Dissertation zur Erlangung des Doktorgrades

der Fakultät für Chemie und Pharmazie

der Ludwig-Maximilians-Universität München

Mass spectrometric analysis of modified nucleotides in embryonic development and disease

Sarah Schiffers

aus

Aachen

2019

„We're all part of the cosmic joke"
- Miranda Bailey

Erklärung

Diese Dissertation wurde im Sinne von §7 der Promotionsordnung vom 28. November 2011 von Prof. Dr. Thomas Carell betreut.

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, den 04.07.2019

Sarah Schiffers

Dissertation eingereicht am: 08.07.2019

1. Gutachter: Prof. Dr. Thomas Carell
2. Gutachter: Dr. Stefanie Kellner

Mündliche Prüfung am: 24.07.2019

Danksagung

Zuallererst möchte ich meinem Doktorvater *Prof. Dr. Thomas Carell* danken, dafür dass er mir interessante Themen und alle Möglichkeiten zur Verfügung gestellt hat, die mich gefördert und gefordert haben.

Des Weiteren bedanke ich mich herzlich bei meiner Zweitgutachterin *Dr. Stefanie Kellner* und meinem restlichen Prüfungskomitee für die Durchführung meiner Promotionsprüfung.

Allen Festangestellten *Markus Müller, Kerstin Kurz, Luis de la Osa de la Rosa* und - zu Beginn meiner Dissertation - *Kristof Hufnagel* gebührt großer Dank für die Hilfe in Organisationsdingen, Vorbereitung des Massenspektrometers und allen Angelegenheiten in der Zellkultur.

Mein größter Dank geht an *Katharina Iwan*. Sie war mir bereits während meiner Masterarbeit eine Mentorin, aber auch während meiner Dissertation hat sie mich stets mit Rat und Tat am Massenspektrometer unterstützt, mir neue Perspektiven aufgezeigt und mir auch auf persönlicher Ebene im Labor eine tolle Zeit ermöglicht.

Natürlich möchte ich mich weiterhin bei all meinen Projektpartnern bedanken: Vielen Dank an meine externen Kooperationspartner *Dr. Stylianos Michalakis, Dr. Alexandra-Viola Bohne, Prof. Dr. Jörg Nickelsen, Laura Bocci, Prof. Dr. Karsten Spiekermann, Dr. Binje Vick, Dr. Irmela Jeremias, Max Emperle* und *Prof. Dr. Albert Jeltsch*. *Charlotte Ebert* und *René Rahimoff* danke ich dafür, dass sie mich in diversen Projekten mit Synthese unterstützt haben. *Olesea Kosmatchev*, *Jessica Steinbacher* und *Thomas M. Wildenhof* haben mir ihr MS-Wissen und ihre Projekte hinterlassen und anvertraut. Gut Ding will Weile haben! *Alexander Schön* und *Ewelina Kaminska* haben mir synthetisch und zellkulturtechnisch mit dem 6-Aza-Projekt geholfen. *Franziska Traube* danke ich dafür, dass sie mit *Katharina Iwan* und mir wochenlang im Konferenzzimmer über unserem Manuskript gebrütet hat und sich mit mir das cAzadC vorgenommen hat. *Dr. Fabio Spada* und *Angie Kirchner* waren mir immer eine große Hilfe in der Zellkultur, bei unserem Desaminierungsprojekt, der Basenexzisionsreparatur und nicht zuletzt im Literaturdschungel. Und last but not least, hat *Eva Korytiaková* mit mir die synthetischen Höhen und Tiefen der Derivatisierung durchgestanden und ist mir eine gute Freundin geworden.

Ich will auch alle anderen Mitglieder der Carell-Gruppe nicht vergessen: Ihr alle habt mich in die Mensa begleitet, mir in der Kaffeeküche und im Computerraum gute Gesellschaft geleistet und wir haben viele witzige Momente auf Parties und Ausflügen erlebt. Wir hatten eine tolle Zeit und ich hoffe, dass wir noch lange Kontakt halten werden.

Natürlich möchte ich auch in meinem privaten Umfeld einigen Personen danken: meinen Eltern *Petra* und *Ulrich Schiffers*, Brüdern *Thomas, Florian* und *Christian Schiffers*, sowie Großeltern *Dr. Ansgar* und

Content

Publications

[1] T. M. Wildenhof, <u>S. Schiffers</u>, F. R. Traube, P. Mayer, T. Carell, Influencing epigenetic information with a hydrolytically stable carbocyclic 5 aza-2'-deoxycytidine., *Angew. Chem. Int. Ed.* **2019**, DOI: 10.1002/anie.201904794.

[2] <u>S. Schiffers</u>[#], T. M. Wildenhof[#], K. Iwan, M. Stadlmeier, M. Müller, T. Carell, Label-Free Quantification of 5-Azacytidines Directly in the Genome., *Helv. Chim. Acta* **2019**, *102,* e1800229*.*

[3] F. R. Traube[#], <u>S. Schiffers</u>[#], K. Iwan[#], F. Spada, M. Müller, T. Carell, Isotope-dilution mass spectrometry for exact quantification of noncanonical DNA nucleosides., *Nat. Protoc*. **2018**, *14*, 283–312.

[4] A. Rajavelu, C. Lungu, M. Emperle, M. Dukatz, A. Brohm, J. Broche, I. Hanelt, E. Parsa, <u>S. Schiffers</u>, R. Karnik, A. Meissner, T. Carell, P. Rathert, R. Z. Jurkowska, A. Jeltsch., Chromatin-dependent allosteric regulation of DNMT3A activity by MeCP2., *Nucleic Acids Res*. **2018,** *46*, 17, 9044-9056.

[5] <u>S. Schiffers</u>, C. Ebert, R. Rahimoff, O. Kosmatchev, J. Steinbacher, A.-V. Bohne, F. Spada, J. Nickelsen, M. Müller, T. Carell. Quantitative LC-MS Provides No Evidence for m^{6}dA or m^{4}dC in the Genome of Mouse Embryonic Stem Cells and Tissues., *Angew. Chem. Int. Ed.* **2017**, *56*, 11268–11271.

[#]These authors contributed equally

Conference Participation

2018 – 26th Pasteur-Weizmann Symposium. Paris, France.

2018 – SFB 1361 Evaluation. Mainz, Germany; poster presentation.

2018 – 52nd ESBOC Symposium "Life and Death of Nucleic Acids". Gregynog, Wales; poster presentation.

2017 – ScienceRocks!. Munich, Germany; talk.

2017 – 17th Symposium on Chemistry of Nucleic Acid Components. Krumlov, Czech Republic; poster presentation.

2016 – SPP1784 PhD Meeting. Dortmund, Germany; poster presentation and short talk.

Zusammenfassung

Diese Arbeit untersucht DNA und RNA Modifikationen und deren Bedeutung in der Epigenetik und der Entstehung von Krankheiten. In einem ersten Projekt lag die Aufklärung des Vorkommens unbekannter DNA Modifikationen im Vordergrund. Mithilfe einer neuen quantitativen massenspektrometrischen Methode konnte nachgewiesen werden, dass die Modifikationen N^4-methyl-2'-desoxycytidin (m^{4}dC) nicht und N^6-methyl-2'-desoxyadenosin (m^{6}dA) in murinen Stammzellen (mESCs) und Geweben maximal zu ein paar hundert Nukleotiden pro Genom vorkommen. Weiterhin wurde bestätigt, dass bakterielle Kontaminationen falsch-positive Quantifizierungen verursacht haben könnten, die aufgrund unpräziser Datenevaluation in mehreren Fachjournalen publiziert worden sind. Zusätzlich sollte aufgeklärt werden, ob eventuell vorhandenes m^{6}dA durch enzymatische Methylierung der DNA-Base zustande kommt, oder ob es alternative Quellen gibt. Hierzu wurden die m^{6}dA und m^{6}A Nukleoside ins Zellmedium gegeben, es wurde auch m^{6}A-enthaltende RNA transfiziert, und die Differenzierung von mESCs wurde induziert. Ziel war es potenzielle endogene und exogene Ursachen und Quellen für m^{6}dA in genomischer DNA zu evaluieren. Die Ergebnisse ergaben Hinweise darauf, dass die Entstehung von m^{6}dA aus dem RNA-Baustein erfolgt, welcher infolge von RNA-Degradierung im Zytosol vorhanden ist.

Ein weiterer Fokus dieser Arbeit lag auf der Erforschung einer aktiven Demethylierung des 5-Methyl-2'-desoxycytidins (m^{5}dC) durch enzymatische Desaminierung, Deformylierung, sowie Basenexzisionsreparatur während der Embryonalentwicklung. Für die Untersuchung wurden *in vivo* Isotopenverfolgungsstudien durchgeführt, um Änderungen der Modifikationslevel in verschiedenen Zellen zu ermitteln. Als Grundlage für die Untersuchung aller biochemischen Experimente mussten geeignete massenspektrometrische Messmethoden entwickelt werden.

Die Untersuchung eines Mechanismus für einen direkten C-C Bindungsbruch an 5-Formyl-2'-desoxycytidin erfolgte durch Applikation von 6-Aza-Nukleosid-Analoga der 2'-Desoxycytidine. Durch das Stickstoff-Atom in der 6-Position sollte ein enzymatischer, nukleophiler Angriff blockiert sein. In der Tat wurde trotz Einbaus des gefütterten 6-Aza-5-formyl-2'-desoxycytidins in die DNA kein unmodifiziertes 6-Aza-2'-desoxycytidin beobachtet. Aufgrund des durch die hohe Zytotoxizität geringen Einbaus des Analogons kann jedoch nicht endgültig ausgeschlossen werden, dass die Eliminierung der Formylgruppe nicht doch einen nukleophilen Angriff an Position 6 erfordert.

Die aktive Demethylierung über Basenexzisionsreparatur wurde auf zwei verschiedene Arten unter Zuhilfenahme von Hydroxylamin-basierten Derivatisierungsreagenzien untersucht. Zuerst wurde ein bereits bekanntes Protokoll zur Derivatisierung abasischer Stellen verwendet, um die Aktivität verschiedener Glykosylasen zu analysieren. Hierbei konnten keinerlei Hinweise auf eine Beteiligung der BER an der Prozessierung von Methylcytosinen festgestellt werden. Allerdings könnten aufgrund

einer schnellen Prozessierung der mutagenen DNA-Läsionen die Intermediate der BER unter dem Detektionslimit liegen. Aus diesem Grund stand die Entwicklung eines Derivatisierungsprotokolls im Vordergrund, welches das Reagenz statt mit der abasischen Stelle mit dem ausgeschnittenen Formylcytosin im löslichen Nukleosidpool reagieren lässt.

Die Untersuchung einer potentiellen Desaminierungsreaktion auf der DNA-Ebene erfolgte durch metabolische Markierung des m^5dC in embryonalen Stammzellen der Maus durch Zugabe von [^{13}C,D$_3$]-markiertem L-Methionin zum Zellmedium. Dieses Molekül kann als Vorstufe des methylübertragenden S-Adenosyl-L-Methionins [^{13}C,D$_3$]-markiertes m^5dC generieren, dessen Desaminierung anschließend durch die Existenz von [^{13}C,D$_3$]-markiertem dT nachgewiesen werden kann. Da S-Adenosyl-L-Methionin nicht an der direkten Synthese von dT aus dU beteiligt ist, kann eine andere Entstehung dieser markierten Modifikation ausgeschlossen werden. Die Untersuchungen führten zur Detektion des [^{13}C,D$_3$]-markierten dT und somit zu einem Nachweis der Entstehung aus dem markierten m^5dC, besonders während des *Priming* von embryonalen Stammzellen der Maus. Des Weiteren konnte festgestellt werden, dass in Abwesenheit von DNMT3b die Bildung des markierten dT reduziert ist und dass auch Komplementierung mit dem Enzym das Ursprungsniveau nicht wiederherstellen kann. Studien an einer *Doppelknockout*-Zelllinie der Desaminasen des löslichen Nukleosid-/Nukleotid-Pools zeigten, dass markiertes dT in verringerter Menge vorkommt. Die Ergebnisse bestätigen, dass der größte Teil der beobachteten Desaminierungen an durch Reparatur herausgeschnittenen m^5dC Nukleosiden stattfindet. Diese Ereignisse stehen im Zusammenhang mit der Aktivität von DNMT3 Enzymen und könnten eine Form der lokalen Demethylierung darstellen. Entsprechende Genabschnitte, auf die dies zutrifft, stellen die Loci der genomischen Prägung dar. Untersuchungen an uni-parentalen Zellen konnten abschließend belegen, dass es zu einer verminderten Entstehung von markiertem dT kommt, wenn vererbbare differentielle Methylierungsmuster nicht vorhanden sind.

In einem letzten Projekt wurde die Wirkung der Nukleosidanaloga 5-Azacytidin und 5-Aza-2'-desoxycytidin an embryonalen Stammzellen der Maus, sowie Zelllinien der akuten myeloiden Leukämie untersucht. Durch die Blockierung der 5-Position können Methyltransferasen kovalent gebunden werden, was eine globale Erniedrigung der DNA-Methylierung hervorruft. Es wurde eine neue Massenspektrometrie-basierte Methode entwickelt, die durch chemische Stabilisierung des Hydrolyse-empfindlichen Moleküls eine direkte und simultane Quantifizierung der Modifikation und der methylierten (2'-Desoxy)cytidine in der DNA und RNA ermöglicht. Mithilfe dieser Analyse-Methode wurde herausgefunden, dass alle Methyltransferasen gleichermaßen empfindlich auf Azacytidine reagieren. Zuvor war postuliert worden, dass die *Maintenance* Methyltransferase DNMT1 besonders empfindlich auf Aza(d)C in der DNA reagieren würde. Zudem konnte gezeigt werden, dass die Sensitivität von Leukämiezellen gegenüber ribo- bzw. desoxyribo-Azacytidinen nicht mit deren

Einbauraten einhergeht, obwohl unterschiedliche Zelllinien klare Unterschiede im Aza(d)C Einbau aufwiesen. Auch die Reduktion der Methylierung scheint nicht proportional zum Einbau des Analogons. Scheinbar besitzen diese Substanzen eine komplexe Wirkung auf unterschiedliche Krebsarten, die sich mit einer direkten Wirkung auf die Methyltransferasen und entsprechende Reparaturmechanismen an der DNA alleine nicht erklären lässt. Die Applikation von 5-Azacytidin auf re-isolierte, von Patienten abgeleitete Xenotransplantat-Zellen gab des Weiteren zusätzliche Hinweise auf die komplexe, von Patient zu Patient unterschiedliche Wirkung des Therapeutikums. Die Analyse-Methode könnte eine schnelle und einfache Möglichkeit darstellen, um bei Untersuchungen in der Klinik eine Inkorporation des Analogons in die DNA und RNA zu belegen und den Effekt auf die Methylierungsgrade zu evaluieren. Dies könnte frühzeitig eine Identifikation einer Resistenz von Patienten gegenüber den Substanzen ermöglichen. Eine Anwendung von anderen Therapie-Optionen könnte so früher in Erwägung gezogen werden. Da Azacytidine sehr Hydrolyse-empfindlich sind und ein Teil ihrer Wirkung damit auf ihre akute Toxizität zurückzuführen ist, wurde zusätzlich ein carbozyklisches Derivat des 5-Aza-2'-desoxycytidins in Hinblick auf dessen Effekt auf die DNA-Methylierung untersucht. Das Derivat ist stabil gegenüber Hydrolyse und wird in großen Mengen in die DNA eingebaut, wo es eine Reduzierung von m^5dC verursacht. Einbau und nachfolgende Effekte werden jedoch erst nach längerer Inkubationszeit und unter Zugabe des Moleküls in höherer Konzentration ersichtlich. Die Ergebnisse indizieren, dass DNA Methyltransferasen zwar vermutlich kovalent gebunden werden, aber das Molekül nicht leicht in die Zelle aufgenommen oder in das Triphosphat umgewandelt werden kann. Nichtsdestotrotz könnte die Stabilität des Moleküls eine geringere Mutagenität aufweisen, da das Nukleosid in der DNA nicht zerfällt und keine Strangbrüche erzeugt werden. Des Weiteren könnte die Stabilität eine lediglich einmalige Applikation des Therapeutikums pro Behandlungszyklus ermöglichen und letztlich zeitverzögert denselben Effekt auf DNA-Methylierung und Genaktivierung verursachen. Dies würde für den Patienten eine angenehmere und für die Krankenkassen eine kosteneffizientere Therapie bedeuten.

Summary

The research presented here analyzes DNA and RNA modifications and their role in epigenetics and disease formation. In the first project, I focused on the evaluation of the existence of unknown DNA modifications. Using a new mass spectrometry based quantitative method, we were able to uncover that N^4-methyl-2'-deoxycytidine does not occur and N^6-methyl-2'-deoxyadenosine might at maximum be present to a few hundred nucleotides per genome in mouse embryonic stem cells and tissues. Furthermore, we verified that bacterial contaminations can easily cause a false-positive quantification, which due to unprecise data evaluation had been published in multiple journals. Additionally, we wanted to investigate whether potentially occurring m^6dA can form through enzymatic methylation of the DNA base, or if there are alternative causes and sources. To this end, we administered the m^6dA and m^6A nucleosides, transfected m^6A-containing RNA and induced differentiation to evaluate potential stimulated endogenous and exogenous causes. Our results give evidence on the formation of m^6dA from the RNA building block, which stems from cytosolic RNA degradation.

A second part of this thesis investigated potential pathways for active demethylation of 5-methyl-2'-deoxycytidine through enzymatic deamination, deformylation, as well as base excision repair during embryonic development. For this investigation, we conducted *in vivo* isotope tracing studies to determine changes in the modification levels of different cells. As a basis for the analysis of the biochemical experiments suitable mass spectrometric methods had to be developed.

Investigation of a mechanism for a direct C-C bond cleaving reaction on 5-formyl-2'-deoxycytidine as part of active demethylation of m^5dC was performed by administration of 6-aza nucleoside analogues of 2'-deoxycytidines. The nitrogen atom in the 6-position was supposed to block an enzymatic nucleophilic attack. Indeed, despite incorporation of the administered 6-aza-5-formyl-2'-deoxycytidine into DNA, no deformylated 6-aza-2'-deoxycytidine species was observed. Since however the levels of the supplemented nucleoside were rather low due to the high cytotoxicity of the analogue, we cannot fully exclude the possibility that elimination of the formyl group does not require attack on position 6. Active demethylation *via* base excision repair was investigated in two different ways utilizing hydroxylamine-based derivatization reagents. First, we applied a previously established protocol for the derivatization of abasic sites to analyze the function of various glycosylases. We could not find evidence for the involvement of base excision repair in processing of methylcytosines. Due to fast processing of the mutagenic DNA lesions intermediates of BER might be present below the limit of detection. Consequently, we started the development of a derivatization protocol that captures the excised formylcytosine from the soluble pool.

Investigation of a potential deamination reaction on the DNA level was performed through labeling of 5-methyl-2'-deoxycytidine (m^5dC) in mouse embryonic stem cells by administration of $[^{13}C,D_3]$-labeled

L-methionine to the cell medium. This molecule as a precursor for the methyl donor *S*-adenosyl-L-methionine can generate $[^{13}C,D_3]$-labeled m^{5}dC, whose subsequent deamination can be proven by the existence of $[^{13}C,D_3]$-labeled dT. Since *S*-adenosyl-L-methionine is not involved in the synthesis of dT from dU, another source for the formation of this modified nucleoside can be excluded. Our results detected $[^{13}C,D_3]$-labeled dT especially during the *priming* of mouse embryonic stem cells, which proves its generation from labeled m^{5}dC. Furthermore, we found out that in absence of DNMT3 the formation of the modification is reduced and that complementation with the enzyme cannot recover original values. Research on double *knockout* mESCs of the deaminases of the soluble nucleoside/nucleotide pool resulted in detection of reduced levels of labeled dT. These results confirm that the major part of the observed deamination happens on m^{5}dC nucleosides that had been excised through DNA repair. These events correlate with the activity of DNMT3 enzymes and could cause local demethylation. Prime candidates for these loci are genomic imprints. Studies on uni-parental cell lines were able to prove a reduced formation of the modification, when parental-specific methylation patterns are not present disrupted.

In a final project, the mode of action of the nucleoside analogues 5-azacytidine and 5-aza-2'-deoxycytidine was analyzed in mESCs and cell lines of acute myeloid leukemia. Due to the blockage of the 5-position the methyltransferases can be bound covalently and induce global reduction of DNA methylation. We developed a new mass spectrometry-based method which enables direct and quantitative measurement of the modifications and methylated (2'-deoxy)cytidines in the DNA and RNA through chemical stabilization of the hydrolysis-sensitive molecule. Using this method, we found out that methyltransferases react equally sensitive towards ribo- and deoxyribo-azacytidines. Previously it was reported that the maintenance methyltransferase DNMT1 was more sensitive towards azacytidines in the DNA. Furthermore, the incorporation of the analogue into different cell lines of acute myeloid leukemia is diverse and does not represent the sensitivity of the cells towards the therapeutics. The reduction of the methylation is also not proportional to the incorporation of the analogues. We assume that the substances possess a complex effect on different cancer types that cannot be explained only by direct effects on methyltransferases and corresponding repair mechanisms. Administration of 5-azacytidine on re-isolated patient-derived xenograft cells gave additional evidence on the complex, patient-dependent effect of the therapeutic. We postulate that our analytical method offers a fast and simple possibility to determine the incorporation of the analogues into DNA and RNA and to evaluate their effect on methylation levels. This could enable timely identification of a resistance of the patient towards the substances. Application of different therapeutic strategies can therefore be considered earlier.

Due to the low hydrolytic stability of azacytidines and the resulting acute toxicity, we furthermore investigated a carbocyclic derivative of 5-aza-2'-deoxycytidine in regards to its effect on DNA

methylation. The derivative is stable towards hydrolysis and is incorporated into DNA to high amounts, where it causes reduction of m^{5}dC. Incorporation and subsequent effects are only observed after longer incubation times and administration of higher concentrations of the molecule. We propose covalent binding of the DNA methyltransferases, but reduced cell uptake or formation of the respective triphosphate. Nevertheless, we presume that the hydrolytic stability causes less mutagenicity due to decreased degradation in the DNA and therefore lesser occurrence of DNA strand breaks. Furthermore, the stability could enable single administration of the therapeutic per treatment cycle, which might be sufficient for the same, potentially time-lagged effect on DNA methylation. Administration of this nucleoside analogue would enable a less strenuous treatment for the patients.

1. Introduction

1.1 Deoxyribonucleic acid

The fate of an organism strongly depends on its composition - on not only what kind of cells it is comprised of, but rather what the cells contain and how they are regulated. As the storage material for every organism's genetic information, deoxyribonucleic acid (DNA) is one of the major macromolecules in every cell. The DNA sequence usually remains the same throughout the whole lifetime of an organism, unless damage or specialized processes occur. People use the expression: "it's written in his genes" thinking of the fact, that the DNA of a person plain in its sequence, and the kinds of genes it holds, determines the looks and behavior of the individual. What most people do not know is the fact that the sequence can be read and interpreted in a multitude of ways leading to different outcomes. Determination of how and why specific regulation of the gene expression occurs, is fundamental to understand developmental processes and to determine possible improvement options for disease states.

DNA does not only consist of the canonical Watson-Crick bases, but depending on the organism and the condition it is in, it additionally contains modified nucleosides. These modifications can have several origins. Enzymatic processes can endogenously alter nucleotides to functionalize them for the recognition of special proteins. Since a cell is exposed to a multitude of environmental influences all the time, damage can and will occur multiple times a day and mutate the DNA. This contributes to genomic variation, and may result in beneficial, neutral, or harmful consequences for an organism. Well-established sources of genomic mutation are enzymatic editing events – processes that create immunogenic diversity[1] –, DNA replication errors, and environmental mutagens, such as certain chemicals,[2] light and ionizing radiation. Radiation leads to radical formation of small molecules, which can add to the DNA, *e.g.* Benzo[*a*]pyrene.[3] Absorption of UV light leads to the reaction of neighboring thymidines producing pyrimidone 6-4 pyrimidine photoproducts.[4] Other known DNA damage products are thymidine glycols[5] and 7,8-dihydro-8-oxoguanines (8-oxodG),[6] which form by oxidation. 2'-Deoxyuridine (dU) is created by deamination,[7] or 1,N^6-ethenoadenosine can form by lipid peroxidation.[8]

1.1.1 DNA repair mechanisms

The maintenance of the DNA's composition is necessary to guarantee the production and survival of normally functioning cells. Therefore, various mechanisms have evolved to repair damages in the DNA. Bulky adducts can distort the DNA helix and are mostly processed by nucleotide excision repair (NER), releasing a 25-30 nucleotide long DNA strand. This process will lead to DNA strand breaks, which have to be repaired with an enormous effort to keep the DNA sequence unaltered. NER can be executed as

global genome repair (GGR), if the helix-distorting lesion is recognized specifically, or transcription-coupled repair (TCR), if a lesion stalls an RNA polymerase on the transcribed strand. Subsequent common steps include dual incisions by structure-specific endonucleases together with NER-specific helicases releasing the DNA strand. Symmetric alterations can cause double strand breaks and have to be repaired by homologous recombination repair (HRR) or non-homologous end joining (NHEJ). The latter of the two can cause severe genetic mutations, because the excised DNA strand might be deleted or not replaced in the right way.

Single nucleotide DNA lesions can be recognized by enzymes of the mismatch repair (MMR) or base excision repair (BER) pathway. Such lesions can be caused *e.g.* by oxidations, misincorporations or deaminations. Excision of the modified or incorrect base from the nucleotide is necessary for maintaining the integrity of the DNA. Enzymatic deglycosylation reactions cleave the N-glycosidic bond, release the base and leave an apurinic/apyrimidinic (abasic, AP) site. To date, eleven glycosylases have been discovered in mammals to work on a variety of DNA modifications. Substrate recognition involves rotation of DNA bases out of the DNA helix into a specific binding pocket with the catalytic active site.[9] The cleavage of the N-glycosidic bond involves nucleophilic attack and acid-base catalysis. Some of the glycosylases are monofunctional with only one glycosylase activity using water to hydrolyze the bond.[10] Bifunctional glycosylases additionally exhibit AP lyase activity, where they use a lysine amino group to form a Schiff-base with the 2'-deoxyribose[11-12] directly. Subsequent proton transfer reactions involving the excised base cleave the DNA strand at the 3'-phosphate.[13-14] Glycosylases have a high affinity for the AP sites they generate.[15-19] The tight binding and steric hindrance of BER enzymes blocks the glycosylase activity on the counter strand and reduces the probability of double strand breaks in symmetrical CpGs.[20] AP sites can also form spontaneously by hydrolysis. In yeast, AP sites mainly stem from misincorporated 2'-deoxyuridine during replication[21] and were shown to be repaired partly by NER.[22] After their generation, the AP sites[23-25] need to be processed further. Since AP sites do not distort the DNA helix, but can stall RNA polymerases, their repair is mediated by TC-NER.[26-27] Translesion synthesis at AP sites can lead to mutations, because the polymerases predominantly incorporate dA in the position of the lesion.[28] The most effective repair pathway for processing of AP sites is direct excision of the free 2'-deoxyribose (see **Figure 1**).
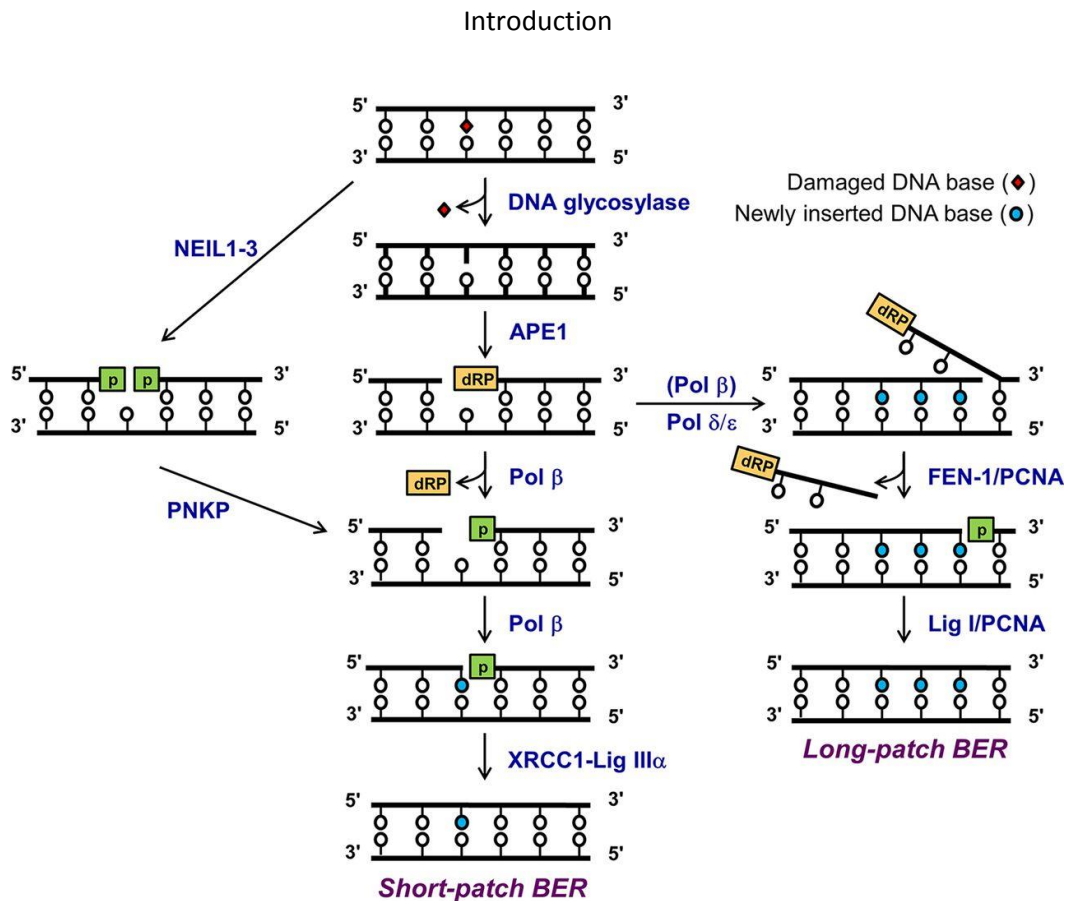
Figure 1: Graphic overview over the enzymatic reactions involved in base excision repair. Taken from *Carter et al*.[29]

First, an *AP endonuclease* (APE 1 and 2), in mice and humans mostly APE1,[30] cleaves the DNA phosphodiester bond at the 5'-end of the lesion[31] leaving a 5'-deoxyribose phosphate (5'-dRP). If AP lyases nick the sugar phosphate backbone, cleavage on the 3'-side of the AP site is also possible.[32] For insertion of a new nucleotide, a DNA polymerase needs a 3'-OH and a 5'-phosphate on the strand break. *DNA polymerase β* (Polβ) possesses an intrinsic 5'-dRPase activity[33-34] to generate the 5'-phosphate, but other polymerases need additional enzymes to process the ends *e.g. polynucleotide 5'-hydroxyl-kinase* (PNK). To finally close the strand break DNA ligases are necessary, which fuse the newly incorporated nucleotide with the other end of the strand. A recent publication reports the existence of a so-called BERosome,[35-36] in which all reactions take place in an orchestrated fashion. *X-ray repair complementing defective repair in Chinese hamster cells 1* (XRCC1) is an interactor with many components of the BER pathway: it is involved in deglycosylation reactions with *Nei-like DNA glycosylases* (NEILs),[37-38] it interacts with APE1, PNK[39] and possibly other end-cleaning enzymes[40] and is associated with Polβ[41] and *DNA ligase* (Lig IIIα).[42] Therefore, it is regarded as a scaffolding and potentially recruiting protein for the BER proteins.

1.1.2 Epigenetics

C. H. WADDINGTON[43] first observed differential cell fates in 1942. He postulated that mechanisms take places, which cannot be explained just with interpretation of genes and termed this "epigenesis" from

the Greek word "epi", which translates to "on top of". Later-on, this lead to an establishment of the term "epigenetics", which describes the dynamic information layer, that exists in addition to the DNA sequence. Now it is well-known, that regulation of gene expression is essential for modulation of biological pathways in the cell. One regulatory mechanism involves installation of modified nucleotides, like 5-methyl-2'-deoxycytidine (m^{5}dC)[44] or 5-hydroxymethyl-dC (hmdC)[45] by specialized enzymes ('writers'). Detection of these modifications by other proteins ('readers') can subsequently induce or suppress a biological reaction. When their effect needs to be reversed, specific enzymes ('erasers') can finally remove the modifications.

1.1.2.1 The role of m^{6}dA and N^4-methyl-dC

In prokaryotes, a common modification is methylation of dC in the N4-position. The so-called *N*^{4}-methyl-dC[46-47] (m^{4}dC) is highly abundant *e.g.* in *Synechocystis*. Its function has, however, so far only been investigated in *Helicobacter pylori,* where it was shown to regulate transcription and pathogenesis.[48] Another very abundant modification in prokaryotes is *N*^{6}-methyl-2'-deoxyadenosine (m^{6}dA).[46] This modification is an important component of the restriction/modification system in bacteria to defend against bacteriophage invasion.[49] In detail, the methylation of the host cell DNA functions as protection against cleavage by restriction endonucleases. This enables selective degradation of invasive unmethylated genomic material. The modification is furthermore involved in regulation of DNA replication, repair and transcriptional control in prokaryotes.

In addition, m^{6}dA was also reported to occur in various unicellular eukaryotes, *e.g. Trypanosoma cruzi,*[50] the ciliates *Tetrahymena pyriformis*[51-52] and *thermophila,*[53-56] and many other uni- and multicellular organisms.[52, 57-61] In *Tetrahymena thermophila*, exclusive association of the modification with RNA Polymerase II transcribed genes was described.[62] Furthermore, several groups report that it occurs in the linker DNA,[63-65] *e.g.* of H2A.Z-containing nucleosomes,[62] where it might direct the positioning of the nucleosomes.[66] The methyltransferase responsible for generation of m^{6}dA in *Tetrahymena* was found to be Tamt-1.[66] In *Blepharisma japonicum*[67] reduction of the modification was correlated with gene activation. In *Chlamydomonas reinhardtii,* the modification might function as a mark of active transcription start sites.[60]

Additionally, the nucleotide was reported to occur in traces in higher eukaryotes like *Aedes albopictus,*[68] and in adult rats.[69] For a long time, however, the existence of m^{4}dC and m^{6}dA in higher eukaryotes had not been proven[70] and with less than one m^{6}dA in one million nucleotides[71] even been attributed to bacterial contaminations. Recent investigations of different groups that established sensitive protocols, however, claim detection of m^{6}dA in *Caenorhabditis elegans,*[72] *Drosophila melanogaster,*[73] *Xenopus laevis*[74] and even human[75] and murine cells and tissues[74, 76] (see **Figure 2**).
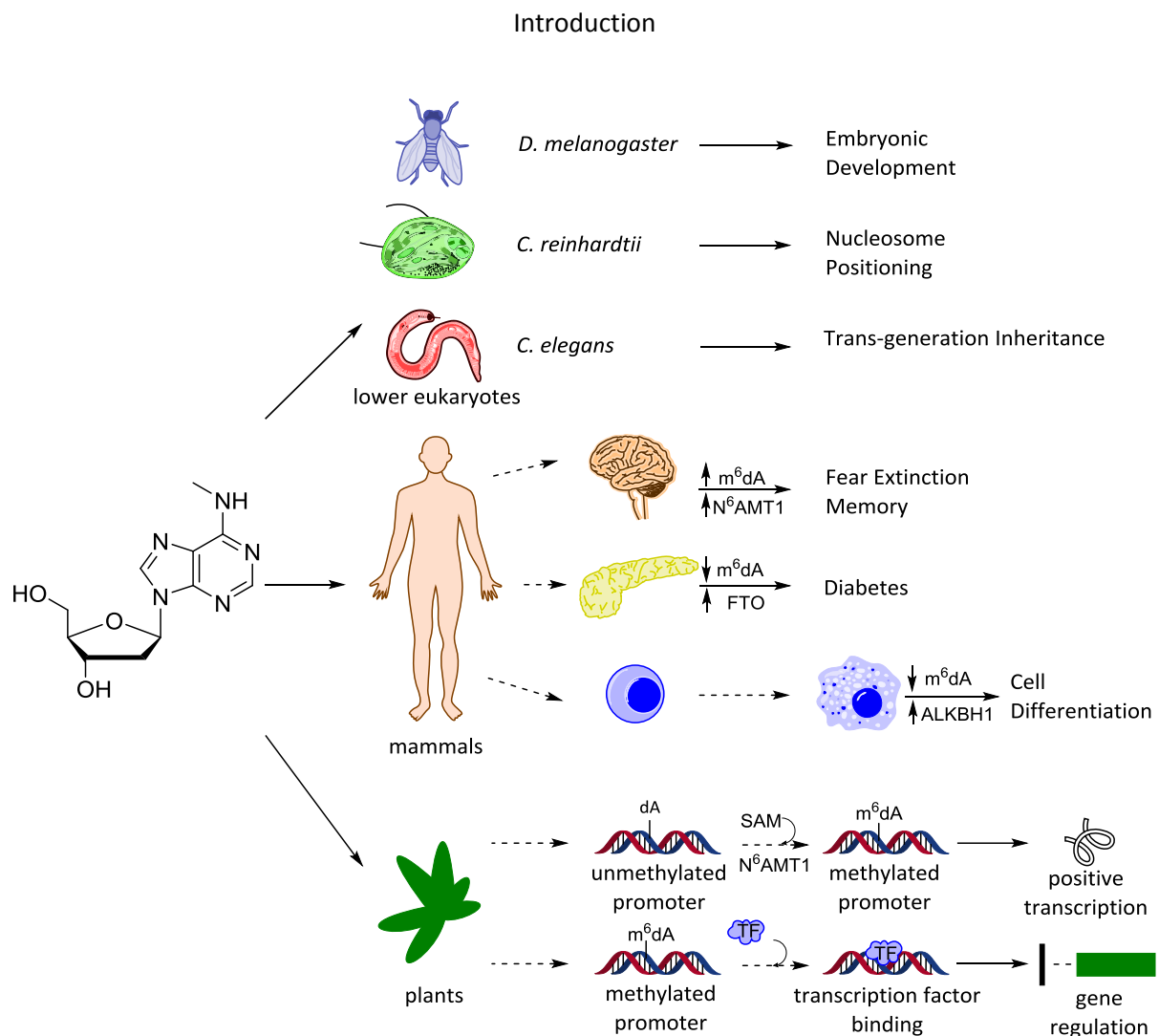
Figure 2: Schematic overview over occurrence and functions of m^{6}dA in various organisms. Adapted and modified from *Parashar et al.*[77]

For some organisms, like *C. elegans,*[72] even the existence of a methyltransferase DAMT-1 and a demethylase NMAD-1 were described. The research suggests a potential crosstalk between the modification and the *histone 3 lysine 4 dimethyl demethylase* Spr-5, which affects inheritance of infertility. These findings sparked even greater interest in the research community and led more groups to investigate different organisms for the existence of m^{6}dA. Single molecule real time (SMRT) sequencing by *Liang et al.*[78] subsequently reported widespread occurrence of the modification in the *Arabidopsis thaliana* genome, mostly in gene bodies and not in intergenic regions.[78] The researchers furthermore claim, that m^{6}dA is associated with actively expressed genes and that it shows dynamics during development.[78] In the *Oryza sativa* genomes, m^{6}dA was reported to be involved in gene expression, plant development and stress response.[79] The increased occurrence of the modification due to mitochondrial stress was furthermore reported for *C. elegans,*[80] confirming the results from *Greer et al.*[72] Similarly, m^{6}dA was reported to be found in *M. musculus* brain[81] upon environmental restraining stress, and to be altered in the serotonin receptor *Htr2a* gene promoter upon early-life stress in rat.[82] The elevated levels of m^{6}dA were claimed to correspond with gene repression. In pig

and zebrafish,[83] the modification was associated with early embryogenesis. Even in humans, the methyltransferase N^6-adenine specific DNA methyltransferase 1 (N6AMT1)[84] and the demethylase *non-heme Fe^{II}/α-ketoglutarate dependent dioxygenase* ALKBH1[76, 85] have recently been suggested to install and remove m⁶dA in cells and were associated with cancer formation.[84, 86] A potential role for the modification in humans was furthermore reported in correlation with *fat mass and obesity associated protein* (FTO) and Diabetes Mellitus Type II,[87] and in systemic lupus erythematosus.[88] Additionally, m⁶dA was discovered to be a more abundant modification in the DNA of mitochondria, where it was suggested to be recognized by *single-stranded DNA-binding protein 1* (SSBP1), a replication factor associated with the heavy strand of mitochondrial DNA (mtDNA). This finding could correlate m⁶dA with the regulation of mtDNA replication.[89] Most recently, even an oxidized derivative of the modification -N^6-hydroxylmethyl-adenine - was reported, that might be installed by ALKBH1 and is thought to be associated with lung carcinoma.[90]

Despite the recent findings, it is still under debate, whether m⁶dA exists and has a function in eukaryotes. In 2006, a sensitive HPLC-MS method did not detect m⁶dA in mouse tissues or human mtDNA and determined the maximum possible amount of the modification to less than one m⁶dA in one million nucleotides.[71] This finding was challenged by researchers, who claimed to have found the modification in the range of 0.0001-0.1% of dA in mESC or human cells.[74, 76, 87] Several groups[91-92] subsequently investigated the robustness of the described methods for the quantification of m⁶dA in different organisms. *Lentini et al.*[92] report overestimations of the nucleoside because IgG-based antibodies often possess binding affinities towards short tandem repeat DNA motifs, that do not contain modified nucleosides and immunoprecipitation data analysis is performed without appropriate controls. The researchers additionally found substantial amounts of reads in the respective sequencing data sets that stem from bacterial contaminations. *O'Brown et al.*[91] furthermore uncovered biased UHPLC-MS data analysis due to nucleoside contaminations in commercially available nucleases and artefact levels that stem from bacterial contamination in DNA preparations.[91] They however still report a detection of the nucleoside in preparations from mice that were raised without germs or microbiome. As a conclusion, trace analysis of nucleosides is a challenging task that requires careful sample preparation with adequate controls and cautious data analysis to exclude impurities from other sources or to identify artefacts.

1.1.2.2 The role of m⁵dC

The most frequent modification in DNA of higher eukaryotes is methylation on dC and it is essential for mammalian development.[44] The underlying reason is stable transcriptional silencing,[93] which affects genome stability, X chromosome inactivation, genomic imprinting, and the silencing of retrotransposons.[94-96] Methylated dC (m⁵dC) is well-studied in vertebrate DNA and occurs widespread

in palindromic CpG dinucleotides.[97] In somatic cells, accumulated CpGs, so-called CpG islands, are however usually unmethylated.[98] A study comparing human somatic and pluripotent embryonic stem cell (ESC) lines found global dC methylation levels of 4.25 and 5.83%, respectively, of which 99.8 and 75.5% occur at CpG sites.[99] Furthermore, relatively high abundance of m^5dC in non-CpG context has been reported in mouse oocytes,[100] pluripotent ESCs[99, 101-102] and mature neurons.[103]

The modification is installed by DNA methyltransferases (DNMT1, 3a and 3b).[104-105] The underlying enzymatic reaction for the methylation is covalent binding of the enzyme to the base in the C6-position and transfer of a methyl group from SAM under acid-base catalysis. DNMT3a and b are so called *de novo* methyltransferases,[105] establishing methylation on completely unmethylated DNA regions. They preferentially associate with the tightly condensed heterochromatin surrounding the central connection points of the sister chromatids (pericentromeric heterochromatin).[106] In comparison, DNMT1 is a maintenance methyltransferase, which installs methyl groups in hemimethylated regions. Therefore, it localizes to replication foci during S phase,[107-108] where the newly synthesized strand is first unmethylated distinguishing it from the parental strand. The hemimethylated sites are recognized by *Ubiquitin-like PHD and RING finger domain-containing protein 1* (UHRF1), which then recruits DNMT1 to methylate the daughter strand. In certain circumstances, DNMT1 can also methylate *de novo*, if the other enzymes are not present and *vice versa*.[109-110] A fourth member of the family, DNMT3l, does not have any catalytic activity, but was shown to increase the binding of the methyl donor *S*-adenosyl-L-methionine (SAM) to the shortened DNMT3a isoform DNMT3a2. Furthermore, it plays an important role in the development of the sexual reproduction system, the germ line.[111-113]

If the levels of m^5dC are altered unspecifically, developmental processes are impaired and cancer can occur.[114-116] The m^5dC nucleotide is considered epigenetic, because its occurrence in promoter regions silences the associated genes,[104] and removal of the modification can reactivate the gene expression (see **Figure 3**).
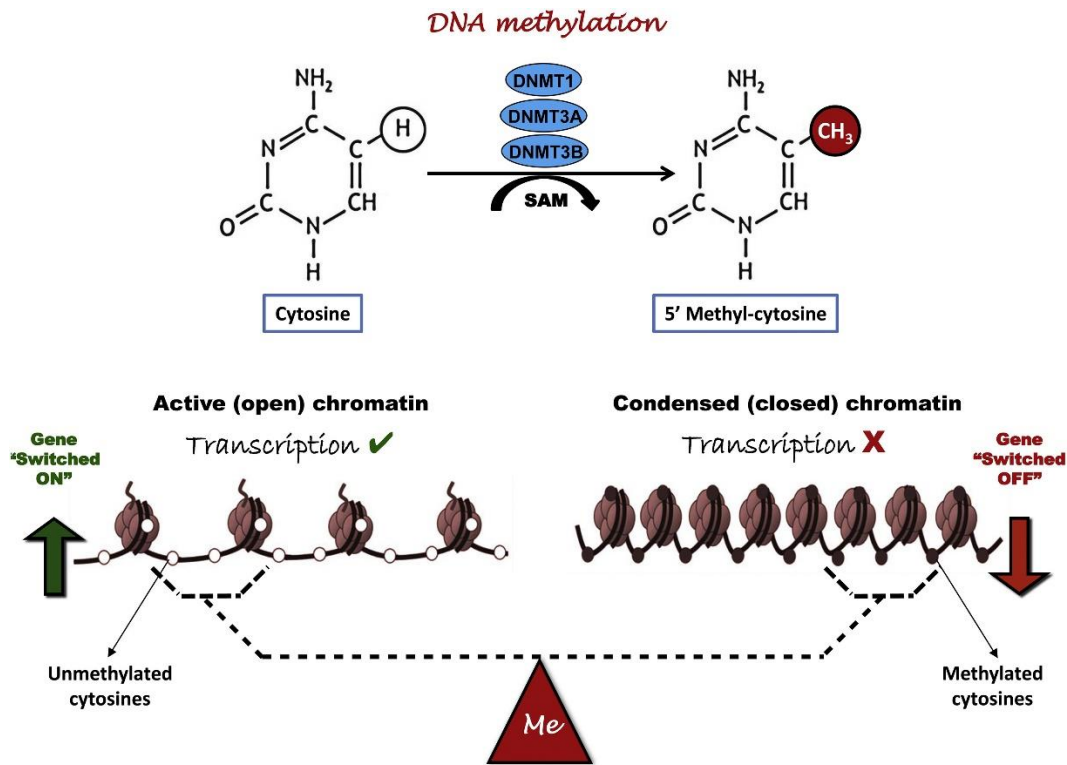
Figure 3: Schematic overview over the establishment and function of m⁵dC in DNA. Taken from *Agrawal et al.*[117]

1.1.2.2.1 Induction of passive demethylation in cancer treatment

Methylation of dC in DNA, forming the epigenetic base m^{5}dC, is an important mechanism for silencing the expression of genes.[93] Cancers have gene-specific increases in methylation (hypermethylation) and global undermethylation (hypomethylation).[114, 118] Furthermore, hypermethylation of CpG islands[119] can lead to cancer formation, because normally expressed genes, *e.g.* tumor suppressor genes, get repressed.[98] Examples for these are *serine protein inhibitors* (SERPINs) involved in gastric cancer.[120] But hypomethylation of CpG islands of oncogenes involved in breast and ovarian cancer can also lead to cancer formation.[121] It has been shown, that in acute myeloid leukemia (AML) and myelodysplastic syndromes (MDS) a great number of genes are hypermethylated.[122-124] Since the absence of maintenance methylation leads to loss of methylation patterns through dilution during replication, targeting of maintenance methylation in leukemia by administration of specific drugs is a common treatment strategy.

MDS comprise various kinds of cancers, in which maturation and development of blood cells in the bone marrow is impaired.[125] Patients show symptoms of tiredness, shortness of breath, easy bruising and bleeding and increased risk of infection. Some types may develop into AML,[125] which is a malignant cancer characterized by clonal proliferation[126-128] of myeloid progenitor cells that build up in the bone marrow and blood,[129] and therefore development of cytopenia. It occurs mostly in elderly

patients[130] with an average on-set of 67 years, and males are more often affected than females. AML patients show the same symptoms as patients with MDS.[129] If it is not treated, AML shows rapid progression and is typically fatal within weeks or months.[129] The disease comprises various subtypes, for which the outcomes and treatments may vary.[129] The typical treatment is chemotherapy aimed at inducing remission. Next treatment options are additional chemotherapy, radiation therapy or a stem cell transplant.[128-129] Genetic mutation analysis can help determine the probability of survival and guide the therapy.[128]

Two of the pharmaceuticals in use for treatment of AML and MDS are 5-azacytidine (Azacitidine, AzaC) and the corresponding DNA analogue 5-aza-2'-deoxycytidine (Decitabine, AzadC),[131-134] because they are known to lower the levels of m⁵dC in cells.[134] In addition, treatment with AzadC or AzaC causes various other changes in cells, including activation of silent genes,[135-136] decondensation of chromatin,[137] and alteration of DNA replication timing.[138] AzaC was furthermore shown to induce apoptotic cell death.[139-141] Clinical usage of AzadC is mostly at low doses.[142] The two compounds do not necessarily show the same clinical results. Depending on the cancer type the drug is administered to, the two azacytidine analogues show different resistance profiles.[143] For the DNA analogue, it has for example been shown, that the *dCTP pyrophosphatase 1* (DCTPP1) and dUTPase enzymes are involved in therapeutic effects, because the deaminated AzadUMP might inhibit *thymidylate synthase* (TS).[144] Nevertheless, both compounds create a hypomethylated state, in which silenced tumor suppressor genes become re-activated.[145-146] Their mechanism of action is complex and not yet fully understood. Both compounds are incorporated into DNA[147-148] as dC analogues, where they form a covalent inhibitory adduct with the DNMTs[149] according to the mechanism shown in **Scheme 1**.
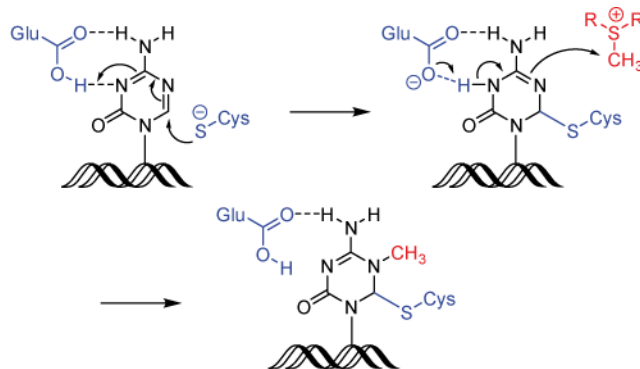
Scheme 1: Proposed mechanism of action of 5-Aza-(2'-deoxy)-cytidine. The blue components are part of the active site of the methyltransferases. The active part of the SAM cofactor is depicted in red.

Methylation of C5 involves addition of a helper nucleophile to the C6-position, followed by electrophilic methylation with SAM and final elimination of the helper nucleophile. Due to the presence of a nitrogen atom at position C5 in AzaC/AzadC, this last elimination step is blocked, which leads to irreversible inhibition of the DNA methyltransferases DNMT1 and DNMT3a/b.[150-153]

Both compounds are rather unstable with half-life times between 3.5 – 21 h depending on conditions, because addition of water to the N5-C6 double bond can compromise the ring structure and hydrolyze the base (see **Scheme 2**).[154-155]
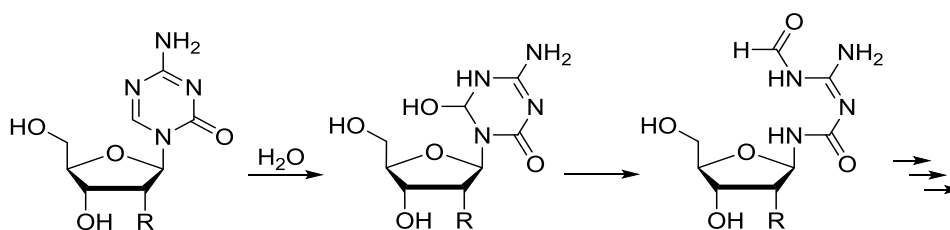
Scheme 2: Depiction of the main hydrolysis pathway of AzadC and AzaC.

This hydrolysis makes the analysis of their mode of action difficult.[156-157] Determination of the extent of Aza(d)C integration into DNA is particularly challenging. Feeding of radioactive AzaC/AzadC for quantitation is only of limited use,[158] because one cannot distinguish degraded from intact material. Several groups[159-160] including our own[161] developed methods for chemically stabilizing the azacytidines in DNA and RNA by hydrogenation with NaBH$_4$ and thereby reduction of the double bond between N5 and C6.

Cells also contain another DNMT family member, DNMT2, which is known to act on cytidine in RNA.[162] In addition, the *NOL1/NOP2/sun domain* (NSUN) enzyme family has been discovered to methylate cytidines in RNA. In particular, it was shown that NSUN2[163] and NSUN6[164] mainly methylate cytidines in tRNA, whereas NSUN4[165] acts on rRNA. Since the DNMTs acting on DNA are known to be affected by Aza(d)C, another possible mechanism of action for AzaC is trapping of the DNMT2 enzyme after incorporation of the nucleoside into RNA. Recently it has been shown by *Aimiuwu et al.,*[166] that AzaC affects the *ribonucleotide reductase* (RNR), because incorporation of the nucleoside into the mRNA of the RRM2-subunit of this enzyme leads to attenuated mRNA stability.

Resistances towards azacytidines

Hepatotropic cancers have never shown strong effects upon azacytidine treatment. The underlying reason for this resistance is increased expression of *cytidine deaminase* (CDA) in liver cells, which deaminates the drugs to unreactive (d)U nucleosides.[167] The major known mechanism of resistance to cytidine nucleoside analogues results from lack of incorporation into DNA[168] for example through *deoxycytidine kinase* (DCK) deficiency as shown *in vitro,* and was reported to be related to *in vivo* resistance in some patients[168-169] and in a rat model.[170] AzaC, however, does not need DCK. Its incorporation is dependent on *uridine-cytidine kinase* (UCK).[171] Indeed, a recent study has shown that some patients can respond to AzadC after showing clinical resistance to AzaC.[172] *Qin et al.*[143] report, that resistance towards AzadC differs depending on the cell line, but the most resistant cells showed a combination of low expression of DCK, high expression of *human equilibrative nucleoside transporter-1*

and -2 (hENT-1, -2), as well as high expression of CDA. Transfection of wt DCK can restore the sensitivity in HL60 cells. Furthermore, they observed that AzadC-treatment induced histone H2A.X phosphorylation, a marker of double strand DNA breaks, and increased the rate of homologous recombination repair, which might give rise to loss of heterozygosity of DCK and resistance to AzadC. Interestingly, resistance to AzadC was, in contrast to previous predictions,[151] unrelated[143] to DNMT levels and long interspersed element (LINE) methylation, a marker of global DNA methylation. Previously, increased resistance towards AzadC was described in Dnmt1$^{-/-}$ mESCs.[151] A different group[173] found out, that naïve Dnmt1$^{-/-}$, as well as Dnmt3a$^{-/-}$ and Dnmt3b$^{-/-}$ mESCs were slightly more resistant (about 4-fold) against AzadC than wt cells, but Dnmt3a$^{-/-}$/Dnmt3b$^{-/-}$ double *knockout* mESCs show strongly increased (200-fold) resistance towards AzadC compared to wt and DNMT single *knockout* cell lines. They furthermore report significantly decreased apoptosis in Dnmt3a$^{-/-}$/Dnmt3b$^{-/-}$ double *knockout* mESCs and therefore a role of these *de novo* methyltransferases in induction of apoptosis upon treatment with AzadC. Differentiating mESCs show hypersensitivity towards AzadC and increased apoptosis, which might be caused by the higher expression of DNMT3a and DNMT3b due to epigenetic reprogramming. Overall, clinical outcome of azacytidine treatment depends on various factors and analysis of genetic variables and protein expression, and the extent of azacytidine incorporation could enable early identification of suitable treatment options.

1.1.2.2.2 Active demethylation of m^{5}dC

Dynamics of the dC modification levels have been described in embryogenesis of mice. Sexual reproduction involves the inheritance of two genome copies, one each from a pair of parental organisms. In higher eukaryotes, this process takes place when one sperm cell enters an oocyte. After this event, the zygote contains the genetic information of both parents and becomes the origin of all cells of the developing organism. Although embryonic development entails progression through a seamless sequence of events, for ease of description distinct stages are usually identified by a snapshot view of the process, with some notable differences between evolutionary distant mammalian species (see **Figure 4**). In mice,[174] the first three cell divisions are symmetrical giving the two-cell, four-cell and eight-cell embryo. This then gets compacted to form the early eight-cell morula. In this period of embryonic development, all cells have high developmental potential and can give rise to all cell types that are present in the final organism. This capability to differentiate into all of the three primary germ cell layers as well as extraembryonic tissues is called naïve pluripotency. Subsequent asymmetric cell divisions form the 16-cell morula and the 32-cell morula, in which the cells are starting to segregate into different lineages. Thereby, they progressively lose the ability to develop into specific tissues, undergoing changes that prepare them for the development of the mesoderm, ectoderm and endoderm. The cells are ,primed' pluripotent and the *priming* procedure involves *e.g.* establishment

of DNA methylation and changes in histone modification patterns. Further asymmetric cell divisions will finally develop the early blastocyst that matures to the late blastocyst by losing the *zona pelucida*. This stage of the embryo is able to implant into the uterus. The cells of the mouse post-implantation epiblast are fully primed at around day 7.5 after fertilization (embryonic day, E) and have lost the potential to generate cells of the germ line.
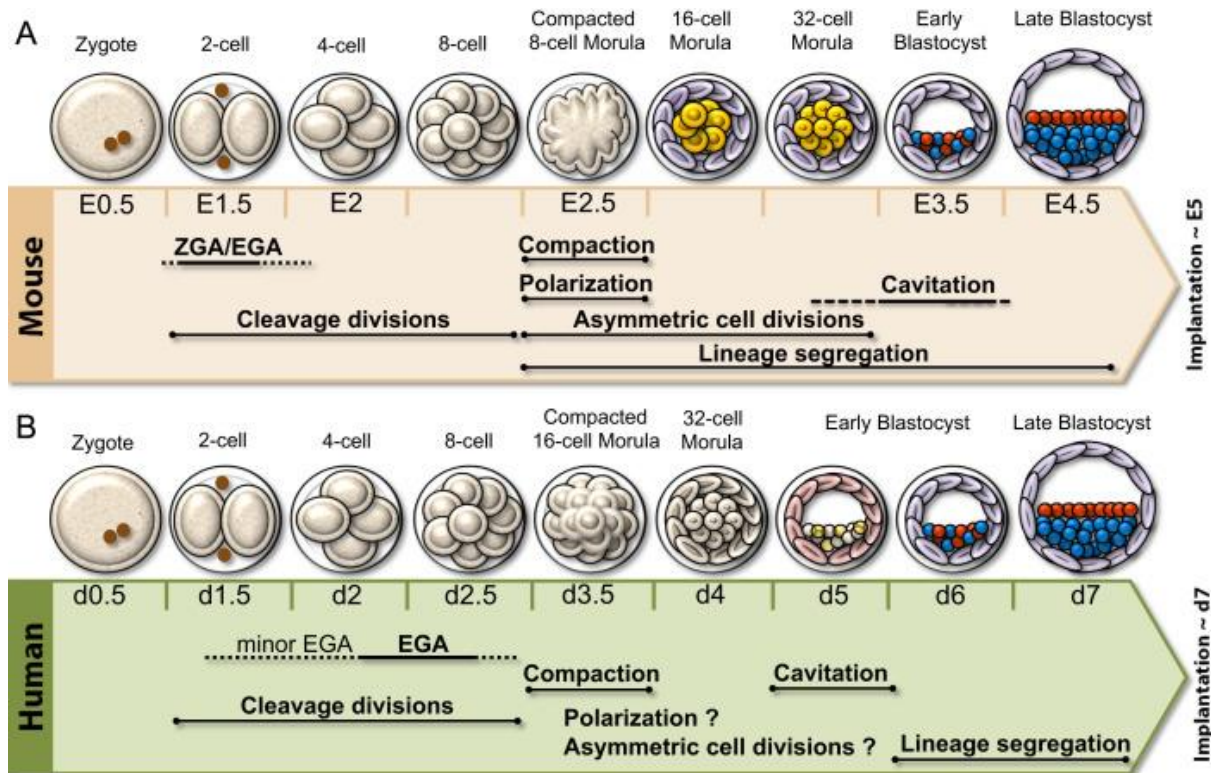
Figure 4: Comparison of the mouse and the human embryonal development. A) Depiction of the mouse blastocyst formation, B) depiction of the human blastocyst development. Taken from *Sozen et al.*[174] 1
ZGA: zygotic genome activation, EGA: embryonic genome activation, E/d: embryonic day; white cell: undifferentiated, yellow: OCT4-expressing ICM precursor, red: GATA6-expressing ICM precursor, blue: NANOG-expressing EPI precursor, light blue: CDX2-expressing TE precursor, light red: partially CDX2-expressing cell, light yellow: partially OCT-4 expressing cell; ICM: inner cell mass; OCT-4, GATA6, NANOG, CDX2: pluripotency marker proteins.

