

Investigation/Characterization of base modifications in mammalian DNA

Dissertation

Zur Erlangung des Grades

Doktor der Naturwissenschaften

Am Fachbereich Biologie

Der Johannes Gutenberg-Universität Mainz

Anne Kijonka-Baumgärtner

geb. am 10.12.1991 in Mainz

Mainz, 2021

Table of Contents

1. Summary.....	1
2. Zusammenfassung.....	2
3. Introduction.....	3
3.1. Nucleic acid base modifications	3
3.1.1. DNA cytosine methylation and demethylation in mammalian DNA.....	5
3.1.2. N6-adenine methylation in vertebrate DNA and RNA	8
3.1.3. Adenosine deamination in mammalian RNA and DNA	9
3.2. DNA:RNA hybrids in the mammalian genome	13
3.3. LC-MS/MS for detection of nucleic acid modifications	15
3.4. Aim of the thesis.....	19
4. Results & Discussion.....	20
4.1. Screening for novel DNA base modifications	20
4.1.1. Gain-of-function screening for enzymatically regulated DNA modifications.....	20
4.1.2. Screen for novel sulfur related DNA modifications.....	31
4.2.1. Screen for R-loop enriched DNA modifications.....	35
4.3. ADAR1 acts as DNA editor in DNA:RNA hybrids.....	40
4.3.1. ADAR1 deaminates R-loops <i>in vivo</i>	40
4.3.2. ADAR1 catalytic activity regulates R-loop levels	47
4.3.3. DNA editing by ADAR possibly regulates R-loop stability and clearance	50
4.4. Origin of m ⁶ dA in mammalian DNA.....	54
5. Material and Methods.....	59
5.1. Material	59
5.1.1. Equipment	59
5.1.2. Chemicals and pre-made buffers	59
5.1.3. Enzymes.....	60
5.1.4. Kits	61

5.1.5.	Buffers and solutions.....	61
5.1.6.	Antibodies.....	62
5.1.7.	Oligonucleotide sequences	62
5.1.8.	siRNAs.....	63
5.1.9.	Plasmids.....	64
5.1.10.	Nucleoside and nucleotide standards	65
5.1.11.	Software	66
5.2.	Methods	66
5.2.1.	Molecular biology.....	66
5.2.2.	Cell culture.....	72
5.2.3.	HPLC-MS.....	75
6.	References.....	84
7.	List of Abbreviations.....	104
8.	Acknowledgements	108
9.	Lebenslauf	109

1. Summary

Deoxyribonucleic acid (DNA) is the macromolecule carrying the majority of heritable information. Beyond the sequence of the four canonical nucleotides, chemical modifications can add another layer of information. In mammalian DNA, so far five enzymatically introduced modifications have been described. DNA modifications can play major biological roles in transcription regulation, chromatin structure and accessibility. In this study, I explored unknown, epigenetic DNA modifications and investigated the regulation and biological functions of N6-methylated deoxyadenosine (m^6dA) and deoxyinosine (dI) in mammalian DNA.

In the first part of my thesis, to identify and characterize novel epigenetic DNA modifications, I used a mass spectrometry-based screening method. However, a gain-of-function screen with potential DNA modifying enzymes and a screen for sulfur-containing DNA modifications did not reveal any new base modifications in human DNA.

In the second part of my thesis, applying the screen approach on DNA:RNA hybrid regions (R-loops) in the mammalian genome showed that deoxyinosine (dI), a deamination product of deoxyadenosine (dA), is enriched within R-loops. Though dI is commonly regarded as DNA damage, its specific enrichment within R-loops suggests a biological significance of this modification. Consistently, my results indicate that adenosine deaminase acting on RNA 1 (ADAR1), an enzyme well characterized for its RNA editing activity, promoted dI formation in DNA:RNA hybrids. Although this study does not provide ultimate proof on the biological consequences, it raises the possibility that ADAR-mediated editing of DNA in DNA:RNA hybrids affects R-loop stability and clearance.

In the third part of my thesis, I contributed to investigating the role and origin of N6-methyl deoxyadenosine (m^6dA) in mammalian DNA, which remains controversial.

2. Zusammenfassung

Der Hauptanteil vererbbarer Informationen ist in Form des Makromoleküls Desoxyribonukleinsäure (DNA) gespeichert. Über die Abfolge der natürlichen Nukleotide hinaus, stellen chemische Modifikationen der Nukleotide eine weitere Informationsebene dar. In Säuger-DNA sind bisher fünf enzymatisch eingefügte Modifikationen beschrieben. DNA-Modifikation üben einen bedeutenden biologischen Einfluss aus, etwa in der RNA-Transkription und der Struktur und Zugänglichkeit von Chromatin. In dieser Studie erforschte ich bisher unbekannte DNA-Modifikationen und untersuchte die Regulation und biologischen Funktionen von N6-Methyldeoxyadenosin (m^6dA) und Desoxyinosin in Säuger-DNA.

Im ersten Teil meiner Arbeit verwendete ich ein Massenspektrometrie-basiertes Screening-Verfahren, um unbekannte, epigenetisch relevante DNA-Modifikationen zu identifizieren und zu charakterisieren. In einem Gain-of-Function Screen mit potenziell DNA-modifizierenden Enzymen, als auch einem Screen für Schwefelhaltige DNA-Modifikationen konnte ich jedoch keine neuen DNA-Modifikationen in humaner DNA identifizieren.

Im zweiten Teil meiner Arbeit, konnte ich durch Anwendung dieses Screening Verfahrens auf DNA:RNA Hybride (R-loops) im Säurgerenom zeigen, dass Desoxyinosin (dI), das Produkt der Desoxyadenosin (dA)-Desaminierung, in R-loops angereichert ist. Obwohl dI allgemein als DNA Schaden betrachtet wird, suggeriert seine spezifische Anreicherung eine biologische Rolle von dI in R-loops. Übereinstimmend damit deuten meine Ergebnisse darauf hin, dass doppelsträngige RNA-spezifische Adenosin-Desaminase 1 (ADAR1), ein Enzym welches für seine Funktion als RNA Editase bekannt ist, die Bildung von dI in DNA:RNA Hybriden fördert. Obwohl diese Studie keine unmittelbaren Beweise zu den biologischen Folgen der Präsenz von dI in R-loops erbringt, eröffnen die hier erlangten Erkenntnisse die Möglichkeit, dass ADAR1-vermittelte Editierung in R-loops deren Stabilität und Beseitigung beeinflussen kann.

Im dritten Teil meiner Arbeit war ich an Untersuchungen hinsichtlich der Rolle und des Ursprungs von N6-Methyldeoxyadenosin (m^6dA) beteiligt, welche weiterhin umstritten sind.

3. Introduction

3.1. Nucleic acid base modifications

The majority of all necessary information for operating a living organism is stored in the cell's nucleus: Deoxyribonucleic acid (DNA) contains the heritable genetic information in a code created by the four nucleotides deoxyadenosine (dA), thymidine (dT), deoxyguanosine (dG) and deoxycytosine (dC). These are building blocks of nitrogenous bases connected to a deoxyribose sugar by an N-glycosidic bond. The single blocks are connected by phosphodiester bonds to build a polyester¹⁻³. Specific hydrogen bonds between the purine and pyrimidine bases, where dA pairs with dT and dG with dC, are the basis for the heritability of the stored information. These so-called Watson-Crick base pairs allow storage and transmission of the genetic sequence in a double-helical DNA duplex of two complementary, antiparallel DNA polynucleotide stands⁴ (Figure 3.1-1). These strands are replicated in a semi-conservative manner to generate two identical DNA duplexes after each circle of replication⁵. The information stored in the DNA can be transcribed into ribonucleic acid (RNA) that consists of the corresponding ribonucleotides A, G, C and Uracil (U). In RNA, the nitrogenous base is linked to a ribose entity instead of deoxyribose. Protein coding genes are transcribed into messenger RNA (mRNA), which is transported to the ribosomes, where the code is translated to proteins⁶.

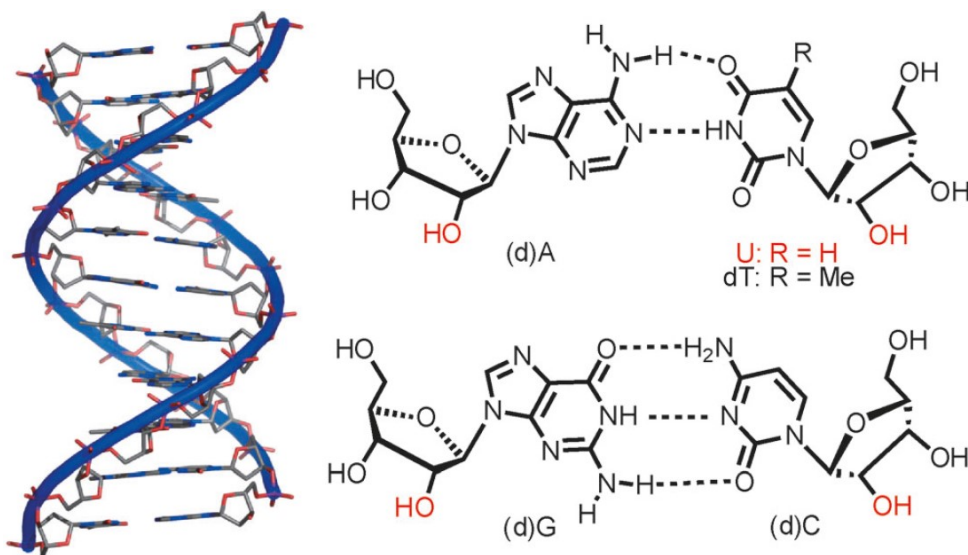


Figure 3.1-1: DNA double helix and Watson-Crick base pairs. The sequence of the antiparallel DNA polymers encodes the genetic information. The genetic code can be replicated or transcribed based on base pairing of (d)A with dT (or U in RNA) and (d)G with (d)C. Common DNA and RNA features are illustrated in black. RNA specific components are drawn in red. Figure from reference: ⁷.

Any heritable phenotypic change that is independent of the DNA or RNA sequence is described by the terms epigenetics⁸ and epitranscriptomics⁹, respectively. These changes can be conducted via chemical modification of histones, proteins that package the DNA into the chromatin, for example by acetylation^{10,11} and methylation¹². Furthermore, covalent modifications of DNA and RNA provide an additional layer of epigenetic and epitranscriptomic information.

In RNA, more than 170 naturally occurring modifications of ribonucleotides are known and some of them have been shown to play important roles in the regulation of RNA functions^{7,13}. These posttranscriptional modifications include the transfer of small chemical groups, resulting for instance in methylated¹⁴, hydroxymethylated^{15,16}, acetylated¹⁷ and thiolated^{18,19} bases. Other RNA modifications include deamination^{20,21}, reduction²², isomerization²³ or hypermodification²⁴ (Figure 3.1-2). The chemical diversity of RNA bases reflects the presence of diverse RNA species with a multitude of functions that go beyond mRNA as a carrier of protein-encoding sequences. For instance, ribosomal RNAs (rRNAs) and transferRNAs (tRNAs) are well known for their contribution to protein synthesis^{25–27}. Long noncoding RNAs (lncRNAs) and several sub-classes of short noncoding RNAs (sncRNAs) are involved in the regulation of gene expression, splicing, and many other processes^{7,28–30}.

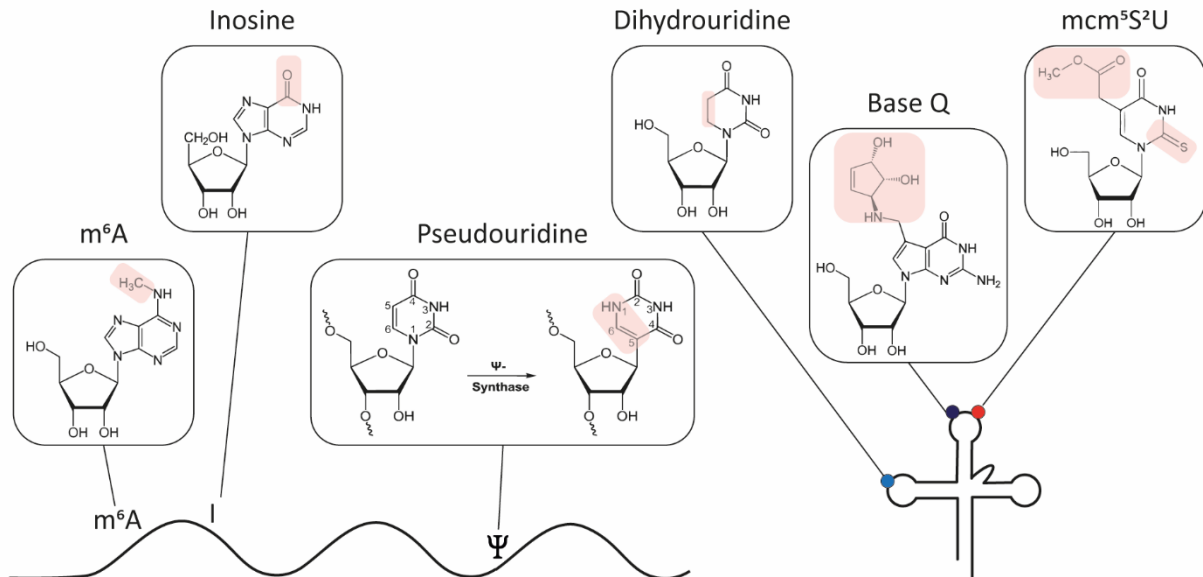


Figure 3.1-2: Base modifications in mammalian RNA. Specific modification groups and the A-to-I edited site in inosine are highlighted. Examples for RNA base modifications include methylation (N6-methyladenosine, m6A), deamination (Inosine, I), isomerization (pseudouridine, ψ), reduction (Dihydrouridine, D), hypermodification (Base Q) and thiolation (mcm⁵S²U).

In contrast to the diverse RNA epitranscriptome, DNA contains only a few acknowledged epigenetic marks. Given the function and structure of the DNA double helix as the carrier of genetic information, the little number of epigenetic DNA base modifications is plausible: The conservation of information relies on unperturbed complementary base pairing, which could be perturbed by chemical modifications, especially when they occur in the base-pairing interface. In mammals, only a hand full of epigenetic modifications have been reported or proposed: 5-methylcytosine (5mC) is involved in the regulation of gene expression by targeted silencing. For long, the 5mC-derived oxidation products 5-hydroxymethyl-, 5-formyl- and 5-carboxyl cytosine (5hmC, 5fC, 5caC)^{13,31} were believed to be merely intermediates of demethylation or even oxidative damage³²⁻³⁴. However, a growing number of studies suggests an epigenetic function of their own^{33,35-37}. In contrast, the function of N6-methyl deoxyadenosine (m⁶dA) in mammalian DNA is controversial. While many studies report its presence and regulation in mammalian DNA^{38,39}, other reports fail to detect genomic m⁶dA^{40,41} or consider it an artifact^{41,42}. This controversy highlights the challenges in the discrimination of epigenetic DNA marks from arbitrary DNA damage. Chemical changes of the DNA arising in a non-targeted fashion, for example by reaction with damaging agents, such as UV irradiation, reactive oxygen species (ROS) and nitric oxide (NO), are referred to as DNA lesions. Such DNA lesions are often mutagenic and cytotoxic, as they can alter or disrupt base pairing and therefore disturb the conservation of genetic information. Consequently, DNA damage can cause emergence of diseases, for instance cancer. To avoid accumulation of DNA lesions, cells possess various DNA repair mechanisms^{43,44}. Importantly, DNA lesions are errors that require repair, while epigenetic DNA modifications are introduced in a targeted fashion and are inherited from one generation to the next.

Beyond the known enzymatically introduced mammalian DNA modifications, other epigenetic modifications might exist in mammalian DNA. As mentioned above, it is challenging to distinguish a spurious DNA lesion from an epigenetic or regulatory DNA modification. I tackled this challenge by screening for unknown DNA modifications in three educated guess approaches. First, I performed a gain-of-function screening with potential DNA modifying enzymes. Second, in analogy to thiolated RNA modifications, I screened for sulfur-containing DNA bases by a SILAC approach. Lastly, I focused on R-loops, genomic regions that require tight regulation, possibly also via DNA modifications.

3.1.1. DNA cytosine methylation and demethylation in mammalian DNA

5-Methylcytosine (5mC) is a well-characterized epigenetic DNA base modification, typically linked to gene silencing. It has been named “the fifth base” of the genome, reflecting its relatively high abundance of 3-5% of all cytosines in mammals⁴⁵⁻⁴⁷. 5mC plays important roles in many biological processes, such as development^{48,49}, genomic imprinting⁵⁰, X-chromosome inactivation⁵¹ and

transposon silencing⁵². In eukaryotes, cytosine methylation is introduced by DNA methyltransferases (DNMTs), enzymes that modify DNA in the context of cytosine-guanine dinucleotides (CpGs) by transferring a methyl group from the cofactor S-adenosyl-L-methionine (SAM) to the 5th atom of the cytosine pyrimidine ring^{53,54}. DNMT1 in complex with Uhrf1 (Ubiquitin Like With PHD And Ring Finger Domains 1) acts as a maintenance methyltransferase by recognizing hemimethylated DNA and introducing the 5mC mark to the complementary DNA strand in a symmetric fashion, assuring the heritability of the epigenetic information^{55,56}. In addition to DNMT1, the *de novo* methyltransferases DNMT3a and DNMT3b can establish methylation at unmethylated DNA⁴⁸.

The majority of eukaryotic cell types display relatively stable CpG methylation level of 70-80%⁵⁷. In promoter regions, the methylation of CpG-rich sequences is associated with transcription repression⁵⁸. Deposition of 5mC in promoters can prevent binding of methylation-sensitive transcription factors, leading to reduced gene expression^{59,60}. The primary mechanism of 5mC-mediated gene silencing, however, is the interaction with methyl-CpG-binding proteins, like MBD2, MeCP1 or MeCP2. Upon binding to methylated DNA, these proteins recruit histone deacetylases and establish heterochromatin formation⁶¹⁻⁶³. Tissue and cell-type specific DNA methylation patterns contribute to the definition of cell identities by repressing unused parts of the genome⁵⁸. Paradoxically, CpG methylation is elevated in gene bodies of highly transcribed genes, where 5mC prevents spurious transcription initiation^{64,65}. Even though the role of 5mC in gene regulation is very important for the cell, it accounts only for a small fraction of 5mC in the genome. Most methylated CpGs are found in repetitive genomic regions, like telomeric and pericentromeric regions, satellite repeats and transposable elements. Permanent silencing of these regions is important for the maintenance of genome integrity^{66,67}. Beyond methylation of CpG sites, DNMT3a and DNMT3b can modify cytosines in non-CpG contexts, such as CpA, CpT or CpC. Methylation of these sites accounts for 0.02% of all 5mC in somatic cells⁶⁸⁻⁷⁰. Yet, the exact role of non-CpG methylation remains obscure.

The carbon-carbon bond between methyl group and pyrimidine ring in 5mC, and the maintenance of DNA methylation by DNMT1, make 5mC chemically and genetically stable. However, passive and active demethylation processes can replace 5mC by unmodified cytosine. Passive DNA demethylation requires the inhibition of maintenance methylation during cell division. Ablation of DNMT1 maintenance methyltransferase activity causes 5mC out-dilution over successive rounds of DNA replication. An example for global passive DNA demethylation is the removal of 5mC from the maternal genome during mouse preimplantation development⁷¹.

Active DNA demethylation involves the function of Ten-eleven translocation (TET) enzymes⁷². TET proteins are iron(II)- α -ketoglutarate-dependent dioxygenases that successively oxidize 5mC to 5hmC,

5fC and 5caC. On the one hand, passive depletion of these oxidized bases occurs during DNA replication. On the other hand, for active DNA demethylation, thymine DNA glycosylase (TDG) excises 5fC and 5caC, rendering an abasic site that is replaced by an unmodified cytosine by base excision repair (BER)⁷³⁻⁷⁵ (Figure 3.1-3).

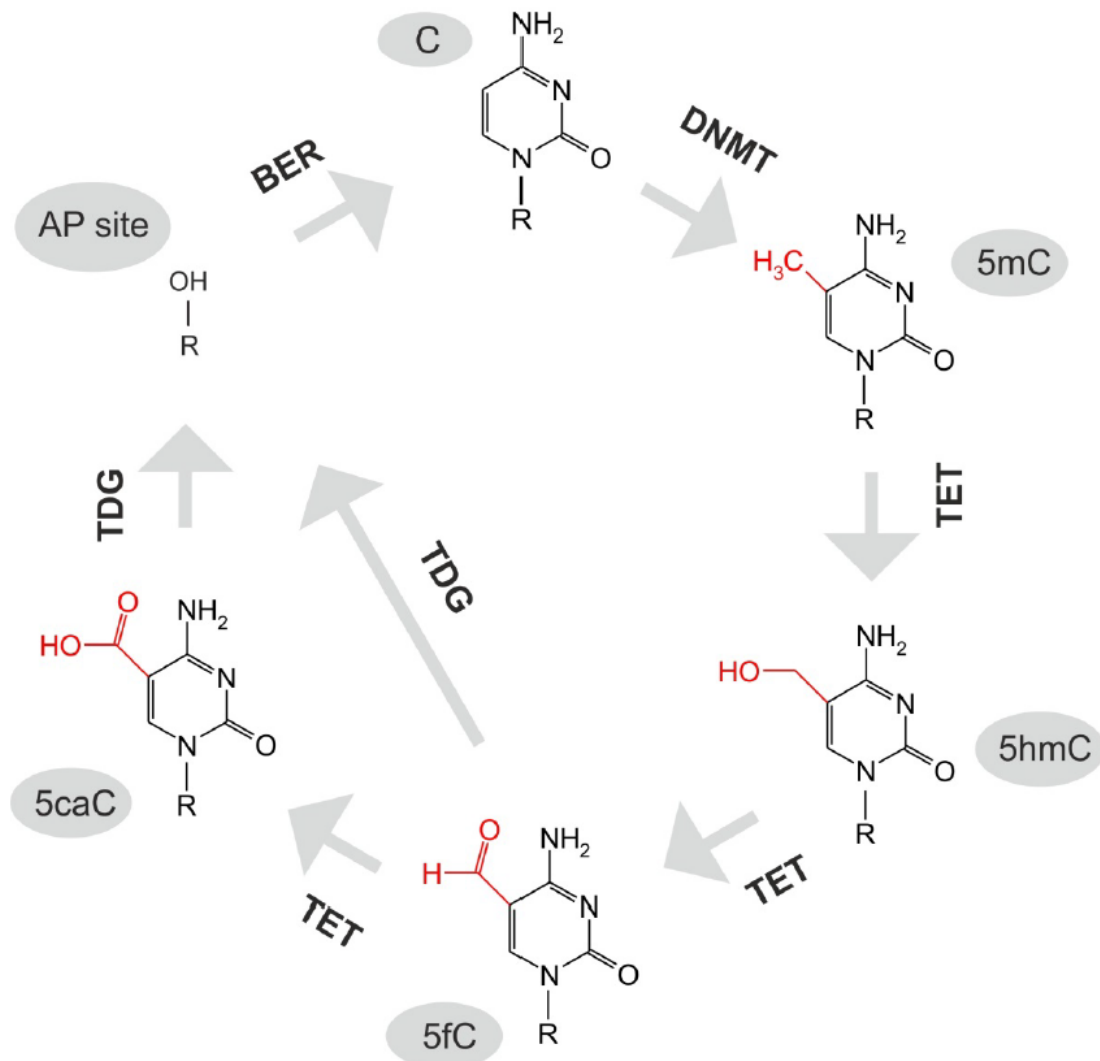


Figure 3.1-3: TET-TDG mediated active DNA demethylation. Unmodified cytosine (C) is methylated by maintenance methyltransferase DNMT1 or *de novo* methyltransferases DNMT3a/b to 5-methylcytosine (5mC). TET proteins sequentially oxidize 5mC to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC). Modified residues are illustrated in red. 5fC and 5caC are excised by TDG and the generated apyrimidinic (AP) site is repaired by base excision repair to restore the DNA with an unmodified cytosine. Figure from reference: ³¹.

While 5hmC, 5fC and 5caC were initially regarded merely as intermediates of active DNA demethylation, a growing body of evidence suggests their own biological function. The most abundant oxidized form of 5mC, 5hmC, appears to be enriched in a tissue-specific manner, with particularly high abundance in the brain^{45,76}. 5hmC also impacts epigenome functions by its differential affinity to methyl-CpG-binding proteins⁷⁷. A recent report on regulation of CA-repeat containing genes by recognition of hydroxymethylated CA-repeats by MeCP2 further supports the postulation of 5hmC as an epigenetic mark⁷⁸. In 2013, two mass spectrometry-based studies showed that there was little overlap between the reader proteins of 5mC, 5hmC, 5fC and 5caC^{79,80}. Hence, the subsets of proteins associated with the respective cytosine marks may cause modification-specific physiological impacts. Other indications for a specific function include the reduced elongation rate of RNA polymerase II (RNAPII) caused by 5fC and 5caC^{35,81}. Biophysical studies suggest that oxidized bases can influence DNA conformation and base-pairing^{81,82}, which might convey epigenetic messages. Further research will elucidate whether and how the oxidized bases exert their biological functions in the genome.

3.1.2. N6-adenine methylation in vertebrate DNA and RNA

While numerous studies established the epigenetic role of 5mC in mammalian DNA, the function of N6-adenine methylation (m⁶dA) in mammalian DNA is controversial. In prokaryotes, m⁶dA is present at levels of 1.5% of total adenine in bacteria⁸³. It protects the bacterial DNA from restriction endonuclease digest in the context of the restriction/modification system that is part of the antiviral host defense^{83,84}. While the invertebrate species *D. melanogaster* and *C. elegans* do not present 5mC in their genomes, m⁶dA was reported to exist and suggested as epigenetic mark in these species^{85,86}. A potential epigenetic role of m⁶dA in vertebrate DNA is controversially disputed. Several studies reported regulatory and potentially epigenetic roles of m⁶dA. Across different species, the reported locations of m⁶dA are variable and range from repetitive elements⁸⁷ over intergenic³⁸ to coding regions⁸⁸. Deposition of m⁶dA in the aforementioned regions correlates with active transcription. On the contrary, m⁶dA in young LINE-1 elements of mouse embryonic stem cells was associated with gene silencing during differentiation³⁹. During zebrafish embryogenesis, m⁶dA levels successively decrease, suggesting a role for m⁶dA in early development. In the zebrafish embryo, m⁶dA was located mostly in repetitive regions⁸⁷. Besides physiological consequences of m⁶dA deposition on developmental processes, it was implicated in the regulation of fear extinction in the mouse brain⁸⁹, as well as tumorigenesis⁸⁸. In human DNA, N6-adenine methyltransferase (N6AMT1) was reported as N6-adenine DNA methyltransferase. Deregulation of N6AMT1 was linked to decreasing m⁶dA levels in cancer⁸⁸. Several studies identified ALKBH1 dioxygenase as the m⁶dA demethylase in human and mouse DNA^{39,88,90,91}. While a number of studies report the presence of m⁶dA in mammalian gDNA, ultrasensitive antibody-free methods failed to detect it^{40,92}. Other researchers consider m⁶dA signals

in mammalian gDNA samples as artifact of handling or bacterial contamination^{41,42}. Ultimate evidence of the presence of m⁶dA in mammalian gDNA and its origin remain elusive.

In RNA, however, ribo-m⁶A is the most common modification in mRNA and lncRNA^{93,94}. The epitranscriptomic regulation and the physiological outcomes of ribo-m⁶A in RNA are widely studied. In mammals, a protein complex with four identified components including methyltransferase-like 3 (METTL3), METTL14, Wilms tumour 1-associated protein (WTAP) and KIAA1429 acts as the writer of ribo-m⁶A⁹⁵⁻⁹⁸. The function of ribo-m⁶A is mediated by reader proteins, including multiple YTH family members (YTHDF1-2 and YTHDC1)⁹⁹⁻¹⁰¹, heterogeneous ribonucleoprotein HNRNPA2B1¹⁰², and eukaryotic initiation factor 3 (eIF3)¹⁰³. The enzymes fat mass and obesity-associated protein (FTO)¹⁰⁴ and alkB homologue 5 (ALKBH5)¹⁰⁵ erase ribo-m⁶A from RNA. Numerous studies describe various functions of ribo-m⁶A in RNA. Generally, ribo-m⁶A destabilizes RNA, for instance by YTHDF2-mediated relocation of RNA to processing bodies (P-bodies), where the RNA is degraded¹⁰⁶. By weakening the A:U base pairing^{107,108}, ribo-m⁶A can affect the secondary structure of RNA¹⁰⁹, which, in turn, can change RNA-protein interactions. These reader protein partners convey other functions of ribo-m⁶A, including promotion of RNA translation^{103,110}, splicing efficiency^{96,102,105} and RNA translocation to the cytosol¹⁰⁵.

While the role of ribo-m⁶A as a typical epitranscriptomic mark is well established, an epigenetic function of its DNA equivalent m⁶dA in eukaryotes remains elusive. The enzymatic methylation and demethylation of adenosine in DNA and the identity of a N6-adenine DNA methyltransferase and demethylase are still under dispute and require further confirmation in order to argue for m⁶dA as an epigenetic mark in mammalian DNA.

3.1.3. Adenosine deamination in mammalian RNA and DNA

Similar to N6-methylation of adenine, the product of adenosine deamination, inosine, is known as widespread epitranscriptomic RNA modification, while no epigenetic function was described for its counterpart in DNA. The enzymatic conversion of adenosine to inosine is referred to as A-to-I editing. This reaction is catalyzed by adenosine deaminases acting on RNA (ADARs) that edit adenosines in structured and double stranded RNA (dsRNA) across all metazoan species^{21,111,112}. A-to-I editing in RNA can affect cell fate and function of RNAs by different mechanisms. The exchange of adenosine by inosine changes RNA folding and alters interactions between proteins and the edited RNA. Changes in protein-RNA interactions, in turn, can affect processing, localization and stability of RNA¹¹³. The translation machinery reads inosine as guanosine; therefore A-to-I editing in protein coding genes changes the amino acid sequence of the encoded protein and can consequently affect their function¹¹⁴. Therefore, RNA editing creates protein diversification. This variation can be tissue-specific, as reported

for editing of the protein-recoding sites of Filamin A and Filamin B, and Glutamate Ionotropic Receptor AMPA Type Subunit 2 (Gria2)^{115–118}. This reflects the role of A-to-I editing in tissue diversification and development. In addition to the structural and recoding function of inosine in RNA, A-to-I editing by ADARs can affect splicing. Accordingly, genome-wide studies investigating the effect of RNA editing on splicing, revealed that ADAR deficiency perturbs the global splicing landscape^{119–121}. Examples for site-specific regulatory connections between A-to-I editing and splicing are the AMPA receptor mRNA and the glutamate receptor B mRNAs^{122,123}.

In mammals, three highly conserved ADAR genes exist,^{124,125} ADAR1, ADAR2 and ADAR3. ADAR1 is expressed in two isoforms, the interferon-inducible ADAR1p150 (150 kDa), and the constitutively expressed ADAR1p110 (110 kDa)^{126,127}. All ADARs share a similar arrangement of functional domains. Figure 3.1-4 illustrates the domain structure for mammalian ADARs. Starting from the C-terminus, all ADARs contain a deaminase domain. However, the catalytic domain of ADAR3 is enzymatically inactive. Next to the catalytic domains, two (ADAR2 and ADAR3) or three (ADAR1) dsRNA binding domains (RBDs) are located, that interact with dsRNA and are required for substrate engagement. Towards the amino terminus of ADAR1, Z-DNA binding domains (ZBD) are located. ADAR1p110 contains one ZBD – Z β – while the ADAR1p150 isoform encloses the Z β domain and an additional Z α domain¹²⁸. All ADARs possess a nuclear localization sequence (NLS). The main location of all ADARs is the nucleus, with the exception of ADAR1p150. An additional nuclear export sequence (NES) within the Z α domain of ADAR1p150 allows shuttling of ADAR1p150 between its main location, the cytosol, and the nucleus¹²⁹.

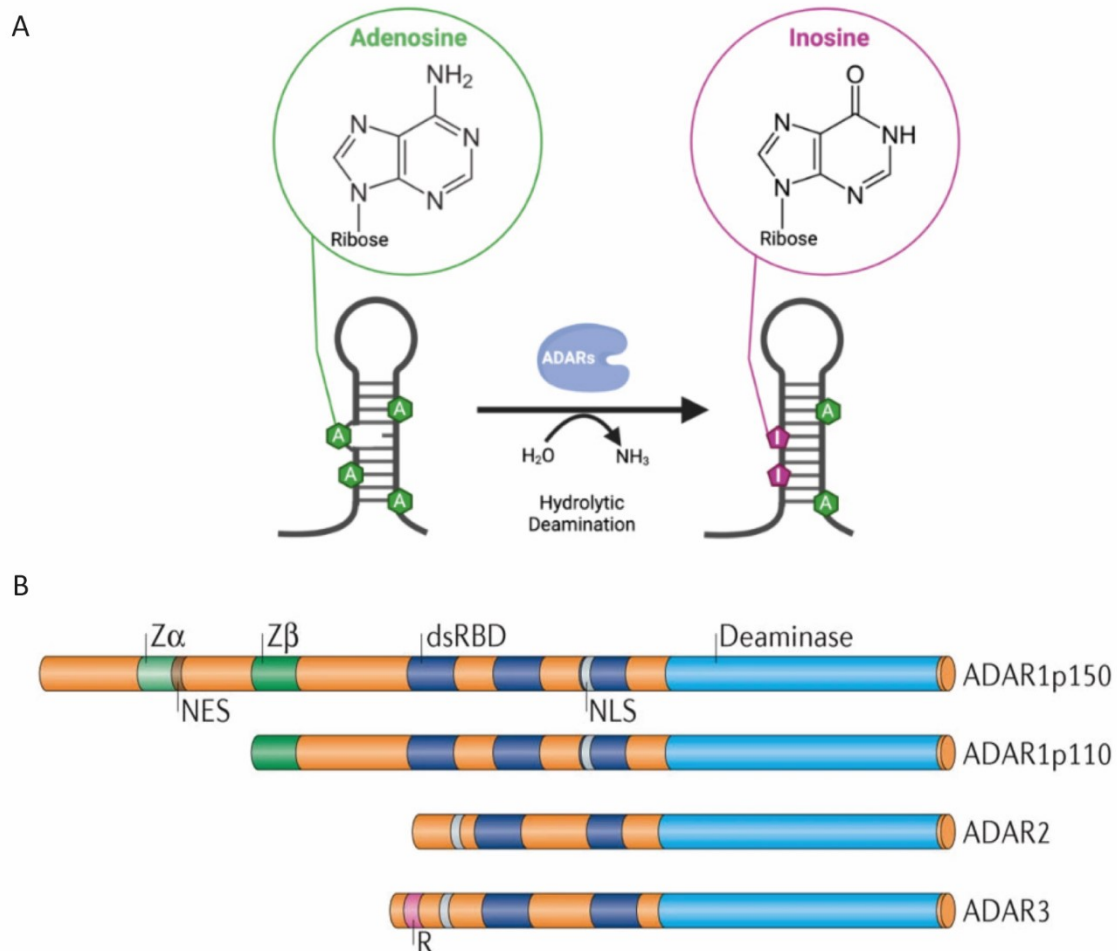


Figure 3.1-4: Activity and architecture of mammalian ADARs. (A) Adenosine deaminases acting on RNA (ADARs) deaminate adenosine to inosine via hydrolytic deamination in dsRNA. (B) The three human ADAR family members (ADAR1, ADAR2 and ADAR3) share common functional domains. These include two or three repeats of the double-stranded RNA (dsRNA)-binding domain (dsRBD) and a catalytic deaminase domain. The deaminase domain of ADAR3 is inactive. Z-DNA-binding domain α ($Z\alpha$) is unique to ADAR1p150 and $Z\beta$ is common in both ADAR1 isoforms. (Figure from references: ^{111,130})

The A-to-I editing activity of ADARs requires homodimerization, which likely allows for the proper formation of the active site^{131,132}. Crystal structures of human ADAR2 bound to dsRNA indicate that the deamination of adenosine requires base flipping of the target adenosine out of the A-form dsRNA helix¹³³. Most A-to-I editing sites are located in introns, untranslated regions and noncoding RNA. Particularly, repetitive elements, like Alu repeats from the SINE family and LINE elements, are substrates for A-to-I editing^{134–138}. Inverted repeats within these repetitive elements are prone to base-pair with themselves and form secondary structures, thereby making them good substrates for ADARs^{111,139–141}. ADAR1 was identified as the primary editing enzyme in repetitive elements, whereas ADAR2 lacks activity on those regions¹³⁷.

Both loss of ADAR1 and ADAR2 cause dramatic physiological defects. Ablation of ADAR1 causes embryonic lethality in mice; accompanied by liver disintegration and increased interferon signaling. The mechanism of this phenotype originates from ADAR1's role as a regulator of the innate immunity. The innate immune system is a nonspecific defense mechanism against foreign viral dsRNA. Viral infections increase dsRNA levels in the cytosol. Co-deletion of the RNA-sensors MDA5¹⁴² and MAVS¹⁴³ or the antiviral nuclease RNaseL¹⁴⁴ along with ADAR1 can rescue the embryonic lethality of the ADAR1-null phenotype. Further, ADAR1 inhibits the dsRNA sensor protein kinase R (PKR), thereby preventing the PKR-responsive downregulation of global protein translation¹⁴⁵. Together, these findings propose a role of ADAR1-mediated RNA editing in the regulation of innate immunity. Hereby, ribo-I serves as a self-recognition mark of endogenous RNA to discriminate from unedited viral RNA. The phenotype of ADAR2-null mutants showing early postnatal lethality and epileptic seizures is caused by the of A-to-I editing in glutamate receptor subunit 2 (Gria2) mRNA, a major substrate for ADAR2^{146,147}.

Despite its many functions in RNA, deoxyinosine (dI) in DNA is regarded as a DNA damage so far. Alike the translation machinery, the DNA replication machinery reads dI as dG. Hence, the deamination of deoxyadenosine (dA) to deoxyinosine (dI) causes a transition dA-to-dG, thereby changing the genetic code^{148,149}. Deoxyinosine in DNA is commonly considered as DNA damage caused by spontaneous or nitric oxide-induced deamination of adenosine^{148,150–153}. The deamination rate of dA to dI is very low, with rates of four times per mammalian cell per day¹⁵⁰. The steady state levels of dI account for approximately 1 base in a million bases^{154,155}, so roughly 3000 bases per cell. Another source of deoxyinosine in DNA is the salvage of free hypoxanthine or inosine¹⁵⁶ to dITP, which can be incorporated into DNA by mammalian DNA polymerases during replication^{157,158}.

Deoxyinosine can be erased from DNA via the base excision repair (BER) pathway – the major repair pathway to remove damaged bases from DNA. To this end, methylpurine-DNA glycosylase (MPG) (also called alkyl-adenine DNA glycosylase (AAG)) cleaves the N-glycosidic bond between the base hypoxanthine and the deoxyribose. The resulting abasic site is replaced with an unmodified deoxyadenosine by the downstream BER machinery^{159–164}. Another pathway for dI repair is the alternative excision repair (AER) via Endonuclease V (EndoV). EndoV incises the DNA backbone one base offset at the 3' side of dI¹⁶⁵, thereby initiating the excision of a small patch of DNA and subsequent gap filling with unmodified nucleotides¹⁶⁶. The incision activity of EndoV at dI sites is well described in *E.coli*, but also *in vitro* experiments with purified human EndoV indicate that this pathway is likely to exist in human DNA¹⁶⁷. Figure 3.1-5 provides an overview of both BER and AER mediated repair pathways of dI.

