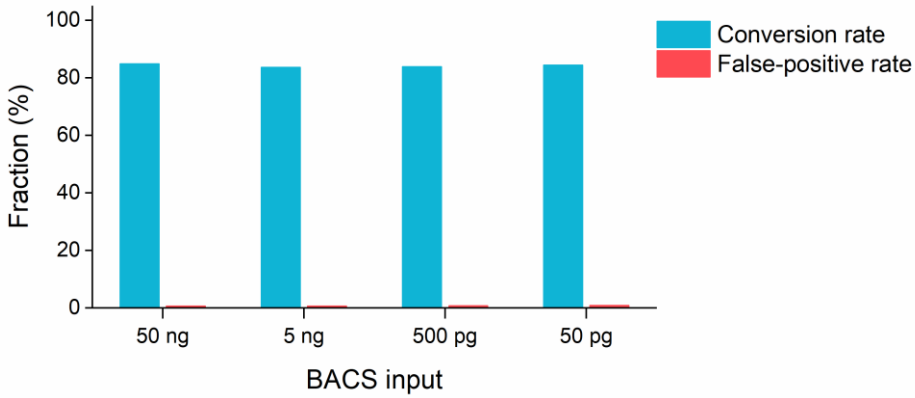


Figure 3.20 | Performances of low-input BACS.

3.10 Summary

In this chapter, I describe the development of a novel method, named BACS, for quantitative and base-resolution sequencing of Ψ. BACS is based on new bromoacrylamide cyclization chemistry and induces Ψ-to-C mutation signatures rather than truncation or deletion signatures, allowing for accurate quantification of Ψ stoichiometry and sequencing of Ψ at absolute single-base resolution. Importantly, BACS overcomes the inherent limitations of BS-based methods in three crucial aspects: (1) it facilitates the precise determination of Ψ sites located adjacent to one or more uridines; (2) it enhances the detection of densely modified Ψ sites with higher accuracy and sensitivity; and (3) it offers significantly more accurate quantification of Ψ in all sequence contexts. These advances make BACS a valuable tool for studying Ψ modifications in cellular RNAs, as it can provide a more comprehensive and accurate picture of the Ψ landscape across various RNA species.

Using BACS, we successfully detected all known Ψ sites in human rRNA and spliceosomal snRNAs and generated the first quantitative Ψ map of human snoRNA and tRNA. Combining BACS with the latest library construction techniques has the potential to further improve its performance on small RNAs [262,263]. We further applied BACS to HeLa mRNA and revealed a rather low level of pseudouridylation. We anticipate BACS to be widely adopted as the new standard to advance our understanding of Ψ modifications and their functional implications in diverse biological processes.

4 Applications of BACS to studying the biological functions and importance of Ψ and PUS enzymes

4.1 Comprehensive PUS-dependent Ψ map across the human transcriptome

To elucidate the PUS-dependent Ψ profile in the human transcriptome, we first generated individual knock out for three key PUS enzymes in HeLa cells: TRUB1, PUS7, and PUS1.

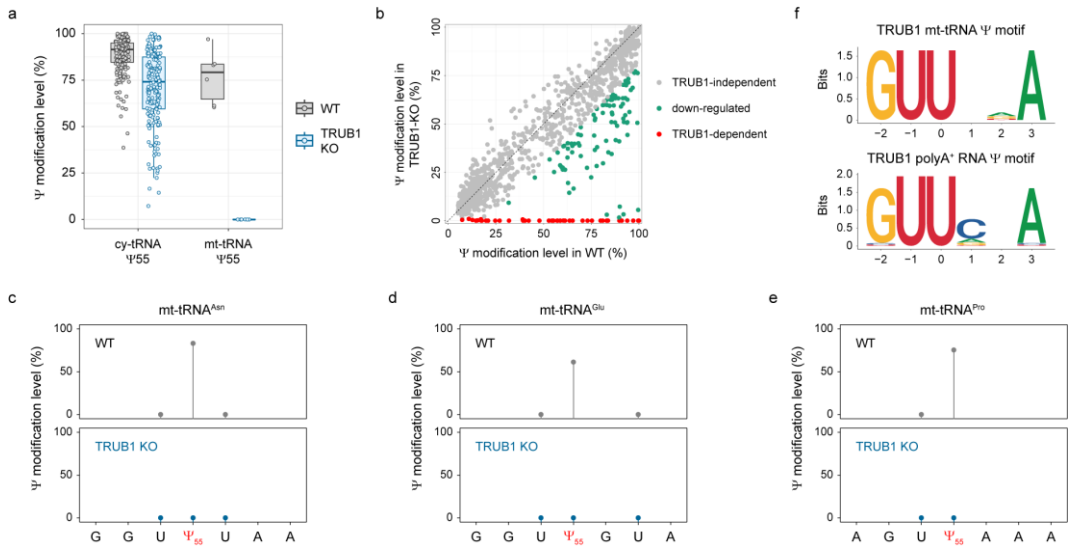


Figure 4.1 | TRUB1-dependent Ψ sites across the HeLa transcriptome. a. Comparison of the modification levels of Ψ55 in HeLa cy-tRNAs and mt-tRNAs upon TRUB1 depletion. Boxplots visualize all Ψ sites at each position; boxes represent the 25th to 75th percentiles with a line at the median; whiskers correspond to 1.5 times the interquartile range (cy-tRNA Ψ55, n = 180; mt-tRNA Ψ55, n = 6). **b.** Scatter plot illustrating all TRUB1-dependant Ψ sites across the HeLa transcriptome. **c–e.** Comparison of the modification levels of Ψ55 in mt-tRNA^{Asn} **(c)**, mt-tRNA^{Glu} **(d)**, and mt-tRNA^{Pro} **(e)** between WT and TRUB1-KO cell lines. **f.** Sequence motifs of TRUB1-dependent Ψ sites in mt-tRNAs (upper panel) and polyA-tailed RNA (lower panel).

In contrast to the conventional view that TRUB1 was the sole PUS enzyme responsible for cy-tRNA Ψ55 [264], depletion of TRUB1 did not eradicate Ψ55 in human cy-tRNAs, suggesting that other PUS enzymes may also participate in the pseudouridylation of this position (Figure 4.1.a,b). Surprisingly, all the Ψ55 sites in mt-tRNA were eliminated upon TRUB1 depletion, challenging another conventional belief that TRUB2 was solely responsible for mt-tRNA Ψ55 [265] (Figure 4.1.a,b). We further discovered that the TRUB1 motif would extend beyond the recognized GUΨCNA (N = A, C, G or U) motif [105,257], since it could modify GUΨUAA in mt-tRNA^{Asn}, GUΨGUA in mt-tRNA^{Glu}, and GUΨAAA in mt-tRNA^{Pro} with high efficiency (Figure 4.1.c–f). We observed similar results in polyA-tailed RNA, confirming that TRUB1 could edit GUΨGNA, GUΨCNA and GUΨUNA motifs (Figure 4.1.f). Therefore, we demonstrate that the TRUB1 motif can be extended from GUΨCNA to GUΨNNA.

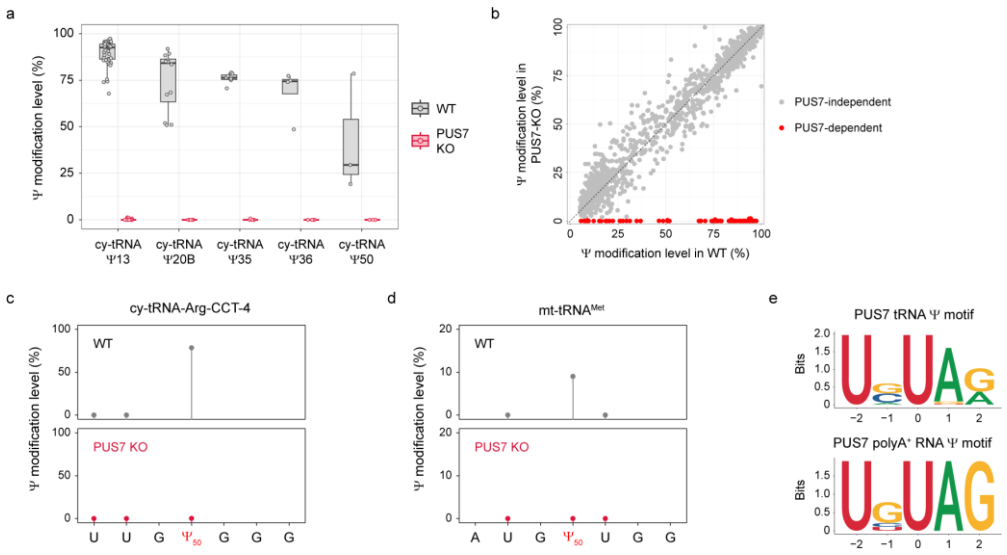


Figure 4.2 | PUS7-dependent Ψ sites across the HeLa transcriptome. a. Comparison of the modification levels of Ψ sites at selected positions of HeLa cy-tRNAs upon PUS7 depletion. Boxplots visualize all Ψ sites at each position; boxes represent the 25th to 75th percentiles with a line at the median; whiskers correspond to 1.5 times the interquartile range (cy-tRNA: Ψ13, *n* =

43; Ψ20B, *n* = 12; Ψ35, *n* = 7; Ψ36, *n* = 4; Ψ50, *n* = 3). **b.** Scatter plot illustrating all PUS7-dependant Ψ sites across the HeLa transcriptome. **c,d.** Comparison of the modification levels of Ψ50 in cy-tRNA$^{Arg(CCT)}$ **(c)** and mt-tRNA^{Met} **(d)** between WT and PUS7-KO cell lines. **e.** Sequence motifs of PUS7-dependent Ψ sites in tRNAs (upper panel) and polyA-tailed RNA (lower panel).

PUS7-knockout HeLa cells allowed us to reveal its role in the formation of Ψ20B, Ψ36, and Ψ50 in cy-tRNAs, in addition to its previously known targets Ψ13 and Ψ35 [266] (Figure 4.2.a,b). It is noted that both Ψ35 and Ψ36 are located within the anticodon, suggesting that PUS7 depletion could induce miscoding events of cy-tRNA$^{Tyr(GTA)}$ and cy-tRNA$^{Arg(TCT)}$, respectively. Moreover, our study uncovers a novel activity for PUS7 in mitochondria by catalyzing the modification of Ψ50 in mt-tRNA^{Met}, expanding its functional repertoire within this organelle for the first time (Figure 4.2.d). The majority of PUS7 targets in tRNA displayed conserved UVΨAR (V = A, C or G; R = A or G) motif (Figure 4.2.e). However, PUS7 displayed comparable activity within the UGΨGG motif (cy-tRNA$^{Arg(CTT)}$ Ψ50) and relatively low activity within the UGΨUG motif (mt-tRNA^{Met} Ψ50) (Figure 4.2.c–e). In polyA-tailed RNA, PUS7 mainly catalyzed the pseudouridylation within UBΨAG (B = C, G or U) motif (Figure 4.2.e). Collectively, the true PUS7 consensus motif would be UNΨAR (N = A, C, G or U; R = A or G) and UGΨKG (K = G or U), which was less strict than previously considered UGΨAR (R = A or G) motif [104,267].

PUS1 depletion resulted in the complete loss of Ψ27/28 in human cy-tRNAs (Figure 4.3.a,b). In mt-tRNAs, PUS1 not only catalyzed the modification of Ψ27/28 but also induced the formation of Ψ66/67/68, consistent with the findings on yeast and mouse PUS1 homologs [268] (Figure 4.3.a,b). We also observed non-canonical activity of PUS1,

exemplified by Ψ25 in mt-tRNA^{Asn} and Ψ20 in mt-tRNA$^{Leu(UUR)}$, thus affirming the precision of BACS in capturing diverse pseudouridylation events (Figure 4.3.d,e). Additionally, PUS1 was found to catalyze all 7 Ψ sites identified in mt-mRNAs (Figure 4.3.c). These results show that PUS1 is the major PUS enzyme in mitochondria with diverse functions. PUS1-dependent Ψ sites did not show any sequence motifs, consistent with an earlier report that its activity is dependent on RNA structure [267] (Figure 4.3.f).

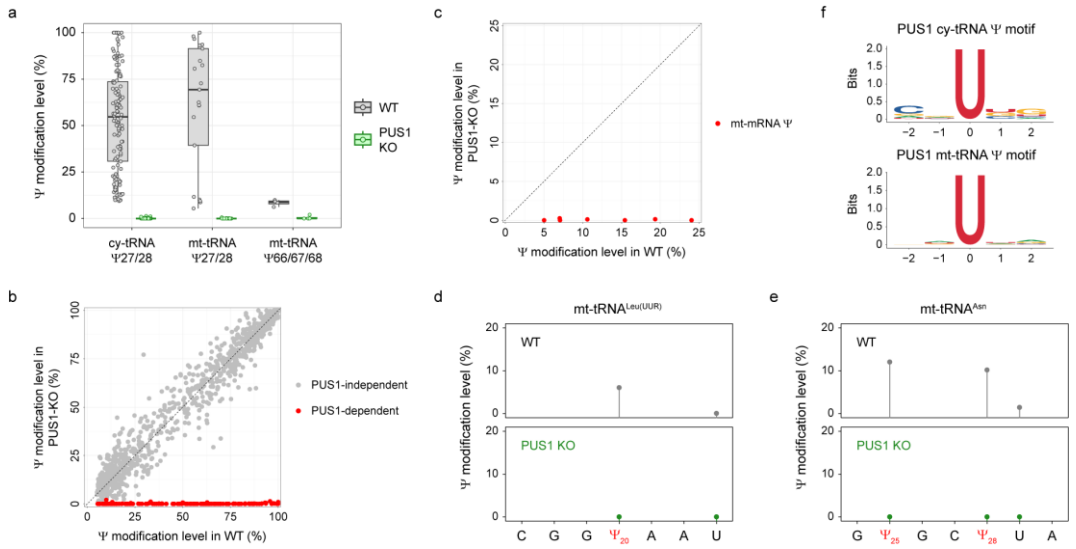


Figure 4.3 | PUS1-dependent Ψ sites across the HeLa transcriptome. a. Comparison of the modification levels of Ψ sites at selected positions of HeLa cy-tRNAs and mt-tRNAs upon PUS1 depletion. Boxplots visualize all Ψ sites at each position; boxes represent the 25th to 75th percentiles with a line at the median; whiskers correspond to 1.5 times the interquartile range (cy-tRNA: Ψ27/28, n = 130; mt-tRNA: Ψ27/28, n = 21; Ψ66/67/68, n = 4). **b.** Scatter plot illustrating all PUS1-dependant Ψ sites across the HeLa transcriptome. **c.** Scatter plot illustrating all PUS1-dependant Ψ sites in HeLa mt-mRNAs. **d,e.** Comparison of the modification levels of Ψ20 in mt-tRNA$^{Leu(UUR)}$ **(d)** and Ψ25 in mt-tRNA^{Asn} **(e)** between WT and PUS1-KO cell lines. **f.** Sequence motifs of PUS1-dependent Ψ sites in cy-tRNAs (upper panel) and mt-tRNAs (lower panel).

Mimicry of tRNA has been considered as a general way for mRNA pseudouridylation [105,257]. Although TRUB1, PUS7, and PUS1 could edit both tRNAs and polyA-tailed

RNA, the Ψ targets in polyA-tailed RNA were significantly less modified than their counterparts in tRNAs (Figure 4.4). These results suggest that mRNA may not be the primary substrate of these stand-alone PUS enzymes, which is consistent with our earlier data showing that the majority of mRNA Ψ sites were modified to a low level.

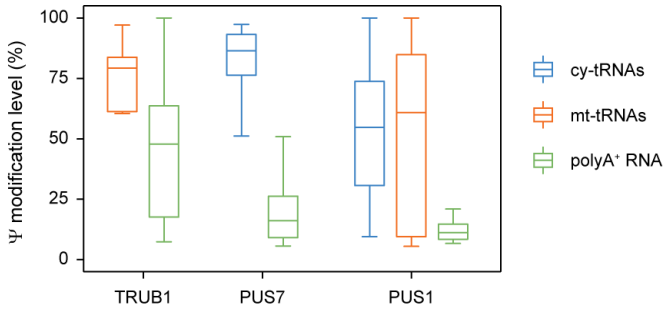


Figure 4.4 | Comparison of the modification levels of PUS-dependent Ψ sites across cy-tRNAs (blue), mt-tRNAs (orange), and polyA-tailed RNA (green). For boxplots, boxes represent the 25th to 75th percentiles with a line at the median; whiskers correspond to 1.5 times the interquartile range (TRUB1: mt-tRNAs, n = 6; polyA-tailed RNA, n = 41; PUS7: cy-tRNAs, n = 69; polyA-tailed RNA, n = 22; PUS1: cy-tRNAs, n = 130; mt-tRNAs, n = 27; polyA-tailed RNA, n = 8).

The individual knockout/knockdown of the other 10 PUS enzymes is in collaboration with Mengjie Li from Parinaz Mehdipour Group. Some of the preliminary results are summarized in Table 4.1 and Figure 4.5.

Table 4.1 | Summary of enzymatic properties of human PUS enzymes.

PUS	cy-rRNA snRNA	mt-rRNA	cy-tRNA	mt-tRNA	Motif
PUS1		12S all	27/28	20 25/27/28 66/67/68	None Structural
PUSL1				38/39/40	None Structural
PUS3			38/39/40		None Structural
PUS7			13 20B 35 36 50	50	UNΨAR UGΨKG
PUS7L			e12/e13/e1		None Structural
PUS10			54, 55		GΨUCRARUC GUΨCRAWWC
TRUB1			55	55	GUΨCRANYC GUΨNNA
TRUB2					No activity ?
RPUSD1			30 72		GKΨRCYW
RPUSD2			31/32 34	31/32	VNΨHWNND
RPUSD3					No activity
RPUSD4		16S 1397			None
DKC1	all				box H/ACA snoRNA dependent

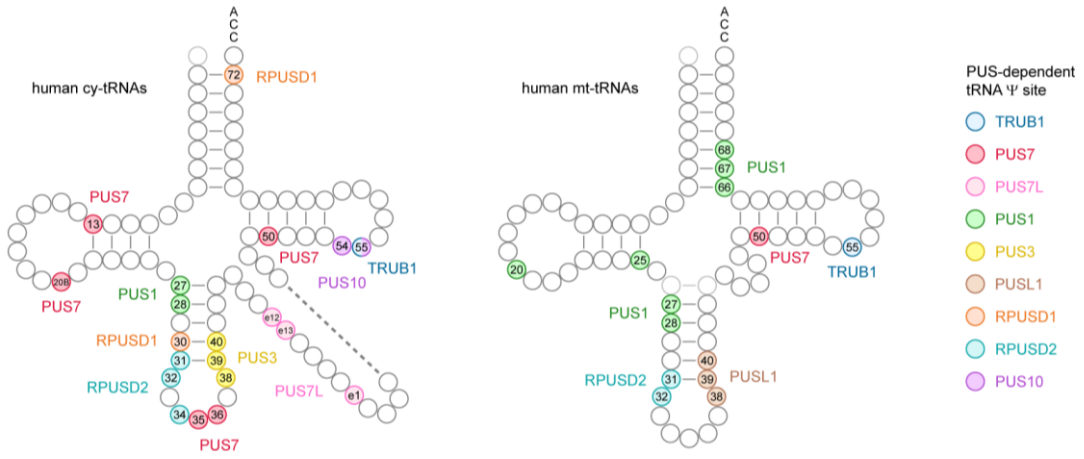


Figure 4.5 | PUS-dependent Ψ landscape across the HeLa transcriptome.

4.2 Ψ is installed at various stages of pre-tRNA processing

In eukaryotes, nuclear tRNA genes are initially transcribed as precursor tRNAs (pre-tRNAs), which will undergo a series of processing steps, including removal of the 5'-leader and 3'-trailer sequences, addition of CCA to the 3'-end, splicing of the intron if presents, and introducing a wide range of chemical modifications [269,270]. Given the high abundance of Ψ in mature cy-tRNAs, understanding the temporal order of pseudouridylation could provide valuable insights into the pre-tRNA processing. However, the Ψ profile of pre-tRNAs is only partially elucidated in yeast [271,272]. To detect the Ψ modification in human pre-tRNAs, we applied stringent filtering to the BACS reads generated from HeLa cells. More specifically, we discarded all the reads which were mapped to mature cy-tRNAs but kept the reads covering at least one of the 5'-leader, 3'-trailer or intron sequences of pre-tRNAs. Benefited from the high sensitivity and accuracy of BACS, we were able to cover all canonical Ψ positions in mature cy-tRNAs, with preliminary results listed in Table 4.2 and Figure 4.6.

Table 4.2 | Ψ modifications at different stages of pre-tRNA processing.

pre-tRNA	Stage 1 (before 5'-/3'- processing)	Stage 2 (after 5'-/3'- processing; before intron splicing)	Stage 3 (after intron splicing)
with intron			
Arg-TCT	27–28 (PUS1) 36 (PUS7) 54, 55 (TRUB1/PUS10)		27–28 (PUS1) 40 (PUS3)
Ile-TAT	34 (RPUSD2) 55 (TRUB1/PUS10)		30 (RPUSD1) 39 (PUS3)
Leu-CAA	20B (PUS7) 55 (TRUB1/PUS10)		38–39 (PUS3) e12 (PUS7L)
Tyr-GTA	35 (PUS7) 55 (TRUB1/PUS10)		39 (PUS3)
intron-free	13, 20B, 50 (PUS7) 27–28 (PUS1) 30 (RPUSD1) 31–32 (RPUSD2) 38–40 (PUS3) e12, e1 (PUS7L) 54 (PUS10) 55 (TRUB1/PUS10)	72 (RPUSD1)	

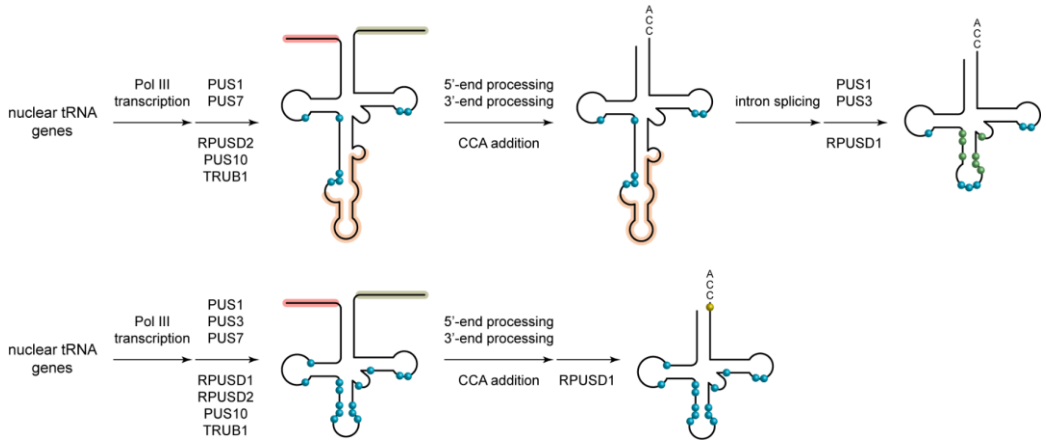


Figure 4.6 | Ψ is installed at three different stages of pre-tRNA processing. Top, intron-containing pre-tRNAs; bottom, intron-free pre-tRNAs. Blue, yellow, and green dots denote Ψ sites modified at Stage 1, 2, and 3, respectively. Red, green, and orange line denote the 5'-leader, 3'-trailer, and intron sequences, respectively.