To date, it is not fully investigated how level changes occur. Since the DNA was found to contain hmdC,[45, 175-176] 5-formyl-dC (fdC)[177-178] and 5-carboxy-dC (cadC),[178] those modifications are proposed to be intermediates of a demethylation process.[179] They can be generated by *Ten-eleven translocation* (TET) enzymes, which utilize α-ketoglutarate, oxygen and Fe^{II} to oxidize m^5dC.[178, 180] The modification hmdC is dynamically regulated during differentiation of mouse (m)ESCs[181] and involved in their transcriptional regulation.[182] Specific functions of fdC and cadC are not yet fully elucidated, but they were implicated to have impact on the structure of the DNA helix[183] and the transcription rate.[184] For fdC, stable,[185] or at least semi-stable levels[186] in DNA have been described and it is specifically

---

1 Reprinted from Developmental Biology, 395/1, Berna Sozen, Alp Can, Necdet Demir, Cell fate regulation during preimplantation development: A view of adhesion-linked molecular interactions, 73-83, Copyright 2014, with permission from Elsevier.

enriched in enhancer and intragenic regions in mESCs.[187] For hmdC, it has furthermore been discovered, that it occurs to high levels[176, 188] in the brain, where it has specific functions.[189-190] It comprises 0.6% of nucleotides in Purkinje and 0.2% in granule cells.[176]

The paternal DNA in mice rapidly loses m^5dC after zygote formation by TET-mediated oxidation[191-193] and the loss is not mediated by *Thymidine DNA Glycosylase* (TDG),[194] whereas the maternal DNA methylation gets passively diluted by replicative DNA synthesis.[195] Some methylation on the DNA of the zygote does, however, not disappear. Specific proteins encoded by the *PGC7/Stella* gene were discovered to protect methylated regions from demethylating events.[196] Methylation in these areas is necessary for normal embryogenesis and the protective proteins are expressed between fertilization and the two-cell stage.[196-198] Furthermore, starting at embryonic day 6.5 after fertilization (E6.5) of the mouse embryogenesis, a second erasure of global methylation takes place in PGCs in two waves (see **Figure 5**). First, maintenance and *de novo* methylation are disabled by depletion of recruiting and targeting enzymes. In early PGCs, CpG island methylation is comparable to that of the epiblast and other somatic tissues, but not like the oocyte, the ICM or other cells of the preimplantation embryo.[199-200] To establish full developmental potential in gametes, a second wave of demethylation occurs around E11.5 additionally affecting imprinted loci, genes in the X chromosome and germ line specific genes.
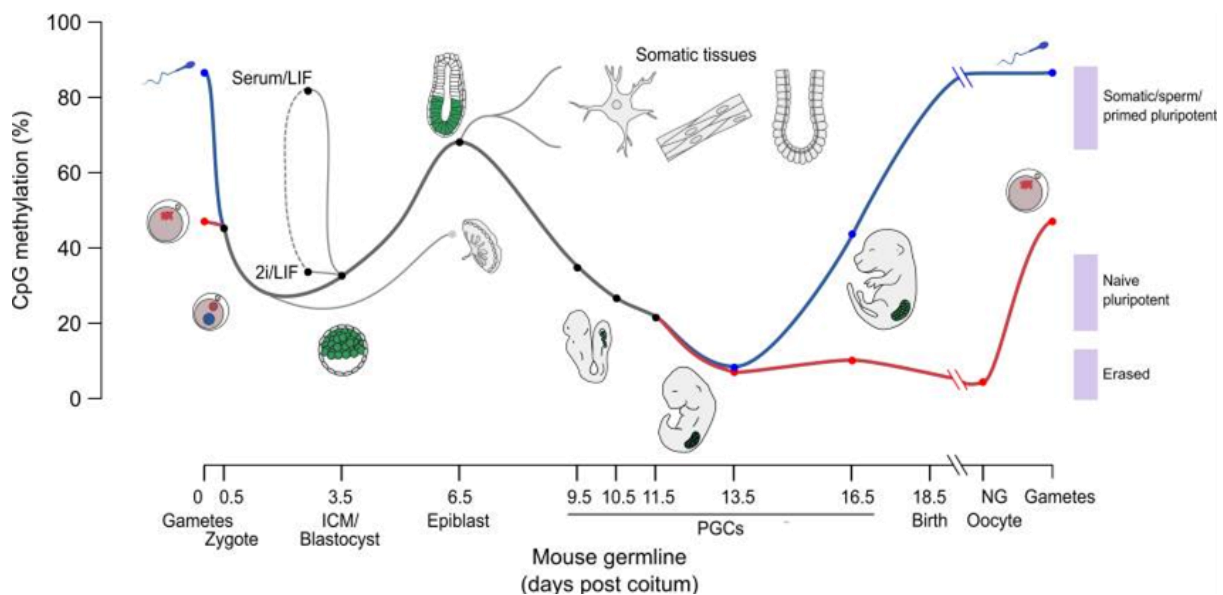
Figure 5: Schematic overview of the correlations between CpG methylation, culturing conditions of mESCs, and the embryonic development of the mouse. Taken from *Lee et al*.[201]

In the paternal genome of mouse zygotes, *elongator protein complex 3* (ELP3) was furthermore identified to be necessary for demethylation.[202] This protein could theoretically attack the methyl group of m^5dC directly *via* its Fe-S radical SAM domain and therefore conduct C-C bond cleavage. *In vitro* studies show, that some DNA methyltransferases are also capable of removing methyl groups,[203]

hydroxymethyl groups[204] and even carboxyl groups[205] in a C-C bond cleaving mechanism in absence of SAM. A decarboxylation reaction was additionally reported in mESCs.[206] Active demethylation through NER involving *growth arrest and DNA damage 45* (GADD45) and *Xeroderma pigmentosum complementation group G* (XPG) was found in *Xenopus*[207] and was controversially discussed,[208] but later confirmed to act in locus specific demethylation.[209-220] In addition, GADD45 was shown to facilitate active demethylation through deamination[221] or oxidation *via* TET1 and TDG.[222-223] In plants, where no TET enzymes are found, active demethylation by direct base excision of 5-methylcytosine was reported.[224-225] However, in mammals, homologues of the responsible proteins do not exist. Involvement of BER in removal of m^{5}dC through its deaminated and oxidized species was furthermore reported in PGCs,[226] but for acute and global demethylation, *e.g.* to reactivate silenced genes, BER is not suitable and other mechanisms are likely (see **Figure 6**):

a) passive demethylation by dilution of the modification during replication,

b) active demethylation by

    I) direct excision of the methyl group by C-C bond cleavage (due to the strong covalent bond between two carbons is it rather unlikely, although ELP3 might use an Fe-S radical mechanism[202]),

    II) direct excision of methylcytosine (confirmed in plants,[224] but there are no protein homologues in mammals),

    III) oxidation of the methyl group to hmdC, fdC or cadC and reestablishment of dC by C-C bond cleavage *via* dehydroxymethylation, deformylation[179] or decarboxylation,[206]

    IV) oxidation of the methyl group to hmdC, fdC or cadC and excision of the base in MMR, BER[180, 227] with further processing of the formed AP site, NER or combinations thereof,[228]

    V) deamination of the 4-amino group of the dC derivatives and excision/processing of the newly-formed dU derivatives.[229-231]

It is likely that most of these pathways can and will take place in various organisms or tissues, but might be gene-specific or depending on the developmental status of the organism.
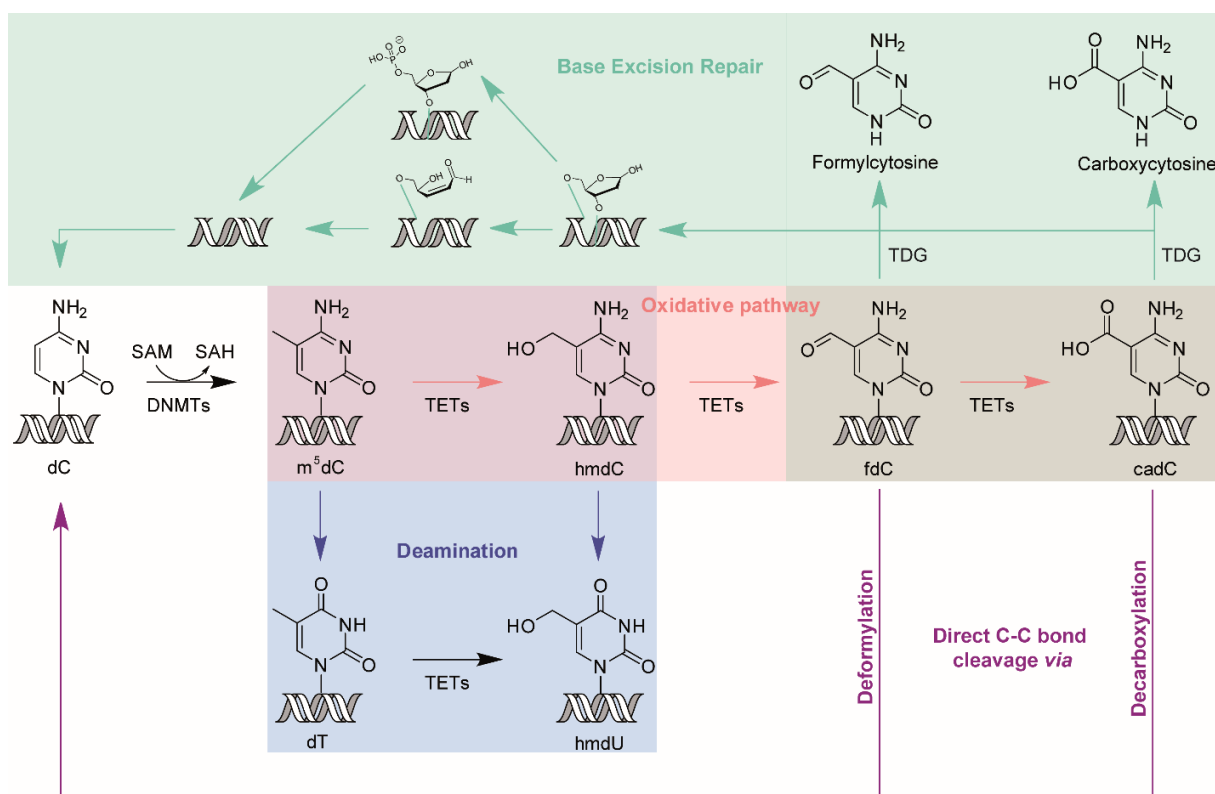
Figure 6: Overview over the possible pathways for formation and erasure of 2'-deoxycytidine derivatives.

Glycosylases involved in the removal of cytosine derivatives

As described in section **1.1.1**, BER is a process to remove unwanted DNA modifications utilizing glycosylases. Their role in removing enzymatic products of m[5]dC as a means for epigenetics has however not been fully elucidated. One glycosylase known to act on a broad variety of substrates that can be generated from m[5]dC is TDG. It was first described to detect dT:dG mismatches and counteract the mutagenic effect of spontaneous m[5]dC deamination.[232] The major function of TDG lies in excision of fdC and cadC[180, 227, 233] and it is therefore involved in active demethylation of m[5]dC through these oxidized species. TDG can furthermore process 5-hydroxymethyl-dU (hmdU)[230-231] and 5-formyl-dU (fdU)[234-235] opposite of dG. TDG was furthermore reported to be the glycosylase whose ablation leads to the most severe phenotype in mice and therefore to be essential for embryonic development in mouse.[231, 236] This importance of the enzyme lies not in global early zygotic demethylation,[194] but rather site-specific TET-TDG directed demethylation[237-240] exhibiting tissue-specific patterns.[187] Later, excision of m[5]dC in hemi-methylated DNA was also described *in vitro*.[241] The m[5]dC glycosylase activity was then associated with the related *mismatch-specific DNA N-glycosylase Methyl-CpG-binding domain protein 4* (MBD4). Nevertheless, the activity was still considered weak and therefore doubted to occur *in vivo*.[230, 242-244] TDG and MBD4 both process various dG-mispaired base lesions, *e.g.* thymine and uracil.[245-246] MBD4 was also shown to act on fdU.[234] Since MBD4 excises dT, that stems from

deamination of m⁵dC, it might be involved in active demethylation processes, although depletion of the enzyme only leads to moderate increase in dC --> dT transversion mutations and no developmental defects.[247-248] MBD4 interacts with *MutL homologue 1* (MLH1), a protein of the post-replicative MMR system.[249-250]

Three different NEIL enzymes form another class of glycosylases working on oxidized dC species. Of these, 1 and 2 express glycosylase activity on oxidized pyrimidines,[251-253] but also on 8-oxodG in DNA bubble structures,[254] whereas NEIL3 shows no activity towards 8-oxodG.[255] *Neil1* gene inactivation in mice exhibits a phenotype similar to metabolic syndrome and increased damage of mitochondrial DNA.[256] NEIL enzymes possess an additional AP lyase function in excising the DNA strand 5' and 3' to the nucleoside and can therefore perform β- and δ-elimination reactions on the sugar.[257] The 3'-phosphate then gets processed by PNK.[37-38] NEIL enzymes are thought to release other glycosylases from the AP sites they generate and to be necessary for the formation of a so-called BERosome,[35-36] in which all reactions of the BER process can take place in an orchestrated fashion. There is a complex relationship between TDG and the NEIL1 and 2 enzymes. The NEILs can partially compensate for loss of TDG, but they show no binding affinity and activity towards fdC or cadC *in vitro*.[258-259] The NEIL enzymes were shown to be involved in substrate turnover of TDG,[259] like APE1,[16, 260] releasing the glycosylase from the tight binding to the AP site and processing it with β,δ-elimination.[261] Redundancy of the functions of NEIL1 and 2 with those of APE1 could explain the unexpectedly mild effects of APE1 *knockdown* on fdC and cadC in HeLa cell extracts and cultured cells, as well as *Xenopus* embryos.[259]

In mice, spontaneous deamination of dC to dU is processed mostly by the *uracil DNA glycosylases* (UNG)[262-263] 1 and 2. UNG1 is the mitochondrial isoform protecting mitochondrial DNA from mutagenesis.[263] Alternative promoter usage and splicing generates the UNG2 isoform, which is located in the nucleus, where it acts on genomic DNA.[263] Additionally, spontaneous deamination to dU is processed by *single-strand selective monofunctional uracil DNA glycosylase* (SMUG1).[264] In humans, SMUG1 was also found to be a back-up enzyme in addition to hUNG2.[265] Nuclear UNG2 is targeted to DNA synthesis sites by *proliferating cell nuclear antigen* (PCNA) and *replication protein A* (RPA), where it excises misincorporated dU.[266] SMUG1 can furthermore process hmdU,[229, 267-269] which is a product of deamination of hmdC or TET-oxidation of dT, as well as further oxidized species like fdU.[229, 270-272] The enzyme acts preferred on ssDNA,[264] but seems to act also on dsDNA in humans.[265] The origin of dU lesions determines its mutagenic potential. If dU derivatives stem from TET-mediated oxidation, they can still be read as dT analogues and base pair with dA, but if they originate from deamination events on dC derivatives,[273] a XdU:dG mismatch will be generated, which is more efficiently processed than XdU:dA mismatches.[7] Transversion mutations of dA --> dC and dT --> dG are elevated when AP endonucleases are disrupted and they additionally depend on expression levels of UNG. Furthermore, these mutations occur frequently with high cellular dU

triphosphate (dUTP) levels, which can be counteracted with overexpression of dUTP pyrophosphatase.[22] An isotope tracing study in mESCs[274] shows that hmdU appears only labeled, when dT and not m^{5}dC is labeled, suggesting a formation of hmdU through oxidation of dT. This explains a correlation between hmdU and hmdC in mice reported by *Alsøe et al.*,[7] since both modifications are affected by TET activity. SMUG1 also exhibits higher activity in the brain in comparison with other organs.[7]

Deamination

Active demethylation is expected to not only happen *via* an oxidative pathway, but might potentially take place through enzymatic deamination of m^{5}dC to dT and further processing of this nucleotide by a repair mechanism. Deaminases transform (d)C to (d)U by replacement of the exocyclic amino group with a hydroxyl group. There are two types of cytidine deaminases in the soluble pool. CDA[275] acts on nucleosides (C or dC), whereas *deoxycytidylate deaminase* (DCTD) is active on the nucleotide 2'-deoxycytidine monophosphate (dCMP). DCTD is a zinc dependent enzyme, which can be allosterically activated through binding of dCTP and inhibited by dTTP.[276] Human CDA is a tetrameric enzyme of 15 kDa per subunit.[277] Investigation of those two enzymes in leukemic cells revealed that upon feeding of radioactively labeled m^{5}dC only labeled dT, not labeled m^{5}dC was found in the genome.[278-279] The observation that only labeled dT, but no m^{5}dC gets incorporated might be caused by the substrate specificity of *nucleoside monophosphate kinase*, which is not active on m^{5}dC monophosphate.[280] The incorporation was analyzed in *knockout* conditions and revealed, that labeled dT was only found, when the CDA/thymidine kinase (TK), but not the DCTD/TK pathway was functional. Another study also reported incorporation of tritiated m^{5}dC in *Chinese hamster ovary* cells with low m^{5}dCMP deaminase activity.[281] Deamination of m^{5}dC is therefore considered to take place on the nucleoside level. Some cancer cell lines overexpress CDA, which causes increased levels of mutagenic dU species upon treatment with hmdC and fdC.[282] This observation poses a potential targeting strategy for resistant cancers. Decreased CDA levels have been implicated in Bloom Syndrome, a disease with chromosome instability, because elevated dC and dCTP pools inhibit the DNA repair enzyme *PAR polymerase-1* (PARP-1).[283-285] In yeast, *knockout* of the DCTD homologue Dcd1 leads to elevated dCTP and decreased dTTP pools and replication fork stalling and collapse.[286] Therefore, ablation of the soluble pool deaminases seems to have similar effects between different species.

In DNA, the 24 kDa *activation-induced cytidine deaminase* (AID, AICDA) can convert dC to dU.[287] This is a desired process for the generation of antibody diversity through:

I)   Somatic Hypermutation (SHM)[1] – minimal mutation of antibody genes generating a library of antibody variants with diverse affinity for a particular antigens.

II) Class Switch Recombination (CSR)[1] - B cells change the expression of their immunoglobulin (Ig).

III) Immunoglobulin gene Conversion (IGC)[288] - exchange of antibody genes by homologous recombination.

The molecular basis for the generation of antibodies is DNA editing. After replication, it leads to formation of a dT:dA basepair in place of dC:dG due to the recognition of dU as dT. Especially dCs in the WRCY context (W=adenosine or thymidine, R=purine, C=cytidine, Y=pyrimidine) were shown to be edited. Cytidine deaminases possess a typical His/Cys-X-Glu-X$_{25-31}$-Pro-Cys-X$_{2-4}$-Cys motif.[289-290] Another class of cytidine deaminases is the *apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like* (APOBEC) family, which is evolutionarily conserved. These enzymes perform mRNA C-to-U editing, which generates protein diversity, and deamination on DNA. The zinc dependent catalytic domain of APOBEC like proteins is located in the N-terminus, while the C-terminus contains a pseudocatalytic domain that is not present in the family members 3A, C and H. Some APOBEC3 enzymes demonstrate antiviral activities[291-293] and misregulation of these enzymes is a major source of mutation in numerous cancer types.[291, 294] APOBEC3 enzymes furthermore act on ssDNA,[295-296] with at least three bases 5', of which the closest is dT, and one base 3' of the dC.[297] A deaminase activity of AID and APOBEC1 on m^{5}dC has previously been reported in DNA *in vitro*,[298] but so far there is no evidence for the activity *in vivo*. Of all family members, the APOBEC3A enzyme has most often been shown to act as a DNA cytidine deaminase that converts m^{5}dC to dT.[299-301] AID and APOBEC were however shown to deaminate hmdC to hmdU *in vivo*,[231, 302] but not *in vitro*.[303-304] Direct action of deaminases on m^{5}dC or other cytidine derivatives in the DNA is therefore not fully investigated and might pose an alternative mechanism for their repair.

1.1.2.2.3   Genomic Imprinting

Diploid organisms can exhibit expression of genes from one allele, while the other one is silenced. This mono-allelic expression can differ between tissues or cells at different developmental stages. Certain genes are differentially expressed depending on the origin of an inherited allele – from the mother, maternal, or from the father, paternal. This phenomenon is termed genomic imprinting. The imprinted genes are necessary to establish normal embryogenesis and postnatal development. Embryos developed with only two maternal alleles, so-called parthenogenotes (gynogenotes), or two paternal alleles, so-called androgenotes, are not viable and die shortly after implantation.[305-306]

First evidence for genomic imprinting came from chromosome mapping studies, a process to determine the location of a certain gene. In more detail, the researchers duplicated the genes of interest and noticed that some genes had different outcomes, when they originated from one allele or

the other. Discovery of the maternally expressed imprinted genes *insulin-like growth factor receptor 2* (*Igfr2*),[307] the paternally expressed imprinted gene *insulin-like growth factor 2* (*Igf2*)[308-309] and the maternally expressed imprinted *H19* gene[310] finally confirmed these observations.

Underlying cause for differential gene expression is DNA methylation. Differentially methylated regions (DMRs) are first established in the germ line. DMRs comprise certain genes that are always methylated in the haploid maternal germ line, while other genes are only methylated in the paternal germ line. Those gametic DMRs (gDMRs) represent inherited DNA methylation and are installed by DNMT3 enzymes.[311-312] They furthermore depend strongly on DNMT3l[313-314] and stay present throughout the life-time of the organism. Only in primordial germ cells (PGCs) the DNA gets globally demethylated including all imprinted methylation.[200] The gDMRs are subsequently reestablished during gametogenesis.[201] Methylation at gDMRs is essential for the formation of further methylation patterns, so-called somatic DMRs (sDMRs), at the implantation stage. Recent investigations report dynamics of sDMRs during tissue development and different cell-types of the adult mouse brain.[315] Furthermore, DNMT1 is essential for maintenance of DMRs in somatic cells.[316-317]

The imprinted *H19* gene comprises an unusually long noncoding RNA (lncRNA), which was shown to cluster close to the *Igf2* gene. This finding lead to the hypothesis, that imprinted genes occur in clusters and was the key discovery for understanding the mechanism controlling genomic imprinting in mammals. To date about 150 different imprinted genes have been identified in mice, of which most cluster into 16 well defined regions.[318] Each cluster contains one gametic DMR, which functions as imprinting control region (ICR) that determines the expression state of the somatic DMRs of the cluster. The activation or silencing of those genes can be conducted in *cis*, utilizing so far not fully elucidated regulatory mechanisms including DNA methylation, histone modifications and ncRNA. Maternally methylated ICRs are well understood, occur all within genes and comprise promoters. Additionally, they often possess an antisense, *cis*-acting ncRNA domain, whose removal will result in bi-allelic expression of normally mono-allelically expressed genes.[319] The ncRNA is therefore considered as an essential element for the repression of paternally inherited genes. Paternally methylated gDMRs comprise fewer CpGs than maternally methylated ones.[320] Furthermore, they are intergenic and do not possess specific features in regards to promoters, but the respective cluster can also contain ncRNA. Two different imprinted gene clusters regulated by maternally imprinted or paternally imprinted ncRNA are depicted in **Figure 7**. The *Dlk1-Dio3* locus is one of few imprinted domains that are controlled by a paternally methylated ICR both in *cis*[321] and *trans.*[322-323]

A *Dlk1-Dio3* Domain
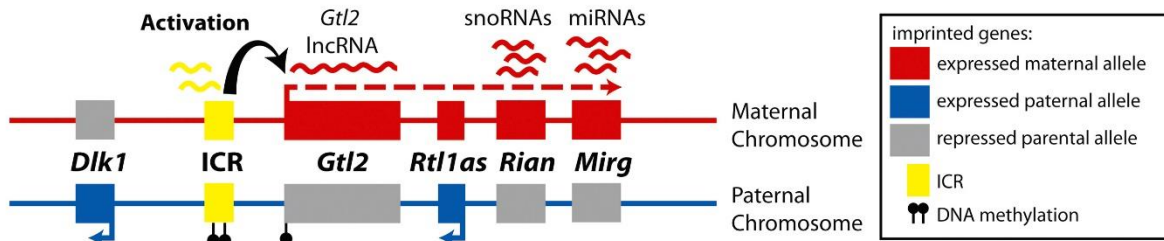
B *Snrpn* Domain
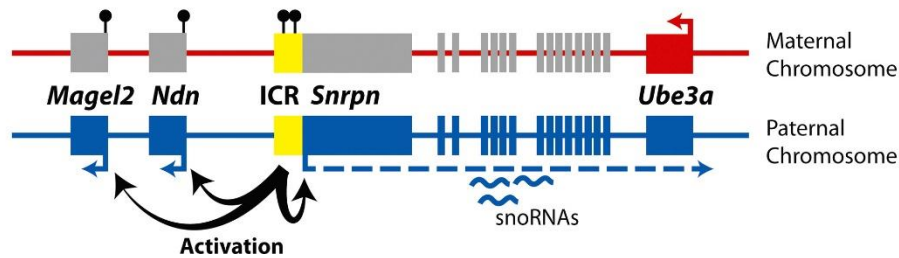
Figure 7: Genome organizations of two different imprinted mouse clusters. A) Illustration of the paternally imprinted *Dlk1-Dio3 domain* cluster containing the *Rtl1as* ncRNA. B) Depiction of the maternally imprinted *Snrpn* cluster. Taken from *Sanli et al.*[324] (https://doi.org/10.1016/j.biocel.2015.04.004).

Imprinted genes are very susceptible to loss-of-function or epimutations because of their mono-allelic nature. Therefore, loss-of-imprinting (LOI) can have mild to severe effects including abnormalities of the development, behavior or organ function, as well as cancer.[44] In humans, several diseases and malignancies have been attributed to LOI.[325-326] One example is the Prader-Willi syndrome (PWS), a disease that is caused by partial (65-75% of cases) or complete deletion of the paternal chromosome 15 and replacement with a second copy of the maternal chromosome (20-30% of cases) or aberrant methylation in the ICR (1-3%).[327] Several imprinted genes are responsible for this effect, but as a major factor the absence of a small nucleolar organizing RNA gene, *SNORD116,* was identified.[327] Ablation, mutation or imprinting defects of the maternal genes of the same chromosomal region are the causes for a disease called Angelman syndrome (AS), because maternal expression and paternal imprinting of the *UBE3A* gene encoding E3 ubiquitin ligase is lost.[328] Duplication of this region has been associated with autistic spectrum disorder, found in >1-2% of cases.

The Silver-Russell syndrome (SRS) is caused partly by aberrant expression of imprinted genes on chromosome 7. Additionally, the imprinted expression of the *H19* and *Igf2* genes can be disrupted due to maternal duplication of chromosome 11, and the respective ICRs were shown to be hypomethylated in 40-60% of cases. Genetically opposite to SRS is the Beckwith-Wiedemann syndrome (BWS).[329] It has been reported, that there is a higher incidence for BWS and AS after *in vitro* fertilization.[330-332]

1.2   Modifications in ribonucleic acid

Not only DNA contains modifications. Ribonucleic acid (RNA) is an even more modified macromolecule. To date, more than 100 modifications have been discovered with a big variety from small methyl groups to large glycosylations. The modifications are thought to be necessary for establishing the right secondary structure and stability of the RNA to enable their unique features. Although modifications on DNA are rare to occur on adenosines (see section **1.1.2.1**), one of the more abundant modifications in RNA is m^6A. This modification is involved in many processes concerning messenger RNA (mRNA). It was also found to be a reversible modification with the corresponding 'writers' *methyltransferase-like* (METTL)-3 and -14[333-334] and 'erasers', the two demethylases FTO[335-336] and ALKBH5,[337] as well as various 'readers' – the YTH domain family[338] enzymes. In comparison, the homologue of the most important DNA modification m^5dC, namely m^5C, is one of the less abundant modifications, although it is installed by various enzymes[162-165] and was also shown to be oxidized by the TET enzymes to hmC, fC and caC.[339-340] Additionally, oxidation of m^5C to hmC and fC in tRNA can be performed by ALKBH1.[341]

RNA furthermore contains hypermodified nucleosides, which carry modifications that require multiple steps of biosynthesis, like 2-methylthio-N^6-isopentenyl-adenosine (ms^2i^6A) and 2-methylthio-N^6-threonylcarbamoyl-adenosine (ms^2t^6A) (see **Figure 8**).
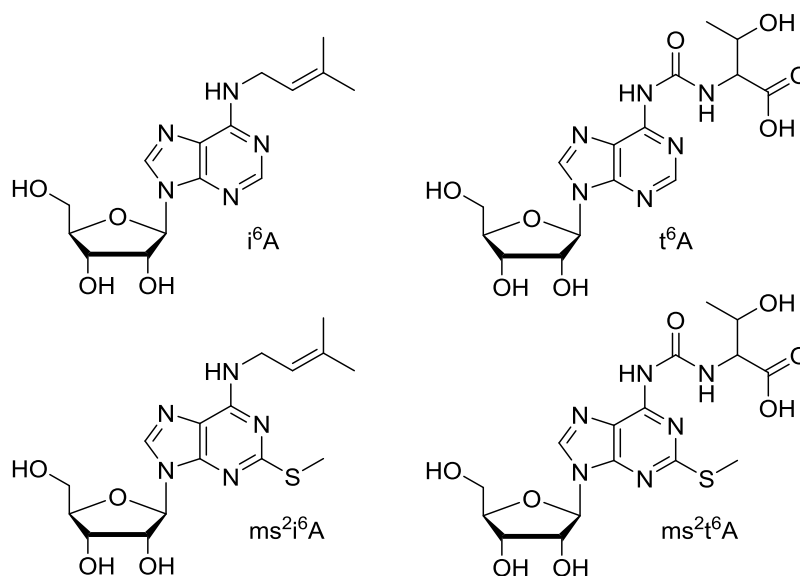
Figure 8: Overview over the structure of some of the hypermodified bases in RNA. Upper row i^6A and t^6A, lower row ms^2i^6A and ms^2t^6A.

The modification N^6-isopentenyl-adenosine (i^6A) was first discovered in tRNA by H. G. ZACHAU in the year 1966.[342] Both isopentenylated modifications are very prominent in prokaryotic tRNA[343] and are especially abundant in tRNA that decode codons with U as the first base. They are usually located in position 37, right next to the anticodon, where they are responsible for proofreading of the codon-

anticodon interaction.[344-345] This effect is especially strong for ms^2i^6A, since it can interact with the base opposite of A36 and increase binding of the tRNA to mRNA and the ribosome in the A-, P- and E-site.[346-347] Therefore, wobble base pairing in the 5'-region is impaired,[345] since the rather weak interaction between A in the anticodon and U in the codon gets increased. This effect is caused mainly by the isopentenyl moiety, but is supported by stacking of the thiomethyl group to the base pair.[348] Furthermore, mismatch bonds of the third base of the codon are enhanced, which enable different base pairing in the wobble position.[346]

In *E. coli*, i^6A is generated by a *tRNA dimethylallyltransferase*, from now on called MiaA.[349] It transfers the isopentenyl group from dimethyl allyl diphosphate (DMAPP) to the N^6-atom of adenosine,[350] releasing pyrophosphate (PP_i) (see **Figure 9**). In a next step, the tRNA methylthiotransferase (MiaB) can synthesize ms^2i^6A by thiolation, and subsequently methylation, of the 2-position of adenosine using a SAM-dependent radical mechanism.
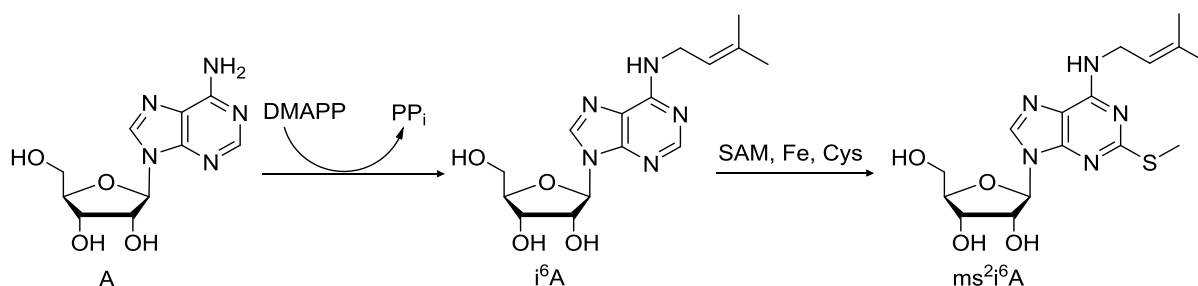
Figure 9: Reaction scheme for the biosynthesis of i^6A and ms^2i^6A; DMAPP: dimethylallylpyrophosphate, PP_i: pyrophosphate, Fe: iron, Cys: cysteine; taken and adapted from *Connolly et al*.[351]

As the human homologue of MiaA, tRNA isopentenyltransferase 1 (TRIT1) was identified.[352] It synthesizes i^6A specifically in A37 of cytosolic and mitochondrial tRNA.[352] The enzyme is thought to have a tumor-suppressing function in human cells.[353] In yeast, the modification is installed by the MiaA homologue Mod5[354] and its absence in position A37 reduces translation efficiency and fidelity.[355] As a homologue of MiaB, the *cyclin-dependent kinase 5 regulatory subunit associated protein 1* (CDK5RAP1) was discovered in humans, which generates ms^2i^6A specifically in mitochondrial tRNA, but also in nuclear and mitochondrial RNA.[356-357] CDK5RAP1 is described as repressor of *cyclin-dependent kinase 5* (CDK5).[357] Without CDK5RAP1, the mitochondrial protein synthesis is significantly impaired and *knockout* mice show strong respiratory defects, *e.g.* myopathy and cardiac dysfunction.[358] In the same publication, a higher sensitivity towards oxidative stress is described for patients with mitochondrial diseases, in which thiomethylated tRNA are not generated. The group of *Zhigao Li*[359] discovered that *knockdown* of CDK5RAP1 through RNA interference in the human cancer cell line MCF-7 leads to cell cycle arrest in the G2/M-phase, apoptosis and generation of reactive oxygen species (ROS). This suggests a new therapy option for cancer through the ROS/JNK signal

transduction pathway. CDK5RAP1 is furthermore suspected to be involved in the formation of the skin disease vitiligo.[360]

Like the isopentenylated bases, N^6-threonyl-carbamoyl-adenosine (t^{6}A) and its derivatives are universally conserved and often found in position A37 of tRNAs, especially in those responsible for ANN codons.[361-366] The bulky side chain also stabilizes the anticodon loop structure by strengthening π-π stacking with adjacent bases. The underlying cause for this phenomenon is intramolecular hydrogen bonding that aligns the atoms in a planar ring. Subsequently, base pairing with U33 is prevented,[367] and tRNA binding to the A-site codon and translocation[368] gets enhanced, thereby helping to maintain the efficiency and accuracy of translation. t^{6}A37 also plays critical roles in aminoacylation of tRNA[369] and prevention of leaky scanning of initiation codons and read-through of stop codons.[370] In mammalian mitochondria, t^{6}A37 is present in mitochondrial (mt-)tRNAs for Ser(AGY), Thr, Asn, Ile, and Lys.[371] To date, however, the biogenesis and functional role of t^{6}A37 in mammalian mt-tRNAs have not been elucidated. Thiomethylation of t^{6}A is generated by *CDK5 regulatory subunit associated protein 1-like* 1 (CDKAL1).[372] The *Cdkal1* gene has been identified as a susceptibility gene for type II diabetes, since it is associated with reduced β-cell function and insulin release.[373]

2. Aim of the research

To date several modifications were found on DNA nucleotides and their role has not yet fully been elucidated. This research is conducted to discover potentially new modifications and investigate their role for the organism. To this end, mass spectrometric method development and corresponding data analysis, as well as biochemical methods were supposed to be applied. A special focus was planned to be acute myeloid leukemia. This disease is frequently treated with a nucleoside analogue, 5-Aza-(2'-deoxy)cytidine, which is incorporated into the DNA and RNA and interferes with the respective methyltransferases. Analysis of modification levels in a standardized protocol could help unravel all mechanisms of function for this drug. Application of the technique with patient samples was supposed to show the effectiveness of the treatment in order to give faster evidence of a resistance and therefore earlier identification of better therapy options.

A second focus was put on the investigation of active demethylation of m^5dC in mouse embryonic stem cells. Current literature describes oxidized derivatives of the nucleoside as intermediates in the major pathway for demethylation. Therefore, we planned to analyze a putative C-C bond cleaving deformylation mechanism as a non-harmful alternative to DNA repair mechanisms. Base excision repair, for example, first cleaves the base moiety of the nucleotide and leaves abasic sites and β-elimination products. Subsequently, the repair enzymes perform incisions in the DNA backbone. As a second potential pathway for the reestablishment of dC in place of m^5dC, we wanted to investigate formylcytosine as a product of the base excision repair. This small molecule, like the other products of base excision repair, does not allow direct quantification due to lack of a labile bond. To this end, extraction and derivatization strategies were planned to be tested and optimized. Finally, this thesis had the aim to uncover the role of deamination on m^5dC to dT as another alternative removal pathway leading to reintroduction of dC into the DNA.

3. Results and Discussion

3.1 Published results

3.1.1 Quantitative LC–MS Provides No Evidence for m6dA or m4dC in the Genome of Mouse Embryonic Stem Cells and Tissues

Sarah Schiffers, Charlotte Ebert, René Rahimoff, Olesea Kosmatchev, Jessica Steinbacher, Alexandra-Viola Bohne, Fabio Spada, Stylianos Michalakis, Jörg Nickelsen, Markus Müller and Thomas Carell

Prologue

After publication of various controversial articles about the existence of m6dA in vertebrate DNA, the sensitive evaluation of this modification was essential. In this project, we investigated the existence of m6dA and m4dC in murine cell culture and tissues. To this end, a triple quadrupole method was developed and gDNA was extracted from murine cells and tissues. Although the established method was verified using gDNA of organisms known to contain the nucleotides of interest, both modifications were not detected in any of the murine material. This result challenges previous findings, which showed low levels of m6dA in murine and human samples. Our data that m6dA does not exist is supported by *Lentini et al.* [92] and *O'Brown et al.*[91] In these two publications, it has been reported that and biased data analysis due to unspecificity of antibodies and artefacts in the LC-MS analysis from bacterial contamination led to overestimations of the m6dA levels in higher eukaryotes.

Author contribution

I performed the method development, gDNA preparation from human and murine sources, as well as gDNA digest and mass spectrometric data analysis. *Charlotte Ebert* and *René Rahimoff* synthesized the nucleosides m6dA and m4dC and their respective isotopologues. *Olesea Kosmatchev* and *Jessica Steinbacher* performed measurements of known DNA modifications. *Dr. Alexandra-Viola Bohne* and *Dr. Stylianos Michalakis* prepared *Synechocystis* and *Chlamydomonas* gDNA and mouse tissues. *Dr. Fabio Spada*, *Prof. Dr. Jörg Nickelsen*, *Dr. Markus Müller* and *Prof. Dr. Thomas Carell* supervised and developed the project. All authors wrote the manuscript.

License

Copy of article with permission of the publishing company; Copyright 2017 *Wiley Company*.

Quantitative LC–MS Provides No Evidence for m⁶dA or m⁴dC in the Genome of Mouse Embryonic Stem Cells and Tissues

Sarah Schiffers, Charlotte Ebert, René Rahimoff, Olesea Kosmatchev, Jessica Steinbacher,
Alexandra-Viola Bohne, Fabio Spada, Stylianos Michalakis, Jörg Nickelsen, Markus Müller, and
*Thomas Carell**

Abstract: *Until recently, it was believed that the genomes of higher organisms contain, in addition to the four canonical DNA bases, only 5-methyl-dC (m⁵dC) as a modified base to control epigenetic processes. In recent years, this view has changed dramatically with the discovery of 5-hydroxymethyl-dC (hmdC), 5-formyl-dC (fdC), and 5-carboxy-dC (cadC) in DNA from stem cells and brain tissue. N⁶-methyldeoxyadenosine (m⁶dA) is the most recent base reported to be present in the genome of various eukaryotic organisms. This base, together with N⁴-methyldeoxycytidine (m⁴dC), was first reported to be a component of bacterial genomes. In this work, we investigated the levels and distribution of these potentially epigenetically relevant DNA bases by using a novel ultrasensitive UHPLC–MS method. We further report quantitative data for m⁵dC, hmdC, fdC, and cadC, but we were unable to detect either m⁴dC or m⁶dA in DNA isolated from mouse embryonic stem cells or brain and liver tissue, which calls into question their epigenetic relevance.*

Figure 1. Depiction of the four canonical DNA bases and the epigenetic DNA marks m⁵dC, hmdC, fdC, and cadC, as well as the bases m⁶dA and m⁴dC together with the synthesized isotopologues.

The genetic material of living organisms is constructed from the four canonical nucleobases dA, dC, dG, and dT, which establish the sequence information that, in multicellular organisms, is stored in the nucleus of every cell (Figure 1). In addition to the canonical bases, the methylated dC base 5-methyldeoxycytidine (m⁵dC) is frequently found.[1] The presence or absence of this base in specific promoter segments determines whether the gene is actively transcribed or silenced.[1] The cell-type-specific distribution of m⁵dC thus determines the identity of a given cell. Recently, 5-hydroxymethyldeoxycytidine (hmdC) was found as a sixth base of

the genetic system[2,3] and in 2011, 5-formyldeoxycytidine (fdC)[4,5] and 5-carboxydeoxycytidine (cadC)[5,6] were also discovered, particularly in DNA isolated from stem cells, but also in brain DNA. It is currently believed that fdC and cadC are intermediates in an active demethylation process that allows cells to change the methylation pattern and hence the activity state of specific genes.[7,8] For fdC, separate epigenetic functions are also envisaged.[9]

While the genomes of bacteria are known to also contain N⁴-methyldeoxycytidine (m⁴dC)[10] and N⁶-methyldeoxyadenosine (m⁶dA),[11] attempts to detect these bases in the DNA of higher organisms have failed until recently.[12–15] m⁶dA has now been found in algae (0.4 mol% m⁶dA/A),[12] fruit flies (0.001%-0.07% m⁶dA/A),[14] and *C. elegans* (0.01%-0.4% m⁶dA/A),[13] and its presence has even been reported in the DNA of vertebrates (0.00009% in *X. laevis*[16] and 0.00019–0.003% of dA in murine cells and tissue[17]). These discoveries, especially concerning the DNA of vertebrates, have spurred a worldwide research interest in unraveling the function of these new bases in human genomic DNA.[18–20]

In this study, we developed an ultrasensitive triple quadrupole mass spectrometry (QQQ-MS) method, which in combination with ultra-high-pressure chromatography (UHPLC) enables m⁴dC and m⁶dA to be searched for and quantified in parallel to the more established new epigenetic DNA marks m⁵dC, hmdC, fdC and cadC.

[*] S. Schiffers, C. Ebert, R. Rahimoff, O. Kosmatchev, J. Steinbacher,
Dr. F. Spada, Dr. M. Müller, Prof. Dr. T. Carell
Center for Integrated Protein Science (CiPSM) at the Department of
Chemistry, LMU München
Butenandtstr. 5–13, 81377 München (Germany)
E-mail: Thomas.Carell@lmu.de
Homepage: http://www.carellgroup.de

Dr. A.-V. Bohne, Prof. Dr. J. Nickelsen
Biocenter of the LMU
Dept. Biologie 1 – Botanik, Molekulare Pflanzenwissenschaften
Grosshaderner Strasse 2–4, 82152 Planegg-Martinsried (Germany)

Dr. S. Michalakis
CiPSᴹ, Department of Pharmacy, LMU München
Pharmacology for Life Sciences
Butenandtstr. 7, 81377 München (Germany)

Supporting information and the ORCID identification number(s) for
the author(s) of this article can be found under:
http://dx.doi.org/10.1002/anie.201700424.

For the quantitative measurements, we first chemically synthesized the two isotopologues of m6dA and m4dC shown in Figure 1 as internal standards for the analytical method. The prepared compounds D3-m6dA and 15N2-m4dC are three and two mass units heavier, respectively, than the natural bases. Despite these molecular-weight differences, they have identical properties during the UHPL chromatography step so that they strictly coelute with their natural counterparts, thus allowing them to enter the mass spectrometer at exactly the same time as the internal standards. The availability of these isotopologues makes the method highly reliable and strictly quantitative. The syntheses of the two compounds, together with all analytical data, are given in the Supporting Information.

We first benchmarked our study with an investigation of genomic DNA isolated from the unicellular green algae *Chlamydomonas reinhardtii* and the cyanobacterium *Synechocystis*. In both cases, DNA was isolated after cell lysis using a standard method (see the Supporting Information). The isolated DNA was subsequently digested with a mixture of three commercially available digestion enzymes (Nuclease S1, Antarctic Phosphatase, and Snake Venom Phosphodiesterase; see the Supporting Information). We next added the isotope-labelled standards D3-m6dA and 15N2-m4dC to the obtained digestion mixture and performed UHPLC-QQQ analysis. For the mass spectrometry detection, we selected fragmentation of the glycosidic bond as the indicative and hence recorded mass transition. This is $m/z = 266.12 \rightarrow 150.08$ for m6dA and $m/z = 269.14 \rightarrow 153.10$ for its isotopologue D3-m6dA. For m4dC, we also used fragmentation of the glycosidic bond, which gives a mass transition of $m/z = 242.11 \rightarrow 126.07$ for the natural compound m4dC and $m/z = 244.11 \rightarrow 128.07$ for its isotopologue 15N2-m4dC (Figure 2 A).

We next modified the reported UHPLC-QQQ method[10] for the simultaneous quantification of m4dC and m6dA, together with the other epigenetically relevant bases m5dC, hmdC, fdC, and cadC. To this end, the UHPLC gradient was fine-tuned to enable full separation of all six compounds. Finally, we measured precise calibration curves for all of the compounds (see Figures S1 and S2 in the Supporting Information). This subsequently allowed exact quantification of all of the discussed epigenetic DNA marks in a given sample (Figure 2 B–D).

Since m4dC and m6dA are well known in bacteria, we first analysed the cyanobacterium *Synechocystis* (PCC6803), and we indeed found both bases (Figure 2 B). The base m6dA was detected at a level of 8.4×10^{-3} per dN and for m4dC we measured a value of 5.9×10^{-3} per dN. The constitutional isomer m5dC and all other dC-derived epigenetic DNA marks were detectable, but were not quantified in this experiment.

Next, we analyzed two different strains of *Chlamydomonas reinhardtii* (CC-3491 and wt 7d+), in which m6dA has just recently been discovered,[12] and the levels of m6dA were determined to be 8.4×10^{-4} per dN for CC-3491 and 6.9×10^{-4} per dN for wt 7d+ (Figure 2 C). This corresponds to about 3000 m6dA bases per *Chlamydomonas* genome (genome size 1.2×10^8), which at 0.7% of the dA is a relatively high number. In both strains, m4dC was not detected, thus showing that this base is unlikely to be a component of the genetic
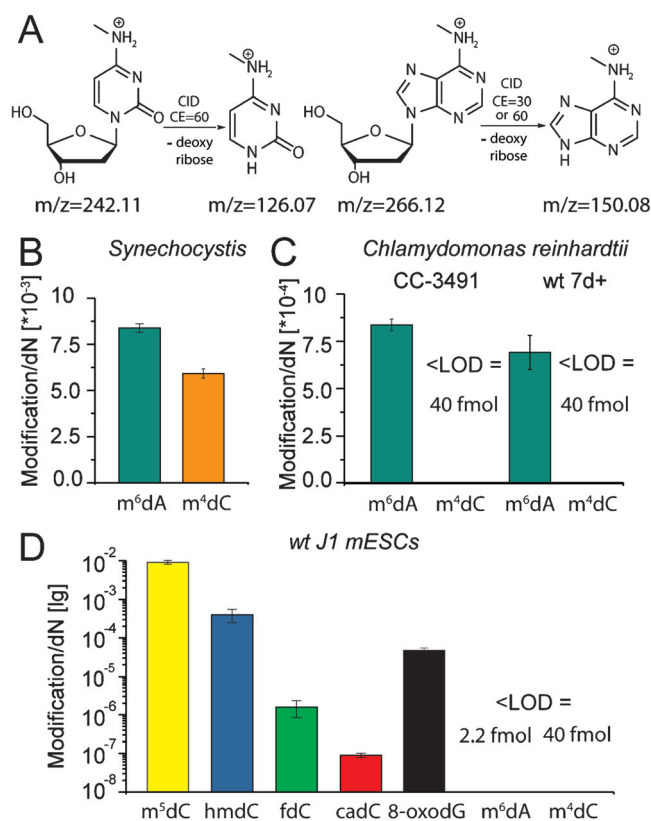
Figure 2. A) Fragmentation patterns of m4dC and m6dA. B–D) Quantitative data of the bases m4dC and m6dA in *Synechocystis* (B), *Chlamydomonas* (C) and of these bases and the other epigenetic DNA marks hmdC, fdC, cadC, and m5dC in mouse embryonic stem cells (D).

material of *Chlamydomonas*. This is interesting because *Synechocystis* is considered a relative of the chloroplasts present in *Chlamydomonas*.

With these positive results in hand, we extended our study to mouse embryonic stem cells (ESCs; wt J1, Figure 2 D). m6dA in particular was recently reported to occur as an epigenetically relevant DNA mark in mouse ESCs (mESC cell line wt TT2).[17] When performing the measurements, we turned the mass spectrometer to maximum sensitivity. But even in this mode, we were unable to detect a signal for m6dA within the detection limits of our system Table S2. In contrast, the other epigenetically relevant bases hmdC, fdC, cadC, and even the oxidative lesion 8-oxodG, which we also quantified in parallel, were clearly detectable. The 8-oxodG level was 4.8×10^{-5} per dN. The rare and difficult to detect cadC was clearly seen even at levels of only 9.0×10^{-8} per dN. For m6dA, in contrast, a signal did not appear. We also re-measured wt TT2 cells as described and still did not detect m6dA over background levels (see Figure S4).[17] To obtain unequivocal proof that m6dA is not present in stem cells, we added 13CD3-methionine to the mESC culture. Methionine provides the methyl group for the biosynthesis of m6dA. With 13CD3-methionine, this would lead to an m/z-shift of $+4$. We tuned the mass spectrometer to the new m/z-transition and again were unable to see any signal for 13CD3-m6dA (Figure S8).

We subsequently turned our attention to adult mouse tissue and analysed DNA isolated from liver and whole brain

(Figure S3) using our UHPLC-QQQ method. Figure 3 shows the data obtained from mouse liver. The middle column shows the data we obtained for hmdC. The already reported D_2-$^{15}N_2$-hmdC standard elutes at a retention time of 2.25 min

The m^{6}dA data are highly interesting (Figure 3 right column). While the D_3-m^{6}dA standard was clearly detectable at a retention time of 10.1 min, the unlabelled m^{6}dA provided a very weak signal. We then performed a control experiment to determine the limit of detection and investigated the digestion solution alone, which contains all of the commercial enzymes but no isolated DNA (red chromatogram). A weak signal for m^{6}dA was again detectable at a retention time of 10.1 min. After subtracting this background signal (red) from the measured chromatogram (blue), we obtained the black line showing that the original signal at 10.1 min is purely caused by background derived from the enzyme mixture. Here it is important to note that most of these proteins are recombinant proteins obtained from bacterial expression systems and bacterial DNA contains plenty of m^{6}dA. To support the evidence that m^{6}dA is not present in vertebrate DNA, we performed the same experiment with HeLa cells and also observed no signal for m^{6}dA (Figure 3B).

Figure 3. A) Chromatograms of the mass signal of mouse liver DNA. UHPLC-QQQ data obtained for m^{5}dC and m^{4}dC (left), hmdC (middle), and m^{6}dA (right, blue line) and their corresponding isotopic standards are shown. Additionally, for m^{6}dA, the chromatogram for the digest blank is shown (red line) and a computed baseline (black line), which was determined by subtracting the digest blank from the sample. B) A chromatogram of the mass signal from UHPLC-QQQ data obtained for m^{6}dA in DNA from HeLa cells.

and shows the expected fragmentation of the glycosidic bond, providing the fragmentation signature $m/z = 262.12 \rightarrow 146.07$, which allowed assignment of the signal. The naturally occurring hmdC is detected at exactly the same retention time with a mass transition of $m/z = 258.11 \rightarrow 142.06$, thus unequivocally demonstrating the presence of hmdC in murine liver DNA. Regarding the different monomethylated dC compounds m^{5}dC and m^{4}dC (left column), the epigenetic DNA mark m^{5}dC is clearly detected at a retention time of 3.2 min, but for m^{4}dC with a retention time of 3.5 min, there is obviously no signal present.

We then determined our limit of detection for m^{6}dA to be 3.5×10^{-7} per dN, which corresponds to 170 m^{6}dA bases per murine genome. This is a very small number and demonstrates the excellent sensitivity of our method. It shows that the maximum number of m^{6}dA that could be theoretically present and would not be detectable by our method is less than 170 m^{6}dA bases per genome, which led us to conclude that m^{6}dA is likely not epigenetically relevant but rather formed as a DNA lesion, perhaps by misguided methyltransferases. Spiking tests with synthetic nucleoside and DNA from *Chlamydomonas* nevertheless confirmed the sensitivity of the method, since the input amount equalled the found amount (see Figure S6).

To find a potential source for m^{6}dA in mESC DNA that could explain previous sequencing data,[17] we thought that m^{6}dA-containing bacterial DNA that gets degraded could provide the m^{6}dA nucleoside, which then might get incorporated into mESC DNA. This is indeed a possibility. When we added the m^{6}dA nucleoside to a mESCs culture, we indeed saw incorporation of some m^{6}dA into the genome (Figure S7).

We also cannot fully exclude the possibility that the presence of few m⁶dA bases, at levels below our detection limit, could have a biological function. In addition, it is possible that at certain stages of organismal development, certain methyltransferases are activated that may induce high m⁶dA levels at specific time points that may have escaped our detection.[18] Our data, however, show clearly that the maximum possible levels of m⁶dA in the analysed organisms and mESCs under normal conditions are far lower than so far believed.

Acknowledgements

Conflict of interest

The authors declare no conflict of interest.

[1] J. A. Law, S. E. Jacobsen, *Nat. Rev. Genet.* **2010**, *11*, 204–220.

[2] M. Tahiliani, K. P. Koh, Y. Shen, W. A. Pastor, H. Bandukwala, Y. Brudno, S. Agarwal, L. M. Iyer, D. R. Liu, L. Aravind, A. Rao, *Science* **2009**, *324*, 930–935.

[3] S. Kriaucionis, N. Heintz, *Science* **2009**, *324*, 929–930.

[4] T. Pfaffeneder, B. Hackner, M. Truss, M. Münzel, M. Müller, C. A. Deiml, C. Hagemeier, T. Carell, *Angew. Chem. Int. Ed.* **2011**, *50*, 7008–7012; *Angew. Chem.* **2011**, *123*, 7146–7150.

[5] S. Ito, L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He, Y. Zhang, *Science* **2011**, *333*, 1300–1303.

[6] Y. F. He, B. Z. Li, Z. Li, P. Liu, Y. Wang, Q. Tang, J. Ding, Y. Jia, Z. Chen, L. Li, Y. Sun, X. Li, Q. Dai, C. X. Song, K. Zhang, C. He, G. L. Xu, *Science* **2011**, *333*, 1303–1307.

[7] R. M. Kohli, Y. Zhang, *Nature* **2013**, *502*, 472–479.

[8] S. Schiesser, T. Pfaffeneder, K. Sadeghian, B. Hackner, B. Steigenberger, A. S. Schröder, J. Steinbacher, G. Kashiwazaki, G. Höfner, K. T. Wanner, C. Ochsenfeld, T. Carell, *J. Am. Chem. Soc.* **2013**, *135*, 14593–14599.

[9] M. Su, A. Kirchner, S. Stazzoni, M. Müller, M. Wagner, A. Schröder, T. Carell, *Angew. Chem. Int. Ed.* **2016**, *55*, 11797–11800; *Angew. Chem.* **2016**, *128*, 11974–11978.

[10] M. Yu, L. Ji, D. A. Neumann, D. H. Chung, J. Groom, J. Westpheling, C. He, R. J. Schmitz, *Nucleic Acids Res.* **2015**, *43*, e148.

[11] M. Ehrlich, M. A. Gama-Sosa, L. H. Carreira, L. G. Ljungdahl, K. C. Kuo, C. W. Gehrke, *Nucleic Acids Res.* **1985**, *13*, 1399–1412.

[12] Y. Fu, G. Z. Luo, K. Chen, X. Deng, M. Yu, D. Han, Z. Hao, J. Liu, X. Lu, L. C. Dore, X. Weng, Q. Ji, L. Mets, C. He, *Cell* **2015**, *161*, 879–892.

[13] E. L. Greer, M. A. Blanco, L. Gu, E. Sendinc, J. Liu, D. Aristizabal-Corrales, C. H. Hsu, L. Aravind, C. He, Y. Shi, *Cell* **2015**, *161*, 868–878.

[14] G. Zhang, H. Huang, D. Liu, Y. Cheng, X. Liu, W. Zhang, R. Yin, D. Zhang, P. Zhang, J. Liu, C. Li, B. Liu, Y. Luo, Y. Zhu, N. Zhang, S. He, C. He, H. Wang, D. Chen, *Cell* **2015**, *161*, 893–906.