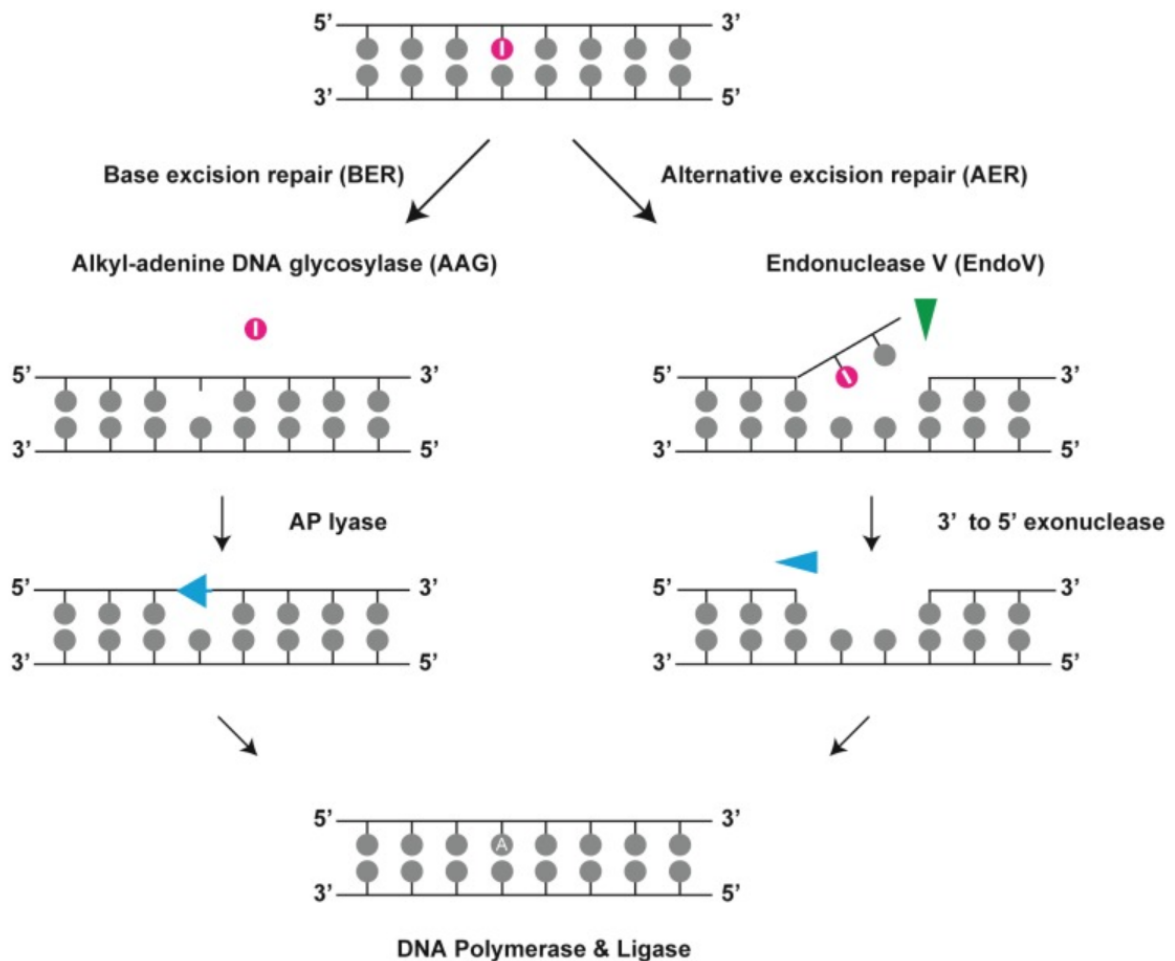


Figure 3.1-5: Base excision repair (BER) and alternative excision repair (AER) erase dI from DNA. The BER pathway requires release of the hypoxanthine base by cleavage of the N-glycosidic bond by AAG/MPG glycosylase. The resulting abasic site (AP) site is repaired with participation of DNA polymerase and ligase. The alternative excision repair pathway involves Endonuclease V that generates a nick in the DNA 3 phosphodiester bonds 3' of dI. The subsequent removal of the lesion in DNA requires the excision of a small patch of DNA containing the lesion (Figure from reference¹⁶⁸).

Structural analyses of ADAR2 bound to substrate dsRNA suggest that ADARs require the A-helical structure of dsRNA to contact the target adenosine¹³³. Similar to dsRNA^{169,170}, DNA:RNA hybrids adopt A-helix conformation^{171,172}. This structural similarity might allow ADAR binding and editing not only of dsRNA, but also of R-loops in the genomic context.

3.2. DNA:RNA hybrids in the mammalian genome

R-loops are three-stranded nucleic acid structures occurring when an RNA strand invades the DNA double helix. They consist of a DNA:RNA hybrid and a displaced single-stranded DNA (ssDNA)¹⁷³.

Initially, they were considered as by-products of transcription, but during the recent years, evidence emerged for the participation of R-loops in steering biological processes. However, when R-loops are

misregulated, they pose a danger for DNA integrity. For instance, aberrant accumulation of R-loops induces DNA double-strand breaks (DSBs) via transcription-coupled nucleotide excision repair (TC-NER). TC-NER endonucleases XPG and XPF cleave at R-loop sites, causing DSBs when misregulated^{174,175}. This dual role of R-loops highlights the need for tight regulation of these genomic structures^{176–178}. R-loops can be resolved by RNase H endonucleases, that specifically degrade the RNA within the DNA:RNA hybrid¹⁷⁹. Other proteins that are known to unwind DNA:RNA hybrids include helicases, such as SETX, aquarius (AQR)^{174,180} and DEAD-box-RNA (DDX) helicase family members^{181–184}.

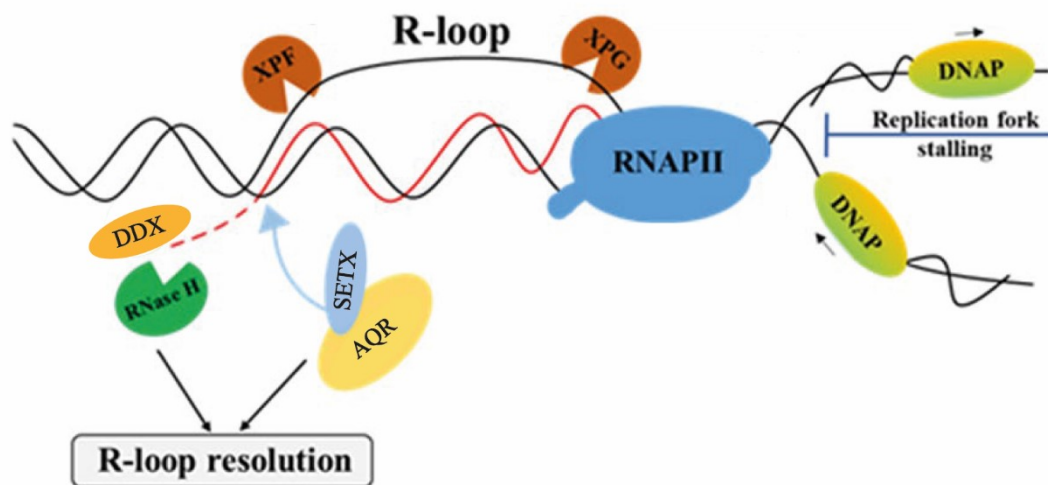


Figure 3.2-1: Mechanisms of R-loop resolution. R-loops can form at sites of active transcription, e.g. by RNAPII. DNA:RNA hybrids can cause replication fork stalling. RNase H specifically degrades the RNA moiety within a DNA:RNA hybrid. Helicases like SETX, AQR and DDX family proteins unwind DNA:RNA hybrids. TC-NER factors XPF and XPG incise the DNA at R-loop sites, potentially causing DSBs. DNAP=DNA polymerase. Figure modified from reference:¹⁸⁵.

R-loops can either form co-transcriptionally in *cis*, behind progressing RNA polymerases (RNAP) or in *trans*, when the RNA is produced at another genomic locus¹⁸⁶. They occur in regions transcribed by the three RNA polymerases I, II and III. In RNAPII-transcribed gene loci, they tend to occur mostly in promoter regions. One might think that the occurrence of R-loops in actively transcribed genes is a consequence of the open chromatin at sites of ongoing transcription, which favors the invasion of the DNA double helix by RNA. However, R-loops are well-characterized regulators of gene expression themselves. R-loops promote transcription by protecting promoters from methylation^{187,188} or recruit factors of the DNA demethylation machinery¹⁸⁹. Another mode of R-loop-mediated regulation of gene expression involves the recruitment of transcription regulators. For instance, an R-loop downstream of the VIM TSS recruits NF- κ B, thereby promoting transcription¹⁹⁰. In addition to site-specific effects

on transcription, R-loops also can regulate chromatin conformation. In mouse embryonic stem cells, R-loop formation at differentiation genes inhibits repressive chromatin-modifying enzymes and recruits activating chromatin-remodelers¹⁹¹. R-loop dependent modulation of promotor-proximal chromatin can involve binding of the polycomb repressive complex (PRC). However, the mode of recruitment does not seem to be universal. While a subset of R-loops recruits PRC2^{192,193}, thereby repressing transcription, other R-loop loci rather repel PRC2¹⁹¹. Which factors determine differential transcriptional outcomes of R-loop formation remains elusive.

Since a significant number of R-loops are mapped also to transcription termination sites (TTS)¹⁸⁷, they might play a regulatory role also in transcription termination. Proposed mechanisms for R-loop mediated transcription termination include backtracking of RNAPII at the R-loop site¹⁹⁴, torsional stress¹⁸⁰ or heterochromatin formation¹⁹⁵.

Genomic DNA serves as a common template for transcription and DNA replication. When replication fork and transcription machineries encounter, the resulting transcription-replication collisions (TRCs) present a source of potentially lethal DNA damage. Indeed, most R-loop-mediated DNA damage occurs during the S-phase and R-loop accumulation is expected to exacerbate TRC-related DNA instability and double strand breaks (DSBs)¹⁹⁶. Despite the danger they pose to DNA stability, R-loops play roles in DNA repair mechanisms as well. R-loops accumulate at DSB¹⁹⁷⁻¹⁹⁹, where they are bound by breast cancer type 1 susceptibility protein (BRCA1) and other repair factors, like BRCA2, RAD52 and XPG¹⁹⁹. After removal of the RNA moiety in the DNA:RNA hybrid by SETX¹⁹⁸, RNase H1²⁰⁰, RNase H2¹⁹⁹ or DDX1¹⁹⁷, the 5' resected DNA end is loaded with Rad51 to allow for homologous recombination^{178,197,198}.

An important tool for the investigation of R-loops is the monoclonal S9.6 antibody, that binds DNA:RNA hybrids with high affinity. It is widely used in approaches including dot blots, imaging and DNA:RNA immunoprecipitation (DRIP)²⁰¹. Since the S9.6 antibody shows affinity also to other non-B-DNA structures, like dsRNA^{202,203}, an important control for the specificity of S9.6 binding is the pretreatment of samples with recombinant RNase H. This endonuclease digests the RNA moiety within DNA:RNA hybrids²⁰⁴, thereby removing the recognized S9.6 epitope. Sensitivity to RNase H pretreatment validates S9.6 signals as R-loop specific. Additionally, pre-treatment of the samples with RNase III, a dsRNA specific endonuclease, improves the performance of R-loop detection by the S9.6 antibody²⁰¹.

3.3. LC-MS/MS for detection of nucleic acid modifications

Due to their low abundance and similarity (i.e. similar base-pairing preferences) to the canonical nucleobases, detection and studying of nucleic acid modifications is challenging. Antibody-based detection can be problematic, since it may not be sufficient to distinguish between modified and regular bases and antibodies are available only for certain modifications. Other techniques require

harsh chemical pre-treatment that can destroy the nucleic acids and introduce artifacts. Approaches relying on enzyme recognition (i.e. inactive nucleases or deaminases) suffer from their dependency on the presence and features of the appropriate enzyme in nature²⁰⁵.

Mass spectrometry allows direct identification of modified DNA and RNA bases. The principle of mass spectrometry (MS) is the generation of ions and measurement of their mass-to-charge (m/z) ratio. For the analysis of DNA and RNA bases, nucleic acids are hydrolyzed and dephosphorylated to single nucleosides using a cocktail of enzymes²⁰⁶. To allow detection of low abundant nucleic acid modifications, the sample requires an additional separation from the bulk of the sample matrix, such as buffers, salts and all other nucleosides. This separation is achieved by coupling the mass spectrometer to liquid chromatography (LC) that separates analytes according to their hydrophobicity. An electrospray ionization (ESI) source allows ionization of samples directly from aqueous solutions and therefore helps to successfully couple LC and MS²⁰⁷. The hydrolyzation of nucleic acid polymers to single nucleosides, however, erases all information of the location and co-occurrence of modifications. This is the major constraint of this otherwise very specific and sensitive method, compared to sequencing-based techniques, for example.

For detection of covalently modified nucleic acids, commonly tandem mass spectrometers (MS/MS) are used. These contain two mass analyzers, separated by a fragmentation reaction of the analyzed molecules – thereby providing information on the intact (precursor) and the fragmented (product) ion. In this study, I used a triple quadrupole mass analyzer (QQQ). A quadrupole (Q) acts as a mass-selective filter consisting of two sets of opposing rods. Radiofrequency (RF) and DC voltages applied to the opposing rods create an oscillating electric field that exclusively allows the passage of ions with a selected m/z ratio. Resonant ions with the designated m/z ratio can pass through the quadrupole, whereas the motion of ions with larger or smaller m/z is destabilized. These ions are not transmitted through the quadrupole²⁰⁸ (Figure 3.3-1).

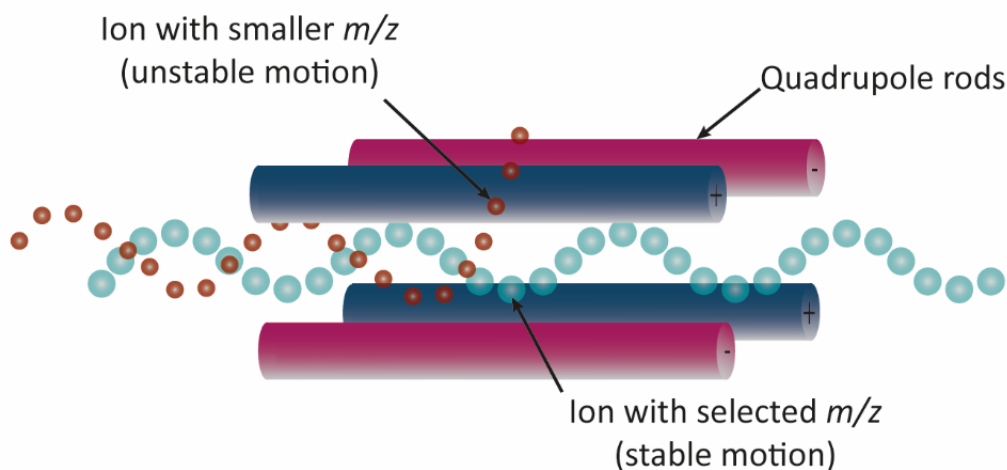


Figure 3.3-1: Principle of quadrupole mass filtering. The quadrupole rods generate an oscillating electric field, through which only resonant ions of selected m/z ratio can pass. Ions with smaller (or larger) m/z ratio fail to pass through the quadrupole, as their motion is distracted. Hence, ions of selected m/z are filtered and can be further fragmented or detected.

For detection of modified DNA and RNA bases, LC-MS/MS takes advantage of their common structural feature of a sugar moiety linked to the nucleobase by an N-glycosidic bond. The most likely fragmentation reaction under low energy collision-activated dissociation (CAD) is the breakage of the N-glycosidic bond. Therefore, in the positive acquisition mode, the most prominent product ion is the ionized base after neutral loss of the sugar moiety²⁰⁹. Figure 3.3-2A illustrates this fragmentation reaction for modified deoxynucleosides: fragmentation of the parental ion causes neutral loss of the deoxyribose (dR), yielding the protonated base (the product ion) that is 116 mass units smaller. For RNA bases, the neutral loss of a ribose accounts for 132 mass units, accordingly. To screen for unknown DNA base modifications, triple-quadrupole mass spectrometers can perform a constant neutral loss scanning. As illustrated in Figure 3.3-2B, quadrupole 1 (Q1) filters the parental ions, and quadrupole 3 (Q3) scans in synchronized offset from Q1 by the defined mass difference of $m/z=116$. Fragmentation occurs in the collision cell (q2), where accelerated ions collide with an inert gas. Only fragment ions that are 116 mass units smaller than the parental ion travel through all three quadrupoles and will be recorded as a signal²¹⁰.

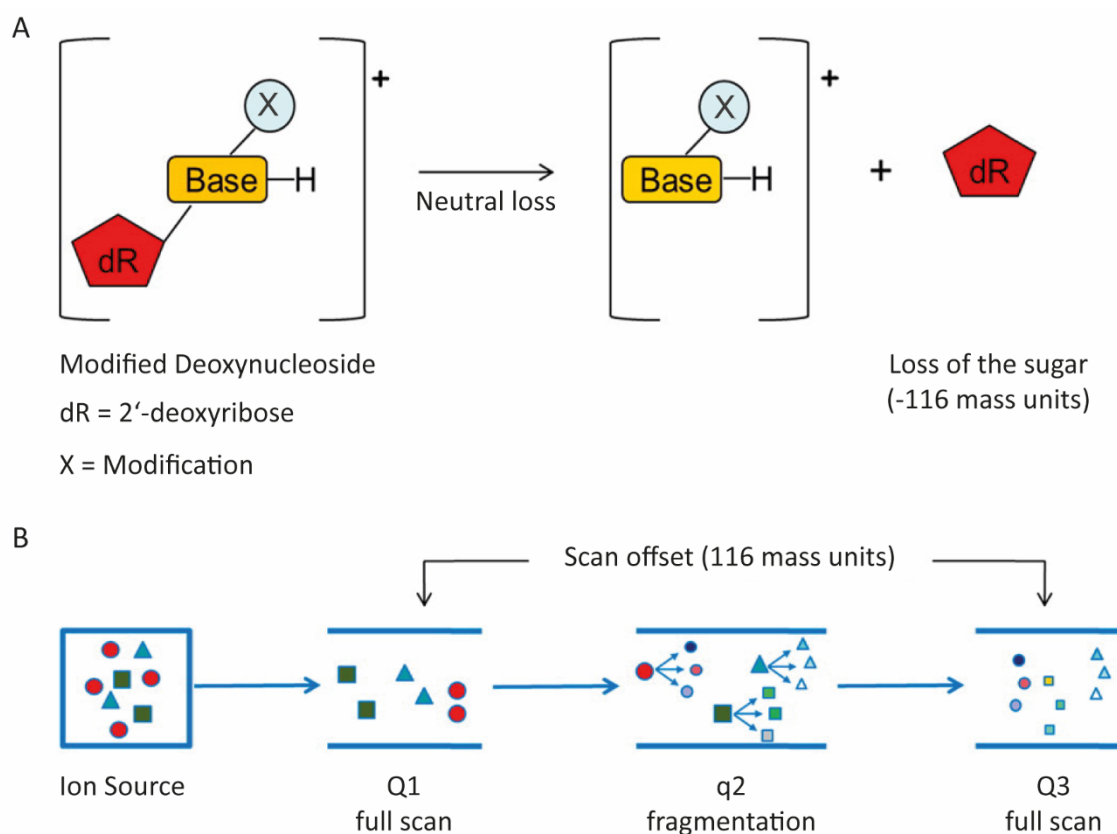


Figure 3.3-2: Neutral loss scanning for detection of modified DNA bases. (A) Ionized deoxyribonucleosides (precursor ions) are fragmented by breakage of the N-glycosidic bond. The product ion (the base) is 116 mass units smaller than the precursor ion, due to the neutral loss of deoxyribose (dR). Base modification is indicated with "X". (B) Schematic of the neutral loss scanning method. Quadrupole 1 (Q1) and quadrupole 3 (Q3) scan with 116 mass units offset to each other. Fragmentation occurs in the collision chamber (q2) by collision with an inert gas (Figure from reference: ²¹⁰).

Stable isotope dilution allows for absolute quantification of known compounds. To this end, the analyte is mixed with a defined concentration of an authentic isotope-labeled analog of the compound of interest. The endogenous and the isotope-labeled compounds are identical except for their mass. Internal standards compensate for numerous perturbations, like ionization efficiency variation across analytes and ion suppression or enhancement due to sample matrix components²¹¹⁻²¹⁴. This provides most specific and precise detection.

3.4. Aim of the thesis

The goal of this study was to identify novel epigenetic modifications of DNA bases in the mammalian genome. As an unbiased screening method, I used a LC-MS/MS based approach to monitor all deoxyribonucleosides between $m/z=228$ and 530. I applied it in different screening approaches:

I. Gain-of-function screening for novel enzymatically introduced DNA modifications:

I used a comparative gain-of-function based screening approach with bioinformatically preselected candidate proteins containing both a DNA binding domain and an enzymatic activity. Upon overexpression of these proteins, I screened for potential novel or increasing deoxyribonucleotides. By linking the candidate enzyme to the modified nucleotide, I could distinguish potential modifications from unspecific DNA damage and artifacts.

II. SILAC screening for sulfur-containing DNA modifications:

While there are a number of RNA modifications that contain the element sulfur, there have been no reports on sulfur-containing deoxyribonucleotides in mammalian DNA so far. The amino acid methionine can serve as sulfur donor for thiolated RNA bases. To this end, I used a SILAC-based labelling approach with a stable heavy isotope of sulfur (^{34}S), administered as ^{34}S -methionine. The DNA from cells treated with this heavy isotope was analyzed for incorporation of ^{34}S by the LC-MS/MS screening method.

III. Screen for DNA modifications in DNA:RNA hybrids:

R-loops were initially considered merely side-products of transcription. Recently, however, many studies certified that R-loops have regulatory functions, while at the same time posing a danger to genome stability. To maintain the balance between beneficial and detrimental roles, R-loops need to be tightly controlled and therefore might be hotspots of DNA modification. To screen for DNA modifications in R-loops, I carried out DNA:RNA immunoprecipitation and analyzed the DNA with the aforementioned LC-MS/MS screening approach.

In addition to the LC-MS/MS screening approach for novel modifications, our lab has been working on unraveling the origin of genomic m^6dA in mammalian DNA. In this context, we investigated the possibility that m^6dA might stem from methylated RNA that is recycled and incorporated into DNA during replication. Therefore, I performed experiments to support that the nucleotide salvage pathway is the major source of m^6dA in mammalian genomic DNA.

4. Results & Discussion

4.1. Screening for novel DNA base modifications

Based on the assumption that there are more epigenetic DNA base modifications in mammalian DNA than the previously described 5mC, 5hmC, 5fC, 5caC and m⁶dA, I used three approaches to conduct a screening for currently unknown DNA modifications. A gain-of-function screening with potential DNA modifying enzymes can provide a direct link between the DNA modification and the corresponding enzymatic modifier. Stable isotope labelling of amino acids in cell culture (SILAC) with sulfur-labelled methionine, a known sulfur-donor for RNA modifications, allows the detection of DNA bases modified with sulfur-containing groups. Lastly, screening for base modifications in DNA:RNA hybrid-enriched DNA can provide insight on DNA modifications that accumulate in genomic R-loops. The three approaches were combined with an LC-MS/MS screening method, that monitors deoxyribonucleosides between $m/z=229$ and $m/z=530$.

4.1.1. Gain-of-function screening for enzymatically regulated DNA modifications

For the gain-of-function screening approach, proteins that have both a DNA binding domain and an enzymatic activity were pre-selected bioinformatically and overexpressed in cultured HEK293T cells. This gain-of-function screening approach not only allows for identification of novel DNA base modifications, but also provides the link to the modifying enzyme, thereby discriminating from chemical DNA damage, like spontaneous or nitric-oxide (NO)-induced deamination. Successful overexpression was confirmed either by fluorescence microscopy to detect expression of turboGFP, encoded on a polycistronic construct together with the coding regions of the enzymes of interest, or by validating overexpression of V5-, HA-, or FLAG-tagged proteins by western blot (data not shown).

To increase the screening power and to avoid false positives from sample contamination and introduction of artifacts during sample preparation, I used two biological replicates and measured in two technical replicates. The filter settings were selected to only count biologically reproducible hits. After overexpression of the candidates, I isolated genomic DNA, hydrolyzed and dephosphorylated it for subsequent analysis by the LC-MS/MS screening approach. The screened mass range between $m/z=229$ and $m/z=530$ covers a large range of potential covalent base modifications. A software detects and aligns identical peaks derived from electropherograms monitoring the same m/z ratio. From the aligned peaks, I filtered noise by excluding those peaks that were not reproduced in biological and technical replicas from the same treatment. From the set of aligned peaks, two major scenarios were considered: novel peaks, and peaks that increase in abundance upon overexpression of one enzyme. (Figure 4.1-1).

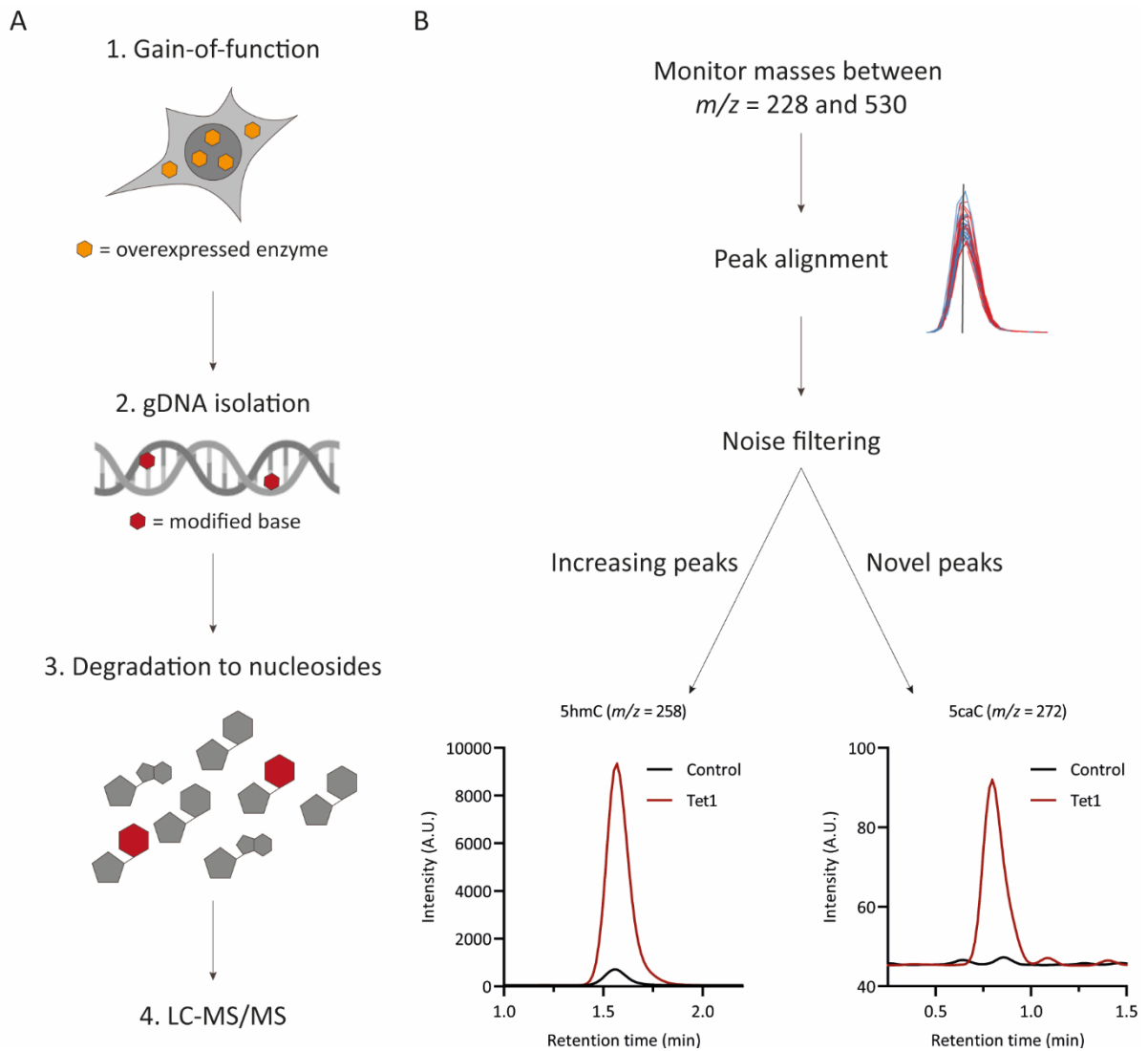


Figure 4.1-1: Gain-of-function screening strategy. (A) Experimental procedure of the gain-of-function screening approach. Enzymes of interest were overexpressed in HEK293T cells, gDNA was isolated 48 h after transfection with the respective plasmids. DNA was degraded to nucleosides and analyzed by LC-MS/MS. (B) Peak filtering approach and validation. m/z between 228 and 530 were monitored by LC-MS/MS. Peaks from all samples were aligned. To distinguish between noise and real increasing or novel peaks, the data was filtered. For filtering, Tet1 oxidation products were used as controls to monitor novel and increasing peaks.

Importantly, to validate the screen filtering approach, I used Tet1 overexpression as a control in each experiment. Tet1 produces three demethylation products: 5hmC, 5fC and 5caC. Due to the different abundance of the three Tet1 oxidation products, this control covers a wide dynamic range and validated the filtering process for identifying both novel peaks and peaks that increase in abundance. To further expand the filtering power, I also considered the naturally occurring isotopologues of Tet1-products and their low abundant salt-adducts (Figure 4.1-2).

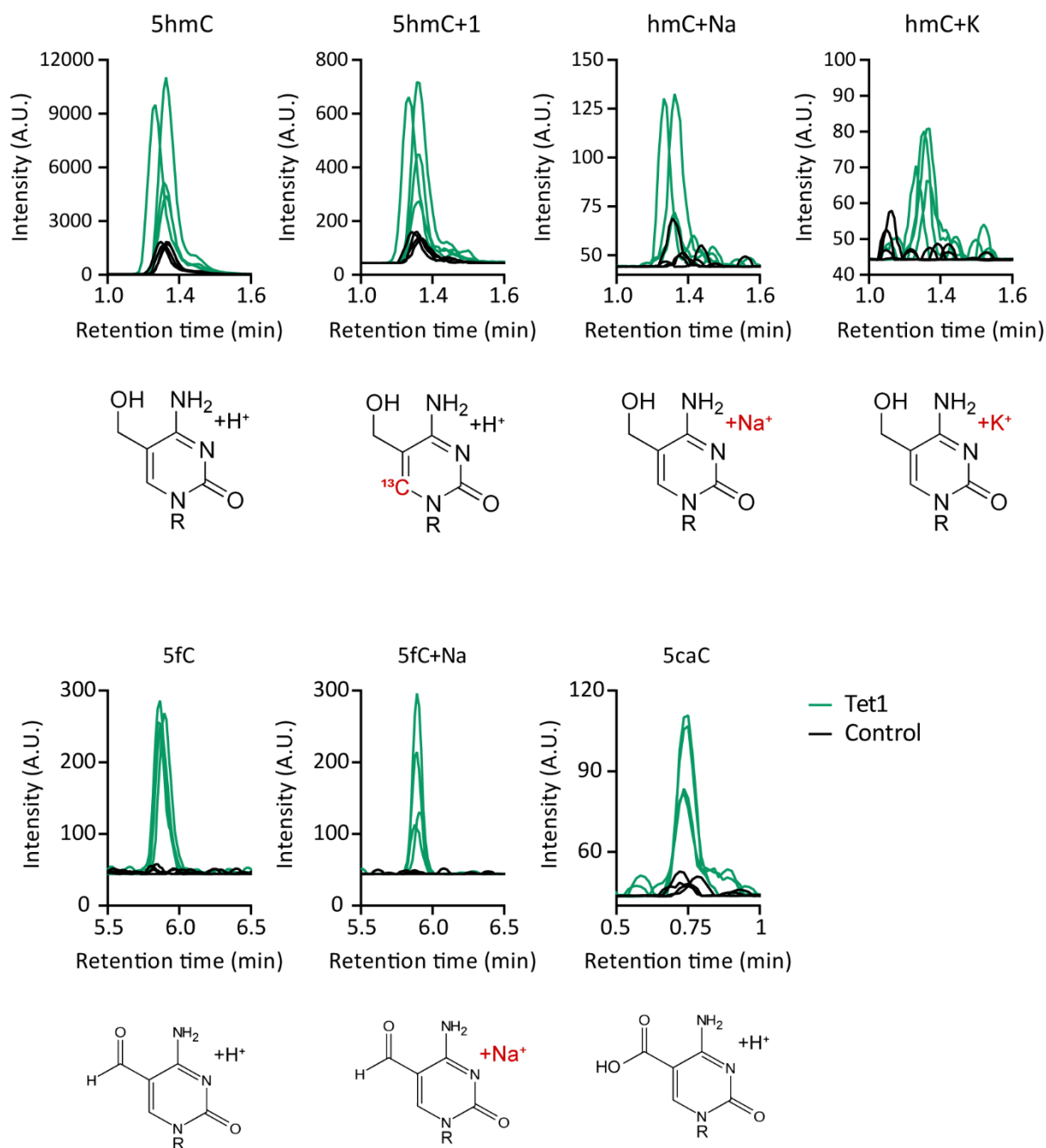


Figure 4.1-2: 5hmC, 5fC and 5caC and their naturally occurring isotopologues and salt adducts serve as validation for the peak filtering. Electropherograms of Tet1 oxidation products, their isotopologues and salt adducts after Tet1 overexpression (green) and in control samples (black). Below each electropherogram the structural formula of the respective ion is illustrated. Atoms that cause changes in the m/z ratio are indicated in red. Screening experiments were performed in two biological and two technical replicates.

Isotopologues are molecules that contain at least one atom with different number of neutrons compared to the parent atom (e.g. ^{13}C or ^{15}N). Salt adducts occur during ionization, when a monovalent salt ion (often sodium or potassium) instead of a proton is added to the molecule²¹⁵. The screening

filtering strategy successfully returned 5hmC, 5fC and 5caC peaks, their isotopologues and salt adducts upon Tet1 overexpression. These controls highlight the power of my screening method to identify novel and increasing peaks (Figure 4.1-2).

After validation of the screen strategy, I proceeded with the overexpression of the 124 candidate proteins listed in Table 4.1.1. and subsequent analysis of the respective gDNA for unknown DNA modifications.

Table 4.1.1: Candidate proteins tested in the gain-of-function screen and their annotated enzymatic activities.

Protein	Enzymatic function	Protein	Enzymatic function
AICDA	Deaminase	METTL7B	Methyltransferase
ALKBH3	Dioxygenase	METTL8	Methyltransferase
ALKBH5	Demethylase	METTL9	Methyltransferase
ALKBH7	Dioxygenase	MINA	Oxygenase
APOBEC3A	Cytidine Deaminase	MRPP1	Methyltransferase
ARNTL	Activator	MTA2	Deacetylase
ASH2L	Methyltransferase	MTFMT	Formyltransferase
CDKAL1	Methylthiotransferase	NAA10	Acetyltransferase
CDY2B	Acetyltransferase	NAT10	Acetyltransferase
CDYL	Crotonyl-coA Hydratase	NAT14	Acetyltransferase
CMTR2	Methyltransferase	NOP2	Methyltransferase
CTCFL	Chromatin Regulator	NSUN2	Methyltransferase
DcpS	Diphosphatase	NTMT1	Methyltransferase
DIMT1	Methyltransferase	OGFOD1	Hydroxylase
DNMT3A	DNA methyltransferase	OGT	Acetylglucosaminyltransferase
DNMT3B	Methyltransferase	OSGEP	Acetyltransferase
EGLN2	Hydroxylase	PAM	Monooxygenase
EHMT2	Methyltransferase	PARP1	Poly(ADP-ribose)-Polymerase
EP300	Acetyltransferase	PARP15	Poly(ADP-ribose)-Polymerase
EZH1	Methyltransferase	PCMTD2	Methyltransferase
EZH2	Methyltransferase	PHF2	Demethylase
FTO	Dioxygenase	PHF8	Demethylase
GLYR1	Reductase	POLM	DNA-directed DNA/RNA polymerase
GTF3C4	Acetyltransferase	PRDM5	Methyltransferase
HCAP1	Phosphatase	PRDM6	Methyltransferase
HEMK1	Methyltransferase	PRDM7	Methyltransferase
HR	Demethylase	PRMT7	Methyltransferase
HTATIP	Oxidoreductase	PUS1	Pseudouridine Synthase/Isomerase
KAT2B	Acetyltransferase	PUS3	Pseudouridine Synthase
KAT7	Acetyltransferase	QTRT1	Glycosyltransferase
KDM1A	Demethylase	Rev1	Deoxycytidyl Transferase
KDM2B	Demethylase	RG9MTD3	Methyltransferase
KDM4A	Demethylase	RNMT	Methyltransferase
KDM4B	Demethylase	RNMTL1	Methyltransferase
KDM4C	Demethylase	RRM2B	Ribonucleotide Reductase
KDM5A	Demethylase	RTCA	Ligase
KDM5C	Demethylase	SETD2	Methyltransferase
KDM6A	Demethylase	SETD3	Methyltransferase
KDM6B	Demethylase	SETD6	Methyltransferase
KDM8	Dioxygenase	SETD7	Methyltransferase

KMT2E	Methyltransferase	SETDB1	Methyltransferase
KMT5C	Methyltransferase	SETDB2	Methyltransferase
LanCL2	ATP/GTP Binding	SETMAR	Methyltransferase
MBD1	Methyl-CpG-binding protein	SMYD2	Methyltransferase
MBD2	Demethylase	SMYD3	Methyltransferase
MBD3L1	Methyl-CpG-binding Protein	SUV39H2	Methyltransferase
MBD3L2	Methyl-CpG-binding protein	SUV420H1	Methyltransferase
MBD4	DNA N-glycosylase	TET3	Dioxygenase
MBD5	Methyl-CpG-binding Protein	TNKS	Poly(ADP-ribose)-Polymerase
MBD6	Methyl-CpG-binding	TNKS2	Poly(ADP-ribose)-Polymerase
METTL1	Methyltransferase	TRDMT1	Methyltransferase
METTL10	Methyltransferase	TRIT1	Dimethylallyltransferase
METTL13	Methyltransferase	TRM5	Methyltransferase
METTL15	Methyltransferase	TRMO	Adenine-N6-Methyltransferase
METTL16	Methyltransferase	TRMT1	Thiouridylase
METTL21A	Methyltransferase	TRMT10C	Methyltransferase
METTL21B	Methyltransferase	TRMT112	Methyltransferase
METTL22	Methyltransferase	TRMT44	Methyltransferase
METTL25	Methyltransferase	TRMT6	Methyltransferase
METTL2A	Methyltransferase	TRMT61A	Methyltransferase
METTL4	Methyltransferase	TYW1	Demethylwyosine Synthase
METTL5	Methyltransferase	TYW5	Dioxygenase

Among 124 overexpressed candidates, three promising novel or increasing peaks were detected upon overexpression of a candidate enzyme.

Overexpression of Histone-lysine N-methyltransferase KMT5C resulted in the increase of $m/z=324$ at retention time of 7.7 min. Figure 4.1-3A shows the electropherograms of the respective peaks in KMT5C overexpression (red lines) and control samples (black lines) in the screening setup, which was carried out in two biological and two technical replicates. The detected signal of $m/z=324$ might originate from a molecule 56 mass units larger than deoxyguanosine (dG). Interestingly, 56 mass units coincide with four additional methylgroups substituting hydrogen atoms on dG. In theory, this molecule could be a fourfold-methylated dG. However, given the fact that large modifications would probably distort genome stability, block replication and transcription, it is unlikely that such a hyper-modification would possess a regulatory biological function. To exclude erroneous detection of this peak, I repeated the experiment. However, the increase of this peak upon KMT5C overexpression was not reproducible in an independent validation experiment (Figure 4.1-3B).

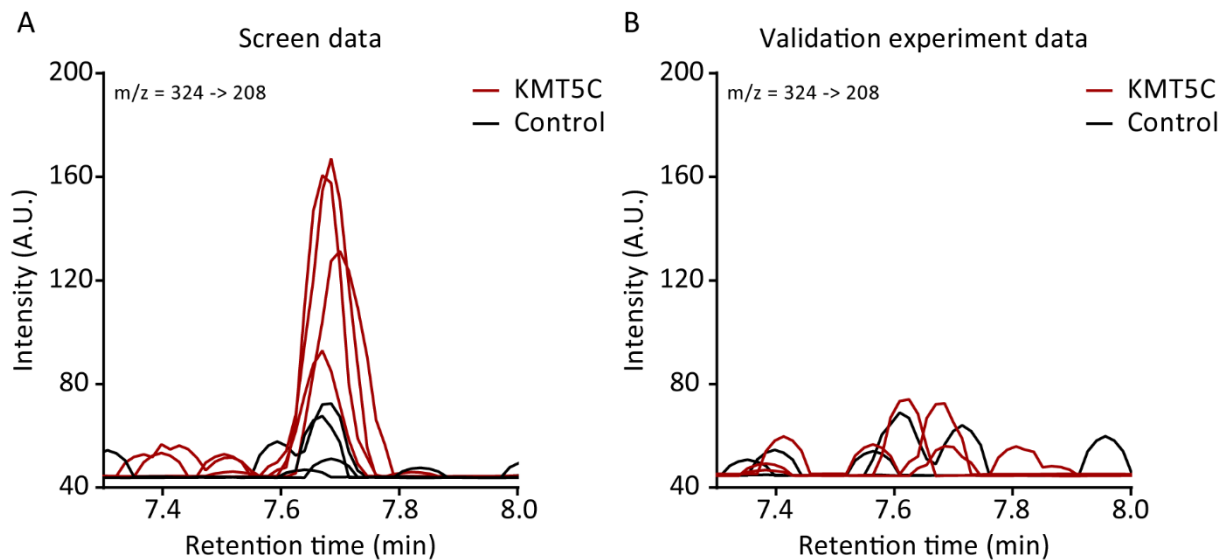


Figure 4.1-3: Overexpression of KMT5C caused peak increase in channel $m/z=324$. Electropherograms showing a peak at 7.7 min retention time. Two technical replicates were measured from two gDNA samples from HEK293T cells overexpressing KMT5C (red) or empty vector control (black). (A) In the screening experiment, the peak area increased in KMT5C overexpression samples. (B) This increase was not reproduced in an independent experiment.