We first focused on intron-containing pre-tRNAs in human cells, including Arg-TCT, Ile-TAT, Leu-CAA, and Tyr-GTA. Interestingly, we found these pre-tRNAs possessed several Ψ sites which could only be modified after the intron splicing, especially for the Ψ localized near the intron sequence, including RPUSD1-dependent Ψ30 and PUS3-depednent Ψ38–40. Moreover, the intron sequence also completely blocked the formation of PUS7L-dependent Ψe12 in pre-tRNA$^{Leu\text{-}CAA}$, which is relatively far from the anticodon. Additionally, Ψ28 was found to be unmodified in pre-tRNA$^{Arg\text{-}TCT}$ with intron, while Ψ27 was partially modified, indicating the PUS1 activity was also affected by the intron sequence. In contrast, the anticodon Ψ34 / Ψ35 / Ψ36 were all introduced before the intron splicing, which is consistent with their yeast counterparts [273,274]. In addition, pre-tRNA$^{Leu\text{-}CAA}$ Ψ20B, and pre-tRNA$^{Arg\text{-}TCT}$ Ψ54 / Ψ55 were introduced before the trimming of 5'-leader and 3'-trailer sequence, respectively, indicating these Ψ sites were installed at a very early stage of pre-tRNA processing.

We further expanded the analysis to intron-free pre-tRNAs, which represents the majority of pre-tRNAs in human cells. Due to the difficulty in mapping, most of the identified pre-tRNA Ψ sites were Ψ55, which is conserved across all human cy-tRNAs except cy-tRNA^{iMet}. As expected, we confirmed TRUB1/PUS10-dependent Ψ55 was introduced before the trimming of 3'-trailer sequence. Similarly, Ψ54, as another PUS10-depedent site, was modified together with Ψ55. More importantly, most of the other Ψ modifications were installed in the early stage of pre-tRNA processing, with the exception of Ψ72, which was only introduced after 3'-end processing of pre-tRNA$^{Arg\text{-}CCT}$ and pre-tRNA$^{Arg\text{-}TCG}$. This phenomenon could be explained by the altered secondary structure of pre-tRNA acceptor arm compared to mature tRNA and the absence of 3'-CCA end.

Notably, m^5C72 was also found to be installed after 3'-CCA addition, where the 3'-CCA end played an important role for NSUN6 recognition [275,276], suggesting a similar way for RPUSD1 recognition of its Ψ72 substrate. Taken together, we demonstrated that PUS enzymes modify Ψ at three different stages of pre-tRNA processing.

4.3 Mapping of Ψ in viral RNAs

It has been widely accepted that Ψ-modified RNAs can suppress innate immune responses and may influence the mRNA vaccine design [277]. Although several studies have reported the presence of Ψ in viral RNAs [278,279], it has not been thoroughly confirmed with the latest sequencing technologies. We therefore applied BACS to various RNA viruses, including severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), hepatitis C virus (HCV), Zika virus (ZIKV), hepatitis delta virus (HDV), and Sindbis virus (SINV), to study the existence of Ψ in viral transcripts and genomes. We first focused on SARS-CoV-2, since earlier studies reported Ψ in its genomic and subgenomic RNAs by Nanopore sequencing [278,279]. Similar to previous RNA-seq results [280], the majority of reads were mapped to the positive strand with a unique pattern at the 3'-end corresponding to the subgenomic RNAs (Figure 4.7.a). However, we could not detect any high-confidence Ψ sites in SARS-CoV-2 RNA (Figure 4.7.b,c). Surprisingly, we did not identify any high-confidence Ψ sites in the other four RNA viruses (Figure 4.7.d–g). Importantly, we confirmed robust viral infection in our model systems, to ensure a high abundance of viral RNAs concomitant with high depth of coverages. These results suggest that Ψ is not directly involved in the modification of these RNA viruses.

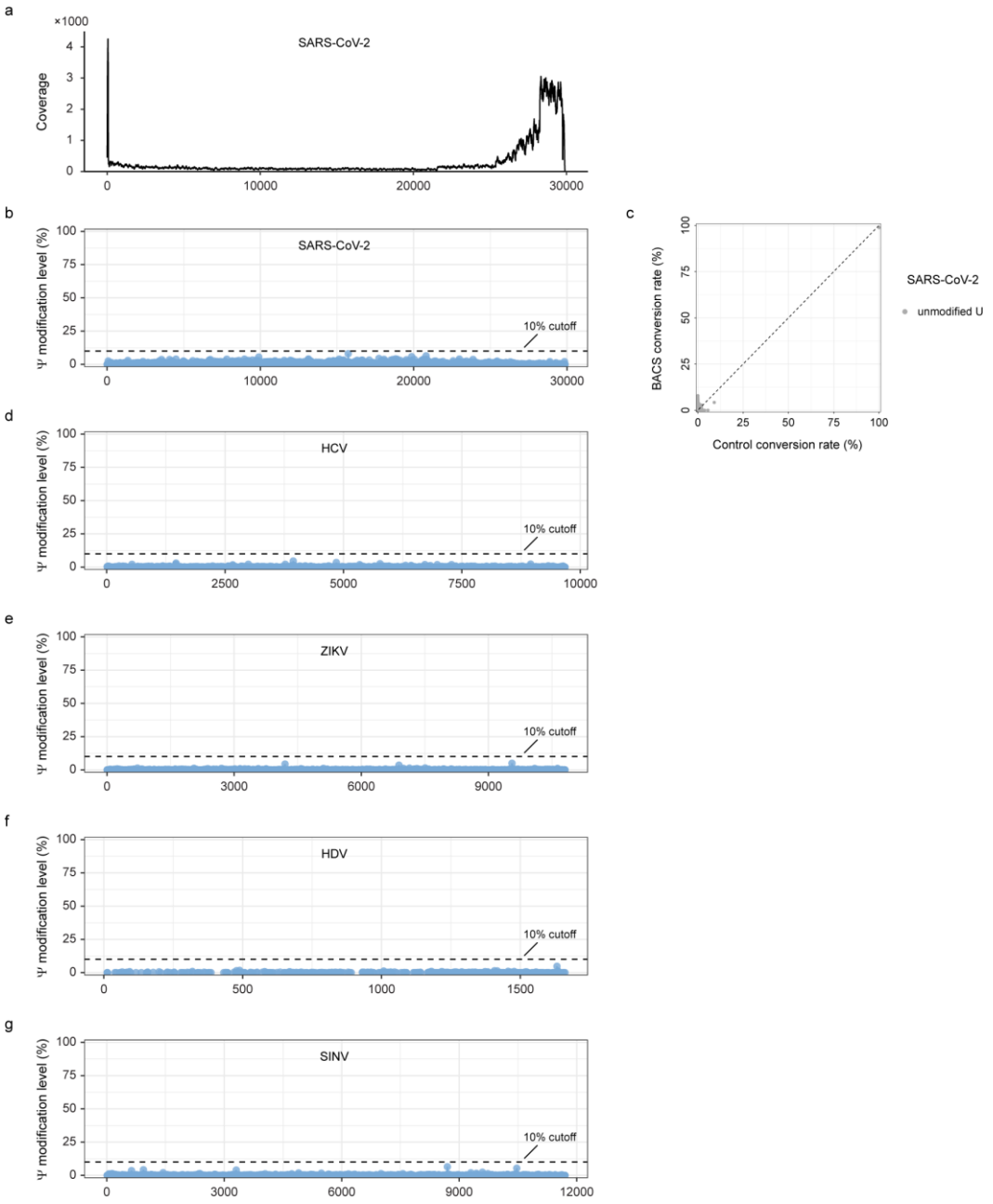


Figure 4.7 | Absence of Ψ in transcripts and genomes of RNA viruses. a. Sequencing coverage of SARS-CoV-2 viral RNA. **b.** Ψ modification levels in SARS-CoV-2 viral RNA. **c.** Comparison of the conversion rates in SARS-CoV-2 viral RNA between BACS and control samples. **d–g.** Ψ modification levels in HCV **(d)**, ZIKV **(e)**, HDV **(f)**, and SINV **(g)** viral RNAs.

In addition to these RNA viruses, we also applied BACS to human cell lines infected by Epstein-Barr virus (EBV), a DNA virus that encodes two highly expressed ncRNAs, EBER1 and EBER2 (EBER, EBV-encoded small RNA), both of which are RNA polymerase III transcripts [281,282]. A previous study using HydraPsi-seq [283] and CMC-based primer extension methods reported one lowly modified Ψ160 site in EBER2 [284] (Figure 4.8.a). However, we detected one highly modified Ψ114 site in EBER2, while no Ψ site was identified in EBER1, indicating the previous results were likely caused by high backgrounds from hydrazine and CMC chemistry (Figure 4.8.a,b). Notably, this novel Ψ site was predicted to be in a loop region accessible for PUS enzymes. The EBER2 Ψ114 site was conserved across all EBV strains and host cell lines tested, including nasopharyngeal carcinoma cell line C666-1 and two Burkitt's lymphoma cell lines Raji and Elijah (Figure 4.8.b).

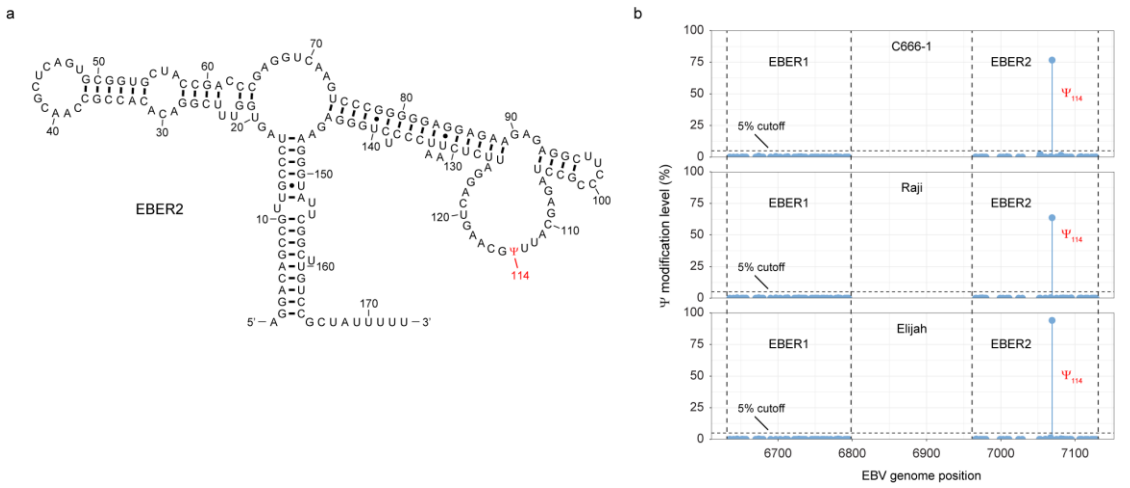


Figure 4.8 | Investigation of Ψ modification in EBV RNAs. a. Canonical EBER2 structure, with Ψ114 site labeled accordingly. **b.** Ψ modification levels in EBER1 and EBER2 from C666-1 (upper panel), Raji (middle panel), and Elijah (lower panel) cell lines.

To validate this novel Ψ site, we applied modified BID-seq to EBERs from Raji and Elijah cell lines. Indeed, we detected a potential Ψ site within the three consecutive uridines (U112–U114) in EBER2, while no Ψ site was found in EBER1 (Figure 4.9.a). However, BS-based methods could not determine the exact position of this site due to the limitations of deletion signals, which further strengthens the advantages of BACS (Figure 4.9.b). Taken together, these results suggest different ways of utilizing Ψ between virus families and further highlight BACS as a highly specific method.

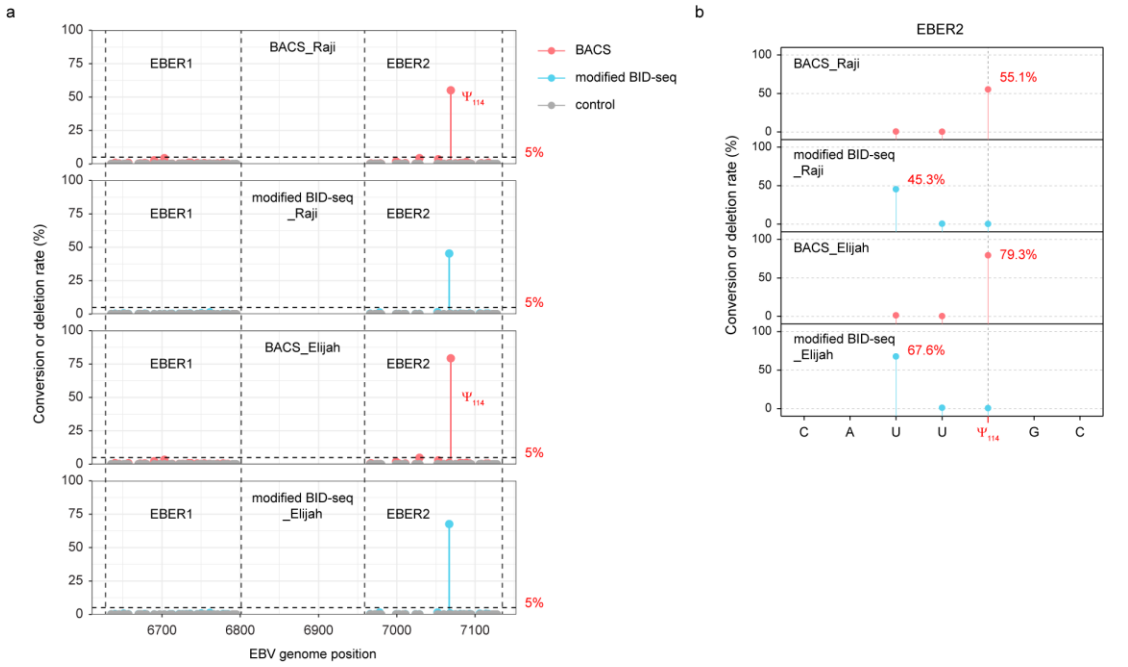


Figure 4.9 | Validation of Ψ114 site in EBER2. a. Comparison of the conversion rates of BACS (pink) with the deletion rates of modified BID-seq (blue) for EBER1 and EBER2. **b.** Comparison of the conversion rates of BACS (pink) with the deletion rates of modified BID-seq (blue) in the EBER2 Ψ114 region. Due to their deletion signals, BS-based methods cannot determine the exact position or identify the number of Ψ within consecutive uridine contexts. By default, the aligner will position the deletion at the 5'-most uridine.

4.4 Repurposing 5-fluorouracil as a hypopseudouridylating agent through inhibition of PUS enzymes

5-Fluorouracil (5-FU) is one of the most commonly used chemotherapy drugs for colorectal cancer [285]. Although its mechanism of action (MoA) has long been attributed to the inhibition of thymidylate synthase [286], there is emerging evidence revealing that RNA modification, processing, and damage pathways play more important roles in 5-FU cytotoxicity [287-291]. Particularly, one recent study using the cellular thermal shift assay (CETSA) screened out several PUS enzymes as the targets of 5-FU treatment [290]. Indeed, 5-fluorouridine (F^5U)-labeled RNA oligonucleotides were characterized as potent inhibitors of multiple *E. coli* and yeast PUS enzymes *in vitro* [292,293], and therefore have been widely used to study the catalytic mechanisms [294-297] and resolve the structures of PUS enzymes [298-300]. However, a comprehensive evaluation of 5-FU inhibitory effects of human PUS enzymes has not been achieved, limiting our understanding of 5-FU MoA.

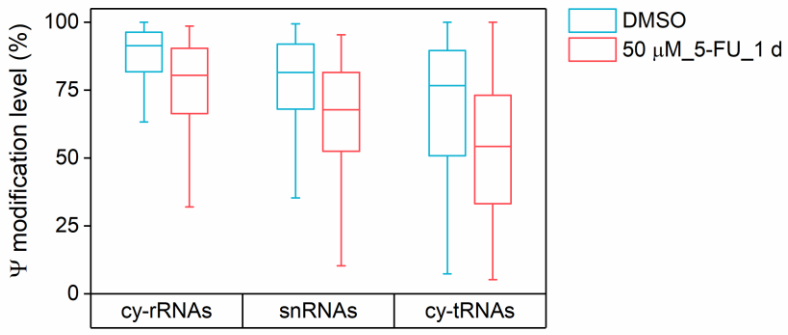


Figure 4.10 | Comparison of the modification levels of Ψ sites in cy-rRNAs, spliceosomal snRNAs, and cy-tRNAs from untreated and 5-FU treated HeLa cells.

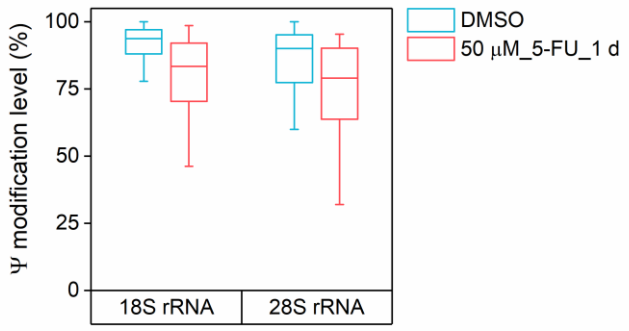


Figure 4.11 | Comparison of the modification levels of Ψ sites in 18S and 28S cy-rRNAs from untreated and 5-FU treated HeLa cells.

To overcome this challenge, we treated HeLa cells with 5-FU and quantified the Ψ modification levels using BACS. First, we observed a global reduction of Ψ modification levels in cy-rRNAs, spliceosomal snRNAs, and cy-tRNAs (Figure 4.10). Generally, Ψ sites in spliceosomal snRNAs and cy-tRNAs were more significantly altered than those in cy-rRNAs, while there was no obvious difference between 18S rRNA and 28S rRNA (Figure 4.10 and 4.11). Next, we analyzed the reduction of Ψ modification levels site-specifically, which could directly reflect the 5-FU inhibitory effects to each of the PUS enzymes. For DKC1-dependent Ψ sites, we found a clear site-specific pattern upon 5-FU treatment, possibly implying the varied catalytic kinetics among different box H/ACA snoRNAs. For example, the modification level of Ψ43 in U2 spliceosomal snRNA was significantly reduced, while that of Ψ54 remained stable (Figure 4.12). In addition to DKC1, we also observed that all stand-alone PUS enzymes that worked on cy-tRNAs (including PUS1, PUS3, PUS7, TRUB1/PUS10, RPUSD1, and RPUSD2) were inhibited to different extents upon 5-FU treatment (Figure 4.13). Among them, RPUSD2 and PUS10 were potently inhibited, while PUS7 was only inhibited in a minor way.

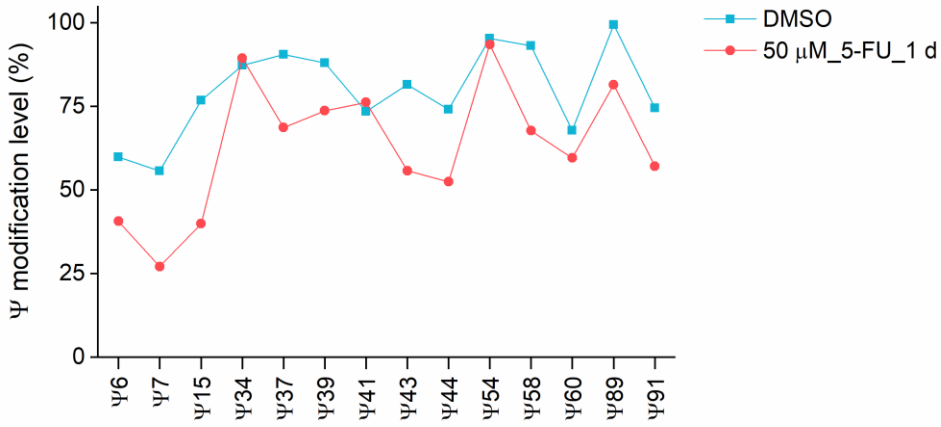


Figure 4.12 | Comparison of the modification levels of Ψ sites in U2 snRNA from untreated and 5-FU treated HeLa cells.

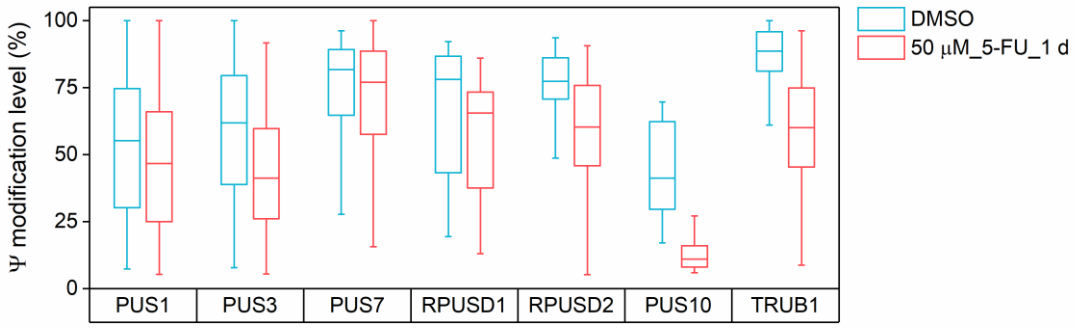


Figure 4.13 | Comparison of the modification levels of PUS-dependent Ψ sites in cy-tRNAs from untreated and 5-FU treated HeLa cells.

We noticed that many stand-alone PUS enzymes could work on both cy-tRNAs and mt-tRNAs. Therefore, we compared the changes of Ψ modification levels in cy-tRNAs with their counterparts in mt-tRNAs to further explore the 5-FU inhibitory effects (Figure 4.14). Surprisingly, we found that PUS1 and TRUB1 were not inhibited in mt-tRNAs, although they were significantly inhibited in cy-tRNAs. However, RPUSD2-dependent Ψ sites

were potently inhibited in both cy-tRNAs and mt-tRNAs. This discrepancy between cy-tRNAs and mt-tRNAs could be explained by several possibilities, such as low concentrations of 5-FU metabolites in mitochondria, or a low incorporation rate of 5-FUTP into mitochondrial RNAs. Nevertheless, the significant inhibition of RPUSD2 in mt-tRNAs indicated a strong (possibly covalent) inhibition of this enzyme in the nucleus (or cytosol) before it is transported into mitochondria, resulting in its deficiency within mitochondria. Indeed, previous studies have suggested a covalent inhibition of its ancestral *E. coli* RluA enzyme by F^5U-labeled RNA [299]. To confirm that RPUSD2 is inhibited outside the mitochondria, we further investigated RPUSD4, another human PUS enzyme belonging to the RluA family, which predominantly works on 16S mt-rRNA and therefore should not be inhibited in the nucleus or cytosol. As anticipated, we did not observe any significant inhibition of RPUSD4 after 5-FU treatment, strongly supporting our hypothesis (Figure 4.15).

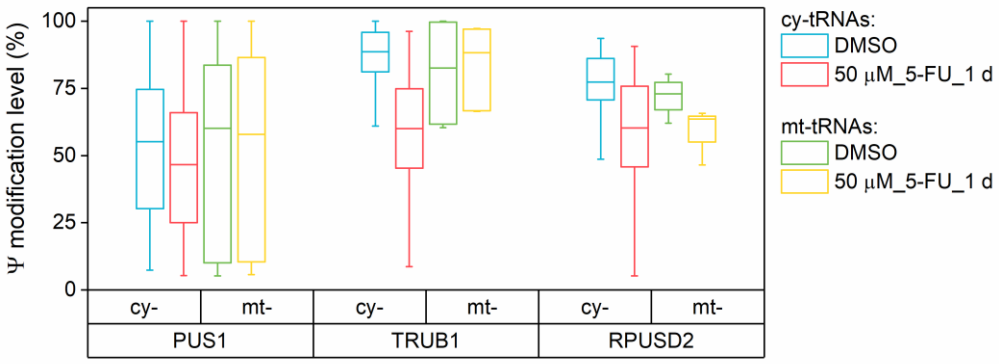


Figure 4.14 | Comparison of the modification levels of PUS-dependent Ψ sites in cy-tRNAs and mt-tRNAs from untreated and 5-FU treated HeLa cells.

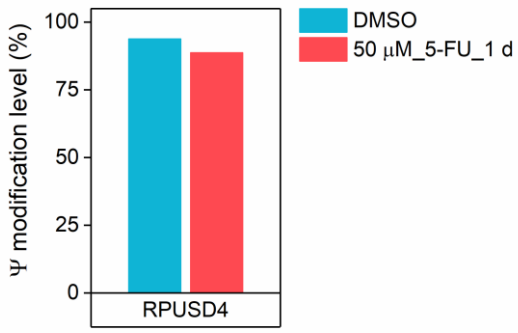


Figure 4.15 | Comparison of the modification levels of RPUSD4-dependent Ψ1397 site in 16S mt-rRNA from untreated and 5-FU treated HeLa cells.