[15] S. Hattman, C. Kenny, L. Berger, K. Pratt, *J. Bacteriol.* **1978**, *135*, 1156–1157.

[16] M. J. Koziol, C. R. Bradshaw, G. E. Allen, A. S. Costa, C. Frezza, J. B. Gurdon, *Nat. Struct. Mol. Biol.* **2016**, *23*, 24–30.

[17] T. P. Wu, T. Wang, M. G. Seetin, S. Lai, S. Zhu, K. Lin, Y. Liu, S. D. Byrum, S. G. Mackintosh, M. Zhong, A. Tackett, G. Wang, L. S. Hon, G. Fang, J. A. Swenberg, A. Z. Xiao, *Nature* **2016**, *532*, 329–333.

[18] J. Liu, Y. Zhu, G. Z. Luo, X. Wang, Y. Yue, X. Wang, X. Zong, K. Chen, H. Yin, Y. Fu, D. Han, Y. Wang, D. Chen, C. He, *Nat. Commun.* **2016**, *7*, 13052.

[19] W. Huang, J. Xiong, Y. Yang, S.-M. Liu, B.-F. Yuan, Y.-Q. Feng, *RSC Adv.* **2015**, *5*, 64046–64054.

[20] G. Z. Luo, F. Wang, X. Weng, K. Chen, Z. Hao, M. Yu, X. Deng, J. Liu, C. He, *Nat. Commun.* **2016**, *7*, 11301.

3.1.2   Isotope-dilution mass spectrometry for exact quantification of noncanonical DNA nucleosides

Franziska R. Traube#, Sarah Schiffers#, Katharina Iwan#, Stefanie Kellner, Fabio Spada, Markus Müller and Thomas Carell

(# These authors contributed equally to this publication)

Prologue

Analysis of DNA for epigenetic events often requires absolute quantification of modifications on a global level. To this end, UHPLC-MS/MS using the isotope-dilution technique can be a very useful tool. This protocol describes the details of this method in comparison with other available techniques. The publication provides a detailed description of all steps necessary to perform the quantification, from an optimized spin-column based DNA extraction protocol, total enzymatic DNA digest, to mass spectrometric analysis of a sample within 14 minutes.

Author contribution

The detailed author contribution is listed in the manuscript.

License

Copy of the article by publishing company; Copyright 2018 *Nature Publishing Group*.

Isotope-dilution mass spectrometry for exact quantification of noncanonical DNA nucleosides

Franziska R. Traube [1,2], Sarah Schiffers[1,2], Katharina Iwan [1,2], Stefanie Kellner[1], Fabio Spada[1], Markus Müller [1] and Thomas Carell [1*]

DNA contains not only canonical nucleotides but also a variety of modifications of the bases. In particular, cytosine and adenine are frequently modified. Determination of the exact quantity of these noncanonical bases can contribute to the characterization of the state of a biological system, e.g., determination of disease or developmental processes, and is therefore extremely important. Here, we present a workflow that includes detailed description of critical sample preparation steps and important aspects of mass spectrometry analysis and validation. In this protocol, extraction and digestion of DNA by an optimized spin-column and enzyme–based method are described. Isotopically labeled standards are added in the course of DNA digestion, which allows exact quantification by isotope dilution mass spectrometry. To overcome the major bottleneck of such analyses, we developed a short (~14-min-per-sample) ultra-HPLC (UHPLC) and triple quadrupole mass spectrometric (QQQ-MS) method. Easy calculation of the modification abundance in the genome is possible with the provided evaluation sheets. Compared to alternative methods, the quantification procedure presented here allows rapid, ultrasensitive (low femtomole range) and highly reproducible quantification of different nucleosides in parallel. Including sample preparation and evaluation, quantification of DNA modifications can be achieved in less than a week.

Introduction

In addition to the canonical nucleotides, both DNA and RNA contain a variety of modifications of the bases. In the DNA of vertebrates, for example, modified cytidines such as 5-methyl-2′-deoxycytidine (m5dC), 5-hydroxymethyl-dC (hm5dC)[1,2], 5-formyl-dC (f5dC)[3,4] and 5-carboxy-dC (ca5dC)[4,5] have been discovered. It is well established that m5dC and hm5dC are particularly epigenetically relevant[6,7]. The levels of hm5dC are often altered by several orders of magnitude in tumor tissues, and this has been shown to correlate with the aggressiveness of tumors[8–10]. In stem cells, f5dC and ca5dC were detected at substantial levels[3,4] and their abundance changes during differentiation[11,12]. Both f5dC[13] and ca5dC are thought to be involved in a process of active demethylation. Whether f5dC has additional distinct epigenetic functions is unclear[14–18] and, for ca5dC, no such function has yet been found. Neither f5dC nor ca5dC has yet been explored in regard to potential level changes in response to disease states. In bacteria, two major modifications are N^4-methyl-dC (m4dC)[19] and N^6-methyl-2′-deoxyadenosine (m6dA)[20]. The latter is a well-studied modification with extreme importance in host defense mechanisms[21]. m4dC has not been found in vertebrate DNA, and conflicting results have been found for research into the presence of m6dA in vertebrate DNA[22–26].

In addition to these noncanonical bases that are actively generated for partially unknown purposes, genomic DNA (gDNA) contains modified bases that are generated as DNA lesions. In particular, oxidative DNA lesions such as 8-oxo-7,8-dihydro-deoxyguanosine (8oxodG) can easily form[27], and, again, the levels of such base lesions can correlate with diseases[28]. Measurement of 8oxodG has proven to be particularly difficult, because its amounts can easily be overestimated because of additional oxidation during sample preparation or when bringing it into the gas phase in the mass spectrometer (gas chromatography MS)[29].

Determination of the exact amount of all noncanonical DNA nucleosides (epigenetic nucleosides or DNA lesions) is consequently an important task and requires a fast, ultrasensitive and highly reproducible approach.
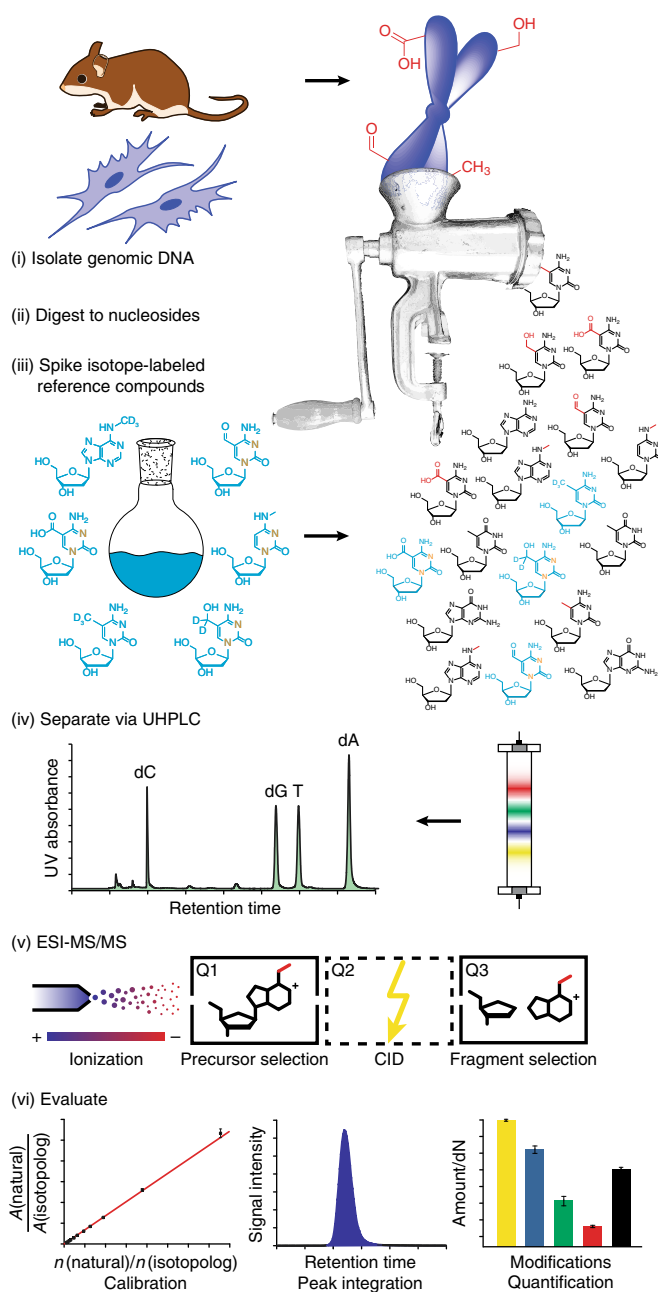
Fig. 1 | Procedure overview. General workflow for rapid DNA extraction, digestion and ultrasensitive quantification of the obtained nucleosides via UHPLC-QQQ-MS. T, deoxythymidine.

Here, we present a detailed workflow for the quantification of noncanonical nucleosides. It includes DNA extraction from cell culture or tissues, total enzymatic digestion of the DNA, and ultrasensitive quantification of the obtained nucleosides via UHPLC and QQQ-MS (UHPLC-QQQ-MS) (Fig. 1). The major advantage of UHPLC-QQQ-MS is its capability of chemically determining the true identity of the modified nucleoside and its exact abundance in a relatively short time frame. Furthermore, the application of the multiple reaction monitoring (MRM) mode enables simultaneous fragmentation of different precursor ions into one or multiple fragment ions and their subsequent detection[30]. The fundament of the method is the isotope dilution technique, in which stable isotopologs of the nucleosides of interest are added to the specimen (spiking) as internal standards. This analytic approach results in highly reproducible parallel quantification of modified DNA nucleosides in only ~14 min per sample and makes this method the ideal choice when sequence resolution is not required and sufficient sample material is available.

Development of the protocol

Quantification of very abundant DNA modifications, such as m⁵dC in vertebrates, has been possible for decades without the need for ultrasensitive quantification methods[31]. Owing to their abundance, small quantification errors do not affect interpretation of the data. However, accurate quantification of DNA nucleosides can be challenging when the modification is very rare. Contaminations at the cell culture level or during tissue dissection are a particular problem if they go unnoticed. These contaminations include bacteria and fungi, which often carry DNA modifications that might be rare or not present in mammals and can have a substantial effect on the resulting values. Therefore, the biological material that is investigated must be carefully tested for the presence of such contaminations. With respect to the investigation of DNA lesions, it must be noted that lesions are also formed during DNA isolation, for example, by releasing reactive oxygen species in the course of cell lysis. Most problematic are unspecific oxidations and deaminations. To overcome these problems, addition of butylated hydroxytoluene (BHT) and deferoxamine (DFOA) during isolation is essential. BHT is a radical scavenger that prevents DNA from being oxidized, and DFOA inhibits deaminases during the isolation process[11,32].

The biggest challenge of exact quantification is the generation of correct calibration curves. Mass spectrometry is a highly sensitive, but inherently not quantitative, method. The signal intensity reflects not only the amount of a molecule of interest (MOI) in the specimen, but also its ionizability, which is, in complex mixtures, strongly affected by other co-eluting molecules. To correlate the signal intensity with the exact amount of the MOI, calibration curves are essential. In principle, there are two different procedures for quantifying the amount of an MOI. One can use either external or internal calibration. Both approaches depend on the availability of the MOI in weighable quantities from, e.g., chemical synthesis. For external calibration, a serial dilution of the synthetic MOI is measured before analysis of the samples. Here, it is important to use the same buffer and the same LC column, and to perform the calibration on the same day, ideally once before the sample measurement, once after the sample set and, in the case of many samples, once during the sample worklist. This is important in order to counteract differences in instrument performance. The advantage of the external method is that it requires only the pure MOI as the standard. However, it also has many disadvantages. For best reproducibility, at least five different dilutions are measured as technical triplicates, which increases the measurement time substantially. Although this extensive calibration effort reduces errors from instrument performance fluctuations, it cannot overcome quantification errors stemming from matrix effects.

Matrix effects are mainly caused by salts, solvents and other undefined components within the sample[33]. It must be noted that a biological sample, e.g., gDNA, is much more complex than a solution of pure, synthetically generated nucleosides and that this complexity impacts the measurement. The matrix of the samples or the pressure on the column may furthermore lead to shifts in retention time (RT) of the MOI, so that a reliable statement on identification of the MOI may be infeasible as well. The sample matrix may also affect the signal intensity of the MOI: certain nucleosides suffer from ion suppression, which occurs when high amounts of interfering ions are present that may co-elute with the MOI and therefore compete for charge and space in the mass spectrometry device[33,34]. Taken together, it is not predictable whether a certain MOI can be reliably quantified using external calibration. Therefore, the gold standard in MS quantification is internal calibration, whereby a specific amount ($n^{\star}$) of a reference molecule is added to the sample and used as an internal standard at all times.

Our group utilizes the isotope dilution technique, in which the reference compound is an isotopolog of the MOI (designated MOI*)[13,26,35–44]. This molecule has very similar chromatographic and fragmentation properties as compared to the MOI, and as such it elutes ideally with an identical RT. The isotopes of choice should be ¹³carbon (¹³C) and ¹⁵nitrogen (¹⁵N), because no RT shifting is observed with ¹²C/¹³C and ¹⁴N/¹⁵N substitutions. By contrast, replacing hydrogen (H) atoms with deuterium (D) atoms in the MOI* affects the physicochemical properties of the molecules[45,46]. This leads to an observable shift in RTs, and, more importantly, the acid–base properties and thus the ionizing efficiency are affected in the MOI* as compared to the MOI.

The chosen MOI* is then used as an internal standard to reference and identify the correct peak of the MOI in the chromatogram. As such, it is essential to achieve accurate quantification. Calibration curves based on internal standards have the advantage that they can be used for multiple measurements on different days, because they are independent of the instrument performance on a specific occasion.

Applications of the method

The described protocol for DNA isolation is optimized for cell culture and vertebrate tissues. Isolation of DNA from other organisms, especially from those consisting of cells with cell walls, require harsher conditions. Once the DNA is isolated, our protocol can be applied to any kind of sample and the robust method reported here provides reliable quantification data. In addition, the method is highly sensitive, which enables quantification of nucleosides in the low femtomole and even attomole range, as described in the following section. Thereby, depending on the amount of DNA available for the measurements and the genome size of the organism of interest, even the determination of nucleosides that are very rare (a couple hundred nucleosides per genome) is possible. With this specificity, one can reliably demonstrate the presence of a modification in a genome[26] and determine even small changes of its abundance in response to disease progression[43]. Furthermore, the effects of stress factors, as well as those of cellular differentiation and mutation/knock-down of involved proteins, can be measured. Such biological conditions influence the biological pathways that lead to the formation or removal of noncanonical nucleosides in the genome[13,44].

Comparison with other methods

The workflow described here involves isolation and total digestion of gDNA with subsequent (parallel) analysis and quantification of nucleosides of interest, even in complex mixtures. This method provides robust and highly reproducible data in a fast manner due to the utilization of stable isotopologs as internal standards. The analysis time for parallel quantification of nucleosides routinely takes <15 min.

DNA digestion can be performed in a variety of ways, depending on the research focus. The method described here is optimized toward native DNA modifications, and other protocols that require harsh conditions or reactive substances will not be discussed in detail in this section. A comparative summary is nevertheless provided in Supplementary Table 1. For the optimized digestion of DNA adducts, see Liu et al.[47].

Antibody-based methods, such as dot blots, often represent the method of choice for nucleoside analysis when no detector-based system is available. These methods always depend on the specificity and binding affinity of the antibody. Antibodies have the major disadvantage that they often show cross reactivity, particularly between RNA and DNA bases[24], and it was recently shown that IgG antibodies have an intrinsic affinity for short DNA repeats[48]. This might obscure the obtained data. Our method is a combination of two different analytical techniques, (i) separation of analytes with an LC device and (ii) detection of these analytes with a detector, and it can distinguish between DNA and RNA nucleosides. In general, LC can be accomplished by applying HPLC or UHPLC. The following detection of the eluting MOI is achieved by using a UV detector, a fluorescence detector (FLD), an electrochemical detector or an MS device.

Chromatographic separation

Separation of the nucleosides of interest can be done by either reversed-phase HPLC or UHPLC. Both methods can apply either isocratic or binary gradient elution. Binary gradients usually shorten the analysis time, and the resolution is similar to that for isocratic elution. Gradient elution requires more regular maintenance, and the accuracy of the peak area and peak height is often questionable, which could impede reliable quantification[49,50]. This does not apply to our isotope dilution method, as a variation in size or shape of the peak will also affect the corresponding isotopologs and is therefore taken into account. On the chromatographic side, the generally high polarity of nucleosides requires the use of reversed-phase C8 or C18 columns to achieve good separation. Recently, hydrophilic interaction chromatography (HILIC) columns were alternatively used for separation of nucleosides[51]. The most recent developments regarding the high-throughput analysis of nucleosides is the application of UHPLC. The UHPLC columns have smaller particle size (typically ≤2 μm, in comparison with >4 μm for HPLC), and they are shorter and have a smaller inner diameter (50–150 mm length, <2.5-mm diameter for UHPLC as compared to 150–250 mm length, 4.6-mm diameter for HPLC). As such, they provide superior separation in a short time (~15 min for UHPLC as compared to ~45 min for HPLC). Most importantly, these columns give extremely narrow and sharp peaks that are essential to achieving the highest possible sensitivity. In comparison to ordinary HPLC, this UHPLC technique has lower solvent consumption due to low flow rates <0.5 mL/min. With less solvent, analytes are also more concentrated after elution and can be more easily vaporized and ionized, which is a major advantage for the subsequent MS measurement. In principle, the injection volumes should be in the

range between 1 and 10 µL; otherwise, this will impair efficient separation. As a rule of thumb, 10% of the flow rate should be used as the maximum injection volume. In our system, the flow rate is 0.35 mL/min, which allows for a maximum injection volume of 35 µL. However, the presented system was found to accept injection volumes up to 40 µL without loss of chromatographic resolution.

Detectors

Previously established methods for detection use, for example, the combination of reversed-phase HPLC and a UV detector. A UV detector is often the detector of choice, because the resulting peak area, and therefore the amount of a nucleoside, depends only on its extinction coefficient. UV-based quantification entails no matrix effects. Determination of the amount of each nucleoside can therefore be reliably accomplished by applying external calibration curves. One disadvantage is that baseline separation of the nucleosides of interest is critical. Moreover, UV detection suffers from lower sensitivity as compared to the MS- and fluorescence-based methods, which is reflected in the limits of quantification (here, lower limit of quantification (LLOQ)) of ~0.3–1.4 pmol. Correspondingly, the limits of detection (LOD) are typically not <0.08–0.42 pmol per nucleoside[52]. The absolute numbers depend on the type of nucleoside.

Higher sensitivity can be achieved when an LC device is coupled to an FLD. This combination results, for example, in an LOD for deoxyguanosine (dG) of 0.24 pmol[53]. Because most nucleosides do not show strong enough autofluorescence, however, they must be derivatized for fluorescence detection. For example, phenylglyoxal can be used to modify guanine-containing compounds[53]. m^5dC can be converted into the corresponding fluorescent $3,N^4$-etheno-5-methyl-2′-deoxycytidine (εm^5C). These compounds can then be quantified by reversed-phase HPLC-FLD with a typical LOD of 0.02 pmol[54]. One major disadvantage of this method is that derivatization chemistry is never quantitative (never 100% efficient). This leads to an underestimation of the abundance of the MOI and may distort the resulting data.

For detection and quantification of a rare MOI, an MS detector with higher sensitivity compared to UV and FLD is required. To analyze the MOI, it must be brought into the gas phase and be ionized so that it can enter the vacuum of the mass spectrometer. This critical step is achieved in the ion source, and it is the first critical parameter that must be optimized for high sensitivity. Next, efficient fragmentation and detection of the specific fragments must be achieved. Our MS device is equipped with an ESI source, which is perfectly suited to the evaporation and ionization of nucleosides and nucleotides. ESI is considered to be a mild ionization source that applies relatively low energy. This results in less in-source fragmentation of the MOI in comparison to those of electron ionization and chemical ionization, and therefore less loss of signal intensity due to a broad distribution of fragment ions[55]. Furthermore, this technique provides the highest possible flow rate to the following interface, e.g., the mass analyzer[56]. This is essential for maximal sensitivity.

We use QQQ mass spectrometers for our analyses because they have the highest sensitivity among the available mass analyzers. Time-of-flight (ToF) analyzers[57], ion-trap analyzers[58,59] or hybrid analyzers[60], which are a combination of different mass analyzers, can also be used in this protocol. Compared to ToF and ion-trap mass detectors, the QQQ detector has high sensitivity but low resolution. This means a QQQ mass spectrometer cannot distinguish between ions that differ in <1 AMU. However, high resolution is not necessary for quantification of known compounds for which synthetic standards are available.

The greatest advantage of the QQQ mass detector is the possibility of monitoring multiple mass/ nucleoside signals in the MRM mode, e.g., examination of fragmentation of the precursor ion and observation of the resulting product ion simultaneously, within a fraction of a second. The first quadrupole can be programmed to rapidly switch between monitoring molecules of various specific m/z values. The selected MOI ions are then fragmented in the second quadrupole, which represents the collision cell. Subsequently, the third quadrupole can be stepped to different m/z values to identify a specific fragment ion generated from the MOI in the collision cell. The detector ascertains the analytes, and the signal is enhanced by an electron multiplier. This technique improves the detection limits of analytes because only the m/z values from the molecules of interest are recorded, instead of scanning across the whole mass spectrum[55].

The method presented here is a combination of a UHPLC device coupled to a UV detector and a QQQ analyzer. The less sensitive UV detector is used to quantify the highly abundant canonical nucleosides, whereas the quantification of the less abundant noncanonical nucleosides is achieved by the QQQ detector. Applying this method leads to LLOQs in the femtomole range or lower for nucleosides of interest and enables the determination of their quantities in the genome.

Experimental design

Our protocol covers all steps to establish the whole procedure from the beginning, when neither expertise nor the material for quantification of DNA nucleosides is available. For development of a quantification method, several machine parameters must be determined, apart from the m/z values for the precursor and product ion of the MOIs and MOI*s.

The Procedure provides detailed steps for two methods:

(i)   Quantification of the so-called epigenetic DNA modifications (m^5dC, hm^5dC, f^5dC and ca^5dC), as well as a common DNA lesion (8oxodG), with highest sensitivity toward the less abundant modifications. This method uses a C8 column and a water/MeCN solvent system (Supplementary Tables 2 and 3).

(ii) Quantification of further DNA modifications (m^4dC and m^6dA) known from bacteria, in which they are highly abundant. For investigation of their existence in higher organisms, in which their abundance is expected to be low, the use of a C18 column enables the detection and distinction of m^4dC and m^5dC. Furthermore, applying a water/MeOH solvent system provides the highest sensitivity for the detection of m^6dA. This method is used in two variations with adjusted collision energies (CEs) (sensitive and insensitive) for m^6dA to enable quantification of the modification in organisms with low or high abundance, respectively, without reaching limitations for other important components of quantification, e.g., the LLOQ of canonical nucleosides (Supplementary Tables 4 and 5).

A compilation of linear equations is given in Supplementary Tables 2 and 4; these can be applied only to the respective solvent systems and columns (in Supplementary Table 2, water/MeCN on a C8 column; in Supplementary Table 4, water/MeOH on a C18 column). For m^4dC, two linear equations are given because they span two ranges (big = values in the pmol range; small = values in the low fmol range), and for m^6dA, two linear equations are shown for the different CEs (sensitive and insensitive), thereby resulting in two respective ranges.

If you want to measure all the nucleosides given in Supplementary Tables 2 and 4 at once, new calibration curves are necessary for the modifications with the different solvent system. In this case, we recommend sticking to the water/MeOH system with a C18 column.

If identical equipment and methods are used, it is possible to use the calibration curves provided in Supplementary Tables 2 and 4.

Internal calibration

We recommend internal calibration with a stable isotopolog MOI* as a reference for the MOI. This MOI* is spiked into the DNA sample. If several MOIs must be quantified in parallel, a mixture of the required MOIs* is spiked into the sample in the form of a spiking mixture. This mixture contains not only the needed MOIs* but also the digestion enzymes, and it is added to all samples of one experiment. At best, the MOI* should be at least two atom units heavier than the MOI to allow a clear MS-based differentiation between the MOI* and the naturally occurring MOI. Natural ^{13}C has an abundance of 1%. Thus, any molecule with 10 C atoms, such as m^5dC, will have a 10% chance to carry at least one ^{13}C atom and a 1% chance to carry two ^{13}C atoms. The presence of these natural [+1] and [+2] isotopologs limits the sensitivity, especially for highly abundant MOIs. Therefore, a mass difference between the MOI and the MOI* of three or more units is recommended for high fidelity.

The mass spectrometer is set to monitor the nucleoside signals of the MOI and the MOI*, which will be displayed in two separate chromatograms: one for the MOI signal and one for the MOI* signal. Integration of the peak areas for MOI and MOI* gives areas A and A^*. To calculate the amount of the MOI from A and A^*, calibration curves are required. These calibration curves are generated before the measurement by mixing the MOI and its MOI* at different ratios, in which the amount of the MOI (n) is stepwise diluted by a factor of two, while the amount of MOI* (n^*) is kept constant.

The ratios of A/A^* and n/n^* correlate in a linear manner. This allows the use of the following linear equation:

$$\frac{A}{A^*} = m \times \frac{n[pmol]}{n^*[pmol]} + t.$$

For the determination of the slope m and the y-intercept t of the calibration curve, the aforementioned ratios for n/n^* are measured in technical triplicates, and the resulting ratio of A/A^* is plotted against n/n^*. The standard deviation for each data point must not exceed 20%, and the accuracy must be between 80 and 120%.

The lowest amount that meets the requirements for the standard deviation and accuracy determines the LLOQ. The same criteria regarding standard deviation and accuracy apply for the highest amount, which is therefore defined as the upper limit of quantification.

Once these prerequisites are met, the specific amount of the MOI (n) can be calculated using the following equation:

$$n[\text{pmol}] = \frac{\frac{A}{A^*} - t}{m} \times n^*[\text{pmol}].$$

If the calibration curve meets those requirements, it is considered reliable, independent of the machine performance, as long as the components are the same. The specific parameters for each compound, e.g., collision energy and cell accelerator voltage, must be set identically for the MOI and the MOI*. This saves measurement time and secures high fidelity. In addition, RT shifts can be accepted if the internal standard shows the same shift.

However, if the MOI and the MOI* do not co-elute, it must be assumed that one of the signals is not caused by either the MOI or MOI*. In particular, in complex mixtures, it may be generated by an unknown analyte that shows the same mass signal. In general, RT differences that are, for example, caused by the isotope difference between the MOI and the MOI*, are tolerated when they fall within a 2.5% limit. In particular, when deuterium is used as a source of labeling, slightly shifted RTs are common because of diverging binding strengths. Still, the MOI and MOI* must meet the above criterion. If the deviation is >2.5%, the apparent peak is not the peak of the MOI[61]. Whether the 2.5% limit is exceeded can be calculated with the following equation:

$$\frac{|\ RT^{MOI}[\text{min}] - RT^{MOI*}[\text{min}]\ |}{RT^{MOI*}[\text{min}]} \times 100\% \overset{!}{\leq} 2.5\%.$$

The limit of detection

The LLOQ does not equal the lower LOD, which is important for evaluation of the data. Even if the LLOQ criterion is fulfilled, the sample data may not meet the LOD criterion when the signal-to-noise ratio (SNR) is too low. The $LOD_{theoretical}$ is set at an SNR of 3 for the pure MOI in water. But analyzing MOIs in biological samples demands harsher constraints because of matrix effects. Therefore, the $LOD_{practical}$ is defined as the SNR of 3 of the MOI in a biological sample. To determine the $LOD_{practical}$ in a biological sample, we always evaluate so-called digestion blanks (in our protocol, samples 1-3) for each measurement. The digestion blanks contain the spiking mixture, but no DNA. For the digestion blanks, we analyze A (in this case, ideally A = 0) at the respective RTs of each MOI and A* of each MOI* of the spiking mixture. Therefore, the $LOD_{practical}$ relates to the background signal of the spiking mixture in the digestion blanks. It can be calculated in two different ways:

(i)   One can calculate the $LOD_{practical}$ as the mean of the ratio A/A* for the digestion blanks multiplied by the factor of 3.

(ii) One can calculate the $LOD_{practical}$ as the mean of A for the digestion blanks multiplied by the factor of 3.

Both definitions can be used for further data evaluation. Method (ii) is even stricter than method (i). However, it is important to choose one method consistently. Applying this $LOD_{practical}$ value will prevent the emergence of false-positive data and is therefore highly recommended. Only the ratios of A/A* (method (i)) or the areas A (method (ii)) of the subsequent samples that exceed this $LOD_{practical}$ represent true peaks. Ratios of A/A* that exceed the LLOQ determined by the calibration curve, but not the $LOD_{practical}$, might be false-positive hits and therefore have to be excluded from further analysis. To overcome this problem in future experiments, we recommend the use of more DNA.

Calibration curves

A calibration curve spans a specific concentration range of at least five data points, in which the curve is linear. The limitations of this so-called dynamic range become problematic if an experiment contains samples in which the abundance of the MOI varies by several orders of magnitude. In this case, a single calibration curve might be insufficient if it does not cover the resulting intensities of signals. Then a second calibration curve with a different range is needed. The two resulting calibration curves will ideally cover the lowest and highest concentrations of the MOI in the specimen. To achieve consistent quantification results across multiple biological samples, it is important to use a single calibration curve for the whole dataset. It may, for example, occur that a knock-out cell line

cannot be quantified with the same calibration curve as its corresponding wild type, because the amount of the MOI in the latter exceeds the values in the knockout by a factor of ten or more. For this biological question, a combination of different calibration curves might seem a valid solution, but it is not recommended. It is better to adjust the amount of DNA by dilution, so that the total amount of the MOI in both samples falls within the same range and hence a single calibration curve can be applied. It is also important that the concentrations of the spiked-in MOI*s are in the range of the expected concentrations of the MOI. If dilution of the DNA is not applicable, one can prepare two different spiking mixtures to meet the requirements of a single calibration curve; one with a higher amount of MOI*, which is added to the biological samples when a higher amount of MOI is expected and one with a lower amount of MOI* for the corresponding sample with a low abundance of MOI. In the subsequent evaluation of the data, the respective amount of MOI* (n^*) must be adjusted.

For handling samples with unknown and potentially strongly deviating MOI content, we recommend generating a calibration curve, in which you start with an amount n that is more than four times greater than n^* and dilute this amount n by a factor of three instead of two. We used this procedure successfully in some cases, but for certain modifications, the smaller dilution factor resulted in a better calibration curve.

Evaluation of UHPLC–MS/MS data

The provided Excel sheet (Supplementary Methods 1 and 2) includes all functions that are needed for evaluating the data. It allows calculation of the ratio of A/A^* and thereby the unknown amount n of the MOI with the following equation, where m represents the slope and t is the y-intercept. This equation is unique for each MOI and only valid within its concentration ranges.

$$\frac{A}{A^*} = m \times \frac{n[\text{pmol}]}{n^*[\text{pmol}]} + t,$$

$$n[\text{pmol}] = \frac{\frac{A}{A^*} - t}{m} \times n^*[\text{pmol}].$$

Note that in this protocol an injection volume of 39 μL is applied. This was chosen because it is the largest amount that can be injected without a loss of chromatographic resolution. The amount of the MOI (pmol/sample), e.g., in 50 μL of digestion mixture, is calculated by multiplication of the upper equation by (50/39 μL). The complete calculation is described in the following equation:

$$n[\text{pmol}] = \frac{\frac{A}{A^*} - t}{m} \times n^*[\text{pmol}] \times \frac{50\,\mu\text{L}}{39\,\mu\text{L}}.$$

If <39 μL is injected, this value must be adjusted in the provided Excel sheet (Supplementary Methods 1 and 2). This additional calculation will provide comparable data for your technical and later biological replicates. The injection volume is automatically recorded in the QQQ quantitative analysis program.

In the case that the GC content of the DNA is known, division of the amount of MOI (in pmol) by the amount of dG (in pmol) will provide the term 'MOI per dG'. For example, for mouse gDNA, further multiplication by 0.21 (adjusting for the 42% GC value of mouse gDNA[62]) will yield the term 'MOI per dN', which is mainly used for presenting quantification data.

$$\frac{n(\text{MOI})[\text{pmol}]}{n(\text{dG})[\text{pmol}]} = \text{MOI/dG},$$

$$\text{MOI/dG} \times 0.21 = \text{MOI/dN}.$$

In the case that the GC content is unknown, the amounts of dA, dC, dG and thymidine (T) must be determined using UV detection, and the amount of MOI divided by the amount of Σ(dA, dC, dG, T) will directly provide the term 'MOI per dN'.

If one is interested in the derivatives of dC, namely m^5dC, hm^5dC, f^5dC and ca^5dC, it is important to perform additional calculations because these modifications are measured in the MS mode, whereas canonical dC is measured in the UV mode. The combination of UV- and MS-derived data often results in summed values for the total amounts of all dC derivatives (dC, m^5dC, hm^5dC, f^5dC, and ca^5dC and/or dC*) that deviate from the dG content, which is quantified by UV detection. Therefore, the expected exact 1:1 ratio of $\Sigma x dC/dG$ is almost never reached. Nevertheless, the amounts of dC and xdC derivatives can be expressed relative to the amount of dG, but a corrective

factor is needed. To this end, the amount of each dC derivative is divided by the amount of dG. Multiplying by 100 results in the percentage of every dC derivative. Summarizing these percentages will lead to a total percentage as a corrective factor (%dC^{total}). Dividing the measured %xdC by this factor will provide the term 'MOI/dG'. For mouse gDNA, further multiplication by 0.21 will lead to the desired term 'MOI/dN'.

$$dC[pmol]/dG[pmol] \times 100 = \%dC,$$
$$m^5dC[pmol]/dG[pmol] \times 100 = \%m^5dC,$$
$$hm^5dC[pmol]/dG[pmol] \times 100 = \%hm^5dC,$$
$$\dots$$
$$\sum \left(\%dC; \%m^5dC; \%hm^5dC; \dots \right) = \%dC^{total}.$$

$$\%dC/\%dC^{total} = dC/dG^{correct},$$
$$\%m^5dC/\%dC^{total} = m^5dC/dG^{correct},$$
$$\%hm^5dC/\%dC^{total} = hm^5dC/dG^{correct},$$
$$\dots$$
$$m^5dC/dG^{correct} \times 0.21 = m^5dC/dN,$$
$$hm^5dC/dG^{correct} \times 0.21 = hm^5dC/dN,$$
$$\dots$$

The values of each MOI per sample can be presented as the mean and standard deviation of a technical triplicate. Combination of at least three biological replicates, e.g., their means, leads to reliable data.

Level of expertise needed to implement the protocol

A trained technician, graduate student or postdoctoral researcher can perform all the steps from DNA isolation to DNA digestion and sample preparation. For working with a UHPLC-QQQ-MS system, at least basic knowledge of how to use the machine is required. Core facilities for MS measurements typically operate the LC–MS/MS instrument and perform the standard LC-MS/MS analysis, but it is recommended to use a facility focused on small molecules, with a dedicated instrument and experienced personnel to avoid contamination with distinct analytes that would impair sensitivity. If this facility is unavailable on-site, the DNA samples can also be shipped to a respective facility on dry ice. But any researcher with an interest in nucleoside research can generate and evaluate data when he or she uses the provided material. Nevertheless, the problems addressed in the Troubleshooting section require a more profound knowledge of the mass spectrometer, so adapting the provided method to different nucleosides of interest or mastering upcoming challenges calls for a well-trained researcher.

Limitations

The high sensitivity of a triple quadrupole mass spectrometer originates from the selection for the specific mass signals of the molecules of interest by disregarding other potential contents of the analysis mixture. Before analysis of a sample of interest, it is therefore critical to define all nucleosides that are to be quantified. Molecules that were not considered in the method are not monitored, and their data can therefore not be extracted retrospectively.

The development of measurement protocols for new molecules requires optimization of the LC-based separation part, adjustment of MS parameters such as the optimal collision energy and validation of the nucleoside signal of the MOI. Furthermore, quantification with internal standards requires the availability of suitable isotopologs with a $\Delta m/z$ of at least 2, preferably more. If such an isotopolog is not available commercially, the corresponding molecule must be synthesized chemically or metabolically[63,64] or one needs a collaborator who can provide it.

Because the origin of the naturally occurring nucleosides is DNA, the extraction and digestion efficiency are critical. Extracting DNA from cell culture usually leads to high yields and can typically also be scaled up. DNA isolation from tissues, however, is more demanding and may not provide a sufficient amount of DNA to detect a modification with low abundance. In this case, it might be necessary to combine several biological samples. The isolation of DNA from only a few cells and subsequent quantification of modified nucleosides has been reported[65], but it is not routinely possible. At the other extreme, excessive amounts of DNA and corresponding nucleosides can result in

so-called matrix effects, which can suppress the signal of a nucleoside. It is therefore required to determine the optimal amount of DNA to be analyzed to obtain the best signal. The reported method provides global quantitative data; sequence information is not available.

The whole method depends on the quality of the input material, as the origin of the resulting DNA, and the respective nucleosides, cannot be determined at a later stage. Contaminations, e.g., from a bacterial or fungal source during cell culture work or from the microbiome of a higher organism, might lead to false-positive results for certain modified nucleosides. In addition, the abundance of modifications may vary substantially during cell differentiation or in response to stress. Therefore, the timing of the cell culture work, e.g., harvesting time points, is very important to obtaining reliable values for the biological replicates.

There are also some challenges when performing the measurements, as not every rare nucleoside of interest can be quantified with every setup. Sometimes it may not be possible to measure all nucleosides with the same setup and conditions. If no further optimization is possible, one needs to process the sample with two different methods, but this of course requires increased instrument time and it also requires more material. For nucleosides that are quantified using the UV trace (dC and dG), it is critical that the peaks be baseline separated. For MS analysis, chromatographic separation of nucleosides is mostly not necessary because of the separation in the mass spectrometer according to their unique m/z values and fragmentation patterns. Only when analyzing samples that might contain isomers, e.g., m⁵dC and m⁴dC, which have identical precursor and product ions, chromatographic separation becomes necessary in order to clearly determine the identity of the detected signal.

To further increase the sensitivity and the number of MS data points, it is beneficial to subdivide the table of analytes into several segments according to their RTs. In each time segment, only certain nucleosides are monitored, which increases the dwell time for each analyte and thus the strength of the signal. The price of this increased MS sensitivity is that, depending on chromatographic performance, the RTs of the nucleosides might change, and then these nucleosides may escape their time segment. Then the corresponding data of the analyte in the time segment are irretrievably lost.

Materials

Biological materials
- Cell line: cell lines of various sorts, as well as animal tissues, have worked well in our experience. Specifically, we have performed this protocol using iNGN cells[66] (hPSCreg no. HVRDi004-B-1); HEK293T cells (ATCC, cat. no. CRL-3216); mES wild-type cell line J1 (ref. [67]); and mouse cerebellum from a C57-BL6/J wild-type genetic background, provided by S. Michalakis (Department of Pharmacy, Ludwig-Maximilians–Universität München). **! CAUTION** The cell lines used in your research should be regularly checked to ensure they are authentic and are not infected with mycoplasma. **! CAUTION** Ensure sterile work in order to avoid cross-contamination of the extracted tissue with bacteria or other organisms from the environment. **! CAUTION** All animal experiments must be performed according to the relevant guidelines and regulations and must be approved by your institutional animal care and use committee.

Reagents
- 2-Mercaptoethanol (βME, CAS no. 60-24-2; Sigma-Aldrich, cat. no. M3148-25mL) **! CAUTION** 2-Mercaptoethanol is toxic, so avoid exposure.
- Acetonitrile (MeCN, 99.95% (vol/vol), LC–MS grade; Roth, cat. no. AE70.2) **▲ CRITICAL** Sensitivity of the mass spectrometer might vary when a different supplier is used. Check new suppliers carefully.
- Antarctic phosphatase (New England Biolabs, cat. no. M0289L)
- BHT (CAS no. 128-37-0; Sigma-Aldrich, cat. no. B1378-100G) **▲ CRITICAL** To keep background oxidation at a minimum, it is recommended to store BHT powder under vacuum.
- Benzonase nuclease, 10 KU (VWR, cat. no. 70746-3)
- Blood & Cell Culture DNA Midi Kit (Qiagen, cat. no. 13343)
- Degradase Plus (Zymo Research, cat. no. E2021)
- DFOA (Sigma-Aldrich, CAS 138-14-7, cat. no. D9533-1G)
- Dimethyl sulfoxide (DMSO, CAS no. 67-68-5; Acros, cat. no. 327182500)
- DNA pre-wash buffer (Zymo Research, cat. no. D3004-5)
- Dulbecco's PBS without MgCl₂, CaCl₂ (DPBS; sterile filtered, suitable for cell culture; Sigma-Aldrich, cat. no. D8537-500 mL)
- EDTA disodium salt (NA₂[EDTA]; CAS no. 6381-92-6; VWR, cat. no. 33600.267)

- Formic acid (CAS no. 64-18-6; Fluka Honeywell Chemicals, cat. no. 94318-50 mL-F) **! CAUTION** Formic acid is highly corrosive and can lead to severe burns when inhaled. Use only in a fume hood or in a highly ventilated area and protect your skin and eyes carefully.
- gDNA wash buffer (Zymo Research, cat. no. D3004-2)
- Genomic lysis buffer (Zymo Research, cat. no. D3004-1) **▲ CRITICAL** Do not add βME in advance.
- Glycerol (CAS no. 56-81-5; Roth, cat. no. 3783.2)
- Magnesium chloride hexahydrate (MgCl$_2$·H$_2$O, CAS no. 7786-30-3; Merck, cat. no. M8266-100g)
- Methanol (MeOH; 99.9% (vol/vol), LC–MS grade VWR, cat. no. HONC34966-1L) **! CAUTION** Methanol is toxic, so avoid exposure. **▲ CRITICAL** Mass sensitivity might vary when a different supplier is used. Check new suppliers carefully.
- Nuclease S1 from *Aspergillus oryzae* (Merck, cat. no. N5661-50 KU) **▲ CRITICAL** Stock solution should be stored at −20 °C and kept on ice while making aliquots for usage.
- Parafilm M (4 × 125 inches, clear; Bemis, cat. no. 52858-000)
- Phosphodiesterase I from *Crotalus adamanteus* venom (Abnova, cat. no. P5263)
- PicoGreen dsDNA Assay Kit (Thermo Fisher Scientific, cat. no. P7589)
- Buffer RLT (Qiagen, cat. no. 79216)
- RNase A (100 mg/mL, 7,000 U/mL; Qiagen, cat. no. 19101)
- Sodium chloride (NaCl; CAS no. 77-86-1; Bernd Kraft, cat. no. 10724344)
- Tetrahydrouridine (THU; CAS no. 18771-50-1; Merck Millipore, cat. no. 584222)
- Tris(hydroxymethyl)aminomethane (Tris base; CAS no. 77-86-1; Fisher Scientific, cat. no. BP152-5)
- Water, LC–MS grade (Honeywell, cat no. 39253-1L) **▲ CRITICAL** Mass sensitivity might vary when a different supplier is used. Check new suppliers carefully.
- Zinc sulfate (ZnSO , CAS no. 7446-20-0; Grüssing, cat. no. 14039)

Nucleosides
- 2′-Deoxyadenosine (dA) (Carbosynth, cat. no. ND04011, CAS 16373-93-6)
- 2′-Deoxycytidine (dC) (Carbosynth, cat. no. ND06286, CAS 951-77-9)
- 2′-Deoxyguanosine (dG) (Carbosynth, cat. no. ND06306, CAS 961-07-9)
- Thymidine (T) (Carbosynth, cat. no. NT02592, CAS 50-89-5)
- 5′-Methyl-2′-deoxycytidine (m5dC) (Carbosynth, cat. no. ND06242, CAS 838-07-3)
- 5′-Hydroxymethyl-2′-deoxycytidine (hm5dC) (Carbosynth, cat. no. NH15898, CAS 7226-77-9)
- 5′-Formyl-2′-deoxycytidine (f5dC) (Carbosynth, cat. no. ND63556, CAS 137017-45-9)
- 5′-Carboxy-2′-deoxycytidine (ca5dC) (Carbosynth, cat. no. ND158446, CAS 46003-72-9)
- 8-Oxo-7,8-dihydro-deoxyguanosine (8oxodG) (Carbosynth, cat. no. ND06344, CAS 88847-89-6)
- 15N5-8oxodG (Cambridge Isotope Laboratories, cat. no. NLM-67 15-0, CAS NA)
- D3-m5dC (synthesis described in ref. 36)
- 15N2-hm5dC (synthesis described in ref. 41)
- 15N2-f5dC (synthesis described in ref. 41)
- 15N2-ca5dC (synthesis described in ref. 41)
- N4-methyl-2′-deoxycytidine (m4dC) (synthesis described in ref. 26)
- 15N2-m4dC (synthesis described in ref. 26)
- N6-methyl-2′-deoxyadenosine (m6dA) (synthesis described in ref. 26)
- D3-m6dA (synthesis described in ref. 26)

Equipment
- Stainless-steel beads (5 mm; Qiagen, cat. no. 69989)
- Syringe filter (0.2-μm cellulose acetate; VWR, cat. no. 514-0061)
- Snap ring cap 11 mm tr. (natural rubber/TEF, 60°, 1.0 mm, HPLC vial cap; VWR cat. no. 548-0014)
- Snap ring micro-vial (0.3 mL, polypropylene, 32 × 11.6 mm, transparent, HPLC vial; VWR, cat. no. 548-0120) **▲ CRITICAL** If using different HPLC vials, they must have a volume-reducing insert.
- Poroshell 120 SB C8 column (2.7 μm, 2.1 × 150 mm; Agilent Technologies, cat. no. 683775-906)
- Poroshell 120 SB-C18 column (2.7 μm, 2.1 × 150 mm; Agilent Technologies, cat. no. 683775-902)
- PCR plate (skirted, 96-well, 0.2 mL; VWR, cat. no. 732-3225)
- Falcon tubes (15 mL; VWR, cat. no. 188271)
- Falcon tubes (50 mL; VWR, cat. no. 227161)
- AcroPrep Advance 96-well, 350-μL, 0.2-μm Supor short-tip natural polypropylene plates (Pall, cat. no. 518-0022)

- Centrifuge tubes (0.5 mL; Eppendorf, cat. no. 211-2140)
- Centrifuge tubes (1.5 mL; Eppendorf, cat. no. 211-2130)
- Centrifuge tubes (2.0 mL; Eppendorf, cat. no. 211-2120)
- −20 °C Freezer (e.g., Bosch, model no. GSN58AW45)
- −80 °C Freezer (e.g., Eppendorf, Innova U725, model no. U9440-0002)
- 0.5- to 10-µL, 2- to 20-µL, 10- to 100-µL, 20- to 200-µL, and 100- to 1,000-µL pipettes (Eppendorf)
- 4 °C Refrigerator (e.g., Liebherr, model no. LCv 4010)
- 37 °C Heat block (Eppendorf, ThermoMixer Comfort model, device, cat. no. 5382000015, plus top, cat. no. 5360000011)
- MM400 bead mill (Retsch, cat. no. 20.745.0001)
- Refrigerated benchtop microcentrifuge (e.g., Centrifuge 5424R; Eppendorf, cat. no. 5404000014)
- Refrigerated swinging-bucket rotor centrifuge (Centrifuge 5810R; Eppendorf, cat. no. 5811000428 with A-4-81 rotor and holders for MTP plates)
- Microscale (e.g., Sartorius, cat. no. RC 210 P)
- Ultrapure water system (e.g., arium pro DI; Sartorius Stedim Biotech, cat. no. H2OPRO-DI-B)
- UV/Vis spectrophotometer (NanoDrop; NanoDrop Technologies, cat. no. ND-1000)
- Triple quadrupole LC/MS system with iFunnel technology (Agilent Technologies, model no. 6490)
- UHPLC system (Agilent Technologies, model no. 1290 Infinity II LC)
- Vortex mixer (Scientific Industries, model no. Vortex-Genie 2)
- Zymo-Spin IIC-XL column (Zymo Research, cat. no. C1102)

Software
- Microsoft Office Excel 2016 (Microsoft, https://products.office.com/en-us/compare-all-microsoft-office-products?activetab=tab%3aprimaryr1)
- OriginPro 2016G b.9.3.226 (https://www.originlab.com/2016)

Reagent setup
▲ CRITICAL Deionized water is used for all solutions, unless otherwise indicated. ▲ CRITICAL For the gDNA isolation, work at room temperature (23 °C) all the time. Some buffers and solutions will freeze at 4 °C. ▲ CRITICAL For each procedure using stock solutions, let all nucleoside and salt dilutions thaw and equilibrate at room temperature and vortex them vigorously (for at least 1 min).
▲ CRITICAL Perform all pipetting steps with well-calibrated pipettes.

Internal standard mastermix
To make the internal standard (ISTD) mastermix, add 58.0 µL of D_3-m^5dC (m^5dC^*, concentration (c) = 264.1 µM), 90.0 µL of D_2,$^{15}N_2$-hm^5dC (hm^5dC^*, c = 25.5 µM), 90.0 µL of $^{15}N_2$-f^5dC (f^5dC^*, c = 0.152 µM), 120.0 µL of $^{15}N_2$-ca^5dC (ca^5dC^*, c = 0.108 µM), 180.0 µL of $^{15}N_5$-8oxodG (8oxodG*, c = 0.181 µM). Sum the real volume and bring the volume to 900.1 µL by adding water. Vortex rigorously for 1 min. The prepared ISTD mastermix, which is sufficient for 300 measurements, can be stored at −20 °C for up to 1 year, and multiple thawing and freezing cycles are acceptable.
▲ CRITICAL To ensure accurate preparation of the ISTD mastermix, the added volume after each pipetting step is controlled by using microscales. For each MOI*, a deviation in weight of ±5% is tolerated. If the deviation is >−5%, one could add the required volume of the MOI*, but if the deviation is >5%, the sample must be discarded. ▲ CRITICAL If you are interested in MOIs other than those listed in the ISTD mastermix, e.g., m^6dA or m^4dC, the MOI*s in the ISTD mastermix can be adjusted accordingly.

1,000× BHT
Prepare a 200 mM stock solution of BHT (1,000× BHT) in DMSO. Make 50-µL aliquots and store them at −80 °C for up to 3 years. 1,000× BHT can be thawed and refrozen up to three times but must be refrozen as soon as possible, and multiple freeze–thaw cycles should be avoided. Before use, the 1,000× BHT must be diluted 1:10 to make it soluble in water (100× BHT). Combine 1 equiv. of 1,000× BHT with 3 equiv. of DMSO and vortex briefly. Add 6 equiv. of H_2O dropwise and vortex thoroughly between drops to make sure that the BHT does not precipitate. 100× BHT should be a clear solution at the end. If the solution is turbid, vortex longer. If too much BHT has already precipitated, discard the dilution and prepare a new 100× BHT dilution from the 1,000× BHT stock. 100× BHT must be prepared immediately before use. ▲ CRITICAL Do not refreeze or store the 100× BHT; instead prepare it fresh each time before the isolation.

1,000× DFOA

Prepare a 200 mM stock solution of DFOA (1,000× DFOA) in degassed water. Make 20-μL aliquots and store them at −80 °C for up to 3 years. ▲ **CRITICAL** 1,000× DFOA should not be refrozen. ▲ **CRITICAL** When DFOA is dissolved in water, the final volume increases. Therefore, dissolve DFOA in only 80% of the calculated amount of degassed water, measure the volume afterward and add the missing volume at the end.

Lysis buffer

If the option to isolate RNA and total protein (denatured) is desirable, use Buffer RLT as the base lysis buffer. If only isolation of gDNA is required, use genomic lysis buffer (already contains RNase) instead. The yield of gDNA will be slightly higher if you use genomic lysis buffer. To 1 equiv. of base lysis buffer, add 0.01 equiv. of βME (14.3 mM final concentration), 0.002 equiv. of 1,000× DFOA (400 μM final concentration) and 0.02 equiv. of 100× BHT (400 μM final concentration). For example, to 1 mL of base lysis buffer, add 10 μL of βME, 2 μL of 1,000× DFOA and 20 μL of 100× BHT. The ready-to-use lysis buffer is referred to as Buffer RLT$^+$ or genomic lysis buffer$^+$, respectively. These buffers without βME, BHT and DFOA can be stored at room temperature until the expiration date given by the manufacturer. The buffers supplemented with βME, BHT and DFOA must be prepared immediately before use and cannot be stored. ▲ **CRITICAL** If you are interested in any kind of deamination, you should consider adding a broadband deamination inhibitor such as tetrahydrouridine (THU).

RNase wash buffer

To 1 equiv. of genomic lysis buffer, add 0.002–0.01 equiv. of RNase A (0.2–1 mg/mL final concentration), 0.002 equiv. of 1,000× DFOA (400 μM final concentration) and 0.02 equiv. of 100× BHT (400 μM final concentration). The amount of RNase A to be added can be adjusted according to the RNA content of the cells. If cells are known to contain a high amount of total RNA, we recommend using 0.01 equiv. of RNase A, but 0.002 equiv. is sufficient in most cases. For example, to 1 mL of genomic lysis buffer, add 2–10 μL of RNase A, 2 μL of 1,000× DFOA and 20 μL of 100× BHT. RNase wash buffer must be prepared immediately before use and cannot be stored. ▲ **CRITICAL** If you are interested in any kind of deamination, you should consider adding a broadband deamination inhibitor such as THU.

Washing of stainless-steel beads

To wash the beads, shake them in a soap and water mix for 10 min and rinse them with plenty of water to remove all the remaining soap. Wash them once with acetone and twice in pure ethanol. If necessary, sonicate between washes. Let the beads dry in an oven and let them equilibrate to room temperature before use. The washed beads should be protected from dust and can be stored at room temperature for an infinte amount of time without further washing.

Phosphodiesterase I buffer

Phosphodiesterase I buffer is 5.5 mM Tris (pH 8.9), 5.5 mM NaCl, 0.7 mM $MgCl_2 \cdot H_2O$, 50% (vol/vol) glycerol and 50% (vol/vol) water. The buffer can be stored at −20 °C for long-term storage (up to 5 years) and should be filtered with a syringe filter (0.2-μm cellulose acetate) before use. ▲ **CRITICAL** Sodium (Na^+) adducts form easily, and these can potentially distort the data (the mass of the MOI-Na^+ adduct is not selected). In addition, an abundance of ions in the mass device could enhance matrix effects. However, for this application, Na^+ cannot be avoided and, in our experience, does not cause any problems.

Snake venom phosphodiesterase I

Use a syringe filter (0.2-μm cellulose acetate) to filter 10 mL of phosphodiesterase I buffer. Add 1 mL of this sterile buffer to the phosphodiesterase I pellet and dissolve it by slowly inverting the vessel (enzyme concentration = 100 U/mL). If the enzyme does not dissolve, vortex briefly. Make 100-μL aliquots and store them at −20 °C until usage for up to 1 year.

$ZnSO_4$ stock solution

Prepare a 4 mM solution of $ZnSO_4$ in water. The stock solution can be stored at −20 °C for up to 1 year.

EDTA stock solution

Prepare a 1 mM solution of $Na_2[EDTA]$ in water (pH 8.0). The stock solution can be stored at $-20\,°C$ for up to 1 year. ▲ CRITICAL Na^+ is critical for sensitivity in mass spectrometry; however, for this application, Na^+ cannot be avoided and, in our experience, does not cause any problems.

Nuclease S1 solution

Dilute the stock nuclease S1 (100,000 U/mL) to a concentration of 18,400 U/mL for usage. Mix this dilution by pipetting up and down. For example, add 20.4 μL of water to 4.6 μL of nuclease S1. Nuclease S1 solution must be prepared immediately before use and cannot be stored. ▲ CRITICAL Pipette the enzyme very slowly because of its high viscosity, which is due to its glycerol-containing storage buffer. The nuclease S1 stock solution should be stored for only a few minutes on ice until dilution. Prepare freshly diluted nuclease S1 for each DNA digestion and prepare at least 1 μL more than needed for the mastermix 1.

Mastermix 1

For each sample to be digested, you need 7.5 μL of mastermix 1. Per 7.5 μL of mastermix 1, add the calculated amount of water, $ZnSO_4$ stock solution to a final amount of 3.6 nmol, nuclease S1 solution to a resulting amount of 18.4 U, antarctic phosphatase to a resulting amount of 5 U and specific amounts of labeled internal standards (pH 6.0) (add the reagents in the specified order). See the provided Excel Sheet (Supplementary Methods 3) for details. ▲ CRITICAL The amount of enzyme can be adjusted according to the amount of DNA you want to digest. It is possible to digest at least 10 μg of DNA per sample using this mastermix. Mastermix 1 must be freshly prepared for each digestion and should be stored on ice only for a few minutes before usage. ▲ CRITICAL If you are interested in any kind of deamination, you should consider adding a broadband deamination inhibitor such as THU. This will ensure that any detected deaminated nucleosides are native and not produced due to the digestion conditions, as deaminases are often contaminations of commercially available nucleases.

Mastermix 2

For each sample to digest, you need 7.5 μL of mastermix 2. Per 7.5 μL of mastermix 2, add the calculated amount of water, EDTA stock solution to a final amount of 3.9 nmol and snake venom phosphodiesterase to a resulting amount of 0.15 U (add in the specified order). See the provided Excel Sheet (Supplementary Methods 3) for details. ▲ CRITICAL The enzyme concentrations can be adjusted according to the amount of DNA you want to digest. It is possible to digest at least 10 μg DNA per sample using this mastermix. Mastermix 2 must be freshly prepared for each digestion and should be stored on ice for only a few minutes before usage.

Procedure

General procedure for dissolution and dilution of nucleosides ● Timing 30 min

1   Dissolve a small amount (typically 1-2 mg in 1,000 μL) of each desired nucleoside (natural and isotopically labeled) in a 1.5-mL centrifuge tube in water.
2   Measure the absorption of the solution at the respective wavelength for the maximum absorption of the nucleoside on a photometer.
    ▲ CRITICAL STEP The extinction coefficient must be known for the respective nucleoside (see Supplementary Table 6 for the extinction coefficients of m^5dC, m^4dC, hm^5dC, f^5dC, ca^5dC and m^6dA). If it is unknown, use microscales (Box 1) instead. When using a photometer, make sure that your absorption is within the linear range.
3   Determine the concentration c of the nucleoside using the Beer–Lambert law:

$$c = \frac{E_\lambda}{\varepsilon_\lambda \times d}$$

    with wavelength-dependent extinction E_λ, wavelength-dependent extinction coefficient ε_λ and path length d.
4   Dilute the nucleoside with water to the desired concentration.
    ▲ CRITICAL STEP Very accurate pipettes are needed. Always use the pipette with the smallest margin of error. If you cannot ensure that your pipettes are accurate, use calibrated microscales (Box 1).
    ■ PAUSE POINT The dissolved nucleoside can be stored at $-20\,°C$ for up to 1 year. If you want to store it longer, we recommend storing it at $-80\,°C$.

<div style="border:1px solid #000">

Box 1 | Use of microscales for dilution of nucleosides ● **Timing** 15–30 min

1. Use microscales to weigh small amounts of the natural nucleoside and the isotopically labeled nucleoside into separate tared 1.5-mL centrifuge tubes.
2. Add an appropriate amount of water to each tube and weigh the tube again.
3. Calculate the respective concentrations.
4. If your stock solution is too concentrated and you want to dilute it further, calculate the necessary volume of water and use microscales to control the addition until the desired concentration is reached.