A unique peak in channel $m/z=283$ at 6.95 min retention time was detected upon overexpression of PRDM7 (Probable Histone-Lysine N-Methyltransferase) (Figure 4.1-4) in three out of four replicates. Intriguingly, the detected mass corresponds with a monomethylated dG – which fits well to the predicted methyltransferase activity of PRDM7. In this case, PRDM7 could act as guanine-DNA methyltransferase. Possibly, the detection of $m/z=283$ could as well be related to contamination with a ubiquitous polyester molecule. It matches a polyethylene glycol ion, a common contaminant in mass spectrometry²¹⁶. Indeed, repetition of the experiment could not validate the presence of this peak upon PRDM7 overexpression, suggesting that this screen hit originated from a sample contamination (Figure 4.1-4).

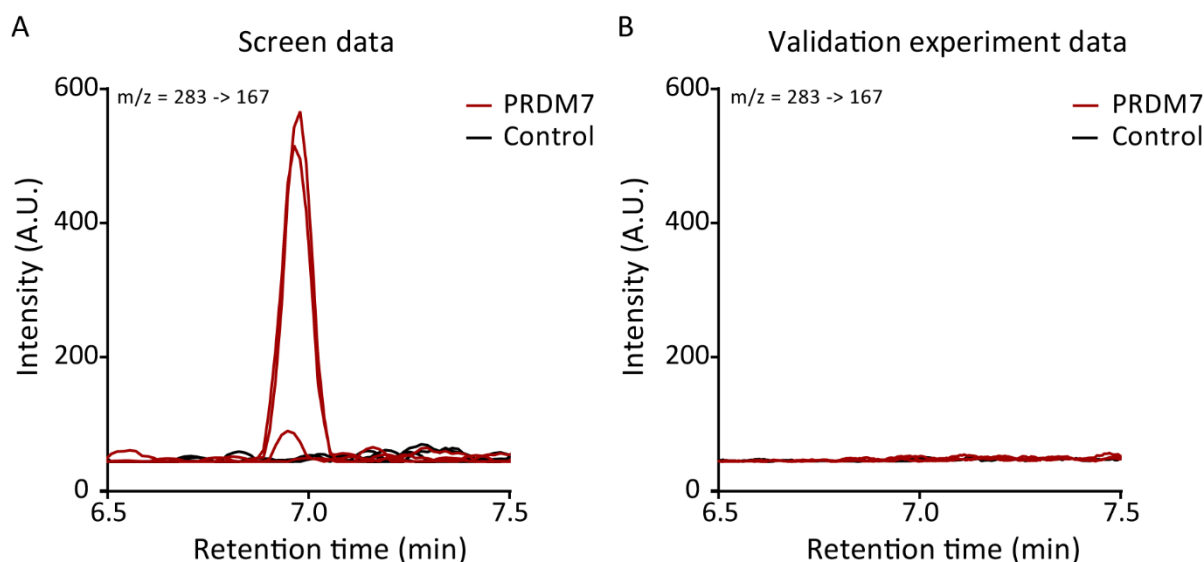


Figure 4.1-4 Overexpression of PRDM7 caused a peak in channel $m/z=283$. Electropherograms showing a peak at 7 min retention time. Two technical replicates were measured from two gDNA samples from HEK293T cells overexpressing PRDM7 (red) or empty vector control (black). (A) In the screening experiment, a unique peak occurred in gDNA from PRDM7 overexpressing HEK293T cells. (B) This peak could not be detected in an independent experiment.

The third screen hit was detected in $m/z=364$ at 1.5 min retention time, upon overexpression of PRMT7 (Protein Arginine N-Methyltransferase 7). The peak area increased in all four replicates with respect to the control. The large mass of this molecule speaks for a very bulky modification, as it is 96 mass units larger than dG. This argues against direct DNA methylation through the enzymatic activity of PRMT7. To confirm, whether the signal intensity is dependent specifically on the overexpression of PRMT7, I repeated the experiment. In the validation experiment, the peak area did not differ from the control samples (Figure 4.1-5). The early retention time suggests that the monitored molecule is rather hydrophilic. Indeed, the mass of $m/z=364$ fits with monophosphorylated dG. However, this base modification is likely an artifact of the enzymatic DNA degradation; free monophosphates could react with the free nucleosides during the dephosphorylation reaction. In the case of dG, the phosphate could attack either the endocyclic nitrogen at position 1, or the exocyclic N-2-aminogroup.

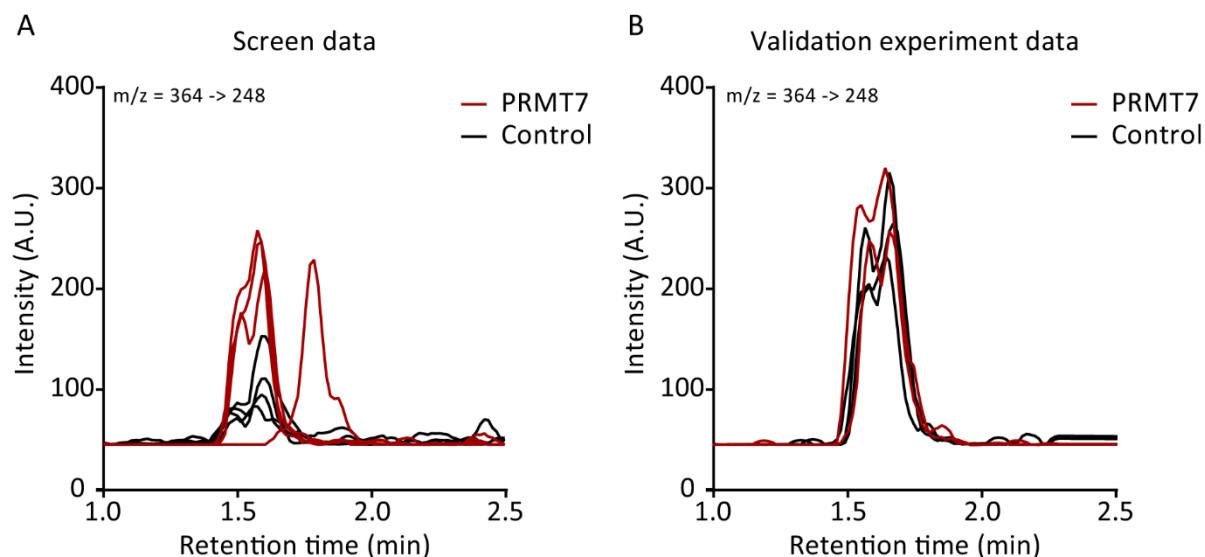


Figure 4.1-5: Overexpression of PRMT7 resulted in a false positive hit on channel $m/z = 364$. Electropherograms showing a peak at 1.5 min retention time. Two technical replicates were measured from two gDNA samples from HEK293T cells transfected with a PRMT7 overexpression construct (red) or an empty vector control (black). In the screen mode the peak area was increased in PRMT7 overexpressing samples (left). This increase was not reproduced in an independent experiment (right).

Indeed, the previous results highlight the importance of additional validations to deal with contaminations and artificial signals produced by spontaneous deamination, phosphorylation, oxidation. The applied quality controls (i.e. Tet1 overexpression as a positive control, extensive cleaning of the instruments and precise normalization of the loaded material by preceding stable isotope dilution LC-MS/MS) were crucial for the validation of the screen. Unfortunately, the three most promising hits could not be reproduced in independent experiments and likely arose from such artifacts and contaminations. Contaminations and artifacts can originate from all reagents that are used during sample processing, sample containers and even volatile compounds in the air²¹⁶. Therefore, in order to confirm the biological relevance of a molecule, it is essential to include appropriate biological mock controls and extensive blank testing to minimize effects of contaminants on the biological conclusions²¹⁶.

A significant constraint of the mass spectrometry based screening approach to find novel DNA modifications is the hydrolysis of DNA to single nucleosides, erasing all information on location and co-occurrence of modified bases. The advances in direct sequencing by third generation sequencing using Oxford Nanopore might overcome these limitations. It identifies nucleotides in a nucleic acid by monitoring electric current intensity across the nanopore surface. Modified nucleotides cause shifts in the monitored current intensity that can be used to identify these modified bases^{217–220}. However, so far, the analysis was restricted to known modifications that can be identified only after extensive

machine learning with known substrates. Recent pre-print studies present software solutions that were able to identify DNA and RNA modifications *a priori* without any training data^{221,222}. This approach could overcome limitations of the bulk analysis of all nucleosides in the genome by LC-MS/MS that might mask low abundance DNA modifications with specific locations.

In summary, using an unbiased LC-MS/MS screening method, I screened human genomic DNA for unknown enzymatically introduced DNA base modifications. Detection and filtering of the known Tet1 oxidation products, 5hmC, 5fC and 5caC, validated and controlled the screening and filtering processes. Among 124 screened candidate enzymes, three false-positive hits could not be reproduced in validation experiments. Nevertheless, the limited number of only three false-positive hits highlights the power of the used screening strategy to distinguish DNA damage and LC-MS/MS artifacts from enzymatically introduced DNA modifications in human DNA.

4.1.1.1. Impact of DNA binding enzymes on the abundance of 5-modified cytosine

The gain-of-function screen for unknown DNA modifications did not reveal novel, enzymatically introduced DNA bases in the human genome. However, since the overexpressed candidate proteins are known, or expected, to interact with DNA, it is possible that they act as regulators of DNA methylation and demethylation. To investigate potential effects of the overexpressed candidate enzymes on the abundance of 5mC, 5hmC, 5fC and 5caC, I quantified these modifications by stable-isotope dilution LC-MS/MS. The absolute quantification of these modifications allows for detection of even small changes in abundance.

Figure 4.1-6 shows the fold-change of 5mC in DNA from HEK293T cells overexpressing the indicated proteins over the empty vector control. Most candidate proteins did not affect the global abundance of 5mC in HEK293T gDNA. Concordant with the literature⁴⁸, overexpression of DNMT3a and DNMT3b caused approximately 20% increase in global 5mC levels over the control sample.

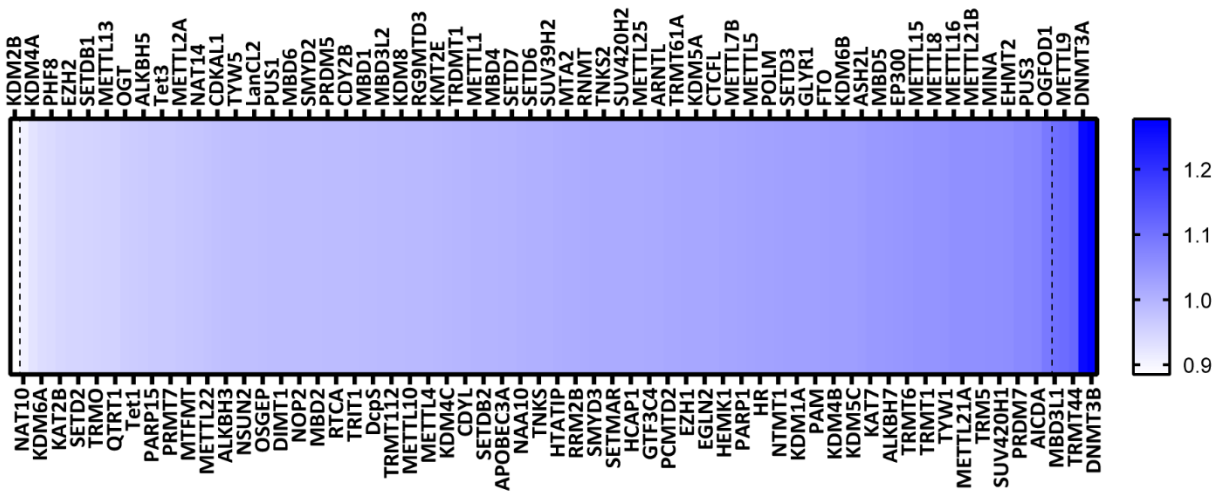


Figure 4.1-6: Fold-change of 5-methylcytosine in HEK293T genomic DNA. 5mC levels, measured by stable-isotope dilution LC-MS/MS in gDNA from HEK293T cells overexpressing annotated candidate enzymes in biological duplicates, were normalized to the control condition to determine the fold change. 5mC values outside the dashed lines differ by 10% or more from the control.

Likewise, the fold change of 5hmC in HEK293T gDNA was measured to investigate if one of the candidate proteins might regulate DNA demethylation (Figure 4.1-7). Overexpression of Tet1 and Tet3, as expected, caused roughly a 10-fold and 4-fold increase of global 5hmC levels over the control, respectively.

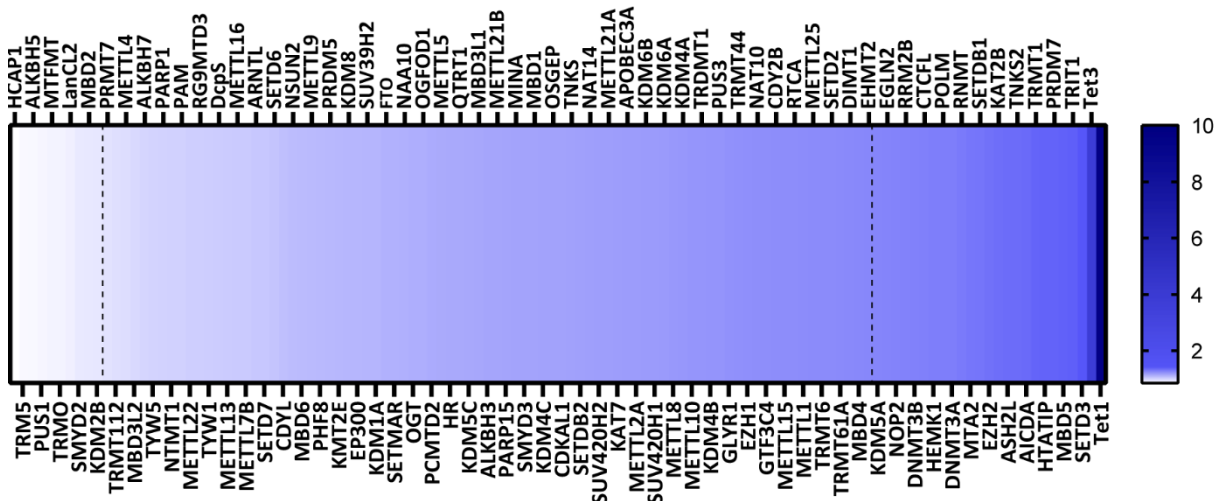


Figure 4.1-7: Fold-change of 5-hydroxymethylcytosine in HEK293T genomic DNA. 5hmC levels, measured by stable-isotope dilution LC-MS/MS in gDNA from HEK293T cells overexpressing annotated candidate enzymes in biological duplicates, were normalized to the control condition to determine the fold change. 5hmC values outside the dashed lines differ by 10% or more from the control.

Thirty-three candidate enzymes changed the 5hmC levels by more than 10% with respect to the control. Intriguingly, Lysine Demethylase 2B (KDM2B) decreased both 5mC and 5hmC and might therefore be a candidate for Tet-independent 5mC demethylation.

Since the focus of this project was the identification of novel DNA modifications, I did not further investigate the changes of 5mC and its oxidative derivatives. However, these effects may be validated and the roles of the respective candidate enzymes could be further studied. Such experiments could help to better understand the regulation of DNA methylation and demethylation in mammalian genomes.

4.1.2. Screen for novel sulfur related DNA modifications

The second approach to identify currently unknown mammalian DNA modifications was based on the hypothesis, that analogously to enzymatically introduced sulfur-containing RNA modifications (i.e. thiocarbonyl groups²²³) also DNA bases could be modified with sulfur-containing groups. I used a SILAC approach and provided cells with methionine containing a stable heavy isotope of sulfur (³⁴S-methionine). While this targeted approach allows conclusions on the chemical properties of a potential DNA modification, it does not directly discriminate between DNA damage and epigenetic DNA modifications. However, analogous to enzymatically introduced sulfur-containing groups in RNA, incorporation of ³⁴S into DNA bases could result from an enzymatic transfer of ³⁴S from the ³⁴S-methionine to DNA.

To deplete cellular stores of sulfur-donor molecules, I cultured HEK293T, HeLa and NIH3T3 cells in low methionine and cysteine medium for seven days. Sulfur-starvation was followed by incubation in light or heavy cell culture medium, containing natural methionine, or heavy ³⁴S-methionine, respectively. By incorporation of the supplied ³⁴S, DNA bases would gain two mass units, which serve as the unique identifier of sulfur. This mass-shift can be monitored by LC-MS/MS. To confirm that the cells metabolize ³⁴S-methionine, I monitored a sulfur-containing RNA modification: 5-methoxycarbonylmethyl-2-thiouridine (mcm⁵S²U) and its isotopologue ³⁴S-mcm⁵S²U (Figure 4.1-8A).

Qualitative LC-MS/MS analysis confirmed that ³⁴S was metabolized and incorporated into nucleic acids, as the peak area corresponding to ³⁴S-mcm⁵S²U increases upon ³⁴S-methionine treatment in HEK293T, HeLa and NIH3T3 cells. Under light medium conditions, the relative area of ³⁴S-mcm⁵S²U (m/z=335) accounted for approximately 3-5% of total mcm⁵S²U. This agrees well with the natural abundance of ³⁴S of 4.21%²²⁴. Treatment with ³⁴S-methionine increased the abundance of ³⁴S-mcm⁵S²U up to 73%, confirming the incorporation of ³⁴S from the provided ³⁴S-methionine (Figure 4.1-8 & Table 4.1.2).

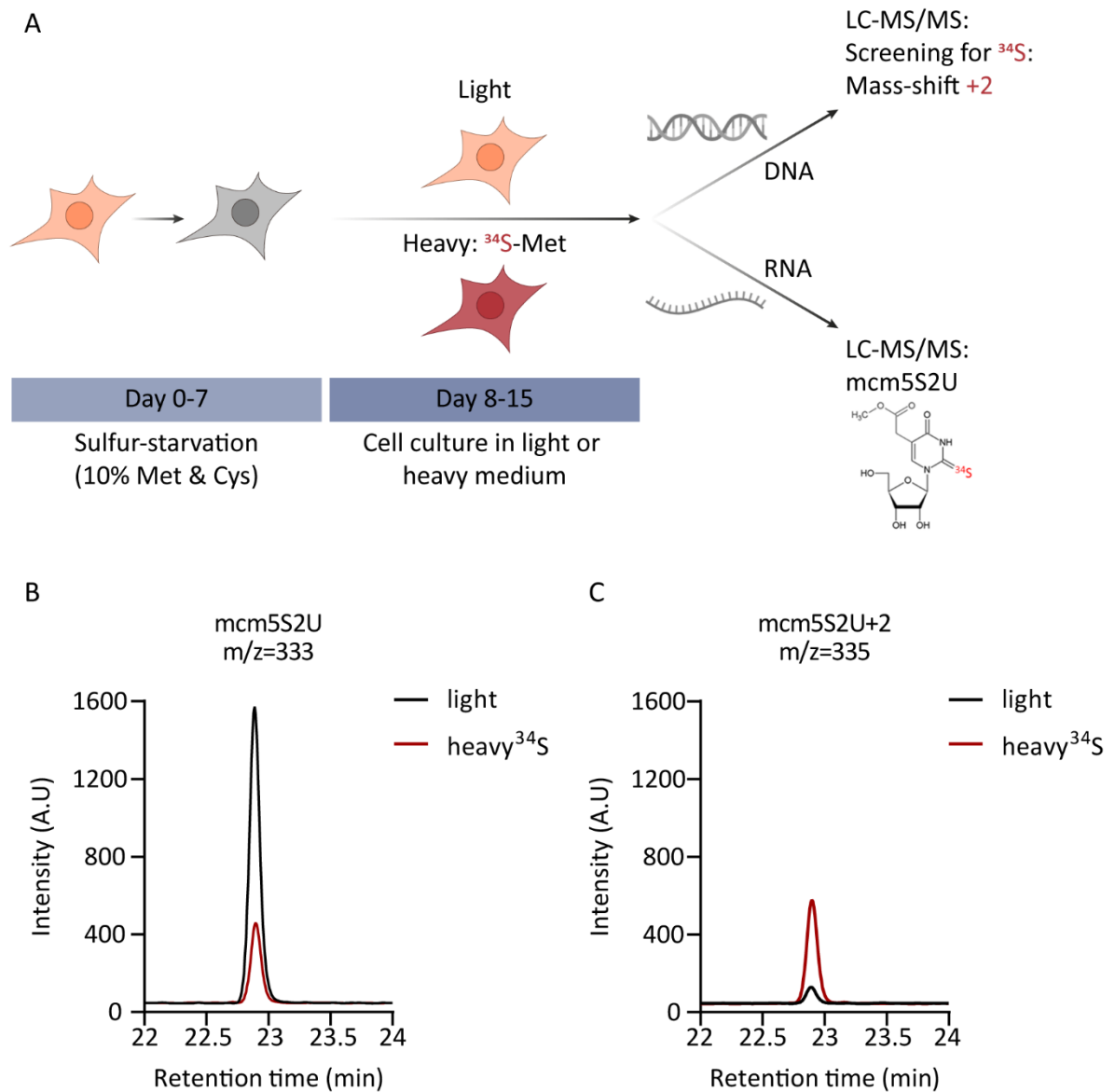


Figure 4.1-8: Incorporation of ^{34}S into sulfur-containing RNA modification $\text{mcm}^5\text{S}^2\text{U}$ in HEK293T cells. (A) Cultured cells were kept at low supply of Met and Cys to deplete their sulfur-stores. After starvation, cells were incubated in light or heavy (^{34}S -met) cell culture medium. DNA from these cells was scanned for incorporation of sulfur into any base. Monitoring of $\text{mcm}^5\text{S}^2\text{U}$ RNA base was used as a control for the incorporation of ^{34}S into nucleic acid modifications. (B-C) Electropherograms of channels monitoring (B) $\text{mcm}^5\text{S}^2\text{U}$ (m/z=333) and (C) ^{34}S - $\text{mcm}^5\text{S}^2\text{U}$ (m/z=335). Black lines: RNA from cells cultured in light medium. Red lines: RNA from cells cultivated with heavy ^{34}S -L-methionine containing medium for three days.

Table 4.1.2: Areas of $mcm^{52}U$ ($m/z=333$) and ^{34}S - $mcm^{52}U$ ($m/z=335$) in HEK293T RNA under light and heavy (^{34}S) conditions. The labelling efficiency was calculated by dividing the peak area of the labelled compound by the sum of the labelled and non-labelled compound in the same sample.

Cell line	Treatment	Non-labelled peak area ($m/z = 333$)	Labelled peak area ($m/z = 335$)	% labelling (labelled area/total area)
HEK293T	Light	660	37	5.31%
	heavy (^{34}S)	183	269	59.51%
HeLa	Light	471	18	3.68%
	heavy (^{34}S)	619	1714	73.47%
NIH3T3	Light	8655	465	5.1%
	heavy (^{34}S)	116	41	26.11%

After confirming that ^{34}S successfully incorporates into nucleic acids on the RNA level, I manually analyzed gDNA for a mass-shift of two additional mass units upon treatment with ^{34}S -methionine. While in HeLa and NIH3T3 cells, no ^{34}S -responsive peaks were found, I identified one peak in HEK293T cells, that responded to ^{34}S labelling by a mass-increase from $m/z=523$ (non-labelled) to $m/z=525$ (^{34}S -labelled) (Figure 4.1-9).

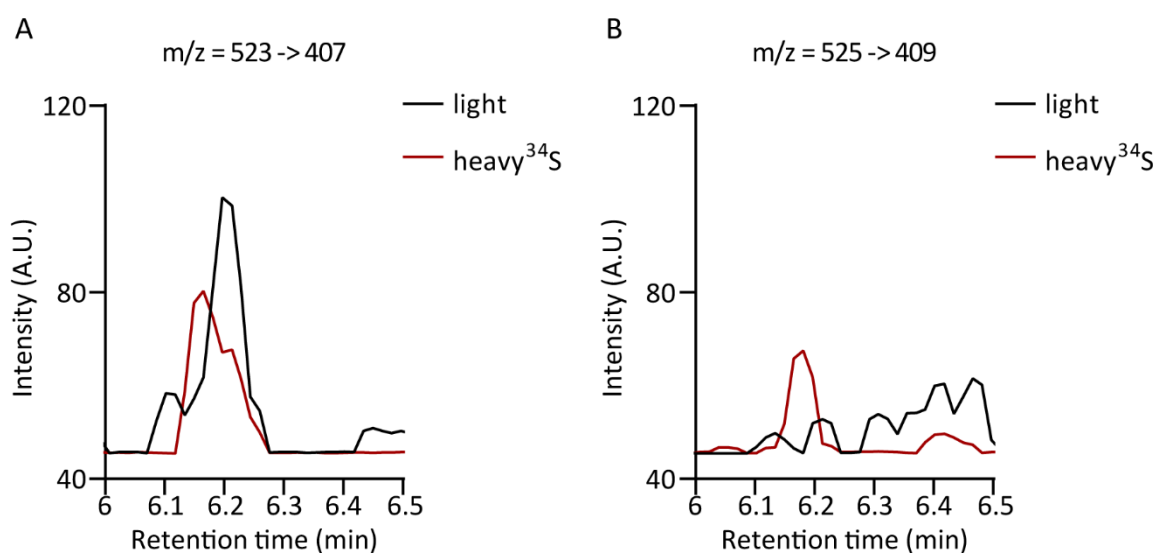


Figure 4.1-9: Sulfur-responsive peak in channel $m/z=523$. Electropherograms of channels monitoring (A) $m/z=523$ and (B) $m/z=525$. Black lines: DNA from cells cultured in light medium. Red lines: DNA from cells cultivated with heavy ^{34}S -L-methionine containing medium for 3 days.

To confirm, whether the signal is indeed dependent specifically on the incorporation of ^{34}S , I repeated the experiment. However, in a validation experiment, the sulfur-dependent increase of the peak in channel $m/z=523$ to $m/z=525$ was not reproducible and looked similar to the control (Figure 4.1-10).

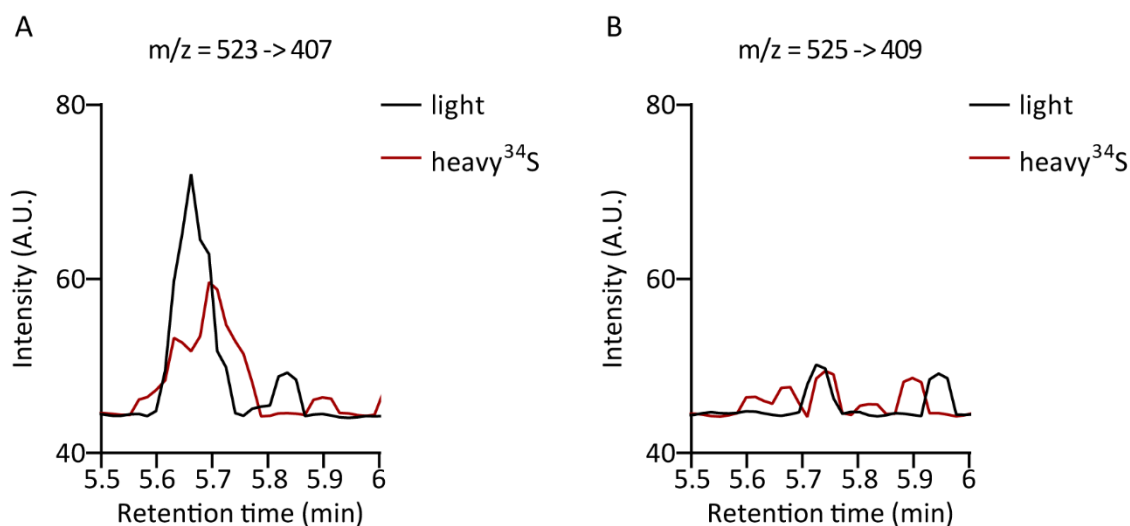


Figure 4.1-10: Sulfur-responsive peak in channel $m/z=523$ was not reproduced. Electropherograms of channels monitoring (A) $m/z=523$ and (B) $m/z=525$. Black lines: DNA from cells cultured in light medium. Red lines: DNA from cells cultivated with heavy ^{34}S -L-methionine containing medium for 3 days.

As mentioned above, it is difficult to control for unknown artifacts during sample processing. Indeed, the m/z -of 525 matches an ion that is commonly detected in mass spectrometry due to PEG contamination²¹⁶.

All in all, the screen for novel DNA modifications that contain sulfur provided by methionine did not reveal any sulfur-dependent DNA modifications in HEK293T, HeLa and NIH3T3 cells. The approach was validated by successful incorporation and detection of ^{34}S -labelling of the thiolated RNA modification $\text{mcm}^5\text{S}^2\text{U}$, however, one sulfur-responsive peak identified in the screen in HEK293T gDNA was not reproducible.

4.2.1. Screen for R-loop enriched DNA modifications

Low abundance DNA modifications might occur only in certain regions of the genome. The enrichment of such regions removes the bulk of unmodified DNA, thereby increasing the sensitivity of the screen. One of such regions are R-loops, three stranded nucleic acid structures that require tight regulation to maintain the balance between their biological functions and the dangers they pose to genome stability^{196–199}. Besides endonucleases¹⁷⁹, helicases^{174,180} and other mechanisms, DNA modifications might provide another layer of R-loop regulation. To screen for novel DNA base modifications in the context of mammalian R-loops, I isolated them by DNA:RNA immunoprecipitation (DRIP), using the S9.6 antibody. An important control for the specificity of the pulldown towards R-loop regions is the pretreatment of the input material with RNase H, which degrades the RNA moiety within DNA:RNA hybrids – thereby removing the S9.6-recognized epitope²⁰⁴. To control for DRIP specificity, I used an DNA:RNA hybrid spike-in oligo and analyzed its recovery and sensitivity to RNase H by subsequent qPCR analysis (Figure 4.2-1).

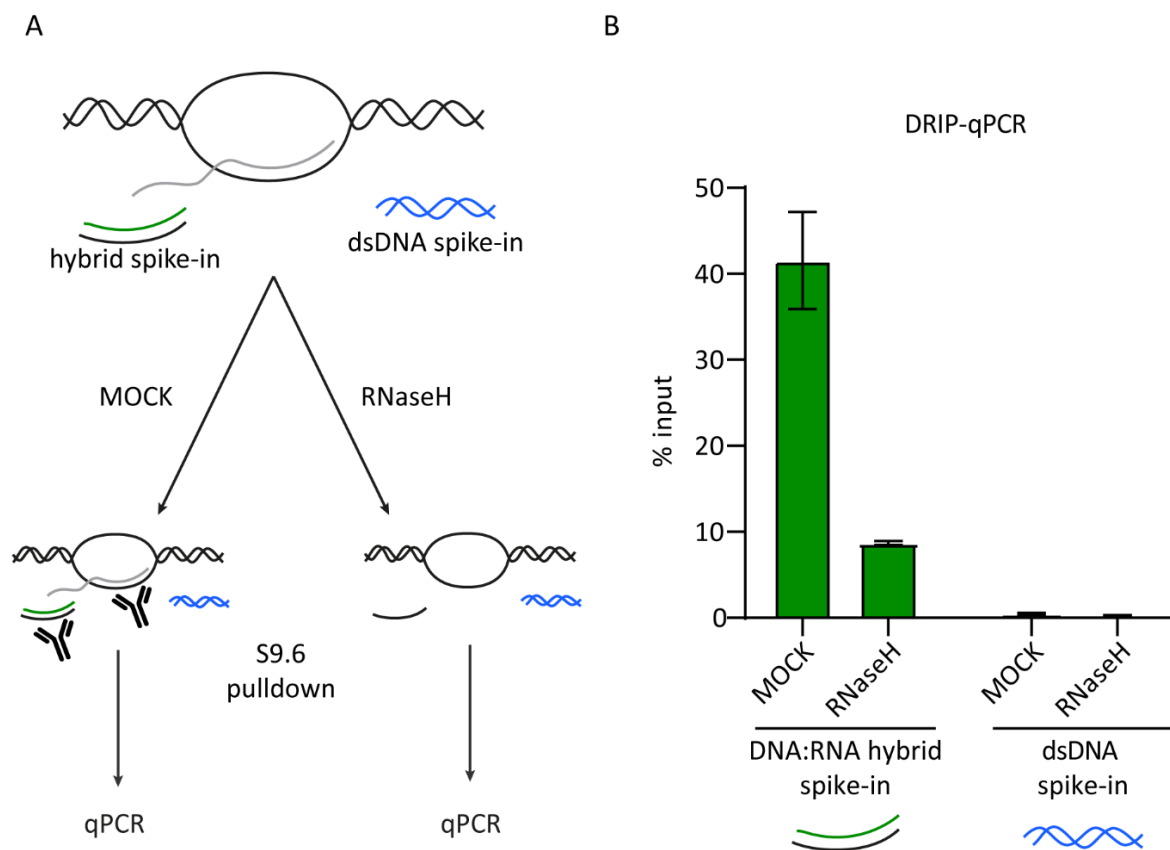


Figure 4.2-1: Validation of DNA:RNA immunoprecipitation by qPCR. (A) A DNA:RNA hybrid and a dsDNA was spiked into the genomic DNA sample. The sample is then split into two. RNase H treatment removes RNA from DNA:RNA hybrids, removing the epitope for the S9.6 antibody. (B) After DRIP, pulldown efficiency was determined by qPCR. The hybrid sequence was recovered in an RNase H-dependent manner, while dsDNA was not recovered at all.

4.2.1.1. Deoxyinosine is enriched within R-loops in mammalian genomic DNA

The DRIP-enriched DNA was degraded to nucleosides and analyzed by the LC-MS/MS screening approach to monitor deoxyribonucleosides between $m/z=229$ and $m/z=530$. The screening returned a reproducible peak of $m/z=253$, which was RNase H sensitive (Figure 4.2-2). Thus, I reasoned that this molecule was specifically enriched within R-loops.

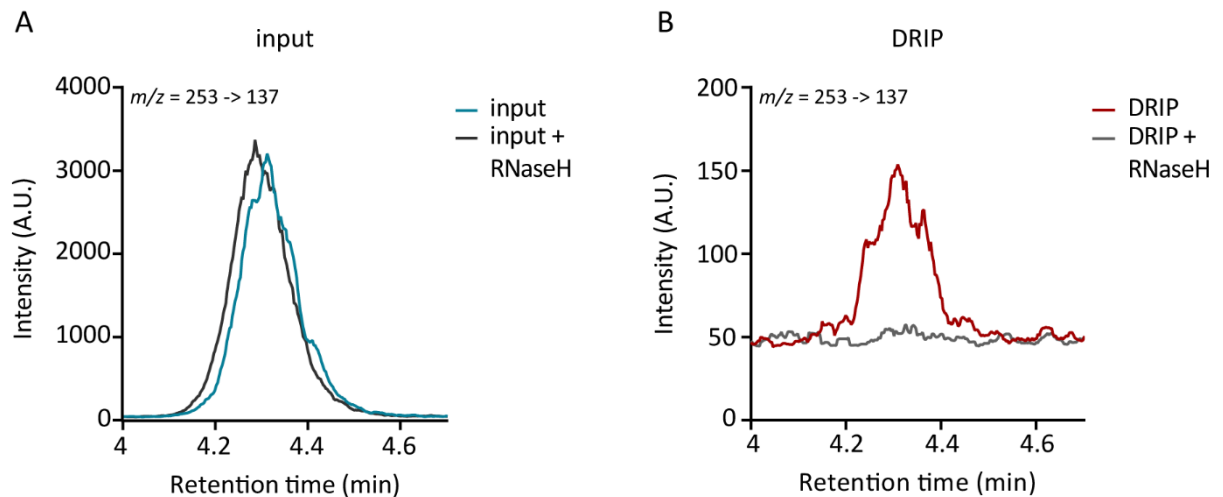


Figure 4.2-2: Peak of $m/z=253$ in DRIP is RNase H-sensitive. (A) Electropherograms of input samples in channel $m/z=253 \rightarrow 137$. The grey line represents the sample that was pre-treated with RNase H to remove RNA:DNA hybrids, the blue line represents samples that were not treated with RNase H. (B) Electropherograms of DRIP enriched DNA samples in channel $m/z=253 \rightarrow 137$. The grey line represents the sample that was pre-treated with RNase H to remove RNA:DNA hybrids, the red line represents samples that were not treated with RNase H. The peak was sensitive to RNase H-enrichment.

Intriguingly, $m/z=253$ coincides with a known nucleoside: deoxyinosine (dI) – essentially the base hypoxanthine linked to deoxyribose¹⁶⁸. LC-MS/MS analysis with a nucleoside standard compound indeed identified it as deoxyinosine (Figure 4.2-3).

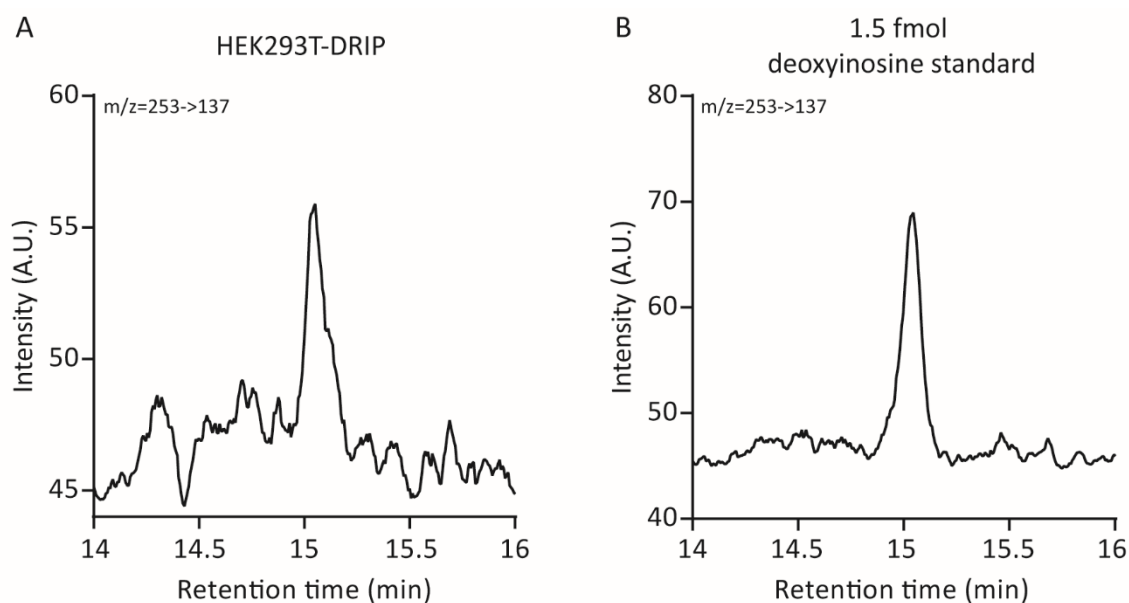


Figure 4.2-3: DRIP-enriched peak is identical to deoxyinosine. (A) LC-MS/MS electropherogram of the peak in channel $m/z=253$ that was found in HEK293T DRIP-enriched DNA. (B) LC-MS/MS electropherograms of deoxyinosine standard.

Deoxyinosine in DNA is also known to exist, but typically is regarded as DNA damage, since it can arise from spontaneous hydrolytic deamination¹⁵⁰ or nitrous deamination through nitric oxide¹⁵¹. Furthermore, dA-to-dI deamination changes the base-pairing preferences from thymidine to cytosine, and results in dA-to-dT to dG-dC transitions after DNA replication^{225,226}. However, the finding of R-loop specific dI-enrichment poses the question of a potential regulatory role of this modification in the genome. Therefore, I decided to further investigate dI in genomic R-loops by LC-MS/MS. This method allows direct detection of DNA bases, providing clear discrimination from ribo-I in RNA, is highly sensitive and provides quantitative readouts.

For accurate quantitative analysis of dI by stable-isotope dilution LC-MS/MS, I generated a ¹⁵N-labelled dI standard and established a quantitative LC-MS/MS method for dI detection. Importantly, to avoid artificial incorporation of inosine during preparation and processing of nucleic acids, all experiments were performed in the presence of 100 nM pentostatin (deoxycoformycin, dCF); an adenosine deaminase inhibitor. Pentostatin prevents spurious adenosine deamination by bacterial adenosine deaminases (ADAs) that are a prevalent contaminant of recombinant enzymes used for processing of nucleic acids for LC-MS/MS; such as alkaline phosphatase²²⁷. Additionally, to prevent hydrolytic deamination of dA to dI, which is known to occur in un-buffered conditions¹⁵³, all processing steps were performed in buffered solution.