Based on the above results, we concluded that 5-FU could serve as a global hypopseudouridylating agent *in vivo* through RNA incorporation and inhibition of various PUS enzymes. Future studies will focus on understanding the relationship between PUS enzymes and 5-FU resistance.

5 Discussions — insights into future directions of Ψ research

In Chapter 4, I introduced the ongoing research on Ψ, covering topics such as PUS-dependent Ψ map, Ψ and pre-tRNA processing, Ψ in viral RNAs, and the inhibition of PUS enzymes by 5-FU. In this Chapter, I will propose future directions of Ψ research, including the evolution of PUS enzymes, the elegant strategy for efficient pseudouridylation, the relationship between Ψ and cancer, and the relationship between Ψ and innate immunity.

5.1 Evolution of PUS enzymes

As mentioned in Chapter 1, the PUS enzyme is predicted to be one of the most conserved protein family across all cellular organisms, like ribosomal proteins and aminoacyl-tRNA synthetases [107]. Indeed, Ψ is prevalent in rRNAs and tRNAs, both of which are the essential components of translation machinery. As we know, translation is a hallmark of emergence of cellular organisms [301]. Therefore, it is highly likely that Ψ was already present at the very beginning of cellular life. Since rRNAs and ribosomal proteins are widely used for evolutionary analysis [302], it is reasonable to postulate that PUS enzymes are co-evolved with rRNAs, tRNAs, and other relevant proteins. Such analysis may help us uncover the mystery of the evolution of life.

To date, all PUS enzymes can be divided into six families, including TruA, TruB, TruD, RluA, RsuA, and Pus10 families, according to their sequence similarity of the catalytic domain [118] (Table 5.1). Although Ψ positions are highly conserved among all

organisms that have been investigated, there are relatively large divergence of PUS enzymes in different organisms, in terms of their substrate preferences and enzymatic activities. For example, bacteria can possess up to 5 families of PUS enzymes (TruA, TruB, TruD, RluA, and RsuA), while only 4 families (TruA, TruB, TruD, and Pus10) have been found in archaea. In particular, the archaeal TruB family possesses one unique type of PUS enzyme, i.e. Cbf5/Dkc1, which utilizes box H/ACA snoRNAs to guide pseudouridylation [303]. Therefore, it is likely that only TruA, TruB, and TruD families of PUS enzymes were present in LUCA. Bacteria and archaea independently evolved the RluA/RsuA families and the Cbf5/Dkc1 subfamily, respectively, to meet the essential requirement for rRNA pseudouridylation. Interestingly, RluA and RsuA families get significantly expanded or lost in different types of bacteria, while the Cbf5/Dkc1 subfamily is highly conserved among all types of archaea [238], possibly suggesting the high importance of Ψ to archaeal ribosomes. Furthermore, archaea also evolved Pus10 family to replace the bacterial TruB family for tRNA Ψ55 modification [304].

As for PUS enzymes in eukaryotes, we can find that they are a combination of bacterial and archaeal PUS enzymes. Indeed, from a modern perspective of evolution, early eukaryotic cells inherited an archaeal origin of nucleus/cytosol and a bacterial origin of mitochondria which is derived from symbiogenesis [305]. Therefore, there should be two sets of PUS enzymes: TruA, Cbf5/Dkc1, TruD, and Pus10 from archaea, and TruA, TruB, RluA, RluC, and RluB from alphaproteobacteria, which are responsible for the pseudouridylation of cytosolic RNAs (cy-RNAs) and mitochondrial RNAs (mt-RNAs), respectively. This hypothesis is supported by the following facts:

Table 5.1 | Comparison of PUS enzymes in different species.

PUS family	LUCA	Bacteria		Archaea	Eukarya	
		E. coli (gamma) [306]	*R. conorii* (alpha)	*P. furiosus*	*S. cerevisiae* [214]	*H. sapiens* [215]
TruA	TruA	TruA	TruA	TruA	PUS3	PUS3
						PUSL1
					PUS1	PUS1
					PUS2	
TruB	TruB	TruB	TruB		PUS4	TRUB1
						TRUB2
				Cbf5	Cbf5	DKC1
TruD	TruD	TruD		TruD	PUS7	PUS7
						PUS7L
RluA		RluA	RluA		PUS8	RPUSD2
					PUS9	
					PUS6	
						RPUSD1
		RluC	RluC		PUS5	RPUSD4
						RPUSD3
		RluD				
		TruC				
RsuA		RsuA				
		RluB	RluB			
		RluE				
		RluF				
Pus10				Pus10		PUS10

Text color denotes PUS enzymes working on rRNAs (red) and tRNAs (green).

Shading color denotes PUS enzymes working on cy-RNAs (blue) and mt-RNAs (orange).

- The highly conserved Cbf5/Dkc1 in eukaryotes, which is responsible for cy-rRNAs pseudouridylation.

- The expansion of TruA family in eukaryotes (Pus3, Pusl1, and Pus1), since at least two PUS enzymes from the TruA family are inherited.

- The conserveness of RluC/Rpusd4, and its responsible Ψ site in mitochondrial large subunit.

- The existence of RluB in early eukaryotes, although it is lost in higher eukaryotes.

This theory can also be applied to the evolution of plant PUS enzymes:

- Symbiogenesis of cyanobacteria explains the further expansion of TruA and RluA families in plants.

- Different sets of PUS enzymes are allocated to mitochondria and chloroplasts [307].

During the evolution of eukaryotes, the PUS enzymes suffer from multiple rounds of gain and loss. This phenomenon is particularly important for Trub1 and Pus10, which have redundant activities for Ψ55 in cy-tRNAs [308]. For example, yeast lost Pus10 during evolution, while fruit fly may have lost Trub1. Additionally, several PUS enzymes emerged only in higher eukaryotes, including Trub2, Pus7l, and Rpusd3, probably due to the further expansion of the corresponding PUS family through gene duplication. Nevertheless, some of them are found to be essential for eukaryotic cells, suggesting their functional importance. A comprehensive summary of all validated and predicted functions of human PUS enzymes and the relevant phenotypes is provided in Table 5.2.

Table 5.2 | Functions of human PUS enzymes.

PUS	Function	Phenotype	Ref.
PUS1	cy-tRNA pseudouridylation		
	mt-tRNA pseudouridylation	Mutation MLASA	[110,111]
PUSL1	mt-tRNA pseudouridylation	Decreased mitochondrial translation	[309]
PUS3	cy-tRNA pseudouridylation	Mutation Intellectual disability	[112]
PUS7	cy-tRNA pseudouridylation	Mutation Intellectual disability	[113,114]
	mt-tRNA pseudouridylation		
	mTOG pseudouridylation	KO Impairs embryogenesis KD Impacts hematopoietic commitment	[115]
	16S mt-rRNA assembly factor ? (*T. brucei* analog mt-LAF4)		[310]
PUS7L	?		
PUS10	cy-tRNA pseudouridylation	KD Reduces TRAIL-induced apoptosis	[311,312]
TRUB1	cy-tRNA pseudouridylation		
	mt-tRNA pseudouridylation		
	tRNA chaperone ? (*E. coli* analog TruB)		[313]
TRUB2	16S mt-rRNA regulatory module	KO Embryonic lethality Essential for trophoblast stem cells	[314] [315,316]
	16S mt-rRNA processing ?		[317,318]

Color denotes enzymatic (green) and non-enzymatic (purple) functions of PUS enzymes.

Table 5.2 continued | Functions of human PUS enzymes.

PUS	Function	Phenotype	Ref.
RPUSD1	cy-tRNA pseudouridylation		
RPUSD2	cy-tRNA pseudouridylation		
	mt-tRNA pseudouridylation		
RPUSD3	16S mt-rRNA regulatory module		[314]
RPUSD4	16S mt-rRNA regulatory module	KO Lethality	[314]
	16S mt-rRNA pseudouridylation		[319]
	16S mt-rRNA assembly factor ? (*T. brucei* analog mt-LAF3)		[310]
DKC1	cy-rRNAs pseudouridylation	Essential	
		KD Defects in oogenesis	[320]
	Spliceosomal snRNAs pseudouridylation	Essential	
	Box H/ACA snoRNA stabilization		[321]
	TERC stabilization	Mutation X-linked dyskeratosis congenita	[109]

Color denotes enzymatic (green) and non-enzymatic (purple) functions of PUS enzymes.

5.2 Elegant strategy for efficient pseudouridylation

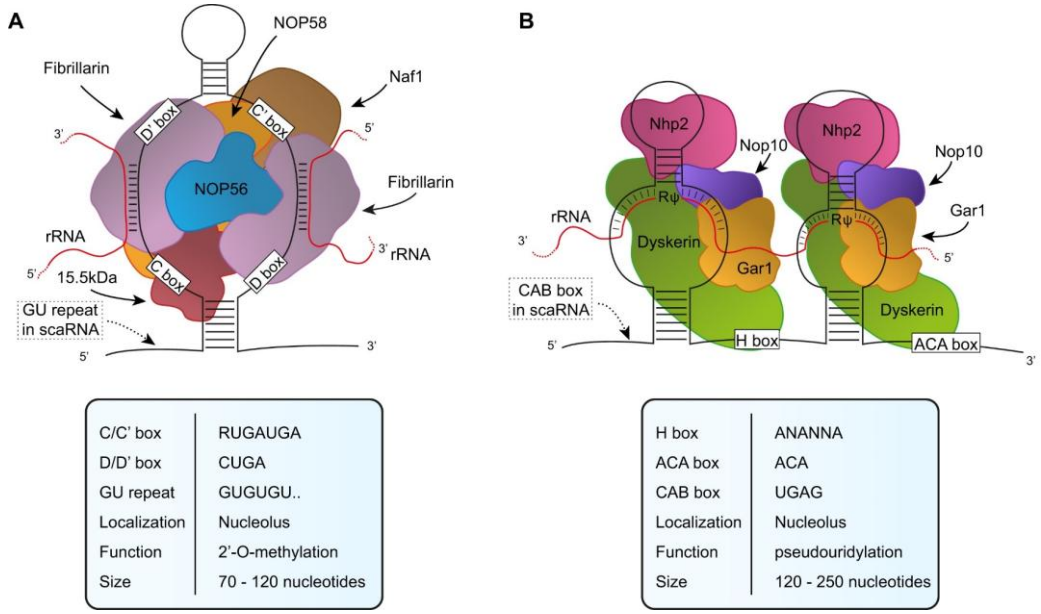


Figure 5.1 | Schematic overview of (a) box C/D snoRNP and (b) box H/ACA snoRNP. Fibrillarin, FBL. Dyskerin, DKC1. Reprinted from [322].

Given the prevalence of Ψ modifications and the high abundance of ncRNAs, it is interesting to understand how this complicated process is achieved in human cells. The first strategy could be well-organized compartmentalization. Previous studies have demonstrated that DKC1, the catalytic subunit of box H/ACA small nucleolar ribonucleoproteins (snoRNPs) and scaRNPs (which possess a subset of snoRNA called scaRNA), is predominantly localized in the nucleolus and the Cajal body, where it co-localizes with FBL, the catalytic subunit of box C/D snoRNPs and scaRNPs [323-325]. These snoRNPs and scaRNPs play crucial roles in the maturation of cy-rRNAs and spliceosomal snRNAs, respectively [326-329]. However, the presence of the nucleolus or Cajal body may not be essential for the formation of Ψ and Nm in these ncRNAs

[330,331], suggesting that these compartments facilitate but do not necessarily determine the modifications. Indeed, larger and more numerous nucleoli and Cajal bodies are often observed in metabolically active cells, such as tumor cells, neurons, and stem cells [332,333], likely due to their increased demand for protein production.

The second strategy is to couple pseudouridylation with other pre-processing machinery. This approach is well known for m^1A/m^1G9 modifications in mt-tRNAs, whose methyltransferase TRMT10C has been identified as a subunit of PRORP (protein-only RNase P, which cleaves the 5'-end of mt-tRNAs) [334-336]. Given the high abundance of Ψ in tRNAs, it is highly likely that PUS enzymes may also be coupled with some pre-processing machinery, like RNase P or RNase Z. Another possibility is that PUS enzymes may serve as chaperones for tRNA folding in addition to their enzymatic activities, a role recently suggested for *E. coli* TruB [313]. Notably, these two approaches may also apply to mt-rRNAs. A recent study on the assembly of *T. brucei* mitoribosome revealed the involvement of RPUSD4-like and PUS7-like proteins in correct folding of mt-rRNAs [310]. The RPUSD4-like protein mt-LAF3 is predicted to function as an active PUS enzyme, while the PUS7-like protein mt-LAF4 has lost its catalytic aspartate residue, which sheds light on the evolution of PUS enzymes — many may have lost or not exhibit their enzymatic activities *in vivo*, instead playing roles similar to other tRNA/rRNA binding proteins. This scenario is suggested for TRUB2, which may form a complex with MTERF3 [318]. Interestingly, MTERF4, a close relative of MTERF3 in the MTERF family, forms a complex with NSUN4, which is bifunctional as both the 12S mt-rRNA m^5C methyltransferase and a ribosome assembly factor [337-340]. NSUN4 is thought to have lost the rRNA binding domain (PUA domain, named after pseudouridine synthase

and archaeosine transglycosylase [341]) present in its ancestor RsmB [342,343], thus requiring binding with MTERF4 to act on rRNA. Similarly, human TRUB2 has also lost its PUA domain compared to *E. coli* TruB [344]. Although there is no direct evidence linking the TRUB2 ancestor TruB to rRNA pseudouridylation and assembly, the phenotypes of TRUB2-knockdown cells strongly suggested its involvement in the early assembly of the mitoribosome large subunit [317]. Indeed, the *in vivo* enzymatic activity of TRUB2 has not been detected yet, further supporting the above hypothesis.

The third strategy is to install Ψ modifications co-transcriptionally, as suggested by a recent study using CMC-based methods [345]. Interestingly, the same study revealed that Ψ is much more abundant in the introns of pre-mRNA, potentially regulating the pre-mRNA splicing. Although these observations need further validation using BACS, this strategy is likely true for Ψ in cy-rRNAs, since earlier research demonstrated that Nm modifications are deposited on nascent cy-rRNA transcripts [346].

5.3 Relationship between Ψ and cancer

As discussed in Section 4.4, we have shown that 5-FU can serve as a global hypopseudouridylating agent, indicating a close relationship between Ψ and cancer. Analysis of public pan-cancer RNA-seq data reveals the global upregulation of PUS enzymes in a wide range of cancer types [347]. Recent studies have increasingly highlighted the emerging role of Ψ in cancer, as reviewed in [108]. Collectively, these findings suggest that the upregulation of Ψ and PUS enzymes is positively corrected with tumorigenesis. In this section, I will discuss the possible underlying mechanisms.

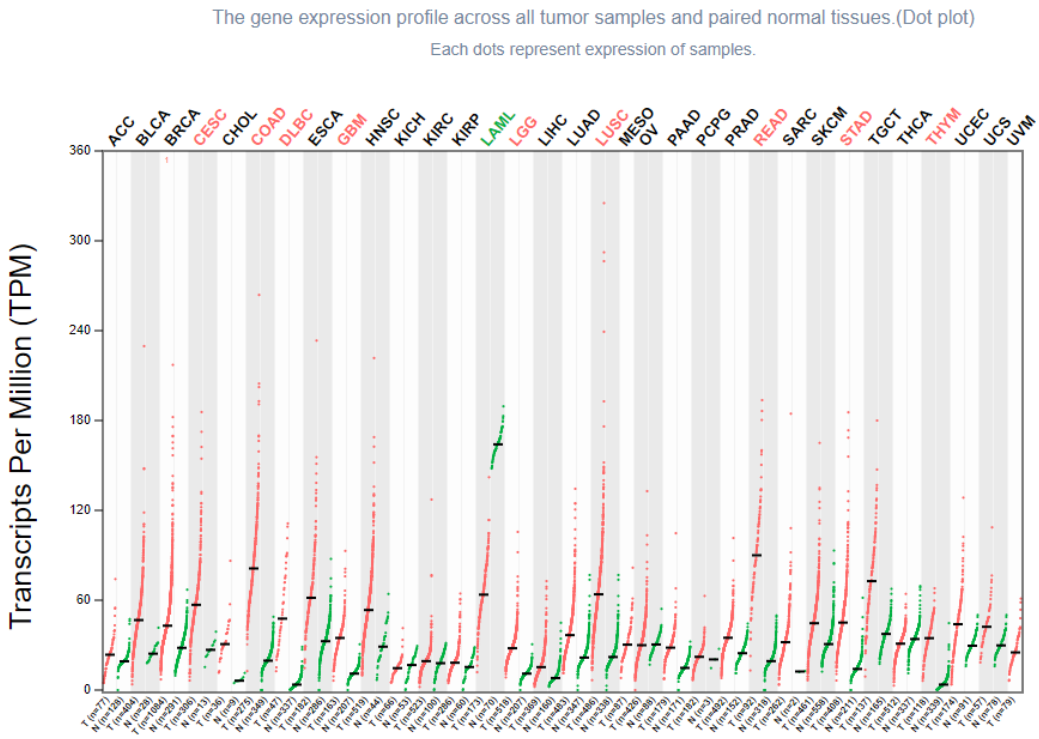


Figure 5.2 | Comparison of DKC1 expression levels between tumor samples (red dots) and normal samples (green dots). Reprinted from [347], web server http://gepia2.cancer-pku.cn/.

DKC1, as the most unique and probably the most important human PUS enzyme, has been linked to cancers in several ways. As mentioned in Section 5.2, cancer cells typically exhibit enlarged nucleoli and Cajal bodies, which are primarily responsible for ribosome and small nuclear ribonucleoproteins (snRNPs)/telomerase biogenesis, respectively [330,331]. Indeed, ribosome, spliceosome, and telomerase functions are often enhanced in tumor cells compared to normal cells [348-350], to meet the requirement of rapid cell division. This phenomenon might be a common mechanism for human cells, since other metabolically active cells, such as oocytes and stem cells, also display similar phenotypes. Interestingly, a recent study revealed the biological importance of box H/ACA snoRNPs during fruit fly oocyte specification, through impacting the ribosome translation of specific mRNAs decoding crucial differentiation factors [320]. Similarly, box H/ACA snoRNPs may be essential or adaptive factors for cancer cell survival by enhancing protein translation. Additionally, DKC1 has been identified as a direct target of c-Myc, one of the most important cancer driver factors [351], which may further explain the widespread abundance of DKC1 in various cancer types. Furthermore, EBV-transformed B cells display significant upregulation of DKC1 expression [352], potentially contributing to the oncogenic nature of EBV. This explanation could also be applied to the other tumor-inducing human herpesviruses, such as Kaposi's sarcoma-associated herpesvirus (KSHV) [353].

Notably, mutations in *DKC1* cause X-linked dyskeratosis congenita [109], a rare genetic disease characterized by increased occurrence of tumors and early mortality due to bone marrow failure, suggesting a dual role of DKC1 in tumorigenesis [354]. Consistent with this finding, AML is one of the few exceptions where DKC1 expression is decreased [347].

These phenotypes can be attributed to the crucial role of DKC1 in maintaining stemness by safeguarding the telomerase activity and ribosome biogenesis. Indeed, dyskeratosis congenita patients typically have short or abnormal telomeres, and genetic mutations of other components in the telomerase or shelterin complexes also lead to dyskeratosis congenita [355]. However, recent studies also point out alterations of ribosomal biogenesis as another underlying mechanism [356].

Based on the above discussions, 5-FU can be considered as a hypopseudouridylating agent with general toxicity to all cells, though it has more significant effects on fast-replicating cells, such as tumor cells, which require higher levels of DKC1. This theory is consistent with the fact that 5-FU has also been used in dermatology to treat warts and in aesthetic medicine to reduce scarring [357,358]. However, a major limitation of 5-FU in cancer treatment is the rapid development of resistance, leading to poor outcomes [286]. This resistance could stem from alterations in PUS expression or from reduced incorporation rate of 5-FU into RNA due to adaptations of nucleoside metabolisms. Therefore, future studies will focus on elucidating the molecular mechanisms of 5-FU resistance to improve its efficacy. In addition, 5-FU treatment often induces severe side effects, ranging from myelosuppression (bone marrow suppression) to hand foot syndrome [359]. 5-FU also needs to be detoxified into dihydro-5-FU by dihydropyrimidine dehydrogenase (DPD) [360]. Patients with partial or complete DPD deficiency due to genetic variations may experience more severe or even fatal side effects [361]. Consequently, developing more selective small-molecule inhibitors or F^5U-lebaled small RNAs to target DKC1 may overcome the limitations of 5-FU treatment and serve as the next-generation of chemotherapy drugs.

Currently, the role of the other 12 PUS enzymes in tumorigenesis is much less understood. It has been suggested that inhibiting these stand-alone PUS enzymes may also partially contribute to the toxicity of 5-FU [291]. Additionally, recent studies have shown that PUS7 is involved in the tumorigenesis of glioblastoma and AML, although the underlying mechanisms remain unclear [116,117]. Further research is needed to uncover the association between these PUS enzymes and cancer.

5.4 Relationship between Ψ and innate immunity

The 2023 Nobel Prize in Physiology or Medicine was awarded to Katalin Karikó and Drew Weissman for their contributions to the development of effective mRNA vaccines using modified nucleobases. Initially, Ψ was used to replace all the uridines in mRNA vaccines, which can suppress innate immune responses while enhancing protein production [277,362]. Although 1-methylpseudouridine ($m^1\Psi$), an archaeal modification, eventually replaced Ψ due to its superior performance [363,364], the relationship between Ψ and innate immunity remains significant. Beyond antiviral applications, mRNA vaccines have also been explored for cancer treatment, demonstrating promising results in Phase II clinical trials [365].

It is noteworthy that utilizing modified nucleic acids is a general strategy to distinguish self from non-self. For example, A-to-I editing predominantly occurs on *Alu* elements that form double-stranded RNA (dsRNA) [366], which loosens the duplex structure and prevents it from being recognized as viral dsRNA [367]. Additionally, DNA with methylated CpG can suppress TLR9 activation [368]. The detailed mechanisms of why Ψ-modified mRNA does not activate toll-like receptors (TLRs) remain unclear. One possibility is that it fine-tunes the mRNA secondary structure, similar to the A-to-I editing scenario.

A recent study on box H/ACA snoRNA and Herpes simplex virus-1 (HSV-1) infection revealed new possibilities for Ψ involvement in innate immunity [369]. Researchers demonstrated that heterozygous mutations in *SNORA31* would result in increased susceptibility to HSV-1 infection of the central neuronal systems, independent of the

TLR3 and IFNAR1/IFNAR2 pathways. Interestingly, *SNORA31* mutations also increase susceptibility to other neurotropic viruses, including varicella zoster virus (VZV), measles virus (MeV), vesicular stomatitis virus (VSV), and encephalomyocarditis virus (EMCV). This observation suggests a novel role for Ψ in modulating innate immunity, although the detailed mechanisms require further investigation.

6 Conclusions

In this thesis, I introduce two novel methods for direct, quantitative, and base-resolution sequencing of DNA and RNA modifications: CAPS+ for 5hmC sequencing and BACS for Ψ sequencing. More importantly, BACS has been applied to elucidate the biological importance of Ψ and PUS enzymes. Future studies will focus on developing single-cell BACS to explore potential cell heterogeneity following 5-FU treatment and applying BACS to circulating cell-free RNAs to investigate the potential of Ψ as a novel biomarker for liquid biopsies. An increasing understanding of Ψ and PUS enzymes could illuminate other ncRNA modifications, potentially transforming the field of 'epitranscriptomics'. The successful development of CAPS+ and BACS underscores the power of chemical tools in biological research. Further advancements in chemistry will likely lead to the creation of more precise sequencing methods for DNA and RNA modifications, which may have profound impacts in both academic research and clinical applications.

7 Materials and methods

7.1 CAPS+

Preparation of model DNA.