</div>

Preparation of ISTD mastermix ● Timing 1–1.5 h

▲ **CRITICAL** Very accurate pipettes are needed. Always use the pipette with the smallest margin of error. If you cannot ensure that your pipettes are accurate, use calibrated microscales.

5. Thaw all labeled nucleosides needed for the ISTD mastermix for at least 30 min at room temperature, vortex rigorously and spin down (5,000g, room temperature, 3 s).
   ▲ **CRITICAL STEP** If the concentrations of the labeled nucleosides do not match the concentration needed for the ISTD mastermix, dilute them further (or if the concentration is too low, repeat Steps 1-4 for the desired nucleoside).
6. Prepare the ISTD mastermix.
   ▲ **CRITICAL STEP** If your ISTD mastermix deviates from the one given in the 'Reagent setup' section, consider adjusting the solvent, column and method for the UHPLC–MS/MS in Steps 7 and 8.
   ■ **PAUSE POINT** If the ISTD mastermix is not evaluated at once, store it at −20 °C for up to 1 year.

Preparation for UHPLC–MS/MS and performance of sensitivity checks ● Timing 1–1.5 h

▲ **CRITICAL** From now on, the procedure is described for the epigenetic modifications and the provided acquisition method (Supplementary Methods 4) must be applied. For the modifications m^4dC and m^6dA, use MeOH as solvent B, the Poroshell 120 SB C18 as a column and the acquisition methods for m^4dC- and m^6dA-sensitive mode (Supplementary Methods 5) or m^4dC- and m^6dA-insensitive mode (Supplementary Methods 6).

7. Set up the UHPLC–MS/MS system by installing the Poroshell 120 SB C8 column and preparing new buffers for LC. Therefore, add 75 µL of formic acid to a full 1.0-L bottle of MS-grade water (solvent A) and 187.5 µL of formic acid to a full 2.5-L bottle of MS-grade acetonitrile (solvent B). (Alternatively, if solvent B is MeOH, use a full 1.0-L bottle of MS-grade MeOH and add 75 µL of formic acid.)
8. Attach the bottles to the UHPLC system (solvent A to port A, solvent B to port B) and purge at 50% A/50% (vol/vol) B for 5 min with a flow of 5 mL/min. To provide reproducible separation efficiency, the columns must be equilibrated. Therefore, you need to perform 20 chromatographic runs using the method for epigenetic modifications (Table 1) without injection for new columns and three runs before measurement of the first sample per set. The flowrate is 0.35 mL/min, the pressure is 600 bar, the temperature of the column oven of the UHPLC is 35 °C and the gas temperature is 80 °C. The UV detector monitors an absorption signal at the 260-nm wavelength. (Alternatively, if the method for m^6dA and m^4dC is used, see Supplementary Table 7. The temperature of the column oven of the UHPLC is 30 °C; flowrate, pressure and gas temperature are not changed.)
9. If this is the first ever produced ISTD mastermix, add 12 µL to an HPLC vial and analyze it three times, each with an injection volume of 3 µL. Evaluate the ISTD mastermix according to Steps 49–53 and Steps 55 and 56. If a previous ISTD mastermix is available, add 12 µL of the previous one and 12 µL of the new one each to separate HPLC vials. Measure both three times, alternating with an injection volume of 3 µL, and evaluate the respective data. Determine the resulting areas for each labeled nucleoside, calculate the mean of this technical triplicate and evaluate the deviation of the mean as a percentage between the previous ISTD mastermix and the new one.
   ▲ **CRITICAL STEP** The deviation of the resulting areas must be smaller than 5% for each labeled nucleoside in order for this ISTD mastermix to qualify for exact quantification.
10. For the sensitivity check, add a few microliters ($3 + 3 \times n$ µL, n = number of performance checks) of the ISTD mastermix to an HPLC vial.
11. Measure 3 µL of the ISTD mastermix with the provided measuring method, integrate the areas of each labeled nucleoside and compare the areas of the labeled nucleosides and the corresponding SNRs from this measurement with those of the measurements from Step 9.
    ? **TROUBLESHOOTING**

Table 1 | UHPLC gradient elution table for the method for epigenetic modifications

	Time (min)	Solvent A (%)	Solvent B (%)
1	0.00	100.0	0.0
2	4.00	96.5	3.5
3	6.90	95.0	5.0
4	7.20	20.0	80.0
5	10.50	20.0	80.0
6	11.30	100.0	0.0
7	14.00	100.0	0.0

Table 2 | Example of a dilution series from ~10 pmol to ~10 fmol with the corresponding calibration levels L11–1

	L11	L10	L9	L8	L7	L6	L5	L4	L3	L2	L1
n(MOI) [pmol]	10	5	2.5	1.25	0.625	0.313	0.156	0.078	0.039	0.020	0.010
n(MOI*) [pmol]	2.48	2.48	2.48	2.48	2.48	2.48	2.48	2.48	2.48	2.48	2.48
n/n^*	4.037	2.018	1.009	0.505	0.252	0.126	0.063	0.032	0.016	0.008	0.004

Table 3 | Example for setting up the dilution mix and the calibration mix (Calmix)

	Calc. for the highest data point (µL)	Dilution mix		Calmix	
		Factor	Vol (µL)	Factor	Vol (µL)
MOI	2	—	—	8	16
MOI*	4	37	148	8	32
Water			925		184
Total volume			1,073		232

Calculation of calibration curves ● Timing 1 d

12 Decide which molar range the calibration curve should span; the calibration curve will consist of 11 different data points (levels, L) with serial 1:2 dilutions of the natural nucleoside while the amount of the labeled nucleoside is kept constant. Note that the highest data point should contain about four times more of the natural nucleoside (MOI) as compated to the isotopically labeled nucleoside (MOI*) and consider this in the next step. Example: if an amount of 500 fmol of the MOI is expected, span the calibration curve from ~10 fmol to ~10 pmol (Table 2).

13 Calculate which volumes (in µL) of the stock solutions (natural and labeled) equal the amount of substance for the highest point of the calibration curve. Example: The calibration curve should span 10 fmol to 10 pmol; thus the highest data point for MOI is 10 pmol. If the concentration of the stock solution for MOI is 5 µM (5 pmol/µL), a volume of 2 µL of stock solution is needed for the highest data point. If the concentration of the stock solution for MOI* is 0.62 µM, a volume of 4 µL of stock solution is needed in this case to obtain 2.48 pmol (~1/4 of 10 pmol) of MOI* (Table 3).

14 Vortex all nucleoside solutions vigorously (for at least 1 min).

15 Prepare the dilution mix and the Calmix with amounts as calculated in Table 3.

16 Prepare eleven 1.5-mL centrifuge tubes and label them from L1 to L11. Add 100 µL of the dilution mix to each of the tubes L1–L10.

17 Add 100 µL each of the Calmix to L10 and L11. Set L11 aside and vortex L10 vigorously (for at least 1 min).

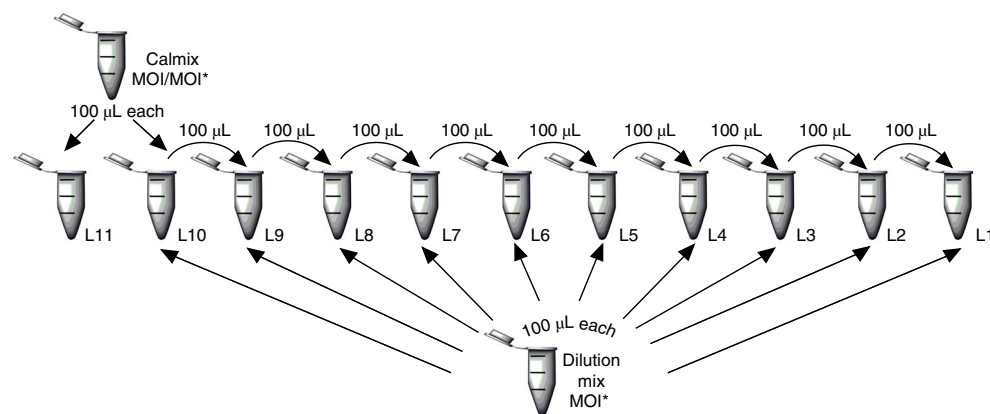18 Make a serial dilution as shown in Fig. 2.

Fig. 2 | Calibration curve preparation. Pipetting scheme for generating a calibration curve of MOI and MOI* with a fixed specific amount of MOI*. L, level.

Table 4 | Example of evaluation of a calibration curve for m⁴dC

Calibration level	n/n^*	$A(m^4dC)$	$A^*(^{15}N_2\text{-}m^4dC)$	A/A^*	Replicate no.
L1	0.004	4,797	1,745,218	0.003	1
L2	0.008	7,214	1,737,978	0.004	
—	—	—	—	—	
L1	0.004	5,094	1,611,819	0.003	2
L2	0.008	6,145	1,658,872	0.004	
—	—	—	—	—	
L1	0.004	4,728	1,689,815	0.003	3
L2	0.008	6,595	1,680,195	0.004	
—	—	—	—	—	

19   Transfer the contents of each tube to an HPLC vial and close the vial with an HPLC vial cap. The injection volume for one measurement is 29 μL. L1 contains the lowest concentration and is therefore the first vial to be measured. Proceed then with measuring L2-11 in their respective order. After measuring each level once, insert a blank measurement into the sample queue without injection. Then repeat the procedure twice, starting from L1 to produce a technical triplicate.

20   For the evaluation, apply Steps 49–53 and 55, integrate the peaks for the MOI (A) and the MOI* (A^*) and transfer the values for A and A^* to an Excel sheet, then calculate A/A^* for each data point. In three additional columns, record the amounts for MOI and MOI* and the respective n/n^* values (Table 4).

21   Calculate the mean, the standard deviation and the standard deviation as percentages (%s.d.) for the A/A^* values of all data points from the technical replicates. If the %s.d. exceeds 20%, the data point is not valid and must be excluded, if it is the highest or lowest data point. Invalid data points in between render the whole calibration curve invalid.

22   Plot the values for the mean of A/A^* value against the respective n/n^* and perform a linear regression of the data points using Origin or similar calculation software. Record the values for m and t.

23   Perform an accuracy check (backfit) by inserting the value of n/n^* into the calculated linear equation and calculating the A/A^* value of every single calibration level. Then determine the percentage of the calculated A/A^* value in comparison to the intended A/A^* value. The resulting percentage provides the accuracy of the linear equation and should not be <80% or >120%.

24   If the accuracy check fails, remove the highest or lowest data point and repeat Steps 22 and 23 with only the remaining data points. Repeat this step until only data points with sufficient accuracy make up the linear equation. A valid linear equation must consist of at least five consecutive levels.
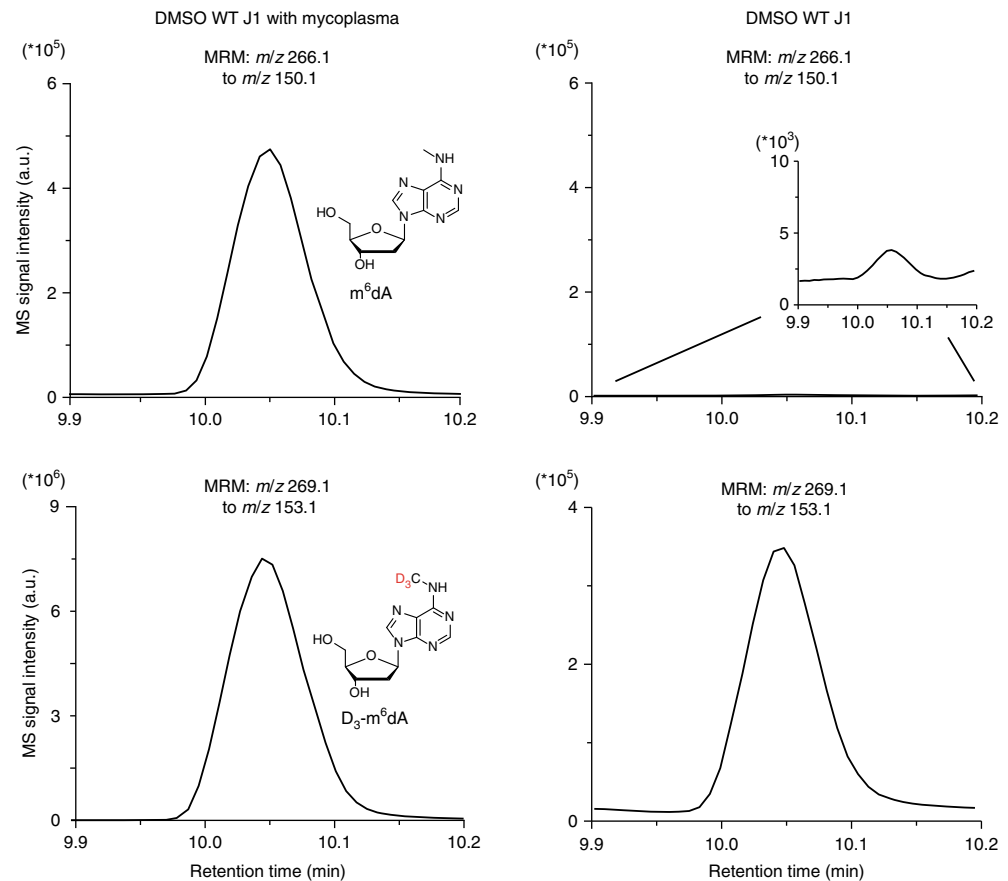
Fig. 3 | Exemplary impairment of results by mycoplasma contamination. Chromatograms with nucleoside signals for the internal standard D$_3$-m^{6}dA (lower row) and natural m^{6}dA (upper row) in 100 ng of DNA from WT J1 mouse embryonic stem cells (mESCs) with a mycoplasma contamination (left column) or 10 µg from mycoplasma-free DNA of WT J1 cultures (right column). a.u., arbitrary units; WT, wild-type.

Isolation of gDNA ● Timing 1.5-3 h

25 Before lysing the cells, wash them once with DPBS and remove all DPBS before proceeding.
▲ CRITICAL STEP If the cells are loosely adherent and will be washed away with DPBS, they can be directly lysed after removing the medium. However, this will lower the yield of isolated gDNA.
▲ CRITICAL STEP Make sure that the cultures are not contaminated with mycoplasma or other microorganisms, as this may substantially alter the abundance of certain DNA modifications (Fig. 3).

26 Add ~200-300 µL of lysis buffer per 10^6 cells. For a typical six-well plate with 75% confluent mouse embryonic stem cells (mESCs; 5–8 × 10^{-6} cells, depending on specific culture conditions), use 1.6 mL of lysis buffer per well. Pipette up and down several times to reduce lysate viscosity and ensure homogeneity. For lysing tissue samples, add ~1.5 mL of lysis buffer to 50 mg of tissue. Transfer the lysate to 2-mL centrifuge tubes.
▲ CRITICAL STEP We recommend direct lysis in the tissue culture vessel because it is convenient and, in direct comparison tests, we have experienced increased quantities of oxidative lesions after harvesting by trypsinization. However, if oxidative lesions are not a major concern, cells can be harvested and further processed, for example, for counting or flow-cytometric analysis, before being lysed as a pellet. Whenever possible, lyse samples before freezing, as the highly chaotropic lysis buffer will inactivate all enzymatic processes. If not avoidable, add lysis buffer directly to frozen cell pellets or frozen soft tissue in very small pieces without letting them thaw. To ensure quick and complete lysis, vortex immediately and thoroughly until no pellet/tissue fragments are visible. Before lysing larger or harder frozen tissue samples, pulverize them using pestle, mortar and liquid nitrogen to prevent them from thawing.
▲ CRITICAL STEP The amount of lysis buffer affects the efficiency of gDNA isolation. Bead milling or sonication is used to further homogenize the lysate and shear the gDNA. If the concentration of

DNA in the lysate is too high (high viscosity), DNA shearing and the final yield of DNA will be reduced because of inefficient elution from the spin column.

■ **PAUSE POINT** Lysate can be stored at −80 °C for up to 5 years. Frozen lysate can be thawed at room temperature before further processing.

27 Add one stainless-steel bead per tube. Further homogenize/shear DNA in the bead mill for 1 min at 30 Hz. For tissue samples, first apply 30 Hz for 1 min and then apply an additional 20 Hz for 4 min.

▲ **CRITICAL STEP** Sonication can also be used for homogenization/DNA shearing. However, we detected increased quantities of abasic sites using this method. If no bead mill or sonicator is available, it is also possible to skip Step 27 and continue directly with Step 29. In this case, use larger volumes of lysis buffer instead and vortex more extensively to provide sufficient DNA shearing, which is important to allow efficient elution from the spin column at the end.

? TROUBLESHOOTING

28 Centrifuge for 5 min at 21,000g at 23 °C to remove the foam (denatured proteins). If there is no foam, the lysis did not work or the number of cells in the tube was very low. If you expect low cell numbers, we highly recommend proceeding directly with Step 29.

29 Load the lysate onto a Zymo-Spin IIC-XL spin column and centrifuge for 1 min at 10,000g at 23 °C. Up to 800 µL of lysate can be loaded at once. If your volume is larger, load several times. If the cells were lysed with Buffer RLT$^+$, the flow-through fraction contains total RNA and denatured proteins, which can be isolated subsequently. In the case of lysis with genomic lysis buffer$^+$, the RNA is already degraded, but the proteins can still be precipitated for subsequent analysis. If desirable, the flow-through can be stored at −80 °C for up to 5 years.

30 Add 400 µL of RNase wash buffer to each spin column and incubate for 10–15 min to make sure that the residual RNA is degraded. After incubation, centrifuge for 2 min at 10,000g at 23 °C, using a new collection tube.

▲ **CRITICAL STEP** The amount of RNA varies greatly from cell type to cell type. RNA impurities can be separated from DNA modifications in the UHPLC–MS/MS analysis. However, it is difficult to estimate the amount of isolated gDNA if the sample is contaminated with a substantial amount of RNA.

? TROUBLESHOOTING

31 From now on, either discard the flow-through by pipetting it out of the collection tube or use a new collection tube for each step.

▲ **CRITICAL STEP** Take care when removing the collecting tubes of the centrifuge. The flow-through should neither touch the end of the spin column nor become trapped between the spin column and the collection tube, as this might lead to alcohol or salt contaminations in the eluate that will massively impair the subsequent UHPLC–MS/MS analysis.

32 Add 400 µL of DNA pre-wash buffer per spin column and centrifuge for 1 min at 10,000g at 23 °C. Discard the flow-through.

? TROUBLESHOOTING

33 Add 600 µL of gDNA wash buffer per spin column and centrifuge for 1 min at 10,000g at 23 °C. Discard the flow-through and repeat this step at least once.

? TROUBLESHOOTING

34 Put the column into a new collection tube and centrifuge for 1 min at 10,000g at 23 °C to remove all the wash buffer.

35 To elute, set the spin column into a new 1.5-mL centrifuge tube, add 50–150 µL of water supplemented with 0.001 equiv. of 100× BHT directly to the matrix of the column and incubate for 10 min. Centrifuge for 2 min at 10,000g at 23 °C.

36 Measure the gDNA concentration with a photometer. The optical density (OD)$_{260}$/OD$_{280}$ ratio should be between 1.85 and 1.90, and the OD$_{260}$/OD$_{230}$ ratio should be >2, ideally between 2.3 and 2.5.

▲ **CRITICAL STEP** If a sample is contaminated with RNA, the OD$_{260}$/OD$_{280}$ ratio is often >1.9. If the OD$_{260}$/OD$_{230}$ ratio is too low, the sample probably contains salt impurities. High concentrations of BHT can also result in a low OD$_{260}$/OD$_{230}$ ratio, but they do not impair the UHPLC–MS/MS analysis. Note: If the amount of RNA impedes determination of DNA content, PicoGreen can be used to clearly determine the content of double-stranded DNA.

? TROUBLESHOOTING

■ **PAUSE POINT** Isolated gDNA can be kept at −80 °C for an infinite amount of time. For short-time storage, −20 °C (1–7 d) or even 4 °C (up to 1 d) is suitable. If you freeze your isolated gDNA, thaw your samples on ice before starting the DNA digestion steps. After thawing, vortex the samples thoroughly before use and do a quick spin (5,000g, 23 °C, 5 s) in order to collect the solution at the bottom.

Table 5 | Example of calculating the amount of gDNA and water for the DNA digestion

Sample ID	c (ng/µL)	m (µg)	V (H$_2$O) (µL)	V (DNA) (µL)	Sample no.
1_Blank 1	—	—	35.0	0	1
2_Blank 2					2
3_Blank 3					3
4_1.DNA_1	444.61	10.0	12.5	22.5	4
5_1.DNA_2					5
6_1.DNA_3					6
7_2.DNA_1	405.34	10.0	10.3	24.7	7
8_2.DNA_2					8
9_2.DNA_3					9

V, volume.

Box 2 | Testing digestion conditions for an unknown sample ● Timing 8 h

1 Produce a single technical replicate, instead of a technical triplicate as described in Step 37, by digesting 1–5 µg of DNA. Then continue with Step 38.
2 Check if all requirements of the protocol are fulfilled:
    (i) Complete digestion of DNA,
    (ii) All MOIs and MOI*s are detectable,
    (iii) All MOIs can be evaluated with the existing calibration curves.
If there are problems with (i), your sample contains impurities (e.g., salt) or shows an unfavorable pH. See the Troubleshooting section (incomplete digestion).
If there are problems with (ii):

Neither the MOI* nor the corresponding MOI* is detectable: see the Troubleshooting section (peak shifting).

Only the MOI*, but not the corresponding MOI, is detectable: increase the amount of DNA and check the pH.

If there are problems with (iii): first, try to adjust the amount of DNA and/or the amount of MOI*. If this does not solve the problem, generate a new calibration curve.
? TROUBLESHOOTING

DNA digestion ● Timing 7 h or overnight

37 Create a digestion sheet (Supplementary Methods 3) to easily calculate the required volume of DNA and water per sample. Each DNA sample should be processed from now on as a technical triplicate. Your sheet starts with samples 1–3 as digestion blanks, which contain no DNA, but to which mastermixes 1 and 2 will be added. For all the other samples, DNA and water should have a total volume of 35 µL (Table 5). Number the samples consecutively, as these will help you to identify your samples afterwards.
    ▲ CRITICAL STEP The volume of DNA must not exceed 35 µL per sample.
38 Start with the calculated volume of water, and add the DNA according to the digestion sheet.
    ▲ CRITICAL STEP Use fresh water for each digestion. Autoclaved water or water that has been stored for a long time might contain dust or additional contaminations that could impair the digestion or the measurements.
    ▲ CRITICAL STEP If this is the first digestion of your DNA sample and/or you are interested in a new DNA modification, you should consider testing your digestion protocol (Box 2). This may help to determine the appropriate addition of labeled nucleosides, e.g., spiking and the proper amount of DNA in general for fulfilling the requirements of the calibration curves.
39 Create a pipetting scheme for your mastermixes 1 and 2, using our provided Excel template (Supplementary Methods 3). Calculate the number of DNA samples (Step 37) you want to digest and add three samples more as spares.
    ▲ CRITICAL STEP Instead of following the two-step digestion protocol, which was designed to give an optimal SNR in the MS, you can use Degradase Plus from Zymo Research, which requires only one digestion step. See Box 3 for details.

> **Box 3 | DNA digestion with Degradase Plus (replaces Steps 39–42)** ● **Timing 5 h**
>
> 1 Create a pipetting scheme for the Degradase mastermix using our provided Excel template for Degradase digestion (see the tabs for 'Degradase digestion' in Supplementary Methods 3). Calculate the number of DNA samples and add at least one more sample as a spare.
> 2 Prepare the Degradase mastermix, add 15.0 µL to each sample and incubate for 4 h at 37 °C.
> When using Degradase Plus, adjust the enzyme concentration if >5 µg of DNA/sample must be digested (Supplementary Methods 3). The additional use of benzonase is required for residual RNA content within your DNA sample. The enzyme mixture in Degradase Plus cannot digest RNA. If benzonase is not used, this will lead to an overpressure in the LC device due to residual undigested RNA, and subsequent clogging of the column. However, there is no need to add the enzyme benzonase in the case of digestion of pure DNA, e.g., synthetic DNA oligonucleotides.

40   Prepare mastermix 1, then add 7.5 µL of it to each sample and incubate for 3 h at 37 °C.

41   Before the incubation time for Step 40 runs out, start preparing mastermix 2.

42   Add 7.5 µL of mastermix 2 to each sample and incubate for 3 h at 37 °C.
    ▲ CRITICAL STEP If >5 µg of DNA/sample is digested, incubate overnight at 37 °C.
    ■ PAUSE POINT Samples can be stored at 4 °C (overnight) or at −20 °C for long-term (up to 5 years) storage.

Sample preparation for UHPLC–MS/MS analysis ● Timing 1 h

43   Tape an AcroPrep Advance 96-well Supor plate (0.2 µm) to a skirted 96-well PCR plate.

44   After the digestion, pipette the full 50 µL of digestion mixture for each sample into a separate well of the assembled filtration plate.

45   Cover the filtration plate with Parafilm M and secure it with tape.
    ▲ CRITICAL STEP Water should not enter the plate from an external source, but the Parafilm M should not seal the plate airtight.

46   Centrifuge the filtration plate, using a plate rotor for 35 min at 3,220g at 4 °C.
    ? TROUBLESHOOTING

47   Transfer 43 µL of the filtered digestion mixture to an HPLC vial and close the vial with an HPLC vial cap. If the volume of the filtered digestion mixture is <43 µL, note which volume is transferred to the vial.
    ▲ CRITICAL STEP The unused wells of the filtration plate can be used for later filtrations. For storage, seal the filtration plate, which is taped to the 96-well PCR tube plate, with Parafilm M and store it at room temperature for an infinite amount of time.

48   For each sample from Step 47, inject 39 µL of sample into the UHPLC–MS/MS system when the machine is ready. Record the signal of the MOIs with the relevant method for your MOIs (Supplementary Methods 4 for epigenetic modifications or Supplementary Methods 5 for m^{6}dA- and m^{4}dC-sensitive or Supplementary Methods 6 for m^{6}dA- and m^{4}dC-insensitive). For details about the measurement methods, see Table 1 and Supplementary Tables 3, 5 and 7.
    ▲ CRITICAL STEP Inject 3 µL less than the transferred volume, if the transferred volume is <43 µL.
    ? TROUBLESHOOTING

Evaluation ● Timing 20 min–1.5 h

49   Create a new folder called 'QuantResults' in the folder with the measured data.

50   Open the program QQQ Quantitative Analysis and click on 'File' > 'New Batch' and label the batch according to your experiment name/number.

51   Import all samples of interest.

52   Set up a method for analysis. Click on 'Method' > 'New' > 'New Method' under 'Acquired MRM Data' > 'Select a sample which was measured with this method' > 'Method Setup Tasks' > 'MRM Compound Setup'. Add UV, UV-dG and UV-dC as new compounds in the table (right-click on the table). Set the time segment 'TS' to '3' and the 'Type' to 'Target'. Enter the appropriate RT in the 'RT setup'. For UV, set it to 6 min with a 'Left RT Delta' of 6 min and a 'Right RT Delta' of 6 min. Go next to 'ISTD Setup'. In the 'ISTD conc.' column, add '1' for the ISTD concentration, make a tick in the 'ISTD Flag' column for each ISTD and select the appropriate ISTD for each target in the 'ISTD Compound Name' column. Finally, select 'Advanced Tasks' > '2D Compound Setup' and then select 'VWD' for the compounds UV, UV-dC and UV-dG. Click on 'Validate' to ensure that
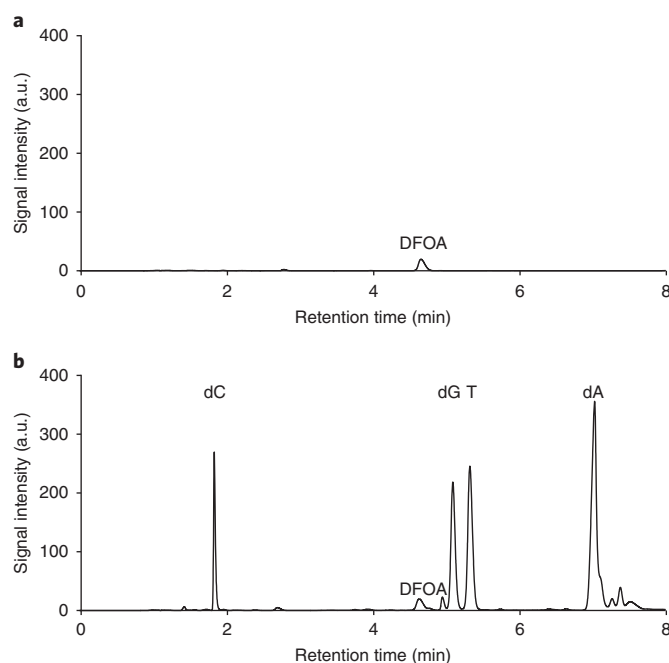
Fig. 4 | Chromatogram of different UV traces as represented in the QQQ quantitative analysis program. a,b, The chromatogram of a digestion blank sample (**a**) and the chromatogram of a gDNA sample from mESC line J1 (**b**).

your method does not have any errors. If errors or warnings occur, the program will tell you where the problem is.

▲ **CRITICAL STEP** The protocol described here uses the provided Excel sheets (Supplementary Methods 1 and 2) for quantification. Therefore, the concentration that you define in the 'ISTD conc.' column is not relevant.

53 Click on 'Exit' and let the program analyze your batch.

54 Select at first for the compound 'UV': click on 'manual integration' and start with the integration of dG (RT ~5.3 min for method with buffer B: MeCN or ~8.3 min for method with buffer B: MeOH) for all samples. Click in the table, then select 'File' > 'Export' > 'Export Table' (this will open an Excel sheet, which should be saved as 'dG'). Continue with the integration of the dC Peak (RT ~1.9 min for method with buffer B: MeCN or ~1.8 min for method with buffer B: MeOH).

▲ **CRITICAL STEP** In the case that you also need dA (RT ~7.3 min for method with buffer B: MeCN or ~9.1 min for method with buffer B: MeOH) and T (RT ~5.5 min for method with buffer B: MeCN or ~8.8 min for method with buffer B: MeOH), additionally integrate those peaks.

▲ **CRITICAL STEP** The digestion blanks should not show any peaks for dA, dC, dG or T in the UV trace. If you see any peaks at the anticipated RT, your blanks are probably contaminated with DNA, and the whole measurement is invalid. Peaks at different RTs, not overlapping with dA, dC, dG and T, in the UV trace of the digestion blank are impurities or small molecules from the digestion mix (e.g., DFOA) that can be ignored (Fig. 4). RNA impurities in the DNA samples should not overlap with dA, dC, dG and T peaks either.

55 In the Batch table, select the first compound measured in the Signal type MS. The upper row shows the peak for the natural nucleoside (MOI), e.g., the one defined as Target and the lower row shows the peak for the labeled standard (MOI*), e.g., the one defined as ISTD (Fig. 5). Integrate both peaks the same way. This is very important for the reliability of the method. Export the Batch table. This will create an Excel sheet automatically, which you should label according to the MOI.

▲ **CRITICAL STEP** Carefully integrate the peaks of your digestion blanks, e.g., samples 1-3, even if they do not contain any DNA. These samples will serve as your negative control and will reveal the limits of quantification and detection; see Experimental setup.

56 Repeat this procedure for all nucleosides (MOI) of interest (Fig. 6).

? **TROUBLESHOOTING**

57 Use the provided Excel sheet (Supplementary Methods 1 or 2) for the evaluation of your data. For this, you need to copy the values of the peak areas (for A (MOI) and A^* (MOI*)) resulting from the
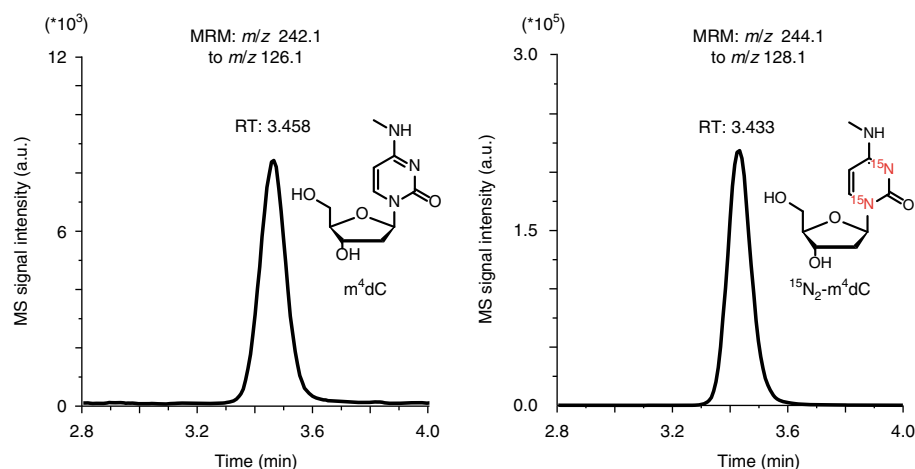
Fig. 5 | Chromatograms for m4dC and 15N2-m4dC. The left panel shows the nucleoside signal for the natural m^4dC and the right panel shows the nucleoside signal for the internal standard $^{15}N_2$-m^4dC.
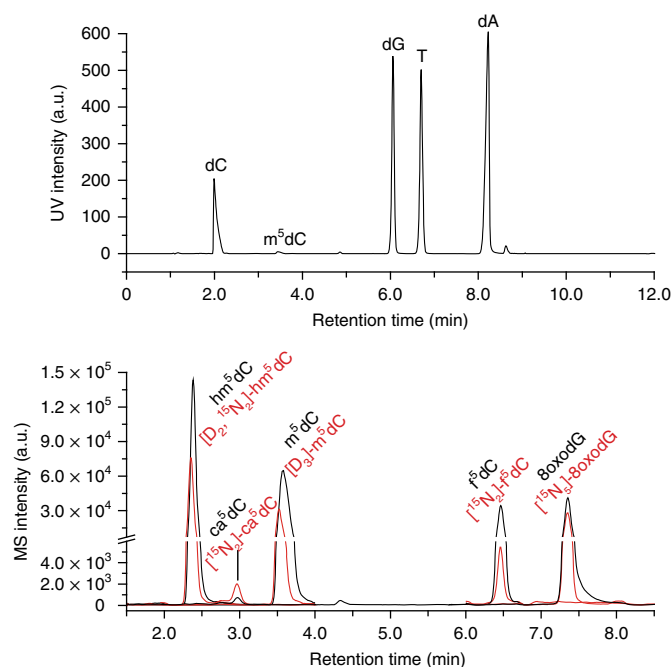
Fig. 6 | Chromatogram showing the analysis of hm^5dC, ca^5dC, m^5dC, f^5dC and 8oxodG with their corresponding isotopically labeled standards. The upper chromatogram shows the UV trace and the lower chromatogram shows the MS trace[11].

integration of Steps 54–56 from the created Excel sheet into the evaluation Excel sheet (Supplementary Methods 1 or 2).

▲ **CRITICAL STEP** Before processing the data, you must calculate the LOD of every single MOI by analyzing the SNR. Furthermore, the ratios of A/A^* and thereby the amount n of the MOI need to meet the requirements of the corresponding calibration curve. Valid data that are consistent with these conditions can be processed further and will provide the amount of the MOI in pmol within one sample, more specifically within the volume of injection (39 μL). If you did not inject 39 μL for the measurement of a sample, you need to adjust this value in the provided Excel sheet for this sample. This additional calculation will provide comparable data for your technical and later biological replicates. The injection volume can be looked up with the QQQ Quantitative Analysis program.

Troubleshooting

Troubleshooting advice can be found in Table 6.

Table 6 | Troubleshooting table

Step	Problem/Observation	Possible reason	Solution
11	The sensitivity of the mass spectrometer is too low: The sensitivity check for the ISTD mastermix shows substantially smaller peaks than in previous measurements	(i) Dirty ionization source (ii) Old buffers (iii) Old column (iv) Molecule insoluble (pH variation)	(i) Clean ionization source (ii) Exchange buffers (iii) Exchange column (iv) Check and adjust pH (v) Check sensitivity between (i)–(iv)
11, 32, 33 and 56; Box 2	Peak shifting: The peak of interest occurs at a different retention time	(i) Impurities (e.g., salt) (ii) pH variation (iii) Column batch shows different properties	(i) Wash samples more rigorously during DNA isolation (ii) Check pH regularly (iii) Adjust time segments, if modifications escape
11, 46 and 56	Peak splitting: The chromatogram for a modification shows an unexpected number of signals; usually one is at the expected retention time and a second one elutes at an earlier retention time. The two signals might be connected	(i) The filtration was not successful (ii) The inlet filter is clogged (iii) The nucleoside was only partially charged during the chromatography and is therefore eluting at two different time points	(i) Optimize the filtration step (ii) Use guard columns or inlet filters (iii) Check the pH of the sample
11 and 56	Peak broadening	(i) Amounts of modification are too high (ii) Not enough data points in peak(s) (iii) Poor performance of the column	(i) Try setting 'MS1 resolution' or 'MS2 resolution' to 'enhanced' in the method (ii) Decrease dwell time to receive more cycles per second (iii) Test a new column
	Two molecules cannot be baseline-separated: The signal for the MOI shows at least a shoulder or a clear second maximum	(i) Similar retention time due to physical properties of the molecule (often in the UV trace, but also possible with MS analysis; e.g., in isomers) (ii) Altered performance of the column, which is possible after change of the column batch	(i) Switch the machine setup by testing a different column or different buffer system (ii) Adjust column temperature and check whether molecules with the same mass signal still elute in the same order
27 and 36	Low yield of DNA: Yield of DNA after isolation is lower than expected	(i) Insufficient DNA shearing (ii) Spin column did not bind DNA (iii) Spin column was overloaded (iv) Insufficient DNA elution from column	(i) Use more lysis buffer (ii) Test a different batch of spin columns (iii) Distribute the lysate on more spin columns (iv) Increase the volume for elution or elute twice (i,iv) Use the Blood & Cell Culture DNA Midi Kit from Qiagen (anion-exchange columns allow elution of larger gDNA fragments than spin columns)
30, 36 and 56	Too much RNA in DNA preparation: (i) The ratios of the absorption at 260 nm and 280 nm are >2 (ii) The UV chromatogram shows additional peaks	(i) Cell line contains higher amounts of RNA (ii) The amount or performance of the RNase was insufficient	Adjust the amount of RNase during the DNA isolation and increase the incubation time
36	Low DNA concentration: DNA volume for digestion exceeds 35 µL	Elution volume too high	Precipitate DNA (Supplementary Methods 7) and re-dissolve in a smaller volume
48, Box 2	The pressure of the column increases during one sample set: Between samples of the same dataset, an increase in the column pressure is observed	(i) Incomplete digestion (ii) Salt impurities	(i) Exchange the inlet filter (ii) Add runs without injection to the worklist
56, Box 2	Incomplete digestion: (i) No or substantially smaller signals for the canonical nucleosides are observed in the UV trace (ii) Sometimes additional UV signals appear at early (~1 min) or late (~10 min) retention times (iii) The pressure of the column increases due to clogging	(i) Enzyme performance is reduced (ii) Enzyme performance is reduced and therefore residual single nucleotides or oligonucleotides exist (iii) pH variation of the sample and/or salt impurities	(i) Test enzymes regularly with appropriate amounts of DNA (ii) See (i) (iii) Test pH compatibility of sample and enzyme and/or wash DNA more often with DNA pre-wash and gDNA wash buffer during the isolation

Timing

Steps 1–4, general procedure for dissolution and dilution of nucleosides: 30 min

Steps 5–11, preparation of ISTD mastermix and UHPLC–MS/MS system, and performance of sensitivity checks: 2–3 h

Steps 12–24, calculation of calibration curves: 1 d

Steps 25–36, isolation of gDNA: 1.5–3 h

Steps 37–42, DNA digestion: 7 h or overnight

Steps 43–48, sample preparation for UHPLC–MS/MS analysis: 1 h

Steps 49–57, evaluation: 20 min–1.5 h

Box 1, use of microscales for dilution of nucleosides: 15–30 min

Box 2, testing digestion conditions for an unknown sample: 8 h

Box 3, DNA digestion with Degradase Plus: 5 h

Anticipated results

The yield of gDNA depends on the cell/tissue type and isolation method. We adapted the procedure delineated in the Quick-gDNA MidiPrep Kit in combination with Zymo-Spin IIC-XL spin columns from Zymo Research. Using this method, ~6 or 10 μg of gDNA can be isolated from 10^6 human embryonic kidney 293T (HEK293T) cells or 10^6 mESCs, respectively. The yield of gDNA from 1 mg of mouse cerebellum will be ~1 μg per mg of tissue. The maximal binding capacity of the Zymo-Spin IIC-XL spin columns is 20–25 μg of gDNA. If more gDNA is required, we recommend using a proportional number of columns and combining the eluates at the end. Earlier, we had successfully used spin columns from Zymo Research with higher binding capacity. These columns yielded DNA of good quality, but recent production batches seemed to contain impurities, which perturbed our LC–MS/MS setup, thus leading to a loss of sensitivity. Note, however, that the sensitivity to impurities may vary with the specific UHPLC–MS/MS equipment. As an alternative to spin column–based isolation, we have successfully used gravity flow columns with anion exchange properties (Qiagen Blood & Cell Culture DNA Midi Kit). This method is recommended for abasic site analysis[44], as it generates lower background quantities of abasic sites, probably as a result of the gentler elution procedure, but it takes much more time than the spin column–based method. An additional advantage of the spin column–based method is that total RNA and (denatured) protein can be subsequently isolated from the very same lysate.

The protocol described above details the steps needed to quantify DNA modifications and lesions such as m^5dC, hmdC or 8oxodG. The levels of 8oxodG in mammalian cells are probably between 0.07 and 0.9×10^{-6} 8oxodG/dN, based on the average of the results from several laboratories and various methods[68]. The background level of 8-oxodG has been estimated at 0.5 lesions per 10^{-6} dN[69,70]. Depending on the modification, there are big differences in the abundance of certain modifications between different cell lines or culture conditions. Whereas the abundance of m^5dC for most
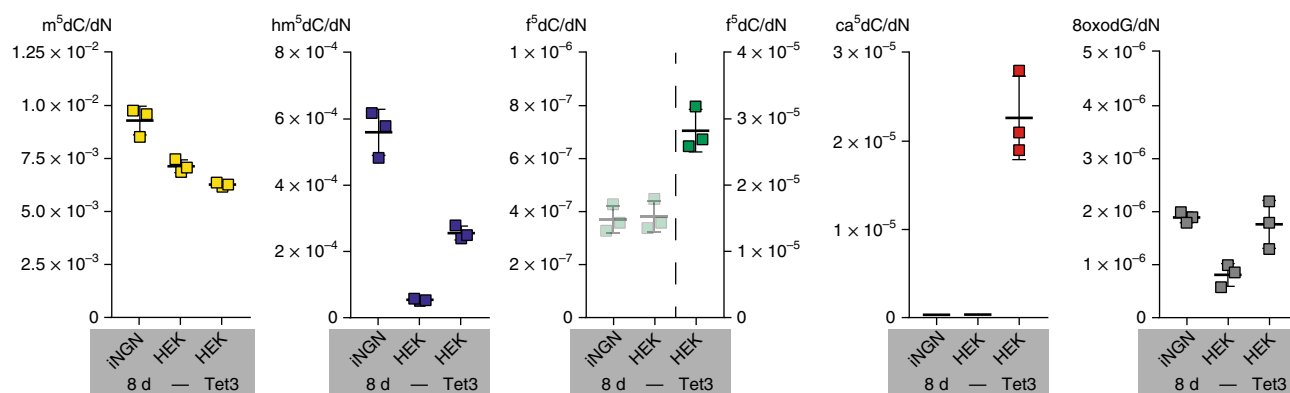
Fig. 7 | Quantities of m^5dC, hm^5dC, f^5dC, ca^5dC and 8oxodG in small molecule–inducible neurogenin human-induced pluripotent stem cells lines (iNGNs) (8 d after induction of differentiation), untransfected HEK293T cells and HEK293T cells that were transfected for 24 h with a plasmid coding for Tet3. The transparent data points at f^5dC for iNGNs and untransfected HEK293T cells indicate that f^5dC was above the LLOQ but did not meet the strict LOD$_{practical}$ criterion. Each data point represents the mean of one biological replicate that was measured in three technical replicates; the bar represents the mean of the biological replicates and error bars represent s.d.

Table 7 | Measurements for an exemplary calibration curve for m⁴dC in technical triplicates

| Calibration level	$A(\text{m}^4\text{dC})$	$A^*(^{15}\text{N}_2\text{-m}^4\text{dC})$	A/A^*	Replicate
L1	4,797	1,745,218	0.0027	1
L2	7,213	1,737,978	0.0042	
L3	10,664	1,692,721	0.0063	
L4	19,426	1,670,953	0.0116	
L5	34,395	1,677,652	0.0205	
L6	74,308	1,646,471	0.0451	
L7	164,901	1,566,133	0.1053	
L8	379,446	1,467,230	0.2586	
L9	869,825	1,340,740	0.6488	
L10	1,960,129	1,181,613	1.6589	
L11	4,164,191	971,267	4.2874	
L1	5,094	1,611,819	0.0032	2
L2	6,145	1,658,872	0.0037	
L3	10,207	1,658,337	0.0062	
L4	18,928	1,586,681	0.0119	
L5	34,678	1,570,039	0.0221	
L6	73,600	1,534,700	0.0480	
L7	163,872	1,558,445	0.1052	
L8	368,404	1,454,277	0.2533	
L9	875,082	1,321,236	0.6623	
L10	1,914,969	1,189,553	1.6098	
L11	4,264,798	885,407	4.8168	
L1	4,728	1,689,815	0.0028	3
L2	6,595	1,680,195	0.0039	
L3	10,841	1,677,259	0.0065	
L4	18,867	1,670,998	0.0113	
L5	35,859	1,696,961	0.0211	
L6	72,444	1,694,990	0.0427	
L7	169,019	1,543,636	0.1095	
L8	386,871	1,510,771	0.2561	
L9	897,343	1,315,416	0.6822	
L10	1,994,610	1,225,289	1.6279	
L11	4,063,788	944,247	4.3037	

mammalian cell lines or tissues is normally in a range between 0.6×10^{-2} and 1.0×10^{-2} m⁵dC/dN, modifications that are less abundant can deviate by factors of hundreds or thousands, depending on the biological background of the sample (e.g., organism, culture conditions).

Figure 7 shows typical quantities of m⁵dC, hm⁵dC, f⁵dC, ca⁵dC and 8oxodG for iNGNs obtained 8 d after differentiation, for untransfected HEK293T cells and HEK293T cells that were transfected with a plasmid coding for Tet3 and incubated for an additional 24 h. For all the cell lines shown, 2 μg of gDNA was used per technical replicate for the digestion. As m⁵dC is a very abundant DNA modification, only very little amounts of gDNA are needed for detection and quantification, and 2 μg is by far sufficient to exceed the LLOQ and the LOD_practical. Also, for hm⁵dC, for which the typical amount is in the range of 10^{-4} hm⁵dC/dN, 2 μg of gDNA is more than enough for the quantification, even in cell lines that have typically low quantities of this modification, such as HEK293T cells. However, for less abundant modifications, here f⁵dC and ca⁵dC, 2 μg is often not enough to meet the LLOQ or LOD_practical criterion. Although the detected f⁵dC levels in iNGNs and untransfected HEK293T cells were above the LLOQ, these results cannot be used for further interpretation, because in this case the background signal was too high and the samples therefore did not meet the LOD_practical criterion. For ca⁵dC, the abundance was above neither the LOD_practical nor the LLOQ. One possible way to overcome this problem is to use more gDNA for the quantification. If you are interested only in a certain modification, you can also try to optimize the UHPLC–MS/MS system for

Table 8 | Calculations for an exemplary m⁴dC calibration curve

Calibration level	Mean ($A/A^\star$)	s.d. ($A/A^\star$)	%s.d. ($A/A^\star$)	Backfit (invalid)	Accuracy (invalid)	Backfit	Accuracy
L1	0.0029	0.0002	7.7	0.0010	35	0.0025	
L2	0.0039	0.0002	5.7	0.0030	76	0.0038	105
L3	0.0063	0.0002	2.4	0.0069	110	0.0064	99
L4	0.0116	0.0003	2.8	0.0148	128	0.0115	101
L5	0.0212	0.0008	3.8	0.0306	144	0.0218	97
L6	0.0453	0.0026	5.8	0.0622	137	0.0424	107
L7	0.1066	0.0025	2.3	0.1254	118	0.0835	
L8	0.2560	0.0026	1.0	0.2519	98	0.1658	
L9	0.6644	0.0168	2.5	0.5047	76	0.3305	
L10	1.6322	0.0248	1.5	1.0103	62	0.6597	
L11	4.4693	0.3010	6.7	2.0216	45	1.3181	

$y = 0.32621x + 0.00123$
$R^2 = 0.99877$
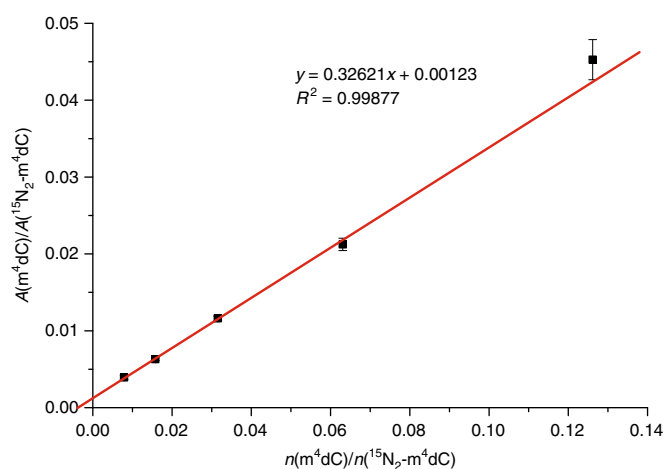
Fig. 8 | Exemplary calibration curve for m4dC. Three technical replicates were measured per data point. The mean is plotted and error bars represent s.d.

this modification; however, this might impair the detection of other modifications. It is not possible to avoid completely the generation of 8oxodG during the whole sample preparation process, and cells always show a low background of oxidative damage. However, the 8oxodG quantity (unless cells were put under (oxidative) stress) should not exceed the low 10^{-6} 8oxodG/dN range.

Evaluation of calibration curve

Table 7 shows an exemplary evaluation of the calibration curve of m⁴dC as described in the Procedure. Consideration of all data points for linear regression would result in the following equation:

$$y = 0.50101x - 9.60772 \times 10^{-4}.$$

Applying a so-called backfit (Step 23) results in the values shown in Table 8, column 5. These values then must be divided by the mean for $A/A^\star$ (Table 8, column 2) and multiplied by 100 to provide the accuracy of the equation (Table 8, column 6) as a percentage. As described in Step 23, the accuracy must be between 80 and 120%, but with consideration of all data points, e.g., levels, the accuracy varies highly and does not meeting the requirements. Therefore Step 23 must be applied.

In this case, Step 24 had to be performed multiple times to yield a calibration curve consisting only of data points with sufficient accuracy and therefore a valid linear equation (Fig. 8). Calibration levels 2-6 (see Table 8, columns 7 and 8) provide the lower and upper limits of quantification for n of 18.9

and 302 fmol, respectively, with the lower and upper limits for A/A^* in the range between 0.0042 and 0.0451.

Ethics statement

Tissue material was provided by the animal facility of the Department of Pharmacy at Ludwig-Maximilians–Universität München, which is approved by the government of Upper Bavaria.

Reporting Summary

Further information on research design is available in the Nature Research Reporting Summary.

References

1. Kriaucionis, S. & Heintz, N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **324**, 929–930 (2009).
2. Tahiliani, M. et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930–935 (2009).
3. Pfaffeneder, T. et al. The discovery of 5-formylcytosine in embryonic stem cell DNA. *Angew. Chem. Int. Ed. Engl.* **50**, 7008–7012 (2011).
4. Ito, S. et al. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333**, 1300–1303 (2011).
5. He, Y.-F. et al. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333**, 1303-1307 (2011).
6. Wu, H. & Zhang, Y. Mechanisms and functions of Tet protein-mediated 5-methylcytosine oxidation. *Gene. Dev.* **25**, 2436–2452 (2011).
7. Branco, M. R., Ficz, G. & Reik, W. Uncovering the role of 5-hydroxymethylcytosine in the epigenome. *Nat. Rev. Genet.* **13**, 7–13 (2011).
8. Langemeijer, S. M. C. et al. Acquired mutations in TET2 are common in myelodysplastic syndromes. *Nat. Genet.* **41**, 838–842 (2009).
9. Ko, M. et al. Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature* **468**, 839–843 (2010).
10. Jones, P. A. & Baylin, S. B. The epigenomics of cancer. *Cell* **128**, 683-692 (2007).
11. Pfaffeneder, T. et al. Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA. *Nat. Chem. Biol.* **10**, 574–581 (2014).
12. Bachman, M. et al. 5-Formylcytosine can be a stable DNA modification in mammals. *Nat. Chem. Biol.* **11**, 555–557 (2015).
13. Iwan, K. et al. 5-Formylcytosine to cytosine conversion by C–C bond cleavage in vivo. *Nat. Chem. Biol.* **14**, 72–78 (2018).
14. Raiber, E.-A. et al. 5-Formylcytosine alters the structure of the DNA double helix. *Nat. Struct. Mol. Biol.* **22**, 44–49 (2014).
15. Song, C. X. et al. Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. *Cell* **153**, 678–691 (2013).
16. Kellinger, M. W. et al. 5-formylcytosine and 5-carboxylcytosine reduce the rate and substrate specificity of RNA polymerase II transcription. *Nat. Struct. Mol. Biol.* **19**, 831–833 (2012).
17. Zhu, C. et al. Single-cell 5-formylcytosine landscapes of mammalian early embryos and ESCs at single-base resolution. *Cell Stem Cell* **20**, 720–731.e5 (2017).
18. Hill, P. W. S., Amouroux, R. & Hajkova, P. DNA demethylation, Tet proteins and 5-hydroxymethylcytosine in epigenetic reprogramming: an emerging complex story. *Genomics* **104**, 324–333 (2014).
19. Yu, M. et al. Base-resolution detection of N(4)-methylcytosine in genomic DNA using 4mC-Tet-assisted-bisulfite sequencing. *Nucleic Acids Res.* **43**, e148 (2015).
20. Ehrlich, M. et al. DNA methylation in thermophilic bacteria: N4-methylcytosine, 5-methylcytosine, and N6-methyladenine. *Nucleic Acids Res.* **13**, 1399–1412 (1985).
21. Arber, W. & Dussoix, D. Host specificity of DNA produced by *Escherichia coli*: I. Host controlled modification of bacteriophage λ. *J. Mol. Biol.* **5**, 18–36 (1962).
22. Wu, T. P. et al. DNA methylation on N6-adenine in mammalian embryonic stem cells. *Nature* **532**, 329–333 (2016).
23. 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 (2015).
24. Koziol, M. J. et al. Identification of methylated deoxyadenosines in vertebrates reveals diversity in DNA modifications. *Nat. Struct. Mol. Biol.* **23**, 24–30 (2015).
25. Liu, J. et al. Abundant DNA 6mA methylation during early embryogenesis of zebrafish and pig. *Nat. Commun.* **7**, 13052 (2016).
26. Schiffers, S. et al. Quantitative LC–MS provides no evidence for m⁶dA or m⁴dC in the genome of mouse embryonic stem cells and tissues. *Angew. Chem. Int. Ed. Engl.* **56**, 11268–11271 (2017).
27. Wang, D., Kreutzer, D. A. & Essigmann, J. M. Mutagenicity and repair of oxidative DNA damage: insights from studies using defined lesions. *Mutat. Res.* **400**, 99–115 (1998).

28. Cooke, M. S., Evans, M. D., Dizdaroglu, M. & Lunec, J. Oxidative DNA damage: mechanisms, mutation, and disease. *FASEB J.* **17**, 1195–1214 (2003).

29. Lunec, J. ESCODD: European standards committee on oxidative DNA damage. *Free Radical Res.* **29**, 601–608 (1998).

30. Murray, K. K. et al. Definitions of terms relating to mass spectrometry (IUPAC Recommendations 2013). *Pure Appl. Chem.* **85**, 1515–1609 (2013).

31. Oakeley, E. J. DNA methylation analysis: a review of current methodologies. *Pharmacol. Ther.* **84**, 389–400 (1999).

32. 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. *Nat. Protoc.* **3**, 1287–1298 (2008).

33. Annesley, T. M. Ion suppression in mass spectrometry. *Clin. Chem.* **49**, 1041–1044 (2003).

34. Buhrman, D. L., Price, P. I. & Rudewiczcor, P. J . Quantitation of SR 27417 in human plasma using electrospray liquid chromatography–tandem mass spectrometry: a study of ion suppression.*J. Am. Soc. Mass Spectrom* **7**, 1099–1105 (1996).

35. Brückl, T., Globisch, D., Wagner, M., Müller, M. & Carell, T. Parallel isotope-based quantification of modified tRNA nucleosides. *Angew. Chem. Int. Ed. Engl.* **48**, 7932–7934 (2009).

36. Münzel, M. et al. Quantification of the sixth DNA base hydroxymethylcytosine in the brain. *Angew. Chem. Int. Ed. Engl.* **49**, 5375–5377 (2010).

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

38. Pearson, D. et al. LC-MS based quantification of 2′-ribosylated nucleosides Ar(p) and Gr(p) in tRNA. *Chem. Commun.* **47**, 5196–5198 (2011).

39. Globisch, D. et al. Systems-based analysis of modified tRNA bases. *Angew. Chem. Int. Ed. Engl.* **50**, 9739–9742 (2011).

40. Wetzel, C. & Limbach, P. A. Global identification of transfer RNAs by liquid chromatography–mass spectrometry (LC–MS). *J. Proteomics* **75**, 3450–3464 (2012).

41. Schiesser, S. et al. Deamination, oxidation, and C–C bond cleavage reactivity of 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxycytosine. *J. Am. Chem. Soc.* **135**, 14593–14599 (2013).

42. Schröder, A. S. et al. Synthesis of a DNA promoter segment containing all four epigenetic nucleosides: 5-methyl-, 5-hydroxymethyl-, 5-formyl-, and 5-carboxy-2′-deoxycytidine. *Angew. Chem. Int. Ed. Engl.* **53**, 315–318 (2014).

43. Thienpont, B. et al. Tumour hypoxia causes DNA hypermethylation by reducing TET activity. *Nature* **537**, 63–68 (2016).

44. Rahimoff, R. et al. 5-Formyl- and 5-carboxydeoxycytidines do not cause accumulation of harmful repair intermediates in stem cells. *J. Am. Chem. Soc.* **139**, 10359–10364 (2017).

45. Turowski, M. et al. Deuterium isotope effects on hydrophobic interactions: the importance of dispersion interactions in the hydrophobic phase. *J. Am. Chem. Soc.* **125**, 13836–13849 (2003).

46. Kanao, E. et al. Isotope effects on hydrogen bonding and CH/CD−π interaction. *J. Phys. Chem.* **122**, 15026–15032 (2018).

47. Liu, S. & Wang, Y. Mass spectrometry for the assessment of the occurrence and biological consequences of DNA adducts. *Chem. Soc. Rev.* **44**, 7829–7854 (2015).

48. Lentini, A. et al. A reassessment of DNA-immunoprecipitation-based genomic profiling. *Nat. Methods* **15**, 499–504 (2018).

49. Shrivastava, A. & Gupta, V. B. HPLC: isocratic or gradient elution and assessment of linearity in analytical methods. *J. Adv. Sci. Res.* **3**, 12–20 (2012).