With a quantitative LC-MS/MS method for detection of dI at hand, I measured total dA and dI levels in each sample, allowing for normalization of dI signals to the total DNA amount. Figure 4.2-4A shows the abundance of dI normalized to dA in input and DRIP-enriched DNA from HEK293T cells. In input DNA, I detected dI levels of roughly 1-2 dI per million bases, this sums up to roughly 3000 dI molecules per human genome. After DRIP-enrichment, this number increased to levels between 10 and 50 dI per million bases. The quantitative analysis confirmed both a global enrichment of dI in R-loop regions over the input and a reduction of dI levels in RNase H-pretreated DRIP samples. Thus, these quantitative analyses verified the specific enrichment of dI within mammalian genomic R-loops.

However, this data does not reveal the location of dI within the R-loop structure. It might be located within in the DNA strand that is engaged in the DNA:RNA hybrid, or in the displaced ssDNA. DNA bases in ssDNA are more prone to spontaneous deamination, while bases within a dsDNA helix are protected from hydrolytic deamination. In dsDNA, spontaneous deamination events occur at 0.5 to 0.7% of the rate of spontaneous deamination within ssDNA^{228,229} and can occur either within the cell or during handling. Accordingly, a main or exclusive location of dI within ssDNA of the R-loop could be the outcome of spontaneous deamination. To investigate, whether dI mostly occurs in ssDNA, I used Nuclease S1 (S1), an endonuclease specific for single-stranded nucleic acids, to degrade the displaced ssDNA in R-loops. I treated total gDNA, as well as input material for DRIP with S1 and quantified dI in the remaining nucleic acids by stable-isotope dilution LC-MS/MS (Figure 4.2-4B). A decrease in global dI levels or reduction of dI-enrichment in DRIPed material after S1 treatment would indicate that the largest part of genomic dI is located within ssDNA. Yet, in gDNA, global dI levels remained unchanged after removal of ssDNA by Nuclease S1 (Figure 4.2-4C). Also in DRIP-enriched DNA, S1 pre-digestion did not reduce the dI-enrichment (Figure 4.2-4D). Hence, the detected dI was present mainly in the DNA:RNA hybrid part of the R-loop, not the ssDNA. The DNA:RNA hybrid-specific location of dI raises the question on how dI is incorporated into R-loops throughout the genome and whether a modifying enzyme can be linked to these findings.

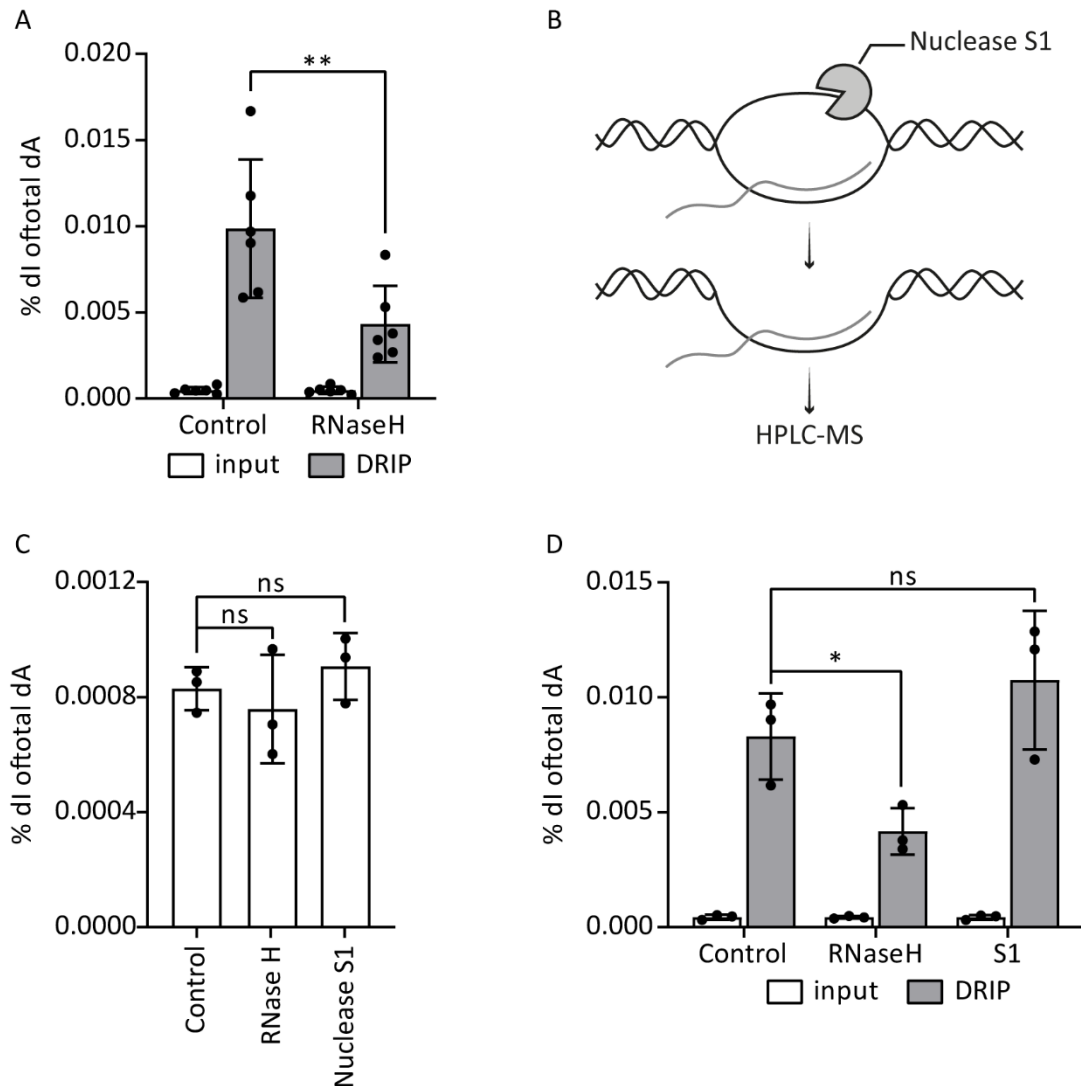


Figure 4.2-4: Deoxyinosine is enriched in DNA:RNA hybrids. (A) Relative dI levels determined by stable isotope dilution LC-MS/MS in input and DRIP-enriched DNA of HEK239T cells. Data are presented as mean, Error bars, s.d., $n = 6$ independent experiments; $**P = 0.0042$, paired two-tailed t-test). (B) Nuclease S1 is used to remove single stranded DNA from R-loops. (C) % dI/dA in genomic DNA of HEK293T pre-treated with RNase H or NucleaseS1. Data are presented as mean, Error bars, s.d., $n = 3$ independent replicates; ns= not significant, unpaired two-tailed t-test. (D) Relative dI levels determined by stable isotope dilution LCMS in input and DRIP-enriched DNA of HEK239T cells. Input material was pretreated with RNase H or Nuclease S1 prior to S9.6-pulldown. Data are presented as mean, Error bars, s.d., $n = 3$ independent experiments; $*P = 0.0294$, ns= not significant, paired two-tailed t-test.

4.3. ADAR1 acts as DNA editor in DNA:RNA hybrids

4.3.1. ADAR1 deaminates R-loops *in vivo*

ADAR1 (Adenosine Deaminase Acting on RNA 1) is an enzyme well-known for its adenosine deamination activity on dsRNA^{111,230,231}. The structural similarity between dsRNA and DNA:RNA hybrids, which both occur in A-helix conformation¹⁶⁹⁻¹⁷², suggests R-loops as promising targets for enzymatic deamination by ADAR1. To address the question, whether ADAR1 is involved in incorporation of dl into R-loops, I performed ADAR1 loss-of-function and gain-of-function experiments in HEK293T cells, followed by DRIP and dl-quantification by stable-isotope dilution LC-MS/MS.

First, I investigated dl levels in DRIP-enriched DNA after knockdown of ADAR1. The knockdown efficiency 72 hours post-transfection with an siRNA pool targeting ADAR1 (siADAR) was determined by RT-qPCR and confirmed a reduction of ADAR1 transcripts by 90-95% (data not shown).

Strikingly, knockdown of ADAR1 reduced dl levels in the DRIP by approximately 50%, compared to the control samples (Figure 4.3-1), supporting the hypothesis that R-loops can be targets for enzymatic deamination by ADAR1.

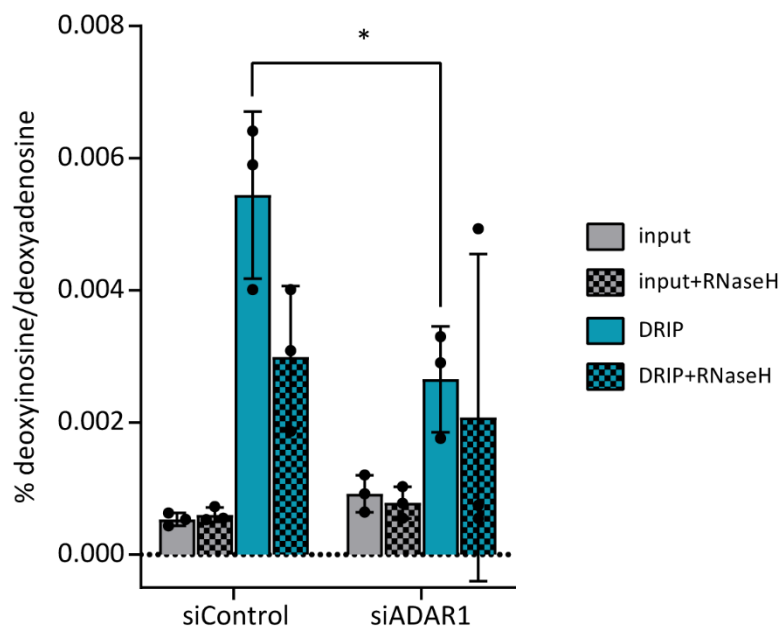


Figure 4.3-1: Knockdown of ADAR1 reduces dl in DRIPs. %dl/dA in input and DRIP-enriched DNA of HEK293T cells after knockdown of ADAR1 or control siRNA treatment. dl and dA levels were quantified by stable-isotope dilution LCMS. Data are presented as mean, Error bars, s.d., n = 3 independent experiments; *P = 0.032, unpaired two-tailed t-test.

One caveat of ADAR1 knockdown experiments is that siRNAs against ADAR1 lack discrimination between its isoforms ADAR1p110 and ADAR1p150. Furthermore, the effects of ADAR1 knockdown may result from an indirect effect of the ADAR1 protein, e.g. by recruiting other factors or blocking the repair of dl. To link the effect of ADAR1 on dl in R-loops to its catalytic activity and to understand, which ADAR1 isoform deaminates DNA within R-loops, I overexpressed both ADAR1 isoforms, ADAR1p110 and ADAR1p150, in HEK293T cells. Moreover, to link the effects of ADAR1 overexpression to its catalytic activity, I also overexpressed hyperactive and inactive mutants of each isoform. Overexpression of these mutants allows discrimination between effects of ADAR1 deamination activity and its other cellular functions. Overexpression of the ADAR1-EGFP fusion proteins was verified by fluorescence microscopy in 80-95% of cells (data not shown).

After confirming the overexpression of the ADAR1 proteins in HEK293T cells, I performed DRIP experiments and quantified dl in the input and DRIP-enriched DNA. I found that overexpression of both wild type ADAR1p110 and ADAR1p150 isoforms did not affect the extent of dl-enrichment in DRIP samples in comparison to the control. However, the overexpression of both hyperactive mutant isoforms ADAR1p110-E713Q and ADAR1p150-E1008Q caused a significant increase of dl in DRIP-enriched DNA. Intriguingly, upon overexpression of the inactive mutants of both isoforms, ADAR1p110-E617A and ADAR1p150-E912A, dl levels in R-loops showed a decreasing trend (Figure 4.3-2).

The observation that overexpression of wild type ADAR1 did not increase the dl enrichment is not surprising, since DNA is not the primary substrate for ADAR1²³². Furthermore, considering the many functions of ADAR1 that cause drastic defects when misregulated^{130,233-235}, ADAR1 activity is likely subject to tight regulation. For instance, SUMOylation of ADAR1 at lysine 418 reduces its editing activity²³⁶, either by sterically interfering with dimerization or by sequestration of SUMOylated ADAR1 to nucleoli, thereby regulating intracellular protein levels²³⁷. Analogously, ADAR2 is released from nucleoli when its substrate RNAs are overexpressed²³⁸ – suggesting sequestration of excessive or idle ADAR proteins to prevent aberrant A-to-I editing. Likely, such regulating mechanisms are sufficient to balance fluctuations of ADAR1 activity, but cannot cope with the additional challenge of overexpression of a hyperactive enzyme. This was reflected in increased dl levels in DRIPs upon expression of hyperactive, but not wild type ADAR1. The trend of decreasing dl enrichment in R-loops after overexpression of inactive ADAR1, raises the possibility that they act as dominant-negative versions. This dominant-negative effect could be exerted by competitive binding to ADAR substrates or by sequestration of endogenous ADAR1 away from potential editing sites via homodimerization with the inactive mutants.

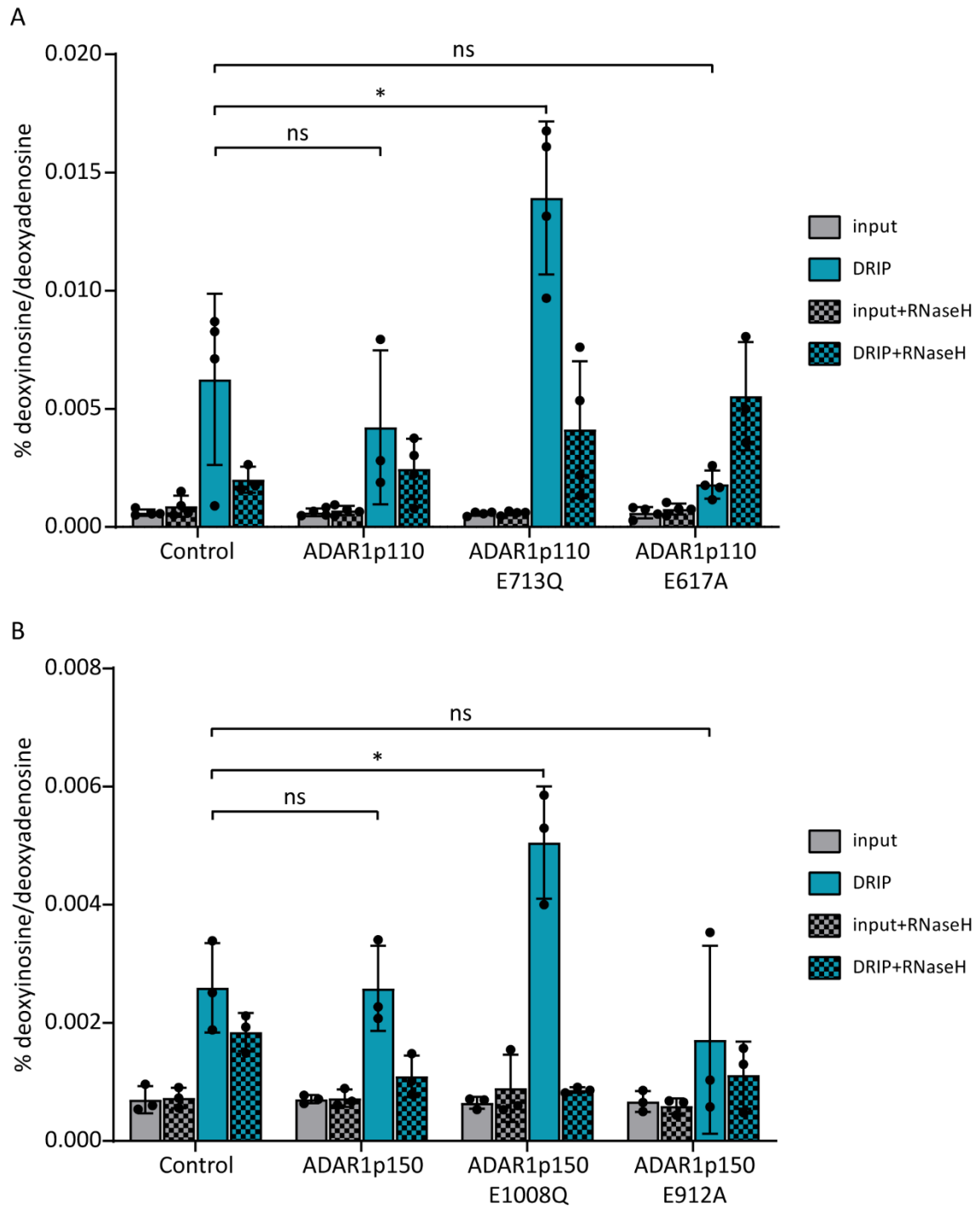


Figure 4.3-2: Catalytically hyperactive ADAR1 overexpression increases dl levels in R-loops. (A) %dl/dA in input and DRIP-enriched DNA of HEK239T cells overexpressing ADAR1p110, hyperactive ADAR1p110-E713Q, and inactive ADAR1p110-E617A. dl levels were quantified by stable-isotope dilution LCMS. Data are presented as mean, Error bars, s.d., n = 4 independent samples; *P=0.0195, ns=not significant, unpaired two-tailed t-test. The experiment was repeated three times with similar results. (B) %dl/dA in input and DRIP-enriched DNA of HEK239T cells overexpressing ADAR1p150, hyperactive ADAR1p150-E1008Q, and inactive ADAR1p150-E912A. dl levels were quantified by stable-isotope dilution LCMS. Data are presented as mean, Error bars, s.d., n = 3 independent experiments; *P=0.0249, ns=not significant, unpaired two-tailed t-test.

Indeed, inactive ADARs can repress the catalytic activity of active ADAR enzymes. Expression of the catalytically inactive ADAR3 in brain cells negatively correlates with overall RNA editing levels¹³⁷. Further, *in vitro* experiments showed that homodimerization of an active ADAR1 with an inactive ADAR1 partner reduced its deamination activity by 75%²³⁹.

The observation that both hyperactive ADAR1 isoforms increased dl in R-loops suggests that the N-terminal Z α domain, which is exclusive to ADAR1p150, is not essential for adenosine deamination within R-loops. Based on the mainly nuclear localization and constitutive expression of ADAR1p110²³⁷ this isoform is more likely to act endogenously as an R-loop targeted adenosine deaminase. This notion is supported by a very recent report that describes ADAR1p110 as regulator of telomeric R-loops via adenosine deamination²⁴⁰. However, analogous the observation of Tasakis and colleagues, dl in R-loops could be a by-product of co-transcriptional RNA-editing by ADAR1²⁴¹.

Of note, neither knockdown, nor overexpression of ADAR1 affected dl levels in input samples, suggesting a locally restricted, and R-loop specific activity of ADAR1.

Overall, the quantitative analysis of dl levels in DRIP-enriched DNA showed reduced dl in R-loops upon ADAR1 knockdown, supported by increased dl abundance in R-loops after overexpression of catalytically hyperactive ADAR1. My data indicates that overexpression of inactive ADAR1 slightly hampers endogenous ADAR1 activity on R-loops by acting as dominant-negative version. Together, these findings suggest that ADAR1 not only is involved in the incorporation of dl in R-loops, but acts as a deoxyadenosine deaminase in R-loops.

Since ADARs are well-described RNA-deaminases, I also measured ribo-I in the input and DRIP-enriched RNA after ADAR1 manipulation, as described above. I found that ribo-I levels in DRIP were not affected by the knockdown of ADAR1. Generally, ribo-I levels were not enriched in the DRIP fraction over the input, but rather slightly decreased (Figure 4.3-3).

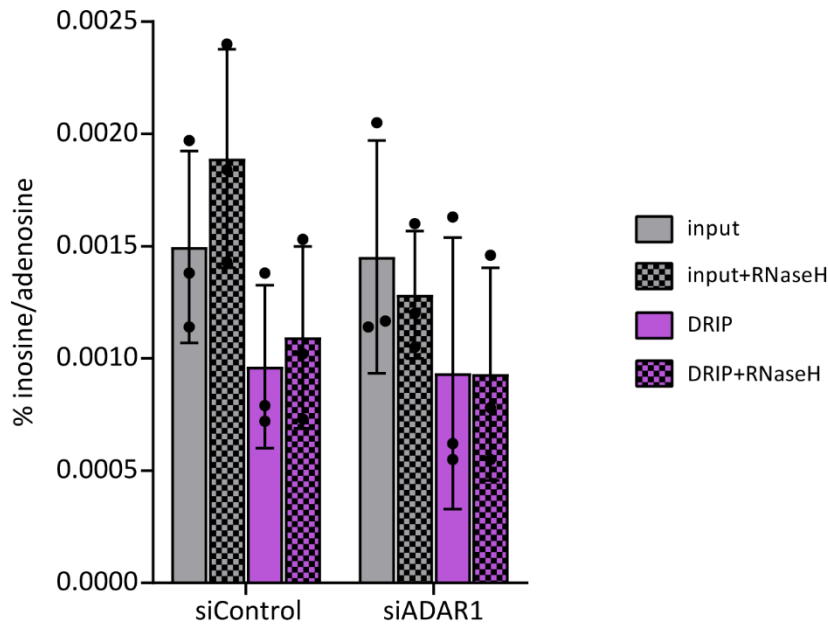


Figure 4.3-3: ADAR1 knockdown does not affect ribo-I levels in DRIP. % ribo-I/ribo-A in input RNA and RNA enriched by DRIP in HEK293T cells after knockdown of ADAR1 and transfection with siControl. Nucleotide amounts were determined by linear interpolation of a standard curve by LC-MS/MS. Data are presented as mean, Error bars, s.d., n = 3 independent experiments.

To investigate, whether wild type, hyperactive or inactive ADAR1 affects RNA editing within R-loops, I quantified ribo-I in the input and DRIP-enriched RNA by LC-MS/MS after overexpression of ADAR1 and its respective mutants. Similar to the ADAR knockdown experiment, ribo-I levels decreased after enrichment for R-loops under control conditions. However, the overexpression of wild type ADAR1 increased ribo-I levels already in the input material and entailed further RNase H-dependent enrichment of ribo-I in DRIP samples. The overexpression of hyperactive ADAR1p110-E713Q and ADAR1p150-E1008Q also increased ribo-I levels in DRIP over the input. However, this enrichment was not RNase H-sensitive. Inactive ADAR1 had no effect on ribo-I levels, which remained unchanged to the control conditions (Figure 4.3-4).

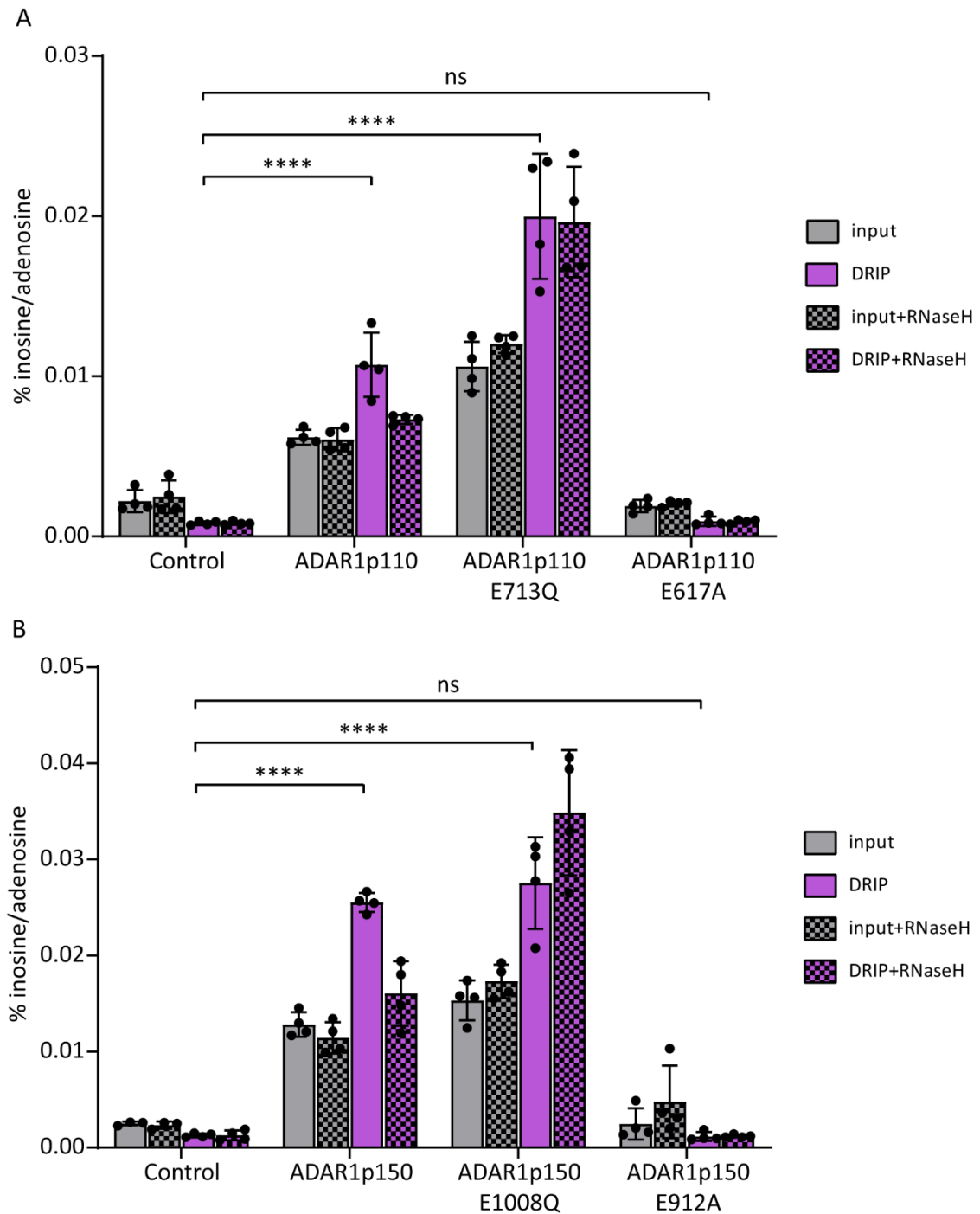


Figure 4.3-4: Overexpression of ADAR1 increases ribo-I levels in R-loops. (A) % ribo-I/ribo-A in input and DRIP-enriched RNA of HEK239T cells overexpressing ADAR1p110, hyperactive ADAR1p110-E713Q, and inactive ADAR1p110-E617A. Inosine content was determined by interpolation from a standard curve by LC-MS/MS. Data are presented as mean, Error bars, s.d., $n = 4$ independent samples; **** $P < 0.0001$, ns=not significant, unpaired two-tailed t-test. The experiment was repeated three times with similar results. (B) %I/A in input and DRIP-enriched RNA of HEK239T cells overexpressing ADAR1p150, hyperactive ADAR1p150-E1008Q, and inactive ADAR1p150-E912A. Inosine content was determined by interpolation from a standard curve by LCMS. Data are presented as mean, Error bars, s.d., $n = 4$ independent samples; **** $P < 0.0001$, ns=not significant, unpaired two-tailed t-test. The experiment was repeated three times with similar results.

Under control conditions, after knockdown of ADAR1 and after overexpression of inactive ADAR1, DRIP-enriched RNA contained less ribo-I than input samples. This indicates lower endogenous editing of R-loop-associated RNA than of other RNA species. Even though I pre-treated the input material with RNase III to eliminate dsRNA, it probably still contains residual RNA, contributing to the measured ribo-I signal. These background levels of ribo-I possibly also mask an RNase H-dependent ribo-I reduction. Since LC-MS/MS in this experimental setup does not discriminate between different RNA species, it was not possible to determine the source of the RNase H-insensitive ribo-I signals.

Upon overexpression of wild type ADAR1p110 and ADAR1p150, ribo-I levels increased already in the input samples, reflecting the generally high editing activity of ADAR1 on RNA substrates. Furthermore, an RNase H-sensitive enrichment of ribo-I in DRIP samples demonstrates an enzymatic activity of ADAR1 on RNA within R-loops. These findings are supported by recent studies, showing that, *in vitro*, ADARs deaminate DNA:RNA hybrids²³². Also upon overexpression of hyperactive ADAR1, ribo-I levels were increased in DRIP over input, however, in an RNase H-independent manner. The RNase H insensitivity could be explained by excessive ribo-I levels in dsRNAs that render them resistant to RNase III treatment²⁴². Due to the absence of DNA:RNA hybrids in RNase H-treated samples, such dsRNAs could be recovered more efficiently, thereby negating an RNase H-dependent reduction of R-loop specific ribo-I. This explanation also takes into account that dl signals in DRIPs after hyperactive ADAR1 overexpression remain RNase H-sensitive (Figure 4.3-2), as the dl signal exclusively originates from DNA hybridized to RNA within R-loop regions.

At first glance, the overall effect size in RNA and DNA seems at odds: In RNA, overexpressing wild type ADAR1 increased ribo-I levels in R-loops (Figure 4.3-4 A&B), while in DNA, it left dl levels in DRIP unaffected (Figure 4.3-2 A&B). Indeed, this is expected, since RNA is the primary substrate for wild type ADAR1. Furthermore, *in vitro* studies showed that the wild type ADAR1 catalytic domain is more active on RNA than on DNA within RNA:DNA hybrids, whereas hyperactive ADAR1 catalytic domain equally deaminates DNA and RNA within DNA:RNA hybrids²³². Hence, mechanistically, both wild type and hyperactive ADAR1 are able to deaminate DNA within DNA:RNA hybrids, with RNA being the preferred substrate. A very important factor preventing accumulation of dl in gDNA are the BER and AER mechanisms that erase the mutagenic dl¹⁵⁹⁻¹⁶⁷. These repair mechanisms might hamper dl versus ribo-I signals in DRIP. One might argue that increased dl levels upon ADAR1 overexpression reflect the higher abundance of ribo-I, that can be recycled to dITP and can be passively integrated into DNA^{158,243}. However, if dITP was passively integrated during DNA replication, its distribution over the genome would be more or less uniform. Hence, the R-loop specific enrichment of dl strongly argues against this explanation.

In summary, DRIP-LC-MS/MS analysis combined with ADAR1 manipulation by knockdown and overexpression experiments supports ADAR1 acting as DNA and RNA deaminase in R-loops *in vivo*. These findings characterize dl as an R-loop specific base modification in human DNA, resulting from catalytic deamination by both ADAR1 isoforms. Supported by *in vitro* studies that provide biochemical evidence of ADAR1's ability to edit DNA within a DNA:RNA hybrid substrate^{232,240}, my results corroborate that, also within the cellular context, ADARs deaminate DNA within R-loops. Importantly, during my studies, further supporting data came from a study published by Shiromoto and colleagues. The authors reported that ADAR1 edits both DNA and RNA in telomeric substrate hybrids *in vitro*²⁴⁰. They showed a regulatory function of ADAR1 in telomere homeostasis in alternative telomere lengthening (ALT)-negative cancer cells via deamination of DNA and RNA in telomeric R-loops²⁴⁰. Similarly, very recent results from Jimeno and colleagues suggest that ADAR-mediated A-to-I editing of RNA within R-loops is required for efficient DNA repair. They propose R-loop-destabilization by ADAR editing, which they found is essential for DNA resection in homologous recombination²⁴⁴. Even though the authors report ADAR activity only on RNA, incorporation of dl into the DNA within an R-loop might serve as a signal for subsequent DNA repair.

Altogether, I demonstrated ADAR-mediated deamination in R-loops of human gDNA, indicating that ADAR1 acts as DNA and RNA deaminase in mammalian genomic R-loops. My findings were supported by several other studies^{232,240,241,244}.

4.3.2. ADAR1 catalytic activity regulates R-loop levels

To understand the biological consequences of R-loop deamination by ADAR1, I measured global R-loop levels in HEK293T gDNA cells after ADAR1 manipulation. To quantify R-loop levels, I developed an S9.6 ELISA assay, since S9.6 dot blots proved unreliable. To do so, I optimized the assay with regards to DNA fragmentation, pretreatment with RNase H and RNase III, DNA immobilization and plate blocking conditions. I found the assay to be linear in a dynamic range between 0.001% and 1% DNA:RNA hybrids within a DNA sample of as little as 300 ng DNA. Thus, the ELISA proved a significant improvement over S9.6 dot blot assays, with greater reproducibility and sensitivity, less DNA amount required, a more feasible protocol and by providing a quantitative readout. A disadvantage of the ELISA assay was that samples, once the signal has been developed, cannot not be probed again, e.g. with an anti-dsDNA antibody as loading control. Once I established the S9.6 ELISA, I proceeded with detection of R-loop levels in HEK293T DNA after knockdown of ADAR1, or overexpression of hyperactive and inactive ADAR1p150 and ADAR1p110. The knockdown efficiency 72 hours post transfection with siADAR was determined by RT-qPCR and confirmed a reduction of ADAR1 transcripts by 90-95% (data not shown).

Detection of EGFP in 80-95% of cells by fluorescence microscopy confirmed overexpression of the EGFP-ADAR1 fusion constructs (data not shown).

ADAR1 knockdown caused a significant reduction in global R-loop levels in HEK293T DNA (Figure 4.3-5A), while leaving the global levels of dl unaffected (Figure 4.3-5D). These results indicate a stabilizing role of dl on R-loops. However, this finding is at odds with other studies, which reported increased R-loop levels upon loss of ADAR1. A recent preprint study showed that ADAR1 deficiency is associated with R-loop accumulation and DNA damage in ovarian cancer cells²⁴⁵. However, the authors quantified R-loop levels by S9.6 immunofluorescence staining, without RNase III digestion to remove dsRNA. This method poses high risks of inaccuracy, since the S9.6 antibody shows affinity to dsRNA that is much more abundant in the cell than DNA:RNA hybrids. Similarly, by S9.6 immunodotblot, Shiromoto and colleagues showed enhanced R-loop levels after knockdown of ADAR1 in HEK293T cells²⁴⁰. However, this study did not include an RNase III treatment to eliminate dsRNAs that can distort detection of DNA:RNA hybrids by the S9.6 antibody. While the other studies only investigated the effects of ADAR1 knockdown, I also investigated the effect of manipulating the catalytic activity of ADAR1 by the overexpression of hyperactive and inactive ADAR1 on global R-loop levels. Supporting my own findings of ADAR1-knockdown experiments, overexpression of catalytically hyperactive ADAR1p110-E713Q and ADAR1p150-E1008Q (Figure 4.3-5 B-C), increased the global R-loop abundance. Importantly, this effect was specific to ADAR1s deaminase activity, as inactive ADAR1p110-E617A and ADAR1p150-E912A did not affect global R-loop levels.

As discussed by Chédin et al., similarly to S9.6 dot blots, interpretation of results obtained by S9.6 ELISA require utmost caution. Both methods only capture global levels of R-loops that can be easily distorted, e.g. by mitochondrial or cytosolic²⁴⁶ R-loops and would miss more nuanced gains and losses in R-loop levels²⁰¹. R-loop formation might as well reflect changes in the cell cycle or nascent transcription^{174,247-249}. This makes it difficult to compare results from different studies, as cell culture conditions, such as cell density might differentially affect global R-loop patterns. Further, in order to obtain conclusive results, elimination of dsRNA species prior to experiments with the S9.6 antibody is required due to affinity of the antibody to dsRNA.

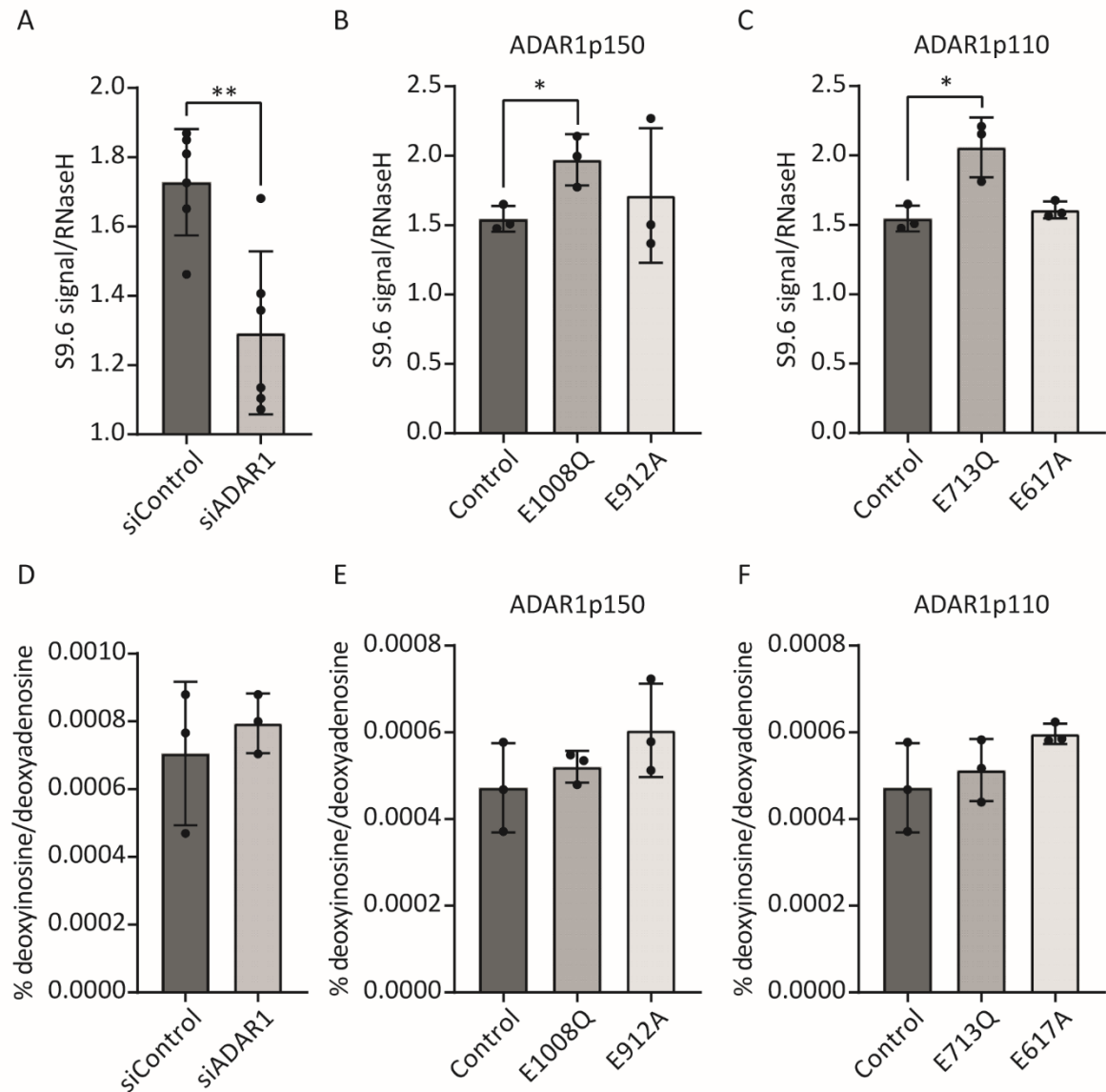


Figure 4.3-5: Aberrant ADAR1 activity coincides with global increase in R-loops. (A-C) Global R-loop levels in HEK293T gDNA determined by S9.6 ELISA. Fold S9.6 signal change over RNase H-treated gDNA. The experiments were repeated three times with similar results. (A) S9.6 signal in gDNA 72h after transfection with siADAR or siControl, as indicated. Data are presented as mean, Error bars, s.d., n = 6 samples; **P = 0.0035, unpaired two-tailed t-test. (B) S9.6 signal in gDNA 48h post transfection with hyperactive (E1008Q) or inactive (E912A) ADAR1p150. Data are presented as mean, Error bars, s.d., n = 3; *P = 0.0233, unpaired two-tailed t-test. (C) S9.6 signal in gDNA 48h post transfection with hyperactive (E713Q) or inactive (E617A) ADAR1p110. Data are presented as mean, Error bars, s.d., n = 3; *P = 0.0191, unpaired two-tailed t-test. (D-F) Genomic relative deoxyinosine content normalized to deoxyadenosine measured by stable-isotope dilution LC-MS/MS. (D) dl in HEK293T gDNA 72h after transfection with siADAR1 or siControl, as indicated. Data are presented as mean, Error bars, s.d., n = 3 samples. (E) dl levels in genomic DNA 48h post transfection with hyperactive (E1008Q) or inactive (E912A) ADAR1p150. Data are presented as mean, Error bars, s.d., n = 3. (F) dl levels in genomic DNA 48h post transfection with hyperactive (E713Q) or inactive (E617A) ADAR1p110. Data are presented as mean, Error bars, s.d., n = 3.

Overall, my results obtained by S9.6 ELISA after ADAR1 gain-and loss-of function indicate a positive effect of ADAR1 catalytic activity on global R-loop levels. Discrepancies with other studies highlight the paramount importance to identify the very genomic R-loops that are edited by ADAR1. Further investigations of such dl-rich regions will help to understand which role ADAR1 plays in mammalian R-loop regulation.

4.3.3. DNA editing by ADAR possibly regulates R-loop stability and clearance

To investigate potential regulatory interaction between R-loops and dl in the cell, I inhibited active transcription via application of known transcriptional inhibitor Dichloro-1- β -D-ribofuranosylbenzimidazole (DRB). This transcription inhibition causes reduction of R-loop levels in total genomic DNA²⁵⁰. After treatment of HEK293T cells with 100 μ M DRB for 30 min, I isolated genomic DNA, measured R-loop levels by S9.6 dot blot (performed before I established the S9.6 ELISA assay), and quantified dl by stable isotope dilution LC-MS/MS. DRB treatment caused global reduction of R-loop levels and, intriguingly, globally decreased dl levels in HEK293T gDNA (Figure 4.3-6).