Regular and 5mC-labeled DNA oligonucleotides were purchased from IDT or Sangon Biotech. 5hmC-labeled DNA oligonucleotides were purchased from IDT or Takara. 5fC- and 5caC-labeled DNA oligonucleotides were purchased from ATDBio. Sequences of short oligonucleotides can be found in Table 7.1.

Table 7.1 | Short DNA oligonucleotide sequences in this work.

Name	Sequence (5' to 3')	Source
for ESI-MS		
11mer C-ODN	TCGAC**C**GGATC	Sangon Biotech
11mer 5mC-ODN	TCGAC**5mC**GGATC	Sangon Biotech
11mer 5hmC-ODN	TCGAC**5hmC**GGATC	Takara
for MALDI		
11mer C-ODN	TCGAC**C**GGATC	IDT
11mer 5mC-ODN	TCGAC**5mC**GGATC	IDT
11mer 5hmC-ODN	TCGAC**5hmC**GGATC	IDT
11mer 5fC-ODN	TCGAC**5fC**GGATC	ATDBio

The 79 bp double-stranded 5hmC-containing DNA used for UHPLC-MS/MS analysis was prepared by annealing 79mer 5hmC-ODN and 79mer ODN-R.

Table 7.2 | DNA oligonucleotide sequences used for preparation of 79 bp double-stranded 5hmC-containing DNA.

Name	Sequence (5' to 3')	Source
79mer 5hmC-ODN	CCTCACCATCTCAACCAATATT**5hmC**GATATTATGTCTACACGTTGGAGTTCCGTGTTATAATATTGAGGGAGAAGTGGTGA	Takara
79mer ODN-R	TCACCACTTCTCCCTCAATATTATAACACGGAACTCCAACGTGTAGACATAATATCGAATATTGGTTGAGATGGTGAGG	Sangon Biotech

The 13mer and 99 bp double-stranded 5fC-containing DNA used for UHPLC-MS/MS analysis was synthesized by primer extension method.

Table 7.3 | DNA oligonucleotide sequences used for preparation of 13mer double-stranded 5fC-containing DNA.

Name	Sequence (5' to 3')	Source
12mer 5fC-F	GTCGACCGGATC	Sangon Biotech
17mer 5fC-R	TTGGATCCGGTCGACTT	Sangon Biotech

Two DNA oligos (10 mM for each) were annealed in 1× annealing buffer containing 5 mM Tris-HCl (pH 7.5), 5 mM $MgCl_2$, and 50 mM NaCl and then incubated with 200 mM 5-formyl-2'-dCTP (Trilink Biotech) and 5 U Klenow Fragment 3'→5' exo– (New England Biolabs) in 50 µl of 1× NEBuffer 2 at 37 °C for 2 h. The product was purified with Micro Bio-Spin P-6 SSC column (Bio-Rad).

Final product sequence (5' to 3'):

GTCGACCGGATC**5fC**

Table 7.4 | DNA oligonucleotide sequences used for preparation of 99 bp double-stranded 5fC-containing DNA.

Name	Sequence (5' to 3')	Source
99mer 5fC template	CTCACTCACCTCCACCCTCTCACTA CCTCACTCTTCCTCCTAACCCTCTCC AACCACCTCTCCACCCTCCTAGATC TCTACCTGACTGAGCGTGTGCGA	Sangon Biotech
5fC primer	TCGCACACGCTCAGTCAGGT	Sangon Biotech

1.5 mg 5fC template were added to a solution containing 10 mM primer, 150 mM 5-formyl-2'-dCTP (Trilink Biotech), dGTP, dATP, dTTP (New England Biolabs), 2.5 U DreamTaq polymerase (Thermo Fisher), and 1× DreamTaq Buffer (Thermo Fisher). The solution was then incubated at 95 °C, 30 s; 68 °C, 60 s; 72 °C, 10 min and purified by Zymo-IC column with DNA Binding Buffer.

Final product sequence (5' to 3'):

TCGCACACGCTCAGTCAGGTAGAGAT**5fC**TAGGAGGGTGGAGAGGTGGTTGG AGAGGGTTAGGAGGAAGAGTGAGGTAGTGAGAGGGTGGAGGTGAGTGAG

Preparation of spike-in controls and carrier DNA for CAPS+.

Detailed preparation protocols and sequences of CpG-methylated lambda DNA, 2 kb unmodified spike-in control and carrier DNA can be found in previous publications [180,183]. Carrier DNA was fragmented by Covaris M220 and size-selected to 150–250 bp with 0.9–1.2× Ampure XP beads according to the manufacturer's protocol. 144mer synthetic 5hmC spike-in control was synthesized by primer extension method.

Table 7.5 | DNA oligonucleotide sequences used for preparation of 144mer synthetic 5hmC spike-in control.

Name	Sequence (5' to 3')	Source
74mer 5hmC-F	AGAGAGCAAGCCGGCTATAGATGCT ACGTACAGTAGCAGCTGATCAAGAC TGCTAAGGCCACAACCAGTTGGCG	IDT
90mer 5hmC-R	ACTCTCACTCTCACCTCCATCTTACT TGTCTACCGAATCCTCACGTACTCA CGATTCGACCATCCACATTCGCCAA CUGGUTGUGGCCTT	IDT

Two DNA oligos (10 mM for each) were annealed in 1× TruSeq annealing buffer containing 10 mM Tris-HCl (pH 7.9), 0.1 mM EDTA, and 50 mM NaCl and then incubated with 400 mM 5-hydroxymethyl-2'-dCTP (Zymo Research), dGTP, dATP, dTTP (New England Biolabs) and 5 U Klenow Fragment 3'→5' exo– (New England Biolabs) in 50 µl of 1× NEBuffer 2 at 37 °C for 1 h. The product was purified by Zymo-IC column with Oligo Binding Buffer.

Final spike-in sequence (5' to 3'):

AGAGAGCAAGCCGGCTATAGATGCTACGTACAGTAGCAGCTGATCAAGACTG CTAAGGCCACAACCAGTTGGCGAATGTGGATGGT**5hmC**GAAT**5hmC**GTGAGT A**5hmC**GTGAGGATT**5hmC**GGTAGA**5hmC**AAGTAAGATGGAGGTGAGAGTGA GAGT

Mass spectrometry analysis of short oligonucleotides.

ESI-MS was performed on a Vion IMS QTof mass spectrometer (Waters). The source-dependent parameters were as follows: capillary voltage 2.5 kV, source temperature

100 °C, desolvation temperature 250 °C, cone gas flow 50 l h^{-1}, desolvation gas flow 800 l h^{-1}. All oligonucleotides were analyzed in negative mode.

MALDI was performed on a Voyager-DE Biospectrometry Workstation (Applied Biosystems) with 2',4',6'-trihydroxyacetophenone (THAP) as matrix. All oligonucleotides were analyzed in positive mode.

Quantification of 5hmC, 5fC, and 5caC level in model DNA by UHPLC-MS/MS.

Table 7.6 | Compound-dependent UHPLC-MS/MS parameters used for nucleoside quantification in model DNA. All nucleosides were analyzed in positive mode.

Compound	Precursor Ion (m/z)	Product Ion (m/z)	RT (min)	Delta RT (min)	Dwell Time (ms)	Q1 Pre Bias (V)	CE (V)	Q3 Pre Bias (V)
cadC+H	272	156	0.76	1.5	100.0	–30.0	–12.0	–27.0
cadC+Na	294	178	0.76	1.5	100.0	–19.0	–13.0	–11.0
dC+H	228	112	2.03	2.4	100.0	–27.0	–14.0	–18.0
dC+Na	250	134	2.03	2.4	100.0	–16.0	–13.0	–12.0
hmdC+H	258	142	2.35	2.4	100.0	–17.0	–11.0	–13.0
hmdC+Na	280	164	2.35	2.4	100.0	–19.0	–15.0	–16.0
mdC+H	242	126	4.70	1.2	100.0	–25.0	–13.0	–12.0
mdC+Na	264	148	4.70	1.2	100.0	–17.0	–14.0	–14.0
dG+H	268	152	5.10	1.2	100.0	–18.0	–11.0	–15.0
fdC+H	256	140	5.60	1.8	100.0	–26.0	–11.0	–12.0
fdC+Na	278	162	5.60	1.8	100.0	–18.0	–12.0	–10.0
dT+H	243	127	5.88	1.6	100.0	–19.0	–11.0	–12.0
dT+Na	265	149	5.88	1.6	100.0	–29.0	–12.0	–24.0
dA+H	252	136	6.37	1.6	100.0	–28.0	–15.0	–13.0

RT: retention time; CE: collision energy.

The oxidized model DNA was digested into nucleosides by Nucleoside Digestion Mix (New England Biolabs) in 20 µl solution according to the manufacturer's protocol. The mixture was diluted to 100 µl and centrifuged at 12,000× g and 4 °C for 20 min. Finally, 2–3 µl of the supernatant was injected into UHPLC-MS/MS.

The nucleosides were separated by ultra-high performance liquid chromatography on an ACQUITY UPLC BEH C18 column (2.1×100 mm, 1.7 µm, Waters) and detected by LCMS-8050 triple quadrupole mass spectrometer (Shimadzu) in the positive ion multiple reaction monitoring (MRM) mode. The column temperature was maintained at 30 °C, and the solvent system was water containing 10 mM ammonium acetate (pH 7.0, solvent A) and methanol (solvent B) with 0.4 ml min^{-1} flow rate. The gradient was: 0–1.50 min, 0% solvent B; 1.50–4.50 min, 0–10% solvent B; 4.50–5.00 min, 10–30% solvent B; 5.00–5.01 min, 30–100% solvent B; 5.01–7.00 min, 100% solvent B; 7.00–7.01 min, 100–0% solvent B; 7.01–9.50 min, 0% solvent B. The source-dependent parameters were as follows: nebulizing gas flow 3 l min^{-1}, heating gas flow 10 l min^{-1}, interface temperature 300 °C, desolvation line temperature 250 °C, heat block temperature 400 °C, drying gas flow 10 l min^{-1}. The ions were monitored in positive mode with mass transitions of m/z 258 to 142 (hmdC+H), m/z 278 to 162 (fdC+Na), and m/z 294 to 178 (cadC+Na) (Table 7.6). Concentrations of nucleosides in DNA samples were deduced by fitting the signal peak areas into the standard curves.

Quantification of 5hmC, 5fC, 5caC, and 8-oxoG level in gDNA by UHPLC-MS/MS.

Table 7.7 | Compound-dependent UHPLC-MS/MS parameters used for nucleoside quantification in gDNA. All nucleosides were analyzed in positive mode.

Compound	Precursor Ion (m/z)	Product Ion (m/z)	RT (min)	Delta RT (min)	CE (V)
cadC+H	272	156	1.5	1.0	12.0
cadC+Na	294	178	1.5	1.0	12.0
dC+H	228	112	3.8	1.0	10.0
dC+Na	250	134	3.8	1.0	10.0
hmdC+H	258	142	4.5	1.0	12.0
hmdC+Na	280	164	4.5	1.0	12.0
mdC+H	242	126	9.1	1.0	10.0
mdC+Na	264	148	9.1	1.0	10.0
dG+H	268	152	9.7	1.0	10.0
fdC+H	256	140	10.8	1.0	12.0
fdC+Na	278	162	10.8	1.0	12.0
dT+H	243	127	11.1	1.0	10.0
dT+Na	265	149	11.1	1.0	10.0
8-oxodG+H	284	168	12.2	1.0	10.0
dA+H	252	136	14.6	1.0	10.0

RT: retention time; CE: collision energy.

K_2RuO_4 oxidation was performed according to previous publications [184,187]. The oxidized gDNA was digested into nucleosides by Nucleoside Digestion Mix (New England Biolabs) in 20 µl solution according to the manufacturer's protocol. After filtering with Amicon Ultra-0.5 mL 3K centrifugal filters (Millipore), the digested samples were subjected to UHPLC–MS/MS analysis as described before [370]. 1290 Infinity LC Systems (Agilent) was equipped with a ZORBAX RRHD SB-C18 column

(2.1 × 150 mm, 1.8 μm, Agilent) coupled with a 6495B Triple Quadrupole Mass Spectrometer (Agilent). The ions were monitored in positive mode with mass transitions of m/z 258 to 142 (hmdC+H), m/z 278 to 162 (fdC+Na), m/z 294 to 178 (cadC+Na), and m/z 284 to 168 (8-oxodG+H) (Table 7.7). Concentrations of nucleosides in DNA samples were deduced by fitting the signal peak areas into the standard curves.

mESCs culture and isolation of genomic DNA.

E14 mESCs (gift from S. Kriaucionis) were cultured on gelatin-coated plates in DMEM (Invitrogen) supplemented with 15% FBS (Gibco), 2 mM L-glutamine (Gibco), 1% nonessential amino acids (Gibco), 1% penicillin/streptavidin (Gibco), 0.1 mM β-mercaptoethanol (Sigma-Aldrich), 1,000 units ml^{-1} leukemia inhibitory factor (Millipore), 1 μM PD0325901 (Stemgent), and 3 μM CHIR99021 (Stemgent). Cultures were maintained at 37 °C and 5% CO_2 and passaged every 2 d. For isolation of genomic DNA, cells were harvested by centrifugation for 5 min at 1,000× g and room temperature. DNA was extracted with Quick-DNA Plus kit (Zymo Research) according to the manufacturer's protocol.

Human normal brain and glioblastoma samples.

The gDNA of human normal brain and glioblastoma samples were obtained from Amsbio. Detailed clinical information for each sample is provided in Table 7.8.

Table 7.8 | CAPS+ sample information.

Sample	Species	Cell-type (s)	Donor sex	Donor Age	Donor race	Donor pathology
mESC_rep1	mouse	embryonic stem cells				
mESC_rep2	mouse	embryonic stem cells				
human normal brain	human	tissue	male	77	Caucasian	normal
glioblastoma	human	tissue	female	60	Asian	glioblastoma

ACT^+ BF_4^- oxidation of 5hmC to 5fC.

100 ng ligated DNA was incubated in 25 µl solution containing 50 mM sodium phosphate buffer (pH 7.5) and 50 mM ACT^+ BF_4^- (TCI) at 37 °C for 4 h. The oxidized DNA was purified with 1.8× AMPure XP Beads.

5fC labeling by hydroxylamine.

14 µl oxidized 11mer 5hmC-containing DNA was added to a 20 µl reaction containing 0.1 M MES buffer (pH 5.0), 10 mM *O*-ethylhydroxylamine (Sigma-Aldrich) and incubated at 37 °C for 2 h. The product was purified using Micro Bio-Spin P-6 SSC column (Bio-Rad, washed four times with water before use).

Pinnick oxidation of 5fC to 5caC.

5 M 2-methyl-2-butene (Acros) was prepared by mixing the pure compound (~9.4 M) with 100% ethanol. The converted DNA was added to a 30 µl reaction containing 0.2 M sodium acetate buffer (pH 4.3), 0.16 M $NaClO_2$ (Alfa Aesar), and 1 M 2-methyl-2-butene. The reaction was then incubated at 25 °C for 16 h and double purified with Micro Bio-

Spin P-6 SSC column (Bio-Rad, washed four times with water before use) and 1.8× AMPure XP Beads.

CAPS+ library construction and sequencing.

mESC, human normal brain and glioblastoma gDNA was spiked with 0.5% of CpG-methylated lambda DNA and 0.25% of unmodified 2 kb spike-in control. DNA samples were fragmented by Covaris M220 and size-selected to 300–500 bp with 0.55–0.85× Ampure XP beads according to the manufacturer's protocol. The sonicated DNA was additionally spiked with 0.05% of 144mer spike-in after size-selection with Ampure XP beads. End-repair and A-tailing reaction and ligation of NEBNext Adaptor for Illumina were prepared with KAPA HyperPrep Kit according to the manufacturer's protocol. The uracil in the loop of NEBNext Adaptor was removed by adding 6 μl of USER Enzyme (New England Biolabs) to the ligation reaction and incubating at 37 °C for 45 min. The reaction was purified with 0.8× Ampure XP beads by washing twice with 80% acetonitrile/water (v/v). ACT^{+} BF_4^{-} oxidation and Pinnick oxidation were performed as described above. Borane reduction was then performed in a 50 μl reaction using optimized condition at 37 °C and 850 r.p.m. for 6 h and purified by Zymo-IC column with Oligo Binding Buffer [184]. Converted DNA was amplified with NEBNext Multiplex Oligos for Illumina (Dual Index Primers Set 1) and KAPA HiFi HotStart Uracil+ ReadyMix PCR Kit for 4 cycles according to the manufacturer's protocols. The PCR product was purified with 0.9× Ampure XP beads and quantified with Qubit dsDNA HS Assay Kit (Thermo Fisher) according to the manufacturer's protocol. Libraries were sequenced on NextSeq 2000 (150 bp paired end) with no PhiX added.

Data pre-processing.

Raw sequencing reads were processed as previously reported [184]. Briefly, fastp (v.0.12.4) [371] was used to filter bad reads, trim low-quality bases and cut adaptors. Trimmed reads were then aligned to either mm9 (for mESC samples) or hg38 (for human normal brain and glioblastoma samples) reference genome using BWA mem (v.0.7.17) [372]. PCR duplicates were removed by MarkDuplicate in Picard tools (v.2.23.0) (http://broadinstitute.github.io/picard/). Methylation calling was performed by asTair (v3.3.2) [180] using reads with MAPQ > 10. Regions prone to cause mapping artefacts were excluded from subsequent analysis (https://sites.google.com/site/anshulkundaje/projects/blacklists) [373,374]. For mESC samples, known single nucleotide variants in E14 cell lines (http://epigenetics.hugef-research.org/data.php) [375] were also excluded. For human normal brain and glioblastoma samples, CpG sites overlapping with common single nucleotide polymorphisms (SNPs) (dbSNP153) and centromeres [376] were further excluded from downstream analysis.

Published data.

Related published data we used were downloaded from the Gene Expression Omnibus (GEO) database: CAPS (GSE155613) [184]; TAB-seq for mESCs (GSE36173) [173]; ACE-seq (GSE116016) [209]; TAB-seq for adult prefrontal cortex (GSE46710) [210]. TAB-seq data for mESCs were reprocessed as previously described to obtain modified and unmodified CpG sites [173]. Blacklisted regions and common single nucleotide variants (SNVs) were also excluded. 5hmCG methylation calls in adult prefrontal cortex

were downloaded from GSE46710 and converted from hg19 to hg38 using CrossMap.py [377].

Pairwise comparison of CAPS+.

To calculate correlation between two mESC replicates, 5hmCG raw signals (T/C+T) were calculated in 10 kb genomic bins. The Pearson correlation coefficient (Pearson's *r*) was calculated by cor function in R. The results were visualized by SmoothScatter in R. Pairwise comparisons between CAPS+ and CAPS, TAB-seq as well as ACE-seq were performed similarly.

Coverage analysis for CAPS+, CAPS, and ACE-seq.

CpG islands were divided into 10 equal-sized bins, while the 4 kb flanking regions were binned into 20 windows. Average per strand CpG coverage was calculated as the sum of modified and unmodified cytosines. Reads from two replicates of CAPS+ were merged and subsampled to the similar coverage of ACE-seq using samtools view with the parameter (–s 0.60).

Statistical testing of high confidence 5hmC sites and genomic annotations.

As previously described, a binomial test was used to estimate high confidence 5hmCG sites [184]. A *p*-value for each site was calculated by comparing the average false-positive rate measured by 2 kb unmodified spike-in between replicates. 5hmCG sites whose per-strand coverage was lower than 5 were filtered. 5hmCG sites with Benjamini-Hochberg (BH) adjusted *p*-value less than 0.01 were identified to be high confidence. For annotations of high confidence 5hmCG sites, genomic features of mouse genome were

downloaded (https://github.com/gireeshkbogu/chromatin_states_chromHMM_mm9) [378]. High confidence 5hmCG sites in each regulatory genomic region were found by bedtools intersect [379] and the counts in each category were normalized by genomic coverages of corresponding category regions. To investigate the enrichment of 5hmCG sites, all detected CpG sites were sampled 10 times to generate the background distribution in different genomic elements.

5hmCG profiles in gene bodies.

Annotated genes were downloaded from RefSeq database (https://www.ncbi.nlm.nih.gov/refseq/MANE/). 5hmCG scores within gene bodies, 5 kb upstream of TSS and 5 kb downstream regions of transcription termination sites (TTS) sites were calculated by computeMatrix in deepTools (v.3.3.1) [380] with parameters (--beforeStartLength 5000 --regionBodyLength 5000 --afterRegionStartLength 5000 --binSize 10). 5 kb flanking regions were binned into 10 bp windows. The gene body regions were fitted into 5 kb and divided into 500 bins. 5hmC scores were calculated with CpG sites covered by at least 5 reads per strand.

Genomic view.

Methylation calls in browser extensible data (BED) format were converted to bigwig (BW) format by bedGraphToBigWig [381]. 5hmCG signals in regions of interest were visualized by Integrative Genomics Viewer (IGV) [382].

Comparison of 5hmCG signals between human normal brain and glioblastoma samples.

5hmCG raw signals (T/T+C) in both normal brain and glioblastoma samples were calculated in 1 kb genomic bins.

7.2 BACS

Preparation of model RNA.

Table 7.9 | RNA oligonucleotide sequences in this work.

Name	Sequence (5' to 3')	Source
for MALDI		
10mer U-ORN	UACUG**U**AGCU	IDT
10mer Ψ-ORN	UACUG**Ψ**AGCU	IDT
for mutation analysis		
72mer Ψ-ORN	GGGAGAACACACCACAACGAAA CCAACGG**Ψ**ACAACAACAGAAA**Ψ** CGAGGACCGAAGCGAAGGCAAA GACAAC	*in vitro* transcription Pseudo-UTP, ATP, CTP, GTP
for UHPLC-MS/MS		
1.8-kb 10% Ψ-modified RNA	T7 *in vitro* transcription using linearized Fluc plasmid (NEB) as template	*in vitro* transcription 10% Pseudo-UTP, 90% UTP, ATP, CTP, GTP
spike-ins		
30mer NNUNN	AUGUCUCGACGU**NN**U**NN**GUUAC AGUACCGU	IDT
30mer NNΨNN	GCUUCAAGUUGA**NN**Ψ**NN**CAUCG CAAGUGCA	IDT

Regular and Ψ-labeled 10mer RNA oligonucleotides and 30mer spike-ins were purchased from Integrated DNA Technologies (IDT). The 72mer Ψ-containing model RNA used for mutation analysis and the 1.8-kb 10% Ψ-modified RNA used for UHPLC-MS/MS were prepared by T7 *in vitro* transcription using HiScribe T7 High Yield RNA Synthesis Kit (NEB) and Pseudo-UTP (Jena Bioscience) according to the manufacturer's protocol. The DNA template was removed by adding 2 µl Turbo DNase (Invitrogen) to the reaction and

incubating at 37 °C for 30 min. The products were finally purified with Monarch RNA Cleanup Kit (NEB). Sequences of RNA oligonucleotides can be found in Table 7.9.

Mass spectrometry analysis of short oligonucleotides.

MALDI was performed on a Voyager-DE Biospectrometry Workstation (Applied Biosystems) with 2',4',6'-trihydroxyacetophenone (THAP) as matrix. All oligonucleotides were analyzed in positive mode.

Quantification of Ψ level by UHPLC-MS/MS.

Table 7.10 | Compound-dependent UHPLC-MS/MS parameters used for nucleoside quantification. All nucleosides were analyzed in positive mode.

Compound	Precursor Ion (m/z)	Product Ion (m/z)	RT (min)	Delta RT (min)	CE (V)
Ψ+H	245	125	1.8	2.0	10.0
rC+H	244	112	2.3	2.0	10.0
rC+Na	266	134	2.3	2.0	10.0
rU+H	245	113	3.3	2.0	10.0
rG+H	284	152	8.1	2.0	10.0
rA+H	268	136	12.9	2.0	8.0

RT: retention time; CE: collision energy.