50. Schellinger, A. P. & Carr, P. W. Isocratic and gradient elution chromatography: a comparison in terms of speed, retention reproducibility and quantitation. *J. Chromatogr. A* **1109**, 253–266 (2006).

51. Zhang, J. J. et al. Analysis of global DNA methylation by hydrophilic interaction ultra high-pressure liquid chromatography tandem mass spectrometry. *Anal. Biochem.* **413**, 164–170 (2011).

52. Contreras-Sanz, A. et al. Simultaneous quantification of 12 different nucleotides and nucleosides released from renal epithelium and in human urine samples using ion-pair reversed-phase HPLC. *Purinergic Signal.* **8**, 741–751 (2012).

53. Yonekura, S., Iwasaki, M., Kai, M. & Ohkura, Y. High-performance liquid chromatographic determination of guanine and its nucleosides and nucleotides in biospecimens with precolumn fluorescence derivatization using phenylglyoxal. *Anal. Sci.* **10**, 247–251 (1994).

54. Giel-Pietraszuk, M. et al. Quantification of 5-methyl-2′-deoxycytidine in the DNA. *Acta Biochim. Pol.* **62**, 281–286 (2015).

55. Pitt, J. J. Principles and applications of liquid chromatography-mass spectrometry in clinical biochemistry. *Clin. Biochem. Rev.* **30**, 19–34 (2009).

56. Niessen, W. M. A. State-of-the-art in liquid chromatography–mass spectrometry. *J. Chromatogr. A* **856**, 179–197 (1999).

57. Williamson, L. N. & Bartlett, M. G. Quantitative liquid chromatography/time-of-flight mass spectrometry. *Biomed. Chromatogr.* **21**, 567–576 (2007).

58. Payne, A. H. & Glish, G. L. Tandem mass spectrometry in quadrupole ion trap and ion cyclotron resonance mass spectrometers. *Methods Enzymol.* **402**, 109–148 (2005).

59. Tozuka, Z. et al. Strategy for structural elucidation of drugs and drug metabolites using (MS)n fragmentation in an electrospray ion trap. *J. Mass Spectrom.* **38**, 793–808 (2003).

60. Ens, W. & Standing, K. G. Hybrid quadrupole/time-of-flight mass spectrometers for analysis of biomolecules. *Methods Enzymol.* **402**, 49–78 (2005).

61. Commission Decision 2002/657/EC. *Implementation of Council Directive 96/23/EC Concerning the Performance of Analytical Methods and the Interpretation of Results* (Brussels, Official Journal of the European Communities, 2002).

62. Zhao, Z. & Zhang, F. Sequence context analysis in the mouse genome: single nucleotide polymorphisms and CpG island sequences. *Genomics* **87**, 68–74 (2006).

63. Kellner, S. et al. Oxidation of phosphorothioate DNA modifications leads to lethal genomic instability. *Nat. Chem. Biol.* **13**, 888–894 (2017).

64. Kellner, S. et al. Absolute and relative quantification of RNA modifications via biosynthetic isotopomers. *Nucleic Acids Res.* **42**, e142 (2014).

65. Amouroux, R. et al. De novo DNA methylation drives 5hmC accumulation in mouse zygotes. *Nat. Cell Biol.* **18**, 225–233 (2016).

66. Busskamp, V. et al. Rapid neurogenesis through transcriptional activation in human stem cells. *Mol. Syst. Biol.* **10**, 760 (2014).

67. Li, E., Bestor, T. H. & Jaenisch, R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915–926 (1992).

68. Collins, A. R., Cadet, J., Mőller, L., Poulsen, H. E. & Viña, J. Are we sure we know how to measure 8-oxo-7,8-dihydroguanine in DNA from human cells? *Arch. Biochem. Biophys.* **423**, 57–65 (2004).

69. ESCODD. Comparative analysis of baseline 8-oxo-7,8-dihydroguanine in mammalian cell DNA, by different methods in different laboratories: an approach to consensus. *Carcinogenesis* **23**, 2129–2133 (2002).

70. Ravanat, J.-L. et al. Cellular background level of 8-oxo-7,8-dihydro-2′-deoxyguanosine: an isotope based method to evaluate artefactual oxidation of DNA during its extraction and subsequent work-up. *Carcinogenesis* **23**, 1911–1918 (2002).

Acknowledgements

Author contributions

F.R.T., S.S and K.I. designed and developed the protocol. S.K. provided expertise on mass spectrometry. F.S. developed the DNA isolation protocol. M.M. designed graphics. T.C. supervised and designed the studies. All authors participated in writing the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41596-018-0094-6.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to T.C.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Published online: 17 December 2018

Related links
Key references using this protocol
Pfaffeneder, T. et al. *Nat. Chem. Biol.* **10**, 574–581 (2014): https://www.nature.com/articles/nchembio.1532
Iwan, K. et al. *Nat. Chem. Biol.* **14**, 72–78 (2018): https://www.nature.com/articles/nchembio.2531
Schiffers, S. et al. *Angew. Chem. Int. Ed. Engl.* **56**, 11268-11271 (2017): https://onlinelibrary.wiley.com/doi/abs/10.1002/anie.201700424
Rahimoff, R. et al. *J. Am. Chem. Soc.* **139**, 10359–10364 (2017): https://pubs.acs.org/doi/10.1021/jacs.7b04131

3.1.3   Chromatin-dependent allosteric regulation of DNMT3A activity by MeCP2

Arumugam Rajavelu, Cristiana Lungu, Max Emperle, Michael Dukatz, Alexander Bröhm, Julian Broche, Ines Hanelt, Edris Parsa, Sarah Schiffers, Rahul Karnik, Alexander Meissner, Thomas Carell, Philipp Rathert, Renata Z. Jurkowska and Albert Jeltsch

Prologue

DNA methylating enzymes are essential for normal embryonic development. Nevertheless, their function is not yet fully evaluated. This project used a systematic domain mapping approach allowing to report a strong and direct interaction between MeCP2 and DNA methyltransferases (DNMT) at pericentromeric heterochromatin. The enzyme binds the ADD domain of DNMT3a with its transcriptional repression domain and thereby inhibits its activity. Underlying mechanism of the inhibition is an allosteric stabilization of the closed, autoinhibitory conformation of DNMT3a. Lysine 4 unmodified histone H3 can replace MeCP2 on the ADD domain. Therefore, depending on the local chromatin context, MeCP2 can control localization and activity of DNMT3a.

Author contribution

Genomic digest, mass spectrometric measurements and data analysis were performed by myself. Further details on the author contribution are described in the manuscript.

License

A. Rajavelu, C. Lungu, M. Emperle, M. Dukatz, A. Bröhm, J. Broche, I. Hanelt, E. Parsa, S. Schiffers, R. Karnik, A. Meissner, T. Carell, P. Rathert, R. Z. Jurkowska and A. Jeltsch, Chromatin-dependent allosteric regulation of DNMT3A activity by MeCP2, *Nucleic Acids Research, 2018, Vol. 46, No. 17,* **9044-9056**, by permission of *Oxford University Press*.

Chromatin-dependent allosteric regulation of DNMT3A activity by MeCP2

Arumugam Rajavelu[1,†], Cristiana Lungu[1,†], Max Emperle[1], Michael Dukatz[1], Alexander Bröhm[1], Julian Broche[1], Ines Hanelt[1], Edris Parsa[2], Sarah Schiffers[2], Rahul Karnik[3,4], Alexander Meissner[3,4], Thomas Carell[2], Philipp Rathert[1], Renata Z. Jurkowska[1] and Albert Jeltsch[1,*]

[1]Department of Biochemistry, Institute of Biochemistry and Technical Biochemistry, Faculty of Chemistry, University Stuttgart, Allmandring 31, 70569 Stuttgart, Germany, [2]Center for Integrated Protein Science (CiPSM) at the Department of Chemistry, Ludwig-Maximilians-University, Butenandtstr. 5–13, 81377 Munich, Germany, [3]Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA 02138, USA and [4]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA

ABSTRACT

Despite their central importance in mammalian development, the mechanisms that regulate the DNA methylation machinery and thereby the generation of genomic methylation patterns are still poorly understood. Here, we identify the 5mC-binding protein MeCP2 as a direct and strong interactor of DNA methyltransferase 3 (DNMT3) proteins. We mapped the interaction interface to the transcriptional repression domain of MeCP2 and the ADD domain of DNMT3A and find that binding of MeCP2 strongly inhibits the activity of DNMT3A *in vitro*. This effect was reinforced by cellular studies where a global reduction of DNA methylation levels was observed after overexpression of MeCP2 in human cells. By engineering conformationally locked DNMT3A variants as novel tools to study the allosteric regulation of this enzyme, we show that MeCP2 stabilizes the closed, autoinhibitory conformation of DNMT3A. Interestingly, the interaction with MeCP2 and its resulting inhibition were relieved by the binding of K4 unmodified histone H3 N-terminal tail to the DNMT3A–ADD domain. Taken together, our data indicate that the localization and activity of DNMT3A are under the combined control of MeCP2 and H3 tail modifications where, depending on the modification status of the H3 tail at the binding sites, MeCP2 can act as either a repressor or activator of DNA methylation.

INTRODUCTION

The correct establishment and maintenance of DNA methylation patterns that are set during mammalian embryogenesis by the *de novo* DNA methyltransferases, DNMT3A and DNMT3B, depend on the accurate targeting and regulation of DNA methyltransferases (DNMTs) (1–3). Mammalian DNMTs comprise two parts, a large multi-domain N-terminal part and a C-terminal catalytic domain (CD) (4,5). The N-terminal parts of DNMTs are important for guiding the nuclear and sub-nuclear localization of the enzymes. They function as an interaction platform with other proteins, DNA and chromatin, thereby regulating the catalytic activity (1,6). Two defined sub-domains are present in the N-terminal part of DNMT3A and DNMT3B (Figure 1A): a PWWP domain, which recognizes H3K36me3-modified H3 tails (7–9), and an ADD domain, which binds the N-terminus of histone H3 if K4 is unmodified (10–12). Structural and biochemical work has shown that the ADD domain of DNMT3A forms contacts with the CD at two distinct interfaces (Supplementary Figure S1) (13), an autoinhibitory site, where ADD binding hinders DNA-binding and thereby inhibits the activity of the CD, and an allosteric site, where binding does not lead to inhibition. Different ADD residues are contacting the CD, depending on the conformation adopted by the enzyme. For instance, Y526 contacts the CD in the allosteric conformation, while D531 contacts the CD in the autoinhibitory conformation. The conversion of DNMT3A from the autoinhibitory to the open conformation was shown to be stimulated by binding of the unmodified H3 peptide to the ADD domain.

---

*To whom correspondence should be addressed. Tel: +49 711 685 64390; Fax: +49 711 685 64392; Email: albert.jeltsch@ibtb.uni-stuttgart.de
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.
Present addresses:
Arumugam Rajavelu, Rajiv Gandhi Center for Biotechnology (RGCB), Trivandrum 695014, Kerala, India.
Renata Z. Jurkowska, BioMed X Innovation Center, Im Neuenheimer Feld 583, D-69120 Heidelberg, Germany.
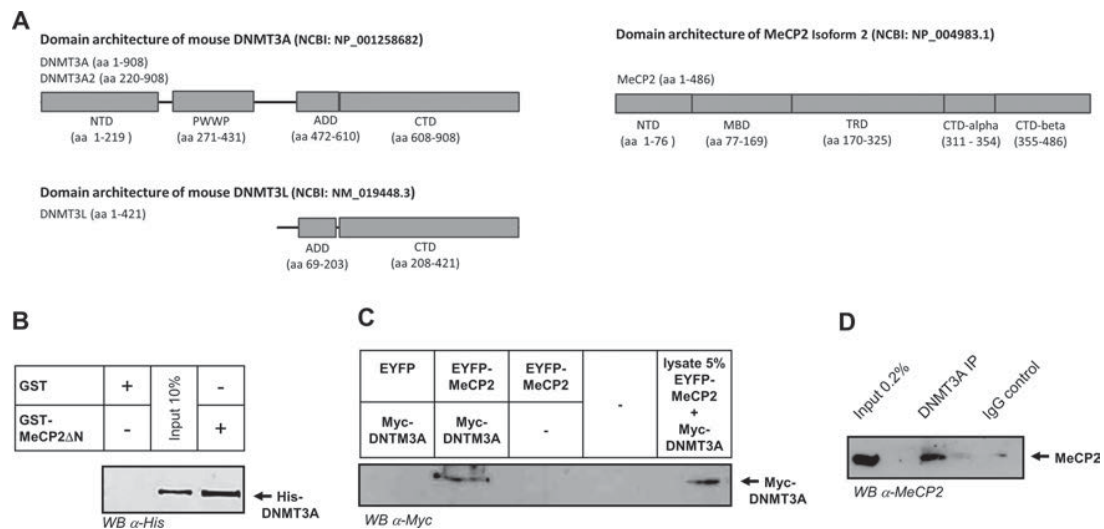
Figure 1. Structures of DNMT3 proteins and MeCP2 and interaction of MeCP2 with DNMT3A. (**A**) Domain architecture of mouse DNMT3A, DNMT3L and MeCP2 annotated with the domain boundaries used in this work. (**B**) Western blot detection of His-DNMT3A after its pull-down with the GST-tagged N-terminal truncated MeCP2 (MeCP2ΔN, residues 104–486). (**C**) Western blot detection of Myc-tagged DNMT3A after its pull-down by EYFP-tagged MeCP2 following transient co-expression of both proteins in human HEK293 cells. As a negative control, the pull-down was performed after co-expression of DNMT3A with EYFP. See also Supplementary Figure S10. (**D**) Western blot detection of endogenous MeCP2 after its pull-down by DNMT3A using mouse brain protein extracts. An immunoprecipitation using an IgG isotype control was included as negative control. NTD, N-terminal domain.

Through its binding, H3 induces a large movement of the ADD domain, which leads to the activation of DNMT3A (12–14). Several studies have recently shown that the activity of DNMT3 enzymes is regulated by the interaction with the H3 tail not only *in vitro,* but also in cells (8,15–18).

The biological role of DNA methylation is mediated by proteins, which specifically bind to DNA carrying methylated cytosine (19). One important reader of DNA methylation is the MeCP2 protein (Figure 1A) (20–23). It is the founding member of a group of proteins containing so called methyl-binding domains (MBDs), which bind to DNA in a methylation specific manner (24). MeCP2 recognizes methylated DNA with a preference for CpG dense islands (22,25), and its binding strength is strongly influenced by the flanking sequence of the methylated CpGs (26,27). Recently, binding of methylcytosine in non-CpG contexts was also reported (28,29). In addition to the MBD, MeCP2 contains a transcriptional repression domain (TRD), which serves as a protein recruitment platform and shows a weak methylation-independent DNA binding (23,30). MeCP2 is known to interact with various transcriptional repressors and co-repressors, including Histone Deacetylases, DNMT1, and the ATRX and Sin3A proteins (31–34). Functionally, MeCP2 is involved in numerous cellular processes, like methylation-induced gene repression (in particular of long genes), control of repetitive elements, chromatin compaction and looping and splice site regulation (23,29,30,35–38). In addition to its role in gene silencing, gene expression studies performed in specific brain subregions found altered expression levels of hundreds of genes after loss of MeCP2, most of which were upregulated by MeCP2 (39–41). These findings indicate that MeCP2 can function as a gene activator or repressor depending on the genomic context. Nevertheless, in spite of its high abundance in adult neurons, important role in chromatin organization and strong clinical relevance, the mechanistic details behind these opposing roles of MeCP2 are not fully understood (42).

Both DNMT3A and MeCP2 are highly expressed in neurons (35,43) (EBI expression atlas http://www.ebi.ac.uk/gxa/) and they have important functions in the brain. DNMT3A has been implicated in neuromuscular control, synaptic plasticity, learning and memory (44–46). MeCP2 functions as a structural protein and forms a specific type of chromatin, which is depleted of histone H1 (23). It plays an essential role in brain plasticity (35) and inactivating mutations of the X-linked *MECP2* gene were shown to cause Rett syndrome, a severe neurodevelopmental disease associated with developmental disorders and autism-like symptoms in females (30,34).

Since functional crosstalks between several readers and writers of epigenetic modifications have been previously reported (47) and DNMT3A and MeCP2 were both known to be targeted to pericentromeric heterochromatin (48), we investigated their potential interaction in this work. We observed a strong binding of MeCP2 to DNMT3A *in vitro,* in cells, and in the mouse brain and mapped the interaction interface to the TRD domain of MeCP2 and the ADD domain of DNMT3A. We found that binding of MeCP2 strongly inhibits the activity of DNMT3A by stabilizing its autoinhibitory conformation. The inhibition of DNMT3A was relieved by binding of an H3 tail peptide unmodified at K4 to the ADD domain suggesting that MeCP2 controls DNMT3A localization and activity depending on the local chromatin context.

MATERIALS AND METHODS

Generation of DNMT3A2 mutants and MeCP2 domain constructs

The murine DNMT3A and MeCP2 proteins and protein domains were prepared as indicated in Figure 1A. The MeCP2 domains and mutant proteins were cloned in pGEX-6P2 vector using BamHI and XhoI cloning sites. All constructs were verified by DNA sequencing. Site-directed mutagenesis were carried out by rolling circle polymerase chain reaction (PCR) using a primer carrying point mutation (49). The presence of the mutations was confirmed by restriction marker analysis and by DNA sequencing.

Expression and purification of MeCP2 and DNMT3A proteins

The MeCP2 domains and mutant proteins were expressed in *Escherichia coli* BL-21 cells. Cells were cultivated in Lysogeny broth (LB) medium at 37°C while shaking until an OD (600 nm) of 0.6–0.7 was reached. Then, protein expression was induced by addition of 1 mM of isopropyl-β-D-thiogalactoside and the culture was incubated at 18°C shaking at 200 rpm overnight. The cells were harvested by centrifugation (15 min at 4600 rpm) and the pellet resuspended in sonication buffer (20 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) (pH 7.5), 500 mM KCl, 1 mM ethylenediaminetetraacetic acid (EDTA), 1 mM dithiothreitol (DTT), 10% glycerol) including protease inhibitor (Sigma). The cells were lysed by sonication and centrifuged at 18 000 rpm for 1 h to prepare a clear lysate, which was applied on a GST-sepharose column (GE Healthcare). After washing with sonication buffer, the protein was eluted with sonication buffer containing 50 mM reduced glutathione and dialyzed first against dialysis buffer I (20 mM HEPES (pH 7.5), 200 mM KCl, 1 mM EDTA, 1 mM DTT, 10% glycerol) for 3 h, then against dialysis buffer II containing 60% glycerol overnight. The purified proteins were analyzed on 12% sodium dodecylsulphate-polyacrylamide gel electrophoresis (SDS-PAGE) gel stained with colloidal Coomassie BB. The murine DNMT3A2 and DNMT3A–C proteins were expressed and purified as described previously (50–52). Since all DNMT3A structures were annotated with numbers for the human proteins, we use human numbering here. The residue numbers corresponding to human Q527, D528 and D531 in murine DNMT3A are Q523, D524 and D527, respectively. Examples of images of the purified proteins used in this work are shown in Supplementary Figures S2–S7.

GST pull-down experiments

For GST pull-down experiments, 20 μl of Glutathione-Sepharose 4B beads were washed with 200 μl of interaction buffer (25 mM Tris (pH 8.0), 100 mM KCl, 5 mM MgCl₂, 10% glycerol, 0.1% Nonident P-40 (NP-40), 200 μM Phenylmethanesulfonyl fluoride (PMSF)). The beads were incubated for 1 h at 4°C with 10–15 μg of the different GST-tagged proteins, washed three times with interaction buffer and incubated with His- or MBD-tagged proteins (15 μg) for 1 h at 4°C with shaking. Then, the beads

were washed three times with wash buffer containing high salt (25 mM Tris (pH 8.0), 5 mM MgCl₂, 300 mM KCl, 10% glycerol, 0.1% NP40, 200 μM PMSF). The interaction of DNMT3A–ADD and MeCP2 TRDs was also investigated using buffer containing up to 600 mM KCl. At last, the beads were resuspended in sodium dodecyl sulphate (SDS) gel loading buffer and incubated for 10 min at 95°C. After centrifugation of the beads at 14 000 rpm for 10 min, the supernatant was loaded on a 12% SDS-PAGE gel. Proteins were detected by western blotting or Coomassie BB staining as indicated. Some experiments were conducted in the presence of recombinant histone H3.1 (Cat. No. M2503S, New England Biolabs), as detailed in the 'Results' section.

Co-immunoprecipitation assay

For co-immunoprecipitation of DNMT3A and MeCP2, pcDNA-DNMT3A (expressing myc tagged DNMT3A) and pEYFP-MeCP2 plasmids were co-transfected in HEK293 cells. The pEYFP plasmid was used as control. After 48 h, the cells were collected and the cell pellets stored at −80°C. The cells were lysed as recommended by the GFP trap protocol (ChromTek). Using GFP trap, YFP-tagged MeCP2 was pulled-down and the complex washed with buffer (10 mM Tris/Cl (pH 7.5), 0.5 mM EDTA, 0.5% NP-40, 200 mM NaCl). The MeCP2 and DNMT3A proteins were separated on a 12% SDS-PAGE and transferred to nitrocellulose membrane. To detect DNMT3A, the blot was probed with anti-myc antibody (Santa Cruz, 1:1000 dilution) for 1 h at room temperature.

For immunoprecipitation of endogenous DNMT3A and MeCP2, whole brains from 16-week-old C57Bl/N female mice were used. Following mechanical disruption, the tissue was lysed following a published protocol (53) with some modifications. Three brains were homogenized in NP-40 lysis buffer (10 mM HEPES (pH 7.9), 3 mM MgCl₂, 10 mM KCl, 10 mM NaF, 1 mM Na₃VO₄, 0.5 mM DTT, 0.5% NP-40, 1× complete EDTA-free protease inhibitor cocktail (Roche)), by douncing 30× with a tight pestle, and pelleted at 1000 g. Lysates were next diluted 1:1 with Benzonase buffer (10 mM HEPES (pH 7.9), 3 mM MgCl₂, 280 mM NaCl, 0.2 mM EDTA, 10 mM NaF, 1 mM Na₃VO₄, 0.5 mM DTT, 0.5% NP-40) supplemented with 1× complete EDTA-free protease inhibitor cocktail (Roche) and sonicated with EpiShear (Active Motif) for 2 min 30 s (15 s ON, 30 s OFF cycles, 20% power, 3.2 mm microtip). The homogenate was then digested with 500 units of Benzonase (Novagen) for 2 h rotating at 4°C. Chromatin proteins were separated by centrifugation at 17 000 g for 20 min at 4°C. For each pull-down, 2.5 mg lysate were pre-cleared for 1 h at 4°C with 20 mg Protein A Sepharose CL-4B (GE Healthcare), followed by overnight incubation with 15 μg anti-DNMT3A antibody (sc-2070, Santa Cruz). For negative control, an equivalent amount of non-related rabbit IgG anti-myc (ab9106, Abcam) antibody was used. The antibody-bound proteins were immobilized to 100 mg Protein A Sepharose CL-4B, blocked in 10% bovine serum albumin (Roth) for 6 h rotating at 4°C. After five washes with immunoprecipitation buffer, the immune complexes were eluted from the beads by boiling in 100 μl Laemmli sample buffer. The samples were next analyzed by western blot

as described above. For detection, anti-MeCP2 monoclonal primary antibody (#3456, Cell Signaling) was used, followed by incubation with HRP-linked anti-rabbit IgG light chain specific secondary antibody (211-032-171, Jackson ImmunoResearch). Western lighting *Ultra* (Perkin Elmer) was used as ECL HRP substrate.

Fluorescence microscopy

For sub-nuclear localization studies, NIH3T3 cells were seeded on glass slides and transfected with plasmids expressing CFP- and YFP-tagged DNMT3L, DNMT3A–ADD and MeCP2 using Fugene HD (Promega) according to the manufacturer's instructions. After 48 h, the cells were fixed with 4% formaldehyde, mounted in Mowiol and Z stacks images were collected using a Zeiss LSM 710 confocal microscope. Fluorescence signals were collected in the YFP and CFP channels after confirming absence of crosstalk (Supplementary Figure S8).

Substrates used for DNA methylation

The following oligonucleotide substrates were used for DNA methylation assays: a biotinylated unmethylated 30 mer containing one CpG site (um30mer: TTG CAC TCT CCT CCC GGA AGT CCC AGC TTC / Bt-GAA GCT GGG ACT TCC GGG AGG AGA GTG CAA), the same sequence hemimethylated at the CpG site with the methylation in the lower DNA strand (hm30mer: TTG CAC TCT CCT CCC GGA AGT CCC AGC TTC / Bt-GAA GCT GGG ACT TCmC GGG AGG AGA GTG CAA), a biotinylated hemimethylated 30 mer with optimized flanks for methylation with DNMT3A (54) (hmF30mer: GAA GCT GGA CAG TAC GTC AAG AGA GTG CAA / Bt-TTG CAC TCT CTT GAmC GTA CTG TCC AGC TTC) and a non-CpG substrate (nonCpG: GAA GCT GGT CCA TTmC GAT GAT GGA GTG CAA / Bt-TTG CAC TCC ATC ATmC GAA TGG ACC AGC TTC). The oligonucleotides were annealed by heating to 86°C for 5 min and slowly cooling down to ambient temperature. In addition, a biotinylated 585-mer DNA substrate obtained by PCR was used that contains eight HpaII sites (CCGG) and 45 CpG sites (um585mer) (Supplementary Figure S9). The 585 mer was amplified from Lambda-phage DNA by PCR using the following primers: Bt-GAA GGA CAA CCT GAA GTC CAG GTTG and GTG TAT GAC CAC CAG AGC CTT TTGC and purified by PCR purification kits (Qiagen). To prepare partially methylated 585 mer (pm585mer), the DNA was methylated with M.HpaII (NEB) following the protocol of the provider and afterward purified by PCR purification kits. Successful pre-methylation at HpaII sites was confirmed by HpaII (NEB) restriction digestion (Supplementary Figure S9).

DNA methylation activity assay

The avidin–biotin microplate DNMT activity assay was used to monitor the activity of different DNMT3A variants in the methylation of biotinylated DNA substrates, basically as described (55,56). Each well of the microplate was coated with 1 μg of avidin dissolved in 100 μl of 100 mM NaHCO$_3$

(pH 9.6) and incubated overnight at 4°C. Before starting the assay, the wells were washed five times with 200 μl of 1× PBST (140 mM NaCl, 2.7 mM KCl, 4.3 mM Na$_2$HPO$_4$, 1.4 mM K$_2$HPO$_4$, 0.05% v/v Tween 50, pH 7.2) containing 0.5 M NaCl. The reaction mixtures were prepared containing 2.5 μM DNMT3A2 or DNMT3A–C and 3 μM MeCP2 (or any of its domains) in methylation buffer (20 mM HEPES (pH 7.2), 1 mM EDTA, 50 mM KCl, 1.25 mg/ml bovine serum albumin). For the control reactions without MeCP2, the same volume of dialysis buffer II was added instead of the MeCP2. The reaction mixtures were incubated on ice for 20 min and the wells of the plate were filled with 5 μl of 0.5 M unlabeled AdoMet (Sigma) dispensed in 35 μl 1× PBST/0.5 M NaCl. Then, 1 μM 30-mer oligonucleotide DNA or 100 nM of 585-mer DNA and 0.76 μM [methyl-3H]-AdoMet (PerkinElmer Life Sciences) were added to the reaction mixture and the samples were incubated at 37°C. In order to follow the time course of the reaction, aliquots of 2 μl were removed from the reaction mixtures in duplicates at time points between 2 and 30 min and applied to one well of the microplate where the incorporation of labeled AdoMet was quenched by an excess of unlabeled AdoMet. This mixture was incubated while slightly shaking for 30 min to 1 h. The wells were washed five times with 200 μl of 1× PBST and 0.5 M NaCl. A buffer (100 μl; 50 mM Tris–HCl (pH 8.0), 5 mM MgCl$_2$) containing 0.7 μg unspecific nuclease from *Serratia marcescens* was added per well and the mixture was incubated for 30–60 min with slight shaking. At last, the released radioactivity was measured using liquid scintillation counting and the average count per minute of the duplicates was plotted against time. Linear regression was used to obtain the slopes of the initial linear parts of the time courses. The data are reported as averages and standard error of the mean (SEM) of at least two independent experiments.

DNA methylation analysis in HCT116 DNMT1 hyphomorph cells

To study the effect of MeCP2 on DNMT3A-mediated methylation in cells, we used HCT116 DNMT1 hyphomorphic cells (HCT116^{D1hypo}, kindly provided by Prof. Bert Vogelstein, HHMI, USA), which have a reduced level of global DNA methylation (57,58). HCT116^{D1hypo} cells were cultivated in McCoy's 5A medium (Gibco catalog no.: 16 600) supplemented with 10% heat-inactivated calf serum, 2 mM L-glutamine (Sigma), 100 U/ml penicillin and 100 μg/ml streptomycin at 37°C in a saturated humidity atmosphere containing 5% CO$_2$. HCT116^{D1hypo} cells were modified to express the ecotropic receptor and rtTA3 using retroviral transduction of pWPXLd-RIEP (pWPXLd-rtTA3-IRES-EcoR-PGK-Puro) followed by drug selection (0.8 ug/ml puromycin for 1 week, respectively) similarly as described (59). The resulting cell line was subsequently transduced with ecotropically packaged retroviruses containing the *mecp2* gene fused to EYFP under control of a TRE3G promoter. Retroviral gene transfer was performed as previously described (60) using 10–20 μg plasmid DNA and 5 μg helper plasmid (pCMV-Gag-Pol, Cell Biolabs) for each calcium phosphate transfection. Retroviral packaging was performed using PlatiniumE cells (Cell

Biolabs). Transduction efficiencies of retroviral constructs (TRE3G-MeCP2-EYFP-PGK-NEO and TRE3G-EYFP-PGK-NEO) (61) were measured 48 h post induction with 1 μg/ml doxycycline by flow cytometry (MACSQuant® VYB, Miltenyi Biotec GmbH, Germany). Transduced cell populations were selected 5 days post infection using 500 μg/ml G418 (Gibco Life technologies). After 14 days of induction, ∼1 million MeCP2 and the EYFP control expressing cells (as judged by being EYFP+) were sorted for each replicate by flow cytometry (FACSAria III, BD, USA). Genomic DNA was isolated using the DNA mini Kit (Qiagen) and followed by LC-ESI-MS/MS analysis of DNA methylation as described (62,63).

Quantitative RT-PCR

Quantitative RT-PCR assays were performed on a CFX96 Connect Real-Time detection system (Bio-Rad, Hercules, CA, USA) using SsoFast EvaGreen supermix (Bio-Rad, Hercules, CA, USA). For gene expression analysis, total RNA from 10^6 cells was isolated for each sample using RNeasy Mini Kit (Qiagen, Limburg, The Netherlands). Complementary DNA (cDNA) was prepared using MultiScribe™ Reverse Transcriptase (Applied Biosystems, Thermo Fisher, USA) with oligo d(T)18 primers (New England Biolabs, Ipswich, MA, USA) using 500 ng of RNA. After this, quantitative PCR (qPCR) was carried out using specific primers-sets for DNMT3A (CGA TTT CTC GAG TCC AAC CCT G**,** ACC GGG AAG GTT ACC CCA), DNMT3B (CAG TGA CAC GGG GCT TGA ATA TG, CTT TGA GGA CTA GGT AGC CTG TCG CG) and DNMT1 (GAG ACA CGA TGT CCG ACC TG, CCA ATG CAC TCA TGT CCT TAC AG), normalized to the housekeeping gene SDHA (TGG GAA CAA GAG GGC ATC TG, CCA CCA CTG CAT CAA ATT CAT). The relative DNMT expression was calculated using ΔC_t (Expression $= 2^{-\Delta C_t}$). Non-RT controls that did not undergo the cDNA synthesis step were included in all experiments. Error analysis was based on independent cDNA preparations.

RESULTS

MeCP2 interacts with DNMT3A

To test whether DNMT3A and MeCP2 interact, we first performed GST-pull-down assays using recombinant murine full-length (FL) proteins. MeCP2 could be obtained with good purity after generating a version lacking the N-terminal unstructured domain (MeCP2ΔN) (Supplementary Figure S2). As shown in Figure 1B, a robust pull-down of DNMT3A by GST-MeCP2ΔN could be detected. Encouraged by the strong and direct interaction detected in the *in vitro* assay, we next tested whether DNMT3A and MeCP2 can also interact in cells. For this, HEK293 cells were transiently co-transfected with Myc-tagged DNMT3A and EYFP-tagged MeCP2 (Supplementary Figure S10). To isolate immunocomplexes, the EYFP-tagged MeCP2 was immunoprecipitated and the pull-down material was tested for the presence of Myc-DNMT3A. As shown in Figure 1C, co-purified Myc-DNMT3A was detected after co-expression with EYFP-MeCP2, but not with the EYFP control. To exclude potential artifacts related to protein

overexpression, we finally performed immunoprecipitation of endogenous DNMT3A from mouse brain, a tissue where both MeCP2 and DNMT3A are abundantly expressed. As shown in Figure 1D, MeCP2 could be specifically detected in the pulled-down material, but not in the IgG control. Together, these results demonstrate that in addition to their direct *in vitro* interaction DNMT3A and MeCP2 also associate after transient co-expression in mammalian cells as well as at endogenous levels in brain extracts.

MeCP2 interacts with the ADD of DNMT3 proteins

Having established that the interaction between DNMT3A and MeCP2 is not an artifact of the *in vitro* pull-down assay, but is also detected in biological relevant context, we were next interested to pinpoint the interaction interfaces between these two proteins. For this, we have undertaken a systematic domain mapping approach by performing *in vitro* pull-down experiments with serially truncated recombinant DNMT3A and MeCP2.

To define the interaction interface on DNMT3A, we first performed GST pull-downs using MeCP2ΔN-GST and DNMT3A2, a naturally occurring isoform of DNMT3A, which lacks 219 N-terminal amino acids reported to be involved in DNA binding (Figure 1A) (64,65). As documented in Figure 2A, a robust pull-down was detected with DNMT3A2 indicating that residues 1–219 of DNMT3A are not required for the interaction with MeCP2. To further define the interacting area, we next resorted to DNMT3L, an important regulatory member of the DNMT3 family, which shares the ADD domain with DNMT3A and DNMT3B, while containing a crippled CD and lacking the N-terminal part and the PWWP domains (Figure 1A). The direct interaction detected between DNMT3L and MeCP2ΔN (Figure 2B) suggested that the association with MeCP2 occurs through an interface that is shared between DNMT3L and DNMT3A. To test this hypothesis, we next performed pull-downs with the CDs of DNMT3A (Figure 2C) and DNMT3L (Figure 2D), respectively. Both of these did not show an interaction with MeCP2ΔN. By contrast, a robust interaction could be detected between the GST-tagged ADD domain of DNMT3A as bait and a GST-cleaved MeCP2ΔN as prey (Figure 2E). No signal was detectable for the NTD and the PWWP domains of DNMT3A in line with the results obtained for DNMT3L and DNMT3A2. Taken together, these results show that MeCP2 interacts with DNMT3 proteins via their ADD, a domain that has been already reported to serve as a protein–protein interaction platform and be essential for the regulation of DNMT3 proteins.

DNMT3A interacts with the TRD of MeCP2

Having successfully mapped the interaction interface on the DNMT3A side, we were next interested to dissect which part of MeCP2 is responsible for the binding to DNMT3A. To this end, we performed pull-downs with GST-tagged MeCP2 domains. As shown in Figure 2F, out of the five tested domains an interaction was detectable only for the TRD. Pull-downs using the isolated MeCP2–TRD and DNMT3A–ADD (Figure 2G) confirmed the direct and
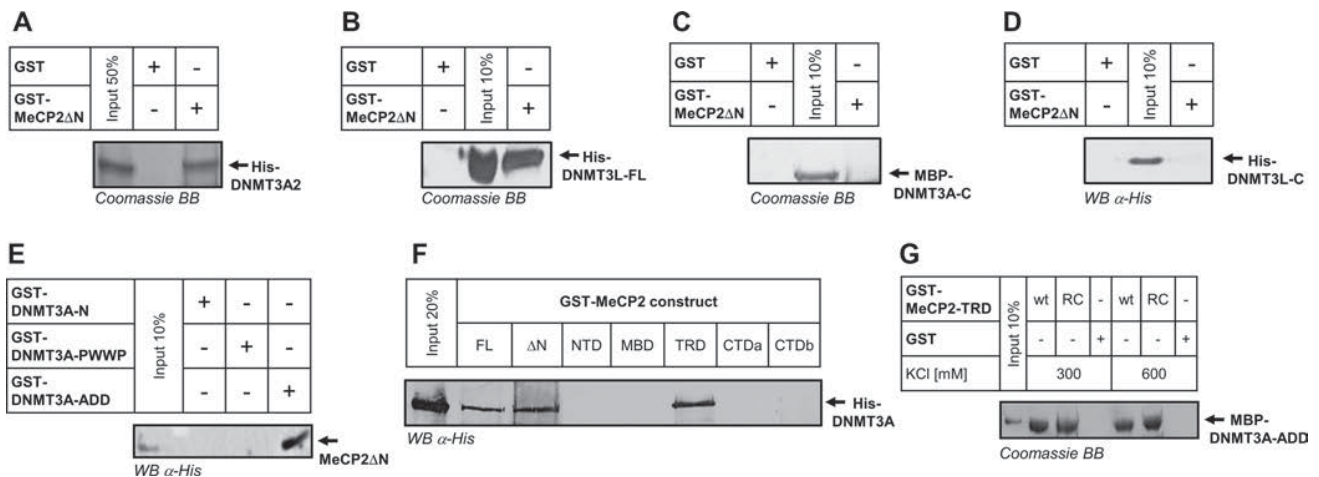
Figure 2. The ADD domain of DNMT3 proteins interacts with the TRD of MeCP2. (**A**) Coomassie staining of the pull-down of His-DNMT3A2 by GST-MeCP2ΔN. (**B**) Coomassie staining of the pull-down of His-DNMT3L by GST-MeCP2ΔN. (**C**) Coomassie staining of the pull-down of MBP-DNMT3A–C by GST-MeCP2ΔN. A signal can be only detected in the input lane. See also Supplementary Figure S15. (**D**) Western blot detection of the pull-down of His-DNMT3L-C by GST-MeCP2ΔN. A signal is visible only in the input lane. (**E**) Western blot detection of the pull-down of MeCP2ΔN with different GST-tagged DNMT3A domains. For this experiment, GST-cleaved MeCP2ΔN was used. The endogenous His-tag in MeCP2-CTDb was used for detection. Out of the three tested DNMT3A domains, only the ADD domain displayed an interaction with MeCP2. (**F**) Western blot detection of the pull-down of DNMT3A by different GST-tagged MeCP2 domains documenting an interaction with the MeCP2 FL, MeCP2ΔN and MeCP2–TRD domain. (**G**) Coomassie detection of the pull-down of MBP–DNMT3A–ADD by wild-type GST-TRD domain (wt) or TRD containing the R306C Rett mutation (RC), under two different salt concentrations. See also Supplementary Figures S11–S14.

self-sufficient interaction between these domains. The interaction of MeCP2–TRD and DNMT3A–ADD was further confirmed by Alpha-assay (Supplementary Figure S11) and gel filtration (Supplementary Figure S12). Having identified the domains responsible for mediating the interaction between these two proteins, we next tested the stability and strength of the association between MeCP2–TRD and DNMT3A–ADD by performing pull-downs under increasingly high salt concentrations. Strikingly, we could retrieve comparable amounts of DNMT3A–ADD at both 300 and 600 mM KCl, indicating that its interaction with MeCP2 TRD is strong and not driven by electrostatic interactions (Figure 2G). A salt resistant and strong interaction of MeCP2–TRD was also observed with DNMT3A2 (Supplementary Figure S13). Pull-down experiments in the presence of the non-specific and highly active nuclease from *S. marcescens* demonstrated that the interaction of both proteins was not mediated by nucleic acids (Supplementary Figure S14). Since MeCP2 is mutated in the Rett syndrome, we also tested the effect of the R306C Rett mutation in the TRD domain of MeCP2 on the interaction with DNMT3A, but did not observe any change when compared with wild-type TRD (Figure 2G).

MeCP2 influences the sub-nuclear localization of DNMT3L and the DNMT3A–ADD domain

Having shown the direct interaction of MeCP2 with DNMT3 proteins, we next aimed to study the influence of this interaction on the cellular localization of these factors. MeCP2 is known to accumulate at pericentromeric heterochromatin (22,48,53), which clusters in characteristic DAPI-dense foci in mouse fibroblasts. Unlike MeCP2, the regulatory factor DNMT3L was shown to have an almost homogenous nuclear distribution (66).

Expressing the fluorophore-tagged DNMT3L and MeCP2 in mouse fibroblasts confirmed the published localization patterns of both proteins (Figure 3A and C). Notably, co-expressing DNMT3L with MeCP2 led to a clear re-targeting of DNMT3L toward chromocenters (Figure 3D and F). This finding confirms the intracellular interaction of DNMT3L and MeCP2 and indicates that the expression of MeCP2 causes a re-distribution of the regulatory factor DNMT3L toward heterochromatin. Moreover, it shows that the MeCP2 interaction influences chromatin targeting of DNMT3L. However, the physiological role of the interaction of DNMT3L and MeCP2 is unclear, because DNMT3L does not appear to play a major role in the brain.

Unlike DNMT3L, DNMT3A is strongly enriched at pericentromeric heterochromatin (52,66) (and references therein), where it co-localizes with MeCP2 (48). This natural co-localization precluded direct studies of the effect of MeCP2 on the localization of DNMT3A. To circumvent this caveat, we performed localization studies with the fluorescently tagged DNMT3A–ADD domain. This showed a diffuse nuclear localization when transfected alone in mouse fibroblasts (Figure 3B). Similar to DNMT3L, after its co-expression with CFP-tagged MeCP2, the ADD domain showed a preferential enrichment at heterochromatic foci, indicating a MeCP2-mediated targeting to heterochromatin (Figure 3E and F). These results confirm that the ADD domain interacts with MeCP2 in cells and demonstrate that through its binding MeCP2 recruits DNMT3L and DNMT3A–ADD to pericentromeric heterochromatin.

The interaction with MeCP2 inhibits the catalytic activity of DNMT3A

To elucidate the function of the interaction between MeCP2 and DNMT3A, we next measured the *in vitro* rates of
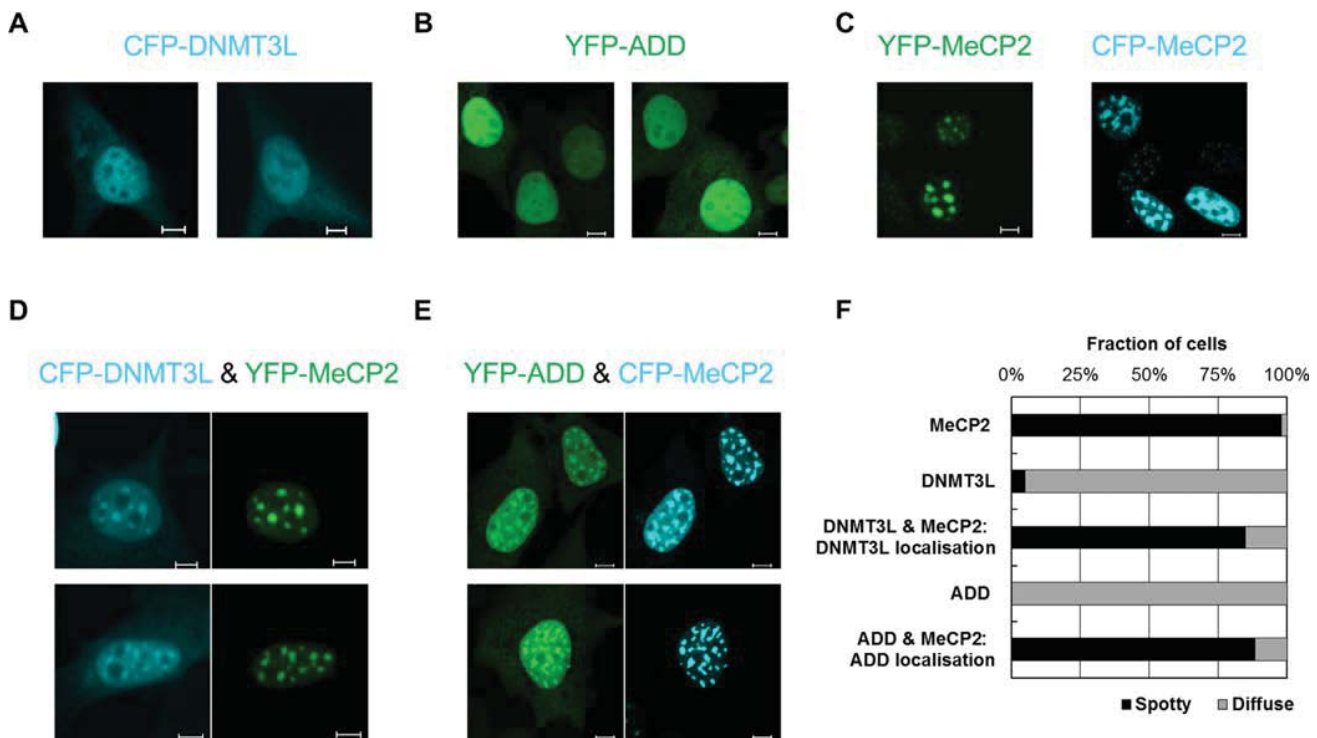
Figure 3. Transient expression of MeCP2 alters the cellular localization of DNMT3L and DNMT3A–ADD in NIH3T3 cells. (**A**) Representative fluorescence microscopy images documenting the localization of CFP-DNMT3L upon its overexpression in mammalian cells. A predominantly homogenous nuclear distribution is observed. (**B**) Representative fluorescence microscopy images documenting the localization of YPF-ADD upon its overexpression in mammalian cells. A predominantly homogenous nuclear distribution is observed. (**C**) Representative fluorescence microscopy images documenting the chromocenter-enriched localization of YFP-MeCP2 (left) and CFP-MeCP2 (right) upon their overexpression in mammalian cells. (**D**) Representative fluorescence microscopy images showing that upon their co-expression DNMT3L is recruited to MeCP2 clusters. (**E**) Representative fluorescence microscopy images showing that upon their co-expression DNMT3A–ADD is recruited to MeCP2 clusters. The scale bars correspond to 10 μm. (**F**) Quantification of the fraction of cells showing spotty and diffuse localization patterns in the experiments shown in panels (A–E) (based on analysis of >20 individual cells in each case). See also Supplementary Figure S8.

DNA methylation by DNMT3A2 in the presence of MeCP2 or its TRD domain. The activity of DNMT3A has been shown to be modulated by the target site, where CpG is preferred over CpA (67–69), the flanking sequence of CpG sites (70,71) and the length of the DNA substrates (72). To study all these different properties, six different DNA substrates were used for our DNA methylation experiments (Figure 4): (i) an unmethylated 30-mer oligonucleotide (um30mer), (ii) the same substrate in hemimethylated form (hm30mer), (iii) a hemimethylated 30 mer with an optimized flank for DNMT3A (hmF30mer) (54), (iv) a 585-mer PCR fragment (um585mer), (v) the 585-mer PCR fragment pre-methylated at HpaII sites (pm585mer) and (vi) a 30-mer oligonucleotide non-CpG substrate (non-CpG 30 mer).

Using 2.5 μM DNMT3A2 and 3 μM MeCP2, we consistently observed that the interaction of MeCP2 with DNMT3A2 resulted in ∼40–60% reduction in DNMT3A2 activity with the unmethylated substrates (Figure 4A and B). Similar results were obtained with a truncated form of DNMT3B corresponding to DNMT3A2 (Supplementary Figure S6). The activity of DNMT3A2 was further reduced by ∼80% with methylated substrates, which can be attributed to the better binding of MeCP2 to methylated DNA via its MBD. We speculated that binding of

MeCP2 to pre-methylated DNA might target DNMT3A2. To test this hypothesis, a partially methylated 585-mer DNA substrate was prepared by methylation with HpaII, an enzyme that exclusively methylates CG sites found within CCGG motifs (Supplementary Figure S9). However, even on this substrate, we observed inhibition of the activity of DNMT3A2 by MeCP2 (Figure 4B, pm585mer). As a control, we used DNMT3A–C, which lacks the ADD domain and does not interact with the MeCP2–TRD (Supplementary Figure S15), and observed that MeCP2 did not inhibit its activity (Figure 4C).

Having mapped the interaction interface to the TRD of MeCP2, we next tested whether this isolated domain can also inhibit the activity of DNMT3A2 and observed 40–70% inhibition with the different substrates (Figure 4D). Non-CpG methylation has recently been detected in considerable amounts in human ES cells and neurons, and it was connected to DNMT3A activity (28,73–75). We, therefore, also investigated the influence of MeCP2–TRD on the non-CpG methylation activity of DNMT3A2 using a 30-mer oligonucleotide substrate that contains one already methylated CpG site, such that additional methylation could only occur at non-CpG sites. As with the other DNA substrates, we observed a similar inhibition of DNMT3A2 activity by TRD (Figure 4D, non-CpG 30 mer), indicating that
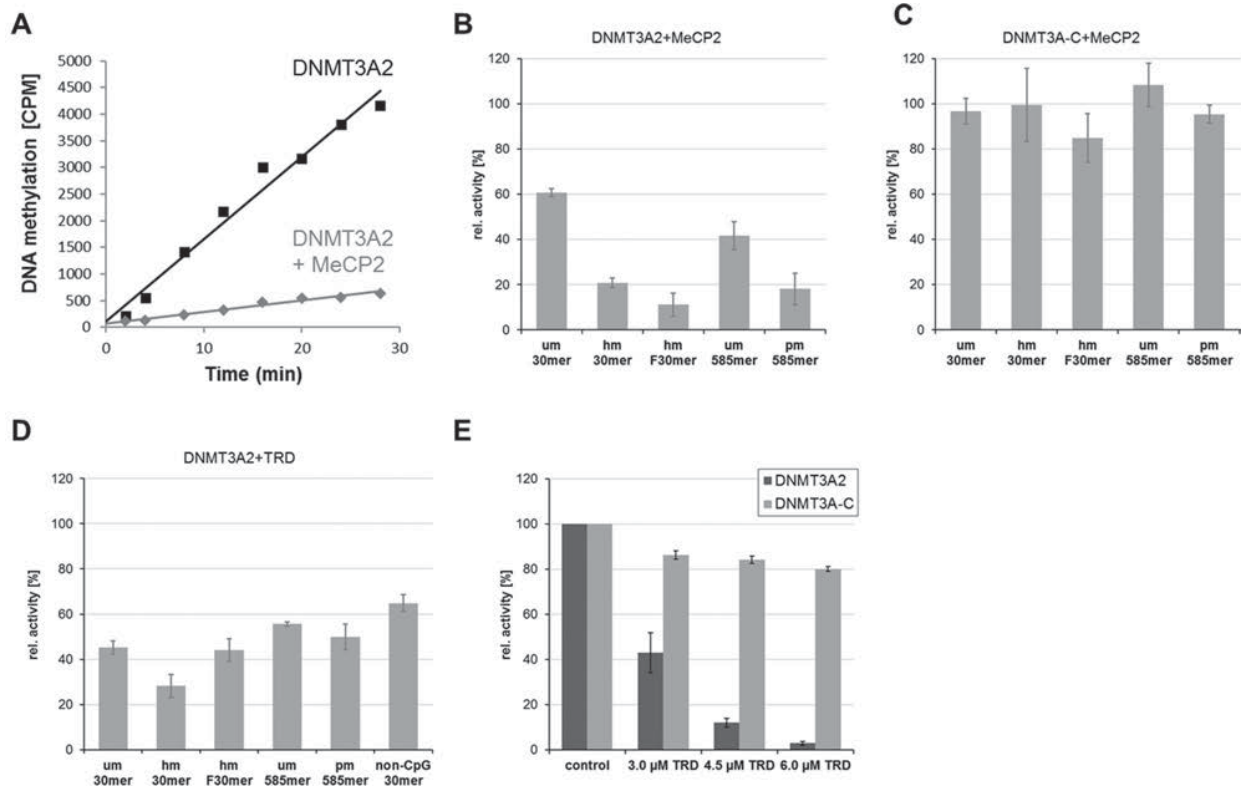
Figure 4. MeCP2 (**A–C**) and MeCP2–TRD (**D** and **E**) inhibit the activity of DNMT3A2 on a broad range of substrates. *In vitro* DNA methylation kinetics was conducted with 2.5 µM DNMT3A2 (A and B) or DNMT3A–C (C) in the presence of 3 µM MeCP2 (A–C) or MeCP2–TRD (D). In each panel, identical control reactions without addition of MeCP2 or MeCP2–TRD were used to calculate relative activities. Different DNA substrates were used as indicated. (E) Inhibition of DNMT3A2 at increasing concentrations of MeCP2–TRD in kinetics using um30mer as substrate. Reactions with DNMT3A–C were performed in parallel since this domain does not interact with MeCP2–TRD. Control refers to reactions without added TRD. Panel (A) shows methylation kinetics of hmF30mer as an exemplary primary data set. Bars show averages and SEM based on 2–3 independent experiments. See also Supplementary Figures S6, S13 and S14.

CpG and non-CpG methylation are equally inhibited by the TRD interaction.

At last, we tested the activity of DNMT3A2 in the presence of increasing amounts of TRD and observed that the methyltransferase activity was strongly inhibited (>95%) using a 2.4-fold excess of TRD (6 µM with 2.5 µM DNMT3A2) (Figure 4E). These results were fitted by a binding constant (K_D) of TRD to DNMT3A2 of 2.8 µM under catalytic conditions. As a control, the same experiments were conducted with DNMT3A–C, but only a very weak reduction of activity was observed (Figure 4E) indicating that the inhibition of DNMT3A2 by TRD is not caused by competition for the DNA substrate. This is an important control, since TRD was reported to weakly bind DNA (76,77). In summary, our results indicate that the interaction between the TRD domain of MeCP2 and the ADD domains of DNMT3A and DNMT3B results in a direct and very strong inhibition of the DNMT activity at both CpG and non-CpG sites.

MeCP2 overexpression reduces DNA methylation in HCT116 cells

Based on the strong influence of MeCP2 on the activity of DNMT3 proteins, we were next interested to see if

the inhibitory effects on DNMT3A activity observed in *in vitro* assays are also re-capitulated in a cellular context. For this we resorted to the HCT116 DNMT1 hypomorphic colon cancer cell line, which contains a truncated DNMT1 with reduced activity, but active copies of DNMT3A and DNMT3B (57,58). Because of the impaired maintenance DNA methylation activity, these cells have an ~20% reduced amount of DNA methylation, which is more dependent on the activity of DNMT3A and DNMT3B. This makes the HCT116^D1hypo cell line a suitable model system to study the effect of the inhibition of DNMT3A by MeCP2. We genomically integrated EYFP-fused MeCP2 or only the fluorophore as control by viral transduction and selected for stably expressing clones. Expression of MeCP2 or fluorophore control was induced for 14 days. Afterward, EYFP-MeCP2 or fluorophore expressing clones were enriched by fluorescence-activated cell sorting (Figure 5A). Genomic DNA was isolated and the global levels of 5-methylcytosine were quantified by liquid chromatography-mass spectrometry (LC-MS/MS). As shown in Figure 5B, we observed a 15% decrease in global DNA methylation. Expression levels of all DNMTs were determined by qPCR indicating a slight increase in DNMT3A expression, and no changes in DNMT1 and DNMT3B (Figure 5C), showing that the reduction in DNA methylation was not caused by reduced ex-
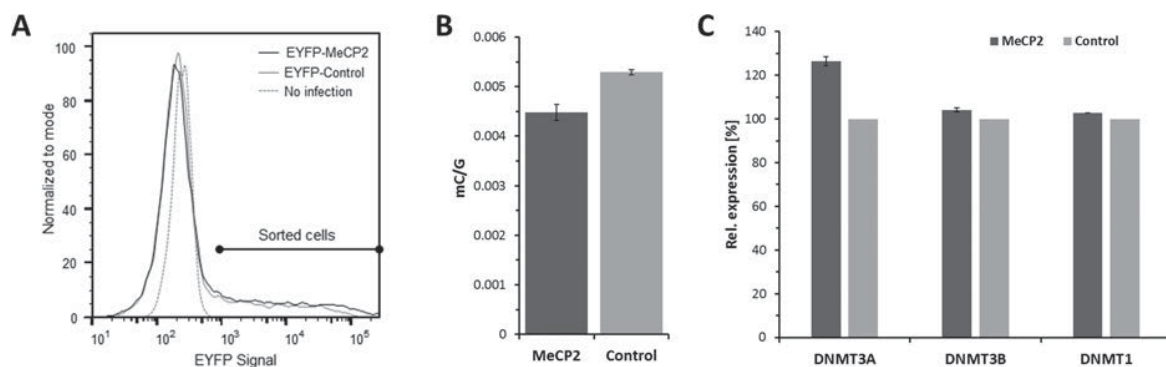
Figure 5. MeCP2 reduces global DNA methylation in HCT116 cells containing a DNMT1 hypomorphic allele. (**A**) Flow cytometry analysis of EYFP-MeCP2 and EYFP expression in HCT116^{D1hypo} cells. The signal range used for sorting is indicated. (**B**) Global DNA methylation levels after overexpression of EYFP-MeCP2 in human HCT116^{D1hypo} cells. EYFP-transfected cells were used as control. (**C**) Expression levels of all DNMTs in the HCT116^{D1hypo} cells expressing EYFP-MeCP2 or EYFP control. SDHA was used as reference gene. The error bars represent the SD based on two repeats.

pression of any of the DNMTs. Since DNMT1 interaction with MeCP2 was reported not to cause a reduction in catalytic activity (32), this result indicates that overexpression of MeCP2 reduces the activity of DNMT3 enzymes in cells.

The TRD domain inhibits DNMT3A2 activity by an allosteric mechanism

We next aimed to mechanistically analyze the striking inhibitory effect of MeCP2 on DNMT3A. As described above, structural studies showed that the ADD domain of DNMT3A can dock on the CD at two alternative sites: an allosteric and an autoinhibitory one (13). Binding of the H3 tail peptide to the ADD domain was shown to stabilize the allosteric conformation and thereby activate DNMT3A (13,14). To investigate the mechanism of the repression of DNMT3A by the TRD domain, we engineered DNMT3A2 variants containing mutations at Y526 or D531 in the ADD domain, two critical residues involved in the two binding sites at the CD (Supplementary Figure S1) in order to selectively disrupt or strongly destabilize one of the two DNMT3A conformations. Y526E was introduced to disrupt the allosteric and D531R to disrupt the autoinhibitory conformation. After confirming that both mutants still interact with MeCP2–TRD (Supplementary Figure S16), we investigated if these conformationally locked DNMT3A variants still respond to the presence of the TRD. As shown in Figure 6A, the inhibitory effect of MeCP2 was specifically lost in the D531R variant that can no longer adopt the autoinhibitory conformation. This finding suggests that MeCP2 reduces the activity of DNMT3A by an allosteric mechanism, in which TRD binding stabilizes the autoinhibitory conformation of the enzyme.