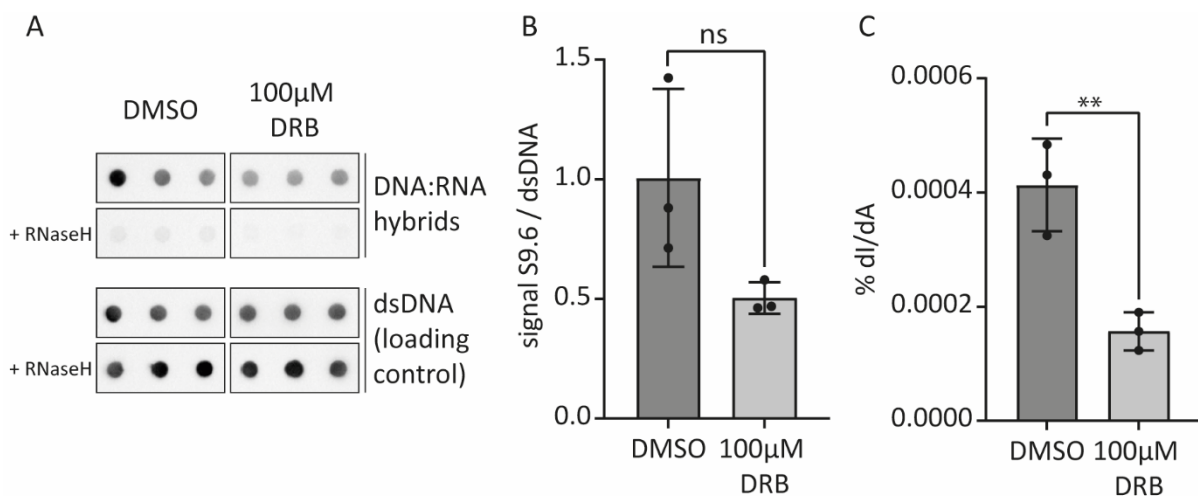


Figure 4.3-6: R-loop reduction by DRB is accompanied by a decrease in genomic dl. (A) S9.6 dotblot with genomic DNA after DRB treatment in HEK293T. (B) Quantification of panel (A) with normalization to loading control. Data are presented as mean, Error bars, s.d., n = 3 biological replicates; ns = not significant, unpaired two-tailed t-test. (C) Relative abundance of dl/dA in total genomic DNA after treatment with DRB, determined by stable isotopologue dilution LC-MS/MS. Data are presented as mean, Error bars, s.d., n = 3 replicates; **P = 0.0072, unpaired two-tailed t-test.

The co-occurrence of reduced R-loop levels with decreasing dl levels, suggests a model in which R-loop formation precedes ADAR-mediated DNA deamination: Reduction of R-loops by DRB causes

withdrawal of DNA:RNA hybrids as substrates for ADAR1 deamination. Consequently, global dl levels rapidly decrease in the absence of R-loops. In turn, manipulation of ADARs failed to change global dl levels (Figure 4.3-5). The fact that rather the presence of R-loops, instead of ADAR1 affects global dl levels, suggests a role for DNA deamination in R-loop clearance rather than in R-loop formation. Concordant with this hypothesis, presence of dl and ribo-I in DNA:RNA hybrids makes them more susceptible to digestion by RNase H²⁴⁰. Additionally, the rapid reduction of dl upon transcription inhibition highlights the very dynamic and transient nature of dl in genomic DNA. This is expected, since deoxyinosine causes undesired mutations that the cell needs to erase, e.g. by base-excision-repair (BER) via the glycosylase MPG or via the alternative excision repair pathway (AER) by EndoV^{150,167,251}. The observed swift reduction of dl upon R-loop depletion by DRB reflects the rapid repair of existing dl.

To understand biochemically, what are the consequences of dl deposition within R-loops, I performed *in vitro* RNase H sensitivity assays with synthetic DNA:RNA hybrids. This experiment can shed light on how dl affects sensitivity of DNA:RNA hybrids to digestion with RNase H. I compared unmodified hybrids to hybrids containing dl:U pairs, as they would arise from deamination of dA in a canonical dA:U base pair. Furthermore, I included hybrids with dl:C pairs, as A:C mismatches are the preferred substrates for ADAR²⁵². After incubation with recombinant *E.coli* RNase H at different concentrations, the reactions were analyzed by agarose gel electrophoresis. The uncleaved band signals were quantified and normalized to the mock control.

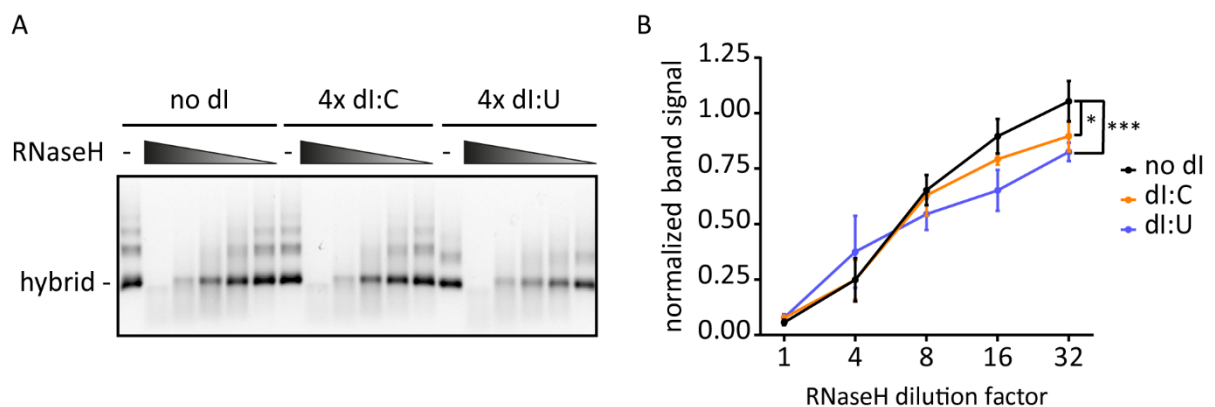


Figure 4.3-7: dl in DNA:RNA hybrids increases sensitivity to RNase H digestion. The indicated DNA:RNA hybrids were incubated with different concentrations of RNase H. (A) The reaction was analyzed by agarose gel electrophoresis to determine the remaining band intensity. (B) Band intensity was normalized to the non-treated control. Data are presented as mean, Error bars, s.d., n = 3 replicates; *P = 0.0387, **P = 0.0066, 2way ANOVA.

As shown in Figure 4.3-7, dl in both dl:U and dl:C contexts rendered the hybrids more sensitive to RNase H digestion, with a stronger effect for the dl:U pairs than for the dl:C pairs. These results suggest a role of dl in hybrid regulation by sensitizing them to degradation by RNase H. Importantly, during my studies, these observations were confirmed by other scientists who reported increased nuclease sensitivity of DNA:RNA hybrids harboring inosine²⁴⁰. The authors showed that telomeric sequence hybrids containing dl:C pairs are more sensitive to digestion by the RNase H2 complex. They corroborate that ADAR1-mediated deamination of dA:C pairs that originate from slipped hybridization with variant transcripts facilitates the resolution of telomeric R-loops by the RNase H2 complex²⁴⁰. To speculate on further R-loop regions that could be affected by ADAR1-deamination, this concept can be interpolated to other repeat regions in the genome that are subjected to slipped hybridization. Such regions could be, for instance, trinucleotide repeats. Intriguingly, in budding yeast, R-loop related CAG repeats are subjected to cytosine deamination, followed by Ung1-dependent BER, causing DNA breaks and subsequent repeat contractions²⁵³. In fact, DSB and DNA mismatch repair are known to cause expansion and contraction of repetitive genomic regions²⁵⁴ and are required for the stability of such regions²⁵⁵. Many human genetic diseases, such as Huntington's disease²⁵⁶, fragile X syndrome²⁵⁷ or the Friedreich's ataxia²⁵⁸ are partially caused by extensive simple repeat amplifications – highlighting the importance of tightly regulating these regions. It is tempting to speculate that in those occasions where dl in R-loops is produced by ADAR1, it might serve as a safeguard of the genome by recruiting the BER machinery, thereby restricting excessive trinucleotide expansion. Intriguingly, Sanz *et.al.* found accumulation of R-loops at repeat regions in patients with Aicardi-Goutières syndrome, a disease also linked to mutations in ADAR1²⁵⁹.

Complementary methods, i.e. deoxyinosine-pulldown sequencing, can be applied to identify whether also other genomic R-loop regions contain dl and are regulated by the enzymatic activity of ADARs. As discussed above, introduction of dl to R-loop regions could serve as a signal for other mechanisms to maintain genomic stability.

In summary, my results support that dl is an ADAR-introduced DNA modification in DNA:RNA hybrids of mammalian R-loops. Supported by other reports on ADAR1 as DNA editor in DNA:RNA hybrids *in vitro*^{232,240,241}, I provide a direct and quantitative read-out of DNA adenosine deamination by ADAR1, using LC-MS/MS. Thereby, I demonstrate that dl is introduced into DNA:RNA hybrids by the catalytic activity of ADAR1 in human genomic R-loops. I further corroborate that the increased catalytic activity of ADAR1 causes elevation of R-loop levels globally, while the inactive enzyme has no effect on R-loop levels. Consistently, I found reduced R-loop levels after knockdown of ADAR1. However, these findings contradict other reports showing a negative effect of ADAR1 on global R-loop levels. Upon depletion

of global R-loops, I found concomitant reduction of dl levels in gDNA, suggesting a role of dl in the resolution of R-loops. Mechanistically, this function could be exerted by R-loop eraser enzymes, supported by my own and other researchers' results²⁴⁰ that showed higher sensitivity of DNA:RNA hybrids to digestion by DNA:RNA hybrid specific endonucleases. Future research may elucidate the biological significance of R-loop-targeted DNA deamination by ADAR1 and test whether dl is a functional factor of R-loop regulation and genome stability.

4.4. Origin of m⁶dA in mammalian DNA

In contrast to the well-studied 5-methylcytosine mark and its oxidative derivatives, the role of N⁶-methyl deoxyadenosine (m⁶dA) in mammalian genomic DNA is rather controversial. On the one hand, some studies report the presence and enzymatic regulation of m⁶dA^{38,39}, while others failed to detect this mark by ultrasensitive antibody-free techniques^{40,41}. Quantitative bulk analysis of m⁶dA in mammalian genomic DNA is complicated by its low abundance of approximately 100 bases per diploid mammalian genome²⁶⁰. Furthermore, m⁶dA levels are 5-6 fold higher in bacterial DNA that can be a main contaminant in sequence-independent LC-MS/MS, thereby falsely inflating the signal.

I was involved in a study to understand how m⁶dA is incorporated into the mammalian gDNA. My colleague Michael Musheev identified the nucleotide salvage pathway as the source of mammalian genomic m⁶dA. My role in this study was to do interfere with different steps of the pathway and to measure the impact on m⁶dA levels in gDNA.

To measure m⁶dA that can be traced back to mammalian DNA specifically, cell culture medium was supplemented with heavy stable isotope labelled methionine (Met+4), which is the precursor of the methyl-donor S-adenosylmethionine (SAM)²⁶¹. Cultured cells incorporate the labelled methyl group into m⁶dA, causing a *m/z* increase of +4, which serves as metabolic identifier for its mammalian origin and discriminates against unlabeled m⁶dA from bacterial DNA.

In fact, ribo-m⁶A is the most abundant internal modification in eukaryotic mRNA and has been studied extensively in recent years, revealing many different regulatory functions of this modification in RNA^{94,262}. We hypothesized that that m⁶dA might be a “recycling product” of its RNA counterpart, ribo-m⁶A. To test this hypothesis, I impaired the mammalian nucleotide salvage pathway, which converts ribonucleotides, e.g. from degraded RNA, to deoxyribonucleotides²⁶³. These are used for DNA replication by DNA polymerases. First, I knocked-down adenosine deaminase-like protein (Adal or ADAL) by RNA interference in NIH3T3 and HEK293T cells. Since Adal transforms cyclic AMP to inosine monophosphate by deamination²⁶⁴, the expectation was that by increasing the pool of one of the precursor molecules, m⁶dA in genomic DNA will be elevated accordingly (See schematics in Figure 4.4-1A). Indeed, transfection with siAdal in mouse NIH3T3 fibroblasts, as well as treatment with siADAL in human HEK293T cells, increased the abundance of m⁶dA in genomic DNA relative to non-targeting siRNA controls (Figure 4.4-1 B and C). Additionally, I targeted one of the master regulators of the nucleotide salvage pathway, ribonucleotide reductase (RNR), by treatment with the RNR inhibitor hydroxyurea (HU). RNR converts ribonucleotides to their corresponding deoxyribonucleotides^{265,266}. Hence, inhibition of RNR may cause depletion of m⁶dADP, the precursor of m⁶dATP, which can be incorporated to DNA by polymerases. Further supporting the hypothesis that m⁶dA is a product of

nucleotide recycling, inhibition of the nucleotide salvage pathway with HU reduced the abundance of m⁶dA in genomic DNA of NIH3T3 mouse fibroblasts (Figure 4.4-1B).

The results presented in Figure 4.4-1 support the idea that m⁶dA in mammalian genomic DNA originates from N6-methylated riboadenosine that is converted to its deoxy-analog via the nucleotide salvage pathway and is then incorporated by DNA polymerases as m⁶dATP.

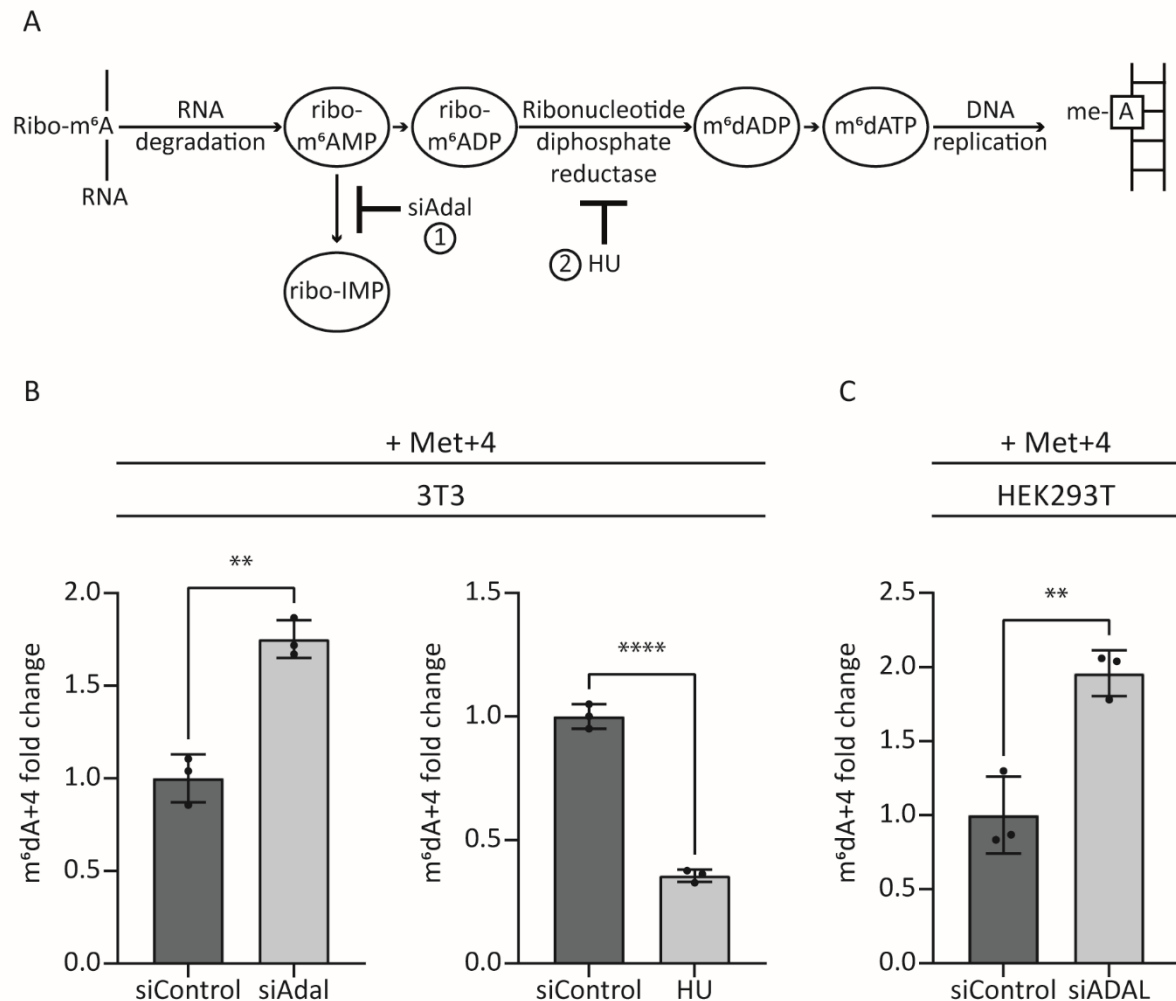


Figure 4.4-1: Inhibiting enzymes of the nucleotide-salvage pathway affects genomic m⁶dA levels. (A) LC–MS quantification of genomic m⁶dA levels in 3T3 cells treated with (left) siRNA targeting adenosine deaminase-like RNA (siAdal) (right), ribonucleotide reductase inhibitor hydroxyurea (HU). Data are presented as mean, Error bars, s.d.; n = 3, **P = 0.0014, ****P = <0.0004 two-tailed, unpaired Student t-test. (B) LC–MS quantification of genomic m⁶dA levels in HEK293-T cells treated with siRNA targeting adenosine deaminase-like RNA (siADAL). Data are presented as mean, Error bars, s.d.; n = 3, **P = 0.005, two-tailed, unpaired Student t-test.

Further, I wanted to confirm *in vitro*, that mammalian DNA polymerases tolerate m⁶dATP as a substrate for DNA synthesis. Therefore, I performed a primer extension assay, in which a DNA template was

incubated with a primer and dNTP cocktails containing either 100% dATP or 75% dATP and 25% m^6dATP in HeLa cell extracts. The DNA polymerases contained in the cell extract can copy the DNA template. After purification, I measured the incorporation of m^6dATP by quantitative LC-MS/MS. The results shown in Figure 4.4-2 demonstrate that mammalian polymerases readily incorporate m^6dATP during DNA replication. The incorporation efficiency, however, is lower than what would be expected considering the ratio of m^6dATP to other nucleotides in the primer extension reaction. This is most likely caused by the proofreading function of mammalian DNA polymerases. These proofreading functions generally minimize incorporation of unmodified bases into the genome, since they might be detrimental to genome stability.

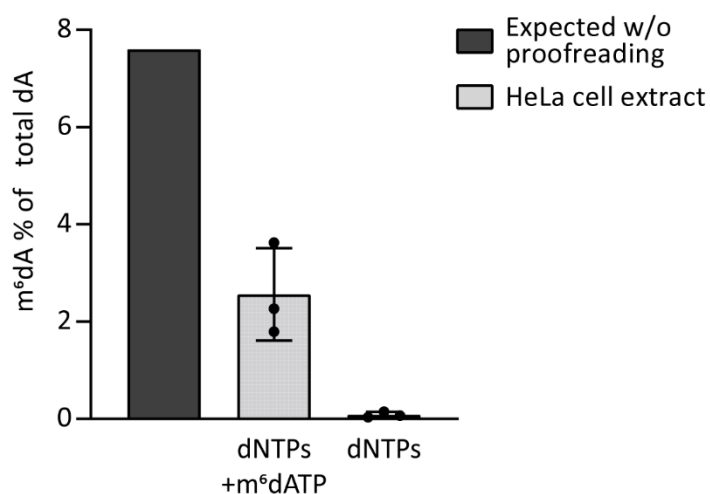


Figure 4.4-2: Mammalian polymerases readily incorporate m^6dATP . Stable isotope dilution LC-MS measurements of m^6dA in a primer extension product using HeLa cell lysate as a source of DNA polymerases. dNTPs with or without m^6dATP were added as indicated. Data are presented as mean, error bars = s.d., $n = 3$ independent reactions.

The previous results comprehensively show that ribo- m^6A from RNA can be recycled into m^6dATP and then incorporated within DNA by mammalian DNA polymerases. This concept, nevertheless, does not exclude an enzymatic N6-methylation of adenine in DNA.

To further address this open question, my colleague Michael Musheev performed double-labelling experiments with Met+4 and isotopically labelled deoxyadenosine, and could not detect a product of direct transfer of SAM to deoxyadenosine²⁶⁰, as it would be expected for a direct DNA N6-adenosine methylation. In a similar stable isotope labelling approach, data from Liu *et.al* supported our model by demonstrating incorporation of m^6dA via DNA polymerase λ and the absence of a N6-DNA methylase²⁶⁷.

Overall, we propose a model, in which the bulk of m^6dA in mammalian genomic DNA originates from free ribo- m^6A that is recycled via the nucleotide salvage pathway and incorporated to DNA as m^6dATP by DNA polymerases²⁶⁰. The model can also explain the changed m^6dA levels reported along with changes in ribo- m^6A , from gain- and loss-of-function experiments with RNA-specific m^6A demethylase FTO^{104,268}. The changes observed in m^6dA levels could be an indirect effect, reflecting fluctuations within nucleotide pools after manipulating ribo- m^6A levels.

Even if these results demonstrate that the bulk of genomic m^6dA originates from recycling of ribo- m^6A they do not completely rule out the possibility of enzymatic methylation of adenosine in genomic DNA in mammals. To test this, I overexpressed N6AMT1 (N-6 Adenine-Specific DNA Methyltransferase 1), the methyltransferase that has been claimed to directly methylate adenosine in mammalian DNA³⁸, and measured the abundance of genomic m^6dA . Overexpression was validated through fluorescence microscopy of nuclear expressed turboGFP, which was encoded on a polycistronic construct with the coding region of N6AMT1. After 48 hours of transfection, 80-90% of the cells were positive for nuclear GFP (Data not shown). However, overexpression of N6AMT1 did not affect the levels of m^6dA in HEK293T DNA compared to the control condition (Figure 4.4-3).

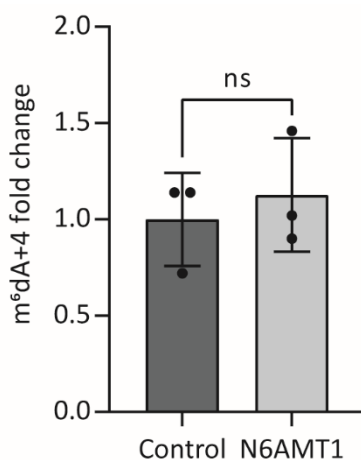


Figure 4.4-3: Overexpression of N6AMT1, a proposed N6-adenine DNA methyltransferase, does not affect m^6dA . HEK293T cells were grown in Met+4 containing medium and transfected with either a control or hN6AMT1 encoding plasmid. Cells were harvested 2 days after transfection and m^6dA+4 levels were measured by stable isotope dilution LC-MS/MS. Data are presented as mean, error bars= s.d., n= 3 biological replicates, ns=not significant, P = 0.6 two-tailed, unpaired Student's t-test.

On the other hand, a direct methylation of DNA by a methyltransferase is still possible; as a recent report indicates that *in vitro* the MettL3-MettL14 complex acts as active adenine DNA methyltransferase in the context of UV-damaged DNA²⁶⁹. Nonetheless, in agreement with the passive

incorporation of m⁶dA via polymerases the proposed mode of adenine methylation in DNA by MettL3-MettL14 at damaged DNA sites does not support a role as an epigenetic mark in DNA, but rather in DNA damage repair. In mouse DNA, MettL4 was proposed another N6-adenine DNA methyltransferase, as its overexpression increased global levels of m⁶dA, thereby preserving Polycomb silencing. Further, the authors found that ALKBH4 dioxygenase erased m⁶dA in mouse genomic DNA²⁷⁰. So far, biochemical evidence on the ability of MettL4 and ALKBH4 to methylate and demethylate mammalian gDNA, respectively, remains elusive. The reported effects of MettL4 and ALKBH4 can be explained by increasing ribo-m⁶A levels subjected to nucleotide salvage and subsequent higher incorporation of m⁶dATP into DNA. Intriguingly, a recent study demonstrates a MettL4 methyltransferase activity on human mitochondrial DNA *in vitro*, thereby repressing mitochondrial gene expression²⁷¹. Accordingly, the decreasing m⁶dA levels after MettL4 knockdown, reported by Kweon and colleagues²⁷⁰ might be due to contaminations with mitochondrial m⁶dA, as well.

Generally, as analyzed by Douvlataniotis and colleagues, detection methods of m⁶dA suffer from multiple sources of error. Bacterial and RNA contaminations, false detection of m⁶dA by SMRT sequencing due to flanking 5mC, and affinity of available m⁶dA antibodies for unmethylated adenosine and repetitive regions pose major challenges for confirming the presence and biological role of m⁶dA in mammalian genomic DNA²⁷². Our metabolic labelling strategy circumvents these caveats and allows discrimination between all contaminating m⁶dA and the labelled m⁶dA+4 derived specifically from mammalian DNA. In support of our results, specific m⁶dA reader proteins required for the mediation of potential epigenetic functions remain elusive²⁷³ and the stalling effect of m⁶dA on DNA polymerases²⁷⁴ make targeted incorporation of m⁶dA into mammalian gDNA questionable.

In summary, our data provide evidence that m⁶dA in genomic mammalian DNA originates from ribo-m⁶A that is converted to m⁶dATP and is passively incorporated by DNA polymerases. Overall, our model constrains models of m⁶dA as an epigenetic mark.

5. Material and Methods

5.1. Material

5.1.1. Equipment

15 cm ReproSil 100 C18 column (Jasco); -150°C freezer (Sanyo); -20°C freezer (Liebherr); -80°C freezer (Sanyo); agarose gel chambers (Bio-Rad); Agilent 1290 Infinity Binary LC system (Agilent Technologies); Agilent 6490 triple quadrupole mass spectrometer (Agilent Technologies); bacterial incubators (Thermo Scientific); bacterial shaker (Infors); balances (Sartorius); Bio-Dot SF Filter Paper (Bio-Rad); Bioruptor pico (Diagenode); Bioruptor plus (Diagenode); cell counter (Bio-Rad); cell culture dishes and flasks (TPP); cell culture incubators (Thermo Scientific); centrifuges (Heraeus); ChemiDoc XR+ System (Bio-Rad); ChemiDoc XRS+ System (Bio-Rad); Criterion TGX Precast Midi Protein Gels (Bio-Rad); cryo tubes (Greiner Bio-One); Cryo-Safe Cooler (Belart); DNA LoBind 1.5 ml Tubes (Eppendorf); Dotblot apparatus (Bio-Rad); DynaMag-2 Magnet rack (Invitrogen); E-Gel electrophoresis system (Invitrogen); extra thick blot filter paper (Bio-Rad); fridges (Liebherr); heating blocks (Eppendorf); Infinite M200 plate reader (Tecan); laminar flow hoods (Dometic); LightCycler 480 (Roche); low Binding 1.5 ml tubes (Eppendorf); magnetic stirrer (Heidolph); MaxiSorp Immuno 96-Well-plates flat transparent (Thermo Scientific); MaXtract High Density, 2mL (Qiagen); micro tube (Covaris); microcentrifuges (Heraeus); Microcon-10 centrifugal filters (Millipore); microscope (Leica); microwave oven (Sharp); multichannel pipettes (Sartorius); multidispenser pipette (Eppendorf); Nanodrop 2000 spectrophotometer (Thermo Scientific); Nitrocellulose membrane (VWR); orbital shaker (Neolab); PAGE midigel chambers (Bio-Rad); PCR thermocyclers (Biometra); pH meter (Mettler Toledo); pipet boy (Integra); pipettes (Eppendorf); power supplies (Bio-Rad); PVDF transfer membrane (Neolab); Qubit (Thermo Fisher); rotator (Neolab); RRHD Eclipse Plus C18, 50 mm (Agilent Technologies); silicone sealing mats (nerbe); SpeedVac concentrator (Eppendorf); test tube rotator (Kisker Biotech); test tubes (Eppendorf, Falcon, Sarstedt); Trans-Blot Turbo (Bio-Rad); Trans-Blot Turbo Midi PVDF (Bio-Rad); tubes with cell-strainer cap (Falcon); ultrapure water purification system (Millipore); ultrapure water purification system (Sartorius); UV crossliner (Stratagene); vortexer (Scientific industries); water baths (Neolab); Chemicals and pre-made buffers

5.1.2. Chemicals and pre-made buffers

0.05 % Trypsin-EDTA (1x) (Gibco); 0.1 % gelatin in ultrapure water (Millipore); 0.25 % Trypsin-EDTA (1x) (Gibco); 1-Step Ultra TMB-ELISA substrate (Thermo Scientific); 8-Azaadenosin (Tocris); Acetonitrile, LCMS grade (Sigma); agarose (Biozym); ammonium acetate (Sigma-Aldrich); ammonium acetate for HPLC>99.0% purity (Honeywell fluka); ampicillin (Sigma-Aldrich); ATP (NEB); boric acid (Sigma-Aldrich); bovine serum albumin (BSA) (Sigma-Aldrich); BSA (Sigma); BSA, Molecular Biology Grade (NEB); buffer

EB (Qiagen); chloroform-isoamylalcohol (Roth); Cutsmart buffer (NEB); dialyzed FBS (Sigma-Aldrich); dimethyl sulfoxide (DMSO) (Sigma-Aldrich); dithiothreitol (DTT) (Sigma-Aldrich); DMEM high glucose, pyruvate, no glutamine (Gibco); DMEM, high glucose, no glutamine, no methionine, no cystine (Gibco); DMEM, high glucose, no glutamine, no pyruvate (Gibco); DNA binding microplate solution (abcam); DPBS, no calcium, no magnesium (Gibco); Dichloro-1- β -D-ribofuranosyl-benzimidazole (DRB) (Sigma); EDTA (Sigma-Aldrich); ES-grade FBS (PAN Biotech); ethanol (Sigma-Aldrich); ethidium bromide (Roth); fetal bovine serum (FBS) (Lonza); Gelred Nucleic Acid Stain (Sigma); GeneAmp PCR Buffer 1, 15 mM mgCl (Thermo Fisher Scientific); glycerol (Sigma-Aldrich); glycogen, RNA grade (Thermo Scientific); hydrochloric acid (Sigma-Aldrich); hydroxyurea (Sigma); isopropanol (Sigma-Aldrich); L-Cysteine (Sigma-Aldrich); L-Glutamine 100x 200 mM (Gibco); Lipofectamine RNAiMAX (Invitrogen); L-Methionine (Sigma); L-Methionine-(methyl- ^{13}C ,d3) (Sigma-Aldrich); L-Methionine-34S (Biozol); lysis buffer AL (Qiagen); magnesium chloride (Sigma-Aldrich); MEM Non-essential amino acids 100x (Gibco); methanol (Sigma-Aldrich); nuclease-free water (Qiagen); NuPAGE LDS Sample Buffer (4X) (Invitrogen); OptiMEM (Gibco); PBS (Gibco); penicillin/streptomycin 10,000 U/ml (PAN); Pentostatin (Sigma); phenol-chloroform-isoamylalcohol (Roth); Protease inhibitor cocktail tablets EDTA-free (Roche); random primers (Invitrogen); restore™ Western Blot Stripping Buffer (Thermo Scientific); Roti C/L (Roth); skim milk powder (Sigma-Aldrich); SOC medium (Thermo Scientific); sodium acetate (Sigma-Aldrich); sodium chloride (Sigma-Aldrich); sodium dodecyl sulfate (Sigma-Aldrich); sodium hydroxide (Sigma-Aldrich); sodium pyruvate 100 mM (Gibco); β -mercaptoethanol (Sigma-Aldrich); stable L-Glytamine (Pan Biotech); SuperSignal West Femto/Pico (Thermo Fisher Scientific); Triton X-100 (Sigma-Aldrich); Trizma hydrochloride (Sigma-Aldrich); Tween-20 (Sigma-Aldrich); XL1-blue chemically competent bacteria (homemade); X-tremeGENE 9 transfection reagent (Roche); yeast tRNA (Roche); Kits

5.1.3. Enzymes

Adenosine Deaminase (Worthington Biochemical); AluI (NEB); BsrGI-HF (NEB); CpG Methyltransferase (M.SssI) (NEB); DNase I (Roche); EcoRI-HF (NEB); FastAp thermosensitive Alkaline phosphatase (Thermo Fisher); Gibson Assembly Master mix (IMB PPCF); Gibson Assembly Master Mix (NEB); HindIII-HF (NEB); M.SssI (NEB); MspI (NEB); ngTet1; Nuclease P1 (Sigma); Nuclease S1 (Invitrogen); Phosphodiesterase I from snake venom (Worthington); Phusion High-Fidelity DNA Polymerase (NEB); ProteinaseK (Qiagen); Q5 Polymerase (NEB); RNase free DNase set (Qiagen); RNaseA (Qiagen); RNase H (NEB); RNase III (IMB PPCF); SspI-HF (NEB); Superscript II reverse transcriptase (IMB PPCF); XbaI (NEB); Buffers and solutions

5.1.4. Kits

Blood & Cell Culture DNA Midi kit (Qiagen); CHIP DNA clean and concentrator (Zymo Research); DNeasy 96 Blood & Tissue kit (Qiagen); DNeasy Blood & Tissue kit (Qiagen); LightCycler 480 Probes Master (Roche); LightCycler 480 SYBR Green I Master (Roche); MEGAscript SP6 (Fisher Scientific); MEGAscript™ T7 Transcription Kit (Thermo Scientific); PCR Mycoplasma Test Kit I/C (Promokine); Protein-A agarose beads (Invitrogen); Protein-A dynabeads (Invitrogen); Protein-G agarose beads (Invitrogen); Protein-G dynabeads (Invitrogen); Qiaprep Maxiprep kit (Qiagen); Qiaprep Midiprep kit (Qiagen); Qiaprep Miniprep kit (Qiagen); Qiaquick Gel extraction kit (Qiagen); QIAquick PCR purification kit (Qiagen); QIAshredder (Qiagen); Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific); RNeasy 96 kit (Qiagen); RNeasy Mini Kit (Qiagen)

5.1.5. Buffers and solutions

Table 5.1.1: Composition of buffers and solutions used in this study

Solution	Components
Adenosine deamination buffer	50 mM sodium phosphate
DNA loading buffer (5x)	for 50 ml total volume in water: 25 ml glycerol, 0.5 ml EDTA (0.5 M), 0.5 ml Tris-HCl pH 8 (0.5 M), pinch of xylene cyanol,
dNTP mix	10 mM dATP, 10 mM dCTP, 10 mM dGTP, 10 mM dTTP
DRIP buffer 10x	Tris-HCl 100mM pH 7.5, 1.5 M NaCl, 100mM KCl, 50 mM mgCl ₂ , 5% TritonX-100
Gelred (10 000x)	Invitrogen
Luria broth (LB)	for 1 l total volume in water: 10 g bactotryptone, 5 g yeast extract, 10 g NaCl pH 7.0, autoclaved
NuPAGE 4x +DTT	4x NuPAGE LDS sample buffer including 400 mM DTT
PBS	140 mM NaCl, 2.7 mM KCl, 1.5 mM KH ₂ PO ₄ , 8.1 mM Na ₂ HPO ₄ , pH 7.4, autoclaved
PBS-0.1 % Tween (PBS-T)	1x TBS, 0.1 % Tween-20
PBS-T 1% BSA	1x PBS, 1% BSA
PBS-T 5% BSA	1x PBS, 5% BSA
PEX-AS2	150 mM NaCl, 20 mM Tris pH 7.5, 2 mM EDTA, 10 % Glycerol, 1 % Triton X-100, 1x Protease inhibitor cocktail
Primer extension buffer	40 mM HEPES/KOH pH 7.8, 70 mM KCl, 5 mM mgCl ₂ , 10 μM ZnCl ₂ , 500 μM DTT, 20 μM dNTPs
Running buffer A	5 mM ammonium acetate pH 6.9
SDS-Running buffer 10x	for 1 l total volume: 10 g SDS, 30.3 g Tris, 144.1 g glycine
SSC buffer (20x)	3.0 M sodium chloride, 0.3 M sodium citrate
TBE (10x)	1 M Trizma base, 1 M boric acid, 20 mM EDTA, autoclaved
TBS (20x)	3 M NaCl, 53.7 mM KCl, 839 M Trizma hydrochloride, 160 mM Trizma base pH 7.4, autoclaved
TBS-T	1x TBS, 0.1 % Tween-20
TE buffer	10 mM Tris-HCl pH 8.0, 1 mM EDTA, autoclaved

5.1.6. Antibodies

Table 5.1.2: Antibodies used in this study

Antibody	Supplier	Catalog number
S9.6 Antibody, 1µg/µl	homemade	
anti-dsDNA antibody, 1.01 mg/ml	abcam	ab27156
Goat-anti-mouse-HRP	dianova	115-035-146

5.1.7. Oligonucleotide sequences

Table 5.1.3: DNA oligonucleotides used for the primer extension N6mdA incorporation assay

Item	Sequence (5'→3')
reverse primer	AGGCTTCTGGACTACCTATGC
masking primer	GGTTCTCAACGAGCAGGAAGGGG
template DNA - 81-mer	CTCCTCTGACTGTAACCACGCCGATCGATCCGATCGATCA CGATCTACGATCTCGATCCGGCATAGGTAGTCCAGAAGCCT

Table 5.1.4: DNA oligonucleotides used for generation of ¹⁵N modified cytidine standards

Item	Sequence (5'→3')
Forward primer	CTCCTCTGACTGTAACCACG
Reverse primer	AGGCTTCTGGACTACCTATGC
83-mer	CTCCTCTGACTGTAACCACGCCGGTACGTTACGATACGATTACGTAATACGA TTTCGAACCGGCATAGGTAGTCCAGAAGCCT

Table 5.1.5: UPL assays used for RT-qPCR

Gene	Sequence forward (5'→3')	Sequence reverse (5'→3')	UPL probe
Mouse Adal	AGTTGGCATTGCATCTTGC	TCCCATGCCCAATTCTGT	#32
Human ADAL	GGAGAGGGAGAGGTTGAAGG	CAGCTTGCTGCAGTCTCAAG	#41
Human ADAR	TTCGAGAATCCCAAACAAGG	CTGGATTCCACAGGGATTGT	#39
Human ADARB1	GTGTAAGCACGCGTTGACTG	CGTAGTAAGTGGGAGGGAACC	#42

Table 5.1.6: UPL assays used for DRIP-qPCR

Gene	Sequence forward (5'→3')	Sequence reverse (5'→3')	UPL probe
DNA:RNA spike-in	GAAGCGGATCACATGGT	CCATGCCGAGAGTGATCC	#67
dsDNA spike-in	ATGCCTGCTTGCCGAATA	CCACAGTCGATGAATCCAGA	#31

Table 5.1.7: Oligonucleotides used for generation of spike-in R-loops and dsDNA

Item	Sequence (5'→3')
CR4-EGFP primer	GCATAGGTAGTCCAGAAGCCTCCTTGACAGCTCGTCCATGCC
SP6-CR4 primer	GATTTAGGTGACACTATAGAATACAAGCTACTTGTTCTTTTTGCAC GTGGTTACAGTCAGAGGAG
KanaR-fw primer	ATG ATT GAA CAA GAT GGA TTG
KanaR-rev primer	CTC AGA AGA ACT CGT CAA G

Table 5.1.8: Oligonucleotides used for generation of synthetic DNA:RNA hybrids

Item	Sequence (5'→3')
T7-RNA transcript	GGGAGACCCAAGCTGGCTAGCCCACCATGGAACAAAACTCATCTCAGAAGAGGA TCTGTCCGGCGGCATGGTGAGCAAGGGCGAGGAGCTGTTACCCGGGGTGGTGCCC ATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCG AGGGCGAGGGCGATGCC
DNA sequence A	GGC ATC GCC CTC GCC CTC GCC GGA CAC GCT GAA CTT GTG GCC GTT TAC GTC GCC GTC CAG CTC GAC CAG GAT GGG CAC CAC CCC GGT GAA CAG CTC CTC GCC CTT GCT CAC CAT GCC GCC GGA CAG ATC CTC TTC TGA GAT GAG TTT TTG TTC CAT GGT GGG CTA GCC AGC TTG GGT CTC CCT ATA GTG AGT CGT ATT A
DNA sequence B (4 I:C pairs)	GGC ATC GCC CTC GCC CTC GCC GGA CAC GCT /ideoxyl/AA CTT GTG GCC GTT TAC GTC GCC GTC CA/ideoxyl/ CTC GAC CAG GAT GGG CAC CAC CCC GGT GAA CAG CTC CTC GCC CTT GCT CAC CAT GCC GCC GGA CAG ATC CTC TTC T/ideoxyl/A GAT GAG TTT TTG TTC CAT GGT GGG CTA GCC A/ideoxyl/C TTG GGT CTC CCT ATA GTG AGT CGT ATT A
DNA sequence C (4 I:T pairs)	GGC ATC GCC CTC GCC CTC GCC GGA CAC GCT G/ideoxyl/A CTT GTG GCC GTT TAC GTC GCC GTC C/ideoxyl/G CTC GAC CAG GAT GGG CAC CAC CCC GGT GAA CAG CTC CTC GCC CTT GCT CAC CAT GCC GCC GGA CAG ATC CTC TTC TG/ideoxyl/ GAT GAG TTT TTG TTC CAT GGT GGG CTA GCC /ideoxyl/GC TTG GGT CTC CCT ATA GTG AGT CGT ATT A