The untreated and treated RNA were digested into nucleosides by Nucleoside Digestion Mix (NEB) in a 50 μl solution according to the manufacturer's protocol. After filtering with Amicon Ultra-0.5 mL 3K centrifugal filters (Millipore), the digested samples were subjected to UHPLC-MS/MS analysis as described before [370]. 1290 Infinity LC Systems (Agilent) was equipped with a ZORBAX RRHD SB-C18 column (2.1 × 150 mm,

1.8 μm, Agilent) coupled with a 6495B Triple Quadrupole Mass Spectrometer (Agilent). The ions were monitored in positive mode with mass transitions of m/z 245 to 125 (Ψ+H) and m/z 245 to 113 (rU+H) (Table 7.10). Concentrations of nucleosides in RNA samples were deduced by fitting the signal peak areas into the standard curves.

Cell culture.

HeLa cells (gifted from Prof. Peter J. Ratcliffe, University of Oxford; originally obtained from ATCC, #CCL-2) were cultured in DMEM medium (Gibco) supplemented with 10% (v/v) FBS (Gibco) and 1% penicillin/streptomycin (Gibco) at 37 °C with 5% CO_2. For isolation of RNA, cells were harvested by centrifugation for 5 min at 1,000× g and room temperature.

C666-1, Raji, and Elijah cells were grown in Roswell Park Memorial Institute 1640 medium (RPMI 1640) (Thermo), complemented with 10% (v/v) fetal bovine serum (FBS) (Biosera), 2 mM L-glutamine (Thermo), and 100 units/ml penicillin and 100 μg/ml streptomycin (Thermo). Cells were grown in humidified incubator at 37 °C with 5% CO_2. C666-1 cell line was gifted from Dr Christopher Dawson (University of Warwick). Raji and Elijah cell lines were gifted from Prof. Paul Farrell (Imperial College London). Cell lines were monthly tested for mycoplasma using MycoAlert Kit (Lonza) and were sent for authentication by Eurofins genomics.

Generation of CRISPR knockout cell lines.

TRUB1-KO, PUS7-KO, and PUS1-KO HeLa cells were generated using CRISPR-Cas9 technology. Briefly, single guide RNA (sgRNA) sequences were cloned into PX459

plasmids [383]. Transfection was performed using Lipofectamine 3000 Transfection Reagent (Invitrogen) following the manufacturer's protocol. Cells were then selected by 2 μg/ml puromycin (Thermo). Serial dilution was performed to achieve clonal isolation. Finally, clones were expanded and picked for western blot validation with TRUB1 antibody (Proteintech, #12520-1-AP) (1:1000 dilution), PUS7 antibody (Abcam, #ab226257) (1:10000 dilution), and PUS1 antibody (Proteintech, #11512-1-AP) (1:1000 dilution), respectively. The sgRNA sequences were listed as follows:

TRUB1: 5'-CACGGCGAACACGCCGCTCAAGG-3';

PUS7-sgRNA1: 5'-TTAATATTGAAACCCCGCTCTGG-3';

PUS7-sgRNA2: 5'-TCGGAATGCAGTCTAACCAAAGG-3';

PUS1: 5'-AATACAGCCTGACCGGACGAGGG-3'.

RNA isolation.

Total RNA was isolated using TRIzol (Invitrogen) and Direct-zol RNA Miniprep Plus (Zymo Research) according to the manufacturer's protocol. Ribo$^-$ RNA was isolated using RiboMinus Eukaryote System v2 (Invitrogen) according to the manufacturer's protocol. PolyA$^+$ RNA was isolated by two rounds of polyA-tailed selection using Dynabeads mRNA DIRECT Purification Kit (Invitrogen) according to the manufacturer's protocol. To remove genomic DNA contamination, RNA was then treated with Turbo DNase and purified by Zymo-IC Column with RNA binding buffer.

Viral infection and RNA isolation.

SARS-CoV-2.

Viral stocks were propagated as previously reported [384]. Briefly, Vero-TMPRSS2 cells (NIBSC, #100978) were infected with SARS-CoV-2 Victoria 02/20 strain at a multiplicity of infection (MOI) of 0.003 and incubated for 48–72 h until visible cytopathic effect was observed. Viral titers were then determined by plaque assay from clarified supernatants. For sequencing, Calu-3 cells (gifted from Prof. Nicole Zitzmann, University of Oxford; originally obtained from ATCC, #HTB-55) were infected at an MOI of 1 for 1 h at 37 °C. The inoculum was then removed, cells washed thrice with PBS and incubated in Advanced DMEM with 10% FCS at 37 °C for 24 h. Total RNA was extracted using the RNeasy Mini Kit (Qiagen) and infection confirmed plaque assay and RT-qPCR using primers for the viral N gene: forward 5'-CACATTGGCACCCGCAATC-3', reverse 5'-GAGGAACGAGAAGAGGCTTG-3'.

HCV and ZIKV.

As previously reported [385], ZIKV MP1751 strain was propagated in Vero cells (ATCC, #CCL-81), and concentrated with 8% PEG in NTE buffer. For infection, Huh7.5 cells (gifted from Prof. Charles Rice, Rockefeller University) were inoculated with either HCV or ZIKV (MOI 1) for 180 min, before extensive washing with PBS. Media was replaced and infected cells were cultured for 72 h before lysing in RLT buffer. RNAs were extracted using RNeasy Mini Kit (Qiagen) and infection confirmed using RT-qPCR quantification of viral RNAs using specific primers pairs. HCV: forward 5'-TCCCGGGAGAGCCATAGTG-3', reverse 5'-TCCAAGAAAGGACCCAGTC-3';

ZIKV: forward 5'-TCGTTGCCCAACACAAG-3', reverse 5'-CCACTAATGTTCTTTTGCAGACAT-3'; and RPLP0: forward 5'-GCAATGTTGCCAGTGTCTG-3', reverse 5'-GCCTTGACCTTTTCAGCAA-3'.

HDV.

HDV inoculum was prepared by concentrating the culture supernatant of Huh-7 cells (gifted from Prof. Arvind Patel, University of Glasgow) transfected with pSVL(D3) and pT7HB2.7 plasmids as previously reported [386]. HepG2-NTCP cells (gifted from Prof. Stefan Urban, University of Heidelberg) were differentiated with 2.5% DMSO-containing culture medium for 72 h prior to viral infection, before inoculation with HDV (MOI 50) in the presence of 4% PEG8000 and 2.5% DMSO for 24 h. After 24 h, cells were washed with PBS and cultured for an additional 5 days in the presence of 2.5% DMSO. Cells were lysed in RLT buffer, and total cellular RNA was purified using RNeasy Mini Kit (Qiagen) and RNase-free DNase Set (Qiagen). Infection was confirmed by RT-qPCR using specific primers to detect HDV transcripts.

SINV.

SINV was produced from pT7-SVwt plasmid [387] that was first linearized with XhoI and purified to use it as a template for *in vitro* RNA transcription with HiScribe T7 ARCA mRNA kit (NEB). Transcribed viral RNA was transfected into BHK-21 cells (ATCC, #CCL-10) using Lipofectamine 3000 reagent (Invitrogen) according to manufacturer's instruction. Viruses were collected from the supernatant 24 h later and cleared by centrifugation at 2,000 r.p.m for 5 min followed by filtration with 0.45 μm PVDF syringe

filter units (Merck) and stored at –80 °C. Cleared supernatants were titrated by plaque assay.

PolyA$^+$ RNA purification was performed based on the previously described protocols [388,389], with the following alterations: A549 cells (ATCC, #CCL-185) were seeded in two 10 cm dishes in DMEM 10% FBS 24 h prior to infection to reach 80% confluence. Next, cells were either mock-infected or infected using 0.1 MOI of SINV for 1 h in serum-free DMEM at 37 °C, followed by the replacement of the medium with DMEM supplemented with 2% FBS and incubated for 18 h. Cells were lysed with 1 ml of lysis buffer (20 mM Tris-HCl pH 7.5, 500 mM LiCl, 0.5% (w/v) LiDS, 1 mM EDTA, 0.1% IGEPAL (NP-40) and 5 mM DTT). Lysates were homogenized by passing the lysate at high speed through a 5 ml syringe with a 27G needle, repeating this process until the lysate was fully homogeneous. Then, the whole lysate was incubated with pre-equilibrated oligo(dT)$_{25}$ magnetic beads (NEB) for 1 h at 4 °C with gentle rotation. Beads were collected in the magnet and washed twice with 2 ml of buffer 1 (20 mM Tris-HCl pH 7.5, 500 mM LiCl, 0.1% (w/v) LiDS, 1 mM EDTA, 0.1% IGEPAL and 5 mM DTT) for 5 min at 4 °C with gentle rotation, followed by two washes with buffer 2 (20 mM Tris-HCl pH 7.5, 500 mM LiCl, 1 mM EDTA, 0.01% IGEPAL and 5 mM DTT). Beads were then washed twice with 2 ml of buffer 3 (20 mM Tris-HCl pH 7.5, 200 mM LiCl, 1 mM EDTA and 5 mM DTT) at room temperature. Finally, beads were resuspended in 50 µl of elution buffer and incubated for 7 min at 55 °C with agitation. Eluates were stored at –80 °C. Approximately 5 µg of polyA$^+$ RNAs were used for further rRNA removal using the Ribo-Zero kit from the TruSeq Stranded Total RNA LT Kit (Illumina). Subsequently, RNAclean XP (Beckman Coulter) purification was conducted, and RNA were finally

eluted into 10–15 µl of nuclease-free water. The concentration and quality of the RNA were assessed using Qubit and RNA Bioanalyzer.

Preparation of IVT RNA control.

100 ng polyA$^+$ RNA was annealed with 2 µl of 10 µM Oligo(dT)$_{30}$VN primer (5'-TTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN-3') and 2 µl of 10 mM dNTP mix (NEB) in 12 µl solution, incubated at 70 °C for 5 min and held at 4 °C. Next, 5 µl 4× Template Switching RT buffer (NEB), 1 µl of 75 µM T7-TSO (5'-/5Biosg/ACTCTAATACGACTCACTATAGGGAGAGGGCrGrGrG-3'), and 2 µl 10× Template Switching RT Enzyme Mix (NEB) were added to the mixture and the reaction was incubated at 42 °C for 90 min followed by 85 °C for 5 min. The second strand synthesis was then performed by adding 100 µl Q5 Hot Start High Fidelity Master Mix (NEB), 10 µl RNase H (NEB), and 70 µl nuclease-free H_2O to the cDNA mixture and incubation at 37 °C for 15 min, 95 °C for 1 min, and 65 °C for 10 min. The double-stranded cDNA product was purified with 0.8× AMPure XP beads (Beckman Coulter). The IVT RNA control was prepared by T7 *in vitro* transcription using the purified cDNA product and HiScribe T7 High Yield RNA Synthesis Kit according to the manufacturer's protocol. To remove cDNA template, IVT RNA was then treated with Turbo DNase and purified by Zymo-IC Column with RNA binding buffer.

BACS for Ψ detection.

50–200 ng ribo$^-$ or polyA$^+$ RNA was fragmented by NEBNext Magnesium RNA Fragmentation Module at 94 °C for 4 min according to the manufacturer's protocol and purified by Zymo-IC Column with RNA binding buffer. The fragmented RNA was mixed

with 5 µl 10× T4 PNK reaction buffer (NEB), 5 µl T4 PNK (NEB), and 2.5 µl SUPERase•In RNase Inhibitor (Invitrogen) in a 50 µl final solution and incubated at 37 °C for 1 h. The 3'-repaired RNA was purified by Zymo-IC Column with RNA binding buffer and eluted with 10 µl nuclease-free H_2O. The eluted RNA was then mixed with 1 µl synthetic 30mer spike-ins (2%) and 1 µl of 20 µM RNA adapter (5'-/5rApp/AGATCGGAAGAGCGTCGTG/3SpC3/-3'), incubated at 70 °C for 2 min and immediately placed on ice. Next, 2.5 µl 10× T4 RNA Ligase reaction buffer (NEB), 1 µl SUPERase•In RNase Inhibitor, 7.5 µl 50% PEG 8000 (NEB), and 2 µl T4 RNA Ligase 2, truncated KQ (NEB) were added to the mixture and the reaction was incubated at 25 °C for 2 h followed by 16 °C for 14 h. To digest excess adapters, the solution was further diluted to 47 µl with nuclease-free H_2O and treated with 2 µl 5'-Deadenylase (NEB) at 30 °C for 1 h followed by adding 1 µl $RecJ_f$ (NEB) and incubation at 37 °C for 1 h. The 3'-ligated RNA was purified by Zymo-IC Column with RNA binding buffer and eluted with 12 µl nuclease-free H_2O. A 10 µl aliquot was subjected to BACS library construction, while the remaining 2 µl was saved as control sample and diluted to 12.5 µl with nuclease-free H_2O. For BACS, 1 M 2-bromoacrylamide (Enamine) was prepared by dissolving the solid in DMSO. 10 µl 3'-ligated RNA was added into a 20 µl solution containing 250 mM 2-bromoacrylamide and 625 mM phosphate buffer (pH 8.5) and incubated at 85 °C for 30 min. The treated RNA was double purified by Micro Bio-Spin P-6 Tris Column (Bio-Rad) and Zymo-IC Column with RNA binding buffer and finally eluted with 12.5 µl nuclease-free H_2O.

Both treated and control RNA samples were mixed with 1 µl of 2 µM RT primer (5'-ACACGACGCTCTTCCGATCT-3') and 1 µl of 10 mM dNTP mix, incubated at 70 °C

for 2 min and immediately placed on ice. Next, 4 µl 5× Maxima H^- RT buffer (Thermo), 0.5 µl RiboLock RNase Inhibitor (Thermo), and 1 µl Maxima H^- Reverse Transcriptase (Thermo) were added to the mixture and the reaction was incubated at 50 °C for 1 h. To digest excess RT primers, the solution was treated with 1 µl Exo I (NEB) and incubated at 37 °C for 30 min followed by adding 1 µl of 0.5 M EDTA (Sigma) to quench the reaction. To hydrolyze the RNA, 2.5 µl of 1 M NaOH (Sigma) was added and the solution was then incubated at 70 °C for 12 min followed by adding 2.5 µl of 1 M HCl (Sigma) to neutralize NaOH. The cDNA was finally purified with Dynabeads MyOne Silane (Invitrogen) and eluted with 13 µl nuclease-free H_2O. The eluted cDNA was then mixed with 2 µl of 25 µM cDNA adapter (5'-/5Phos/NNNNNNAGATCGGAAGAGCACACGTCTG/3SpC3/-3'), incubated at 70 °C for 2 min and immediately placed on ice. Next, 5 µl 10× T4 RNA Ligase reaction buffer, 25 µl 50% PEG 8000, 0.5 µl of 100 mM ATP (NEB), 3.5 µl DMSO (Thermo), and 1 µl T4 RNA Ligase 1, high concentration (NEB) were added to the mixture and the reaction was incubated at 25 °C for 16 h. The ligated cDNA was purified with Dynabeads MyOne Silane and eluted with 15 µl nuclease-free H_2O. The eluted DNA was amplified with NEBNext Multiplex Oligos for Illumina (96 Unique Dual Index Primer Pairs) and NEBNext Ultra II Q5 Master Mix for 10–12 cycles according to the manufacturer's protocol. The PCR products were purified with 0.8× AMPure XP beads and quantified with Qubit dsDNA HS Assay Kit (Thermo) according to the manufacturer's protocol. BACS and control libraries were sequenced on NextSeq 2000 (60 bp paired end) with no PhiX added.

Modified BID-seq.

Library construction of modified BID-seq was conducted similarly with BACS except for the chemical conversion step. For BS treatment, revised BS and desulphonation reaction conditions of BID-seq were used [230]. Briefly, the 3'-ligated RNA was eluted in 10 µl nuclease-free H_2O. An 8.5 µl aliquot was subjected to modified BID-seq library construction, while the remaining 1.5 µl was saved as control sample. The 8.5 µl aliquot was mixed with 45 µl freshly prepared BS reagent (2.4 M Na_2SO_3 and 0.36 M $NaHSO_3$, Sigma) and incubated at 70 °C for 3 h. The treated RNA was purified by Zymo-IC Column with RNA binding buffer. In-column desulphonation was performed using RNA desulphonation buffer (Zymo), with the column incubated at room temperature for 75 min.

Data pre-processing.

Raw sequencing reads were processed by Cutadapt (v.4.2) [390] to remove low-quality bases (-q 20) and short reads (-m 18), as well as to trim adaptors. 6mer UMI were extracted by UMI-tools extract (v.1.0.1) [391] and used for deduplication. Paired reads were then merged into single reads using fastp (v.1.0.1) [371].

Read alignment.

Cleaned reads were first mapped to synthetic spike-ins and rRNA references using bowtie2 (v.2.4.4) [392]. The key parameters are as follows: bowtie2 -p 2 --no-unal --local -L 16 -N 1 --mp 4. Human rRNA sequences (NR_023363.1, NR_003285.3, NR_003286.4, and NR_003287.4) were downloaded from the National Center for Biotechnology

Information (NCBI). The unaligned reads were subsequently mapped to snoRNA references and then to tRNA references, using the same parameters. Human snoRNA sequences that belong to HUGO Gene Nomenclature Committee (HGNC) "Small nucleolar RNAs" gene group (https://www.genenames.org/) were downloaded from RefSeq (https://www.ncbi.nlm.nih.gov/refseq/). Duplicate snoRNA sequences were removed. High-confidence human tRNA sequences (hg38) were downloaded from GtRNAdb [393]. Only non-redundant tRNA sequences were kept and appended with a "3'-CCA" end. Finally, unmapped reads were aligned to human genome (hg38) with GENCODE v.43 annotation by STAR (v.2.7.9a) [394].

For RNA viruses, the following reference genomes were used: Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1 (NCBI, NC_045512.2), Recombinant Hepatitis C virus J6(5'UTR-NS2)/JFH1 (NCBI, JF343782.1), Zika virus isolate ZIKV/*H. sapiens*/Brazil/Natal/2015 (NCBI, NC_035889.1), Hepatitis Delta Virus sequence from the pSVL(D3) plasmid [395] (Addgene plasmid #29335) (https://www.addgene.org/29335/), and Sindbis virus (NCBI, NC_001547.1). For EBV samples, reads were aligned to Epstein-Barr virus (EBV) genome, strain B95-8 (NCBI, V01555.2).

The aligned reads were then filtered and sorted using samtools (v.1.16.1) [396]. For synthetic spike-ins and rRNA, only reads with MAPQ ≥10 were kept. For snoRNA and tRNA, only reads with MAPQ ≥1 were kept. For mRNA, only uniquely mapped reads (-q 30) with a maximum of 3 mutation counts were kept. Deduplication was performed using UMI-tools dedup (v.1.0.1) [391]. Finally, mutations are counted by samtools mpileup (v.1.16.1) [396] and cpup (v.0.1.0) (https://github.com/y9c/cpup).

Calling Ψ sites.

BACS raw conversion rates were calculated as C/(T+C). The Ψ modification levels were calculated using the linear equation: Ψ modification level = (R–F)/(C–F), where R, F, and C indicated raw conversion rates, motif-specific false-positive rates (from NNUNN spike-in for ncRNAs or IVT-BACS library for polyA-tailed RNA), and motif-specific conversion rates (from NNΨNN spike-in), respectively. The following criteria were used to call Ψ sites in ncRNAs: (1) coverage ≥20 in both BACS and control libraries; (2) background conversion rate ≤0.01 or T-to-C mutation counts ≤2 in control libraries; (3) background T-to-R (R = A or G) mutation ratio ≤0.10 in control libraries; (4) Ψ modification level ≥0.05; (5) a *p*-value was calculated for each site using the motif-specific false-positive rates and then adjusted following the Benjamini-Hochberg (BH) procedure; the adjusted *p*-value is required to be <0.001; (6) consistently detected in all replicates. For calling cy-tRNA Ψ sites, criteria (4) and (6) were modified to require a Ψ modification level ≥0.10 in at least one out of two replicates. Only Ψ sites identified in expressed cy-tRNA isodecoders were reported. The following criteria were used to call Ψ sites in polyA-tailed RNA: (1) coverage ≥20 in BACS, control, and IVT-BACS libraries; (2) background conversion rate ≤0.01 or T-to-C mutation counts ≤2 in control libraries; conversion rate in BACS libraries higher than that in IVT-BACS libraries; (3) background T-to-R (R = A or G) mutation ratio ≤0.10 in control libraries; (4) Ψ modification level ≥0.05; (5) a contingency table test was performed for each site between BACS and IVT-BACS libraries; the *p*-value is required to be <0.01. Statistical analyses were performed in R (v.4.0.3).

Calling A-to-I editing sites.

A-to-I editing sites were called based on ICE-seq protocol [397], with minor modifications. Candidate A-to-I editing sites are required to meet the following criteria: (1) coverage ≥10 in control libraries; (2) A-to-G mutation ratio ≥0.05 in control libraries; (3) A-to-Y (Y = C or T) mutation ratio <0.05 in control libraries. Supporting reads for the candidate sites are required to meet the following criteria: (1) supporting reads containing an indel call within 5 bp upstream or downstream to the candidate sites were filtered out; (2) supporting reads containing mismatches other than A-to-G mutation were filtered out; (3) total counts of supporting reads ≥3. A ΔG score for each candidate site was calculated as follows:

$$\Delta G = \log_2\left(\frac{N_G(BACS) + 1}{N_G(control) + 1} \cdot \frac{N(control)}{N(BACS)}\right)$$

where N_G and N denote the counts of A-to-G mutation and total mapped bases, respectively. The ΔG score is required to be ≤–1. Finally, candidate A-to-I editing sites registered as common SNPs in the dbSNP [398] were removed.

RNA structure visualization.

The RNA-RNA interactions were visualized using r2r (v.1.0.6) [399]. The snoRNA-rRNA interactions were adapted from snoRNA Atlas [241].

Downstream analysis.

The snoRNA box and guide sequences were downloaded from snoDB 2.0 [400]. In the metagene analysis, snoRNA sequences that displayed considerable similarity were

streamlined, retaining only one representative snoRNA. The annotation of Ψ sites identified in polyA-tailed RNA was performed using bedtools intersect (v.2.30.0) [379] with GENCODE v.43 annotation. Read counts obtained from featureCounts (v.1.6.4) [401] were normalized based on sequencing depth and gene length using the transcripts per million (TPM) method. GO analysis was performed with mRNA Ψ sites using enrichR (v.3.2) [402]. Sequence logos were generated using ggseqlogo (v.0.1) [403] in R (v.4.3.1).

Calling PUS-dependent Ψ sites.

The following criteria were used to call PUS-dependent Ψ sites: (1) coverage ≥20 in both WT and KO samples; (2) Ψ modification level ≥0.05 in WT samples; (3) T-to-C mutation counts ≥10 in WT samples; (4) Ψ modification level ≤0.01 in KO samples. For down-regulated Ψ sites, we required the reduction in Ψ modification level to be ≥0.20.

Published data.

Related published data were downloaded from the Gene Expression Omnibus (GEO) database: BID-seq for HeLa cells (GSE179798) [228]. BID-seq data were processed using the original pipeline (BID-seq [228]) and updated pipeline (BID-pipe [230]), respectively.

8 References

1. Grosjean, H. Nucleic acids are not boring long polymers of only four types of nucleotides: a guided tour. in *DNA and RNA Modification Enzymes: Structure, Mechanism, Function and Evolution.* (ed. H. Grosjean) 1–18 (Landes Bioscience, Austin; 2009).
2. Raiber, E.-A., Hardisty, R., van Delft, P. & Balasubramanian, S. Mapping and elucidating the function of modified bases in DNA. *Nat. Rev. Chem.* **1**, 0069 (2017).
3. Cappannini, A. et al. MODOMICS: a database of RNA modifications and related information. 2023 update. *Nucleic Acids Res.* **52**, D239–D244 (2024).
4. Robertson, K. D. DNA methylation and human disease. *Nat. Rev. Genet.* **6**, 597–610 (2005).
5. Kulis, M. & Esteller, M. DNA Methylation and Cancer. *Adv. Genet.* **70**, 27–56 (2010).
6. Jonkhout, N. et al. The RNA modification landscape in human disease. *RNA* **23**, 1754–1769 (2017).
7. Barbieri, I. & Kouzarides, T. Role of RNA modifications in cancer. *Nat. Rev. Cancer* **20**, 303–322 (2020).
8. Jordheim, L. P., Durantel, D., Zoulim, F. & Dumontet, C. Advances in the development of nucleoside and nucleotide analogues for cancer and viral diseases. *Nat. Rev. Drug Discov.* **12**, 447–464 (2013).
9. Khvorova, A. & Watts, J. K. The chemical evolution of oligonucleotide therapies of clinical utility. *Nat. Biotechnol.* **35**, 238–248 (2017).