Since through its binding to the ADD domain, unmodified histone H3 was reported to allosterically activate DNMT3A, we next investigated if TRD and histone H3 binding to the ADD domain influence each other. For this, we conducted pull-down experiments using GST-TRD and DNMT3A2 in the presence of increasing concentrations of recombinant histone H3. As shown in Figure 6B, addition of histone H3 abolished the ADD–TRD interaction, suggesting that H3 and TRD binding to the ADD

domain is mutually exclusive. To investigate whether histone H3 can rescue the TRD-mediated inhibition, we next conducted DNA methylation experiments with DNMT3A2 and DNMT3A2 pre-incubated with the unmodified H3 (1–19) peptide in the absence and presence of GST-TRD (Figure 6C). In line with the previous experiments, we observed that the TRD-mediated inhibition is alleviated in the presence of H3 peptide (5 μM). Correspondingly, the activation of DNMT3A2–TRD complexes (which are predominantly in the autoinhibitory conformation) is stronger than the activation of free DNMT3A2 (which is in a mixed conformation state). Moreover, we observed that the inhibitory effect of the TRD domain on DNMT3A activity was completely lost at higher concentrations of the H3 peptide (25 μM, Supplementary Figure S17). This indicates that the binding of H3 to the ADD domain can disrupt the DNMT3A–TRD interaction and relieve the associated enzymatic inhibition.

DISCUSSION

During the past decade compelling experimental evidence has accumulated, indicating that DNA methylation patterns are highly dynamic and result from ongoing *de novo* methylation and demethylation events (3). This dynamic landscape plays particularly important roles in nondividing cells, such as terminally differentiated neurons (78–80). In the absence of cell division and DNA replication, the DNA methylation profiles in these cells can only be controlled through a tight regulation of the targeting and activity of DNA methylating and demethylating enzymes. However, despite their importance, the details of this regulatory network have remained mysterious so far. In this work, we took a closer look at the DNMT3 methyltransferases, factors that play essential roles in mammalian development and disease (1,2,81), their targeting and allosteric regulation (6). We find that recombinant DNMT3A and DNMT3L proteins directly and strongly interact with the chromatin regulator MeCP2 *in vitro*. We confirmed this interaction under overexpression conditions in mammalian cells, as well as at endogenous expression levels in mouse brain lysates. By performing systematic domain mapping, we find that the interaction of MeCP2 and DNMT3 proteins is mediated by their TRD and the ADD domains, respectively. Based on
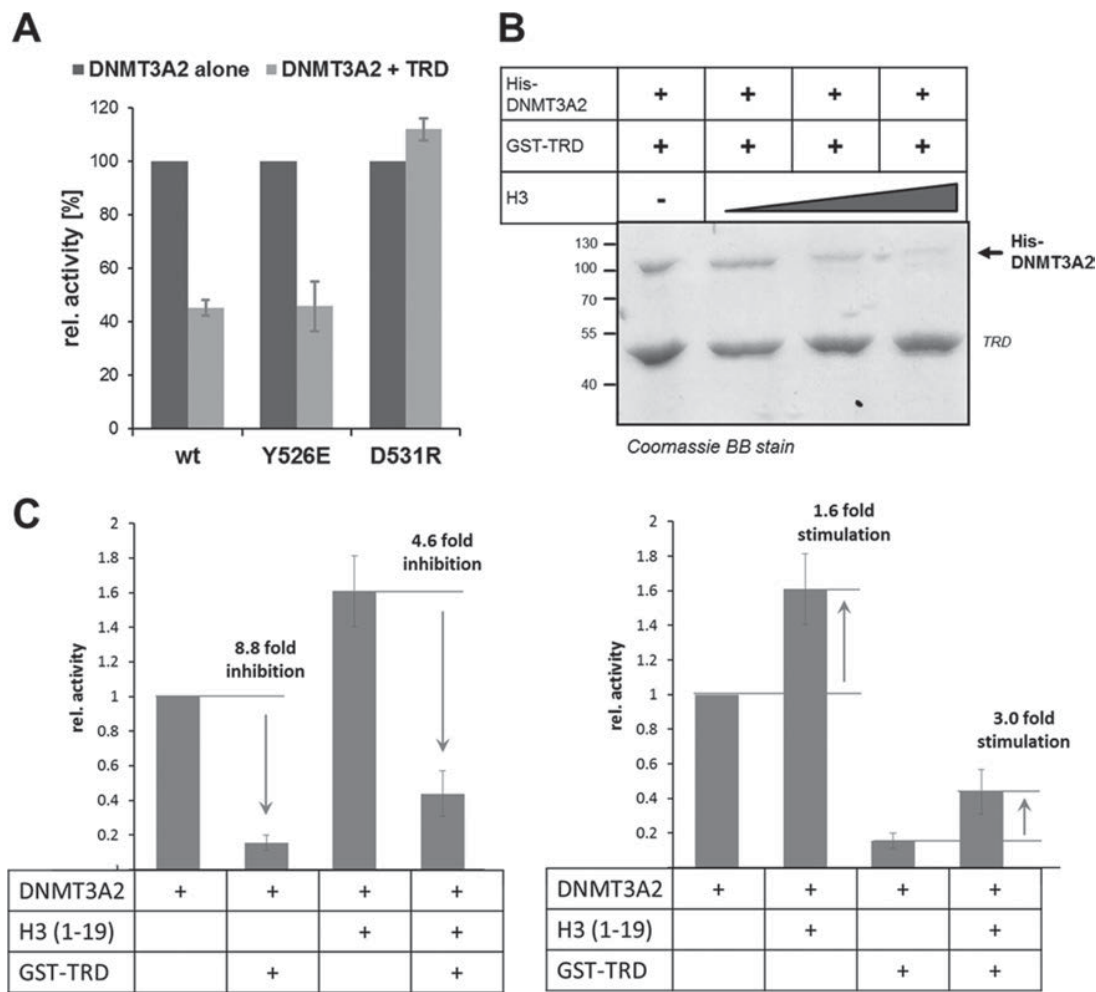
Figure 6. Mechanism of the inhibition of DNMT3A by the MeCP2–TRD. (**A**) Relative *in vitro* activity of DNMT3A2 wild-type and its conformational variants in the absence (dark gray) or presence (light gray) of MeCP2–TRD. The inhibition by the TRD is lost in the DNMT3A2 D531R variant, which carries a mutation that disrupts the autoinhibitory conformation. Error bars indicate the SEM based on three independent experiments. (**B**) Coomassie staining of the pull-down of His-DNMT3A2 (0.25 μM) by GST-TRD in the presence of increasing concentrations of recombinant histone H3 (0, 0.26, 1.3 and 3.9 μM) indicating that the H3 binding to DNMT3A2 interferes with the TRD interaction. (**C**) DNA methylation activity of DNMT3A2 (1 μM) in the absence or presence of the H3 peptide (amino acid sequence 1–19, 5 μM) or TRD (1.2 μM). The two panels show the same data in different representation. Error bars indicate the SEM based on three independent experiments. See also Supplementary Figures S16 and S17.

the fact that MeCP2 is highly expressed in neurons and it has important functions in this cell type, the newly discovered DNMT3A–MeCP2 interaction likely plays an important role in controlling DNA methylation patterns in the brain.

By employing *in vitro* methyltransferase assays using recombinant proteins and a variety of DNA substrates, we observed an almost complete, concentration-dependent inhibitory effect caused by MeCP2 binding to DNMT3A. Inhibition of DNMT3A was observed on both CpG and non-CpG substrates. Furthermore, DNMT3B activity was comparably reduced, proposing a conserved mode of action. To our knowledge, MeCP2 is the first interactor of DNMT3 proteins shown to have a direct inhibitory effect on the enzymatic activity of these proteins. As MeCP2 is an important reader of 5mC and 5hmC, this interaction might be required for mediating the crosstalk between 5mC/hmC sites and DNMT3 proteins and for preventing ectopic *de novo*

methylation. By using engineered conformationally locked DNMT3A variants as a novel tool to investigate DNMT3A regulation, we show that the inhibition of DNMT3A by MeCP2 occurs by an allosteric mechanism, in which binding of MeCP2 stabilizes the autoinhibitory conformation of DNMT3A. Interestingly, binding of the unmodified H3 N-terminal tail peptide to the ADD domain of DNMT3A was shown to have the opposite effect, by precluding the autoinhibitory conformation and leading to the activation of DNMT3A (Figure 7A) (13,14). We mechanistically addressed this crosstalk and show that binding of H3 and TRD to DNMT3A are mutually exclusive and the MeCP2-mediated inhibition of DNMT3A2 can be overcome by addition of the unmodified H3 tail peptide.

In summary, our data unravel one part of the intricate regulatory network, which controls DNA methylation by suggesting a model in which DNMT3A is under the combined control of MeCP2 and the modification state of his-
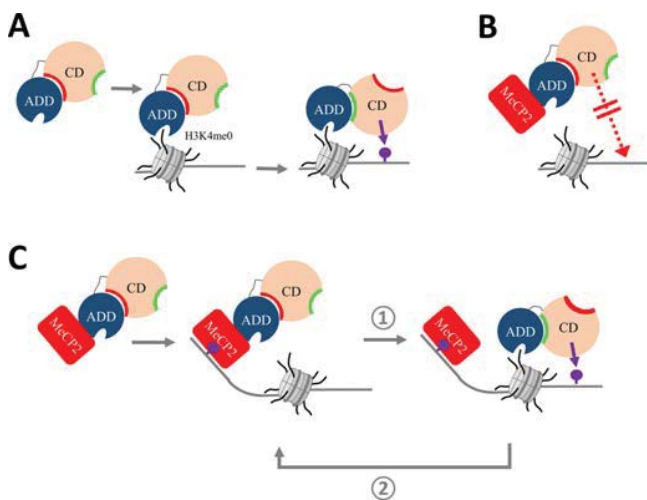
Figure 7. Model of the dual role of MeCP2 in the regulation and targeting of DNMT3A. (**A**) Binding of unmodified H3K4 to the DNMT3A–ADD domain triggers a conformational change where the ADD moves from the autoinhibitory (red) to the allosteric (green) interaction site and DNA methylation can take place (purple lollipop). Model based on (13). (**B**) Inhibitory role of MeCP2 on DNMT3A. Binding of MeCP2 (red) to DNMT3A–ADD inhibits the methyltransferase by stabilizing the autoinhibitory conformation, thereby preventing untargeted activity. (**C**) Role of MeCP2 in targeting of DNA methylation. At genomic sites with unmodified H3K4, H3 binding to the ADD disrupts the interaction between MeCP2 and DNMT3A leading to the activation of the enzyme and DNA methylation (step ①). MeCP2 can next bind to the methylated CpG sites and recruit additional DNMT3A, thereby initiating a positive feedback loop (step ②).

tone H3 tails. On the one hand, the interaction with MeCP2 globally inhibits DNMT3A activity after overexpression of MeCP2 in tissue culture (Figure 7B). This may act as a safeguard mechanism to protect the genome from aberrant DNA methylation. On the other hand, at specific target sites such as repetitive sequences, where histone H3 lacks activating marks, MeCP2 can function as a recruiter of DNMT3 enzymes (Figure 7C). As shown by our biochemical data, unmodified histone H3 can disrupt the MeCP2–DNMT3A interaction, subsequently leading to the relief of the allosteric inhibition. Therefore, the specific delivery of DNMT3A to such regions by MeCP2 as visualized in the cellular localization experiments can target DNA methylation. Afterward, the elevated DNA methylation may increase the methylcytosine-dependent MeCP2 recruitment to these loci, initiating a positive feedback loop, which can contribute to the stable maintenance of methylation at these sites. In neurons, this process may be further supported by the non-CpG (mainly CpA) methylation introduced by DNMT3A, which is bound by MeCP2 as well. The opposing effect of MeCP2 on DNMT3A as potential inhibitor and stimulator depending on the genomic context agrees well with the dual role of MeCP2 in gene control, either as gene repressor or activator (23,30,39,42).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

REFERENCES

1. Jurkowska,R.Z., Jurkowski,T.P. and Jeltsch,A. (2011) Structure and function of mammalian DNA methyltransferases. *ChemBioChem*, **12**, 206–222.
2. Bergman,Y. and Cedar,H. (2013) DNA methylation dynamics in health and disease. *Nat. Struct. Mol. Biol.*, **20**, 274–281.
3. Jeltsch,A. and Jurkowska,R.Z. (2014) New concepts in DNA methylation. *Trends Biochem. Sci.*, **39**, 310–318.
4. Jeltsch,A. (2002) Beyond Watson and Crick: DNA methylation and molecular enzymology of DNA methyltransferases. *ChemBioChem*, **3**, 274–293.
5. Cheng,X. (1995) Structure and function of DNA methyltransferases. *Annu. Rev. Biophys. Biomol. Struct.*, **24**, 293–318.
6. Jeltsch,A. and Jurkowska,R.Z. (2016) Allosteric control of mammalian DNA methyltransferases - a new regulatory paradigm. *Nucleic Acids Res.*, **44**, 8556–8575.
7. Dhayalan,A., Rajavelu,A., Rathert,P., Tamas,R., Jurkowska,R.Z., Ragozin,S. and Jeltsch,A. (2010) The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. *J. Biol. Chem.*, **285**, 26114–26120.
8. Baubec,T., Colombo,D.F., Wirbelauer,C., Schmidt,J., Burger,L., Krebs,A.R., Akalin,A. and Schubeler,D. (2015) Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. *Nature*, **520**, 243–247.
9. Rondelet,G., Dal Maso,T., Willems,L. and Wouters,J. (2016) Structural basis for recognition of histone H3K36me3 nucleosome by human de novo DNA methyltransferases 3A and 3B. *J. Struct. Biol.*, **194**, 357–367.
10. Ooi,S.K., Qiu,C., Bernstein,E., Li,K., Jia,D., Yang,Z., Erdjument-Bromage,H., Tempst,P., Lin,S.P., Allis,C.D. *et al.* (2007)

DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature*, **448**, 714–717.

11. Otani,J., Nankumo,T., Arita,K., Inamoto,S., Ariyoshi,M. and Shirakawa,M. (2009) Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX-DNMT3-DNMT3L domain. *EMBO Rep.*, **10**, 1235–1241.

12. Zhang,Y., Jurkowska,R., Soeroes,S., Rajavelu,A., Dhayalan,A., Bock,I., Rathert,P., Brandt,O., Reinhardt,R., Fischle,W. *et al.* (2010) Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail. *Nucleic Acids Res.*, **38**, 4246–4253.

13. Guo,X., Wang,L., Li,J., Ding,Z., Xiao,J., Yin,X., He,S., Shi,P., Dong,L., Li,G. *et al.* (2015) Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. *Nature*, **517**, 640–644.

14. Li,B.Z., Huang,Z., Cui,Q.Y., Song,X.H., Du,L., Jeltsch,A., Chen,P., Li,G., Li,E. and Xu,G.L. (2011) Histone tails regulate DNA methylation by allosterically activating de novo methyltransferase. *Cell Res.*, **21**, 1172–1181.

15. Morselli,M., Pastor,W.A., Montanini,B., Nee,K., Ferrari,R., Fu,K., Bonora,G., Rubbi,L., Clark,A.T., Ottonello,S. *et al.* (2015) In vivo targeting of de novo DNA methylation by histone modifications in yeast and mouse. *eLife*, **4**, e06205.

16. Noh,K.M., Wang,H., Kim,H.R., Wenderski,W., Fang,F., Li,C.H., Dewell,S., Hughes,S.H., Melnick,A.M., Patel,D.J. *et al.* (2015) Engineering of a histone-recognition domain in Dnmt3a alters the epigenetic landscape and phenotypic features of mouse ESCs. *Mol. Cell*, **59**, 89–103.

17. Stewart,K.R., Veselovska,L., Kim,J., Huang,J., Saadeh,H., Tomizawa,S., Smallwood,S.A., Chen,T. and Kelsey,G. (2015) Dynamic changes in histone modifications precede de novo DNA methylation in oocytes. *Genes Dev.*, **29**, 2449–2462.

18. Petell,C.J., Alabdi,L., He,M., San Miguel,P., Rose,R. and Gowher,H. (2016) An epigenetic switch regulates de novo DNA methylation at a subset of pluripotency gene enhancers during embryonic stem cell differentiation. *Nucleic Acids Res.*, **44**, 7605–7617.

19. Long,H.K., Blackledge,N.P. and Klose,R.J. (2013) ZF-CxxC domain-containing proteins, CpG islands and the chromatin connection. *Biochem. Soc. Trans.*, **41**, 727–740.

20. Lewis,J.D., Meehan,R.R., Henzel,W.J., Maurer-Fogy,I., Jeppesen,P., Klein,F. and Bird,A. (1992) Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. *Cell*, **69**, 905–914.

21. Meehan,R.R., Lewis,J.D. and Bird,A.P. (1992) Characterization of MeCP2, a vertebrate DNA binding protein with affinity for methylated DNA. *Nucleic Acids Res.*, **20**, 5085–5092.

22. Nan,X., Tate,P., Li,E. and Bird,A. (1996) DNA methylation specifies chromosomal localization of MeCP2. *Mol. Cell. Biol.*, **16**, 414–421.

23. Ausio,J., de Paz,A.M. and Esteller,M. (2014) MeCP2: the long trip from a chromatin protein to neurological disorders. *Trends Mol. Med.*, **20**, 487–498.

24. Klose,R.J. and Bird,A.P. (2006) Genomic DNA methylation: the mark and its mediators. *Trends Biochem. Sci.*, **31**, 89–97.

25. Jones,P.L., Veenstra,G.J., Wade,P.A., Vermaak,D., Kass,S.U., Landsberger,N., Strouboulis,J. and Wolffe,A.P. (1998) Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat. Genet.*, **19**, 187–191.

26. Klose,R.J., Sarraf,S.A., Schmiedeberg,L., McDermott,S.M., Stancheva,I. and Bird,A.P. (2005) DNA binding selectivity of MeCP2 due to a requirement for A/T sequences adjacent to methyl-CpG. *Mol. Cell*, **19**, 667–678.

27. Hansen,J.C., Ghosh,R.P. and Woodcock,C.L. (2010) Binding of the Rett syndrome protein, MeCP2, to methylated and unmethylated DNA and chromatin. *IUBMB Life*, **62**, 732–738.

28. Guo,J.U., Su,Y., Shin,J.H., Shin,J., Li,H., Xie,B., Zhong,C., Hu,S., Le,T., Fan,G. *et al.* (2014) Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. *Nat. Neurosci.*, **17**, 215–222.

29. Gabel,H.W., Kinde,B., Stroud,H., Gilbert,C.S., Harmin,D.A., Kastan,N.R., Hemberg,M., Ebert,D.H. and Greenberg,M.E. (2015) Disruption of DNA-methylation-dependent long gene repression in Rett syndrome. *Nature*, **522**, 89–93.

30. Lyst,M.J. and Bird,A. (2015) Rett syndrome: a complex disorder with simple roots. *Nat. Rev. Genet.*, **16**, 261–275.

31. Fuks,F., Hurd,P.J., Wolf,D., Nan,X., Bird,A.P. and Kouzarides,T. (2003) The methyl-CpG-binding protein MeCP2 links DNA methylation to histone methylation. *J. Biol. Chem.*, **278**, 4035–4040.

32. Kimura,H. and Shiota,K. (2003) Methyl-CpG-binding protein, MeCP2, is a target molecule for maintenance DNA methyltransferase, Dnmt1. *J. Biol. Chem.*, **278**, 4806–4812.

33. Nan,X., Hou,J., Maclean,A., Nasir,J., Lafuente,M.J., Shu,X., Kriaucionis,S. and Bird,A. (2007) Interaction between chromatin proteins MECP2 and ATRX is disrupted by mutations that cause inherited mental retardation. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 2709–2714.

34. Bienvenu,T. and Chelly,J. (2006) Molecular genetics of Rett syndrome: when DNA methylation goes unrecognized. *Nat. Rev. Genet.*, **7**, 415–426.

35. Guy,J., Cheval,H., Selfridge,J. and Bird,A. (2011) The role of MeCP2 in the brain. *Annu. Rev. Cell Dev. Biol.*, **27**, 631–652.

36. Hite,K.C., Adams,V.H. and Hansen,J.C. (2009) Recent advances in MeCP2 structure and function. *Biochem. Cell Biol.*, **87**, 219–227.

37. Muotri,A.R., Marchetto,M.C., Coufal,N.G., Oefner,R., Yeo,G., Nakashima,K. and Gage,F.H. (2010) L1 retrotransposition in neurons is modulated by MeCP2. *Nature*, **468**, 443–446.

38. Adkins,N.L. and Georgel,P.T. (2011) MeCP2: structure and function. *Biochem. Cell Biol.*, **89**, 1–11.

39. Chahrour,M., Jung,S.Y., Shaw,C., Zhou,X., Wong,S.T., Qin,J. and Zoghbi,H.Y. (2008) MeCP2, a key contributor to neurological disease, activates and represses transcription. *Science*, **320**, 1224–1229.

40. Ben-Shachar,S., Chahrour,M., Thaller,C., Shaw,C.A. and Zoghbi,H.Y. (2009) Mouse models of MeCP2 disorders share gene expression changes in the cerebellum and hypothalamus. *Hum. Mol. Genet.*, **18**, 2431–2442.

41. Sugino,K., Hempel,C.M., Okaty,B.W., Arnson,H.A., Kato,S., Dani,V.S. and Nelson,S.B. (2014) Cell-type-specific repression by methyl-CpG-binding protein 2 is biased toward long genes. *J. Neurosci.*, **34**, 12877–12883.

42. Della Ragione,F., Vacca,M., Fioriniello,S., Pepe,G. and D'Esposito,M. (2016) MECP2, a multi-talented modulator of chromatin architecture. *Brief. Funct. Genomics*, **15**, 420–431.

43. Feng,J., Chang,H., Li,E. and Fan,G. (2005) Dynamic expression of de novo DNA methyltransferases Dnmt3a and Dnmt3b in the central nervous system. *J. Neurosci. Res.*, **79**, 734–746.

44. Nguyen,S., Meletis,K., Fu,D., Jhaveri,S. and Jaenisch,R. (2007) Ablation of de novo DNA methyltransferase Dnmt3a in the nervous system leads to neuromuscular defects and shortened lifespan. *Dev. Dyn.*, **236**, 1663–1676.

45. Feng,J., Zhou,Y., Campbell,S.L., Le,T., Li,E., Sweatt,J.D., Silva,A.J. and Fan,G. (2010) Dnmt1 and Dnmt3a maintain DNA methylation and regulate synaptic function in adult forebrain neurons. *Nat. Neurosci.*, **13**, 423–430.

46. Morris,M.J., Adachi,M., Na,E.S. and Monteggia,L.M. (2014) Selective role for DNMT3a in learning and memory. *Neurobiol. Learn. Mem.*, **115**, 30–37.

47. Zhang,T., Cooper,S. and Brockdorff,N. (2015) The interplay of histone modifications - writers that read. *EMBO Rep.*, **16**, 1467–1481.

48. Bachman,K.E., Rountree,M.R. and Baylin,S.B. (2001) Dnmt3a and Dnmt3b are transcriptional repressors that exhibit unique localization properties to heterochromatin. *J. Biol. Chem.*, **276**, 32282–32287.

49. Jeltsch,A. and Lanio,T. (2002) Site-directed mutagenesis by polymerase chain reaction. *Methods Mol. Biol.*, **182**, 85–94.

50. Jia,D., Jurkowska,R.Z., Zhang,X., Jeltsch,A. and Cheng,X. (2007) Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. *Nature*, **449**, 248–251.

51. Jurkowska,R.Z., Anspach,N., Urbanke,C., Jia,D., Reinhardt,R., Nellen,W., Cheng,X. and Jeltsch,A. (2008) Formation of nucleoprotein filaments by mammalian DNA methyltransferase Dnmt3a in complex with regulator Dnmt3L. *Nucleic Acids Res.*, **36**, 6656–6663.

52. Rajavelu,A., Jurkowska,R.Z., Fritz,J. and Jeltsch,A. (2012) Function and disruption of DNA methyltransferase 3a cooperative DNA binding and nucleoprotein filament formation. *Nucleic Acids Res.*, **40**, 569–580.

53. Ebert,D.H., Gabel,H.W., Robinson,N.D., Kastan,N.R., Hu,L.S., Cohen,S., Navarro,A.J., Lyst,M.J., Ekiert,R., Bird,A.P. *et al.* (2013) Activity-dependent phosphorylation of MeCP2 threonine 308 regulates interaction with NCoR. *Nature*, **499**, 341–345.

54. Jurkowska,R.Z., Siddique,A.N., Jurkowski,T.P. and Jeltsch,A. (2011) Approaches to enzyme and substrate design of the murine Dnmt3a DNA methyltransferase. *ChemBioChem*, **12**, 1589–1594.

55. Roth,M. and Jeltsch,A. (2000) Biotin-avidin microplate assay for the quantitative analysis of enzymatic methylation of DNA by DNA methyltransferases. *Biol. Chem.*, **381**, 269–272.

56. Jurkowska,R.Z., Ceccaldi,A., Zhang,Y., Arimondo,P.B. and Jeltsch,A. (2011) DNA methyltransferase assays. *Methods Mol. Biol.*, **791**, 157–177.

57. Rhee,I., Bachman,K.E., Park,B.H., Jair,K.W., Yen,R.W., Schuebel,K.E., Cui,H., Feinberg,A.P., Lengauer,C., Kinzler,K.W. *et al.* (2002) DNMT1 and DNMT3b cooperate to silence genes in human cancer cells. *Nature*, **416**, 552–556.

58. Egger,G., Jeong,S., Escobar,S.G., Cortez,C.C., Li,T.W., Saito,Y., Yoo,C.B., Jones,P.A. and Liang,G. (2006) Identification of DNMT1 (DNA methyltransferase 1) hypomorphs in somatic knockouts suggests an essential role for DNMT1 in cell survival. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 14080–14085.

59. Rathert,P., Roth,M., Neumann,T., Muerdter,F., Roe,J.S., Muhar,M., Deswal,S., Cerny-Reiterer,S., Peter,B., Jude,J. *et al.* (2015) Transcriptional plasticity promotes primary and acquired resistance to BET inhibition. *Nature*, **525**, 543–547.

60. Fellmann,C., Hoffmann,T., Sridhar,V., Hopfgartner,B., Muhar,M., Roth,M., Lai,D.Y., Barbosa,I.A., Kwon,J.S., Guan,Y. *et al.* (2013) An optimized microRNA backbone for effective single-copy RNAi. *Cell Rep.*, **5**, 1704–1713.

61. Liu,G.J., Cimmino,L., Jude,J.G., Hu,Y., Witkowski,M.T., McKenzie,M.D., Kartal-Kaess,M., Best,S.A., Tuohey,L., Liao,Y. *et al.* (2014) Pax5 loss imposes a reversible differentiation block in B-progenitor acute lymphoblastic leukemia. *Genes Dev.*, **28**, 1337–1350.

62. Schiesser,S., Pfaffeneder,T., Sadeghian,K., Hackner,B., Steigenberger,B., Schroder,A.S., Steinbacher,J., Kashiwazaki,G., Hofner,G., Wanner,K.T. *et al.* (2013) Deamination, oxidation, and C-C bond cleavage reactivity of 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxycytosine. *J. Am. Chem. Soc.*, **135**, 14593–14599.

63. Bashtrykov,P., Rajavelu,A., Hackner,B., Ragozin,S., Carell,T. and Jeltsch,A. (2014) Targeted mutagenesis results in an activation of DNA methyltransferase 1 and confirms an autoinhibitory role of its RFTS domain. *ChemBioChem*, **15**, 743–748.

64. Chen,T., Ueda,Y., Xie,S. and Li,E. (2002) A novel Dnmt3a isoform produced from an alternative promoter localizes to euchromatin and its expression correlates with active de novo methylation. *J. Biol. Chem.*, **277**, 38746–38754.

65. Suetake,I., Mishima,Y., Kimura,H., Lee,Y.H., Goto,Y., Takeshima,H., Ikegami,T. and Tajima,S. (2011) Characterization of DNA-binding activity in the N-terminal domain of the DNA methyltransferase Dnmt3a. *Biochem. J.*, **437**, 141–148.

66. Jurkowska,R.Z., Rajavelu,A., Anspach,N., Urbanke,C., Jankevicius,G., Ragozin,S., Nellen,W. and Jeltsch,A. (2011) Oligomerization and binding of the Dnmt3a DNA methyltransferase to parallel DNA molecules: heterochromatic localization and role of Dnmt3L. *J. Biol. Chem.*, **286**, 24200–24207.

67. Ramsahoye,B.H., Biniszkiewicz,D., Lyko,F., Clark,V., Bird,A.P. and Jaenisch,R. (2000) Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 5237–5242.

68. Gowher,H. and Jeltsch,A. (2001) 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. *J. Mol. Biol.*, **309**, 1201–1208.

69. Dodge,J.E., Ramsahoye,B.H., Wo,Z.G., Okano,M. and Li,E. (2002) De novo methylation of MMLV provirus in embryonic stem cells: CpG versus non-CpG methylation. *Gene*, **289**, 41–48.

70. Lin,I.G., Han,L., Taghva,A., O'Brien,L.E. and Hsieh,C.L. (2002) Murine de novo methyltransferase Dnmt3a demonstrates strand asymmetry and site preference in the methylation of DNA in vitro. *Mol. Cell. Biol.*, **22**, 704–723.

71. Handa,V. and Jeltsch,A. (2005) Profound flanking sequence preference of Dnmt3a and Dnmt3b mammalian DNA methyltransferases shape the human epigenome. *J. Mol. Biol.*, **348**, 1103–1112.

72. Emperle,M., Rajavelu,A., Reinhardt,R., Jurkowska,R.Z. and Jeltsch,A. (2014) Cooperative DNA binding and protein/DNA fiber formation increases the activity of the Dnmt3a DNA methyltransferase. *J. Biol. Chem.*, **289**, 29602–29613.

73. Lister,R., Pelizzola,M., Dowen,R.H., Hawkins,R.D., Hon,G., Tonti-Filippini,J., Nery,J.R., Lee,L., Ye,Z., Ngo,Q.M. *et al.* (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, **462**, 315–322.

74. Arand,J., Spieler,D., Karius,T., Branco,M.R., Meilinger,D., Meissner,A., Jenuwein,T., Xu,G., Leonhardt,H., Wolf,V. *et al.* (2012) In vivo control of CpG and non-CpG DNA methylation by DNA methyltransferases. *PLoS Genet.*, **8**, e1002750.

75. Lister,R., Mukamel,E.A., Nery,J.R., Urich,M., Puddifoot,C.A., Johnson,N.D., Lucero,J., Huang,Y., Dwork,A.J., Schultz,M.D. *et al.* (2013) Global epigenomic reconfiguration during mammalian brain development. *Science*, **341**, 1237905.

76. Ghosh,R.P., Nikitina,T., Horowitz-Scherer,R.A., Gierasch,L.M., Uversky,V.N., Hite,K., Hansen,J.C. and Woodcock,C.L. (2010) Unique physical properties and interactions of the domains of methylated DNA binding protein 2. *Biochemistry*, **49**, 4395–4410.

77. Baker,S.A., Chen,L., Wilkins,A.D., Yu,P., Lichtarge,O. and Zoghbi,H.Y. (2013) An AT-hook domain in MeCP2 determines the clinical course of Rett syndrome and related disorders. *Cell*, **152**, 984–996.

78. Heyward,F.D. and Sweatt,J.D. (2015) DNA Methylation in Memory Formation: Emerging Insights. *Neuroscientist*, **21**, 475–489.

79. Weaver,I.C. (2014) Integrating early life experience, gene expression, brain development, and emergent phenotypes: unraveling the thread of nature via nurture. *Adv. Genet.*, **86**, 277–307.

80. Shin,J., Ming,G.L. and Song,H. (2014) DNA modifications in the mammalian brain. *Philos. Trans. R. Soc. Lond. Series B, Biol. Sci.*, **369**, 20130512.

81. Yang,L., Rau,R. and Goodell,M.A. (2015) DNMT3A in haematological malignancies. *Nat. Rev. Cancer*, **15**, 152–165.

3.1.4 Label-Free Quantification of 5-Azacytidines Directly in the Genome

Sarah Schiffers#, Thomas M. Wildenhof#, Katharina Iwan, Michael Stadlmeier, Markus Müller, Thomas Carell

(# These authors contributed equally to this publication)

Prologue

Azacytidines have long been used for the treatment of acute myeloid leukemia and myelodysplastic syndromes. The instability of 5-azacytidines, however, compromises the analysis of their mode of action in patient samples. Chemical stabilization of the drug through reduction to dihydro-azacytidine, and optimized digestion conditions overcome this challenge. The reported new UHPLC-MS/MS method enables exact quantification of the dihydro-nucleoside and the respective methyl-cytidine in DNA and RNA. Analysis of cancer and mouse embryonic stem cells reveals similar effects of the drugs on m^5dC, although the azacytidines get incorporated to four-fold different amounts. In addition, different Dnmt KO mouse embryonic stem cells show an overall comparable reduction of m^5dC. In contrast, the m^5C levels were not affected.

Author contribution

Method development on the hydrogenated 5-azacytidine derivative, optimization of the digest protocols and measurements of all RNA samples were performed by *Thomas M. Wildenhof* and me. Furthermore, I conducted all experiments with Dnmt KO cell lines. Additional details on the author contribution are listed in the manuscript.

License

Copy of the article by publishing company; Copyright 2019 *Wiley Company*.

FULL PAPER

Label-Free Quantification of 5-Azacytidines Directly in the Genome

Sarah Schiffers,[a] Thomas M. Wildenhof,[a] Katharina Iwan,[a] Michael Stadlmeier,[a] Markus Müller,[a] and Thomas Carell*[a]

[a] Center for Integrated Protein Science at the Department of Chemistry, Ludwig-Maximilians-Universität München, Butenandtstr. 5 – 13, DE-81377 München, e-mail: Thomas.Carell@lmu.de

Dedicated to Prof. *François Diederich*

Azacytidines (AzaC and AzadC) are clinically relevant pharmaceuticals that operate at the epigenetic level. They are integrated into the genome as antimetabolites to block DNA methylation events. This leads to a reduction of the 5-methyl-2′-deoxycytidine (m^5dC) level in the genome, which can activate epigenetically silenced genes. Because of the inherent chemical instability of Aza(d)Cs, their incorporation levels in DNA and RNA are difficult to determine, which hinders correlation of therapeutic effects with incorporation and removal processes. Existing methods involve radioactive labeling and are therefore unsuitable to monitor levels from patients. We report here a new direct chemical method that allows absolute quantification of the levels of incorporated AzaC and AzadC in both RNA and DNA. Furthermore, it clarifies that Aza(d)C accumulates to high levels (up to 12.9 million bases per genome). Although RNA-based antimetabolites are often 2′-deoxygenated *in vivo* and incorporated into DNA, for AzaC we see only limited incorporation into DNA. It accumulates predominantly in RNA where it, however, only leads to insignificant demethylation.

Keywords: mass spectrometry, leukemia, DNA methylation, azacytidine, DNA methyltransferases.

Introduction

Methylation of deoxycytidines in genomic CpG context creates methylated palindromic (^mCpG) sites, which trigger the silencing of gene expression.[1] Silencing of tumor suppressor genes in turn is a hallmark of cancer.[2] Others and us could recently show, that the inability to remove methyl marks from ^mCpG-islands is a problem in many tumors that helps maintaining uncontrolled cell division and hence tumor growth.[3] The RNA nucleoside 5-Azacytidine (AzaC) and its corresponding DNA analogue 5-Aza-2′-deoxycytidine (Decitabine, AzadC) are pharmaceuticals, which are in clinical use for the treatment of myelodysplastic syndromes (MDS) and acute myeloid leukemia (AML).[4–6] These compounds are prodrugs, which are converted into the corresponding triphosphates and incorporated into both DNA and RNA at levels that are difficult to measure and therefore often unknown.[7,8] Once incorporated, they function as suicide inhibitors of methyltransferases (Dnmt1 and 3a/3b) as depicted in *Scheme 1,a*.[9–12] This inhibitory effect leads to a global reduction of the m^5dC levels in DNA[13] and consequently to a reactivation of silenced tumor-suppressor genes.[14,15] Methylation of cytidine bases (m^5C) occurs also in RNA, and is performed by Dnmt2[16] and specific NOP2/Sun domain family proteins (NSUN2[17], 4,[18] and 6[19]). The question to which extent inhibition of methylation in RNA contributes to the clinical effect of AzaC treatment is unanswered. A recent meta-analysis seems to back up previous findings,[20–22] that AzaC gives slightly better clinical results than AzadC.[23] The molecular cause however remains elusive, especially considering the fact that AzaC first has to be reduced by ribonucleotide reductase to enter DNA. Accordingly, it is important to investigate the levels at which AzadC and AzaC are integrated into nucleic acids.
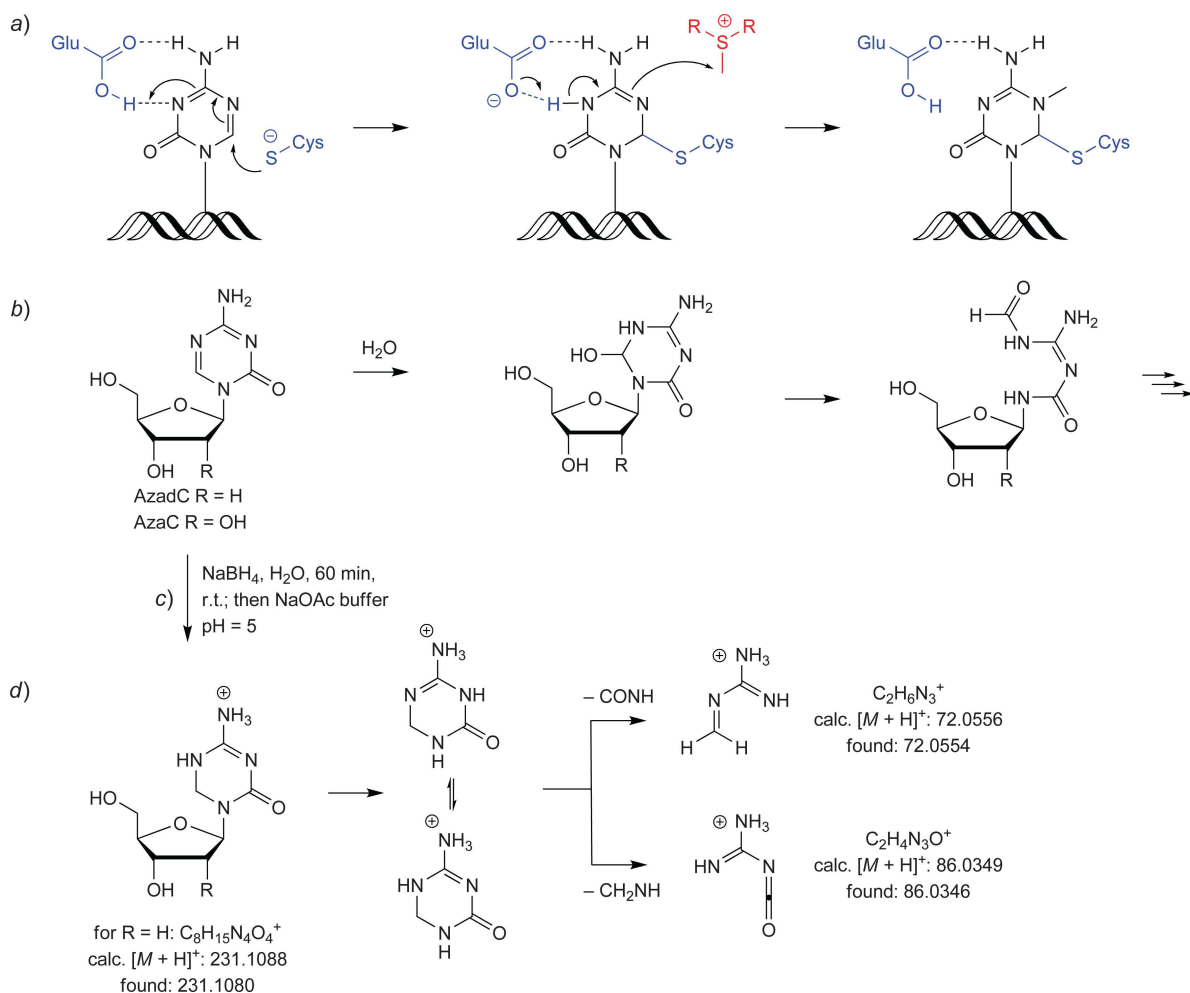
Scheme 1. *a*) Proposed mechanism of action of 5-Azacytidine (Aza(d)C). The blue components are part of the active site of the DNA methyltransferases (DNMTs). The active part of the SAM cofactor is depicted in red. *b*) Depiction of the main hydrolysis pathway of Aza(d)C. *c*) Stabilization of AzadC by NaBH$_4$ reduction and *d*) fragmentation pathway of H$_2$—AzadC with the calculated and found *m/z* values in MS^2 experiments.

Results and Discussion

A major problem associated with the analysis of the incorporation efficiencies is the hydrolytic instability of the Aza(d)C compounds, which feature reported half-life times between 3.5 h and 21 h.[24,25] This makes the direct measurement of the compounds in DNA and RNA impossible. Feeding of radioactive Aza(d)C is only of limited use due to its instability because one is unable to distinguish intact integrated material from chemically unreactive fragments that are still present in DNA and RNA.[26] In light of observed resistance phenomena in treated patients, there is a great need for a direct analytic method that can give levels of intact AzadC and AzaC in DNA and RNA.[27,28]

The instability of Aza(dC) is caused by its electrophilic character, which allows water to attack the C(6) position as depicted in *Scheme 1,b*. This is followed by opening of the hemiaminal substructure and subsequent deformylation and deribosylation.[24][25] We rationalized that any analysis of intact Aza(d)C after DNA or RNA isolation would require immediate stabilization of the incorporated compounds to stop further degradation during DNA and RNA isolation and handling.

We found that treatment of Aza(d)C with NaBH$_4$[29–32], is a very efficient reaction that leads to the formation of the corresponding dihydro—Aza(d)C (H$_2$—Aza(d)C) compounds. We furthermore discovered that despite the lack of any aromaticity, these

compounds are surprisingly stable. When we reacted the Aza(d)C nucleoside with aqueous $NaBH_4$ followed by elimination of borate with acetate buffer (pH = 5, *Scheme 1,c*) we noted full conversion to the corresponding stabilized H_2—Aza(d)C versions already after 60 min reaction at room temperature (*Figure SI-6 A* in the *Supporting Information*). A long-term NMR study showed that H_2—AzadC is stable in D_2O at 37 °C for several hours (*Figure SI-4*), while the original AzadC shows 16% decomposition already after 6 h. In order to enable LC—MS based quantification, we next analyzed the MS fragmentation patterns of H_2—AzadC (*Scheme 1,d*). The positively charged precursor ion with a mass-to-charge ratio (*m/z*) of 231 fragments first through cleavage of the glycosidic bond. The base heterocycle seems to exist in two tautomeric forms in the gas phase, which undergo retro-*Diels-Alder* fragmentations under elimination of either —HNCO or —CH_2NH. This leads to clearly detectable fragment ions with *m/z* of 72 and *m/z* 86 (*Figure SI-7*). This mechanistic assumption is supported by a study with a monodeutero-H_2—AzadC derivative (*Figure SI–8*, *Scheme SI–1*). We then developed a UHPLC method to separate H_2—Aza(d)C from the canonical nucleosides and modified the enzymatic digestion protocol[33] enabling digestion of DNA and RNA and liberating H_2—Aza(d)C completely (*Figure SI-10 – SI-12*). This method was further validated with different amounts of DNA. A similar method that uses a different sample preparation and a different mass spectrometric method was recently published.[31]

Next, we evaluated the new method in realistic scenarios using AzadC-treated cells. For the study, we used leukemia model cell lines and AzadC concentrations of up to 1 μM. The first study was performed with the leukemia model cell line HL60, which are promyeloblast cells derived from acute promyelocytic leukemia. A second study was performed with the AML cell line MOLM-13 (*Figure 1*). In both cases, the cells were cultured in the presence of increasing concentrations of AzadC for 24 h, which is longer than the half-life of the compound in solution.[24,25] The DNA was isolated, treated with $NaBH_4$, subsequently fully digested and analyzed by UHPLC-MS^2 using the developed protocol.

To our delight, we detected a clear and strong signal for the H_2—AzadC using our method. The signal intensity nicely increased in a dose-dependent manner (*Figure 1*, green bars), proving the presence of intact AzadC in the genome. Using external calibration curves of the H_2—AzadC standard subjected to our optimized digestion conditions, it was now possible to
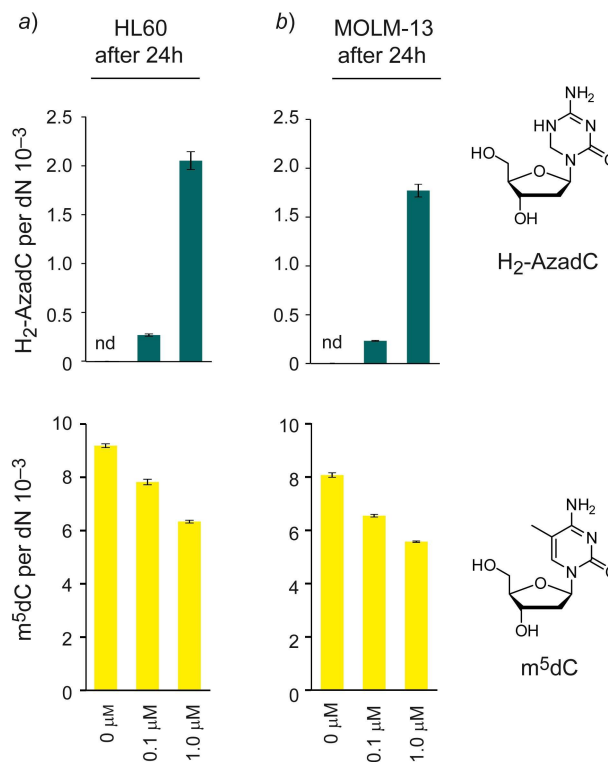
Figure 1. Levels of H_2—AzadC after a 24 h treatment with different concentrations of AzadC. H_2—AzadC (green), m^5dC (yellow) per dN in *a*) HL60 cells, *b*) MOLM-13 cells. nd: not detected. Error bars indicate standard deviation of three independent biological replicates.

perform exact quantification (*Figure SI-10*). We found about 2000 AzadC per million nucleotides when we supplemented with AzadC (1 μM). This corresponds to a rather high level of 12.9 million AzadCs per genome. These high levels may be due to a higher stability of the genome-incorporated AzadC compared to the corresponding free nucleoside, possibly due to shielding of the compound inside the genomic DNA duplex from reaction with water. Indeed, the AzadC content in isolated DNA was more stable (half-life time of 68.7 h at r.t. (*Figure SI-15*)) than the reported half-life time of the nucleoside.[25] However, we strongly recommend performing the $NaBH_4$ treatment as early as possible. We quantified m^5dC in parallel (*Figure 1*, yellow bars), and confirmed that the increase of AzadC goes in hand with a decline of m^5dC as expected, due to the suicide inhibition of the DNA methyl transferases.

We next investigated the effect of AzadC on cells that undergo significant epigenetic reprogramming to see if we could obtain time-dependent data. For these studies, we treated J1 mouse embryonic stem cells

Figure 2. Levels of H$_2$—AzadC (*a*) and H$_2$—AzaC (*b*) in mESC after drug treatment (1 μM Aza(d)C) over a period of 24 h. After the indicated time points the cells were harvested and the DNA was isolated and analyzed as described. H$_2$—AzadC (dark green), m^{5}dC (yellow), H$_2$—AzaC (light green) and m^{5}C (light yellow). nd: not detected; blq: below limit of quantification. Error bars indicate standard deviation obtained from three independent biological replicates.
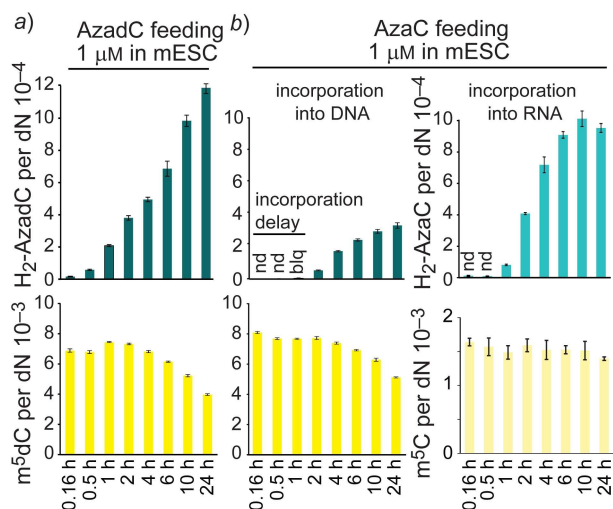
(mESC) with AzadC (1 μM) during the shift from the naïve to the primed state.[34–36] We analyzed the incorporation at different time points (*Figure 2,a*). Indeed, isolation and analysis of the DNA shows an immediate sharp increase of AzadC in the genome that did not reach saturation even after 24 h. Interestingly we noted that despite the immediate integration of AzadC into the genome, a decline of the m^{5}dC values is first observed after 4 h. This observation is difficult to rationalize but it may explain why also in the clinic, long treatment times are essential for therapeutic success. The biochemical reason for this lag phase needs further investigation.

Furthermore, we started to study if our new method would also allow us to determine the incorporation of the ribo-version AzaC into DNA and RNA. We therefore investigated the effect of the AzaC (1 μM) in the mESC model. We first detected again the deoxygenated version AzadC in the genome, showing that 2′-deoxygenation of AzaC occurs and leads to incorporation of AzadC (*Figure 2,b*). However, the detected levels of H$_2$—AzadC are reduced by a factor of about four (330 instead of 1200 H$_2$—AzadC per million nucleotides) and the incorporation is time delayed. The lag phase of 2 h may be attributed to the time needed by the cells to deoxygenate AzaC.[7] The low incorporation yield is most likely to be caused by

decreased availability of the 2′-deoxygenated nucleoside due to the incorporation into RNA and therefore reduction of the soluble pool. Our finding also matches a recent publication[37], in which the ribonucleotide reductase was identified as an AzaC target leading to reduction of the 2′-deoxy-nucleoside pool and therefore probably arrest of the replication.

To our surprise, despite the lower incorporation level, the onset of m^{5}dC reduction is again observed after 4 h and the total decline of m^{5}dC is similar compared to feeding of AzadC. This observation is very interesting and it raises the question of why the observed demethylation is not dose-dependent. From the same sample, we also investigated the levels of the non-deoxygenated AzaC in RNA using a slightly modified UHPLC-MS^2 method. Here we exploit again that the reduction with NaBH$_4$ gives a stable derivative (*Figure SI-5*, *SI-6B*) with a unique high resolution MS fragmentation pattern (*Figure SI-9*).

The detected levels of H$_2$—AzaC in RNA correspond to 1000 AzaC per million nucleotides after 24 h, which is comparable to the H$_2$—AzadC levels in DNA after AzadC treatment. The total amount of incorporated nucleotides is consequently very similar, independent of the supplemented Aza(d)C compound, arguing that maybe proper triphosphate generation could be rate determining *in vivo*. In RNA, m^{5}C does not appear to be significantly reduced upon AzaC feeding, clarifying that demethylation of RNA is likely not responsible for any therapeutic effects.

We conclude that ribosomal m^{5}C represents the vast bulk of the detected material since knockout of Dnmt2, which is known to methylate tRNA, shows only a slight and insignificant effect on global m^{5}C levels (*Figure 3,a*). It seems that in RNA, efficient inhibition of the NSUN-methyltransferases does not occur.

In DNA however, Aza(d)C mediated demethylation appears largely unaffected by the available methyltransferase. We find that equally strong demethylation occurs in all investigated DNMT knockouts, despite efficient incorporation of AzadC into their genomes (*Figure 3,b*). This is somewhat surprising, since inhibition of the maintenance methyltransferase DNMT1 is considered to have the strongest impact on global m^{5}dC levels.[38]

Conclusions

Here, we report two new mass spectrometry-based methods for the exact quantification of AzadC and AzaC in DNA and RNA. Importantly, the methods allow
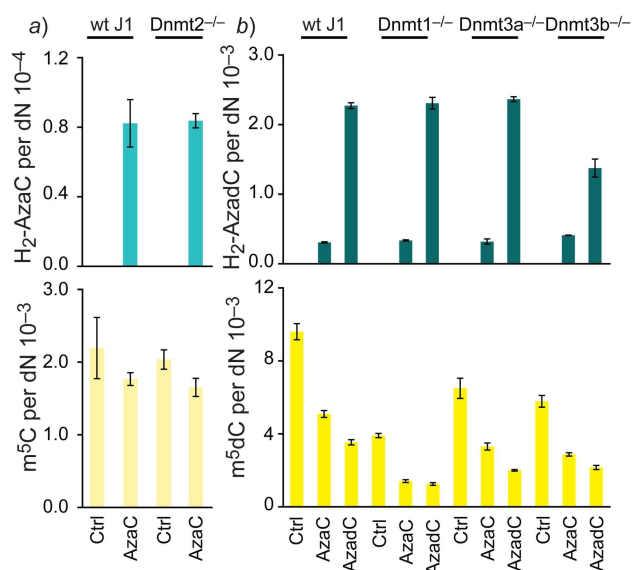
Figure 3. Incorporation of AzadC and AzaC into RNA *a*) and DNA *b*) of various DNMT knock-out mESCs 24 h after drug treatment (1 μM). H_2—AzadC (dark green), m^5dC (yellow), H_2—AzaC (light green), and m^5C (light yellow). nd: not detected. Error bars indicate standard deviation obtained from three independent biological replicates.

quantification of m^5dC in parallel with H_2—AzadC and m^5C in parallel with H_2—AzaC, respectively. This now enables us to correlate incorporation efficiencies with the expected biochemical effect, namely reduction of the $m^5(d)C$ values. Using the new method, we learned that AzaC is efficiently 2′-deoxygenated to AzadC. Both compounds lead to comparable reductions of the m^5dC levels, but not in a dose-dependent manner and only after a significant lag time. Interestingly, we did not see a significant reduction of the m^5C values upon feeding of AzaC, despite its efficient incorporation into RNA, arguing that depletion of RNA methyltransferase may not be accomplished as easily as of DNA methyltransferases. It may also hint at an alternative mechanism of Aza(d)C-mediated demethylation, which acts through replacement of m^5dC by DNA repair processes, rather than inhibition of maintenance methylation. The absence of equivalent repair mechanisms in RNA adds further support to this hypothesis. Given the generally lower DNA incorporation rate in case of AzaC, its particular beneficial effects[23] may arise from the incorporation into RNA rather than DNA. Most importantly, the new methods do not rely on radioactive labeling and can consequently be used to monitor the effects directly in samples from patients treated with Aza(d)C and moreover allow distinguishing catabolic by-products from

the intact drug. This now paves the way to study the pressing resistance problems associated with epigenetic Aza(d)C therapy.

Experimental Section

Chemical Synthesis – General Methods

Preparative HPLC: *Waters 1525* Binary HPLC Pump, *2487* Dual λ Absorbance Detector; *Macherey-Nagel VP 250/10 Nucleosil 100-7-C18*; flow rate 5 mL/min.

Analytical HPLC: *Waters 2695* Separation Module, *2996* Photodiode Array Detector; *Macherey-Nagel EC 250/4 Nucleosil 120-3-C18*; flow rate 0.5 mL/min.

^{1}H- and ^{13}C-NMR spectra were recorded with a *Bruker Avance III HD 400* MHz spectrometer equipped with a CryoProbe. Chemical shifts are expressed in parts per million [ppm] and indicated relative to tetramethylsilane (TMS). The deuterated solvents D_2O served thereby as internal standards. Spin multiplicities are indicated as follows: *s* (*singlet*), *d* (*doublet*), *t* (*triplet*), *q* (*quartet*), *m* (*multiplet*) and combinations thereof. Signals were assigned to their respective source through informally allocated atom numbers. Structural analysis was conducted with ^{1}H- and ^{13}C-NMR spectra under the aid of additional 2D spectra (COSY, HMBC, and HSQC). Spectral analysis was conducted with the software MestReNova v.9.1.0 – 14011 from *Mestrelab Research S.L.*

The high-resolution mass spectrometer for chemicals was operated by the section for mass spectrometry of the department chemistry and pharmacy, LMU Munich. The spectra were acquired through electro spray ionization (ESI) with a *Finnigan LTQ FT* from *Thermo Finnigan GmbH*. Either the molecule ion signal or the signal of another characteristic fragment is indicated in the analysis section of each product.

Melting temperatures were acquired with a *BÜCHI* Melting point *B-450* from *BÜCHI Labortechnik AG* and are uncorrected.

Synthesis of 5,6-Dihydro-5-aza-2′-deoxycytidine (=**4-Amino-1-[(2*R*,4*S*,5*R*)-4-hydroxy-5-(hydroxymethyl)-oxolan-2-yl]-5,6-dihydro-1,3,5-triazin-2(1*H*)-one**; H_2—AzadC). The procedure was modified from a previous publication.[30] In detail, a freshly prepared solution of $NaBH_4$ (9.95 mg, 262.9 μmol, 4 equiv.) in H_2O (1.5 mL) was added to AzadC (15.0 mg, 65.7 μmol,

1.0 equiv., *Carbosynth Limited*) and stirred 1.5 h at r.t. The reaction was quenched with aqueous NaOAc-Buffer (0.5 mL, 750 mM, pH = 5), and the resulting mixture was stirred for 1 h. The mixture was filtered with a syringe filter (*Acrodisc*® 13 mm, 0.2 μm, *GHP Membrane*, *PALL Laboratory*) and directly subjected to preparative HPLC purification and collected as broad peak at *t* = 4.8 min (100% H$_2$O in 25 min). The obtained fractions were pooled and the solvent was removed by lyophilization on a *Christ Alpha L–D* plus to afford H$_2$–AzadC as colorless powder. (14.9 mg, 64.7 μmol, 98%). M.p.: over 230 °C, decomposition. ^{1}H-NMR (400 MHz, D$_2$O): 6.22 (*dd*, *J* = 8.2, 6.5, 1 H); 4.66–4.53 (*m*, 2H); 4.38–4.31 (*m*, 1 H); 3.94–3.82 (*m*, 1 H); 3.75 (*dd*, *J* = 12.3, 4.0, 1 H); 3.67 (*dd*, *J* = 12.3, 5.3, 1 H); 2.25 (*ddd*, *J* = 14.6, 8.2, 6.6, 1 H); 2.06 (*ddd*, *J* = 14.2, 6.5, 3.4, 1 H). ^{13}C-NMR (101 MHz, D$_2$O): 160.3; 159.3; 84.9; 83.7; 70.9; 61.6; 50.9; 35.1. HR-ESI-MS: 231.1088 ([C$_8$H$_{15}$N$_4$O$_4$]$^+$, [*M* + H]$^+$; calc. 231.1088).

Synthesis of 6-Deutero-5-hydro-5-aza-2'-deoxycytidine (= **4-Amino-1-[(2R,4S,5R)-4-hydroxy-5-(hydroxymethyl)oxolan-2-yl](6-^{2}H$_1$)-5,6-dihydro-1,3,5-triazin-2(1H)-one**; MH$_2$–AzadC). The synthesis was performed analogous to H$_2$–AzadC with NaBD$_4$ (11.0 mg, 262.9 μmol, 4.0 equiv.) as reducing agent and afforded 6-monodeutero-5-hydro-5-aza-2'-deoxycytidine (MH$_2$–AzadC; 14.6 mg, 63.1 μmol, 96%) as colorless powder. M.p.: over 230 °C, decomposition. ^{1}H-NMR (400 MHz, D$_2$O): 6.17 (*dd*, *J* = 8.1, 6.6, 1 H); 4.52 (*d*, *J* = 9.2, 1 H); 4.34–4.25 (*m*, 1 H); 3.88–3.79 (*m*, 1 H); 3.69 (*dd*, *J* = 12.3, 3.9, 1 H); 3.61 (*dd*, *J* = 12.2, 5.3, 1 H); 2.19 (*ddd*, *J* = 14.6, 8.2, 6.7, 1 H); 2.04 (*ddd*, *J* = 14.1, 6.4, 3.4, 1 H). ^{13}C-NMR (101 MHz, D$_2$O): 160.5; 159.4; 84.8; 83.6; 70.9; 61.5; 50.5 (*t*, 1 C); 34.97. HR-ESI-MS: 232.1151 ([C$_8$H$_{14}$DN$_4$O$_4$]$^+$, [*M* + H]$^+$; calc. 232.1151).