5.1.8. siRNAs

Table 5.1.9: siRNAs used in this study

Target	Supplier	Catalog number
mouse Adal	Horizon Discovery	M-160109-00-0005
human ADAL	Horizon Discovery	M-022232-01-0005
human ADAR	Horizon Discovery	M-008630-01-0005
human ADARB1	Horizon Discovery	M-009263-01-0005

5.1.9. Plasmids

pBS,IMB-ID: 373; pEGFP-C1,IMB-ID: 377; pEF-FH-hFL-Tet1,IMB-ID: 730; pLOC-DNMT3B,IMB-ID: 1853; pLOC-AICDA,IMB-ID: 1854; pLOC-HEMK1,IMB-ID: 1855; pLOC-SMYD3,IMB-ID: 1856; pLOC-KAT7,IMB-ID: 1858; pLOC-PARP1,IMB-ID: 1859; pLOC-PUS3,IMB-ID: 1860; pLOC-OGT,IMB-ID: 1862; pLOC-TNKS2,IMB-ID: 1864; pLOC-EZH1,IMB-ID: 1868; pLOC-NAA10,IMB-ID: 1872; pLOC-DIMT1,IMB-ID: 1887; pLOC-SUV39H2,IMB-ID: 1888; pLOC-ALKBH3,IMB-ID: 1890; pLOC-TRMT61A,IMB-ID: 1891; pLOC-TRIT1,IMB-ID: 1892; pLOC-QTRT1,IMB-ID: 1893; pLOC-SETMAR,IMB-ID: 1894; pLOC-PCMTD2,IMB-ID: 1895; pLOC-SETD7,IMB-ID: 1896; pLOC-POLM,IMB-ID: 1897; pLOC-CDYL,IMB-ID: 1898; pLOC-SETDB1,IMB-ID: 1899; pLOC-PRDM5,IMB-ID: 1900; pLOC-SETDB2,IMB-ID: 1902; pLOC-TRDMT1,IMB-ID: 1903; pLOC-RNMT,IMB-ID: 1904; pLOC-GTF3C4,IMB-ID: 1905; pLOC-CDY2B,IMB-ID: 1906; pLOC-MTA2,IMB-ID: 1908; pLOC-NOP2,IMB-ID: 1909; pLOC-RRM2B,IMB-ID: 1910; pLOC-APOBEC3A,IMB-ID: 1912; pLOC-MBD4,IMB-ID: 1914; pLOC-RTCA,IMB-ID: 1917; pLOC-ARNTL,IMB-ID: 1918; pLOC-SETD3,IMB-ID: 1921; pLOC-MBD5,IMB-ID: 1922; pLOC-DNMT3A,IMB-ID: 1924; pLOC-ASH2L,IMB-ID: 1926; pLOC-OSGEP,IMB-ID: 1929; pLOC-KDM5C,IMB-ID: 1933; pLOC-PHF8,IMB-ID: 1947; pLOC-NAT14,IMB-ID: 1948; pLOC-PRDM7,IMB-ID: 1950; pLOC-METTTL10,IMB-ID: 1951; pLOC-METTTL22,IMB-ID: 1952; pLOC-METTTL21A,IMB-ID: 1953; pLOC-SETD6,IMB-ID: 1954; pLOC-NTMT1,IMB-ID: 1955; pLOC-GLYR1,IMB-ID: 1956; pLOC-TET2,IMB-ID: 1964; pLOC-CTCFL,IMB-ID: 1965; pLOC-MBD3L1,IMB-ID: 1966; pLOC-METTTL4,IMB-ID: 1967; pLOC-CMTR2,IMB-ID: 1968; pLOC-TRMT44,IMB-ID: 1970; pLOC-METTTL2A,IMB-ID: 1971; pLOC-KDM8,IMB-ID: 1972; pFLAG-CMV-D11-METTTL3,IMB-ID: 2001; pLX304-Blast-V5-EHMT2,IMB-ID: 2065; pLX304-Blast-V5-METTTL1,IMB-ID: 2066; pLX304-Blast-V5-SUV420H1,IMB-ID: 2067; pLX304-Blast-V5-TRMT1,IMB-ID: 2068; pLX304-Blast-V5-METTTL8,IMB-ID: 2069; pLX304-Blast-V5-METTTL14,IMB-ID: 2070; pLX304-Blast-V5-METTTL16,IMB-ID: 2071; pLX304-Blast-V5-METTTL21B,IMB-ID: 2072; pLX304-Blast-V5-METTTL15,IMB-ID: 2073; pLX304-Blast-V5-KDM4B,IMB-ID: 2074; pLX304-Blast-V5-NSUN2,IMB-ID: 2075; pLX304-Blast-V5-TRMT6,IMB-ID: 2076; pLX304-Blast-V5-CDKAL1,IMB-ID: 2077; pLX304-Blast-V5-METTTL25,IMB-ID: 2078; pLX304-Blast-V5-METTTL9,IMB-ID: 2079; pLX304-Blast-V5-ALKBH7,IMB-ID: 2080; pLX304-Blast-V5-PARP15,IMB-ID: 2081; pLX304-Blast-V5-MINA,IMB-ID: 2082; pLX304-Blast-V5-TET3,IMB-ID: 2083; pLX304-Blast-V5-KMT2E,IMB-ID: 2084; pLX304-Blast-V5-KDM1A,IMB-ID: 2085; pLX304-Blast-V5-EGLN2,IMB-ID: 2086; pLX304-Blast-V5-OGFOD1,IMB-ID: 2087; pLX304-Blast-V5-FTO,IMB-ID: 2088; pCMV6-KMT5C,IMB-ID: 2089; pCMV6-HR,IMB-ID: 2090; pCMV6-HTATIP,IMB-ID: 2091; pCMV6-MBD6,IMB-ID: 2092; pCMV6-PAM,IMB-ID: 2093; pCMV6-METTTL5,IMB-ID: 2095; pCMV6-METTTL7B,IMB-ID: 2096; pCMV6-METTTL13,IMB-ID: 2097; pCMV6-TYW1,IMB-ID: 2098; pFLAG-CMV2-TNKS,IMB-ID: 2099; pCMV-KDM6A,IMB-ID: 2100; pCMV-KDM6B,IMB-ID: 2101; pCMV-KDM4A,IMB-ID: 2102; pCI-KAT2B,IMB-ID: 2103; pcDNA3.1-MBD2,IMB-ID: 2104; pcDNA3.1-MBD1,IMB-ID: 2105; pCMV-KDM4C,IMB-ID: 2106;

pcDNA3-KDM5A,IMB-ID: 2107; pCMVHA-EZH2,IMB-ID: 2108; pCMVb-EP300-myc,IMB-ID: 2110; pEGFP-C1-SETD2,IMB-ID: 2113; pEGFP-C1-NAT10,IMB-ID: 2114; p3×FLAG-CMV-14-LanCL2,IMB-ID: 2442; pCS2-PHF2-NLStGFP,IMB-ID: 3107; pCS2-ALKBH5-NLStGFP,IMB-ID: 3108; pCS2-SMYD2-NLStGFP,IMB-ID: 3109; pCS2-MRPP1-NLStGFP,IMB-ID: 3110; pCS2-MBD3L2-NLStGFP,IMB-ID: 3111; pCS2-PRDM6-NLStGFP,IMB-ID: 3112; pCMV-SPORT6-MTFMT,IMB-ID: 3124; pCMV-SPORT6-RNMTL1,IMB-ID: 3125; pLX304-RG9MTD3,IMB-ID: 3127; pLX304-C9orf156,IMB-ID: 3128; pLX304-C2orf60,IMB-ID: 3129; pCS2-N6AMT1-IRES-NLS-tGFP,IMB-ID: 3412; pCS2-PRMT7-IRES-NLS-tGFP,IMB-ID: 3511; pCS2-TRMT10C-IRES-NLS-tGFP,IMB-ID: 3512; pCS2-Kdm2b-IRES-NLS-tGFP,IMB-ID: 3513; pCS2-DcpS-IRES-NLS-tGFP,IMB-ID: 3514; pCS2-Rev1-IRES-NLS-tGFP,IMB-ID: 3515; pCMV-D11-HCAP1,IMB-ID: 3516; pCMV-D11-TRM5,IMB-ID: 3517; pCMV-D11-PUS1,IMB-ID: 3518; pCMV-D11-TRMT112,IMB-ID: 3519; pmGFP-ADAR1-p150,IMB-ID: 3522; pmGFP-ADAR1-p110,IMB-ID: 3523; pmGFP-ADAR2,IMB-ID: 3524; Ph034_pmGFP-ADAR1-p150_E1008Q_hyper,IMB-ID: 3752; Ph035_pmGFP-ADAR1-p150_E912A_inact,IMB-ID: 3753; Ph036_pmGFP-ADAR1-p110_E713Q_hyper,IMB-ID: 3754; Ph037_pmGFP-ADAR1-p110_E617A_inact,IMB-ID: 3755

5.1.10. Nucleoside and nucleotide standards

Table 5.1.10: Nucleosides and nucleotides used in this study

Item	Supplier
Inosine-5'-triphosphate	Jena Biosciences
2'-Deoxyinosine	Sigma
2'-Deoxyadenosine monohydrate	Sigma
5-Hydroxy-2-deoxycytidine	Berry & Associates
5-Formyl-2'-deoxycytidine	Berry & Associates
5-Carboxy-2'-deoxycytidine	Jena Biosciences

Table 5.1.11: Stable isotopologue-labelled nucleosides and nucleotides used in this study

Item	Supplier
¹⁵ N ₃ -dCTP	Silantes
¹³ C ₉ -dCTP	Silantes
D ₃ -5mC	TRC
¹³ C ₁₀ -dAMP	Silantes
¹⁵ N ₅ -dAMP	Silantes

5.1.11. Software

Name	Version
Agilent MassHunter Quantitative Analysis	B.05.02
Agilent MassHunter Quantitative Analysis	B.09.00
Agilent MassHunter Qualitative Analysis	B.06.00
Agilent Mass Profiler Professional	V13.0
BioRad Image Lab	5.2.1
Fiji Image J	V5
GraphPad Prism	9.2.0
Tecan i-control	2.0.10.0

5.2. Methods

5.2.1. Molecular biology

Standard molecular biology methods including preparation of chemically competent XL1-blue *Escherichia coli* (*E. coli*) bacteria, plasmid amplification in *E. coli*, spectrophotometric quantification of DNA and RNA, restriction digests, PCR, agarose gel electrophoresis, and SDS-PAGE were carried out as previously described²⁷⁵. All oligonucleotides used in this study were synthesized by Sigma-Aldrich or Integrated DNA Technologies (IDT). Plasmid DNA was sequenced by GATC Biotech or StarSEQ.

5.2.1.1. Amplification and purification of plasmids from bacteria

To amplify plasmids, 40 μ l of freshly thawed chemically competent XL1-blue *E. coli* were transformed with 100 ng of plasmid DNA by a 45-second heat shock at 42 °C. This was followed by a 2 min incubation on ice, subsequent incubation in 200 μ l SOC medium for 1 h at 37 °C. 100 μ l of bacteria suspension was plated on an LB-Agar plate containing the appropriate selection antibiotic and incubated at 37 °C overnight. On the next day, a single clone was picked with a sterile pipet tip and used to inoculate 5 ml (for Miniprep) or 50 ml (for Midiprep) LB medium, containing an appropriate selection antibiotic, and incubated overnight. Plasmid DNA was purified using QIAprep Spin Miniprep or QIAGEN Plasmid Midi kits according to the manufacturer's instructions. DNA amount and purity were determined with a Nanodrop 2000 spectrophotometer.

5.2.1.2. Cloning of plasmids

Plasmids generated for this study were cloned using the Gibson Assembly kit from New England Biolabs (NEB), according to the manufacturer's recommendations. In brief, PCR primers were designed using the NEBuilder Assembly Tool, PCRs were performed with Q5 polymerase. PCR-amplified inserts were purified using QIAquick Gel extraction kit, and ligated with the backbone vector using the Gibson Assembly Master Mix. Plasmids were then amplified in XL1-blue competent *E. coli*, purified and sequenced to confirm correct insertion.

5.2.1.3. Western blot

Cultured cells were harvested by cell scraping in PBS, transferred to a 15 ml or 50 ml falcon tube, and pelleted by centrifugation for 5 min at 300 x g.

For whole cell lysates, cell pellets were resuspended in protein lysis buffer. Cells were incubated on ice, and cell disruption was achieved by sonication for 20 cycles 15 sec on/off, high energy, using a Bioruptor Plus. Samples were cleared by centrifugation for 10 min at 21 000 g at 4°C. Cell lysates were mixed with 4x NuPAGE sample buffer, containing 100 mM DTT and incubated for 10 min at 70°C. Samples were separated on a precast SDS-PAGE gel by applying 90V for 2 hours. For blotting, a PVDF membrane was activated in methanol for 1 min and whatman papers were soaked in transfer buffer for 1 min. The blotting setup was assembled as followed (from bottom to top): 2 whatman papers, PVDF membrane, PAGE gel, 2 whatman papers. The transfer was performed at 0.2 A per gel for 2 hours. Signals were developed with SuperSignal West Pico or Femto Chemiluminescent Substrate and analyzed using a ChemiDoc with Image Lab software. Quantification of band intensity was performed using Image Lab Software.

5.2.1.4. Genomic DNA isolation for general Mass Spectrometry

DNA was isolated using the DNeasy Blood & Tissue kit from QIAGEN according to the manufacturer's instructions with the following modifications:

If intended for mass spectrometry, up to 5×10^6 cells were resuspended in 360 μ l lysis buffer ATL and incubated with 40 μ l Proteinase K for 2 h at 56°C, shaking at 900 rpm. Next, 3 μ l RNase A + 27 μ l buffer ATL were added and incubated for 15 min at room temperature. Subsequently, 400 μ l buffer AL and 400 μ l Ethanol were added and further steps were performed as per manufacturer's instructions.

After elution of DNA with 200 μ l AE buffer, DNA was precipitated with 3 volumes of 100% ice-cold ethanol, 0.1 volume of 7.5 M ammonium acetate, followed by centrifugation at 21 000 x g, 4°C for 1 h. The DNA pellet was washed once with 70% ethanol, the supernatant was removed and the pellet was dried for 20 min at room temperature and resuspended in nuclease-free water. 1-2 μ g DNA was degraded to nucleosides with 0.003 U nuclease P1, 0.02 U snake venom phosphodiesterase and 0.2 U alkaline phosphatase.

5.2.1.5. Genomic DNA isolation for m⁶dA quantification by Mass Spectrometry

gDNA from cultured cells was isolated using blood and cell culture DNA preparation kit (Qiagen) according to the manufacturer's protocol. Precipitated DNA was dissolved in deionized water, DNA concentration was determined with Nanodrop spectrophotometer, DNA was degraded to nucleosides

using 0.003 U nuclease P1, 0.02 U snake venom phosphodiesterase and 0.2 U alkaline phosphatase per μg of DNA.

5.2.1.6. Genomic DNA isolation for dl quantification by Mass Spectrometry

DNA was isolated using the DNeasy Blood & Tissue kit or DNeasy 96 Blood & Tissue kit from QIAGEN according to the manufacturer's instructions with the following modifications:

If intended for detection of deoxyinosine by mass spectrometry, all buffers and solutions were supplemented with pentostatin at a final concentration of 100 nM to avoid deamination by free adenosine deaminases. Up to 5×10^6 cells were resuspended in 200 μl PBS and incubated with 200 μl lysis buffer AL, 20 μl Proteinase K for 20 min at 56°C, shaking at 900 rpm. After adding 200 μl of 100% Ethanol, lysates were transferred to the spin column and further steps were performed as per manufacturer's instructions.

After elution of DNA with 200 μl AE buffer, DNA was precipitated with 3 volumes of 100% ice-cold ethanol, 0.1 volume of 7.5 M ammonium acetate, followed by centrifugation at 21 000 g, 4°C for 1 h. The DNA pellet was washed once with 70% ethanol, the supernatant was removed and the pellet was dried for 20 min at room temperature and resuspended in 0.5 mM Tris-Cl. DNA was degraded to nucleosides with 0.003 U nuclease P1, 0.02 U snake venom phosphodiesterase and 0.2 U alkaline phosphatase.

5.2.1.7. RNA isolation

RNA was isolated using the QIAshredder and RNeasy Mini kits from Qiagen, according to the manufacturer's instructions, including DNase I on-column digestion. RNA was eluted in nuclease-free water or 0.5 mM Tris-Cl and the RNA concentration was measured on a Nanodrop 2000.

5.2.1.8. RT-qPCR

cDNA synthesis was carried out with Superscript II reverse transcriptase according to the manufacturer's instructions. Briefly, 1000 ng RNA, 1 μl 10 mM dNTPs and 360 ng random primers were mixed in a final volume of 12 μl . After denaturation for 5 min at 65°C the mixture was cooled on ice for 2 min. 4 μl of 5x first-strand buffer, 2 μl of 0.1 M DTT, 1 μl Ribolock and 1 μl Superscript II Polymerase were added and samples were incubated for 10 min at 25°C, 90 min at 42°C for cDNA synthesis and 15 min at 70°C for enzyme inactivation. All incubation steps were carried out in a PCR thermocycler. cDNA was then diluted 1:50 in nuclease-free water. Per 11 μl qPCR reaction, 5 μl cDNA, 5.5 μl 2x UPL Probes Master, 0.11 μl 100 μM forward/reverse primer mixture, 0.11 μl UPL probe and 0.28 μl ddH₂O were mixed. PCR reactions were performed in technical duplicates in a 384-well format in the Roche Light Cycler 480 using the following PCR program:

Table 5.2.1: qPCR cycling program:

Temperature (°C)	Time	Ramp rate	50 Cycles
95	10 min	4.8 °C/sec	
95	10 sec	4.8 °C/sec	
60	20 sec	2.5 °C/sec	
72	1 sec	4.8 °C/sec	
4	1 min	2.5 °C/sec	

5.2.1.9. *In vitro* primer extension assay

HeLa cell extracts, prepared as described previously²⁷⁶ were kindly provided by Dr. Lars Schomacher. Primer extension reactions were performed under the following conditions: 40 mM HEPES/KOH pH 7.8, 70 mM KCl, 5 mM MgCl₂, 10 μM ZnCl₂, 500 μM DTT, 20 μM dNTPs (either with 20 μM dATP or with 16 μM dATP and 4 μM m⁶dATP), 1.5 μM reverse primer, 15 μM masking primer, 1 μM template DNA and 1 mg/ml HeLa cell extract. The reactions were incubated at 37°C for 45 min. Nucleic acids were purified by phenol-chloroform extraction, ethanol precipitated and dissolved in 400 μl of deionized water. To remove nucleotides and primers, the samples were ultra-filtrated on Microcon-10 centrifugal filters, concentrated to ~20 μl by centrifugation, diluted with 500 μl of deionized water and concentrated again to a final volume of ~20 μl. The amplicon was PAGE-purified, ethanol precipitated using glycogen as a carrier. The DNA pellets were dissolved in deionized water, digested to nucleosides and m⁶dA levels were measured as described in section 0.

5.2.1.10. DNA:RNA immunoprecipitation

Per IP sample, genomic DNA from a 10-cm dish of cultured cells was isolated using the QIAGEN DNeasy Blood and Tissue Kit as per manufacturers' instructions, with the following changes:

DNA was purified without RNase A treatment. To keep R-loops intact, excessive pipetting and vortexing was avoided. DNA was eluted in 150 μl per spin column 0.5 mM Tris-HCl. DNA from half 10 cm dish was incubated with 5 cU RNase III, 10 U *EcoRI*-HF, 10 U *BsrGI*-HF, 10 U *SspI*-HF, 10 U *XbaI*, 10 U AluI, 100 μg/ml BSA in RNase H buffer overnight. RNase H-treated controls were incubated with 10 U RNase H. DNA fragmentation was checked on a 1% agarose TBE gel. Fragmented DNA was re-purified using the QIAGEN PCR purification kit. 10 μg of DNA were used for immunoprecipitation with 10 μg S9.6 antibody in 200 μl DRIP buffer. 10% of the volume were kept as input. Per IP reaction, 40 μL of a 1:1 mixture of Protein-A and Protein-G agarose beads were blocked overnight at 4°C with 50 ng/μL yeast tRNA in DRIP buffer. The beads were washed 3 times in ice-cold DRIP buffer and incubated with the samples for 3 hours at 4°C. After 3 washing steps of 10 min at 4°C, DNA:RNA hybrids were eluted from the beads by incubation at 56°C for 30 min with 240 μg proteinaseK in DRIP buffer with 0.5% SDS and purified by CHIP DNA clean and concentrator kit. DNA was eluted in 8 μl 0.5 mM Tris-HCl.

5.2.1.11. Generation of R-loop and dsDNA spike-in oligos

The dsDNA oligo was generated by polymerase chain reaction, using NEB Q5 polymerase according to the manufacturer's instructions. A 806 base-pair sequence from the Kanamycin resistance gene encoded on the pEGFP-C1 plasmid was amplified with the KanaR-forward and KanaR-reverse primers (Table 5.1.7), using the following cycling program:

Table 5.2.2: PCR cycling program:

Temperature (°C)	Time	
95	1 min	30 Cycles
95	30 sec	
60	30 sec	
72	1 min	
72	2 min	
12	∞	

The PCR product was purified by Qiagen PCR purification kit, according to the instructions.

To generate a synthetic R-loop, first, a dsDNA oligo was generated by polymerase chain reaction, using NEB Q5 polymerase according to the manufacturer's instructions. A 805 base-pair sequence from the EGFP gene encoded on the pCS2-EGFP plasmid was amplified with the CR4-EGFP and SP6-CR4 primer (Table 5.1.7), using the cycling program indicated in Table 5.2.2. The PCR product was purified by Qiagen PCR purification kit, according to the instructions. This PCR product was used as template for in vitro transcription to generate a complementary RNA, using the MEGAscript™ SP6 Transcription Kit (Invitrogen) according to manufacturer's instructions. RNA was purified by the RNeasy Mini Kit, according to the QIAGEN's clean-up protocol. Double-stranded DNA and single-stranded RNA were mixed in buffer EB with 20-fold RNA excess, heated to 95°C for 5 min, then the thermocycler was switched off and let cool down to room temperature for 4 hours. R-loops were purified using QIAGEN PCR purification kit and eluted in QIAGEN buffer EB.

5.2.1.12. DNA:RNA Dotblot

Genomic DNA was isolated from cultured cells using QIAGEN DNeasy Blood and Tissue Kit as per manufacturers' instructions, without RNase A treatment. To keep R-loops intact, excessive pipetting and vortexing was avoided. DNA was eluted in 150 µl 0.5 mM Tris-HCl and incubated with 5 cU RNase III in RNase H buffer overnight. RNase H-treated controls were additionally incubated with 10 U RNase H. The enzymes were inactivated by incubation at 65°C for 10 min. DNA concentration was determined using Qubit dsDNA HS kit (Invitrogen), as per manufacturer's instructions. Serial dilutions between 2 µg and 6 ng DNA were prepared in 120 µl 2x SSC buffer. For blotting, a nitrocellulose membrane and whatman papers were soaked in 2x SSC buffer for 1 min and assembled in a Bio-Dot

Apparatus with the membrane on top of 2 whatman papers. The wells were washed 2 times with 200 μ l 2xSSC buffer by applying vacuum. 110 μ l of the samples were loaded to the wells and transferred to the membrane by applying vacuum. The membrane was dried for 10 min at room temperature and UV-crosslinked using a stratalinker at 3x1400 kJ. After crosslinking, the membrane was blocked for 1 hour with 5% BSA in PBS-T and then incubated on a rocker at 4°C overnight with 1:1000 S9.6 antibody in 1% BSA PBS-T. After washing 3 times 5 min at room temperature with PBS-T, the secondary antibody (goat-anti-mouse-HRP conjugated antibody from Dianova) was diluted 1:5000 in 1% BSA PBS-T and incubated for 1 hour at on a sample rocker. The membrane was washed 3 times for 10 min with PBS-T and signals were developed with SuperSignal West Pico or Femto Chemiluminescent Substrate and analyzed using a ChemiDoc with Image Lab software. Quantification of spot intensity was performed using Fiji ImageJ software. For detection of dsDNA as a loading control, the membrane was stripped by rocking with restore™ Western Blot Stripping Buffer for 30 min, blocked with 5% BSA in PBS-T for 1 hour, incubated with 1:5000 anti-dsDNA antibody in 1% BSA PBS-T for 1 hour secondary antibody incubation and signal detection were performed as described above.

5.2.1.13. Enzyme-linked Immunosorbent Assay (ELISA) for DNA:RNA hybrid detection

Genomic DNA was isolated from cultured cells using QIAGEN DNeasy Blood and Tissue Kit as per manufacturers' instructions, without RNase A treatment. To keep R-loops intact, excessive pipetting and vortexing was avoided. DNA was eluted in 150 μ l 0.5 mM Tris-HCl and incubated with 5 cU RNase III, 10 U *EcoRI*-HF, 10 U *BsrGI*-HF, 10 U *SspI*-HF, 10 U *XbaI*, 10 U *AluI*, 100 μ g/mL BSA in RNase H buffer overnight. RNase H-treated controls were additionally incubated with 10 U RNase H. Fragmented DNA was re-purified using DNA clean and concentrator kit according to the manufacturers' instructions. 350 ng DNA was bound to each well of a MaxiSorp Immuno 96-Well-plate using DNA binding microplate solution as per manufacturers' instructions. After binding, the wells were washed 3 times with 200 μ L PBS-T and binding was confirmed by measuring fluorescence at 600 nm after addition of 100 μ L 1xGelred in PBS-T with a TECAN infinite m200 plate reader. The wells were washed 3 times with 200 μ L PBS-T and blocked for 1 hour with 100 μ L 5% BSA in PBS-T. Wells were incubated overnight at 4°C shaking at 200 rpm with 100 μ L 1:1000 S9.6 antibody in 1% BSA PBS-T. The wells were washed 4 times with 200 μ L PBS-T and then incubated with 100 μ L 1:5000 goat-anti-mouse-HRP conjugated antibody (Dianova) in PBS-T for 1 hour at room temperature, shaking at 200 rpm. The wells were washed 4 times and 100 μ L 1-Step Ultra TMB-ELISA substrate was added per well and incubated in the dark at room temperature for 20 min. The colorimetric reaction was stopped by adding 100 μ L 2M sulfuric acid. Absorbance at 450 nm was measured by a TECAN infinite m200 plate reader.

5.2.1.14. Generation of synthetic DNA:RNA hybrids for RNase H sensitivity assay

Single-stranded DNA was synthesized by Integrated DNA Technologies (see Table 5.1.8). Complementary RNA was produced from a PCR product containing a T7 promoter (sequence: CGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCA CTGCTTACTGGCTTATCGAAATTAATACGACTCACTATAGGGAGACCCAAGCTGGCTAGCCCACCATGGAACA AAAACTCATCTCAGAAGAGGATCTGTCCGGCGGCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGT GCCCATCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCG ATGCC), using the MEGAscript™ T7 Transcription Kit (Invitrogen) according to manufacturer's instructions. RNA was purified by the RNeasy Mini Kit, according to the QIAGEN's clean-up protocol. Single-stranded DNA and RNA were mixed at equimolar concentrations in buffer EB, heated to 95°C for 5 min, then the thermocycler was switched off and let cool down to room temperature for 4 hours. DNA:RNA hybrids were purified using QIAGEN PCR purification kit and eluted in QIAGEN buffer EB.

5.2.1.15. *In vitro* RNase H sensitivity assay

500 ng DNA:RNA hybrid in buffer EB was treated with serial dilutions of recombinant RNase H (NEB) in RNase H buffer for 20 min at 37°C in a total volume of 10 µL. The reaction was stopped by adding 1 µL of 0.5 M EDTA. The reaction was run on a 1% agarose TBE gel, supplemented with Gelred dye. The band intensity was quantified by Fiji ImageJ software.

5.2.2. Cell culture

5.2.2.1. Thawing and freezing of cells

Frozen cells were quickly thawed in a water bath and added to 10 ml of the respective cell culture medium in a 15 ml falcon. The cell suspension was centrifuged 3 min at 300 x g, then cells were resuspended in fresh cell culture medium and plated on a cell culture dish.

For cryopreservation, cells were detached from their culture dish with Trypsin-EDTA, resuspended in the respective cell culture medium and pelleted for 3 min at 300 x g. Cells were resuspended in a volume of freezing medium (80% FBS, 20% DMSO). 1 ml aliquots of cells in freezing medium were transferred to cryo tubes and slowly frozen to -80°C at a cooling rate of 1°C/min in a Cryo-Safe Cooler filled with isopropanol. Cells were transferred to -150°C for long-term storage.

5.2.2.2. General culture of HEK293T

HEK293T were cultured in supplemented DMEM medium (DMEM with 10% FBS, 2 mM L-Glutamine) at 37°C, 5% CO₂ and 21% O₂. Cells were passaged every other day, when reaching approximately 80% confluency, by washing cells with PBS once, detaching cells with 0.05% trypsin-EDTA for 2 min, quenching trypsin with supplemented DMEM and plating 1/6th on a new cell-culture flask.

5.2.2.3. DNA transfection of HEK293T

For plasmid DNA transfection, HEK293T cells were transfected with X-tremeGENE-9 16 h after plating 0.5×10^6 HEK293T cells per cavity of a 6-well plate in supplemented DMEM medium without antibiotics. For a 6-well format, 1-2 μg of plasmid DNA were diluted in 30 μl OptiMEM, and 3.6 μl X-tremeGENE-9 were mixed with 26.4 μl OptiMEM. Both solutions were vortexed and incubated for 5 min. Afterwards, both solutions were mixed and incubated for 20 min. Subsequently, the transfection mix was added dropwise to the cells and distributed by gentle swirling of the cell culture dish. Cells were harvested 48 h post DNA transfection.

5.2.2.4. siRNA transfection of HEK293T

For siRNA transfection, HEK293T cells were transfected with Lipofectamine RNAiMAX. For a 6-well format, 0.5 μl of 20 μM siRNA were mixed with 50 μl OptiMEM, and 3 μl Lipofectamine RNAiMAX were mixed with 50 μl OptiMEM. Both solutions were vortexed and incubated for 5 min. Afterwards, the two solutions were mixed and incubated for 20 min. Meanwhile, 0.4×10^6 HEK293T cells per well were plated in 1000 μl supplemented DMEM medium without antibiotics. The transfection mix was added dropwise to the cells and distributed by gentle swirling of the cell culture dish. The cell culture medium was exchanged after 24 hours. Cells were usually harvested 72 hours *post* transfection.

5.2.2.5. Stable isotope labeling by amino acids in cell culture (SILAC) of HEK293T

In order to detect sulfur-containing DNA and RNA bases via mass spectrometry, HEK293T cells were cultured for 7 days (see 6.2.2.2) in “sulfur starvation medium” followed by 7 days with stable isotopes of L-Methionine, using “light” or “heavy” DMEM:

Sulfur starvation medium: DMEM (high glucose, no glutamine, no methionine, no cysteine) 10% dialyzed FBS, 1 mM Sodium Pyruvate, 2 mM L-Glutamine, 0.02 mM L-Cystein, 0.02 mM L-Methionine

Light DMEM: DMEM (high glucose, no glutamine, no methionine, no cysteine) 10% dialyzed FBS, 1 mM Sodium Pyruvate, 2 mM L-Glutamine, 0.02 mM L-Cystein, 0.2 mM L-Methionine

Heavy DMEM: DMEM (high glucose, no glutamine, no methionine, no cysteine) 10% dialyzed FBS, 1 mM Sodium Pyruvate, 2 mM L-Glutamine, 0.02 mM L-Cystein, 0.2 mM ^{34}S -L-Methionine

5.2.2.6. Isotopic labelling of m^6dA in HEK293T

To isotopically label m^6dA in cells, HEK293T cells were cultured for at least 6 days (see 6.2.2.2) with the following medium: DMEM (high glucose, no glutamine, no methionine, no cysteine) supplemented with 0.2 mM L-Methionine-(methyl- ^{13}C , d_3), 0.2 mM L-Cysteine, 1 mM sodium pyruvate, 10% dialyzed FBS, 2 mM stable L-Glutamine.

5.2.2.7. General culture of NIH3T3

NIH3T3 cells were cultured in supplemented DMEM medium (DMEM with 10% FBS, 2 mM L-Glutamine) at 37°C, 5% CO₂ and 21% O₂. Cells were passaged every other day, when reaching approximately 80% confluency, by washing cells with PBS once, detaching cells with 0.25% trypsin-EDTA for 2 min, quenching trypsin with supplemented DMEM and plating 1/6th on a new, cell-culture flask.

5.2.2.8. siRNA transfection of NIH3T3

For siRNA transfection, NIH3T3 cells were transfected with Lipofectamine RNAiMAX. For a 6-well format, 0.5 µl of 20 µM siRNA were mixed with 50 µl OptiMEM, and 3 µl Lipofectamine RNAiMAX were mixed with 50 µl OptiMEM. Both solutions were vortexed and incubated for 5 min. Afterwards, the two solutions were mixed and incubated for 20 min. Meanwhile, 0.4x10⁶ NIH3T3 cells per well were plated on 6-well dishes in 1000 µl supplemented DMEM medium without antibiotics and subsequently, the transfection mix was added dropwise to the cells and distributed by gentle swirling of the cell culture dish. The cell culture medium was exchanged after 24 hours. Cells were usually harvested 72 hours post siRNA transfection.

5.2.2.9. Stable isotope labeling by amino acids in cell culture (SILAC) of NIH3T3

In order to detect sulfur-containing DNA and RNA bases via mass spectrometry, NIH3T3 cells were cultured for 7 days (see 6.2.2.7) in “sulfur starvation medium” followed by 7 days with stable isotopes of L-Methionine, using “light” or “heavy” DMEM:

Sulfur starvation medium: DMEM (high glucose, no glutamine, no methionine, no cysteine) 10% dialyzed FBS, 1 mM Sodium Pyruvate, 2 mM L-Glutamine, 0.02 mM L-Cystein, 0.02 mM L-Methionine

Light DMEM: DMEM (high glucose, no glutamine, no methionine, no cysteine) 10% dialyzed FBS, 1 mM Sodium Pyruvate, 2 mM L-Glutamine, 0.02 mM L-Cystein, 0.2 mM L-Methionine

Heavy DMEM: DMEM (high glucose, no glutamine, no methionine, no cysteine) 10% dialyzed FBS, 1 mM Sodium Pyruvate, 2 mM L-Glutamine, 0.02 mM L-Cystein, 0.2 mM ³⁴S-L-Methionine

5.2.2.10. Isotopic labelling of m⁶dA in NIH3T3

To isotopically label m⁶dA in cells, NIH3T3 cells were cultured for at least 6 days (see 6.2.2.2) with the following medium: DMEM (high glucose, no glutamine, no methionine, no cysteine) supplemented with 0.2 mM L-Methionine-(methyl-¹³C,₃), 0.2 mM L-Cysteine, 1 mM sodium pyruvate, 10% dialyzed FBS, 2 mM stable L-Glutamine.

5.2.2.11. General culture of HeLa

HeLa cells were cultured in supplemented DMEM medium (DMEM with 10% FBS, 2 mM L-Glutamine) at 37°C, 5% CO₂ and 21% O₂. Cells were passaged every other day, when reaching approximately 80% confluency, by washing cells with PBS once, detaching cells with 0.25% trypsin-EDTA for 2 min, quenching trypsin with supplemented DMEM and plating 1/6th on a new, cell-culture flask.

5.2.2.12. Stable isotope labeling by amino acids in cell culture (SILAC) of HeLa

In order to detect sulfur-containing DNA and RNA bases via mass spectrometry, HeLa cells were cultured for 7 days (see 6.2.2.7) in “sulfur starvation medium” followed by 7 days with stable isotopes of L-Methionine, using “light” or “heavy” DMEM:

Sulfur starvation medium: DMEM (high glucose, no glutamine, no methionine, no cysteine) 10% dialyzed FBS, 1 mM Sodium Pyruvate, 2 mM L-Glutamine, 0.02 mM L-Cystein, 0.02 mM L-Methionine

Light DMEM: DMEM (high glucose, no glutamine, no methionine, no cysteine) 10% dialyzed FBS, 1 mM Sodium Pyruvate, 2 mM L-Glutamine, 0.02 mM L-Cystein, 0.2 mM L-Methionine

Heavy DMEM: DMEM (high glucose, no glutamine, no methionine, no cysteine) 10% dialyzed FBS, 1 mM Sodium Pyruvate, 2 mM L-Glutamine, 0.02 mM L-Cystein, 0.2 mM ³⁴S-L-Methionine

5.2.3. HPLC-MS

5.2.3.1. Generation and purification of ¹³C-labeled 5hmC, 5fC and 5caC

Stable isotope labeling of DNA modifications was performed by a series of enzymatic reactions. Firstly, a DNA strand of 83 bp length was PCR amplified using ¹³C₉-labeled dCTP (Silantes) instead of unlabeled dCTP in the reaction mixture. Purified PCR product was digested with *MspI* (NEB) to remove unlabeled primer sequences and PAGE purified on a non-denaturing 20% polyacrylamide gel. Subsequently, DNA was *in vitro* methylated with *M.SssI* (NEB), phenol/chloroform-purified and incubated with purified ngTet1 for 30 min as described previously²⁷⁷. Finally, DNA was degraded to nucleosides with nuclease P1 (Roche), snake venom phosphodiesterase (Worthington) and alkaline phosphatase (Fermentas) as described previously²⁷⁸. Individual nucleosides were separated on an Agilent 1290 Infinity Binary LC system (Agilent technologies) using a ReproSil 100 C18 column (Jasco). Isotopically labeled 5hmC, 5fC and 5caC were identified by analytical HPLC in tandem with triple quadruple mass spectrometry (Agilent 6490, Agilent Technologies) and purified by preparative HPLC. An aliquot was mixed with known concentrations of corresponding unlabeled 5hmC, 5fC and 5caC nucleosides and the concentrations of ¹³C₉-labeled nucleosides were determined after LC-MS/MS by comparing the areas of the labeled and unlabeled compounds.

5.2.3.2. Quantitative measurement of 5mC, 5hmC, 5fC, 5caC

An equal volume of isotopic standard mixture ($^{15}\text{N}_3\text{-C}$ (Silantes), $^2\text{H}_3\text{-5mC}$ (TRC) and self-synthesized $^{13}\text{C}_9\text{-5hmC}$, $^{13}\text{C}_9\text{-5fC}$ and $^{13}\text{C}_9\text{-5caC}$, (see above) was added to the DNA and about 100 ng of total DNA was injected for LC-MS/MS analysis. Quantitative analysis was performed on an Agilent 1290 Infinity Binary LC system (Agilent technologies) using a ReproSil 100 C18 column (Jasco) coupled to an Agilent 6490 triple quadrupole mass spectrometer (Agilent technologies). Running buffers were 5 mM ammonium acetate pH 6.9 (A) and Acetonitrile (B). Separations were performed at a flow rate of 0.5 ml/min using the following gradient: 8 min 0% solvent B, 8 min linear increase to 15% solvent B. Washing and reconditioning of the column was performed with a flow rate of 1.0 ml/min with 15% solvent B for 1 min and 100% buffer A for additional 5 min. During the last minute the flow rate was linearly decreased to the initial value of 0.5 ml/min. The detailed mass spectrometer settings as well as the multiple reaction monitoring (MRM) transitions are listed in Table 5.2.3.