10. Johnson, T. B. & Coghill, R. D. Researches on pyrimidines. C111. The discovery of 5-methyl-cytosine in tuberculinic acid, the nucleic acid of the tubercle bacillus. *J. Am. Chem. Soc.* **47**, 2838–2844 (1925).

11. Vischer, E., Zamenhof, S. & Chargaff, E. Microbial nucleic acids: the desoxypentose nucleic acids of avian tubercle bacilli and yeast. *J. Biol. Chem.* **177**, 429–438 (1949).

12. Hotchkiss, R. D. The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. *J. Biol. Chem.* **175**, 315–332 (1948).

13. Wyatt, G. R. Occurrence of 5-methyl-cytosine in nucleic acids. *Nature* **166**, 237–238 (1950).

14. Mattei, A. L., Bailly, N. & Meissner, A. DNA methylation: a historical perspective. *Trends Genet.* **38**, 676–707 (2022).

15. Globisch, D. et al. Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS One* **5**, e15367 (2010).

16. Smith, Z. D. & Meissner, A. DNA methylation: roles in mammalian development. *Nat. Rev. Genet.* **14**, 204–220 (2013).

17. Lyko, F. The DNA methyltransferase family: a versatile toolkit for epigenetic regulation. *Nat. Rev. Genet.* **19**, 81–92 (2018).

18. Mayer, W., Niveleau, A., Walter, J., Fundele, R. & Haaf, T. Demethylation of the zygotic paternal genome. *Nature* **403**, 501–502 (2000).

19. Oswald, J. et al. Active demethylation of the paternal genome in the mouse zygote. *Curr. Biol.* **10**, 475–478 (2000).

20. Kohli, R. M. & Zhang, Y. TET enzymes, TDG and the dynamics of DNA demethylation. *Nature* **502**, 472–479 (2013).

21. Kriaucionis, S. & Heintz, N. The nuclear DNA base 5-hydroxymethylcytosine is present in purkinje neurons and the brain. *Science* **324**, 929–930 (2009).

22. Tahiliani, M. et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930–935 (2009).

23. Ito, S. et al. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333**, 1300–1303 (2011).

24. He, Y. F. et al. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333**, 1303–1307 (2011).

25. Wu, X. & Zhang, Y. TET-mediated active DNA demethylation: mechanism, function and beyond. *Nat. Rev. Genet.* **18**, 517–534 (2017).

26. Waddington, C. H. The epigenotype. *Endeavour* **1**, 18–20 (1942).

27. Goldberg, A. D., Allis, C. D. & Bernstein, E. Epigenetics: a landscape takes shape. *Cell* **128**, 635–638 (2007).

28. Allis, C. D. & Jenuwein, T. The molecular hallmarks of epigenetic control. *Nat. Rev. Genet.* **17**, 487–500 (2016).

29. Jaenisch, R. & Bird, A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat. Genet.* **33 (Suppl 3)**, 245–254 (2003).

30. Cedar, H. & Bergman, Y. Linking DNA methylation and histone modification: patterns and paradigms. *Nat. Rev. Genet.* **10**, 295–304 (2009).

31. Li, E. & Zhang, Y. DNA methylation in mammals. *Cold Spring Harb. Perspect. Biol.* **6**, a019133 (2014).

32. Branco, M. R., Ficz, G. & Reik, W. Uncovering the role of 5-hydroxymethylcytosine in the epigenome. *Nat. Rev. Genet.* **13**, 7–13 (2011).

33. Luo, C. Y., Hajkova, P. & Ecker, J. R. Dynamic DNA methylation: in the right place at the right time. *Science* **361**, 1336–1340 (2018).

34. Greenberg, M. V. C. & Bourc'his, D. The diverse roles of DNA methylation in mammalian development and disease. *Nat. Rev. Mol. Cell Biol.* **20**, 590–607 (2019).

35. Jones, P. A. & Baylin, S. B. The epigenomics of cancer. *Cell* **128**, 683–692 (2007).

36. Esteller, M. Cancer epigenomics: DNA methylomes and histone-modification maps. *Nat. Rev. Genet.* **8**, 286–298 (2007).

37. Jeschke, J., Collignon, E. & Fuks, F. Portraits of TET-mediated DNA hydroxymethylation in cancer. *Curr. Opin. Genet. Dev.* **36**, 16–26 (2016).

38. Egger, G., Liang, G. N., Aparicio, A. & Jones, P. A. Epigenetics in human disease and prospects for epigenetic therapy. *Nature* **429**, 457–463 (2004).

39. Yoo, C. B. & Jones, P. A. Epigenetic therapy of cancer: past, present and future. *Nat. Rev. Drug Discov.* **5**, 37–50 (2006).

40. Kelly, T. K., De Carvalho, D. D. & Jones, P. A. Epigenetic modifications as therapeutic targets. *Nat. Biotechnol.* **28**, 1069–1078 (2010).

41. Jones, P. A., Issa, J. P. & Baylin, S. Targeting the cancer epigenome for therapy. *Nat. Rev. Genet.* **17**, 630–641 (2016).

42. Wan, J. C. M. et al. Liquid biopsies come of age: towards implementation of circulating tumour DNA. *Nat. Rev. Cancer* **17**, 223–238 (2017).

43. Chan, K. C. A. et al. Noninvasive detection of cancer-associated genome-wide hypomethylation and copy number aberrations by plasma DNA bisulfite sequencing. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 18761–18768 (2013).

44. Song, C. X. et al. 5-Hydroxymethylcytosine signatures in cell-free DNA provide information about tumor types and stages. *Cell Res.* **27**, 1231–1242 (2017).

45. Li, W. et al. 5-Hydroxymethylcytosine signatures in circulating cell-free DNA as diagnostic biomarkers for human cancers. *Cell Res.* **27**, 1243–1257 (2017).

46. Cohn, W. E. & Volkin, E. Nucleoside-5'-phosphates from ribonucleic acid. *Nature* **167**, 483–484 (1951).

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

48. Cohn, W. E. 5-Ribosyl uracil, a carbon-carbon ribofuranosyl nucleoside in ribonucleic acids. *Biochim. Biophys. Acta* **32**, 569–571 (1959).

49. Cohn, W. E. Pseudouridine, a carbon-carbon linked ribonucleoside in ribonucleic acids: isolation, structure, and chemical characteristics. *J. Biol. Chem.* **235**, 1488–1498 (1960).

50. Adler, M., Weissmann, B. & Gutman, A. B. Occurrence of methylated purine bases in yeast ribonucleic acid. *J. Biol. Chem.* **230**, 717–723 (1958).

51. Amos, H. & Korn, M. 5-Methyl cytosine in the RNA of *Escherichia coli*. *Biochim. Biophys. Acta* **29**, 444–445 (1958).

52. Littlefield, J. W. & Dunn, D. B. The occurrence and distribution of thymine and three methylated-adenine bases in ribonucleic acids from several sources. *Biochem. J.* **70**, 642–651 (1958).

53. Smith, J. D. & Dunn, D. B. An additional sugar component of ribonucleic acids. *Biochim. Biophys. Acta* **31**, 573–575 (1959).

54. Smith, J. D. & Dunn, D. B. The occurrence of methylated guanines in ribonucleic acids from several sources. *Biochem. J.* **72**, 294–301 (1959).

55. Hall, R. H., Robins, M. J., Stasiuk, L. & Thedford, R. Isolation of N^6-(γ,γ-dimethylallyl)adenosine from soluble ribonucleic acid. *J. Am. Chem. Soc.* **88**, 2614–2615 (1966).

56. Schweizer, M. P., Chheda, G. B., Baczynskyj, L. & Hall, R. H. Aminoacyl nucleosides. VII. *N*-(purin-6-ylcarbamoyl)threonine. A new component of transfer ribonucleic acid. *Biochemistry* **8**, 3283–3289 (1969).

57. Blobstein, S. H., Grunberger, D., Weinstein, I. B. & Nakanishi, K. Isolation and structure determination of the fluorescent base from bovine liver phenylalanine transfer ribonucleic acid. *Biochemistry* **12**, 188–193 (1973).

58. Kasai, H., Kuchino, Y., Nihei, K. & Nishimura, S. Distribution of the modified nucleoside Q and its derivatives in animal and plant transfer RNA's. *Nucleic Acids Res.* **2**, 1931–1940 (1975).

59. Nishimura, S. Structure, biosynthesis, and function of queuosine in transfer RNA. *Prog. Nucleic Acid Res. Mol. Biol.* **28**, 49–73 (1983).

60. Mandal, D. et al. Agmatidine, a modified cytidine in the anticodon of archaeal $tRNA^{Ile}$, base pairs with adenosine but not with guanosine. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 2872–2877 (2010).

61. Ikeuchi, Y. et al. Agmatine-conjugated cytidine in a tRNA anticodon is essential for AUA decoding in archaea. *Nat. Chem. Biol.* **6**, 277–282 (2010).

62. Krog, J. S. et al. 3-(3-amino-3-carboxypropyl)-5,6-Dihydrouridine is one of two novel post-transcriptional modifications in $tRNA^{Lys}$(UUU) from *Trypanosoma brucei*. *FEBS J.* **278**, 4782–4796 (2011).

63. Chen, P., Crain, P. F., Näsvall, S. J., Pomerantz, S. C. & Björk, G. R. A 'gain of function' mutation in a protein mediates production of novel modified nucleosides. *EMBO J.* **24**, 1842–1851 (2005).

64. Dumelin, C. E., Chen, Y., Leconte, A. M., Chen, Y. G. & Liu, D. R. Discovery and biological characterization of geranylated RNA in bacteria. *Nat. Chem. Biol.* **8**, 913–919 (2012).

65. Suzuki, T. The expanding world of tRNA modifications and their disease relevance. *Nat. Rev. Mol. Cell Biol.* **22**, 375–392 (2021).

66. Sloan, K. E. et al. Tuning the ribosome: the influence of rRNA modification on eukaryotic ribosome biogenesis and function. *RNA Biol.* **14**, 1138–1152 (2017).

67. Grosjean, H. Fine-tuning of RNA functions by modification and editing in *Topics in current genetics*, Vol. 12. (ed. H. Grosjean) (Springer Berlin, Heidelberg; 2005).

68. Jia, G. et al. N^6-Methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat. Chem. Biol.* **7**, 885–887 (2011).

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

70. Meyer, K. D. & Jaffrey, S. R. The dynamic epitranscriptome: N^6-methyladenosine and gene expression control. *Nat. Rev. Mol. Cell Biol.* **15**, 313–326 (2014).

71. Mauer, J. et al. Reversible methylation of m^6A_m in the 5' cap controls mRNA stability. *Nature* **541**, 371–375 (2017).

72. Mauer, J. et al. FTO controls reversible m^6Am RNA methylation during snRNA biogenesis. *Nat. Chem. Biol.* **15**, 340–347 (2019).

73. He, C. Grand challenge commentary: RNA epigenetics? *Nat. Chem. Biol.* **6**, 863–865 (2010).

74. Saletore, Y. et al. The birth of the epitranscriptome: deciphering the function of RNA modifications. *Genome Biol.* **13**, 175 (2012).

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

76. Meyer, K. D. & Jaffrey, S. R. Rethinking m^6A readers, writers, and erasers. *Annu. Rev. Cell Dev. Biol.* **33**, 319–342 (2017).

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

78. Shi, H., Wei, J. & He, C. Where, when, and how: context-dependent functions of RNA methylation writers, readers, and erasers. *Mol. Cell* **74**, 640–650 (2019).

79. Roundtree, I. A., Evans, M. E., Pan, T. & He, C. Dynamic RNA modifications in gene expression regulation. *Cell* **169**, 1187–1200 (2017).

80. Zhao, B. S., Roundtree, I. A. & He, C. Post-transcriptional gene regulation by mRNA modifications. *Nat. Rev. Mol. Cell Biol.* **18**, 31–42 (2017).

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

82. NCT05584111. https://clinicaltrials.gov/study/NCT05584111.

83. Cayir, A. RNA modifications as emerging therapeutic targets. *Wiley Interdiscip. Rev. RNA* **13**, e1702 (2022).

84. Pomaville, M. M. & He, C. Advances in targeting RNA modifications for anticancer therapy. *Trends Cancer* **9**, 528–542 (2023).

85. Wiener, D. & Schwartz, S. The epitranscriptome beyond m^6A. *Nat. Rev. Genet.* **22**, 119–131 (2021).

86. Lewis, C. J. T., Pan, T. & Kalsotra, A. RNA modifications and structures cooperate to guide RNA–protein interactions. *Nat. Rev. Mol. Cell Biol.* **18**, 202–210 (2017).

87. Flamand, M. N., Tegowski, M. & Meyer, K. D. The proteins of mRNA modification: writers, readers, and erasers. *Annu. Rev. Biochem.* **92**, 145–173 (2023).

88. Fagre, C. & Gilbert, W. Beyond reader proteins: RNA binding proteins and RNA modifications in conversation to regulate gene expression. *Wiley Interdiscip. Rev. RNA* **15**, e1834 (2024).

89. Gilbert, W. V., Bell, T. A. & Schaening, C. Messenger RNA modifications: form, distribution, and function. *Science* **352**, 1408–1412 (2016).

90. Sun, H., Li, K., Liu, C. & Yi, C. Regulation and functions of non-m^6A mRNA modifications. *Nat. Rev. Mol. Cell Biol.* **24**, 714–731 (2023).

91. Gilbert, W. V. & Nachtergaele, S. mRNA regulation by RNA modifications. *Annu. Rev. Biochem.* **92**, 175–198 (2023).

92. Meagher, R. B. 'Memory and molecular turnover,' 30 years after inception. *Epigenetics Chromatin* **7**, 37 (2014).

93. Piha, R. S., Cuénod, M. & Waelsch, H. Metabolism of histones of brain and liver. *J. Biol. Chem.* **241**, 2397–2404 (1966).

94. Commerford, S. L., Carsten, A. L. & Cronkite, E. P. Histone turnover within nonproliferating cells. *Proc. Natl. Acad. Sci. U. S. A.* **79**, 1163–1165 (1982).

95. Savas, J. N., Toyama, B. H., Xu, T., Yates, J. R. & Hetzer, M. W. Extremely long-lived nuclear pore proteins in the rat brain. *Science* **335**, 942–942 (2012).

96. Mathieson, T. et al. Systematic analysis of protein turnover in primary cells. *Nat. Commun.* **9**, 689 (2018).

97. Barth, T. K. & Imhof, A. Fast signals and slow marks: the dynamics of histone modifications. *Trends Biochem. Sci.* **35**, 618–626 (2010).

98. Loeb, J. N., Howell, R. R. & Tomkins, G. M. Turnover of ribosomal RNA in rat liver. *Science* **149**, 1093–1095 (1965).

99. Abelson, H. T., Johnson, L. F., Penman, S. & Green, H. Changes in RNA in relation to growth of the fibroblast: II. The lifetime of mRNA, rRNA, and tRNA in resting and growing cells. *Cell* **1**, 161–165 (1974).

100. Sharova, L. V. et al. Database for mRNA half-life of 19 977 genes obtained by DNA microarray analysis of pluripotent and differentiating mouse embryonic stem cells. *DNA Res.* **16**, 45–58 (2008).

101. Weiss, M. C. et al. The physiology and habitat of the last universal common ancestor. *Nat. Microbiol.* **1**, 16116 (2016).

102. Ge, J. & Yu, Y. T. RNA pseudouridylation: new insights into an old modification. *Trends Biochem. Sci.* **38**, 210–218 (2013).

103. Karijolich, J., Yi, C. & Yu, Y.-T. Transcriptome-wide dynamics of RNA pseudouridylation. *Nat. Rev. Mol. Cell Biol.* **16**, 581–585 (2015).

104. Carlile, T. M. et al. Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature* **515**, 143–146 (2014).

105. Schwartz, S. et al. Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. *Cell* **159**, 148–162 (2014).

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

107. Weiss, M. C., Preiner, M., Xavier, J. C., Zimorski, V. & Martin, W. F. The last universal common ancestor between ancient Earth chemistry and the onset of genetics. *PLoS Genet.* **14**, e1007518 (2018).

108. Cerneckis, J., Cui, Q., He, C., Yi, C. & Shi, Y. Decoding pseudouridine: an emerging target for therapeutic development. *Trends Pharmacol. Sci.* **43**, 522–535 (2022).

109. Heiss, N. S. et al. X-linked dyskeratosis congenita is caused by mutations in a highly conserved gene with putative nucleolar functions. *Nat. Genet.* **19**, 32–38 (1998).

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

111. Patton, J. R., Bykhovskaya, Y., Mengesha, E., Bertolotto, C. & Fischel-Ghodsian, N. Mitochondrial myopathy and sideroblastic anemia (MLASA): missense mutation in the pseudouridine synthase 1 (*PUS1*) gene is associated with the loss of tRNA pseudouridylation. *J. Biol. Chem.* **280**, 19823–19828 (2005).

112. Shaheen, R. et al. A homozygous truncating mutation in *PUS3* expands the role of tRNA modification in normal cognition. *Hum. Genet.* **135**, 707–713 (2016).

113. de Brouwer, A. P. M. et al. Variants in *PUS7* cause intellectual disability with speech delay, microcephaly, short stature, and aggressive behavior. *Am. J. Hum. Genet.* **103**, 1045–1052 (2018).

114. Shaheen, R. et al. *PUS7* mutations impair pseudouridylation in humans and cause intellectual disability and microcephaly. *Hum. Genet.* **138**, 231–239 (2019).

115. Guzzi, N. et al. Pseudouridylation of tRNA-derived fragments steers translational control in stem cells. *Cell* **173**, 1204–1216 (2018).

116. Guzzi, N. et al. Pseudouridine-modified tRNA fragments repress aberrant protein synthesis and predict leukaemic progression in myelodysplastic syndrome. *Nat. Cell Biol.* **24**, 299–306 (2022).

117. Cui, Q. et al. Targeting PUS7 suppresses tRNA pseudouridylation and glioblastoma tumorigenesis. *Nat. Cancer* **2**, 932–949 (2021).

118. Spenkuch, F., Motorin, Y. & Helm, M. Pseudouridine: still mysterious, but never a fake (uridine)! *RNA Biol.* **11**, 1540–1554 (2014).

119. Gillingham, D., Geigle, S. & Anatole von Lilienfeld, O. Properties and reactivity of nucleic acids relevant to epigenomics, transcriptomics, and therapeutics. *Chem. Soc. Rev.* **45**, 2637–2655 (2016).

120. Behm-Ansmant, I., Helm, M. & Motorin, Y. Use of specific chemical reagents for detection of modified nucleotides in RNA. *J. Nucleic Acids* **2011**, 408053 (2011).

121. Helm, M., Schmidt-Dengler, M. C., Weber, M. & Motorin, Y. General principles for the detection of modified nucleotides in RNA by specific reagents. *Adv. Biol. (Weinh).* **5**, e2100866 (2021).

122. Dizdaroglu, M., Jaruga, P., Birincioglu, M. & Rodriguez, H. Free radical-induced damage to DNA: mechanisms and measurement. *Free Radic. Biol. Med.* **32**, 1102–1115 (2002).

123. Dizdaroglu, M. & Jaruga, P. Mechanisms of free radical-induced damage to DNA. *Free Radic. Res.* **46**, 382–419 (2012).

124. Cadet, J., Anselmino, C., Douki, T. & Voituriez, L. Photochemistry of nucleic acids in cells. *J. Photochem. Photobiol. B* **15**, 277–298 (1992).

125. Görner, H. Photochemistry of DNA and related biomolecules: quantum yields and consequences of photoionization. *J. Photochem. Photobiol. B* **26**, 117–139 (1994).

126. Schreier, W. J., Gilch, P. & Zinth, W. Early events of DNA photodamage. *Annu. Rev. Phys. Chem.* **66**, 497–519 (2015).

127. Improta, R., Santoro, F. & Blancafort, L. Quantum mechanical studies on the photophysics and the photochemistry of nucleic acids and nucleobases. *Chem. Rev.* **116**, 3540–3593 (2016).

128. Lawley, P. & Brookes, P. Further studies on the alkylation of nucleic acids and their constituent nucleotides. *Biochem. J.* **89**, 127–138 (1963).

129. Lee, Y.-H., Yu, E. & Park, C.-M. Programmable site-selective labeling of oligonucleotides based on carbene catalysis. *Nat. Commun.* **12**, 1681 (2021).

130. Shapiro, R. & Hachmann, J. The reaction of guanine derivatives with 1,2-dicarbonyl compounds. *Biochemistry* **5**, 2799–2807 (1966).

131. Shapiro, R., Cohen, B. I., Shiuey, S.-J. & Maurer, H. Reaction of guanine with glyoxal, pyruvaldehyde, and kethoxal, and the structure of the acylguanines. Synthesis of N^2-alkylguanines. *Biochemistry* **8**, 238–245 (1969).

132. Okamoto, A., Tainaka, K. & Kamei, T. Sequence-selective osmium oxidation of DNA: efficient distinction between 5-methylcytosine and cytosine. *Org. Biomol. Chem.* **4**, 1638–1640 (2006).

133. Sugizaki, K., Ikeda, S., Yanagisawa, H. & Okamoto, A. Facile synthesis of hydroxymethylcytosine-containing oligonucleotides and their reactivity upon osmium oxidation. *Org. Biomol. Chem.* **9**, 4176–4181 (2011).

134. Okamoto, A., Sugizaki, K., Nakamura, A., Yanagisawa, H. & Ikeda, S. 5-Hydroxymethylcytosine-selective oxidation with peroxotungstate. *Chem. Commun.* **47**, 11231–11233 (2011).

135. Jones, J. W. & Robins, R. K. Purine nucleosides. III. Methylation studies of certain naturally occurring purine nucleosides. *J. Am. Chem. Soc.* **85**, 193–201 (1963).

136. Macon, J. B. & Wolfenden, R. 1-Methyladenosine. Dimroth rearrangement and reversible reduction. *Biochemistry* **7**, 3453–3458 (1968).

137. Wintermeyer, W. & Zachau, H. G. A specific chemical chain scission of tRNA at 7-methylguanosine. *FEBS Lett.* **11**, 160–164 (1970).

138. Wintermeyer, W. & Zachau, H. G. Tertiary structure interactions of 7-methylguanosine in yeast $tRNA^{Phe}$ as studied by borohydride reduction. *FEBS Lett.* **58**, 306–309 (1975).

139. Cui, J., Liu, Q., Sendinc, E., Shi, Y. & Gregory, R. I. Nucleotide resolution profiling of m^3C RNA modification by HAC-seq. *Nucleic Acids Res.* **49**, e27 (2020).

140. Shapiro, R. & Pohl, S. H. Reaction of ribonucleosides with nitrous acid. Side products and kinetics. *Biochemistry* **7**, 448–455 (1968).

141. Shapiro, R. & Shiuey, S.-J. Reaction of nitrous acid with alkylaminopurines. *Biochim. Biophys. Acta* **174**, 403–405 (1969).

142. Booth, M. J. et al. Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. *Science* **336**, 934–937 (2012).

143. Matsushita, T., Moriyama, Y., Nagae, G., Aburatani, H. & Okamoto, A. DNA-friendly Cu(II)/TEMPO-catalyzed 5-hydroxymethylcytosine-specific oxidation. *Chem. Commun.* **53**, 5756–5759 (2017).

144. Song, C. X. et al. Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. *Cell* **153**, 678–691 (2013).

145. Lu, X. et al. Chemical modification-assisted bisulfite sequencing (CAB-Seq) for 5-carboxylcytosine detection in DNA. *J. Am. Chem. Soc.* **135**, 9315–9317 (2013).