Synthesis of 5,6-dihydro-5-azacytidine (= **4-Amino-1-[(2R,3R,4S,5R)-3,4-dihydroxy-5-(hydroxymethyl)oxolan-2-yl]-5,6-dihydro-1,3,5-triazin-2(1H)-one**; H$_2$–AzaC). The procedure was modified from a previous publication.[30] In detail, a freshly prepared solution of NaBH$_4$ (9.29 mg, 245.6 μmol, 4 equiv.) in H$_2$O (1.5 mL) was added to AzaC (15.0 mg, 61.4 μmol, 1.0 equiv., *Sigma-Aldrich*) and stirred 1.5 h at r.t. The reaction was quenched with aqueous HCl (0.5 mL, 0.5 M), and the resulting mixture was stirred for 30 min. The mixture was brought to pH = 10 with aqueous NH$_4$OH (5%) and stirred for another 30 min. After filtration with a syringe filter (*Acrodisc*® 13 mm, 0.2 μm, *GHP Membrane*, *PALL Laboratory*), the mixture was directly subjected to preparative HPLC purification

and collected as broad peak at *t* = 5.2 min. (100% H$_2$O in 25 min) The obtained fractions were pooled and the solvent was removed by lyophilization on a *Christ Alpha L–D* plus to afford 5,6-dihydro-5-azacytidine (H$_2$–AzaC) as colorless powder (14.7 mg, 59.7 μmol, 97%). M.p.: over 219 °C, decomposition. ^{1}H-NMR (400 MHz, D$_2$O): 5.64 (*d*, *J* = 6.8, 1 H); 4.52–4.44 (*m*, 2H); 4.14–4.08 (*m*, 1 H); 4.00 (*dd*, *J* = 5.7, 3.6, 1 H); 3.88–3.82 (*m*, 1 H); 3.64 (*dd*, *J* = 12.5, 3.5, 1 H); 3.56 (*dd*, *J* = 12.5, 4.7, 1 H). ^{13}C-NMR (101 MHz, D$_2$O): 160.0; 159.0; 87.1; 83.1; 70.06; 70.04; 61.3; 51.3. HR-ESI-MS: 247.1037 ([C$_8$H$_{15}$DN$_4$O$_5$]$^+$, [*M* + H]$^+$; calc. 247.1037).

Long-Term NMR Study of the Stability of the Compound

A sample of the synthetic H$_2$–AzadC or AzadC, or H$_2$–AzaC and AzaC respectively (*ca.* 1 mg) was dissolved in D$_2$O (1 mL, *Eurisotop*) and immediately subjected to ^{1}H-NMR. The NMR-tube with the sample was then incubated in a water bath at 37 °C and measured after indicated times. After that, the sample was left at r.t. and measured again. For AzadC, integrals of peaks from H–C(1') are indicated, after 6 h at 37 °C, a decay of approximately 16% of the compound can be observed (*Figure SI-4*). For AzaC, integrals of peaks from H–C(1') are indicated, after 18 h at 37 °C, a decay of approximately 37% of the compound can be observed (*Figure SI-5*).

HPLC Conversion Studies

A freshly prepared solution of NaBH$_4$ in H$_2$O (0.5 mL, 250 mM) was added to AzadC (*A*) or AzaC (*B*) (1.00 mg, 4.4 μmol, 1.0 equiv., *Carbosynth Limited*) and stirred at r.t. After 60 min, the mixture (100 μL) was quenched with NaOAc-Buffer (50 μL, 750 mM, pH = 5) and placed in an *Eppendorf Thermomix* comfort at 22 °C and 600 rpm shaking to remove hydrogen bubbles from the reaction. After short-spin centrifugation, the mixture was diluted 1:10 and subjected to analytical HPLC. 0→5% MeCN in H$_2$O, 0→25 min; 5%→80%, 25 min→28 min; 80%→80%, 28 min→38 min; 80%→0%, 38 min→45 min. As control, AzadC or AzaC was diluted in H$_2$O and immediately subjected to analytical HPLC with the same mobile phase gradient (*Figure SI-6*).

HR-MS-Fragmentation

Fragmentation experiments were conducted on an *Orbitrap XL* mass spectrometer (*Thermo Fisher Scientific*), equipped with a *HESI-II-ESI* source (*Thermo Fisher*

Scientific). A solution of the sample in water was directly injected using a syringe pump with a flow rate of 3 μL/min (H$_2$—AzaC) or 5 μL/min (H$_2$—AzadC, MH$_2$—AzadC). Spray parameters are given in *Table SI-5*. The isolation window was set to 1 *m/z*. High-resolution mass spectra were recorded manually with a resolution of 30000 in a mass range from 60 *m/z* to 250 *m/z*. MS^2 and MS^3 spectra for H$_2$—AzadC and MH$_2$—AzadC were recorded with a normalized collision-induced dissociation energy of 20% with a resolution of 30000 in a mass range from 50 *m/z* to 250 *m/z*. In the case of H$_2$—AzaC, MS^2 spectra were acquired with a normalized higher-energy collisional dissociation energy of 30% with a resolution setting of 30000 in a mass range from 65 *m/z* to 300 *m/z*.

To gain a more accurate mass, after acquisition several spectra were summarized in the Xcalibur QualBrowser (*Thermo Fisher Scientific*; *Figure SI-7*). The proposed fragmentation pathway of MH$_2$—AzadC is depicted in *Scheme SI-1*, and complementary to the unlabeled H$_2$—AzaC (*Scheme SI-2*).

Cell Culture and Drug Treatment

5-Azacytidine (*Sigma-Aldrich*) and 5-Aza-2′-deoxycytidine (*Carbosynth*) were dissolved as dimethyl sulfoxide (DMSO) stocks (100 mM) and stored frozen at −80 °C. For treatment of mESC, this stock was diluted to a concentration, that when applied to cell culture medium, the final DMSO concentration did not exceed 1%. Due to the sensitivity of the HL60 cells to DMSO, the DMSO stocks (100 mM) were diluted with ddH$_2$O (to 100× the final concentration) and then directly applied to the culture medium.

Cancer Cell Lines

HL60 cells (*ATCC*) and MOLM-13 (*Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures*) were cultured in *RPMI-1640* Medium (*Sigma-Aldrich*) supplemented with 20% fetal bovine serum (FBS, *Life Technologies*), L-alanyl-L-glutamine (2 mM, *Sigma-Aldrich*) and a mixture of penicillin and streptomycin (100 U/mL, 100 μg/mL, 1×, *Life Technologies*). HL60 cells are promyeloblast cells derived from an acute promyelocytic leukemia, MOLM-13 is an acute myeloid leukemia cell line.[39–41] Cells were incubated in a humidified 37 °C incubator supplied with 5% CO$_2$. For drug treatment, 4×10^6 cells were suspended in culture medium (4 mL) with Aza(d)C and incubated for 24 h in a P60 cell culture dish (*Sarstedt*). The medium

was removed by centrifugation (3 min, 260 *g*) and the cells were washed with phosphate buffered saline (PBS, *Sigma-Aldrich*) and centrifuged again. The pellet was lysed with guanidinium isothiocyanate buffer (1.6 mL, RLT Buffer, *Qiagen*) supplemented with β-mercaptoethanol (final concentration 142 μM, *Sigma-Aldrich*) and subjected to DNA isolation. All cell culture experiments were done in independent biological triplicates.

Mouse Embryonic Stem Cells (mESC)

J1 wt mESCs[42] were maintained in DMEM high glucose (4500 mg/L glucose, sodium pyruvate, and sodium bicarbonate, without L-glutamine, *Sigma-Aldrich*) supplemented with 10% ESC tested FBS (*PAN Biotech*), 1× MEM nonessential amino acids (*Sigma-Aldrich*), L-alanyl-L-glutamine (2 mM, *Sigma-Aldrich*), β-mercaptoethanol (0.1 mM), leukemia inhibitory factor (LIF 1000 U/mL, *ORF Genetics*), and 100 U/mL penicillin with 100 μg/mL streptomycin (*Life Technologies*). For maintaining mESC in the undifferentiated, naive pluripotent state, so called 2i conditions, MEK and GSK3 pathway inhibitors were applied. Therefore, the mESC medium was supplemented with *PD 0325901* (1 μM) and *CHIR 99021* (3 μM, *Axon Medchem*). For all experiments, mESCs were trypsinized with trypsin (0.1%, *Gibco, LifeTechnologies*) in phosphate buffered saline (*Sigma-Aldrich*) containing EDTA (0.02%, *Sigma-Aldrich*), D-glucose (0.01%, *Sigma-Aldrich*) and chicken serum (1%, *Gibco, LifeTechnologies*) and plated in culture dishes pretreated with gelatin (0.2%). mESCs were incubated in a humidified 37 °C incubator supplied with 5% CO$_2$. For drug treatment, cells were moved into the primed state by removing 2i from the medium. Cells were incubated 2 d in medium without 2i in *P60* cell culture dishes (*Sarstedt*). After splitting, 4×10^5 cells were transferred into a 6-well plate culture dish (*VWR*) and incubated additional 2 d without 2i. Then the medium was replaced with Aza(d)C supplemented medium and incubated for another 24 h. The medium was removed and cells were washed with PBS. Then they were directly lysed with RLT Buffer (*Qiagen*) supplemented with β-mercaptoethanol (final concentration 142 μM, *Sigma-Aldrich*) and subjected to DNA isolation.

For comparison of J1 wt mESCs with the respective Dnmt1,[43] Dnmt2,[44] Dnmt3a,[44] and Dnmt3b[44] knockout cell lines the culturing procedure was the same as described above, but after the first passage in

serum/LIF conditions without 2i only 2×10^5 cells were plated per 6-well.

DNA and RNA Isolation

DNA was isolated from cell lysates using *Zymo-Spin™ V* spin columns (*ZymoResearch*), according to the manufacturers' manual with following variation. After DNA binding columns were incubated 5 min with Genomic Lysis Buffer (*ZymoResearch*) supplemented with RNAse A (35 U/mL, *Qiagen*). After the washing steps, DNA was eluted from the column with ddH$_2$O (100–150 μL) containing 3,5-di-*tert*-butyl-4-hydroxy-toluene (BHT, 0.2 μM) and the concentration was determined on a *NanoDrop ND-1000* Spectrophotometer (*NanoDrop Technologies Inc.*).

RNA was isolated from the flow-through of DNA isolations using *ZR-Duet™* DNA/RNA *MiniPrep Kit* (*ZymoResearch*) according to the manufacturer's manual.

NaBH$_4$ Reduction of Isolated DNA and RNA

The DNA or RNA (10 μg) was diluted with ddH$_2$O (up to 75 μL). A freshly prepared solution of NaBH$_4$ (25 μL, 1 M) was added to the sample (final concentration of 250 mM NaBH$_4$) and incubated in the dark in a *Eppendorf Thermomix* comfort at 22 °C and 600 rpm interval shaking (20 s shake, 9 min 40 s interval) for 4 h. This treatment was previously reported[30] in different conditions and was modified to work in water. Its compatibility with genomic DNA was previously shown.[45] Then NaOAc buffer (50 μL, 750 mM, pH = 5) was added carefully to each sample to quench the excess of borohydride and incubated at 22 °C for 2 h at 600 rpm. Remaining hydrogen bubbles were removed by short-spin centrifugation and the DNA was purified and re-isolated using *Zymo-Spin™ IIC–XL* spin columns (*ZymoResearch*) according to the manufacturer's manual. The DNA was eluted in ddH$_2$O (60 μL) and the DNA concentration was determined. RNA was re-isolated accordingly using the *Zymo-Spin™ IIC* spin columns (*ZymoResearch*) according to the manufacturer's manual.

Enzymatic DNA and RNA Digest

DNA and RNA samples were digested to give a nucleoside mixture and spiked with specific amounts of the corresponding isotopically labeled standards before LC–MS/MS analysis. The enzymatic digest method was slightly modified from our previous reported method.[33] Especially TRIS-buffer salts contributed heavily to ion suppression of H$_2$–AzadC and H$_2$–AzaC MS-signals. DNA or RNA (1 μg) was incubated for 3 h at 37 °C in technical triplicate with S1 nuclease (*Sigma-Aldrich*) and Antarctic Phosphatase (*New England BioLabs*) as stated in our reported methods. For RNA, the amount of ZnSO$_4$ was increased (1.6 mM in 7.5 μL), and MgCl$_2$ (2.67 mM in 7.5 μL) was added additionally. Subsequently, for DNA Snake Venom Phosphodiesterase I (*Abnova*) in a glycerol stock was added according to the manufacturer omitting the addition of TRIS-buffer salts and the solution was incubated for another 3 h at 37 °C. For RNA, the first S1/Antarctic Phosphatase addition and incubation was repeated for another 12 h at 37 °C.

UHPLC-MS/MS Analysis

Experimental procedures for synthesis, purification, stock solution preparation, and determination of extinction coefficients for the isotopic nucleoside standards were reported earlier by our group.[33,46–48] In brief, LC-ESI-MS/MS analysis was performed using an *Agilent 1290 UHPLC* system, equipped with an UV-detector, and an *Agilent 6490* triple quadrupole mass spectrometer coupled with the stable isotope dilution technique. The nucleosides were analyzed in the positive ion selected reaction monitoring mode (SRM). In the positive ion mode $[M+H]^+$ species were measured. The optimized general source-dependent parameters were as follows: Gas temp. 80 °C, gas flow 15 L/min (N$_2$), nebulizer 30 psi, sheath gas heater 275 °C, sheath gas flow 11 L/min (N$_2$), capillary voltage 2500 V and nozzle voltage 500 V. The fragmentor voltage was 380 V. Delta EMV was set to 500. For the analysis, we used a *Poroshell 120 SB-C8* column from *Agilent* (2.7 μm, 2.1 mm × 150 mm). The column temperature was maintained at 30 °C. The flow rate was 0.35 mL min^{-1}, and the injection volume amounted to 39 μL. The effluent up to 1.0 min and after 9 min was diverted to waste by a *Valco* valve in order to protect the mass spectrometer. The auto-sampler was cooled to 4 °C.

The dC- and dG-content of DNA samples was determined by LC-UV-detection. The compounds were separated by a gradient using water (0.0090 % *v/v* formic acid) and MeCN (0.0075 % *v/v* formic acid): 0→5 min; 0→3.5 % (*v/v*) MeCN; 5→6.9 min; 3.5→5% MeCN; 6.9→7.2 min; 5→80 % MeCN; 7.2→10.5 min; 80 % MeCN; 10.5→11.3 min; 80→0% MeCN; 11.3→

13 min; 0% MeCN. In addition to our previously reported UHPLC-MS parameters, we implemented parameters for H_2–AzadC in time segment 1.0–4.0 min with a Quantifier and a Qualifier fragmentation (*Table SI-1*).

For RNA samples, the amount of the canonical RNA nucleosides A, C, G, and U was determined by LC-UV detection. For analysis of the H_2–AzaC content of RNA, the compounds were separated by a gradient using water and MeCN, each containing 0.0090% (*v/v*) formic acid: 0→2 min, 0→0% (*v/v*) MeCN; 2→6 min, 0→6% MeCN; 6→7 min, 6→40% MeCN; 7→10.8 min, 40→80% MeCN; 10.8→20 min, 80% MeCN; 20→20.8 min, 80→0% MeCN; 20.8→22 min, 0% MeCN. We implemented UHPLC-MS parameters for H_2–AzaC in time segment 0.8–2.7 min with a Quantifier and a Qualifier fragmentation (*Table SI-2*).

For absolute quantification of H_2–AzadC and H_2–AzaC, we used calibration curves of diluted standards that were measured in technical triplicate prior to every batch (*Figure SI-10*). Each dilution was subjected to the same digest conditions as the DNA or RNA samples to compensate for strong ion suppression of the MS-signal. The resulting calibration curves were then used for quantitation of the samples from the according batch. For each calibration curve, the lower limit of quantification (LLOQ) was defined as the limit, where backfit of the calibration equation was out of a 80%–120% range (*Table SI-3*), and %CV of the median MS-signal was below 15%.

Stability of DNA-Integrated AzadC

In a stability test, we aliquoted a treated, but not reduced DNA sample into two vials, of which we froze one at −20°C and incubated the other one for 24 h at r.t. After the incubation time, both samples were treated with $NaBH_4$ as described and the re-isolated DNA was digested in technical triplicates. The levels of H_2-AzadC were determined (*Figure SI-15*).

Statistical Analysis

UHPLC-ESI-MS/MS data were obtained from three independent biological experiments (unless stated otherwise). Each biological data point was measured as technical triplicate. Error bars represent standard deviation of three independent experiments. Statistical analysis (*Tables SI-8–SI-10*) was performed with Sigma-Plot® software version 11.0 (*Systat Software Inc.*, Chicago, USA), using One-Way ANOVA *Holm-Sidak* as test. Statistical significance is assumed if as * for $p \leq 0.05$, ** for $p \leq 0.01$ and as *** for $p \leq 0.001$.

Author Contribution Statement

S. S. and *T. M. W.* contributed equally to this publication and conducted the experiments. *M. S.* helped with Orbitrap MS and *K. I.* developed the method for quantification of m⁵C. *M. M.* and *T. C.* designed the experiments. *S. S.*, *T. M. W.*, *M. M.* and *T. C.* wrote the manuscript.

References

[1] A. Bird, 'DNA methylation patterns and epigenetic memory', *Genes Dev.* **2002**, *16*, 6–21.

[2] A. Gupta, A. K. Godwin, L. Vanderveer, A. Lu, J. Liu, 'Hypomethylation of the synuclein gamma gene CpG island promotes its aberrant expression in breast carcinoma and ovarian carcinoma', *Cancer Res.* **2003**, *63*, 664–673.

[3] B. Thienpont, J. Steinbacher, H. Zhao, F. D'Anna, A. Kuchnio, A. Ploumakis, B. Ghesquière, L. Van Dyck, B. Boeckx, L. Schoonjans, E. Hermans, F. Amant, V. N. Kristensen, K. Peng Koh, M. Mazzone, M. Coleman, T. Carell, P. Carmeliet, D. Lambrechts, 'Tumour hypoxia causes DNA hypermethylation by reducing TET activity', *Nature* **2016**, *537*, 63–68.

[4] N. Gangat, M. M. Patnaik, A. Tefferi, 'Myelodysplastic syndromes: Contemporary review and how we treat', *Am. J. Hematol.* **2016**, *91*, 76–89.

[5] H. Kantarjian, J. P. J. Issa, C. S. Rosenfeld, J. M. Bennett, M. Albitar, J. DiPersio, V. Klimek, J. Slack, C. de Castro, F. Ravandi, R. Helmer III, L. Shen, S. D. Nimer, R. Leavitt, A. Raza, H. Saba, 'Decitabine improves patient outcomes in myelodysplastic syndromes: results of a phase III randomized study', *Cancer* **2006**, *106*, 1794–1803.

[6] P. Fenaux, G. J. Mufti, E. Hellstrom-Lindberg, V. Santini, C. Finelli, A. Giagounidis, R. Schoch, N. Gattermann, G. Sanz, A. List, S. D. Gore, J. F. Seymour, J. M. Bennett, J. Byrd, J. Backstrom, L. Zimmerman, D. McKenzie, C. L. Beach, L. R. Silverman, 'Efficacy of azacitidine compared with that of conventional care regimens in the treatment of higher-risk myelodysplastic syndromes: a randomised, open-label, phase III study', *Lancet Oncol.* **2009**, *10*, 223–232.

[7] L. H. Li, E. J. Olin, H. H. Buskirk, L. M. Reineke, 'Cytotoxicity and mode of action of 5-azacytidine on L1210 leukemia', *Cancer Res.* **1970**, *30*, 2760–2769.

[8] E. Flatau, F. A. Gonzales, L. A. Michalowsky, P. A. Jones, 'DNA methylation in 5-aza-2′-deoxycytidine-resistant variants of C3H 10T1/2 C18 cells', *Mol. Cell. Biol.* **1984**, *4*, 2098–2102.

[9] J. K. Christman, '5-Azacytidine and 5-aza-2′-deoxycytidine as inhibitors of DNA methylation: mechanistic studies and their implications for cancer therapy', *Oncogene* **2002**, *21*, 5483–5495.

[10] R. Jüttermann, E. Li, R. Jaenisch, 'Toxicity of 5-aza-2′-deoxycytidine to mammalian cells is mediated primarily by covalent trapping of DNA methyltransferase rather than DNA demethylation', *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 11797–11801.

[11] D. Kuch, L. Schermelleh, S. Manetto, H. Leonhardt, T. Carell, 'Synthesis of DNA Dumbbell Based Inhibitors for the Human DNA Methyltransferase Dnmt1', *Angew. Chem. Int. Ed.* **2008**, *47*, 1515–1518.

[12] L. Schermelleh, F. Spada, H. P. Easwaran, K. Zolghadr, J. B. Margot, M. C. Cardoso, H. Leonhardt, 'Trapped in action: direct visualization of DNA methyltransferase activity in living cells', *Nat. Methods* **2005**, *2*, 751–756.

[13] P. A. Jones, S. M. Taylor, 'Cellular differentiation, cytidine analogs and DNA methylation', *Cell* **1980**, *20*, 85–93.

[14] M. Daskalakis, T. T. Nguyen, C. Nguyen, P. Guldberg, G. Köhler, P. Wijermans, P. A. Jones, M. Lübbert, 'Demethylation of a hypermethylated P15/INK4B gene in patients with myelodysplastic syndrome by 5-Aza-2′-deoxycytidine (decitabine) treatment', *Blood* **2002**, *100*, 2957–2964.

[15] M. S. Soengas, P. Capodieci, D. Polsky, J. Mora, M. Esteller, X. Opitz-Araya, R. McCombie, J. G. Herman, W. L. Gerald, Y. A. Lazebnik, C. Cordón-Cardó, S. W. Lowe, 'Inactivation of the apoptosis effector *Apaf-1* in malignant melanoma', *Nature* **2001**, *409*, 207–211.

[16] M. G. Goll, F. Kirpekar, K. A. Maggert, J. A. Yoder, C.-L. Hsieh, X. Zhang, K. G. Golic, S. E. Jacobsen, T. H. Bestor, 'Methylation of tRNA^{Asp} by the DNA Methyltransferase Homolog Dnmt2', *Science* **2006**, *311*, 395–398.

[17] S. Hussain, A. A. Sajini, S. Blanco, S. Dietmann, P. Lombard, Y. Sugimoto, M. Paramor, J. G. Gleeson, D. T. Odom, J. Ule, M. Frye, 'NSun2-Mediated Cytosine-5 Methylation of Vault Noncoding RNA Determines Its Processing into Regulatory Small RNAs', *Cell Rep.* **2013**, *4*, 255–261.

[18] Y. Cámara, J. Asin-Cayuela, C. B. Park, M. D. Metodiev, Y. Shi, B. Ruzzenente, C. Kukat, B. Habermann, R. Wibom, K. Hultenby, T. Franz, H. Erdjument-Bromage, P. Tempst, B. M. Hallberg, C. M. Gustafsson, N. G. Larsson, 'MTERF4 Regulates Translation by Targeting the Methyltransferase NSUN4 to the Mammalian Mitochondrial Ribosome', *Cell Metab.* **2011**, *13*, 527–539.

[19] S. Haag, A. S. Warda, J. Kretschmer, M. A. Günnigmann, C. Höbartner, M. T. Bohnsack, 'NSUN6 is a human RNA methyltransferase that catalyzes formation of m^{5}C72 in specific tRNAs', *RNA* **2015**, *21*, 1532–1543.

[20] R. Itzykson, O. Kosmider, T. Cluzeau, V. Mansat-De Mas, F. Dreyfus, O. Beyne-Rauzy, B. Quesnel, N. Vey, V. Gelsi-Boyer, S. Raynaud, C. Preudhomme, L. Adès, P. Fenaux, M. Fontenay, 'Impact of *TET2* mutations on response rate to azacitidine in myelodysplastic syndromes and low blast count acute myeloid leukemias', *Leukemia* **2011**, *25*, 1147–1152.

[21] R. Itzykson, O. Kosmider, A. Renneville, V. Gelsi-Boyer, M. Meggendorfer, M. Morabito, C. Berthon, L. Adès, P. Fenaux, O. Beyne-Rauzy, N. Vey, T. Braun, T. Haferlach, F. Dreyfus, N. C. P. Cross, C. Preudhomme, O. A. Bernard, M. Fontenay, W. Vainchenker, S. Schnittger, D. Birnbaum, N. Droin, E. Solary, 'Prognostic Score Including Gene Mutations in Chronic Myelomonocytic Leukemia', *J. Clin. Oncol.* **2013**, *31*, 2428–2436.

[22] R. Bejar, A. Lord, K. Stevenson, M. Bar-Natan, A. Pérez-Ladaga, J. Zaneveld, H. Wang, B. Caughey, P. Stojanov, G. Getz, G. Garcia-Manero, H. Kantarjian, R. Chen, R. M. Stone, D. Neuberg, D. P. Steensma, B. L. Ebert, '*TET2* mutations predict response to hypomethylating agents in myelodysplastic syndrome patients', *Blood* **2014**, *124*, 2705–2712.

[23] M. Xie, Q. Jiang, Y. Xie, 'Comparison Between Decitabine and Azacitidine for the Treatment of Myelodysplastic Syndrome: A Meta-Analysis With 1392 Participants', *Clin. Lymphoma Myeloma Leuk.* **2015**, *15*, 22–28.

[24] K.-T. Lin, R. L. Momparlerm, G. E. Rivard, 'High-performance Liquid Chromatographic Analysis of Chemical Stability of 5-aza-2′-Deoxycytidine', *J. Pharm. Sci.* **1981**, *70*, 1228–1232.

[25] D. K. Rogstad, J. L. Herring, J. A. Theruvathu, A. Burdzy, C. C. Perry, J. W. Neidigh, L. C. Sowers, 'Chemical Decomposition of 5-aza-2′-deoxycytidine (Decitabine): Kinetic Analyses and Identification of Products by NMR, HPLC, and Mass Spectrometry', *Chem. Res. Toxicol.* **2009**, *22*, 1194–1204.

[26] S. Öz, G. Raddatz, M. Rius, N. Blagitko-Dorfs, M. Lübbert, C. Maercker, F. Lyko, 'Quantitative determination of decitabine incorporation into DNA and its effect on mutation rates in human cancer cells', *Nucleic Acids Res.* **2014**, *42*, e152.

[27] T. Qin, J. Jelinek, J. Si, J. Shu, J.-P. Issa, 'Mechanisms of resistance to 5-aza-2′-deoxycytidine in human cancer cell lines', *Blood* **2009**, *113*, 659–667.

[28] N. M. Anders, J. Liu, T. Wanjiku, H. Giovinazzo, J. Zhou, A. Vaghasia, W. G. Nelson, S. Yegnasubramanian, M. A. Rudek, 'Simultaneous quantitative determination of 5-aza-2′-deoxycytidine genomic incorporation and DNA demethylation by liquid chromatography tandem mass spectrometry as exposure-response measures of nucleoside analog DNA methyltransferase inhibitors', *J. Chromatogr. B* **2016**, *1022*, 38–45.

[29] M. Matoušová, I. Votruba, M. Otmar, E. Tloušťová, J. Günterová, H. Mertlíková-Kaiserová, '2′-deoxy-5,6-dihydro-5-azacytidine – a less toxic alternative of 2′-deoxy-5-azacytidine: a comparative study of hypomethylating potential', *Epigenetics* **2011**, *6*, 769–776.

[30] J. A. Beisler, M. M. Abbasi, J. A. Kelley, J. S. Driscoll, 'Synthesis and antitumor activity of dihydro-5-azacytidine, a hydrolytically stable analogue of 5-azacytidine', *J. Med. Chem.* **1977**, *20*, 806–812.

[31] A. Unnikrishnan, A. N. Q. Vo, R. Pickford, M. J. Raftery, A. C. Nunez, A. Verma, L. B. Hesson, J. E. Pimanda, 'AZA-MS: a novel multiparameter mass spectrometry method to determine the intracellular dynamics of azacitidine therapy *In Vivo*', *Leukemia* **2018**, *32*, 900–910.

[32] K. S. Harris, W. Brabant, S. Styrchak, A. Gall, R. Daifuku, 'KP-1212/1461, a nucleoside designed for the treatment of HIV by viral mutagenesis', *Antiviral Res.* **2005**, *67*, 1–9.

[33] T. Pfaffeneder, F. Spada, M. Wagner, C. Brandmayr, S. K. Laube, D. Eisen, M. Truss, J. Steinbacher, B. Hackner, O. Kotljarova, D. Schuermann, S. Michalakis, O. Kosmatchev, S. Schiesser, B. Steigenberger, N. Raddaoui, G. Kashiwazaki, U. Müller, C. G. Spruijt, M. Vermeulen, H. Leonhardt, P. Schär, M. Müller, T. Carell, 'Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA', *Nat. Chem. Biol.* **2014**, *10*, 574–581.

[34] G. Ficz, T. A. Hore, F. Santos, H. J. Lee, W. Dean, J. Arand, F. Krueger, D. Oxley, Y.-L. Paul, J. Walter, S. J. Cook, S. Andrews, M. R. Branco, W. Reik, 'FGF Signaling Inhibition in ESCs Drives Rapid Genome-wide Demethylation to the Epigenetic Ground State of Pluripotency', *Cell Stem Cell* **2013**, *13*, 351–359.

[35] E. Habibi, A. B. Brinkman, J. Arand, L. I. Kroeze, H. H. Kerstens, F. Matarese, K. Lepikhov, M. Gut, I. Brun-Heath, N. C. Hubner, R. Benedetti, L. Altucci, J. H. Jansen, J. Walter, I. G. Gut, H. Marks, H. G. Stunnenberg, 'Whole-Genome Bisulfite Sequencing of Two Distinct Interconvertible DNA Methylomes of Mouse Embryonic Stem Cells', *Cell Stem Cell* **2013**, *13*, 360–369.

[36] H. G. Leitch, K. R. McEwen, A. Turp, V. Encheva, T. Carroll, N. Grabole, W. Mansfield, B. Nashun, J. G. Knezovich, A. Smith, M. A. Surani, P. Hajkova, 'Naive pluripotency is associated with global DNA hypomethylation', *Nat. Struct. Mol. Biol.* **2013**, *20*, 311–316.

[37] J. Aimiuwu, H. Wang, P. Chen, Z. Xie, J. Wang, S. Liu, R. Klisovic, A. Mims, W. Blum, G. Marcucci, K. K. Chan, 'RNA-dependent inhibition of ribonucleotide reductase is a major pathway for 5-azacytidine activity in acute myeloid leukemia', *Blood* **2012**, *119*, 5229–5238.

[38] D. Biniszkiewicz, J. Gribnau, B. Ramsahoye, F. Gaudet, K. Eggan, D. Humpherys, M.-A. Mastrangelo, Z. Jun, J. Walter, R. Jaenisch, '*Dnmt1* Overexpression Causes Genomic Hypermethylation, Loss of Imprinting, and Embryonic Lethality', *Mol. Cell. Biol.* **2002**, *22*, 2124–2135.

[39] S. J. Collins, R. C. Gallo, R. E. Gallagher, 'Continuous growth and differentiation of human myeloid leukaemic cells in suspension culture', *Nature* **1977**, *270*, 347–349.

[40] R. Gallagher, S. Collins, J. Trujillo, K. McCredie, M. Ahearn, S. Tsai, R. Metzgar, G. Aulakh, R. Ting, F. Ruscetti, R. Gallo, 'Characterization of the continuous, differentiating myeloid cell line (HL-60) from a patient with acute promyelocytic leukemia', *Blood* **1979**, *54*, 713–733.

[41] Y. Matsuo, R. A. F. MacLeod, C. C. Uphoff, H. G. Drexler, C. Nishizaki, Y. Katayama, G. Kimura, N. Fujii, E. Omoto, M. Harada, K. Orita, 'Two acute monocytic leukemia (AML–M5a) cell lines (MOLM-13 and MOLM-14) with interclonal phenotypic heterogeneity showing MLL-AF9 fusion resulting from an occult chromosome insertion, ins (11;9)(q23;p22p23)', *Leukemia* **1997**, *11*, 1469–1477.

[42] E. Li, T. H. Bestor, R. Jaenisch, 'Targeted mutation of the DNA methyltransferase gene results in embryonic lethality', *Cell* **1992**, *69*, 915–926.

[43] H. Lei, S. P. Oh, M. Okano, R. Jüttermann, K. A. Goss, R. Jaenisch, E. Li, 'De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells', *Development* **1996**, *122*, 3195–3205.

[44] M. Okano, S. Xie, E. Li, 'Dnmt2 is not required for *de novo* and maintenance methylation of viral DNA in embryonic stem cells', *Nucleic Acids Res.* **1998**, *26*, 2536–2540.

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

[46] D. Globisch, M. Münzel, M. Müller, S. Michalakis, M. Wagner, S. Koch, T. Brückl, M. Biel, T. Carell, 'Tissue Distribution of 5-Hydroxymethylcytosine and Search for Active Demethylation Intermediates', *PLoS One* **2010**, *5*, e15367.

[47] M. Münzel, D. Globisch, T. Brückl, M. Wagner, V. Welzmiller, S. Michalakis, M. Müller, M. Biel, T. Carell, 'Quantification of the Sixth DNA Base Hydroxymethylcytosine in the Brain', *Angew. Chem. Int. Ed.* **2010**, *49*, 5375–5377.

[48] S. Schiesser, T. Pfaffeneder, K. Sadeghian, B. Hackner, B. Steigenberger, A. S. Schröder, J. Steinbacher, G. Kashiwazaki, G. Höfner, K. T. Wanner, C. Ochsenfeld, T. Carell, 'Deamination, Oxidation, and C–C Bond Cleavage Reactivity of 5-Hydroxymethylcytosine, 5-Formylcytosine, and 5-Carboxycytosine', *J. Am. Chem. Soc.* **2013**, *135*, 14593–14599.

3.1.5 Influencing epigenetic information with a hydrolytically stable carbocyclic 5 aza-2'-deoxycytidine

Thomas M. Wildenhof, Sarah Schiffers, Franziska R. Traube, Peter Mayer and Thomas Carell

Prologue

During the Ph.D. of *Thomas M. Wildenhof*, an analogue of AzadC, the carbocyclic compound (cAzadC, **Figure 10**) was synthesized. This molecule does not show the same hydrolysis reaction as AzadC. A comparison of the NMR signals immediately after dissolution and after 3 d at r.t., as well as after 14 d showed the same spectrum without any additional signals (see thesis *Thomas M. Wildenhof*). The missing hetero atom in cAzadC must therefore alter the electronic properties of the molecule and hence disable the nucleophilic attack that is known to occur on the C6-position of AzadC.
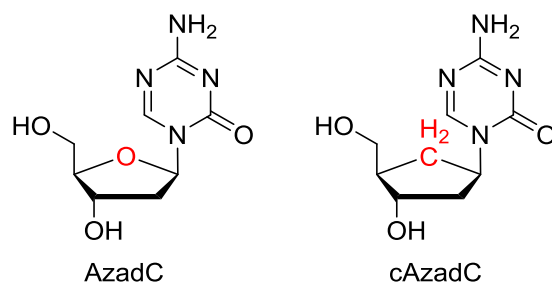

Figure 10: Structures of AzadC (left) and the carbocyclic AzadC derivative (right). The differing atoms are marked in red.

The compound was subsequently added to wt J1 mESCs for 24 h and the effects compared with those of AzadC. Interestingly, the molecule caused a less pronounced decrease of the m^5dC levels (data from *Thomas M. Wildenhof*, not shown), when administered in the same concentration as AzadC. Unfortunately, during this investigation the incorporation of cAzadC was not analyzed. Therefore, direct correlation of the decrease in the m^5dC levels with presence of the drug in the DNA was not possible.

With AzadC, where the enzymes are trapped, a higher amount of covalently bound enzyme leads to a lesser amount of enzyme that is available for maintenance of methylation and subsequently to a quicker decrease of the m^5dC levels. Since with cAzadC the m^5dC levels decrease only to a lesser extent, we had to consider the possibility, that installation of m^5dC is maintained and DNMT enzymes are generally available and active. This would suggest that the enzymes do not get bound covalently to the molecule and would subsequently indicate an inability of the enzymes to act on cAzadC. The decrease of the m^5dC levels could indirectly be caused by the large amount of cAzadC in place of potentially to be methylated dC. As an additional effect, *Thomas M. Wildenhof* observed lower cytotoxicity of the compound in comparison with AzadC.

Project contribution

In a follow-up experiment, *Franziska R. Traube* and I wanted to investigate the decrease of the m^{5}dC levels in wt mESCs that were treated with cAzadC over a longer period of time. We decided to administer the drug in two different concentrations (1 µM and 5 µM, with 0.01% DMSO and 0.05% DMSO, respectively) for 72 h at the end of a 5 d *priming* period. Due to the higher stability of the compound, we did not expect significant hydrolysis of the molecule in the medium. To exclude effects of a potential decrease in the concentration, we set up a second experiment in parallel, where we added fresh supplemented medium every 24 h. In the first 24 h, the previously observed cytotoxic effect is not yet visible under microscopic evaluation. After the 5 d *priming* period, however, the cells treated for 72 h with cAzadC showed a round morphology and significant cell death. With 5 µM of the drug, even the majority of cells had suffered from apoptosis. The DMSO controls were however healthily primed and we can therefore attribute the amount of cell death to effects stemming from the compound. All cells were harvested with RLT$^+$ buffer, the gDNA was subsequently isolated and subjected to total enzymatic digestion as described previously.[161]

The UHPLC-MS method used in the experiments by *Thomas M. Wildenhof* did not record any information for cAzadC. Since this obscured interpretation of the data towards a potential mechanism of action, I developed a new method including the mass transition of cAzadC. Interestingly, the molecule did not only fragment in the typical way for nucleosides (see **Figure 11**) by breaking the N-glycosidic bond, but additionally exhibited elimination of water from the C2'-C3' bond of the carbocycle. This fragmentation was used as a qualifier for the identification of the molecule.
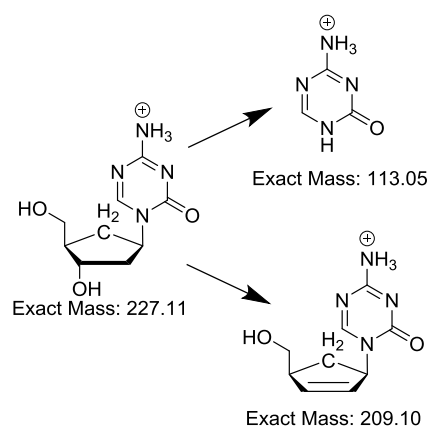
Figure 11: Fragmentation pattern for cAzadC; the upper fragmentation shows the typical N-glycosidic bond cleavage; the lower fragmentation depicts an elimination of water in the carbocycle.

Since no isotope standard of cAzadC was available, exact quantification of the compound was performed using external calibration by measuring a serial dilution of the nucleoside prior to each measurement. Previous analysis of the hydrogenated AzadC compound had revealed significant ion

suppression caused by the digestion conditions. An optimized, but more costly and time-consuming protocol was able to overcome the effects. Although the new compound eluted at a retention time of 2.9 min, where ion suppression is expected to be lower than at the retention time of 1.3 min for H_2-AzadC, we wanted to make sure that the external calibration curve is accurate and that the digestion conditions do not compromise the MS signal. To this end, we measured a single pure serial dilution of the nucleoside, as well as a technical triplicate of the serial dilution that was subjected to the digestion conditions and compared the results. Indeed, the signal intensities and resulting calibration curves did not differ significantly. We therefore decided to perform external calibration by only measuring a pure serial dilution of the nucleoside in bidistilled water, without incubation of the nucleoside with the digestion enzymes.

In a first test, we used 1 µg of gDNA for our sensitive measurements and found out that high amounts of cAzadC were found in the gDNA. The MS signal with the optimized *collision energy (CE)* exceeded standard ranges and was not proportional to the administered amounts. Since the substantial amounts of the nucleoside in the sample saturated the detector, we decided that we did not need maximum sensitivity. Consequently, we selected a different *CE* that gives a less enhanced signal to ensure accurate quantification. With this method in hand, we were able to analyze a full data set with biological triplicates and technical duplicates of m^5dC, hmdC and cAzadC levels in parallel.
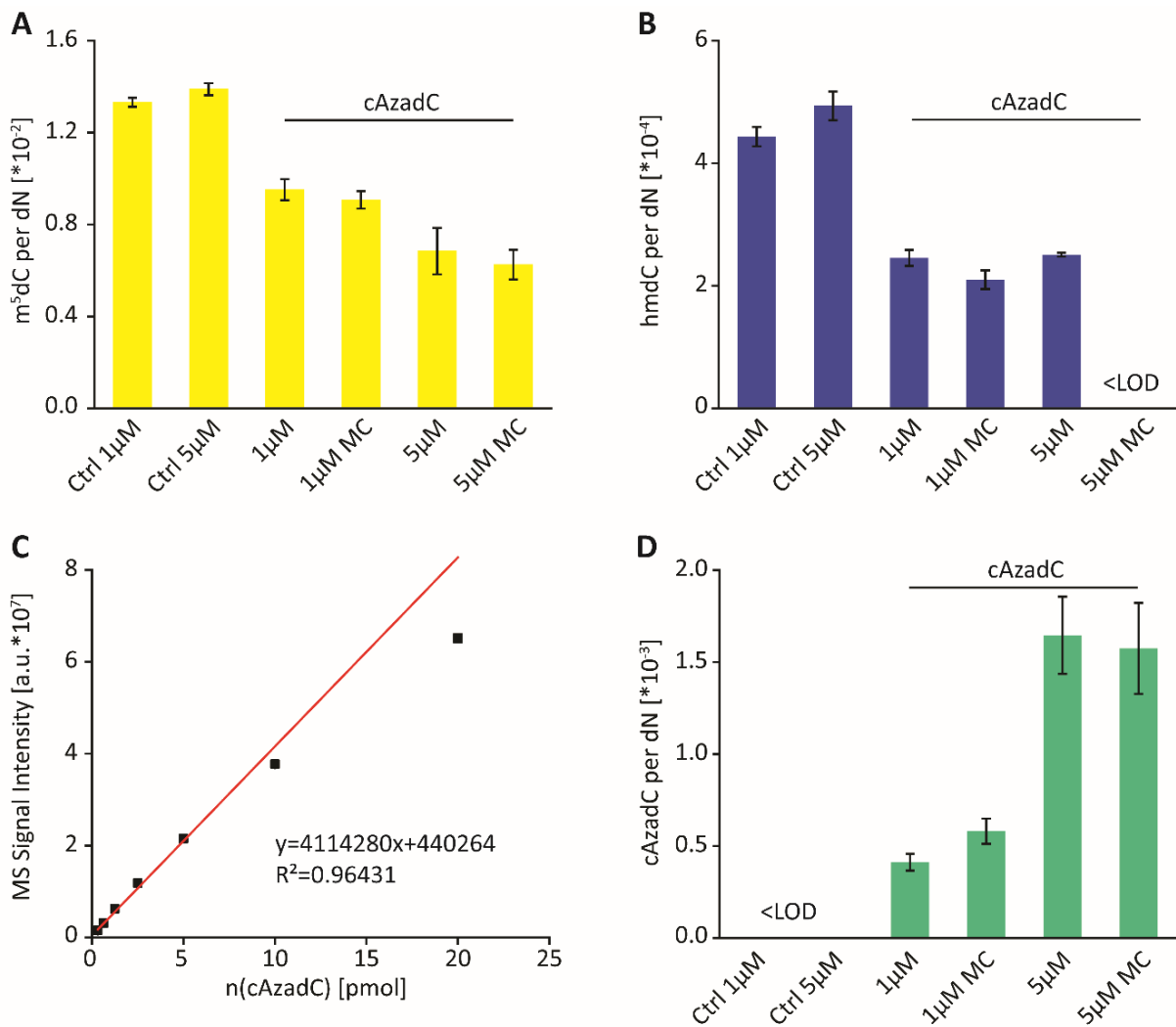
Figure 12: Modification levels upon administration of cAzadC to primed (serum/L, 5 d) wt J1 mESCs. A) m^{5}dC levels, B) hmdC levels, C) external calibration curve for cAzadC, D) levels of cAzadC; Ctrl 1 µM/5 µM=DMSO Ctrl of 0.01% or 0.05%, respectively; MC = medium change.

The result of all samples is displayed in **Figure 12**. The two control samples were supplemented with 0.01% and 0.05% DMSO, respectively, and show only insignificant differences for all analyzed modifications. After administration of the drug for 72 h, a reduction of the m^{5}dC level to 75% of the DMSO control is observed already at 1 µM concentration. The result is the same whether the drug is administered once and the cells are incubated with this solution for 72 h, or the medium is changed and supplemented with fresh cAzadC every 24 h. With 5 µM cAzadC, the decrease is even more pronounced, but only up to 50% of the control levels. In comparison, *Thomas M. Wildenhof* reported a decrease of m^{5}dC to 75% of the controls after 24 h with 5 µM cAzadC, but a reduction to 50% was never observed even at 10 µM. Analyzing the hmdC levels, we observe a drop to 50% of the control levels at both concentrations of the drug. It has to be noted, however, that the hmdC signals were only above LOD in one replicate without medium change and not at all with change of the medium. This inability to quantify above LOD might therefore represent the real effect, hence erasure of hmdC.

Analysis of the cAzadC levels finally reveals incorporation of the drug to substantial amounts of $0.5\text{-}1.5\bullet10^{-3}$ per dN, which is $1/25^{th}$, or $1/5^{th}$ respectively, of the m^{5}dC levels. Interestingly, upon feeding with 5 µM cAzadC, in comparison to 1 µM only a three-fold higher level of the compound is found in the gDNA. In relation to the m^{5}dC levels, administration of 5 µM cAzadC therefore shows a five-fold change that mirrors the difference in the concentration. When we compared these data with the effective amounts of integrated AzadC after administration of 1 µM of the compound for 24 h,[161] we realized that they match those observed with 5 µM cAzadC administered for 72 h. The new compound is therefore able to achieve the same effect on DNA methylation as AzadC. This observation strongly suggests, that the compound also causes entrapment of the DNMT enzymes, but simply needs more time to accumulate in the DNA. A possible explanation for the delayed incorporation is a decreased activity of the nucleotide kinase DCK towards the compound or inefficient uptake of the drug into the cells. Furthermore, it is possible that the DNA polymerases do not accept the cAzadCTP as readily. Due to the hydrolytic stability of the compound, a delay in its incorporation into DNA is not considered a disadvantage.

To summarize the results, cAzadC can be incorporated and quantified to substantial amounts after administration of the drug for 72 h. A decrease of the m^{5}dC levels to 75% and 50% of the control levels can be explained by a direct effect of the drug on the DNMT enzymes. This effect is however strongly time-delayed (72 h instead of 24 h) and depends on a higher concentration of the drug (5 µM instead of 1 µM). Possible explanations for the delayed incorporation comprise reduced activity of DCK, inefficient uptake into the cells or bias of the DNA polymerases. The significant reduction of the hmdC levels remains elusive. In data from *Thomas M. Wildenhof* (see Ph.D. thesis), the levels increased slightly during a time course experiment with administration of AzadC (1 µM). Since this time course was performed during a *priming* procedure, the effects might be due to action of the TET enzymes rather than the drug. In our investigation of cAzadC, the drug was administered earlier during the *priming* procedure (already at day three). A drop of the hmdC levels might indicate an additional effect of the compound on the TET enzymes or other processes that take place during *priming*. The significant occurrence of cell death, and therefore cytotoxicity, of the compound could have a variety of reasons: it might be attributed to effects on the *priming* procedure, but could also hint at efficient activation of normally silenced tumor suppressor genes or interference with other signaling pathways in mESCs. Further analysis of the drug could reveal its mode of action. We conclude that a low dosage of the drug administered over a longer period of time might be a more gentle and cost-effective alternative to treatment with AzadC.

Influencing epigenetic information with a hydrolytically stable carbocyclic 5-aza-2'-deoxycytidine

Thomas M. Wildenhof,[a] Sarah Schiffers,[a] Franziska R. Traube,[a] Peter Mayer[a] and Thomas Carell*[a]

Dedicated to Prof. J. Rebek, Jr. on the occasion of his 75th birthday

Abstract: *5-Aza-2'-deoxycytidine (AzadC) is an antimetabolite in clinical use, which reduces the level of the epigenetic modification 5-methyl-2'-deoxycytidine (mdC). AzadC is incorporated into the genome of proliferating cells, where it inhibits the DNA methyltransferases (DNMTs) in a suicide process leading to a reduction of mdC. The loss of mdC, which is a transcriptional silencer in promoters, leads to the reactivation of genes including tumor suppressor genes, which elicits a beneficial effect. The problem associated with AzadC is that the compound is hydrolytically unstable. It decomposes during treatment to a variety of poorly characterized hydrolysis products. After its incorporation into the genome, this hydrolytic instability generates abasic sites. It is consequently difficult to dissect if the activity of the compound is caused by DNMT inhibition or more generally by DNA lesion formation. We now discovered that a disarmed version of AzadC, in which the ribose oxygen was replaced by a CH2-group, is surprisingly stable under a variety of pH values while keeping the epigenetic activity against the DNMTs.*

5-Aza-2'-deoxycytidine (decitabine, AzadC) is a nucleoside analogue that is able to manipulate epigenetic information.[1-5] Epigenetic information in DNA is associated with the formation of 5-methyl-2'-deoxycytidine (mdC) from 2'-deoxycytidine (dC) with the help of DNA methyltransferases (DNMTs) and *S*-adenosylmethionine (SAM) as the methylating cofactor.[6-7, 4] Methylation of dC to mdC in promoter regions is typically associated with transcriptional silencing of genes.[8-9] AzadC is a prodrug that is inside cells converted into the corresponding active triphosphate and subsequently incorporated into the genome during cell division. The mode of action of AzadC involves reaction of its electrophilic C6 positions with a DNMT active site thiol nucleophile (Fig. 1a).[10-11] This generates a covalent intermediate that is methylated by the SAM cofactor as depicted in Fig 1a. Due to the N-atom at position 5 of the triazine heterocycle, the final β-elimination reaction, which would usually release mdC from the DNMT enzyme, is not possible anymore. The consequence is the formation of a covalent DNA-DNMT crosslink. As a result of administering AzadC, a large drop of the mdC levels (hypomethylating effect) is observed, which leads to the reactivation of silenced tumor suppressor genes in cancer cells.[1] This epigenetic effect is hoped to re-differentiate cancer cells back into normally proliferating cells. AzadC is currently in use as one of the first pharmaceuticals that operates at the epigenetic level for the treatment of myelodysplastic syndromes (MDS)[2] and for acute myeloid leukemia (AML)[4]. Clinically, it is administered in several cycles, with each cycle involving one week of treatment and three weeks of pausing.

The problem associated with AzadC is that the compound hydrolyses in aqueous solution following the path depicted in Fig. 1b. This hydrolysis compromises the activity of AzadC, particularly over the long treatment times. In order to circumvent this problem, it is necessary to generate an AzadC compound that can demethylate (and hence react with an S-nucleophile), while hydrolysis (reaction with an O-nucleophile) should be blocked. Such a compound may allow to dissect how demethylation and lesion formation contribute to the anti-cancer activity, which is an information needed for the design of new epigenetically acting antimetabolites.

Here, we report that replacing the oxygen of the ribose by a CH2-group has a surprisingly large remote effect on the reactivity of the heterocycle. The created carbocyclic version of AzadC (cAzadC, **1**) still inhibits DNMTs but is hydrolytically stable (Fig. 1c).
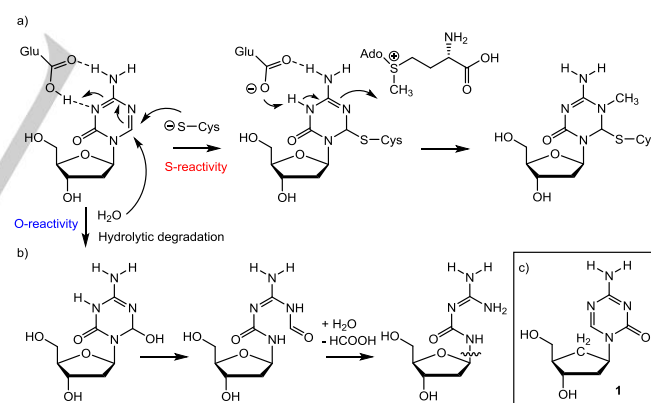

Figure 1. Depiction of 5-aza-2'-deoxycytidine (decitabine, AzadC) together with its mode of action. a) Active site thiol reacts with the C6-position of AzadC. b) Hydrolytic degradation pathway that goes in hand with reaction of a water molecule with the C6-position (O-reactivity) of AzadC. This leads to a final base loss and formation of an abasic site. c) Boxed, depiction of the carbocyclic version cAzadC **1**.

The synthesis of cAzadC **1** is depicted in Scheme 1. It starts with the Boc-protected aminocyclopentane derivative **2** that we used previously to synthesize DNA lesion analogues.[12-15] Compound **2** was first benzyl-protected to **3**, Boc-deprotected to **4**, and then reacted with carbimidazole **5**, which was prepared in two steps from isomethylurea **6** after generation of the free base **7** with potassium hydroxide and reaction of **7** with carbonyldiimidazole.

[a] M. Sc. T. M. Wildenhof, M. Sc. S. Schiffers, M. Sc. F. R. Traube, Dr. P. Mayer, Prof. Dr. T. Carell
Department of Chemistry
Ludwig-Maximilians-Universität, Munich
Butenandtstrasse 5-13
E-mail: Thomas.carell@lmu.de

Supporting information for this article is given via a link at the end of the document

This provides the carbamoylurea-cyclopentane nucleoside analogue **8**. Cyclization to the triazine base **9** was subsequently performed with triethylorthoformate. Reaction of **9** with NH₃ in methanol and deprotection of the benzyl groups with BCl₃ in dichloromethane furnished the final compound cAzadC **1** as the free nucleoside.
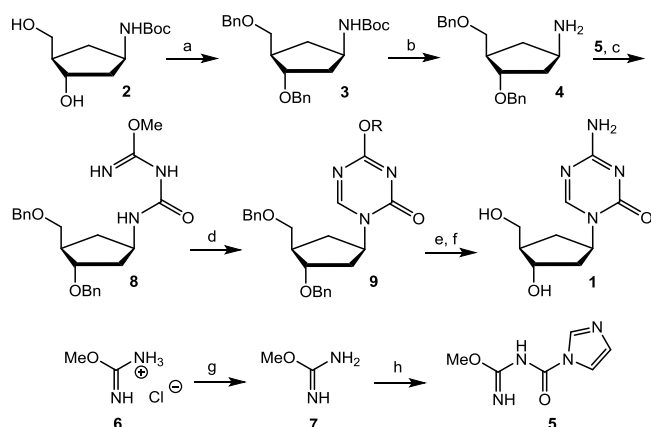
Scheme 1. Synthesis of the carbocyclic 5-aza-2'-deoxycytidine (cAzadC **1**). a) NaH, BnBr, DMF, 0 °C, 1.5 h and stirred for additional 2 h at r.t.; b) TFA (30%), CH₂Cl₂ then Na₂CO₃, 10 min r.t.; c) CH₃CN, reflux, 2 h d) HC(OEt)₃, TFA cat., reflux, 3 h; e) NH₃ (7 N, MeOH), 3 h, r.t., then H₂O ;f) CH₂Cl₂, -78 °C, BCl₃, 1 h, then → r.t., 2 h, MeOH, 20 min. g) KOH, Et₂O:H₂O (39:1), -15 °C, 30 min, h) carbonyldiimidazole, THF, r.t., 3 h; R = Me or Et.

Recrystallization of compound cAzadC **1** from hot methanol gave colourless needles, which allowed us to solve the crystal structure that is depicted in Fig. 2. Interesting is the observation that cAzadC **1** exists with two different cyclopentane conformations in the crystal (Fig. 2; Fig.SI 1). One conformer adopts a C6'-endo ($P = 88.2$ °, $\nu_{max} = 47.8$ °) conformation (Fig. 2a), while the second exists as the C2'-endo-C3'-exo (South, $P = 150.8$ °, $\nu_{max} = 45.4$ °) conformer (Fig. 2b). The latter conformation is typical for 2'-deoxynucleosides in DNA. This shows that the cAzadC **1** nucleoside can adopt the correct DNA-type conformation, fueling hope that the analogue has the potential to get phosphorylated and integrated into the genome.
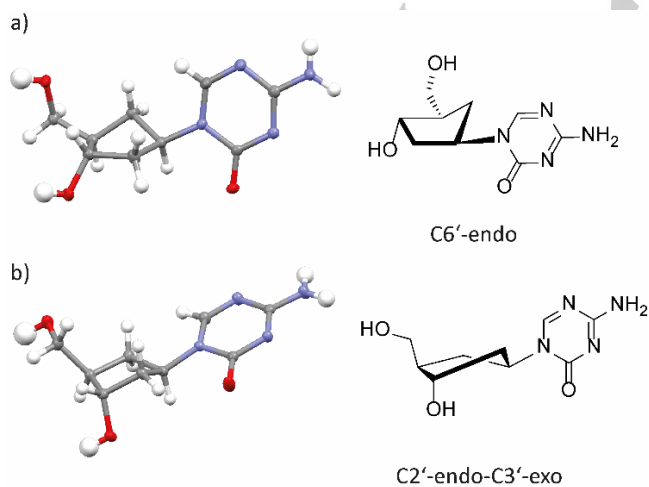

Figure 2. Crystal structure of carbocyclic 5-aza-2'-deoxycytidine (cAzadC **1**) showing the molecule in the observed C6'-endo conformation (a) and the C2'-endo-C3'-exo conformation ($^2T_1$) (b).