Table 5.2.3: LC-MS/MS settings used for quantitative measurement of 5mC, 5hmC, 5fC, 5caC

Agilent 6490 triple quadrupole mass spectrometer settings

Source parameters	
Ion Mode	positive
Gas temperature	110°C
Gas flow	19 L/min
Nebulizer	25 psi
Sheath gas temperature	375°C
Sheath gas flow	11 L/min
Capillary voltage	3000 V
Nozzle voltage	0 V
iFunnel parameters	
High pressure radiofrequency	70 V
Low pressure radiofrequency	80 V

MRM transitions

Nucleoside	Precursor ion (m/z)	Fragment ion (m/z)	Collision energy	Cell accelerator voltage
C natural isotopologue +1*	229	113	5	8
5mC	241.9	126	6	8
5hmC	257.9	142	8	7
5fC	256.1	140	6	8
5caC	272	156	6	8
$^{15}\text{N}_3\text{-C}$	231	115	5	8
$\text{D}_3\text{-5mC}$	244.9	129	6	8
$^{13}\text{C}_9\text{-5hmC}$	260.9	145	8	7
$^{13}\text{C}_9\text{-5fC}$	259.1	143	6	8
$^{13}\text{C}_9\text{-5caC}$	275	159	6	8

* Detection of natural isotopomers increases the dynamic range of the method as described previously²⁷⁹

Quantification of highly abundant cytidine and 5mC was performed using 100x diluted samples. The data were analyzed with the Agilent MassHunter Quantitative Analysis software version B.05.02 (Agilent technologies) using isotopic standards to confirm the peak identity. Areas of the integrated peaks were exported into Microsoft Excel with which the areas were normalized to the area of the corresponding isotopic standard. Absolute amounts of the nucleosides were calculated using linear interpolation from a standard curve. Linear interpolation was performed using the two closely matching data points from the standard curve. Isotopic standards were spiked into the mixture of isotopic standards to normalize for ionization variability. The standard curve for every nucleoside was prepared to cover the amount of the corresponding nucleoside in the DNA sample analyzed.

5.2.3.3. LC-MS/MS screening for novel DNA base modifications by gain-of-function

Hydrolyzed gDNA samples from transfected HEK293T cells were obtained as described in section 5.2.1.4. The concentration of nucleosides in each sample was determined by quantification of cytidine by stable-isotope dilution LC-MS/MS, as described in section 5.2.3.2. Exactly 100 ng of degraded gDNA samples from HEK293T cells overexpressing different candidate enzymes were injected. Nucleosides were separated on an Agilent 1290 Infinity Binary LC system (Agilent Technologies) with a 5 cm RRHD Eclipse Plus C18 column (Agilent Technologies) and detected by a triple-quadrupole mass spectrometer (Agilent 6490, Agilent Technologies). Running solutions were 5 mM ammonium acetate, pH 6.9 (A) and acetonitrile (B). Separations were performed with the following gradient: 3 min 0% B, 4.5 min linear increase to 5% B, 3 min 5% B, 2 min linear increase to 50% B, 2.5 min 0% B. The flow rate was 10.5 min 0.3 ml/min, 2 min 0.38 ml/min, 2 min 0.5 ml min⁻¹ and 0.5 min 0.3 ml/min. The detailed mass spectrometer settings are listed in Table 5.2.4. Loss of the deoxyribose moiety was monitored as mass transitions between $m/z=229 \rightarrow 113$ and $m/z=530 \rightarrow 414$ in intervals of one Dalton (Table 5.2.4).

Table 5.2.4: LC-MS/MS settings used for screening for DNA base-modifications

Agilent 6490 triple quadrupole mass spectrometer settings

Source parameters	
Ion Mode	positive
Gas temperature	110°C
Gas flow	19 L/min
Nebulizer	25 psi
Sheath gas temperature	375°C
Sheath gas flow	11 L/min
Capillary voltage	2000 V
Nozzle voltage	0 V
iFunnel parameters	
High pressure radiofrequency	70 V
Low pressure radiofrequency	80 V

MRM transitions

Precursor ion (m/z)	Fragment ion (m/z)	Collision energy	Cell accelerator voltage
229	113	6	8
230	114	6	8
...	...	6	8
...	...	6	8
529	413	6	8
530	414	6	8

In all MRM channels, chromatograms were extracted by Agilent MassHunter Qualitative Analysis B.06.00, using the function “Find Compounds by Targeted MS/MS”, according to the settings listed in Table 5.2.5.

Table 5.2.5: Settings for extraction of chromatograms in Agilent MassHunter Qualitative Analysis B.06.00.

Integrator - Options	
Integrator selection	ChemStation
Tangent skim mode	0
Front skim height ratio	0
Skim valley ratio	20
Baseline correction mode	Advanced
Peak-to-Valley ratio	20
Integrator - Events	
Slope Sensitivity	5
Peak Width	0.1
Area Reject	30
Height Reject	20
Shoulders Mode	OFF
Baseline Now	ON
Processing	
Maximum chromatogram peak width	0.25
Cpd TIC Peak Filters	
Filter on	Peak area
Area Filters: Absolute Area	10
Peak Spectrum	
Spectra to include: Average scans >	10 % of peak height
Peak spectrum background: MS/MS	none
Results	
Previous results	Delete previous results
New results	Highlight first compound
Chromatograms and spectra	Extract MS/MS chromatogram

The integration peak list was exported as CSV file which was imported to Agilent Mass Profiler Professional V13. Changes in peak area in all test samples were identified with respect to the control sample to identify masses that changed upon overexpression of one candidate protein.

5.2.3.4. LC-MS/MS screening for novel Sulfur-dependent DNA base modifications

Hydrolyzed gDNA samples from transfected HEK293T cells were obtained as described in section 5.2.1.4. The concentration of nucleosides in each sample was determined by quantification of cytidine by stable-isotope dilution LC-MS/MS, as described in section 5.2.3.2. Exactly 100 ng of degraded genomic DNA samples from cells treated with heavy or light medium (as described in sections 5.2.2.5, 5.2.2.9 and 5.2.2.12) were injected. Nucleosides were separated on an Agilent 1290 Infinity Binary LC system (Agilent Technologies) with a 5 cm RRHD Eclipse Plus C18 (Agilent Technologies) and detected by a triple-quadrupole mass spectrometer (Agilent 6490, Agilent Technologies). Running solutions were 5 mM ammonium acetate, pH 6.9 (A) and acetonitrile (B). Separations were performed with the following gradient: 3 min 0% B, 4.5 min linear increase to 5% B, 3 min 5% B, 2 min linear increase to 50% B, 2.5 min 0% B. The flow rate was 10.5 min 0.3 ml/min, 2 min 0.38 ml/min, 2 min 0.5 ml min⁻¹ and 0.5 min 0.3 ml/min. The detailed mass spectrometer settings are listed in Table 5.2.4. Mass transitions of -116 Da were monitored between precursor ions of $m/z=229$ and $m/z=530$ in intervals of one Dalton (Table 5.2.4).

The chromatograms were opened in Agilent Mass Hunter Qualitative Analysis B.8.00 and were visually inspected for mass increase of $m/z=2$ or $m/z=4$ after treatment with heavy medium.

5.2.3.5. Qualitative measurement of mcm5S2U RNA modification by LC-MS/MS

1 μg of total degraded RNA from SILAC-treated cells (see sections 5.2.2.5, 5.2.2.9 and 5.2.2.12) was injected and nucleosides were separated with a 15 cm ReproSil 100 C18 column (Jasco) and detected by a triple-quadrupole mass spectrometer (Agilent 6490, Agilent Technologies). Running solutions were 5 mM ammonium acetate, pH 6.9 (A) and acetonitrile (B). Separations were performed with the following gradient: 8 min 0% B, 16 min linear increase to 15% B, 1 min 15% B, 4 min linear increase to 60% B, 6 min 0% B. The flow rate was 24 min 0.5 ml/min, 1 min linear increase to 1 ml/min, 4 min 1 ml/min, 5 min linear decrease to 0.5 ml/min and 1 min 0.5 ml/min. The LC and MS instrument settings and MS monitored mass transitions are listed in Table 5.2.7: LC-MS/MS settings used for quantitative measurement of of m6dATable 5.2.6.

Table 5.2.6: LC-MS/MS settings used for detection of mcm5S2U

Agilent 6490 triple quadrupole mass spectrometer settings

Source parameters	
Ion Mode	positive
Gas temperature	110°C
Gas flow	19 L/min
Nebulizer	25 psi
Sheath gas temperature	375°C
Sheath gas flow	11 L/min
Capillary voltage	3000 V
Nozzle voltage	0 V
iFunnel parameters	
High pressure radiofrequency	70 V
Low pressure radiofrequency	80 V

MRM transitions

Nucleoside	Precursor ion (m/z)	Fragment ion (m/z)	Collision energy	Cell accelerator voltage
mcm5S2U	333	201	6	8
mcm5 ³⁵ S2U	335	203	6	8

5.2.3.6. Quantitative measurement of genomic m⁶dA

Up to 30 µg of degraded DNA was injected and nucleosides were separated on an Agilent 1290 Infinity Binary LC system (Agilent Technologies) with a 15 cm ReproSil 100 C18 column (Jasco) and detected by a triple-quadrupole mass spectrometer (Agilent 6490, Agilent Technologies). Running solutions were 5 mM ammonium acetate, pH 6.9 (A) and acetonitrile (B). Separations were performed with the following gradient: 8 min 0% B, 16 min linear increase to 15% B, 4 min 15% B, 5 min 0% B. The flow rate was 24 min 0.5 ml/min, 4 min 1 ml/min and 1 min 0.5 ml/min. The LC and MS instrument settings and MS monitored mass transitions are listed in Table 5.2.7: LC-MS/MS settings used for quantitative measurement of m⁶dA Table 5.2.7.

Table 5.2.7: LC-MS/MS settings used for quantitative measurement of m⁶dA

Agilent 6490 triple quadrupole mass spectrometer settings

Source parameters	
Ion Mode	positive
Gas temperature	110°C
Gas flow	19 L/min
Nebulizer	25 psi
Sheath gas temperature	375°C
Sheath gas flow	11 L/min
Capillary voltage	3000 V
Nozzle voltage	0 V
iFunnel parameters	
High pressure radiofrequency	70 V
Low pressure radiofrequency	80 V

MRM transitions

Nucleoside	Precursor ion (m/z)	Fragment ion (m/z)	Collision energy	Cell accelerator voltage
dA	252	136	6	8
m ⁶ dA	266	150.1	6	8
¹³ CD ₃ -m ⁶ dA	270	154.1	6	8
¹³ C ₁₀ -dA	262	141	6	8
(¹⁵ N, ¹³ C) ₅ -m ⁶ dA	271	155.1	6	8
(¹⁵ N, ¹³ C) ₈ -m ⁶ dA	274	158.1	6	8

For LC-MS-based quantification, the samples were first mixed with isotopic standards ¹³C₁₀-dA, which contain (¹⁵N,¹³C)₅-m⁶dA and (¹⁵N,¹³C)₈-m⁶dA as impurities. Quantification of highly abundant dA was performed using 100× diluted samples. Data was analyzed with Agilent MassHunter Quantitative Analysis software v.B.05.02 (Agilent technologies) using isotopic standards for normalization and to confirm the peak identity as described previously²⁷⁶.

5.2.3.7. Generation and purification of ¹⁵N-labeled deoxyinosine

¹⁵N-labelled dAMP was dephosphorylated using Fast Alkaline Phosphatase according to the manufacturer's instructions for 1 hour at 37°C. Afterwards, 1 nmol ¹⁵N-dA was deaminated by 0.8 U of recombinant Adenosine Deaminase in 50 mM sodium phosphate buffer for 1 hour at 37°C. Individual nucleosides were separated on an Agilent 1290 Infinity Binary LC system (Agilent technologies) using a ReproSil 100 C18 column (Jasco). Isotopically labeled deoxyinosine was identified by analytical HPLC in tandem with triple quadrupole mass spectrometry (Agilent 6490, Agilent Technologies) and purified by preparative HPLC. An aliquot was mixed with known concentrations in the range of unlabeled deoxyinosine and deoxyadenosine nucleosides.

5.2.3.8. Quantitative measurement of deoxyinosine

Up to 4 µg of degraded DNA was injected and nucleosides were separated on an Agilent 1290 Infinity Binary LC system (Agilent Technologies) with a 15 cm ReproSil 100 C18 column (Jasco) and detected by a triple-quadrupole mass spectrometer (Agilent 6490, Agilent Technologies). Running solutions were 5 mM ammonium acetate, pH 6.9 (A) and acetonitrile (B). Separations were performed with the following gradient: 1 min 0% B, 19 min linear increase to 10% B, 1 min 10% B, 5 min 50% B. The flow rate was 20 min 0.5 ml/min, 1 min gradual increase from 0.5 ml/min to 1 ml/min, 9 min 1 ml/min and 3 min 0.5 ml/min. The MS monitored mass transitions and settings used in this study are listed in Table 5.2.8.

Table 5.2.8: LC-MS/MS settings used for quantitative measurement of dI

Agilent 6490 triple quadrupole mass spectrometer settings

Source parameters	
Ion Mode	positive
Gas temperature	110°C
Gas flow	19 L/min
Nebulizer	25 psi
Sheath gas temperature	375°C
Sheath gas flow	11 L/min
Capillary voltage	2000 V
Nozzle voltage	0 V
iFunnel parameters	
High pressure radiofrequency	60 V
Low pressure radiofrequency	55 V

MRM transitions

Nucleoside	Precursor ion (m/z)	Fragment ion (m/z)	Collision energy	Cell accelerator voltage
dA	252	136	6	8
dI	253	137	6	8
rA	268.2	136.2	5	8
rl	269.1	137	5	8
¹⁵ N ₅ -dA	257	141	6	8
¹⁵ N ₄ -dI	257	141	6	8

For LC-MS-based quantification, the samples were first mixed with the isotopic standard mix containing, ¹⁵N₅-dA and ¹⁵N₄-dI. Quantification of highly abundant dA was performed using 100× diluted samples. Data was analyzed with Agilent MassHunter Quantitative Analysis software v.B.09.00 (Agilent technologies) using isotopic standards to confirm the peak identity. Areas of the integrated peaks were exported into Microsoft Excel with which the areas were normalized to the area of the corresponding isotopic standard. Absolute amounts of the nucleosides were calculated using linear interpolation from a standard curve. Linear interpolation was performed using the two closely

matching data points from the standard curve. Isotopic standards were spiked into the mixture of isotopic standards to normalize for ionization variability. The standard curve for every nucleoside was prepared to cover the amount of the corresponding nucleoside in the DNA sample analyzed.

5.2.3.9. Quantitative measurement of inosine in RNA

1 µg of degraded total RNA was injected and nucleosides were separated on an Agilent 1290 Infinity Binary LC system (Agilent Technologies) with a 15 cm ReproSil 100 C18 column (Jasco) and detected by a triple-quadrupole mass spectrometer (Agilent 6490, Agilent Technologies). Running solutions were 5 mM ammonium acetate, pH 6.9 (A) and acetonitrile (B). Separations were performed with the following gradient: 1 min 0% B, 19 min linear increase to 10% B, 1 min 10% B, 5 min 50% B. The flow rate was 20 min 0.5 ml/min, 1 min gradual increase from 0.5 ml/min to 1 ml/min, 9 min 1 ml/min and 3 min 0.5 ml/min. The MS monitored mass transitions and settings used in this study are listed in Table 5.2.8. Data was analyzed with Agilent MassHunter Quantitative Analysis software v.B.09.00 (Agilent technologies). Areas of the integrated peaks were exported into Microsoft Excel with. Inosine and adenosine concentrations were quantified using linear interpolation from a standard curve. Linear interpolation was performed using the two closely matching data points from the standard curve. The standard curves for inosine and adenosine were prepared to cover the amount of the corresponding nucleoside in the DNA sample analyzed.

6. References

1. HERSHEY, A. D. & CHASE, M. Independent functions of viral protein and nucleic acid in growth of bacteriophage. *The Journal of general physiology* **36**, 39–56; 10.1085/jgp.36.1.39 (1952).
2. Avery, O. T., Macleod, C. M. & McCarty, M. STUDIES ON THE CHEMICAL NATURE OF THE SUBSTANCE INDUCING TRANSFORMATION OF PNEUMOCOCCAL TYPES : INDUCTION OF TRANSFORMATION BY A DESOXYRIBONUCLEIC ACID FRACTION ISOLATED FROM PNEUMOCOCCUS TYPE III. *The Journal of experimental medicine* **79**, 137–158; 10.1084/jem.79.2.137 (1944).
3. Chargaff, E., Lipshitz, R. & Green, C. COMPOSITION OF THE DESOXYRIBONUCLEIC ACIDS OF FOUR GENERA OF SEA-URCHIN. *The Journal of biological chemistry* **195**, 155–160; 10.1016/S0021-9258(19)50884-5 (1952).
4. WATSON, J. D. & CRICK, F. H. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171**, 737–738; 10.1038/171737a0 (1953).
5. Meselson, M. & Stahl, F. W. THE REPLICATION OF DNA IN ESCHERICHIA COLI. *Proceedings of the National Academy of Sciences of the United States of America* **44**, 671–682; 10.1073/pnas.44.7.671 (1958).
6. BRENNER, S., JACOB, F. & Meselson, M. An unstable intermediate carrying information from genes to ribosomes for protein synthesis. *Nature* **190**, 576–581; 10.1038/190576a0 (1961).
7. Carell, T. *et al.* Structure and function of noncanonical nucleobases. *Angewandte Chemie (International ed. in English)* **51**, 7110–7131; 10.1002/anie.201201193 (2012).
8. Berger, S. L., Kouzarides, T., Shiekhattar, R. & Shilatifard, A. An operational definition of epigenetics. *Genes & development* **23**, 781–783; 10.1101/gad.1787609 (2009).
9. Dominissini, D. & Rechavi, G. Epitranscriptome regulation. *Nature structural & molecular biology*; 10.1038/s41594-018-0140-7 (2018).
10. Durrin, L. K., Mann, R. K., Kayne, P. S. & Grunstein, M. Yeast histone H4 N-terminal sequence is required for promoter activation in vivo. *Cell* **65**, 1023–1031; 10.1016/0092-8674(91)90554-C (1991).
11. Brownell, J. E. *et al.* Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. *Cell* **84**, 843–851; 10.1016/s0092-8674(00)81063-6 (1996).
12. Rea, S. *et al.* Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406**, 593–599; 10.1038/35020506 (2000).
13. Grosjean, H. Nucleic Acids Are Not Boring Long Polymers of Only Four Types of Nucleotides: A Guided Tour. *Madame Curie Bioscience Database [Internet]* (2009).
14. Zhou, Y. *et al.* Principles of RNA methylation and their implications for biology and medicine. *Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie* **131**, 110731; 10.1016/j.biopha.2020.110731 (2020).

15. Fu, L. *et al.* Tet-mediated formation of 5-hydroxymethylcytosine in RNA. *J. Am. Chem. Soc.* **136**, 11582–11585; 10.1021/ja505305z. (2014).
16. Huber, S. M. *et al.* Formation and abundance of 5-hydroxymethylcytosine in RNA. *Chembiochem : a European journal of chemical biology* **16**, 752–755; 10.1002/cbic.201500013 (2015).
17. Sharma, S. *et al.* Yeast Kre33 and human NAT10 are conserved 18S rRNA cytosine acetyltransferases that modify tRNAs assisted by the adaptor Tan1/THUMP1. *Nucleic acids research* **43**, 2242–2258; 10.1093/nar/gkv075 (2015).
18. Leidel, S. *et al.* Ubiquitin-related modifier Urm1 acts as a sulphur carrier in thiolation of eukaryotic transfer RNA. *Nature* **458**, 228–232; 10.1038/nature07643 (2009).
19. Noma, A., Sakaguchi, Y. & Suzuki, T. Mechanistic characterization of the sulfur-relay system for eukaryotic 2-thiouridine biogenesis at tRNA wobble positions. *Nucleic acids research* **37**, 1335–1352; 10.1093/nar/gkn1023 (2009).
20. Paul, M. S. & Bass, B. L. Inosine exists in mRNA at tissue-specific levels and is most abundant in brain mRNA. *The EMBO journal* **17**, 1120–1127; 10.1093/emboj/17.4.1120 (1998).
21. Bass, B. L. & Weintraub, H. An unwinding activity that covalently modifies its double-stranded RNA substrate. *Cell* **55**, 1089–1098; 10.1016/0092-8674(88)90253-x (1988).
22. Dalluge, J. J., Hashizume, T., Sopchik, A. E., McCloskey, J. A. & Davis, D. R. Conformational flexibility in RNA: the role of dihydrouridine. *Nucleic acids research* **24**, 1073–1079; 10.1093/nar/24.6.1073 (1996).
23. Ganot, P., Bortolin, M.-L. & Kiss, T. Site-Specific Pseudouridine Formation in Preribosomal RNA Is Guided by Small Nucleolar RNAs. *Cell* **89**, 799–809; 10.1016/S0092-8674(00)80263-9 (1997).
24. 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 research* **2**, 1931–1939; 10.1093/nar/2.10.1931 (1975).
25. Crick, F. The origin of the genetic code. *Journal of molecular biology* **38**, 367–379; 10.1016/0022-2836(68)90392-6 (1968).
26. Altman, S. Enzymatische Spaltung der RNA durch RNA (Nobel-Vortrag). *Angew. Chem.* **102**, 735–744; 10.1002/ange.19901020704 (1990).
27. Cech, T. R. The efficiency and versatility of catalytic RNA: implications for an RNA world. *Gene* **135**, 33–36; 10.1016/0378-1119(93)90046-6 (1993).
28. Mattick, J. S. RNA driving the epigenetic bus. *The EMBO journal* **31**, 515–516; 10.1038/emboj.2011.479 (2012).
29. Statello, L., Guo, C.-J., Chen, L.-L. & Huarte, M. Gene regulation by long non-coding RNAs and its biological functions. *Nature reviews. Molecular cell biology* **22**, 96–118; 10.1038/s41580-020-00315-9 (2021).

30. Brosnan, C. A. & Voinnet, O. The long and the short of noncoding RNAs. *Current opinion in cell biology* **21**, 416–425; 10.1016/j.ceb.2009.04.001 (2009).
31. Schomacher, L. & Niehrs, C. DNA repair and erasure of 5-methylcytosine in vertebrates. *BioEssays : news and reviews in molecular, cellular and developmental biology* **39**; 10.1002/bies.201600218 (2017).
32. Valinluck, V. *et al.* Oxidative damage to methyl-CpG sequences inhibits the binding of the methyl-CpG binding domain (MBD) of methyl-CpG binding protein 2 (MeCP2). *Nucleic acids research* **32**, 4100–4108; 10.1093/nar/gkh739 (2004).
33. Münzel, M., Globisch, D. & Carell, T. 5-Hydroxymethylcytosine, the sixth base of the genome. *Angewandte Chemie (International ed. in English)* **50**, 6460–6468; 10.1002/anie.201101547 (2011).
34. Song, C.-X. & He, C. Potential functional roles of DNA demethylation intermediates. *Trends in biochemical sciences* **38**, 480–484; 10.1016/j.tibs.2013.07.003 (2013).
35. Kellinger, M. W. *et al.* 5-formylcytosine and 5-carboxylcytosine reduce the rate and substrate specificity of RNA polymerase II transcription. *Nature structural & molecular biology* **19**, 831–833; 10.1038/nsmb.2346 (2012).
36. Song, C.-X. *et al.* Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. *Cell* **153**, 678–691; 10.1016/j.cell.2013.04.001 (2013).
37. Uribe-Lewis, S. *et al.* 5-hydroxymethylcytosine and gene activity in mouse intestinal differentiation. *Scientific reports* **10**, 546; 10.1038/s41598-019-57214-z (2020).
38. Xie, Q. *et al.* N6-methyladenine DNA Modification in Glioblastoma. *Cell* **175**, 1228-1243.e20; 10.1016/j.cell.2018.10.006 (2018).
39. Wu, T. P. *et al.* DNA methylation on N(6)-adenine in mammalian embryonic stem cells. *Nature* **532**, 329–333; 10.1038/nature17640 (2016).
40. Schiffers, S. *et al.* Quantitative LC-MS Provides No Evidence for m6 dA or m4 dC in the Genome of Mouse Embryonic Stem Cells and Tissues. *Angewandte Chemie (International ed. in English)* **56**, 11268–11271; 10.1002/anie.201700424 (2017).
41. O'Brown, Z. K. *et al.* Sources of artifact in measurements of 6mA and 4mC abundance in eukaryotic genomic DNA. *BMC genomics* **20**, 445; 10.1186/s12864-019-5754-6 (2019).
42. Lentini, A. *et al.* A reassessment of DNA immunoprecipitation-based genomic profiling (2017).
43. Basu, A. K. DNA Damage, Mutagenesis and Cancer. *IJMS* **19**; 10.3390/ijms19040970 (2018).
44. Chatterjee, N. & Walker, G. C. Mechanisms of DNA damage, repair, and mutagenesis. *Environmental and molecular mutagenesis* **58**, 235–263; 10.1002/em.22087 (2017).
45. Globisch, D. *et al.* Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PloS one* **5**, e15367; 10.1371/journal.pone.0015367 (2010).

46. Breiling, A. & Lyko, F. Epigenetic regulatory functions of DNA modifications: 5-methylcytosine and beyond. *Epigenetics & chromatin* **8**, 24; 10.1186/s13072-015-0016-6 (2015).
47. Ehrlich, M. *et al.* Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells. *Nucleic acids research* **10**, 2709–2721; 10.1093/nar/10.8.2709 (1982).
48. Okano, M., Bell, D. W., Haber, D. A. & Li, E. DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development. *Cell* **99**, 247–257; 10.1016/s0092-8674(00)81656-6 (1999).
49. Li, E., Bestor, T. H. & Jaenisch, R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915–926; 10.1016/0092-8674(92)90611-f (1992).
50. Li, E., Beard, C. & Jaenisch, R. Role for DNA methylation in genomic imprinting. *Nature* **366**, 362–365; 10.1038/366362a0 (1993).
51. Panning, B. & Jaenisch, R. DNA hypomethylation can activate Xist expression and silence X-linked genes. *Genes & development* **10**, 1991–2002; 10.1101/gad.10.16.1991 (1996).
52. Walsh, C. P., Chaillet, J. R. & Bestor, T. H. Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nature genetics* **20**, 116–117; 10.1038/2413 (1998).
53. Cedar, H., Solage, A., Glaser, G. & Razin, A. Direct detection of methylated cytosine in DNA by use of the restriction enzyme MspI. *Nucleic acids research* **6**, 2125–2132; 10.1093/nar/6.6.2125 (1979).
54. Chen, L. *et al.* Direct identification of the active-site nucleophile in a DNA (cytosine-5)-methyltransferase. *Biochemistry* **30**, 11018–11025; 10.1021/bi00110a002 (1991).
55. Hermann, A., Goyal, R. & Jeltsch, A. The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites. *The Journal of biological chemistry* **279**, 48350–48359; 10.1074/jbc.M403427200 (2004).
56. Bostick, M. *et al.* UHRF1 plays a role in maintaining DNA methylation in mammalian cells. *Science (New York, N.Y.)* **317**, 1760–1764; 10.1126/science.1147939 (2007).
57. Ziller, M. J. *et al.* Charting a dynamic DNA methylation landscape of the human genome. *Nature* **500**, 477–481; 10.1038/nature12433 (2013).
58. Bird, A. DNA methylation patterns and epigenetic memory. *Genes & development* **16**, 6–21; 10.1101/gad.947102 (2002).
59. Watt, F. & Molloy, P. L. Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. *Genes & development* **2**, 1136–1143; 10.1101/gad.2.9.1136 (1988).
60. Choy, M.-K. *et al.* Genome-wide conserved consensus transcription factor binding motifs are hypermethylated. *BMC genomics* **11**, 519; 10.1186/1471-2164-11-519 (2010).

61. Jones, P. L. *et al.* Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nature genetics* **19**, 187–191; 10.1038/561 (1998).
62. Nan, X. *et al.* Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* **393**, 386–389; 10.1038/30764 (1998).
63. Ng, H. H. *et al.* MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex. *Nature genetics* **23**, 58–61; 10.1038/12659 (1999).
64. Yang, X. *et al.* Gene body methylation can alter gene expression and is a therapeutic target in cancer. *Cancer cell* **26**, 577–590; 10.1016/j.ccr.2014.07.028 (2014).
65. Neri, F. *et al.* Intragenic DNA methylation prevents spurious transcription initiation. *Nature* **543**, 72–77; 10.1038/nature21373 (2017).
66. Janssen, A., Colmenares, S. U. & Karpen, G. H. Heterochromatin: Guardian of the Genome. *Annual review of cell and developmental biology* **34**, 265–288; 10.1146/annurev-cellbio-100617-062653 (2018).
67. Pappalardo, X. G. & Barra, V. Losing DNA methylation at repetitive elements and breaking bad. *Epigenetics & chromatin* **14**, 25; 10.1186/s13072-021-00400-z (2021).
68. Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315–322; 10.1038/nature08514 (2009).
69. Laurent, L. *et al.* Dynamic changes in the human methylome during differentiation. *Genome research* **20**, 320–331; 10.1101/gr.101907.109 (2010).
70. Gowher, H. & Jeltsch, A. Enzymatic properties of recombinant Dnmt3a DNA methyltransferase from mouse: the enzyme modifies DNA in a non-processive manner and also methylates non-CpG correction of non-CpA sites. *Journal of molecular biology* **309**, 1201–1208; 10.1006/jmbi.2001.4710 (2001).
71. Rougier, N. *et al.* Chromosome methylation patterns during mammalian preimplantation development. *Genes & development* **12**, 2108–2113; 10.1101/gad.12.14.2108 (1998).
72. Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science (New York, N.Y.)* **324**, 930–935; 10.1126/science.1170116 (2009).
73. Dalton, S. R. & Bellacosa, A. DNA demethylation by TDG. *Epigenomics* **4**, 459–467; 10.2217/epi.12.36 (2012).
74. Shen, L. *et al.* Genome-wide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics. *Cell* **153**, 692–706; 10.1016/j.cell.2013.04.002 (2013).
75. Kohli, R. M. & Zhang, Y. TET enzymes, TDG and the dynamics of DNA demethylation. *Nature* **502**, 472–479; 10.1038/nature12750 (2013).

76. Mellén, M., Ayata, P. & Heintz, N. 5-hydroxymethylcytosine accumulation in postmitotic neurons results in functional demethylation of expressed genes. *Proceedings of the National Academy of Sciences of the United States of America* **114**, E7812-E7821; 10.1073/pnas.1708044114 (2017).
77. Kinde, B., Gabel, H. W., Gilbert, C. S., Griffith, E. C. & Greenberg, M. E. Reading the unique DNA methylation landscape of the brain: Non-CpG methylation, hydroxymethylation, and MeCP2. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 6800–6806; 10.1073/pnas.1411269112 (2015).
78. Ibrahim, A. *et al.* MeCP2 is a microsatellite binding protein that protects CA repeats from nucleosome invasion. *Science (New York, N.Y.)* **372**; 10.1126/science.abd5581 (2021).
79. Spruijt, C. G. *et al.* Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* **152**, 1146–1159; 10.1016/j.cell.2013.02.004 (2013).
80. Iurlaro, M. *et al.* A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation. *Genome biology* **14**, R119; 10.1186/gb-2013-14-10-r119. (2013).
81. Wang, L. *et al.* Molecular basis for 5-carboxycytosine recognition by RNA polymerase II elongation complex. *Nature* **523**, 621–625; 10.1038/nature14482 (2015).
82. Szulik, M. W. *et al.* Differential stabilities and sequence-dependent base pair opening dynamics of Watson-Crick base pairs with 5-hydroxymethylcytosine, 5-formylcytosine, or 5-carboxylcytosine. *Biochemistry* **54**, 1294–1305; 10.1021/bi501534x (2015).
83. Marinus, M. G. & Løbner-Olesen, A. DNA Methylation. *EcoSal Plus* **3**; 10.1128/ecosalplus.4.4.5 (2009).
84. Vasu, K. & Nagaraja, V. Diverse functions of restriction-modification systems in addition to cellular defense. *Microbiology and molecular biology reviews : MMBR* **77**, 53–72; 10.1128/MMBR.00044-12 (2013).
85. Greer, E. L. *et al.* DNA Methylation on N6-Adenine in *C. elegans*. *Cell* **161**, 868–878; 10.1016/j.cell.2015.04.005 (2015).
86. Zhang, G. *et al.* N6-methyladenine DNA modification in *Drosophila*. *Cell* **161**, 893–906; 10.1016/j.cell.2015.04.018 (2015).
87. Liu, J. *et al.* Abundant DNA 6mA methylation during early embryogenesis of zebrafish and pig. *Nature communications* **7**, 13052; 10.1038/ncomms13052 (2016).
88. Xiao, C.-L. *et al.* N6-Methyladenine DNA Modification in the Human Genome. *Molecular cell* **71**, 306–318.e7; 10.1016/j.molcel.2018.06.015 (2018).
89. Li, X. *et al.* The DNA modification N6-methyl-2'-deoxyadenosine (m6dA) drives activity-induced gene expression and is required for fear extinction. *Nature neuroscience* **22**, 534–544; 10.1038/s41593-019-0339-x (2019).

90. Zhang, M. *et al.* Mammalian ALKBH1 serves as an N6-mA demethylase of unpairing DNA. *Cell research* **30**, 197–210; 10.1038/s41422-019-0237-5 (2020).
91. Zhou, C., Liu, Y., Li, X., Zou, J. & Zou, S. DNA N6-methyladenine demethylase ALKBH1 enhances osteogenic differentiation of human MSCs. *Bone research* **4**, 16033; 10.1038/boneres.2016.33 (2016).
92. Abakir, A. *et al.* N6-methyladenosine regulates the stability of RNA:DNA hybrids in human cells. *Nature genetics* **52**, 48–55; 10.1038/s41588-019-0549-x (2020).
93. Meyer, K. D. *et al.* Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell* **149**, 1635–1646; 10.1016/j.cell.2012.05.003 (2012).
94. Dominissini, D. *et al.* Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* **485**, 201–206; 10.1038/nature11112 (2012).
95. Liu, J. *et al.* A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. *Nature chemical biology* **10**, 93–95; 10.1038/nchembio.1432 (2014).
96. Ping, X.-L. *et al.* Mammalian WTAP is a regulatory subunit of the RNA N6-methyladenosine methyltransferase. *Cell research* **24**, 177–189; 10.1038/cr.2014.3 (2014).
97. Schwartz, S. *et al.* Perturbation of m6A writers reveals two distinct classes of mRNA methylation at internal and 5' sites. *Cell reports* **8**, 284–296; 10.1016/j.celrep.2014.05.048 (2014).
98. Wang, Y. *et al.* N6-methyladenosine modification destabilizes developmental regulators in embryonic stem cells. *Nature cell biology* **16**, 191–198; 10.1038/ncb2902 (2014).
99. Wang, X. *et al.* N(6)-methyladenosine Modulates Messenger RNA Translation Efficiency. *Cell* **161**, 1388–1399; 10.1016/j.cell.2015.05.014 (2015).
100. Xu, C. *et al.* Structural basis for selective binding of m6A RNA by the YTHDC1 YTH domain. *Nature chemical biology* **10**, 927–929; 10.1038/nchembio.1654 (2014).
101. Xiao, W. *et al.* Nuclear m(6)A Reader YTHDC1 Regulates mRNA Splicing. *Molecular cell* **61**, 507–519; 10.1016/j.molcel.2016.01.012 (2016).
102. Alarcón, C. R. *et al.* HNRNPA2B1 Is a Mediator of m(6)A-Dependent Nuclear RNA Processing Events. *Cell* **162**, 1299–1308; 10.1016/j.cell.2015.08.011 (2015).
103. Meyer, K. D. *et al.* 5' UTR m(6)A Promotes Cap-Independent Translation. *Cell* **163**, 999–1010; 10.1016/j.cell.2015.10.012 (2015).
104. Jia, G. *et al.* N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nature chemical biology* **7**, 885–887; 10.1038/nchembio.687 (2011).
105. Zheng, G. *et al.* ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Molecular cell* **49**, 18–29; 10.1016/j.molcel.2012.10.015 (2013).

106. Wang, X. *et al.* N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* **505**, 117–120; 10.1038/nature12730 (2014).
107. Engel, J. D. & Hippel, P. H. von. Effects of methylation on the stability of nucleic acid conformations: studies at the monomer level. *Biochemistry* **13**, 4143–4158; 10.1021/bi00717a013 (1974).
108. Engel, J. D. & Hippel, P. H. von. Effects of methylation on the stability of nucleic acid conformations. Studies at the polymer level. *The Journal of biological chemistry* **253**, 927–934 (1978).
109. Spitale, R. C. *et al.* Structural imprints in vivo decode RNA regulatory mechanisms. *Nature* **519**, 486–490; 10.1038/nature14263 (2015).
110. Zhou, J. *et al.* Dynamic m(6)A mRNA methylation directs translational control of heat shock response. *Nature* **526**, 591–594; 10.1038/nature15377 (2015).
111. Nishikura, K. A-to-I editing of coding and non-coding RNAs by ADARs. *Nature reviews. Molecular cell biology* **17**, 83–96; 10.1038/nrm.2015.4 (2016).
112. Wagner, R. W., Smith, J. E., Cooperman, B. S. & Nishikura, K. A double-stranded RNA unwinding activity introduces structural alterations by means of adenosine to inosine conversions in mammalian cells and *Xenopus* eggs. *Proceedings of the National Academy of Sciences of the United States of America* **86**, 2647–2651; 10.1073/pnas.86.8.2647 (1989).
113. Wang, Q., Zhang, Z., Blackwell, K. & Carmichael, G. G. Vigilins bind to promiscuously A-to-I-edited RNAs and are involved in the formation of heterochromatin. *Current biology : CB* **15**, 384–391; 10.1016/j.cub.2005.01.046. (2005).
114. Rosenthal, J. J. C. The emerging role of RNA editing in plasticity. *The Journal of experimental biology* **218**, 1812–1821; 10.1242/jeb.119065 (2015).
115. Balik, A., Penn, A. C., Nemoda, Z. & Greger, I. H. Activity-regulated RNA editing in select neuronal subfields in hippocampus. *Nucleic acids research* **41**, 1124–1134; 10.1093/nar/gks1045 (2013).
116. Czermak, P., Amman, F., Jantsch, M. F. & Cimatti, L. Organ-wide profiling in mouse reveals high editing levels of Filamin B mRNA in the musculoskeletal system. *RNA biology* **15**, 877–885; 10.1080/15476286.2018.1480252 (2018).
117. Penn, A. C., Balik, A. & Greger, I. H. Reciprocal regulation of A-to-I RNA editing and the vertebrate nervous system. *Frontiers in neuroscience* **7**, 61; 10.3389/fnins.2013.00061 (2013).
118. Stulić, M. & Jantsch, M. F. Spatio-temporal profiling of Filamin A RNA-editing reveals ADAR preferences and high editing levels outside neuronal tissues. *RNA biology* **10**, 1611–1617; 10.4161/rna.26216 (2013).
119. Kapoor, U. *et al.* ADAR-deficiency perturbs the global splicing landscape in mouse tissues. *Genome research* **30**, 1107–1118; 10.1101/gr.256933.119 (2020).

120. Solomon, O. *et al.* Global regulation of alternative splicing by adenosine deaminase acting on RNA (ADAR). *RNA (New York, N.Y.)* **19**, 591–604; 10.1261/rna.038042.112 (2013).
121. Rueter, S. M., Dawson, T. R. & Emeson, R. B. Regulation of alternative splicing by RNA editing. *Nature* **399**, 75–80; 10.1038/19992 (1999).
122. Penn, A. C., Balik, A. & Greger, I. H. Steric antisense inhibition of AMPA receptor Q/R editing reveals tight coupling to intronic editing sites and splicing. *Nucleic acids research* **41**, 1113–1123; 10.1093/nar/gks1044 (2013).
123. Schoft, V. K., Schopoff, S. & Jantsch, M. F. Regulation of glutamate receptor B pre-mRNA splicing by RNA editing. *Nucleic acids research* **35**, 3723–3732; 10.1093/nar/gkm314 (2007).
124. Slavov, D., Clark, M. & Gardiner, K. Comparative analysis of the RED1 and RED2 A-to-I RNA editing genes from mammals, pufferfish and zebrafish. *Gene* **250**, 41–51; 10.1016/s0378-1119(00)00174-8 (2000).
125. Slavov, D., Crnogorac-Jurčević, T., Clark, M. & Gardiner, K. Comparative analysis of the DRADA A-to-I RNA editing gene from mammals, pufferfish and zebrafish. *Gene* **250**, 53–60; 10.1016/s0378-1119(00)00175-x (2000).
126. Patterson, J. B. & Samuel, C. E. Expression and regulation by interferon of a double-stranded-RNA-specific adenosine deaminase from human cells: evidence for two forms of the deaminase. *Molecular and cellular biology* **15**, 5376–5388; 10.1128/mcb.15.10.5376 (1995).
127. Kawakubo, K. & Samuel, C. E. Human RNA-specific adenosine deaminase (ADAR1) gene specifies transcripts that initiate from a constitutively active alternative promoter. *Gene* **258**, 165–172; 10.1016/s0378-1119(00)00368-1 (2000).
128. Herbert, A. *et al.* A Z-DNA binding domain present in the human editing enzyme, double-stranded RNA adenosine deaminase. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 8421–8426; 10.1073/pnas.94.16.8421 (1997).
129. Fritz, J. *et al.* RNA-regulated interaction of transportin-1 and exportin-5 with the double-stranded RNA-binding domain regulates nucleocytoplasmic shuttling of ADAR1. *Molecular and cellular biology* **29**, 1487–1497; 10.1128/MCB.01519-08 (2009).
130. Vesely, C. & Jantsch, M. F. An I for an A: Dynamic Regulation of Adenosine Deamination-Mediated RNA Editing. *Genes* **12**; 10.3390/genes12071026 (2021).
131. Cho, D.-S. C. *et al.* Requirement of dimerization for RNA editing activity of adenosine deaminases acting on RNA. *The Journal of biological chemistry* **278**, 17093–17102; 10.1074/jbc.M213127200 (2003).
132. Chilibeck, K. A. *et al.* FRET analysis of in vivo dimerization by RNA-editing enzymes. *The Journal of biological chemistry* **281**, 16530–16535; 10.1074/jbc.M511831200 (2006).