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

147. Burton, K. Oxidation of pyrimidine nucleosides and nucleotides by osmium tetroxide. *Biochem. J.* **104**, 686–694 (1967).

148. Ziff, E. B. & Fresco, J. R. Chemical transformation of 4-thiouracil nucleosides to uracil and cytosine counterparts. *J. Am. Chem. Soc.* **90**, 7338–7342 (1968).

149. Hayatsu, H. & Yano, M. Permanganate oxidation of 4-thiouracil derivatives. Isolation and properties of 1-substituted-2-pyrimidone-4-sulfonates. *Tetrahedron Lett.* **10**, 755–758 (1969).

150. Yano, M. & Hayatsu, H. Permanganate oxidation of 4-thiouracil derivatives: isolation and properties of 1-substituted 2-pyrimidone 4-sulfonates. *Biochim. Biophys. Acta* **199**, 303–315 (1970).

151. Ofengand, J. The function of pseudouridylic acid in transfer ribonucleic acid: I. The specific cyanoethylation of pseudouridine, inosine, and 4-thiouridine by acrylonitrile. *J. Biol. Chem.* **242**, 5034–5045 (1967).

152. Yoshida, M. & Ukita, T. Modification of nucleosides and nucleotides: VII. Selective cyanoethylation of inosine and pseudouridine in yeast transfer ribonucleic acid. *Biochim. Biophys. Acta.* **157**, 455–465 (1968).

153. Krasheninina, O. A., Thaler, J., Erlacher, M. D. & Micura, R. Amine-to-azide conversion on native RNA via metal-free diazotransfer opens new avenues for RNA manipulations. *Angew. Chem. Int. Ed.* **60**, 6970–6974 (2021).

154. Xia, B. et al. Bisulfite-free, base-resolution analysis of 5-formylcytosine at the genome scale. *Nat. Methods* **12**, 1047–1050 (2015).

155. Robins, M. J., Hall, R. H. & Thedford, R. N^6-(Δ^2-isopentenyl)adenosine. A component of the transfer ribonucleic acid of yeast and of mammalian tissue, methods of isolation, and characterization. *Biochemistry* **6**, 1837–1848 (1967).

156. Merino, E. J., Wilkinson, K. A., Coughlan, J. L. & Weeks, K. M. RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). *J. Am. Chem. Soc.* **127**, 4223–4231 (2005).

157. Birkedal, U. et al. Profiling of ribose methylations in RNA by high-throughput sequencing. *Angew. Chem. Int. Ed.* **54**, 451–455 (2015).

158. Dai, Q. et al. Nm-seq maps 2'-*O*-methylation sites in human mRNA with base precision. *Nat. Methods* **14**, 695–698 (2017).

159. Voorhees, R. M. et al. The structural basis for specific decoding of AUA by isoleucine tRNA on the ribosome. *Nat. Struct. Mol. Biol.* **20**, 641–643 (2013).

160. Agris, P. F., Vendeix, F. A. P. & Graham, W. D. tRNA's wobble decoding of the genome: 40 years of modification. *J. Mol. Biol.* **366**, 1–13 (2007).

161. Grosjean, H., de Crécy-Lagard, V. & Marck, C. Deciphering synonymous codons in the three domains of life: co-evolution with specific tRNA modification enzymes. *FEBS Lett.* **584**, 252–264 (2010).

162. Agris, P. F. et al. Celebrating wobble decoding: half a century and still much is new. *RNA Biol.* **15**, 537–553 (2018).

163. Rozov, A. et al. Novel base-pairing interactions at the tRNA wobble position crucial for accurate reading of the genetic code. *Nat. Commun.* **7**, 10457 (2016).

164. Shapiro, R., Servis, R. E. & Welcher, M. Reactions of uracil and cytosine derivatives with sodium bisulfite. A specific deamination method. *J. Am. Chem. Soc.* **92**, 422–424 (1970).

165. Hayatsu, H., Wataya, Y. & Kai, K. Addition of sodium bisulfite to uracil and to cytosine. *J. Am. Chem. Soc.* **92**, 724–726 (1970).

166. Hayatsu, H., Wataya, Y., Kai, K. & Iida, S. Reaction of sodium bisulfite with uracil, cytosine, and their derivatives. *Biochemistry* **9**, 2858–2865 (1970).

167. Sono, M., Wataya, Y. & Hayatsu, H. Role of bisulfite in the deamination and the hydrogen isotope exchange of cytidylic acid. *J. Am. Chem. Soc.* **95**, 4745–4749 (1973).

168. Shapiro, R., DeFate, V. & Welcher, M. Deamination cytosine derivatives by bisulfite. Mechanism of the reaction. *J. Am. Chem. Soc.* **96**, 906–912 (1974).

169. Shapiro, R., Braverman, B., Louis, J. B. & Servis, R. E. Nucleic acid reactivity and conformation: II. Reaction of cytosine and uracil with sodium bisulfite. *J. Biol. Chem.* **248**, 4060–4064 (1973).

170. Hayatsu, H. Discovery of bisulfite-mediated cytosine conversion to uracil, the key reaction for DNA methylation analysis — A personal account. *Proc. Jpn. Acad. Ser. B Phys. Biol. Sci.* **84**, 321–330 (2008).

171. Hayatsu, H. & Shiragami, M. Reaction of bisulfite with the 5-hydroxymethyl group in pyrimidines and in phage DNAs. *Biochemistry* **18**, 632–637 (1979).

172. Wu, Q., Amrutkar, S. M. & Shao, F. W. Sulfinate based selective labeling of 5-hydroxymethylcytosine: application to biotin pull down assay. *Bioconjugate Chem.* **29**, 245–249 (2018).

173. Yu, M. et al. Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell* **149**, 1368–1380 (2012).

174. Isono, K., Suzuki, S., Tanaka, M., Nanbata, T. & Shibuya, K. Reaction of 5-carboxypyrimidine derivatives with sodium bisulfite. A facile decarboxylation method. *Agric. Biol. Chem.* **36**, 1571–1579 (1972).

175. Frommer, M. et al. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 1827–1831 (1992).

176. Cokus, S. J. et al. Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nature* **452**, 215–219 (2008).

177. Booth, M. J., Marsico, G., Bachman, M., Beraldi, D. & Balasubramanian, S. Quantitative sequencing of 5-formylcytosine in DNA at single-base resolution. *Nat. Chem.* **6**, 435–440 (2014).

178. Wu, H., Wu, X., Shen, L. & Zhang, Y. Single-base resolution analysis of active DNA demethylation using methylase-assisted bisulfite sequencing. *Nat. Biotechnol.* **32**, 1231–1240 (2014).

179. Tanaka, K. & Okamoto, A. Degradation of DNA by bisulfite treatment. *Bioorg. Med. Chem. Lett.* **17**, 1912–1915 (2007).

180. Liu, Y. et al. Bisulfite-free direct detection of 5-methylcytosine and 5-hydroxymethylcytosine at base resolution. *Nat. Biotechnol.* **37**, 424–429 (2019).

181. Liu, Y. et al. Accurate targeted long-read DNA methylation and hydroxymethylation sequencing with TAPS. *Genome Biol.* **21**, 54 (2020).

182. Chen, J. et al. Whole-genome long-read TAPS deciphers DNA methylation patterns at base resolution using PacBio SMRT sequencing technology. *Nucleic Acids Res.* **50**, e104 (2022).

183. Siejka-Zielinska, P. et al. Cell-free DNA TAPS provides multimodal information for early cancer detection. *Sci. Adv.* **7**, eabh0534 (2021).

184. Liu, Y. et al. Subtraction-free and bisulfite-free specific sequencing of 5-methylcytosine and its oxidized derivatives at base resolution. *Nat. Commun.* **12**, 618 (2021).

185. Sas-Chen, A. et al. Dynamic RNA acetylation revealed by quantitative cross-evolutionary mapping. *Nature* **583**, 638–643 (2020).

186. Zerk, T. J. et al. Elucidating the mechanism of the Ley-Griffith (TPAP) alcohol oxidation. *Chem. Sci.* **8**, 8435–8442 (2017).

187. Zeng, H. et al. Bisulfite-free, nanoscale analysis of 5-hydroxymethylcytosine at single base resolution. *J. Am. Chem. Soc.* **140**, 13190–13194 (2018).

188. Liu, P. Y. et al. The oxidative damage of plasmid DNA by ascorbic acid derivatives *in vitro*: the first research on the relationship between the structure of ascorbic acid and the oxidative damage of plasmid DNA. *Chem. Biodivers.* **3**, 958–966 (2006).

189. Abel, G. R., Jr., Calabrese, Z. A., Ayco, J., Hein, J. E. & Ye, T. Measuring and suppressing the oxidative damage to DNA during Cu(I)-catalyzed azide-alkyne cycloaddition. *Bioconjug. Chem.* **27**, 698–704 (2016).

190. Fukuzawa, S., Tachibana, K., Tajima, S. & Suetake, I. Selective oxidation of 5-hydroxymethylcytosine with micelle incarcerated oxidants to determine it at single base resolution. *Bioorg. Med. Chem. Lett.* **25**, 5667–5671 (2015).

191. Fukuzawa, S., Takahashi, S., Tachibana, K., Tajima, S. & Suetake, I. Simple and accurate single base resolution analysis of 5-hydroxymethylcytosine by catalytic oxidative bisulfite sequencing using micelle incarcerated oxidants. *Biorg. Med. Chem.* **24**, 4254–4262 (2016).

192. Golubev, V. A., Rozantsev, É. G. & Neiman, M. B. Some reactions of free iminoxyl radicals with the participation of the unpaired electron. *Izv. Akad. Nauk SSSR, Ser. Khim.* **11**, 1927–1936 (1965).

193. Bobbitt, J. M. & Flores, M. C. L. Organic nitrosonium salts as oxidants in organic-chemistry. *Heterocycles* **27**, 509–533 (1988).

194. Bobbitt, J. M. Oxoammonium salts. 6. 4-Acetylamino-2,2,6,6-tetramethylpiperidine-1-oxoammonium perchlorate: a stable and convenient reagent for the oxidation of alcohols. Silica gel catalysis. *J. Org. Chem.* **63**, 9367–9374 (1998).

195. Merbouh, N., Bobbitt, J. M. & Brückner, C. Preparation of tetramethylpiperidine-1-oxoammonium salts and their use as oxidants in organic chemistry. A review. *Org. Prep. Proced. Int.* **36**, 1–31 (2004).

196. Qiu, J. C., Pradhan, P. P., Blanck, N. B., Bobbitt, J. M. & Bailey, W. F. Selective oxoammonium salt oxidations of alcohols to aldehydes and aldehydes to carboxylic acids. *Org. Lett.* **14**, 350–353 (2012).

197. Mercadante, M. A., Kelly, C. B., Bobbitt, J. M., Tilley, L. J. & Leadbeater, N. E. Synthesis of 4-acetamido-2,2,6,6-tetramethylpiperidine-1-oxoammonium tetrafluoroborate and 4-acetamido-(2,2,6,6-tetramethyl-piperidin-1-yl)oxyl and their use in oxidative reactions. *Nat. Protoc.* **8**, 666–676 (2013).

198. Leifert, D. & Studer, A. Organic synthesis using nitroxides. *Chem. Rev.* **123**, 10302–10380 (2023).

199. Bailey, W. F., Bobbitt, J. M. & Wiberg, K. B. Mechanism of the oxidation of alcohols by oxoammonium cations. *J. Org. Chem.* **72**, 4504–4509 (2007).

200. Hamlin, T. A. et al. Toward a unified mechanism for oxoammonium salt-mediated oxidation reactions: a theoretical and experimental study using a hydride transfer model. *J. Org. Chem.* **80**, 8150–8167 (2015).

201. Maiti, A., Michelson, A. Z., Armwood, C. J., Lee, J. K. & Drohat, A. C. Divergent mechanisms for enzymatic excision of 5-formylcytosine and 5-carboxylcytosine from DNA. *J. Am. Chem. Soc.* **135**, 15813–15822 (2013).

202. Link, C. N. et al. Protonation-dependent sequencing of 5-formylcytidine in RNA. *Biochemistry* **61**, 535–544 (2022).

203. Zott, F. L., Korotenko, V. & Zipse, H. The pH-dependence of the hydration of 5-formylcytosine: an experimental and theoretical study. *ChemBioChem* **23**, e202100651 (2022).

204. Kraus, G. A. & Taschner, M. J. Model studies for the synthesis of quassinoids. 1. Construction of the BCE ring system. *J. Org. Chem.* **45**, 1175–1176 (1980).

205. Kraus, G. A. & Roth, B. Synthetic studies toward verrucarol. 2. Synthesis of the AB ring system. *J. Org. Chem.* **45**, 4825–4830 (1980).

206. Bal, B. S., Childers, W. E. & Pinnick, H. W. Oxidation of α,β-unsaturated aldehydes. *Tetrahedron* **37**, 2091–2096 (1981).

207. Raach, A. & Reiser, O. Sodium chlorite-hydrogen peroxide — a mild and selective reagent for the oxidation of aldehydes to carboxylic acids. *J. Prakt. Chem.* **342**, 605–608 (2000).

208. Grollman, A. P. & Moriya, M. Mutagenesis by 8-oxoguanine: an enemy within. *Trends Genet.* **9**, 246–249 (1993).

209. Schutsky, E. K. et al. Nondestructive, base-resolution sequencing of 5-hydroxymethylcytosine using a DNA deaminase. *Nat. Biotechnol.* **36**, 1083–1090 (2018).

210. Wen, L. et al. Whole-genome analysis of 5-hydroxymethylcytosine and 5-methylcytosine at base resolution in the human brain. *Genome Biol.* **15**, R49 (2014).

211. Raiber, E.-A. et al. Base resolution maps reveal the importance of 5-hydroxymethylcytosine in a human glioblastoma. *NPJ Genom. Med.* **2**, 6 (2017).

212. Irie, N. et al. DMRT1 regulates human germline commitment. *Nat. Cell Biol.* **25**, 1439–1452 (2023).

213. Li, X., Ma, S. & Yi, C. Pseudouridine: the fifth RNA nucleotide with renewed interests. *Curr. Opin. Chem. Biol.* **33**, 108–116 (2016).

214. Rintala-Dempsey, A. C. & Kothe, U. Eukaryotic stand-alone pseudouridine synthases – RNA modifying enzymes and emerging regulators of gene expression? *RNA Biol.* **14**, 1185–1196 (2017).

215. Borchardt, E. K., Martinez, N. M. & Gilbert, W. V. Regulation and function of RNA pseudouridylation in human cells. *Annu. Rev. Genet.* **54**, 309–336 (2020).

216. Bakin, A. V. & Ofengand, J. Mapping of pseudouridine residues in RNA to nucleotide resolution. *Methods Mol. Biol.* **77**, 297–309 (1998).

217. Gilham, P. T. An addition reaction specific for uridine and guanosine nucleotides and its application to the modification of ribonuclease action. *J. Am. Chem. Soc.* **84**, 687–688 (1962).

218. Ho, N. W. Y. & Gilham, P. T. The reversible chemical modification of uracil, thymine, and guanine nucleotides and the modification of the action of ribonuclease on ribonucleic acid. *Biochemistry* **6**, 3632–3639 (1967).

219. Ho, N. W. Y. & Gilham, P. T. Reaction of pseudouridine and inosine with *N*-cyclohexyl-*N*'-β-(4-methylmorpholinium) ethylcarbodiimide. *Biochemistry* **10**, 3651–3657 (1971).

220. Lovejoy, A. F., Riordan, D. P. & Brown, P. O. Transcriptome-wide mapping of pseudouridines: pseudouridine synthases modify specific mRNAs in *S. cerevisiae*. *PLoS One* **9**, e110799 (2014).

221. Li, X. et al. Chemical pulldown reveals dynamic pseudouridylation of the mammalian transcriptome. *Nat. Chem. Biol.* **11**, 592–597 (2015).

222. Incarnato, D., Neri, F., Anselmi, F. & Oliviero, S. Genome-wide profiling of mouse RNA secondary structures reveals key features of the mammalian transcriptome. *Genome Biol.* **15**, 491 (2014).

223. Wang, P. Y., Sexton, A. N., Culligan, W. J. & Simon, M. D. Carbodiimide reagents for the chemical probing of RNA structure in cells. *RNA* **25**, 135–146 (2019).

224. Singhal, R. P. Chemical probe of structure and function of transfer ribonucleic acids. *Biochemistry* **13**, 2924–2932 (1974).

225. Everett, D. W. Part I: Reaction of pseudouridine with bisulfite. Part II: Reaction of glyoxal with guanine derivatives: A spectrophotometric probe of molecular structure. (New York University, 1980).

226. Khoddami, V. et al. Transcriptome-wide profiling of multiple RNA modifications simultaneously at single-base resolution. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 6784–6789 (2019).

227. Fleming, A. M. et al. Structural elucidation of bisulfite adducts to pseudouridine that result in deletion signatures during reverse transcription of RNA. *J. Am. Chem. Soc.* **141**, 16450–16460 (2019).

228. Dai, Q. et al. Quantitative sequencing using BID-seq uncovers abundant pseudouridines in mammalian mRNA at base resolution. *Nat. Biotechnol.* **41**, 344–354 (2023).

229. Zhang, M. et al. Quantitative profiling of pseudouridylation landscape in the human transcriptome. *Nat. Chem. Biol.* **19**, 1185–1195 (2023).

230. Zhang, L.-S. et al. BID-seq for transcriptome-wide quantitative sequencing of mRNA pseudouridine at base resolution. *Nat. Protoc.* **19**, 517–538 (2024).

231. Chambers, R. W., Kurkov, V. & Shapiro, R. The chemistry of pseudouridine. Synthesis of pseudouridine-5'-diphosphate. *Biochemistry* **2**, 1192–1203 (1963).

232. Chambers, R. W. The chemistry of pseudouridine. IV. Cyanoethylation. *Biochemistry* **4**, 219–226 (1965).

233. Yoshida, M. & Ukita, T. Modification of nucleosides and nucleotides: VIII. The reaction rates of pseudouridine residues with acrylonitrile and its relation to the secondary structure of transfer ribonucleic acid. *Biochim. Biophys. Acta.* **157**, 466–475 (1968).

234. Knutson, S. D., Ayele, T. M. & Heemstra, J. M. Chemical labeling and affinity capture of inosine-containing RNAs using acrylamidofluorescein. *Bioconjug. Chem.* **29**, 2899–2903 (2018).

235. Emmerechts, G., Herdewijn, P. & Rozenski, J. Pseudouridine detection improvement by derivatization with methyl vinyl sulfone and capillary HPLC–mass spectrometry. *J. Chromatogr. B* **825**, 233–238 (2005).

236. Mengel-Jorgensen, J. & Kirpekar, F. Detection of pseudouridine and other modifications in tRNA by cyanoethylation and MALDI mass spectrometry. *Nucleic Acids Res.* **30**, e135 (2002).

237. Singer, B. & Dosanjh, M. K. Site-directed mutagenesis for quantitation of base-base interactions at defined sites. *Mutat. Res.* **233**, 45–51 (1990).

238. Ofengand, J. Ribosomal RNA pseudouridines and pseudouridine synthases. *FEBS Lett.* **514**, 17–25 (2002).

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

240. Lestrade, L. & Weber, M. J. snoRNA-LBME-db, a comprehensive database of human H/ACA and C/D box snoRNAs. *Nucleic Acids Res.* **34**, D158–D162 (2006).

241. Jorjani, H. et al. An updated human snoRNAome. *Nucleic Acids Res.* **44**, 5068–5082 (2016).

242. Tan, K. T., Ding, L. W., Wu, C. S., Tenen, D. G. & Yang, H. Repurposing RNA sequencing for discovery of RNA modifications in clinical cohorts. *Sci. Adv.* **7**, eabd2605 (2021).

243. Yu, A. T., Ge, J. H. & Yu, Y. T. Pseudouridines in spliceosomal snRNAs. *Protein & Cell* **2**, 712–725 (2011).

244. Yamaki, Y. et al. Direct determination of pseudouridine in RNA by mass spectrometry coupled with stable isotope labeling. *Anal. Chem.* **92**, 11349–11356 (2020).

245. Massenet, S. & Branlant, C. A limited number of pseudouridine residues in the human atac spliceosomal UsnRNAs as compared to human major spliceosomal UsnRNAs. *RNA* **5**, 1495–1503 (1999).

246. Deryusheva, S., Choleza, M., Barbarossa, A., Gall, J. G. & Bordonne, R. Post-transcriptional modification of spliceosomal RNAs is normal in SMN-deficient cells. *RNA* **18**, 31–36 (2012).

247. Padgett, R. A. & Shukla, G. C. A revised model for U4atac/U6atac snRNA base pairing. *RNA* **8**, 125–128 (2002).

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

249. Nagpal, N. & Agarwal, S. Telomerase RNA processing: implications for human health and disease. *Stem Cells* **38**, 1532–1543 (2020).

250. Kim, N. K., Theimer, C. A., Mitchell, J. R., Collins, K. & Feigon, J. Effect of pseudouridylation on the structure and activity of the catalytically essential P6.1 hairpin in human telomerase RNA. *Nucleic Acids Res.* **38**, 6746–6756 (2010).

251. Zhang, Q., Kim, N. K. & Feigon, J. Architecture of human telomerase RNA. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 20325–20332 (2011).

252. de Crecy-Lagard, V. et al. Matching tRNA modifications in humans to their known and predicted enzymes. *Nucleic Acids Res.* **47**, 2143–2159 (2019).

253. Song, J. et al. Differential roles of human PUS10 in miRNA processing and tRNA pseudouridylation. *Nat. Chem. Biol.* **16**, 160–169 (2020).

254. Suzuki, T. et al. Complete chemical structures of human mitochondrial tRNAs. *Nat. Commun.* **11**, 4269 (2020).

255. Sprinzl, M., Horn, C., Brown, M., Ioudovitch, A. & Steinberg, S. Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res.* **26**, 148–153 (1998).

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

257. Safra, M., Nir, R., Farouq, D., Slutzkin, I. V. & Schwartz, S. TRUB1 is the predominant pseudouridine synthase acting on mammalian mRNA via a predictable and conserved code. *Genome Res.* **27**, 393–406 (2017).

258. Karijolich, J. & Yu, Y. T. Converting nonsense codons into sense codons by targeted pseudouridylation. *Nature* **474**, 395–398 (2011).

259. Sakurai, M., Yano, T., Kawabata, H., Ueda, H. & Suzuki, T. Inosine cyanoethylation identifies A-to-I RNA editing sites in the human transcriptome. *Nat. Chem. Biol.* **6**, 733–740 (2010).

260. Sakurai, M. et al. A biochemical landscape of A-to-I RNA editing in the human brain transcriptome. *Genome Res.* **24**, 522–534 (2014).

261. Safra, M. et al. The m^1A landscape on cytosolic and mitochondrial mRNA at single-base resolution. *Nature* **551**, 251–255 (2017).

262. Watkins, C. P., Zhang, W., Wylder, A. C., Katanski, C. D. & Pan, T. A multiplex platform for small RNA sequencing elucidates multifaceted tRNA stress response and translational regulation. *Nat. Commun.* **13**, 2491 (2022).

263. Behrens, A., Rodschinka, G. & Nedialkova, D. D. High-resolution quantitative profiling of tRNA abundance and modification status in eukaryotes by mim-tRNAseq. *Mol. Cell* **81**, 1802–1815 (2021).

264. Becker, H. F., Motorin, Y., Planta, R. J. & Grosjean, H. The yeast gene YNL292w encodes a pseudouridine synthase (Pus4) catalyzing the formation of Ψ_{55} in both mitochondrial and cytoplasmic tRNAs. *Nucleic Acids Res.* **25**, 4493–4499 (1997).

265. Mukhopadhyay, S., Deogharia, M. & Gupta, R. Mammalian nuclear TRUB1, mitochondrial TRUB2, and cytoplasmic PUS10 produce conserved pseudouridine 55 in different sets of tRNA. *RNA* **27**, 66–79 (2021).

266. Behm-Ansmant, I. et al. The *Saccharomyces cerevisiae* U2 snRNA: pseudouridine-synthase Pus7p is a novel multisite-multisubstrate RNA: Ψ-synthase also acting on tRNAs. *RNA* **9**, 1371–1382 (2003).