133. Matthews, M. M. *et al.* Structures of human ADAR2 bound to dsRNA reveal base-flipping mechanism and basis for site selectivity. *Nature structural & molecular biology* **23**, 426–433; 10.1038/nsmb.3203 (2016).
134. Blow, M., Futreal, P. A., Wooster, R. & Stratton, M. R. A survey of RNA editing in human brain. *Genome research* **14**, 2379–2387; 10.1101/gr.2951204 (2004).
135. Peng, Z. *et al.* Comprehensive analysis of RNA-Seq data reveals extensive RNA editing in a human transcriptome. *Nature biotechnology* **30**, 253–260; 10.1038/nbt.2122 (2012).
136. Neeman, Y., Levanon, E. Y., Jantsch, M. F. & Eisenberg, E. RNA editing level in the mouse is determined by the genomic repeat repertoire. *RNA (New York, N.Y.)* **12**, 1802–1809; 10.1261/rna.165106 (2006).
137. Tan, M. H. *et al.* Dynamic landscape and regulation of RNA editing in mammals. *Nature* **550**, 249–254; 10.1038/nature24041 (2017).
138. Kim, D. D. Y. *et al.* Widespread RNA editing of embedded alu elements in the human transcriptome. *Genome research* **14**, 1719–1725; 10.1101/gr.2855504 (2004).
139. Eisenberg, E. & Levanon, E. Y. A-to-I RNA editing - immune protector and transcriptome diversifier. *Nature reviews. Genetics* **19**, 473–490; 10.1038/s41576-018-0006-1 (2018).
140. Nishikura, K. Functions and regulation of RNA editing by ADAR deaminases. *Annual review of biochemistry* **79**, 321–349; 10.1146/annurev-biochem-060208-105251 (2010).
141. Mannion, N., Arieti, F., Gallo, A., Keegan, L. P. & O'Connell, M. A. New Insights into the Biological Role of Mammalian ADARs; the RNA Editing Proteins. *Biomolecules* **5**, 2338–2362; 10.3390/biom5042338 (2015).
142. Liddicoat, B. J. *et al.* RNA editing by ADAR1 prevents MDA5 sensing of endogenous dsRNA as nonself. *Science (New York, N.Y.)* **349**, 1115–1120; 10.1126/science.aac7049 (2015).
143. Mannion, N. M. *et al.* The RNA-editing enzyme ADAR1 controls innate immune responses to RNA. *Cell reports* **9**, 1482–1494; 10.1016/j.celrep.2014.10.041 (2014).
144. Li, Y. *et al.* Ribonuclease L mediates the cell-lethal phenotype of double-stranded RNA editing enzyme ADAR1 deficiency in a human cell line. *eLife* **6**; 10.7554/eLife.25687 (2017).
145. Jha, B. K. *et al.* Inhibition of RNase L and RNA-dependent protein kinase (PKR) by sunitinib impairs antiviral innate immunity. *The Journal of biological chemistry* **286**, 26319–26326; 10.1074/jbc.M111.253443 (2011).
146. Brusa, R. *et al.* Early-onset epilepsy and postnatal lethality associated with an editing-deficient GluR-B allele in mice. *Science (New York, N.Y.)* **270**, 1677–1680; 10.1126/science.270.5242.1677 (1995).
147. Higuchi, M. *et al.* Point mutation in an AMPA receptor gene rescues lethality in mice deficient in the RNA-editing enzyme ADAR2. *Nature* **406**, 78–81; 10.1038/35017558 (2000).

148. Yasui, M. *et al.* Miscoding properties of 2'-deoxyinosine, a nitric oxide-derived DNA Adduct, during translesion synthesis catalyzed by human DNA polymerases. *Journal of molecular biology* **377**, 1015–1023; 10.1016/j.jmb.2008.01.033 (2008).
149. Martin, F. H., Castro, M. M., Aboul-ela, F. & Tinoco, I. Base pairing involving deoxyinosine: implications for probe design. *Nucleic acids research* **13**, 8927–8938; 10.1093/nar/13.24.8927 (1985).
150. Karran, P. & Lindahl, T. Hypoxanthine in deoxyribonucleic acid: generation by heat-induced hydrolysis of adenine residues and release in free form by a deoxyribonucleic acid glycosylase from calf thymus. *Biochemistry* **19**, 6005–6011; 10.1021/bi00567a010 (1980).
151. Nguyen, T. *et al.* DNA damage and mutation in human cells exposed to nitric oxide in vitro. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 3030–3034; 10.1073/pnas.89.7.3030 (1992).
152. Levy, M. & Miller, S. L. The stability of the RNA bases: implications for the origin of life. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 7933–7938; 10.1073/pnas.95.14.7933 (1998).
153. Wang, S. & Hu, A. Comparative study of spontaneous deamination of adenine and cytosine in unbuffered aqueous solution at room temperature. *Chemical Physics Letters* **653**, 207–211; 10.1016/j.cplett.2016.05.001 (2016).
154. Pang, B. *et al.* Lipid peroxidation dominates the chemistry of DNA adduct formation in a mouse model of inflammation. *Carcinogenesis* **28**, 1807–1813; 10.1093/carcin/bgm037 (2007).
155. Taghizadeh, K. *et al.* Quantification of DNA damage products resulting from deamination, oxidation and reaction with products of lipid peroxidation by liquid chromatography isotope dilution tandem mass spectrometry. *Nature protocols* **3**, 1287–1298; 10.1038/nprot.2008.119 (2008).
156. Sakumi, K. *et al.* ITPA protein, an enzyme that eliminates deaminated purine nucleoside triphosphates in cells. *Mutation research* **703**, 43–50; 10.1016/j.mrgentox.2010.06.009 (2010).
157. Dierick, H., Stul, M., Kever, W. de, Marynen, P. & Cassiman, J. J. Incorporation of dITP or 7-deaza dGTP during PCR improves sequencing of the product. *Nucleic acids research* **21**, 4427–4428; 10.1093/nar/21.18.4427 (1993).
158. Myrnes, B., Guddal, P. H. & Krokan, H. Metabolism of dITP in HeLa cell extracts, incorporation into DNA by isolated nuclei and release of hypoxanthine from DNA by a hypoxanthine-DNA glycosylase activity. *Nucleic acids research* **10**, 3693–3701; 10.1093/nar/10.12.3693 (1982).
159. Karran, P. & Lindahl, T. Enzymatic excision of free hypoxanthine from polydeoxynucleotides and DNA containing deoxyinosine monophosphate residues. *The Journal of biological chemistry* **253**, 5877–5879; 10.1016/S0021-9258(17)34545-3 (1978).

160. Dianov, G. & Lindahl, T. Preferential recognition of I.T base-pairs in the initiation of excision-repair by hypoxanthine-DNA glycosylase. *Nucleic acids research* **19**, 3829–3833; 10.1093/nar/19.14.3829 (1991).
161. Miao, F., Bouziane, M. & O'Connor, T. R. Interaction of the recombinant human methylpurine-DNA glycosylase (MPG protein) with oligodeoxyribonucleotides containing either hypoxanthine or abasic sites. *Nucleic acids research* **26**, 4034–4041; 10.1093/nar/26.17.4034 (1998).
162. Sapparbaev, M. & Laval, J. Excision of hypoxanthine from DNA containing dIMP residues by the *Escherichia coli*, yeast, rat, and human alkylpurine DNA glycosylases. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 5873–5877; 10.1073/pnas.91.13.5873 (1994).
163. Sapparbaev, M., Mani, J. C. & Laval, J. Interactions of the human, rat, *Saccharomyces cerevisiae* and *Escherichia coli* 3-methyladenine-DNA glycosylases with DNA containing dIMP residues. *Nucleic acids research* **28**, 1332–1339; 10.1093/nar/28.6.1332 (2000).
164. Kow, Y. W. Repair of deaminated bases in DNA 1 2 1Guest Editor: Miral Dizdaroglu 2This article is part of a series of reviews on “Oxidative DNA Damage and Repair.” The full list of papers may be found on the homepage of the journal. *Free Radical Biology and Medicine* **33**, 886–893; 10.1016/S0891-5849(02)00902-4 (2002).
165. Yao, M., Hatahet, Z., Melamed, R. J. & Kow, Y. W. Purification and characterization of a novel deoxyinosine-specific enzyme, deoxyinosine 3' endonuclease, from *Escherichia coli*. *The Journal of biological chemistry* **269**, 16260–16268 (1994).
166. Lieb, M. & Bhagwat, A. S. Very short patch repair: reducing the cost of cytosine methylation. *Molecular microbiology* **20**, 467–473; 10.1046/j.1365-2958.1996.5291066.x (1996).
167. Mi, R., Alford-Zappala, M., Kow, Y. W., Cunningham, R. P. & Cao, W. Human endonuclease V as a repair enzyme for DNA deamination. *Mutation research* **735**, 12–18; 10.1016/j.mrfmmm.2012.05.003 (2012).
168. Kuraoka, I. Diversity of Endonuclease V: From DNA Repair to RNA Editing. *Biomolecules* **5**, 2194–2206; 10.3390/biom5042194 (2015).
169. SPENCER, M., FULLER, W., WILKINS, M. H. & BROWN, G. L. Determination of the helical configuration of ribonucleic acid molecules by X-ray diffraction study of crystalline amino-acid-transfer ribonucleic acid. *Nature* **194**, 1014–1020; 10.1038/1941014a0 (1962).
170. Brahms, J. & Mommaerts, W. A study of conformation of nucleic acids in solution by means of circular dichroism. *Journal of molecular biology* **10**, 73–88; 10.1016/S0022-2836(64)80029-2 (1964).
171. Egli, M., Usman, N. & Rich, A. Conformational influence of the ribose 2'-hydroxyl group: Crystal structures of DNA-RNA chimeric duplexes. *Biochemistry* **32**, 3221–3237; 10.1021/bi00064a004 (1993).
172. Egli, M., Usman, N., Zhang, S. G. & Rich, A. Crystal structure of an Okazaki fragment at 2-Å resolution. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 534–538; 10.1073/pnas.89.2.534 (1992).

173. Thomas, M., White, R. L. & Davis, R. W. Hybridization of RNA to double-stranded DNA: formation of R-loops. *Proceedings of the National Academy of Sciences of the United States of America* **73**, 2294–2298; 10.1073/pnas.73.7.2294 (1976).
174. Sollier, J. *et al.* Transcription-coupled nucleotide excision repair factors promote R-loop-induced genome instability. *Molecular cell* **56**, 777–785; 10.1016/j.molcel.2014.10.020 (2014).
175. Cristini, A. *et al.* Dual Processing of R-Loops and Topoisomerase I Induces Transcription-Dependent DNA Double-Strand Breaks. *Cell reports* **28**, 3167–3181.e6; 10.1016/j.celrep.2019.08.041 (2019).
176. Sollier, J. & Cimprich, K. A. Breaking bad. R-loops and genome integrity. *Trends in cell biology* **25**, 514–522; 10.1016/j.tcb.2015.05.003 (2015).
177. Richard, P. & Manley, J. L. R Loops and Links to Human Disease. *Journal of molecular biology* **429**, 3168–3180; 10.1016/j.jmb.2016.08.031 (2017).
178. Niehrs, C. & Luke, B. Regulatory R-loops as facilitators of gene expression and genome stability. *Nature reviews. Molecular cell biology* **21**, 167–178; 10.1038/s41580-019-0206-3 (2020).
179. Wahba, L., Amon, J. D., Koshland, D. & Vuica-Ross, M. RNase H and multiple RNA biogenesis factors cooperate to prevent RNA:DNA hybrids from generating genome instability. *Molecular cell* **44**, 978–988; 10.1016/j.molcel.2011.10.017 (2011).
180. Skourti-Stathaki, K., Proudfoot, N. J. & Gromak, N. Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Molecular cell* **42**, 794–805; 10.1016/j.molcel.2011.04.026 (2011).
181. Cargill, M., Venkataraman, R. & Lee, S. DEAD-Box RNA Helicases and Genome Stability. *Genes* **12**, 1471; 10.3390/genes12101471 (2021).
182. Yu, Z. *et al.* DDX5 resolves R-loops at DNA double-strand breaks to promote DNA repair and avoid chromosomal deletions. *NAR cancer* **2**, zcaa028; 10.1093/narcan/zcaa028 (2020).
183. Pérez-Calero, C. *et al.* UAP56/DDX39B is a major cotranscriptional RNA-DNA helicase that unwinds harmful R loops genome-wide. *Genes & development* **34**, 898–912; 10.1101/gad.336024.119 (2020).
184. Song, C., Hotz-Wagenblatt, A., Voit, R. & Grummt, I. SIRT7 and the DEAD-box helicase DDX21 cooperate to resolve genomic R loops and safeguard genome stability. *Genes & development* **31**, 1370–1381; 10.1101/gad.300624.117 (2017).
185. Mikolaskova, B. *et al.* Maintenance of genome stability: the unifying role of interconnections between the DNA damage response and RNA-processing pathways. *Current genetics* **64**, 971–983; 10.1007/s00294-018-0819-7 (2018).

186. Wahba, L., Gore, S. K. & Koshland, D. The homologous recombination machinery modulates the formation of RNA-DNA hybrids and associated chromosome instability. *eLife* **2**, e00505; 10.7554/eLife.00505 (2013).
187. Ginno, P. A., Lott, P. L., Christensen, H. C., Korf, I. & Chédin, F. R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Molecular cell* **45**, 814–825; 10.1016/j.molcel.2012.01.017 (2012).
188. Grunseich, C. *et al.* Senataxin Mutation Reveals How R-Loops Promote Transcription by Blocking DNA Methylation at Gene Promoters. *Molecular cell* **69**, 426–437.e7; 10.1016/j.molcel.2017.12.030 (2018).
189. Arab, K. *et al.* GADD45A binds R-loops and recruits TET1 to CpG island promoters. *Nature genetics* **51**, 217–223; 10.1038/s41588-018-0306-6 (2019).
190. Boque-Sastre, R. *et al.* Head-to-head antisense transcription and R-loop formation promotes transcriptional activation. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 5785–5790; 10.1073/pnas.1421197112 (2015).
191. Chen, P. B., Chen, H. V., Acharya, D., Rando, O. J. & Fazio, T. G. R loops regulate promoter-proximal chromatin architecture and cellular differentiation. *Nature structural & molecular biology* **22**, 999–1007; 10.1038/nsmb.3122 (2015).
192. Skourti-Stathaki, K. *et al.* R-Loops Enhance Polycomb Repression at a Subset of Developmental Regulator Genes. *Molecular cell* **73**, 930–945.e4; 10.1016/j.molcel.2018.12.016 (2019).
193. Beckedorff, F. C. *et al.* The intronic long noncoding RNA ANRASSF1 recruits PRC2 to the RASSF1A promoter, reducing the expression of RASSF1A and increasing cell proliferation. *PLoS genetics* **9**, e1003705; 10.1371/journal.pgen.1003705 (2013).
194. Kireeva, M. L., Komissarova, N. & Kashlev, M. Overextended RNA:DNA hybrid as a negative regulator of RNA polymerase II processivity. *Journal of molecular biology* **299**, 325–335; 10.1006/jmbi.2000.3755 (2000).
195. Skourti-Stathaki, K., Kamieniarz-Gdula, K. & Proudfoot, N. J. R-loops induce repressive chromatin marks over mammalian gene terminators. *Nature* **516**, 436–439; 10.1038/nature13787 (2014).
196. Gan, W. *et al.* R-loop-mediated genomic instability is caused by impairment of replication fork progression. *Genes & development* **25**, 2041–2056; 10.1101/gad.17010011 (2011).
197. Li, L. *et al.* DEAD Box 1 Facilitates Removal of RNA and Homologous Recombination at DNA Double-Strand Breaks. *Molecular and cellular biology* **36**, 2794–2810; 10.1128/MCB.00415-16 (2016).
198. Cohen, S. *et al.* Senataxin resolves RNA:DNA hybrids forming at DNA double-strand breaks to prevent translocations. *Nature communications* **9**, 533; 10.1038/s41467-018-02894-w (2018).

199. D'Alessandro, G. *et al.* BRCA2 controls DNA:RNA hybrid level at DSBs by mediating RNase H2 recruitment. *Nature communications* **9**, 5376; 10.1038/s41467-018-07799-2 (2018).
200. Britton, S. *et al.* DNA damage triggers SAF-A and RNA biogenesis factors exclusion from chromatin coupled to R-loops removal. *Nucleic acids research* **42**, 9047–9062; 10.1093/nar/gku601 (2014).
201. Chédin, F., Hartono, S. R., Sanz, L. A. & Vanoosthuysse, V. Best practices for the visualization, mapping, and manipulation of R-loops. *The EMBO journal*, e106394; 10.15252/embj.2020106394 (2021).
202. Hartono, S. R. *et al.* The Affinity of the S9.6 Antibody for Double-Stranded RNAs Impacts the Accurate Mapping of R-Loops in Fission Yeast. *Journal of molecular biology* **430**, 272–284; 10.1016/j.jmb.2017.12.016 (2018).
203. Silva, S., Camino, L. P. & Aguilera, A. Human mitochondrial degradosome prevents harmful mitochondrial R loops and mitochondrial genome instability. *Proceedings of the National Academy of Sciences of the United States of America* **115**, 11024–11029; 10.1073/pnas.1807258115 (2018).
204. Stein, H. & Hausen, P. Enzyme from calf thymus degrading the RNA moiety of DNA-RNA Hybrids: effect on DNA-dependent RNA polymerase. *Science (New York, N.Y.)* **166**, 393–395; 10.1126/science.166.3903.393 (1969).
205. Chen, L.-Q., Zhao, W.-S. & Luo, G.-Z. Mapping and editing of nucleic acid modifications. *Computational and structural biotechnology journal* **18**, 661–667; 10.1016/j.csbj.2020.03.010 (2020).
206. Villalta, P. & Balbo, S. The Future of DNA Adductomic Analysis. *IJMS* **18**, 1870; 10.3390/ijms18091870 (2017).
207. Brink, A., Lutz, U., Völkel, W. & Lutz, W. K. Simultaneous determination of O6-methyl-2'-deoxyguanosine, 8-oxo-7,8-dihydro-2'-deoxyguanosine, and 1,N6-etheno-2'-deoxyadenosine in DNA using on-line sample preparation by HPLC column switching coupled to ESI-MS/MS. *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences* **830**, 255–261; 10.1016/j.jchromb.2005.10.046 (2006).
208. Paul, W. & Steinwedel, H. Notizen: Ein neues Massenspektrometer ohne Magnetfeld. *Zeitschrift für Naturforschung A* **8**, 448–450; 10.1515/zna-1953-0710 (1953).
209. Wolf, S. M. & Vouros, P. Application of capillary liquid chromatography coupled with tandem mass spectrometric methods to the rapid screening of adducts formed by the reaction of N-acetoxy-N-acetyl-2-aminofluorene with calf thymus DNA. *Chem. Res. Toxicol.* **7**, 82–88; 10.1021/tx00037a013 (1994).
210. Balbo, S., Turesky, R. J. & Villalta, P. W. DNA adductomics. *Chem. Res. Toxicol.* **27**, 356–366; 10.1021/tx4004352 (2014).
211. Tretyakova, N., Villalta, P. W. & Kotapati, S. Mass spectrometry of structurally modified DNA. *Chemical reviews* **113**, 2395–2436; 10.1021/cr300391r (2013).

212. Annesley, T. M. Ion suppression in mass spectrometry. *Clinical chemistry* **49**, 1041–1044; 10.1373/49.7.1041 (2003).
213. Chan, C. T. Y. *et al.* A quantitative systems approach reveals dynamic control of tRNA modifications during cellular stress. *PLoS genetics* **6**, e1001247; 10.1371/journal.pgen.1001247 (2010).
214. Brückl, T., Globisch, D., Wagner, M., Müller, M. & Carell, T. Parallele isotopenbasierte Quantifizierung modifizierter tRNA-Nucleoside. *Angew. Chem.* **121**, 8074–8077; 10.1002/ange.200902740 (2009).
215. Krueve, A. & Kaupmees, K. Adduct Formation in ESI/MS by Mobile Phase Additives. *Journal of the American Society for Mass Spectrometry* **28**, 887–894; 10.1007/s13361-017-1626-y (2017).
216. Keller, B. O., Sui, J., Young, A. B. & Whittall, R. M. Interferences and contaminants encountered in modern mass spectrometry. *Analytica chimica acta* **627**, 71–81; 10.1016/j.aca.2008.04.043 (2008).
217. Kasianowicz, J. J., Brandin, E., Branton, D. & Deamer, D. W. Characterization of individual polynucleotide molecules using a membrane channel. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 13770–13773; 10.1073/pnas.93.24.13770 (1996).
218. Branton, D. *et al.* The potential and challenges of nanopore sequencing. *Nature biotechnology* **26**, 1146–1153; 10.1038/nbt.1495 (2008).
219. Manrao, E. A., Derrington, I. M., Pavlenok, M., Niederweis, M. & Gundlach, J. H. Nucleotide discrimination with DNA immobilized in the MspA nanopore. *PLoS one* **6**, e25723; 10.1371/journal.pone.0025723 (2011).
220. Laszlo, A. H. *et al.* Detection and mapping of 5-methylcytosine and 5-hydroxymethylcytosine with nanopore MspA. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 18904–18909; 10.1073/pnas.1310240110 (2013).
221. Leger, A. *et al.* RNA modifications detection by comparative Nanopore direct RNA sequencing (2019).
222. Stoiber, M. *et al.* De novo Identification of DNA Modifications Enabled by Genome-Guided Nanopore Signal Processing (2016).
223. Lentini, J. M., Ramos, J. & Fu, D. Monitoring the 5-methoxycarbonylmethyl-2-thiouridine (mcm5s2U) modification in eukaryotic tRNAs via the γ -toxin endonuclease. *RNA (New York, N.Y.)* **24**, 749–758; 10.1261/rna.065581.118 (2018).
224. Hoefs, J. *Stable Isotope Geochemistry* (Springer International Publishing, Cham, 2015).
225. Hill-Perkins, M., Jones, M. D. & Karran, P. Site-specific mutagenesis in vivo by single methylated or deaminated purine bases. *Mutation research* **162**, 153–163; 10.1016/0027-5107(86)90081-3 (1986).
226. Schuster, H. Die Reaktionsweise der Desoxyribonucleinsäure mit salpetriger Säure. *Zeitschrift für Naturforschung B* **15**, 298–304; 10.1515/znb-1960-0507 (1960).

227. Dong, M., Wang, C., Deen, W. M. & Dedon, P. C. Absence of 2'-Deoxyoxanosine and Presence of Abasic Sites in DNA Exposed to Nitric Oxide at Controlled Physiological Concentrations. *Chem. Res. Toxicol.* **16**, 1044–1055; 10.1021/tx034046s (2003).
228. Lindahl, T. Instability and decay of the primary structure of DNA. *Nature* **362**, 709–715; 10.1038/362709a0 (1993).
229. Frederico, L. A., Kunkel, T. A. & Shaw, B. R. A sensitive genetic assay for the detection of cytosine deamination: determination of rate constants and the activation energy. *Biochemistry* **29**, 2532–2537; 10.1021/bi00462a015 (1990).
230. Polson, A. G., Crain, P. F., Pomerantz, S. C., McCloskey, J. A. & Bass, B. L. The mechanism of adenosine to inosine conversion by the double-stranded RNA unwinding/modifying activity: a high-performance liquid chromatography-mass spectrometry analysis. *Biochemistry* **30**, 11507–11514; 10.1021/bi00113a004 (1991).
231. Bass, B. L. RNA editing by adenosine deaminases that act on RNA. *Annual review of biochemistry* **71**, 817–846; 10.1146/annurev.biochem.71.110601.135501 (2002).
232. Zheng, Y., Lorenzo, C. & Beal, P. A. DNA editing in DNA/RNA hybrids by adenosine deaminases that act on RNA. *Nucleic acids research* **45**, 3369–3377; 10.1093/nar/gkx050 (2017).
233. Gannon, H. S. *et al.* Identification of ADAR1 adenosine deaminase dependency in a subset of cancer cells. *Nature communications* **9**, 5450; 10.1038/s41467-018-07824-4 (2018).
234. Slotkin, W. & Nishikura, K. Adenosine-to-inosine RNA editing and human disease. *Genome medicine* **5**, 105; 10.1186/gm508 (2013).
235. Song, B., Shiromoto, Y., Minakuchi, M. & Nishikura, K. The role of RNA editing enzyme ADAR1 in human disease. *Wiley interdisciplinary reviews. RNA*, e1665; 10.1002/wrna.1665 (2021).
236. Desterro, J. M. P. *et al.* SUMO-1 modification alters ADAR1 editing activity. *Molecular biology of the cell* **16**, 5115–5126; 10.1091/mbc.e05-06-0536 (2005).
237. Desterro, J. M. P. *et al.* Dynamic association of RNA-editing enzymes with the nucleolus. *Journal of cell science* **116**, 1805–1818; 10.1242/jcs.00371 (2003).
238. Sansam, C. L., Wells, K. S. & Emeson, R. B. Modulation of RNA editing by functional nucleolar sequestration of ADAR2. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 14018–14023; 10.1073/pnas.2336131100 (2003).
239. Valente, L. & Nishikura, K. RNA binding-independent dimerization of adenosine deaminases acting on RNA and dominant negative effects of nonfunctional subunits on dimer functions. *The Journal of biological chemistry* **282**, 16054–16061; 10.1074/jbc.M611392200 (2007).

240. Shiromoto, Y., Sakurai, M., Minakuchi, M., Ariyoshi, K. & Nishikura, K. ADAR1 RNA editing enzyme regulates R-loop formation and genome stability at telomeres in cancer cells. *Nature communications* **12**, 1654; 10.1038/s41467-021-21921-x (2021).
241. Tasakis, R. N. *et al.* ADAR1 can drive Multiple Myeloma progression by acting both as an RNA editor of specific transcripts and as a DNA mutator of their cognate genes (2020).
242. Hundley, H. A. & Bass, B. L. ADAR editing in double-stranded UTRs and other noncoding RNA sequences. *Trends in biochemical sciences* **35**, 377–383; 10.1016/j.tibs.2010.02.008 (2010).
243. Pang, B. *et al.* Defects in purine nucleotide metabolism lead to substantial incorporation of xanthine and hypoxanthine into DNA and RNA. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 2319–2324; 10.1073/pnas.1118455109 (2012).
244. Jimeno, S. *et al.* ADAR-mediated RNA editing of DNA:RNA hybrids is required for DNA double strand break repair. *Nature communications* **12**, 5512; 10.1038/s41467-021-25790-2 (2021).
245. cui, h. *et al.* ADAR1 Prevents R-loop Accumulation-Driven ATR Pathway Activation in Ovarian Cancer (2020).
246. Shen, Y. J. *et al.* Genome-derived cytosolic DNA mediates type I interferon-dependent rejection of B cell lymphoma cells. *Cell reports* **11**, 460–473; 10.1016/j.celrep.2015.03.041 (2015).
247. Hamperl, S., Bocek, M. J., Saldivar, J. C., Swigut, T. & Cimprich, K. A. Transcription-Replication Conflict Orientation Modulates R-Loop Levels and Activates Distinct DNA Damage Responses. *Cell* **170**, 774–786.e19; 10.1016/j.cell.2017.07.043 (2017).
248. Stork, C. T. *et al.* Co-transcriptional R-loops are the main cause of estrogen-induced DNA damage. *eLife* **5**; 10.7554/eLife.17548 (2016).
249. San Martin-Alonso, M., Soler-Oliva, M. E., García-Rubio, M., García-Muse, T. & Aguilera, A. Harmful R-loops are prevented via different cell cycle-specific mechanisms. *Nature communications* **12**, 4451; 10.1038/s41467-021-24737-x (2021).
250. Sanz, L. A. *et al.* Prevalent, Dynamic, and Conserved R-Loop Structures Associate with Specific Epigenomic Signatures in Mammals. *Molecular cell* **63**, 167–178; 10.1016/j.molcel.2016.05.032. (2016).
251. Hang, B., Singer, B., Margison, G. P. & Elder, R. H. Targeted deletion of alkylpurine-DNA-N-glycosylase in mice eliminates repair of 1,N6-ethenoadenine and hypoxanthine but not of 3,N4-ethenocytosine or 8-oxoguanine. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 12869–12874; 10.1073/pnas.94.24.12869 (1997).
252. Wong, S. K., Sato, S. & Lazinski, D. W. Substrate recognition by ADAR1 and ADAR2. *RNA (New York, N.Y.)* **7**, 846–858; 10.1017/s135583820101007x (2001).

253. Su, X. A. & Freudenreich, C. H. Cytosine deamination and base excision repair cause R-loop-induced CAG repeat fragility and instability in *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences of the United States of America* **114**, E8392-E8401; 10.1073/pnas.1711283114 (2017).
254. Pâques, F., Leung, W. Y. & Haber, J. E. Expansions and contractions in a tandem repeat induced by double-strand break repair. *Molecular and cellular biology* **18**, 2045–2054; 10.1128/mcb.18.4.2045 (1998).
255. Strand, M., Prolla, T. A., Liskay, R. M. & Petes, T. D. Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *Nature* **365**, 274–276; 10.1038/365274a0 (1993).
256. MACDONALD, M. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* **72**, 971–983; 10.1016/0092-8674(93)90585-E (1993).
257. Kremer, E. J. *et al.* Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence p(CCG)_n. *Science (New York, N.Y.)* **252**, 1711–1714; 10.1126/science.1675488 (1991).
258. Campuzano, V. *et al.* Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. *Science (New York, N.Y.)* **271**, 1423–1427; 10.1126/science.271.5254.1423 (1996).
259. Rice, G. I. *et al.* Mutations in ADAR1 cause Aicardi-Goutières syndrome associated with a type I interferon signature. *Nature genetics* **44**, 1243–1248; 10.1038/ng.2414 (2012).
260. Musheev, M. U., Baumgärtner, A., Krebs, L. & Niehrs, C. The origin of genomic N6-methyl-deoxyadenosine in mammalian cells. *Nature chemical biology* **16**, 630–634; 10.1038/s41589-020-0504-2 (2020).
261. Cantoni, G. L. THE NATURE OF THE ACTIVE METHYL DONOR FORMED ENZYMATIALLY FROM L-METHIONINE AND ADENOSINETRIPHOSPHATE 1,2. *J. Am. Chem. Soc.* **74**, 2942–2943; 10.1021/ja01131a519 (1952).
262. Roignant, J.-Y. & Soller, M. m6A in mRNA: An Ancient Mechanism for Fine-Tuning Gene Expression. *Trends in genetics : TIG* **33**, 380–390; 10.1016/j.tig.2017.04.003 (2017).
263. Lane, A. N. & Fan, T. W.-M. Regulation of mammalian nucleotide metabolism and biosynthesis. *Nucleic acids research* **43**, 2466–2485; 10.1093/nar/gkv047 (2015).
264. Chen, M. *et al.* m6A RNA Degradation Products Are Catabolized by an Evolutionarily Conserved N6-Methyl-AMP Deaminase in Plant and Mammalian Cells. *The Plant cell* **30**, 1511–1522; 10.1105/tpc.18.00236 (2018).
265. Löffler, M., Fairbanks, L. D., Zameitat, E., Marinaki, A. M. & Simmonds, H. A. Pyrimidine pathways in health and disease. *Trends in molecular medicine* **11**, 430–437; 10.1016/j.molmed.2005.07.003 (2005).
266. Wang, L. Mitochondrial purine and pyrimidine metabolism and beyond. *Nucleosides, nucleotides & nucleic acids* **35**, 578–594; 10.1080/15257770.2015.1125001 (2016).

267. Liu, X. *et al.* N6-methyladenine is incorporated into mammalian genome by DNA polymerase. *Cell research*; 10.1038/s41422-020-0317-6 (2020).
268. Huang, W. *et al.* Determination of DNA adenine methylation in genomes of mammals and plants by liquid chromatography/mass spectrometry. *RSC Adv.* **5**, 64046–64054; 10.1039/C5RA05307B (2015).
269. Yu, D. *et al.* Human MettL3-MettL14 RNA adenine methyltransferase complex is active on double-stranded DNA containing lesions. *Nucleic acids research*; 10.1093/nar/gkab460 (2021).
270. Kweon, S.-M. *et al.* An Adversarial DNA N6-Methyladenine-Sensor Network Preserves Polycomb Silencing. *Molecular cell* **74**, 1138-1147.e6; 10.1016/j.molcel.2019.03.018 (2019).
271. Hao, Z. *et al.* N6-Deoxyadenosine Methylation in Mammalian Mitochondrial DNA. *Molecular cell* **78**, 382-395.e8; 10.1016/j.molcel.2020.02.018 (2020).
272. Douvlataniotis, K., Bensberg, M., Lentini, A., Gylemo, B. & Nestor, C. E. No evidence for DNA N6-methyladenine in mammals. *Science advances* **6**, eaay3335; 10.1126/sciadv.aay3335 (2020).
273. Bochtler, M. & Fernandes, H. DNA adenine methylation in eukaryotes: Enzymatic mark or a form of DNA damage? *BioEssays : news and reviews in molecular, cellular and developmental biology* **43**, e2000243; 10.1002/bies.202000243 (2021).
274. Wang, W. *et al.* Epigenetic DNA Modification N6-Methyladenine Causes Site-Specific RNA Polymerase II Transcriptional Pausing. *J. Am. Chem. Soc.* **139**, 14436–14442; 10.1021/jacs.7b06381 (2017).
275. Sambrook, J. & Russell, D. W. *Molecular cloning. A laboratory manual.* 3rd ed. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2001).
276. Schomacher, L. *et al.* Neil DNA glycosylases promote substrate turnover by Tdg during DNA demethylation. *Nature structural & molecular biology* **23**, 116–124; 10.1038/nsmb.3151. (2016).
277. Hashimoto, H. *et al.* Structure of a Naegleria Tet-like dioxygenase in complex with 5-methylcytosine DNA. *Nature* **506**, 391–395; 10.1038/nature12905 (2014).
278. Kellner, S. *et al.* Absolute and relative quantification of RNA modifications via biosynthetic isotopomers. *Nucleic acids research* **42**, e142; 10.1093/nar/gku733 (2014).
279. Tsuji, M. *et al.* A validated quantitative liquid chromatography-tandem quadrupole mass spectrometry method for monitoring isotopologues to evaluate global modified cytosine ratios in genomic DNA. *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences* **953-954**, 38–47; 10.1016/j.jchromb.2014.01.050 (2014).

7. List of Abbreviations

5caC	5-carboxyl cytosine
5fC	5-formyl cytosine
5hmC	5-hydroxymethyl cytosine
5mC	5-methyl cytosine
A.U.	Arbitrary units
AAG	Alkyl-adenine DNA glycosylase
Adal	Adenosine deaminase-like protein
ADAR	Adenosine deaminase acting on RNA
ADAR1p110	Adenosine deaminase acting on RNA 1, 110 kDa isoform
ADAR1p150	Adenosine deaminase acting on RNA 1, 150 kDa isoform
AER	Alternative excision repair
ALKBH	alkB homologue
ALT	Alternative telomere lengthening
AMP	Adenosine monophosphate
AMPA	α -amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid receptor
amu	Atomic mass units
AQR	Aquarius
A-to-I editing	Adenosine-to-inosine editing
BER	Base excision repair
bp	Base pairs
BRCA	Breast cancer type 1 susceptibility protein
C	Cytosine
CAD	Collision activated dissociation
CpA	Cytosine-adenine dinucleotides
CpC	Cytosine-cytosine dinucleotides
CpG	Cytosine-guanine dinucleotides
CpT	Cytosine-thymine dinucleotides
dA	Deoxyadenosine
Da	Daltons
dC	Deoxycytosine
DC	Direct current
dCF	2'Deoxycoformycin/Pentostatin
DDX	DEAD-box-RNA

dG	Deoxyguanosine
dI	Deoxyinosine
dITP	Deoxyriboinosine triphosphate
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase
dNTP	Deoxyribonucleotide triphosphate
dR	Deoxyribose
DRB	Dichloro-1- β -d-ribofuranosyl-benzimidazole
DRIP	DNA:RNA immunoprecipitation
DSB	Double-strand break
dsDNA	Double-stranded DNA
dsRBD	Double stranded RNA binding domain
dsRNA	Double-stranded RNA
EGFP	Enhanced green fluorescent protein
eIF3	Eukaryotic initiation factor 3
ELISA	Enzyme-linked immunosorbent assay
EndoV	Endonuclease V
ESI	Electrospray ionization
FTO	Fat mass and obesity-associated protein
gDNA	Genomic DNA
GFP	Green fluorescent protein
Gria2	Glutamate Ionotropic Receptor AMPA Type Subunit 2
h	Hours
hCNT2	Human concentrative nucleoside transporter 2
HEK293T	Human embryonic kidney cells
HeLa	Henrietta Lacks cervical cancer cell line
HNRNPA2B1	Heterogeneous Nuclear Ribonucleoprotein A2/B1
HU	Hydroxyurea
IMP	Inosine monophosphate
KMT5C	Histone-lysine N-methyltransferase 5C
LC	Liquid chromatography
LC/MS	Liquid chromatography mass spectrometry
LINE	Long interspersed nuclear elements
lncRNA	Long noncoding RNA

m/z	Mass to charge ratio
m6AMP	N6-methyl adenosine monophosphate
m6ATP	N6-methyl adenosine triphosphate
m6dA	N6-methyl adenosine
m6dAMP	N6-methyl deoxy adenosine monophosphate
m6dATP	N6-methyl deoxyadenosine triphosphate
mcm5S2U	5-methoxycarbonylmethyl-2-thiouridine
MeCP2	Methyl CpG binding protein 2
Met+4	Heavy stable isotope labelled methionine, 4 additional mass units
MettL	Methyltransferase-like protein
min	Minutes
MPG	N-methylpurine DNA glycosylase
MS	Mass spectrometry
N6AMT1	N-6 adenine-specific DNA methyltransferase 1
NER	Nucleotide excision repair
NES	Nuclear export sequence
NIH3T3	Murine embryo fibroblast cell line NIH-3T3
NLS	Nuclear localization sequence
NO	Nitric oxide
PKR	protein kinase R
PRC	Polycomb repressive complex
PRDM7	Probable histone-lysine N-methyltransferase 7
PRTM7	Protein arginine N-methyltransferase 7
Q	Quadrupole
RAD52	DNA repair protein RAD52 homolog
RF	Radiofrequency
ribo-A	Adenosine
ribo-I	Inosine
RNA	Ribonucleic acid
RNAP	RNA Polymerase
RNH	RNase H
RNR	Ribonucleotide reductase
ROS	Reactive oxidative species
rRNA	Ribosomal RNA

RT-qPCR	Reverse transcription followed by quantitative pcr
S1	Nuclease S1
SAM	S-adenosyl-L-methionine
SETX	Senataxin
SILAC	Stable isotope labelling by amino acids in cell culture
siRNA	Small interfering RNA
ssDNA	Single-stranded DNA
ssRNA	Single-stranded RNA
SUMO	Small ubiquitin-related modifier
T	Thymidine
TDG	Thymine-DNA glycosylase
Tet1	Ten-eleven-translocation enzyme 1
TOF	Time-of-flight
TRC	Transcription replication collision
tRNA	Transfer RNA
TTS	Transcription terminaton site
U	Uracil
Uhrf1	Ubiquitin Like With PHD And Ring Finger Domains 1
WTAP	Wilms tumour 1-associated protein
XPF	Xeroderma pigmentosum group F
XPG	Xeroderma pigmentosum group G
YTHDF	YTH domain-containing family protein
Z α	Z-DNA binding domain α
Z β	Z-DNA binding domain β
ZBD	Z-DNA binding domain

9. Lebenslauf

Angaben zur Person

Name: Anne Kijonka-Baumgärtner, geb. Baumgärtner

Geburtsdatum: 10.12.1991

Geburtsort: Mainz

Staatsangehörigkeit: Deutsch

Anschrift: Spielbergstraße 1, 55271 Stackeden-Elsheim

Schulische Ausbildung

1998 – 2002 Grundschule Stackeden-Elsheim

2002 – 2011 Kurt-Schumacher Integrierte Gesamtschule Ingelheim

Abschluss: Abitur

Studium und wissenschaftliche Ausbildung

2011 – 2014 Studium Biochemie an der Goethe-Universität Frankfurt

Abschluss: Bachelor of Science

2014 – 2016 Studium Biochemie, Schwerpunkt Virologie an der Universität Ulm

Abschluss: Master of Science

2016 – 2021 Doktorandin, Arbeitsgruppe Prof. Dr. Christof Niehrs, Institut für Molekulare Biologie (IMB), Mainz