267. Carlile, T. M. et al. mRNA structure determines modification by pseudouridine synthase 1. *Nat. Chem. Biol.* **15**, 966–974 (2019).

268. Behm-Ansmant, I. et al. A previously unidentified activity of yeast and mouse RNA: pseudouridine synthases 1 (Pus1p) on tRNAs. *RNA* **12**, 1583–1593 (2006).

269. Phizicky, E. M. & Hopper, A. K. tRNA biology charges to the front. *Genes Dev.* **24**, 1832–1860 (2010).

270. Phizicky, E. M. & Hopper, A. K. The life and times of a tRNA. *RNA* **29**, 898–957 (2023).

271. Hopper, A. K. Transfer RNA post-transcriptional processing, turnover, and subcellular dynamics in the yeast *Saccharomyces cerevisiae*. *Genetics* **194**, 43–67 (2013).

272. Li, J., Zhu, W.-Y., Yang, W.-Q., Li, C.-T. & Liu, R.-J. The occurrence order and cross-talk of different tRNA modifications. *Sci. China Life Sci.* **64**, 1423–1436 (2021).

273. Johnson, P. F. & Abelson, J. The yeast $tRNA^{Tyr}$ gene intron is essential for correct modification of its tRNA product. *Nature* **302**, 681–687 (1983).

274. Simos, G. et al. Nuclear pore proteins are involved in the biogenesis of functional tRNA. *EMBO J.* **15**, 2270–2284 (1996).

275. Haag, S. et al. NSUN6 is a human RNA methyltransferase that catalyzes formation of m^5C72 in specific tRNAs. *RNA* **21**, 1532–1543 (2015).

276. Long, T. et al. Sequence-specific and shape-selective RNA recognition by the human RNA 5-methylcytosine methyltransferase NSun6. *J. Biol. Chem.* **291**, 24293–24303 (2016).

277. Karikó, K., Buckstein, M., Ni, H. P. & Weissman, D. Suppression of RNA recognition by Toll-like receptors: the impact of nucleoside modification and the evolutionary origin of RNA. *Immunity* **23**, 165–175 (2005).

278. Fleming, A. M., Mathewson, N. J., Manage, S. A. H. & Burrows, C. J. Nanopore dwell time analysis permits sequencing and conformational assignment of pseudouridine in SARS-CoV-2. *ACS Cent. Sci.* **7**, 1707–1717 (2021).

279. Giambruno, R. et al. Unveiling the role of PUS7-mediated pseudouridylation in host protein interactions specific for the SARS-CoV-2 RNA genome. *Mol. Ther. Nucleic Acids* **34** (2023).

280. Wing, P. A. C. et al. Hypoxia inducible factors regulate infectious SARS-CoV-2, epithelial damage and respiratory symptoms in a hamster COVID-19 model. *PLoS Pathog.* **18** (2022).

281. Moss, W. N., Lee, N., Pimienta, G. & Steitz, J. A. RNA families in Epstein-Barr virus. *RNA Biol.* **11**, 10–17 (2014).

282. Lee, N. The many ways Epstein-Barr virus takes advantage of the RNA tool kit. *RNA Biol.* **18**, 759–766 (2021).

283. Marchand, V. et al. HydraPsiSeq: a method for systematic and quantitative mapping of pseudouridines in RNA. *Nucleic Acids Res.* **48**, e110 (2020).

284. Henry, B. A. et al. Pseudouridylation of Epstein-Barr virus noncoding RNA EBER2 facilitates lytic replication. *RNA* **28**, 1542–1552 (2022).

285. Heidelberger, C. et al. Fluorinated pyrimidines, a new class of tumour-inhibitory compounds. *Nature* **179**, 663–666 (1957).

286. Longley, D. B., Harkin, D. P. & Johnston, P. G. 5-Fluorouracil: mechanisms of action and clinical strategies. *Nat. Rev. Cancer* **3**, 330–338 (2003).

287. Pritchard, D. M., Watson, A. J. M., Potten, C. S., Jackman, A. L. & Hickman, J. A. Inhibition by uridine but not thymidine of p53-dependent intestinal apoptosis initiated by 5-fluorouracil: evidence for the involvement of RNA perturbation. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 1795–1799 (1997).

288. Pettersen, H. S. et al. UNG-initiated base excision repair is the major repair route for 5-fluorouracil in DNA, but 5-fluorouracil cytotoxicity depends mainly on RNA incorporation. *Nucleic Acids Res.* **39**, 8430–8444 (2011).

289. Chalabi-Dchar, M. et al. A novel view on an old drug, 5-fluorouracil: an unexpected RNA modifier with intriguing impact on cancer cell fate. *NAR Cancer* **3**, zcab032 (2021).

290. Liang, Y. Y. et al. CETSA interaction proteomics define specific RNA-modification pathways as key components of fluorouracil-based cancer drug cytotoxicity. *Cell Chem. Biol.* **29**, 572–585.e578 (2022).

291. Chen, J.-K. et al. An RNA damage response network mediates the lethality of 5-FU in clinically relevant tumor types. *bioRxiv*, DOI:10.1101/2023.1104.1128.538590 (2023).

292. Samuelsson, T. Interactions of transfer RNA pseudouridine synthases with RNAs substituted with fluorouracil. *Nucleic Acids Res.* **19**, 6139–6144 (1991).

293. Spedaliere, C. J. & Mueller, E. G. Not all pseudouridine synthases are potently inhibited by RNA containing 5-fluorouridine. *RNA* **10**, 192–199 (2004).

294. Gu, X., Liu, Y. & Santi, D. V. The mechanism of pseudouridine synthase I as deduced from its interaction with 5-fluorouracil-tRNA. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 14270–14275 (1999).

295. Hamilton, C. S. et al. Mechanistic investigations of the pseudouridine synthase RluA using RNA containing 5-fluorouridine. *Biochemistry* **45**, 12029–12038 (2006).

296. Miracco, E. J. & Mueller, E. G. The products of 5-fluorouridine by the action of the pseudouridine synthase TruB disfavor one mechanism and suggest another. *J. Am. Chem. Soc.* **133**, 11826–11829 (2011).

297. Veerareddygari, G. R., Singh, S. K. & Mueller, E. G. The pseudouridine synthases proceed through a glycal intermediate. *J. Am. Chem. Soc.* **138**, 7852–7855 (2016).

298. Hoang, C. & Ferré-D'Amaré, A. R. Cocrystal structure of a tRNA Ψ55 pseudouridine synthase: nucleotide flipping by an RNA-modifying enzyme. *Cell* **107**, 929–939 (2001).

299. Hoang, C. et al. Crystal structure of pseudouridine synthase RluA: indirect sequence readout through protein-induced RNA structure. *Mol. Cell* **24**, 535–545 (2006).

300. Czudnochowski, N. et al. The mechanism of pseudouridine synthases from a covalent complex with RNA, and alternate specificity for U2605 versus U2604 between close homologs. *Nucleic Acids Res.* **42**, 2037–2048 (2013).

301. Bowman, J. C., Petrov, A. S., Frenkel-Pinter, M., Penev, P. I. & Williams, L. D. Root of the tree: the significance, evolution, and origins of the ribosome. *Chem. Rev.* **120**, 4848–4878 (2020).

302. Hug, L. A. et al. A new view of the tree of life. *Nat. Microbiol.* **1**, 16048 (2016).

303. Kuhn, J. F., Tran, E. J. & Maxwell, E. S. Archaeal ribosomal protein L7 is a functional homolog of the eukaryotic 15.5kD/Snu13p snoRNP core protein. *Nucleic Acids Res.* **30**, 931–941 (2002).

304. Watanabe, Y.-i. & Gray, M. W. Evolutionary appearance of genes encoding proteins associated with box H/ACA snoRNAs: Cbf5p in *Euglena gracilis*, an early diverging eukaryote, and candidate Gar1p and Nop10p homologs in archaebacteria. *Nucleic Acids Res.* **28**, 2342–2352 (2000).

305. Roger, A. J., Muñoz-Gómez, S. A. & Kamikawa, R. The origin and diversification of mitochondria. *Curr. Biol.* **27**, R1177–R1192 (2017).

306. Hamma, T. & Ferré-D'Amaré, A. R. Pseudouridine synthases. *Chem. Biol.* **13**, 1125–1135 (2006).

307. Niu, Y. & Liu, L. RNA pseudouridine modification in plants. *J. Exp. Bot.* **74**, 6431–6447 (2023).

308. Fitzek, E., Joardar, A., Gupta, R. & Geisler, M. Evolution of eukaryal and archaeal pseudouridine synthase Pus10. *J. Mol. Evol.* **86**, 77–89 (2018).

309. Busch, J. D. et al. MitoRibo-Tag mice provide a tool for *in vivo* studies of mitoribosome composition. *Cell Rep.* **29**, 1728–1738.e1729 (2019).

310. Jaskolowski, M. et al. Structural insights into the mechanism of mitoribosomal large subunit biogenesis. *Mol. Cell* **79**, 629–644.e624 (2020).

311. Aza-Blanc, P. et al. Identification of modulators of TRAIL-induced apoptosis via RNAi-based phenotypic screening. *Mol. Cell* **12**, 627–637 (2003).

312. Jana, S., Hsieh, A. C. & Gupta, R. Reciprocal amplification of caspase-3 activity by nuclear export of a putative human RNA-modifying protein, PUS10 during TRAIL-induced apoptosis. *Cell Death Dis.* **8**, e3093 (2017).

313. Keffer-Wilkes, L. C., Veerareddygari, G. R. & Kothe, U. RNA modification enzyme TruB is a tRNA chaperone. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 14306–14311 (2016).

314. Arroyo, J. D. et al. A genome-wide CRISPR death screen identifies genes essential for oxidative phosphorylation. *Cell Metab.* **24**, 875–885 (2016).

315. Perez-Garcia, V. et al. Placentation defects are highly prevalent in embryonic lethal mouse mutants. *Nature* **555**, 463–468 (2018).

316. Dong, C. et al. A genome-wide CRISPR-Cas9 knockout screen identifies essential and growth-restricting genes in human trophoblast stem cells. *Nat. Commun.* **13**, 2548 (2022).

317. Antonicka, H. et al. A pseudouridine synthase module is essential for mitochondrial protein synthesis and cell viability. *EMBO Rep.* **18**, 28–38 (2017).

318. Pei, J., Zhang, J. & Cong, Q. Human mitochondrial protein complexes revealed by large-scale coevolution analysis and deep learning-based structure modeling. *Bioinformatics* **38**, 4301–4311 (2022).

319. Zaganelli, S. et al. The pseudouridine synthase RPUSD4 is an essential component of mitochondrial RNA granules. *J. Biol. Chem.* **292**, 4519–4532 (2017).

320. Breznak, S. M. et al. H/ACA snRNP–dependent ribosome biogenesis regulates translation of polyglutamine proteins. *Sci. Adv.* **9**, eade5492 (2023).

321. Lafontaine, D. L. J., Bousquet-Antonelli, C., Henry, Y., Caizergues-Ferrer, M. & Tollervey, D. The box H+ACA snoRNAs carry Cbf5p, the putative rRNA pseudouridine synthase. *Genes Dev.* **12**, 527–537 (1998).

322. Zacchini, F., Barozzi, C., Venturi, G. & Montanaro, L. How snoRNAs can contribute to cancer at multiple levels. *NAR Cancer* **6**, zcae005 (2024).

323. Kiss, T., Fayet, E., Jády, B. E., Richard, P. & Weber, M. Biogenesis and intranuclear trafficking of human box C/D and H/ACA RNPs. *Cold Spring Harb. Symp. Quant. Biol.* **71**, 407–417 (2006).

324. Kiss, T., Fayet-Lebaron, E. & Jády, B. E. Box H/ACA small ribonucleoproteins. *Mol. Cell* **37**, 597–606 (2010).

325. Watkins, N. J. & Bohnsack, M. T. The box C/D and H/ACA snoRNPs: key players in the modification, processing and the dynamic folding of ribosomal RNA. *Wiley Interdiscip. Rev. RNA* **3**, 397–414 (2012).

326. Lewis, J. D. & Tollervey, D. Like attracts like: getting RNA processing together in the nucleus. *Science* **288**, 1385–1389 (2000).

327. Henras, A. K. et al. The post-transcriptional steps of eukaryotic ribosome biogenesis. *Cell. Mol. Life Sci.* **65**, 2334–2359 (2008).

328. Henras, A. K., Plisson-Chastang, C., O'Donohue, M.-F., Chakraborty, A. & Gleizes, P.-E. An overview of pre-ribosomal RNA processing in eukaryotes. *Wiley Interdiscip. Rev. RNA* **6**, 225–242 (2015).

329. Will, C. L. & Lührmann, R. Spliceosomal UsnRNP biogenesis, structure and function. *Curr. Opin. Cell Biol.* **13**, 290–301 (2001).

330. Boisvert, F.-M., van Koningsbruggen, S., Navascués, J. & Lamond, A. I. The multifunctional nucleolus. *Nat. Rev. Mol. Cell Biol.* **8**, 574–585 (2007).

331. Morris, G. E. The Cajal body. *Biochim. Biophys. Acta* **1783**, 2108–2115 (2008).

332. Corman, A., Sirozh, O., Lafarga, V. & Fernandez-Capetillo, O. Targeting the nucleolus as a therapeutic strategy in human disease. *Trends Biochem. Sci.* **48**, 274–287 (2023).

333. Gall, J. G. The centennial of the Cajal body. *Nat. Rev. Mol. Cell Biol.* **4**, 975–980 (2003).

334. Holzmann, J. et al. RNase P without RNA: identification and functional reconstitution of the human mitochondrial tRNA processing enzyme. *Cell* **135**, 462–474 (2008).

335. Bhatta, A., Dienemann, C., Cramer, P. & Hillen, H. S. Structural basis of RNA processing by human mitochondrial RNase P. *Nat. Struct. Mol. Biol.* **28**, 713–723 (2021).

336. Meynier, V. et al. Structural basis for human mitochondrial tRNA maturation. *Nat. Commun.* **15**, 4683 (2024).

337. Cámara, Y. et al. MTERF4 regulates translation by targeting the methyltransferase NSUN4 to the mammalian mitochondrial ribosome. *Cell Metab.* **13**, 527–539 (2011).

338. Spåhr, H., Habermann, B., Gustafsson, C. M., Larsson, N.-G. & Hallberg, B. M. Structure of the human MTERF4–NSUN4 protein complex that regulates mitochondrial ribosome biogenesis. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 15253–15258 (2012).

339. Yakubovskaya, E. et al. Structure of the essential MTERF4:NSUN4 protein complex reveals how an MTERF protein collaborates to facilitate rRNA modification. *Structure* **20**, 1940–1947 (2012).

340. Metodiev, M. D. et al. NSUN4 is a dual function mitochondrial protein required for both methylation of 12S rRNA and coordination of mitoribosomal assembly. *PLoS Genet.* **10**, e1004110 (2014).

341. Pérez-Arellano, I., Gallego, J. & Cervera, J. The PUA domain – a structural and functional overview. *FEBS J.* **274**, 4972–4984 (2007).

342. Andersen, N. M. & Douthwaite, S. YebU is a m^5C methyltransferase specific for 16 S rRNA nucleotide 1407. *J. Mol. Biol.* **359**, 777–786 (2006).

343. Hallberg, B. M. et al. The structure of the RNA m^5C methyltransferase YebU from *Escherichia coli* reveals a C-terminal RNA-recruiting PUA domain. *J. Mol. Biol.* **360**, 774–787 (2006).

344. Zucchini, C. et al. The human *TruB* family of pseudouridine synthase genes, including the *Dyskeratosis Congenita 1* gene and the novel member *TRUB1*. *Int. J. Mol. Med.* **11**, 697–704 (2003).

345. Martinez, N. M. et al. Pseudouridine synthases modify human pre-mRNA co-transcriptionally and affect pre-mRNA processing. *Mol. Cell* **82**, 645–659 (2022).

346. Koš, M. & Tollervey, D. Yeast pre-rRNA processing and modification occur cotranscriptionally. *Mol. Cell* **37**, 809–820 (2010).

347. Tang, Z., Kang, B., Li, C., Chen, T. & Zhang, Z. GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic Acids Res.* **47**, W556–W560 (2019).

348. Pelletier, J., Thomas, G. & Volarević, S. Ribosome biogenesis in cancer: new players and therapeutic avenues. *Nat. Rev. Cancer* **18**, 51–63 (2018).

349. Bradley, R. K. & Anczuków, O. RNA splicing dysregulation and the hallmarks of cancer. *Nat. Rev. Cancer* **23**, 135–155 (2023).

350. Harley, C. B. Telomerase and cancer therapeutics. *Nat. Rev. Cancer* **8**, 167–179 (2008).

351. Alawi, F. & Lee, M. N. DKC1 is a direct and conserved transcriptional target of c-MYC. *Biochem. Biophys. Res. Commun.* **362**, 893–898 (2007).

352. GTEx Portal – DKC1. https://gtexportal.org/home/gene/DKC1.

353. Mottram, T. J., Harper, K. L., Vasconcelos, E. J. R., Anene, C. A. & Whitehouse, A. Pseudouridine prevalence in Kaposi's sarcoma-associated herpesvirus transcriptome reveals an essential mechanism for viral replication. *bioRxiv*, DOI:10.1101/2023.1101.1131.526461 (2023).

354. Kirwan, M. & Dokal, I. Dyskeratosis congenita, stem cells and telomeres. *Biochim. Biophys. Acta* **1792**, 371–379 (2009).

355. Tummala, H., Walne, A. & Dokal, I. The biology and management of dyskeratosis congenita and related disorders of telomeres. *Expert Rev. Hematol.* **15**, 685–696 (2022).

356. Nachmani, D. et al. Germline *NPM1* mutations lead to altered rRNA 2'-O-methylation and cause dyskeratosis congenita. *Nat. Genet.* **51**, 1518–1529 (2019).

357. Yen Moore, A. Clinical applications for topical 5-fluorouracil in the treatment of dermatological disorders. *J. Dermatol. Treat.* **20**, 328–335 (2009).

358. Shah, V. V. et al. 5-Fluorouracil in the treatment of keloids and hypertrophic scars: a comprehensive review of the literature. *Dermatol. Ther. (Heidelb)* **6**, 169–183 (2016).

359. Cancer Research UK – Fluorouracil (5FU). https://www.cancerresearchuk.org/about-cancer/treatment/drugs/fluorouracil.

360. Diasio, R. B. & Harris, B. E. Clinical pharmacology of 5-fluorouracil. *Clin. Pharmacokinet.* **16**, 215–237 (1989).

361. van Kuilenburg, A. B. P. Dihydropyrimidine dehydrogenase and the efficacy and toxicity of 5-fluorouracil. *Eur. J. Cancer* **40**, 939–950 (2004).

362. Karikó, K. et al. Incorporation of pseudouridine into mRNA yields superior nonimmunogenic vector with increased translational capacity and biological stability. *Mol. Ther.* **16**, 1833–1840 (2008).

363. Pardi, N., Hogan, M. J., Porter, F. W. & Weissman, D. mRNA vaccines — a new era in vaccinology. *Nat. Rev. Drug Discov.* **17**, 261–279 (2018).

364. Chaudhary, N., Weissman, D. & Whitehead, K. A. mRNA vaccines for infectious diseases: principles, delivery and clinical translation. *Nat. Rev. Drug Discov.* **20**, 817–838 (2021).

365. Weber, J. S. et al. Individualised neoantigen therapy mRNA-4157 (V940) plus pembrolizumab versus pembrolizumab monotherapy in resected melanoma (KEYNOTE-942): a randomised, phase 2b study. *Lancet* **403**, 632–644 (2024).

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

367. Eisenberg, E. & Levanon, E. Y. A-to-I RNA editing — immune protector and transcriptome diversifier. *Nat. Rev. Genet.* **19**, 473–490 (2018).

368. Hemmi, H. et al. A Toll-like receptor recognizes bacterial DNA. *Nature* **408**, 740–745 (2000).

369. Lafaille, F. G. et al. Human *SNORA31* variations impair cortical neuron-intrinsic immunity to HSV-1 and underlie herpes simplex encephalitis. *Nat. Med.* **25**, 1873–1884 (2019).

370. Muller, C. A. et al. Capturing the dynamics of genome replication on individual ultra-long nanopore sequence reads. *Nat. Methods* **16**, 429–436 (2019).

371. Chen, S. F., Zhou, Y. Q., Chen, Y. R. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, 884–890 (2018).

372. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

373. Dunham, I. et al. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).

374. Amemiya, H. M., Kundaje, A. & Boyle, A. P. The ENCODE blacklist: identification of problematic regions of the genome. *Sci. Rep.* **9**, 9354 (2019).

375. Incarnato, D., Krepelova, A. & Neri, F. High-throughput single nucleotide variant discovery in E14 mouse embryonic stem cells provides a new reference genome assembly. *Genomics* **104**, 121–127 (2014).

376. Rosenbloom, K. R. et al. The UCSC Genome Browser database: 2015 update. *Nucleic Acids Res.* **43**, D670–D681 (2015).

377. Zhao, H. et al. CrossMap: a versatile tool for coordinate conversion between genome assemblies. *Bioinformatics* **30**, 1006–1007 (2014).

378. Bogu, G. K. et al. Chromatin and RNA maps reveal regulatory long noncoding RNAs in mouse. *Mol. Cell. Biol.* **36**, 809–819 (2016).

379. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

380. Ramirez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).

381. Kent, W. J., Zweig, A. S., Barber, G., Hinrichs, A. S. & Karolchik, D. BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics* **26**, 2204–2207 (2010).

382. Robinson, J. T. et al. Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).

383. Ran, F. A. et al. Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, 2281–2308 (2013).

384. Wing, P. A. C. et al. Hypoxic and pharmacological activation of HIF inhibits SARS-CoV-2 infection of lung epithelial cells. *Cell Rep.* **35**, 109020 (2021).

385. Zhuang, X. D. et al. The circadian clock components BMAL1 and REV-ERBα regulate flavivirus replication. *Nat. Commun.* **10**, 377 (2019).

386. Sureau, C. The use of hepatocytes to investigate HDV infection: the HDV/HepaRG model. *Methods Mol. Biol.* **640**, 463–473 (2010).

387. Sanz, M. A. & Carrasco, L. Sindbis virus variant with a deletion in the 6K gene shows defects in glycoprotein processing and trafficking: lack of complementation by a wild-type 6K gene in *trans*. *J. Virol.* **75**, 7778–7784 (2001).

388. Castello, A. et al. Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* **149**, 1393–1406 (2012).

389. Castello, A. et al. System-wide identification of RNA-binding proteins by interactome capture. *Nat. Protoc.* **8**, 491–500 (2013).

390. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12 (2011).

391. Smith, T., Heger, A. & Sudbery, I. UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res.* **27**, 491–499 (2017).

392. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

393. Chan, P. P. & Lowe, T. M. GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. *Nucleic Acids Res.* **44**, D184–D189 (2016).

394. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

395. Kuo, M. Y. P., Chao, M. & Taylor, J. Initiation of replication of the human hepatitis delta virus genome from cloned DNA: role of delta antigen. *J. Virol.* **63**, 1945–1950 (1989).

396. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

397. Suzuki, T., Ueda, H., Okada, S. & Sakurai, M. Transcriptome-wide identification of adenosine-to-inosine editing using the ICE-seq method. *Nat. Protoc.* **10**, 715–732 (2015).

398. Sherry, S. T. et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* **29**, 308–311 (2001).

399. Weinberg, Z. & Breaker, R. R. R2R - software to speed the depiction of aesthetic consensus RNA secondary structures. *BMC Bioinformatics* **12**, 3 (2011).

400. Bergeron, D. et al. snoDB 2.0: an enhanced interactive database, specializing in human snoRNAs. *Nucleic Acids Res.* **51**, D291–D296 (2022).

401. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).

402. Kuleshov, M. V. et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–W97 (2016).

403. Wagih, O. ggseqlogo: a versatile R package for drawing sequence logos. *Bioinformatics* **33**, 3645–3647 (2017).