We next investigated the stability of cAzadC **1** in direct comparison to the pharmaceutical AzadC (Fig. 3). Since one treatment cycle goes over four weeks we decided to measure the stability at a time point related to a half cycle (14 d). We dissolved AzadC and cAzadC **1** at a concentration of 100 mM in a phosphate buffer (100 mM) at three different pH values (7.4, 5.5 and 8.5) and measured NMR spectra after keeping the solutions at r.t. Since tumour cells often provide an acidic micro-environment,[16] the stability under slightly acidic pH is particularly informative. As evident from the data shown in Fig. 3, the pharmaceutical AzadC strongly degraded within these 14 d. Importantly, at pH = 5.5 and at pH = 8.5, intact AzadC was only hardly detectable anymore. At physiological pH (7.4), AzadC was still present after 14 d but the level of degradation is dramatic. In contrast to these results, we observed for cAzadC **1** surprisingly no degradation at all tested pH values, including pH = 5.5. This result led to the surprising discovery that the simple O → CH₂ exchange causes a strong remote disarming effect that seems to change the properties of the triazine ring so that reaction with water is stopped.
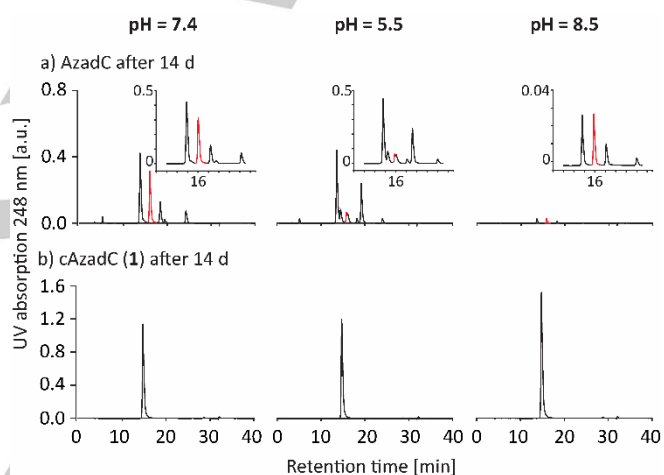

Figure 3. HPLC-based stability measurements showing (a) severe hydrolytic decomposition of 5-aza-2'-deoxycytidine (AzadC) solutions at different pH values, while (b) the carbocyclic compound cAzadC **1** was stable at all three pH values. The inset table in (a) shows the chromatogram between $t_1 = 10$ min and $t_2 = 20$ min for AzadC. The AzadC signal is depicted in red.

We next investigated if this disarming effect would influence the biological functions. We used for this purpose mouse embryonic stem cells (mESC) that were primed in serum/LIF as a model system, since mdC levels increase from naïve to primed state.[17] We added cAzadC **1** in two different concentrations (1 μM and 5 μM) to mESC that have been primed for 48 h and allowed the cells to further proliferate under priming conditions in the presence of cAzadC **1** for additional 72 h. After the 72 h, we harvested the cells, isolated the DNA and digested the DNA down to the nucleoside level using our described protocol.[18] The levels of mdC were finally precisely quantified using isotope dilution UHPLC-MS². To this end, isotopically labelled standards of the nucleosides were spiked in for exact quantification.[19, 18] In addition to mdC, we quantified the levels of 5-hydroxymethyl-2'-deoxycytidine (hmdC), which is formed from mdC by the action of TET enzymes.[20-21] The absolute levels of hmdC are in mESC

COMMUNICATION

more than ten times lower than the mdC levels[20, 22]. The consequence is that even after a substantial reduction of mdC, there should be sufficient mdC to keep the hmdC levels constant. The question if and by how much the hmdC level is affected can therefore inform us about how epigenetic reprogramming is organized. Parallel to the quantification of mdC and hmdC we also quantified to which extent cAzadC **1** itself was incorporated into the genome of the mESC. Detection of AzadC in the genome of treated cells is only possible after treatment of the DNA with $NaBH_4$. Application of $NaBH_4$ reduces the C(5)=C(6) double bond, which stabilizes the compound so that its quantification becomes possible.[23, 19] To our delight, we noted that the stability of cAzadC **1** allowed its quantification without this pre-treatment. We also noted that the applied enzymatic digestion protocol allowed to digest genomic DNA (gDNA) even in the presence of large amounts of cAzadC **1**. Taken together, quantification of cAzadC **1** by UHPLC-MS² using an external calibration curve (Fig. SI2) was possible in parallel to quantification of canonical and epigenetic bases.
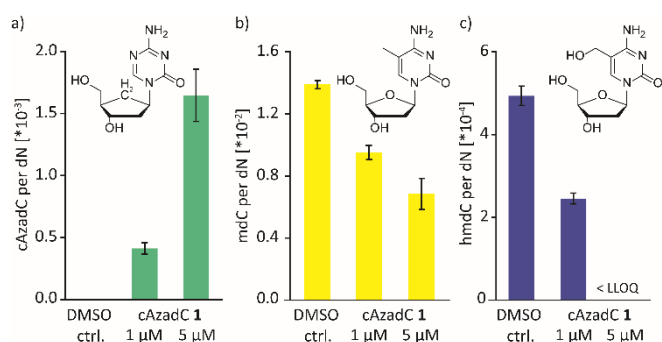

Figure 4. Depiction of the quantification data of DNA modifications of carbocyclic 5-aza-2'-deoxycytidine-treated (cAzadC **1**) mouse embryonic stem cells (mESC) obtained by UHPLC-MS². For each condition, three biological replicates were measured in technical triplicates. For each technical replicate, 0.5 µg of DNA were digested. Bar graphs represent mean, error bars represent standard deviation. LLOQ indicates the lower limit of quantification.

At 1 µM cAzadC **1** concentration, we detected a cAzadC **1** level of 5×10^{-4} cAzadC per dN (Fig. 4a). This amounts to almost 3 million cAzadC nucleotides integrated into the genome. At the higher concentration of 5 µM cAzadC **1**, the level increased 3-fold to 1.7×10^{-3} cAzadC per dN and consequently to more than 8 million integrated cAzadCs per genome. Compared to the incorporation of AzadC, which reaches 1.2×10^{-3} AzadC per dN, when applied with 1 µM[19], the levels of cAzadC **1** reaches about a third of this level. The data clearly show that the carbocyclic version of AzadC (cAzadC **1)** is incorporated and that it reaches in the genome finally comparable levels at 5 µM concentration.

Importantly, after exposing the mESC for 72 h at 1 µM cAzadC in the medium, we detected a reduction of the mdC values by almost 30% (Fig. 4b). At 5 µM concentration in the medium, the mdC levels dropped even to about 50% of the original value. A decrease to 50% is observed for AzadC as well. Here, however, the 50%-reduction is reached faster (24 h) and already with lower AzadC concentration (1 µM).[19] The data show that the carbocyclic version cAzadC **1** needs simply more time to affect the mdC levels by the same amount. We believe that this effect is caused by a potentially slower conversion of cAzadC **1** into the triphosphate. The slower kinetics of cAzadC **1**, however, is not necessarily a disadvantage given the long treatment times that are applied in the clinic.

Very interesting is also the discovery that the hmdC levels were reduced to about 50% already in the 1 µM experiment. At 5 µM, we were even unable to detect hmdC above background levels using 0.5 µg of genomic DNA. The result shows that the hmdC level dropped even faster than the mdC levels, although hmdC is more than ten times less abundant in the genome. This result is interesting. It indicates that hmdC might be potentially predominantly generated in the mdC maintenance process during cell division. We see here that compound cAzadC **1** is a perfect tool molecule that now allows to gain further insight into the interplay between methylation of dC to mdC and oxidation of mdC to hmdC. With the new compound cAzadC **1** in hand we can now begin to clearly correlate demethylation of the genome with the corresponding cellular effects without compromising DNA damaging effects. Finally, cAzadC **1** may not only be a valuable tool compound but potentially even a next generation epigenetic pharmaceutical.

In summary, we show that the replacement of the in-ring O-atom by a CH_2-unit stabilizes the pharmaceutical so that its nucleophilic reaction with water is stopped. The new nucleoside cAzadC **1** is accepted by the phosphorylating enzymes in cells and the corresponding cAzadC-triphosphates are efficiently incorporated into the genome. cAzadC **1** is incorporated in the genome with several million nucleotides and it causes the mdC level to decrease to 70% relative to the control levels.

Experimental Section

Synthesis
All synthetic procedures are described in detail in the supplementary material.

Cell culture of mESC for cAzadC treatment
Feeder independent wt J1 (strain 129S4/SvJae)[24] cells were cultured in the presence of serum and LIF as previously described[25]. They were routinely maintained on gelatinized plates in 2i/L medium. For priming experiments, 2i cultures were passaged when applicable in DMEM supplemented with FBS and LIF as above but lacking the inhibitors. For drug treatment, cells were moved into the primed state by removing 2i from the medium. Cells were incubated 2 d in DMEM supplemented with FBS and LIF in 6-well plates (*VWR*). After splitting, 2×10^5 cells were transferred into a 6-well plate culture dish and supplemented with either 1 µM (in 0.01% DMSO) or 5 µM cAzadC (in 0.05% DMSO) and treated for 72 h. After removal of the medium and washing the cells with DPBS, they were directly lysed with RLT^+ buffer as described in a previous publication[18]

gDNA isolation, total enzymatic digest and UHPLC-MS²
The gDNA was isolated as described previously[18]. Due to the higher stability of cAzadC **1**, a hydrogenation procedure was not necessary and the gDNA was directly subjected to a total enzymatic digest and analyzed using UHPLC-MS² as described in a previous publication.[19]

COMMUNICATION

[1]   M. Daskalakis, T. T. Nguyen, C. Nguyen, P. Guldberg, G. Kohler, P. Wijermans, P. A. Jones, M. Lubbert, *Blood* **2002**, *100*, 2957-2964.

[2]   H. Kantarjian, J. P. Issa, C. S. Rosenfeld, J. M. Bennett, M. Albitar, J. DiPersio, V. Klimek, J. Slack, C. de Castro, F. Ravandi, R. Helmer, 3rd, L. Shen, S. D. Nimer, R. Leavitt, A. Raza, H. Saba, *Cancer* **2006**, *106*, 1794-1803.

[3]   C. Stresemann, F. Lyko, *Int. J. Cancer* **2008**, *123*, 8-13.

[4]   Y. Koh, Y. A. Kim, K. Kim, J.-A. Sim, S.-S. Yoon, S. M. Park, Y. H. Yun, *Blood* **2016**, *128*, 2381-2381.

[5]   M. Nieto, P. Demolis, E. Béhanzin, A. Moreau, I. Hudson, B. Flores, H. Stemplewski, T. Salmonson, C. Gisselbrecht, D. Bowen, F. Pignatti, *Oncologist* **2016**, *21*, 692-700.

[6]   S. S. Smith, B. E. Kaplan, L. C. Sowers, E. M. Newman, *Proc. Natl. Acad. Sci.* **1992**, *89*, 4744-4748.

[7]   G. G. Wilson, R. J. Roberts, S. Kumar, J. Posfai, M. Sha, S. Klimasauskas, X. Cheng, *Nucleic Acids Res.* **1994**, *22*, 1-10.

[8]   A. Bird, *Genes Dev.* **2002**, *16*, 6-21.

[9]   R. J. Klose, A. P. Bird, *Trends Biochem. Sci.* **2006**, *31*, 89-97.

[10]   R. Jüttermann, E. Li, R. Jaenisch, *Proc. Natl. Acad. Sci.* **1994**, *91*, 11797-11801.

[11]   S. Gabbara, A. S. Bhagwat, *Biochem. J.* **1995**, *307*, 87-92.

[12]   M. Ober, H. Müller, C. Pieck, J. Gierlich, T. Carell, *J. Am. Chem. Soc.* **2005**, *127*, 18143-18149.

[13]   H. Müller, T. Carell, *Eur. J. Org. Chem.* **2007**, *2007*, 1438-1445.

[14]   F. Büsch, J. C. Pieck, M. Ober, J. Gierlich, G. W. Hsu, L. S. Beese, T. Carell, *Chem. Eur. J.* **2008**, *14*, 2125-2132.

[15]   T. H. Gehrke, U. Lischke, K. L. Gasteiger, S. Schneider, S. Arnold, H. C. Müller, D. S. Stephenson, H. Zipse, T. Carell, *Nat. Chem. Bio.* **2013**, *9*, 455.

[16]   Y. Kato, S. Ozawa, C. Miyamoto, Y. Maehata, A. Suzuki, T. Maeda, Y. Baba, *Cancer Cell Int.* **2013**, *13*, 89.

[17]   S. Takahashi, S. Kobayashi, I. Hiratani, *Cell. Mol. Life Sci.* **2018**, *75*, 1191-1203.

[18]   F. R. Traube, S. Schiffers, K. Iwan, S. Kellner, F. Spada, M. Müller, T. Carell, *Nat. Protoc.* **2019**, *14*, 283-312.

[19]   S. Schiffers, T. M. Wildenhof, K. Iwan, M. Stadlmeier, M. Müller, T. Carell, *Helv. Chim. Acta* **2019**, *102*, e1800229.

[20]   M. Tahiliani, K. P. Koh, Y. Shen, W. A. Pastor, H. Bandukwala, Y. Brudno, S. Agarwal, L. M. Iyer, D. R. Liu, L. Aravind, A. Rao, *Science* **2009**, *324*, 930-935.

[21]   S. Ito, L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He, Y. Zhang, *Science* **2011**, *333*, 1300-1303.

[22]   M. Münzel, D. Globisch, T. Carell, *Angew. Chem. Int. Ed. Engl.* **2011**, *50*, 6460-6468.

[23]   A. Unnikrishnan, A. N. Q. Vo, R. Pickford, M. J. Raftery, A. C. Nunez, A. Verma, L. B. Hesson, J. E. Pimanda, *Leukemia* **2018**, *32*, 900-910.

[24]   E. Li, T. H. Bestor, R. Jaenisch, *Cell* **1992**, *69*, 915-926.

[25]   T. Pfaffeneder, B. Hackner, M. Truss, M. Münzel, M. Müller, C. A. Deiml, C. Hagemeier, T. Carell, *Angew. Chem. Int. Ed. Engl.* **2011**, *50*, 7008-7012.

3.2 Unpublished results

3.2.1 Investigation of the formation of m⁶dA in gDNA upon exogenous stimuli

We discovered previously[374] (see section **3.1.1**) that m⁶dA does not seem to be a highly abundant modification in human or murine cells or tissues. The question as to why this modification is frequently found in these organisms still remained. In the article, we also report an incorporation of the nucleoside into the gDNA upon administration of the free nucleoside. We therefore suggest that the occurrence of the modification can be explained from a dynamic endogenous or exogenous source. According to a publication by *Liu et al.*,[83] m⁶dA is found in early development of zebrafish 0.5-3.5 h post fertilization (hpf), while m⁶A-containing RNA is degraded in this developmental stage (2-8 hpf).[375] These results led us to hypothesize that there might be a connection between m⁶A-containing RNA degradation and the occurrence of m⁶dA in DNA.

3.2.1.1 Administration of free m⁶A to different cell lines

By supplementing growth media with 1 µM/1 mM m⁶dA, we have previously shown that genomic m⁶dA may stem from external sources, such as media components or bacterial degradation products. We now wanted to elucidate also internal sources of this modification. As a first experiment, we administered the free RNA nucleoside m⁶A to wt J1 mESCs and the somatic cells HeLa and HEK293T, since ribonucleotide reductase (RNR) may lead to the conversion of phosphorylated m⁶A to the 2'-deoxy nucleoside. Subsequently, we analyzed the DNA for m⁶dA. Feeding of HeLa and HEK293T cells was only performed once.
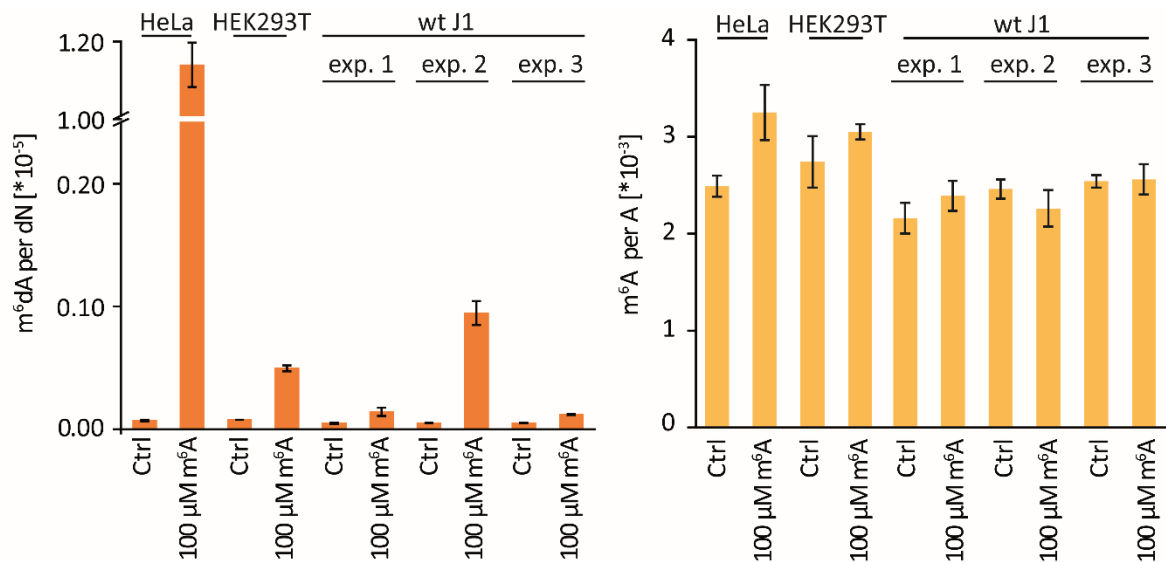
Figure 13: Levels for DNA incorporated m⁶dA and respective m⁶A levels in RNA after administration of m⁶A to the medium of mESCs and cancer cells. For wt J1 three independent biological samples are depicted (Exp.: experiment). As a control, the levels of m⁶dA in unfed cells were analyzed. The results for DNA are depicted on the left side and those for the RNA are on the right.

Indeed, administration of a 100 µM solution (0.25% DMSO for wt J1 and 2% for the somatic cell lines) led to clearly detectable elevated levels of m⁶dA in DNA in HEK293T and wt J1 cells, and even high levels in HeLa cells (see **Figure 13**, right side). It is also clear that administration of the free RNA nucleoside leads to a significant incorporation of m⁶A into RNA of HeLa cells (see **Figure 13**, left side), whereas HEK293T and wt J1 cells show no difference.

Interestingly, the incorporation of m⁶dA does not seem to occur to the same levels in different replicates of supplemented mESCs, whereas the m⁶A levels seem to be rather stable between replicates.

To summarize the results, administration of m⁶A to various cell lines gave us evidence for the ability of various cells to process the free m⁶A nucleoside and finally to incorporate it as m⁶dA into the DNA. Due to slight contamination of the commercially available adenosine as the precursor for the m⁶A synthesis with 2'-deoxyadenosine, we could however not exclude the possibility that the incorporated m⁶dA nucleoside was actually an impurity of the administered nucleoside. Nevertheless, the concentration of m⁶dA that was administered in a previous publication[374] lead to barely detectable levels of m⁶dA in the DNA, and the concentrations of potential m⁶dA impurities in the administered m⁶A solution are still expected to differ by several orders of magnitude. The potential impurity of the administered solution would be expected to be present in a constant concentration leading to a constant background signal and not cause fluctuating levels for m⁶dA. The incorporated m⁶dA is therefore likely to stem from the converted m⁶A nucleoside.

3.2.1.2 Transfection of m^{6}A-containing RNA

Like detailed in the introduction, m^{6}A is the most abundant modification in mRNA. Since mRNA is constantly turned over, we hypothesized that this m^{6}A represents a source also for genomic m^{6}dA. Therefore, we thought of simulating degradation of m^{6}A-containing RNA *in vivo*. To this end, we designed the following RNA strand containing six m^{6}A nucleosides:

5'-(6-FAM)-UGm^{6}ACCGm^{6}AUGGm^{6}AGGUm^{6}AGUm^{6}AUm^{6}AG-3'

As a control, we used the same strand with unmodified A. Both forward strands had a 6-Carboxyfluoresceine (6-FAM)-tag at the 5'-end. The tag was used as a fluorescent marker for the transfection efficiency of the RNA. Since transfection is usually performed with dsRNA, we constructed a complementary strand that contained only A. All strands were synthesized by *Matthias Q. Kurz*.

We transfected the 6-FAM-tagged RNA containing m^{6}A into mESCs and monitored the levels of m^{6}dA in DNA. With an increased abundance of free m^{6}A in the cytoplasm, we subsequently expected a higher level of m^{6}dA in DNA. As a control, DNA from untransfected cells and from cells transfected with the RNA strand only containing unmodified A was analyzed.
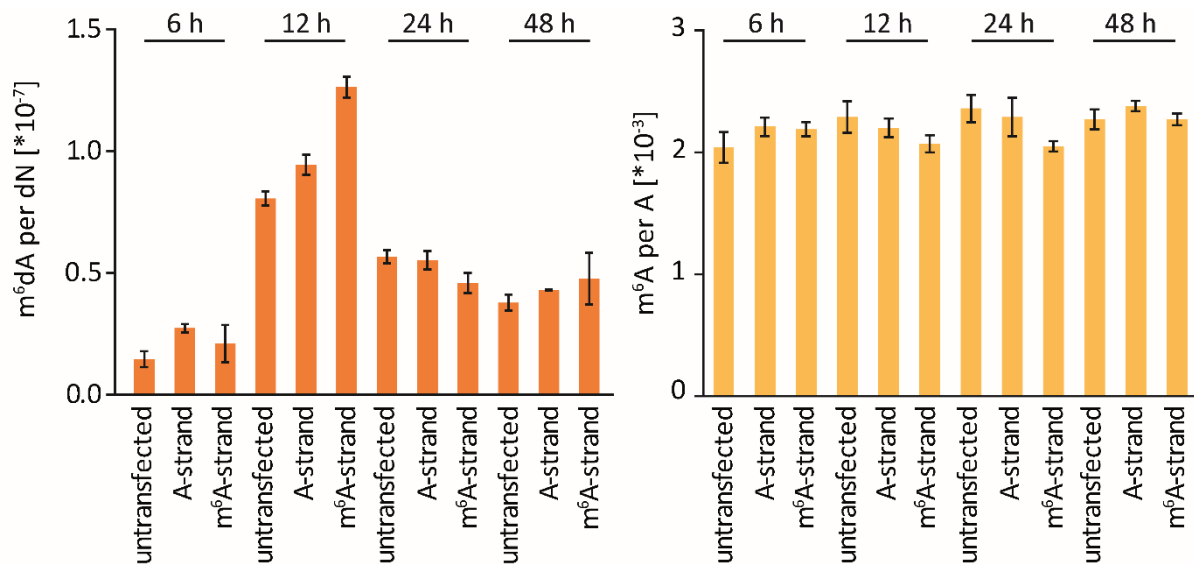
Figure 14: Levels for DNA incorporated m^{6}dA and respective m^{6}A levels in RNA after transfection of m^{6}A-containing RNA at different time points. As a control, the levels of m^{6}dA in untransfected cells and cells transfected with only A-containing RNA were analyzed. The results for DNA are depicted on the left side and those for the RNA are displayed on the right.

The cells were lyzed at the different time points and the gDNA was isolated and enzymatically digested. Analysis of a first biological replicate of DNA indeed showed higher levels of m^{6}dA for the m^{6}A-containing strand 12 h post-transfection (see **Figure 14**). Since the strand has to be degraded to

release free m6A nucleoside, which then needs to be phosphorylated and 2'-deoxygenated to be finally incorporated into the DNA, an increase of m6dA is expected to occur time-delayed to the microscopic observation. The results therefore seem to fit our expectations and hint at a formation of m6dA in the DNA upon degradation of RNA that contains m6A. How this degradation process would take place and in which biological process remains elusive. Curiously, increased signals for the nucleoside were also found in the untransfected and A-strand control, potentially indicating an additional internal source of m6dA nucleosides. It has to be clarified, that the culturing conditions for the untransfected control were not different from those previously described.[374] Longer culturing of the cells without change of medium might therefore lead to higher background levels for the m6dA nucleoside.

Additional analysis of the RNA for changes in the m6A levels was conducted to determine potential effects upon transfection of RNA. After the first 6 h, the levels are the same in the controls and the cells transfected with the m6A-strand. After 12 h and 24 h, a slight and insignificant decrease of m6A can be observed. Upon administration of m6A-containing RNA, one could expect more m6A, but our result could reflect other biochemical processes as reaction towards the foreign RNA. Furthermore, the slight variations might be fluctuations due to the high abundance of the nucleoside.

Repetition of the experiment unfortunately did not show any detectable levels of m6dA above background level. Analysis of the m6A levels in the RNA isolated from the transfected cells also did not show altered levels for the m6A-containing strand. We therefore decided to try a different approach.

3.2.1.3 Induction of differentiation of wt mESCs with all-trans retinoic acid

The cause of the different results from the DNA analysis in section **3.2.1.1** might be explained by an additional interesting observation, which was not recorded visually: we found that in the biological replicate with the highest m6dA level, the cell morphology was different from the controls and the cells seemed to be partly differentiated. Since we only saw this effect once, we were not sure whether this was indeed a consequence of the administration of the nucleoside. In this replicate, the cells might have been seeded at a lower density than in the other experiments and therefore there might have been a higher local concentration of the nucleoside. The DNA results nevertheless suggest that the observed morphology might indeed have been caused by the administered nucleoside. Recent literature also reports a differentiating effect of m6dA on pluripotent C2C12 cells[376] and cancer cells,[377] whereas this was not present with m6A.[377] Whether the cells differentiate upon feeding or incorporate more of the 2'-deoxygenated m6dA nucleoside due to differentiation remains elusive.

Therefore, we wanted to analyze, whether m6dA could potentially be formed during a differentiation process. To differentiate cells, transcription is regulated differentially and many signaling cascades are altered. Decay of RNA containing m6A was reported during maternal-to-zygotic transition in zebrafish.[375] Additional processes involving degradation of m6A-containing RNA might be present

106

during differentiation of mESCs. These events could provide an endogenous source for free m^6A, which might subsequently lead to m^6dA in DNA. It was reported previously, that administration of *all-trans* retinoic acid (ATRA, see **Figure 15**) can induce differentiation of mESCs into neural cells.[378]
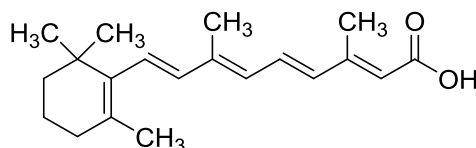
Figure 15: Structure of *all-trans* retinoic acid (ATRA).

Naïve mESCs express hallmark pluripotency factors like *octamer-binding protein 4* (OCT-4),[379-380] also known as *Pou domain, class 5, transcription factor 1* (POU5F1), which is a homeodomain transcription factor of the Pou family and is critically involved in the self-renewal of undifferentiated embryonic stem cells.[381] We decided to administer ATRA to OCT-4 reporter *knockin* mESCs expressing yellow fluorescent protein (YFP) from one allele of the *Oct-4* gene. The cells are from now on referred to as Oly2-1. Excitation of YFP (excitation maximum at 512 nm) with an excitation filter of 500/24 nm and monitoring the fluorescence of YFP (emission maximum at 527 nm) with an emission filter of 524/27 nm, allowed the determination of the expression of OCT-4 and hence the differentiation state of the stem cells.

In a first experiment, we tested various seeding densities and concentrations of ATRA to find suitable conditions for our study. The basal medium for the experiment was serum/L. A 2i/L culture served as additional control on how undifferentiated, naïve cells look. To monitor changes of the developmental potential, we analyzed the fluorescence at 48 h and 72 h after administration of the drug.
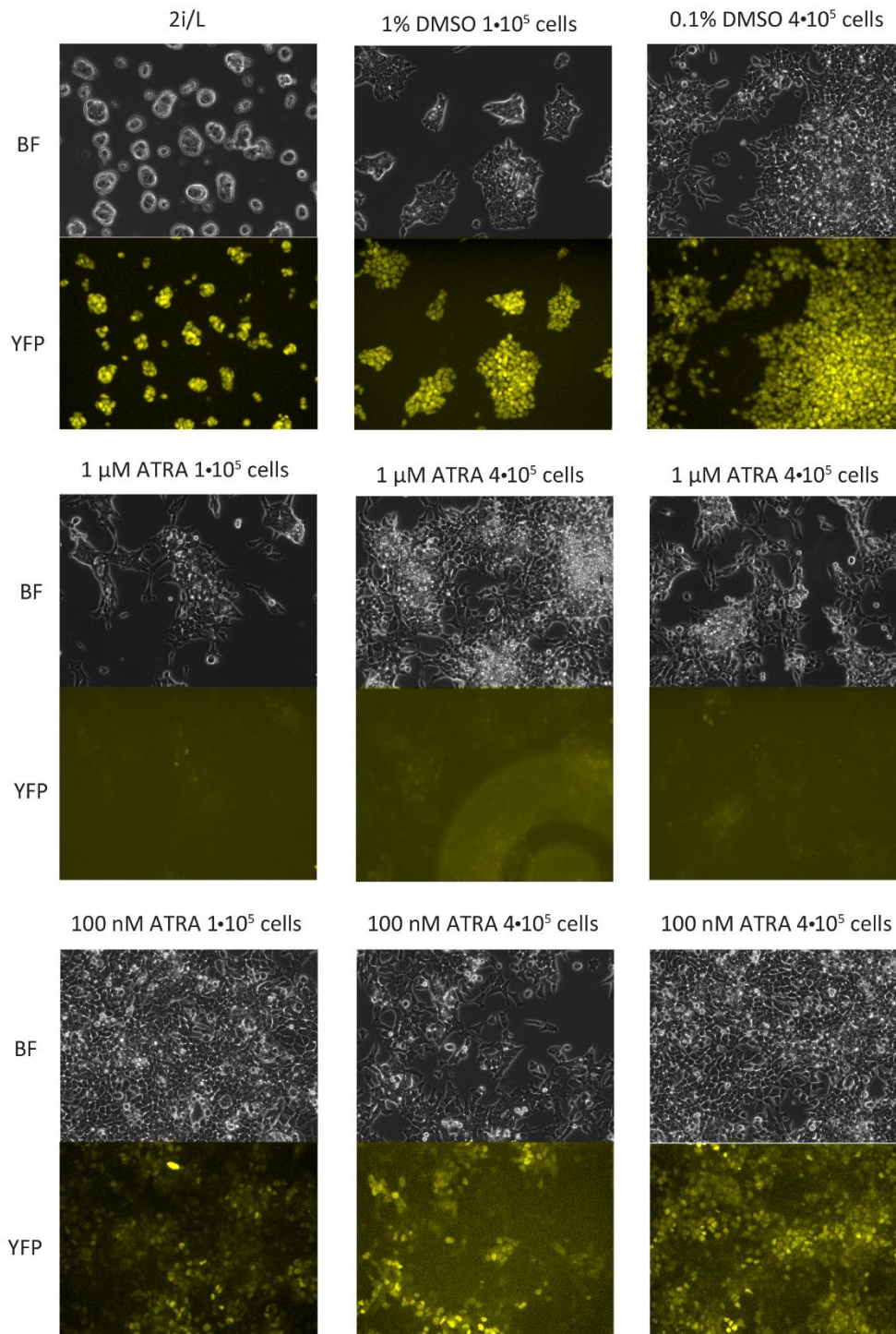
Figure 16: Induction of differentiation of Oly2-1 cells with *all-trans* retinoic acid (ATRA). The cells used are OCT-4 reporter cells expressing YFP. The figure shows the morphology (BF: bright field) and fluorescence (YFP) of the cells with different seeding densities and different concentrations of ATRA after 48 h.

After 48 h (see **Figure 16**), the cells already show very little fluorescence at 527 nm for a concentration of 1 µM ATRA. This indicates substantial loss of pluripotency. The uninduced 2i/L and DMSO controls exhibit strong fluorescence and hence expression of YFP-tagged OCT-4, and the cells treated with 100 nM ATRA show a weakened signal and therefore decreased pluripotency.
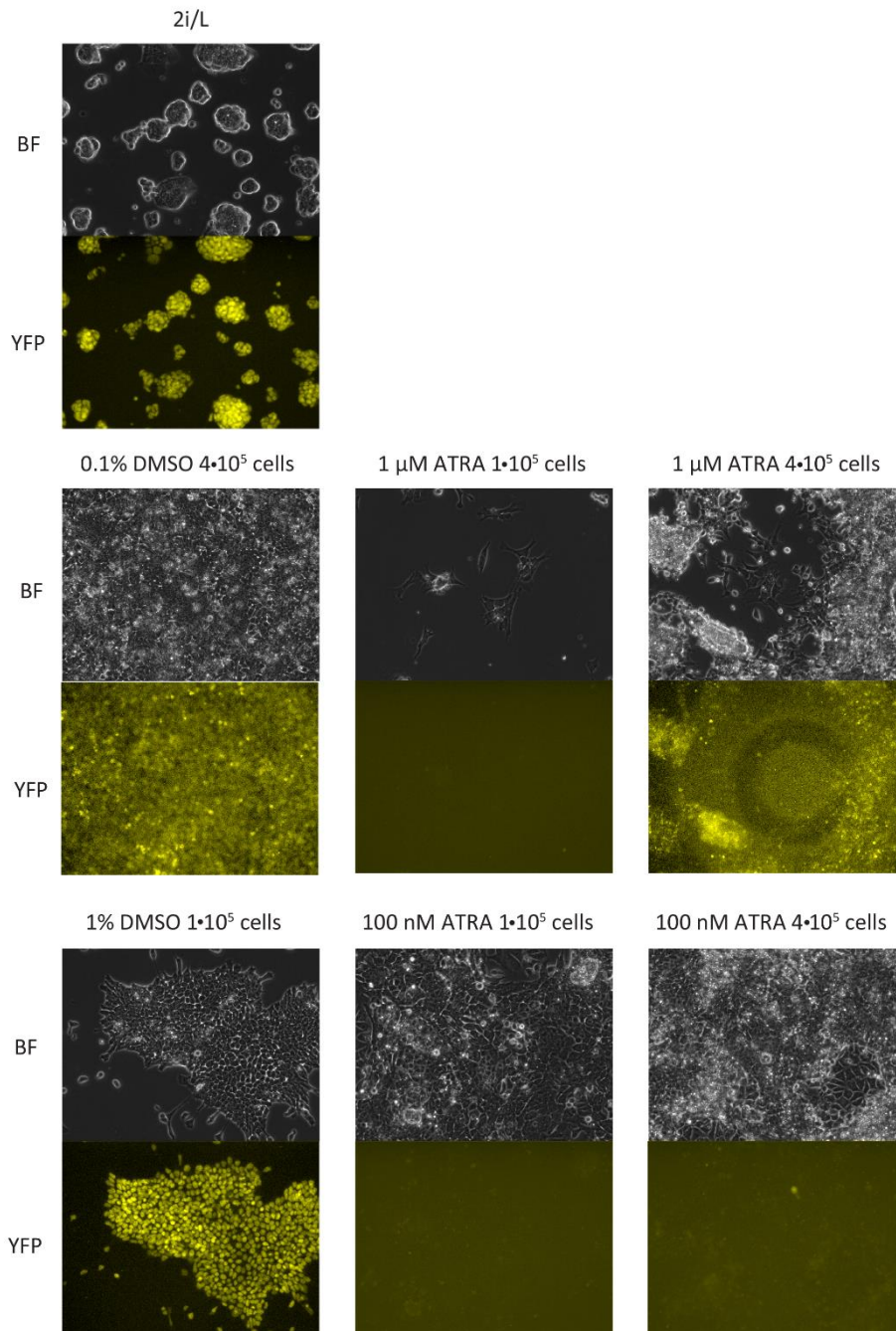
Figure 17: Induction of differentiation of Oly2-1 cells with *all-trans* retinoic acid (ATRA). The cells used are OCT-4 reporter cells expressing YFP. The figure shows the morphology (BF: bright field) and fluorescence (YFP) of the cells with different seeding densities and different concentrations of ATRA after 72 h.

After 72 h (see **Figure 17**), we observed significant amounts of cell death at 1 µM ATRA and concluded that this concentration is too high for our investigation. Administration of 100 nM ATRA with a seeding density of $4 \cdot 10^5$ cells per 6-well, however, produced healthy, differentiated cells with significant loss of pluripotency. We therefore decided to apply these conditions in our subsequent experiments.

Figure 18: Induction of differentiation of Oly2-1 cells with 100 nM ATRA after 24 h. The figure shows the morphology (BF: bright field) and fluorescence (YFP) of the cells in different p100 plates (for harvests at 48 h and 60 h) or p60 plates (for harvest at 72 h).

The results of our experiments performed with the optimized seeding density and administered ATRA concentration (see **Figure 18**) show that the cells exhibit a strong signal for YFP after 24 h. This suggests expression of the pluripotency marker OCT-4 and we conclude that the cells are as expected still pluripotent in all administered conditions at this time point.
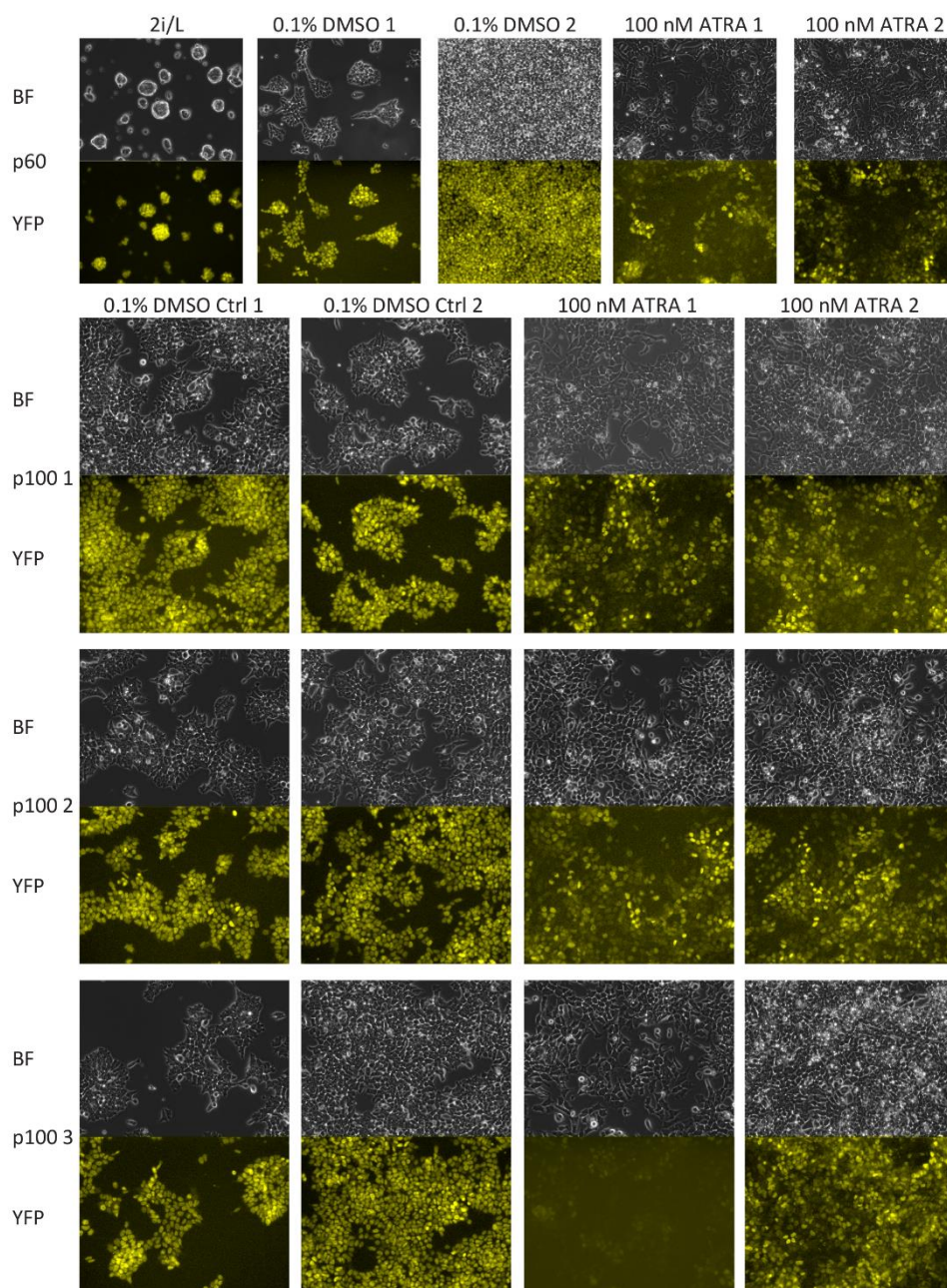
Figure 19: Induction of differentiation of Oly2-1 cells with 100 nM ATRA after 48 h. The figure shows the morphology (BF: bright field) and fluorescence (YFP) of the cells in different p100 plates (for harvests at 48 h and 60 h) or p60 plates (for harvests at 72 h).

After 48 h at the first harvesting time point (see **Figure 19**), the cells treated with 100 nM ATRA already show a decreased signal for YFP, which indicates their differentiation. For the treated and DMSO control, two p100 plates each were lyzed and the lysate stored at -80 °C for subsequent gDNA and RNA isolation and UHPLC-MS/MS analysis.
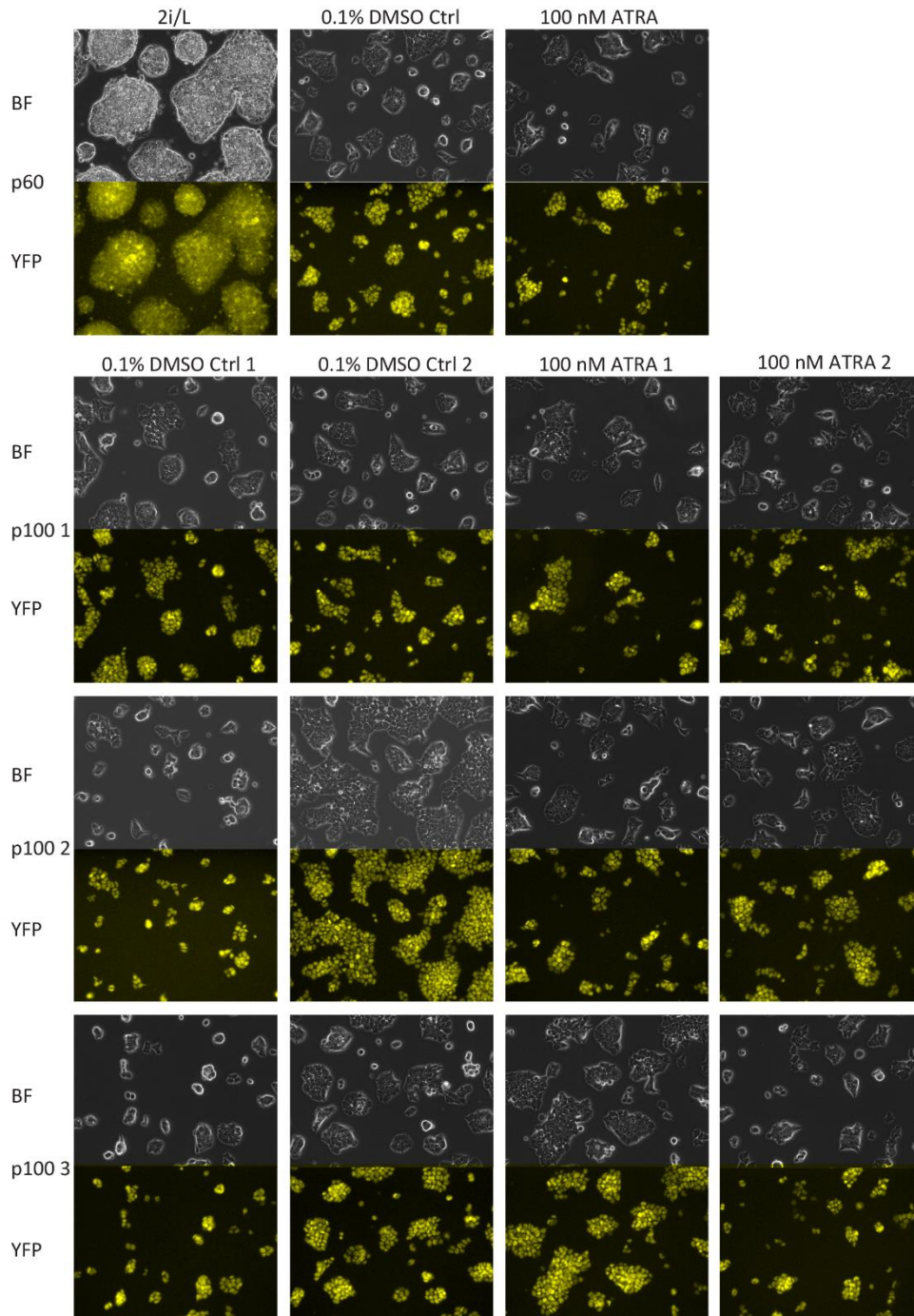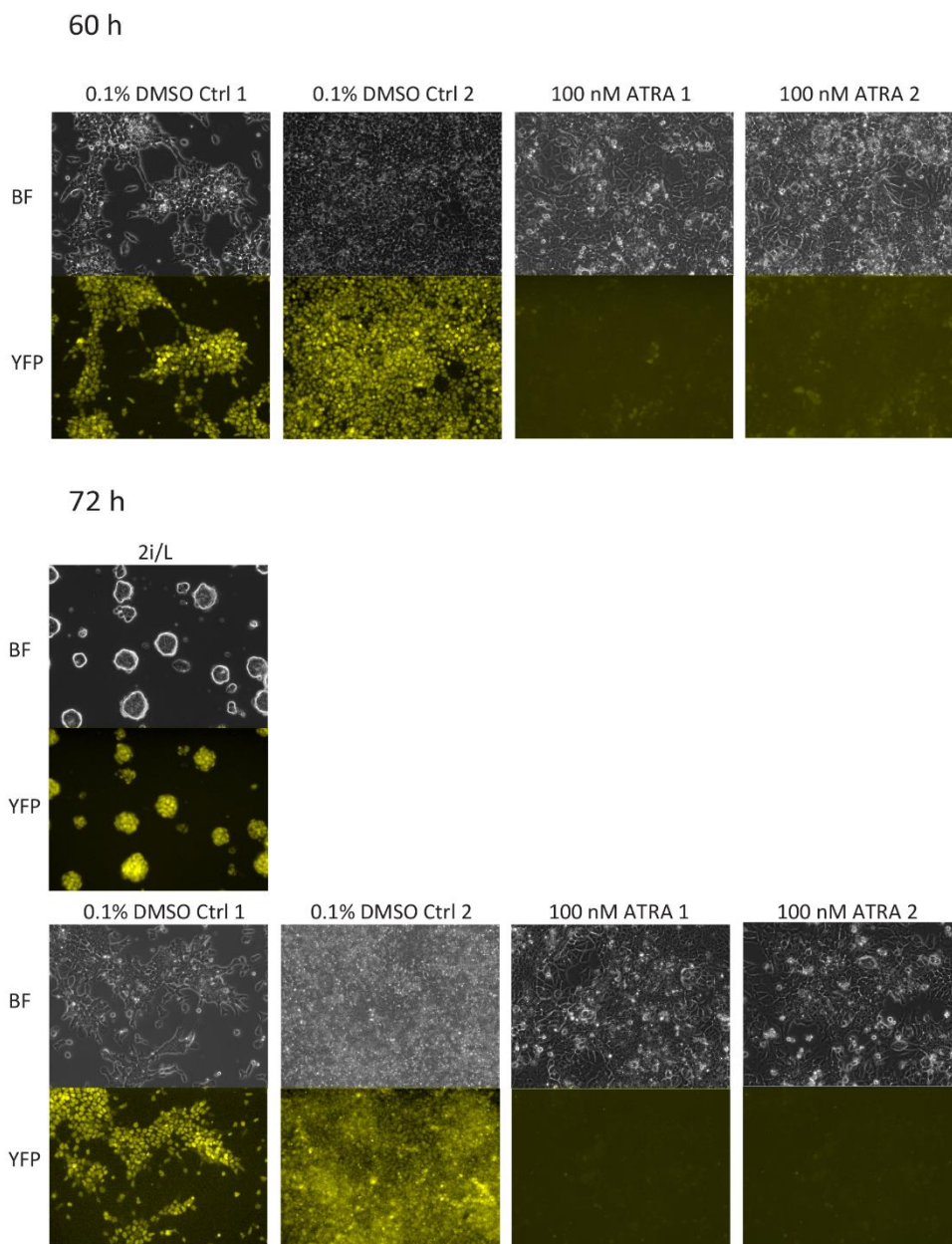
Figure 20: Induction of differentiation of Oly2-1 cells with 100 nM ATRA after 60 h and 72 h. The figure shows the morphology (BF: bright field) and fluorescence (YFP) of the cells in a p100 (60 h) or a p60 (72 h) plate at the harvesting time points.

After 60 h at the second harvesting time point (see **Figure 20**), the cells treated with 100 nM ATRA show a strongly decreased signal for YFP, which indicates progressive loss of pluripotency and therefore differentiation. For the treated and DMSO control, one p100 plate each was lyzed and the lysate stored at -80 °C for subsequent gDNA and RNA isolation and UHPLC-MS/MS analysis. At the last time point after 72 h, almost no fluorescence is detected for the ATRA-treated cells indicating near-complete differentiation. With bright field microscopy, the morphology of the cells is also rather flat and spread-out. Although the DMSO control has a similar morphology and is also cultured in the absence of 2i which leads to maturation of the cells, evaluation of the YFP signal clearly shows expression of OCT-4 and therefore pluripotency.

Subsequent isolation of the gDNA and RNA enabled investigation of the occurrence of m⁶dA and potential effects on the RNA.
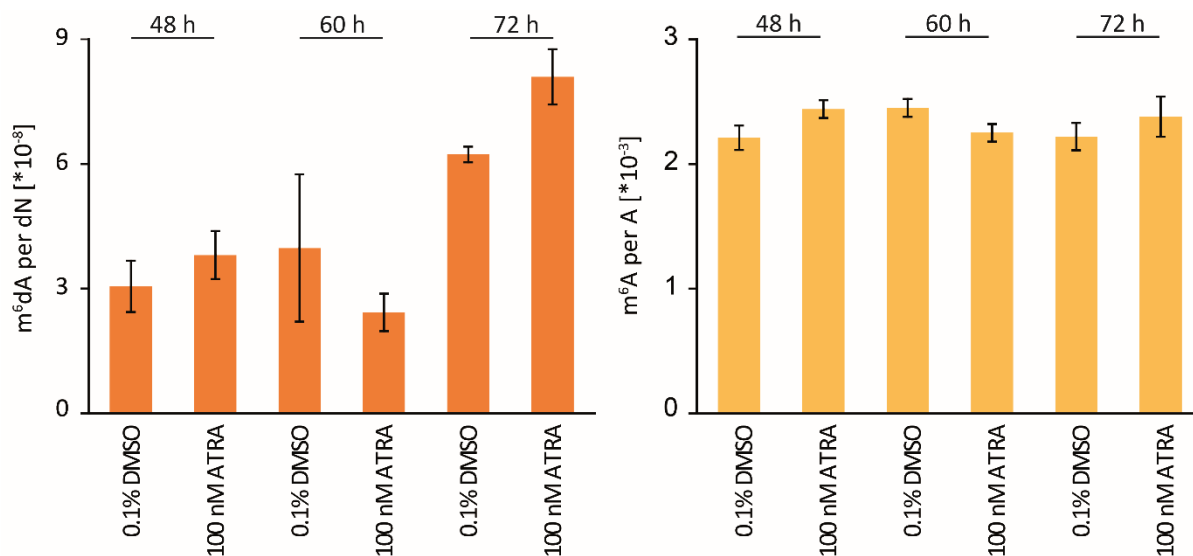
Figure 21: Levels for m⁶dA in gDNA and m⁶A in RNA after induction of differentiation with ATRA at different time points. As a control, the levels of m⁶dA in untreated cells with only administration of DMSO were analyzed. The results for gDNA are depicted on the left side and those for the RNA are on the right.

Analysis of the gDNA (see **Figure 21**, left) reveals very small amounts of m⁶dA after 48 h and 60 h, but these do not differ between the control and the induced cells and might be artefacts. After 72 h however both set-ups show increased levels, and for the induced cells this increase also seems to be significantly different from the control. Therefore, it is possible that differentiation upon ATRA treatment and corresponding events in the cells are responsible for the generation of m⁶dA in stem cells. The m⁶A content of the RNA (see **Figure 21**, right) shows slight fluctuations with differences of about 2.5% between the untreated control and the induced cells, but at the 72 h time point they are similar. If m⁶dA originates from m⁶A-containing RNA turnover, however, a lag time would be expected. Therefore, the slight decrease of m⁶A at 60 h after start of the treatment is a potential indicator of liberated free m⁶A nucleoside and therefore potential source for m⁶dA in the gDNA.

The experiment was performed only once. Because of this reason, a biological effect cannot be fully evaluated and needs further investigation.

3.2.1.4 Treatment of wt mESCs with Trichostatin A

As a last experiment, we wanted to investigate the influence of histone deacetylase (HDAC) inhibitors on m⁶dA formation. HDAC enzymes remove acetyl groups from lysines on histone tails. The subsequently positively charged amino group enables tighter binding of those proteins to the negatively charged DNA backbone. As a result, the DNA gets compacted more tightly into heterochromatin and the transcription decreases. Inhibition of the enzymes leads to less dense

packaging of the DNA, so-called euchromatin, and therefore increased transcription. We hypothesized that less densely packed DNA might be more accessible to m⁶dA writers and wanted to study this condition by applying the HDAC inhibitor Trichostatin A (TSA, see **Figure 22**). This inhibitor[382] is selective for Zn^{2+} dependent Class I, II and IV histone deacetylases, but not NAD^+ dependent HDACs.[383]
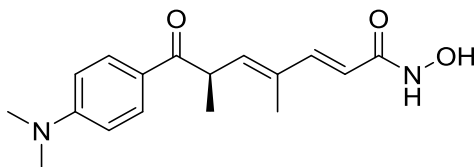
Figure 22: Structure of the HDAC inhibitor Trichostatin A (TSA).

Furthermore, it was previously reported,[384] that HDAC1 and 2 inhibition leads to degradation of poly(A)-RNA. If this includes many RNA containing m⁶A, a transformation into phosphorylated m⁶dA and subsequent incorporation into DNA would be possible. Additionally, recent literature reports that TSA positively affects the development of neuronal[385] and embryonic cells *in vitro*.[386] Investigation of the effect of TSA on mESCs might result in detection of m⁶dA stemming from development-related degradation of m⁶A-containing RNA.

In detail, we administered TSA (10 nM and 500 nM, respectively) to wt Kindlin3⁺/⁺ mESCs.[387] After lysis of the cells at different time points after treatment, subsequent gDNA isolation and enzymatic gDNA digest, we analyzed the gDNA for m⁶dA.

Unfortunately, for none of the conditions applied m⁶dA was found. This could have three reasons: a) chromatin density and therefore transcription-coupled processes do not have an effect on formation of m⁶dA in gDNA, b) the treatment with TSA has to be performed as a shorter and maybe more concentrated pulse for effects to take place or c) the inhibitor was not functional and the experiment should be repeated with parallel analysis of the histone acetylation and the m⁶A levels in the RNA. If m⁶dA occurs as a lesion due to processes during transcription, it might furthermore be repaired rather quickly and we might not have been able to capture the effects.

3.2.2 Investigation of active demethylation of m^{5}dC *via* deamination

Active demethylation is expected to not only happen *via* an oxidative pathway, but might potentially take place through enzymatic deamination of m^{5}dC to dT and further processing of this nucleotide by a repair mechanism.

Previous research in this group by *Jessica Steinbacher, Olesea Kosmatchev, Angie Kirchner* and *Dr. Fabio Spada* has indeed revealed a formation of dT through deamination of m^{5}dC. The underlying experimental set-up for the investigation included isotope tracing studies of "heavy" [^{13}C,D$_3$]-methionine on mESCs and analysis *via* UHPLC-MS/MS. The isotopically labeled methionine as a fundamental precursor for the DNMT cofactor SAM can establish [^{13}C,D$_3$]-methyl groups on dC (see **Figure 23**). TET-mediated oxidation of this subsequently labeled [^{13}C,D$_3$]-m^{5}dC leads to the formation of [^{13}C,D$_2$]-hmdC and [^{13}C,D]-fdC. A putative deamination reaction of [^{13}C,D$_3$]-m^{5}dC is the only origin for [^{13}C,D$_3$]-dT. Endogenous dT in the DNA is generated through methylation of dUMP by TS with 5,10-methylenetetrahydrofolate (5,10-CH$_2$-THF) as the methyl donor and further phosphorylation to dTTP. The dT biosynthesis is therefore SAM-independent and will not introduce a [^{13}C,D$_3$]-label in dT. This has previously been proven by investigation of DNMT triple *knockout* mESCs, in which upon administration of [^{13}C,D$_3$]-methionine [^{13}C,D$_3$]-dT was not detected.
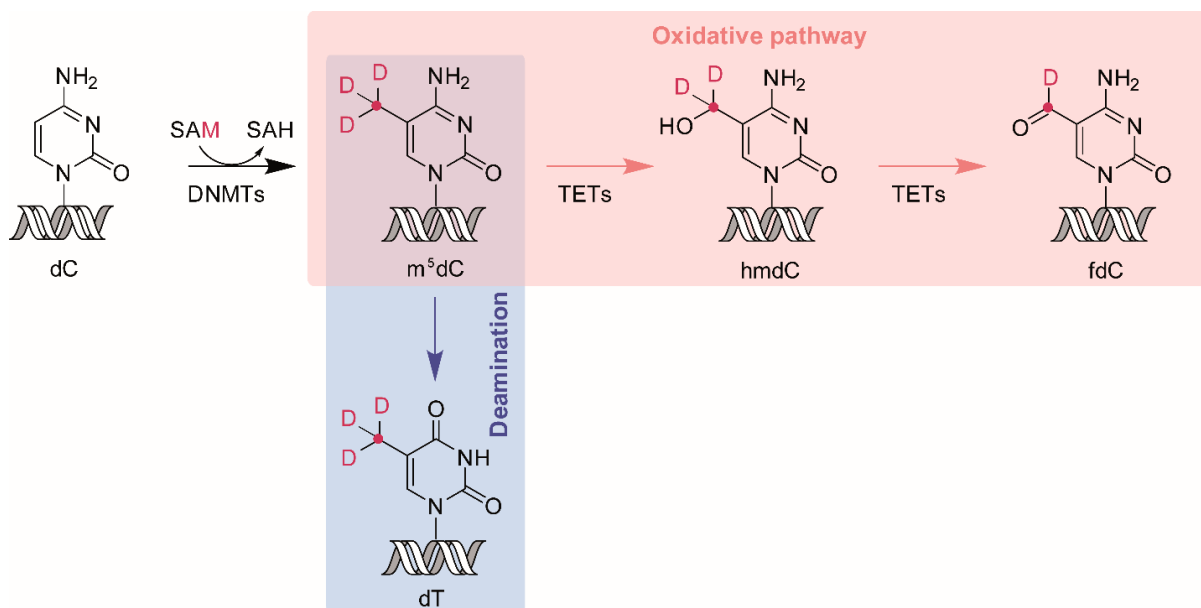
Figure 23: Overview over the labeling of DNA nucleotides upon administration of [^{13}C,D$_3$]-methionine.

The UHPLC-MS/MS data from the previous research were evaluated *via* relative quantification. Absolute quantification was not possible, because [^{13}C,D$_3$]-m^{5}dC differs by only 1 amu from [D$_3$]-m^{5}dC, which was the only available isotope standard at that time. This difference in the mass-to-charge ratio is not sufficient for mass spectrometric distinction on a Triple Quadrupole mass spectrometer. Co-culture of the same experiment supplemented with unlabeled methionine allowed then for a

projection of the percentage of the labeled or unlabeled species onto the global levels of the modification. This projection was nevertheless not accurate due to small isotope effects.

Therefore, new standards with more isotopes $[^{15}N_2,D_4]$-m^{5}dC, $[^{13}C_5,^{15}N_2]$-hmdC, $[^{13}C_5,^{15}N_2]$-fdC and $[^{13}C_5,^{15}N_2]$-dT were synthesized by *Charlotte Ebert* and *René Rahimoff* to enable exact quantification. For exact quantification of $[^{15}N_5,^{13}C_{10}]$-dG, which was co-fed as a marker for replication, we bought commercially available $[^{15}N_5]$-dG. With these standards in hand, *Fabio Spada* and I could continue the investigation *via* absolute quantification.
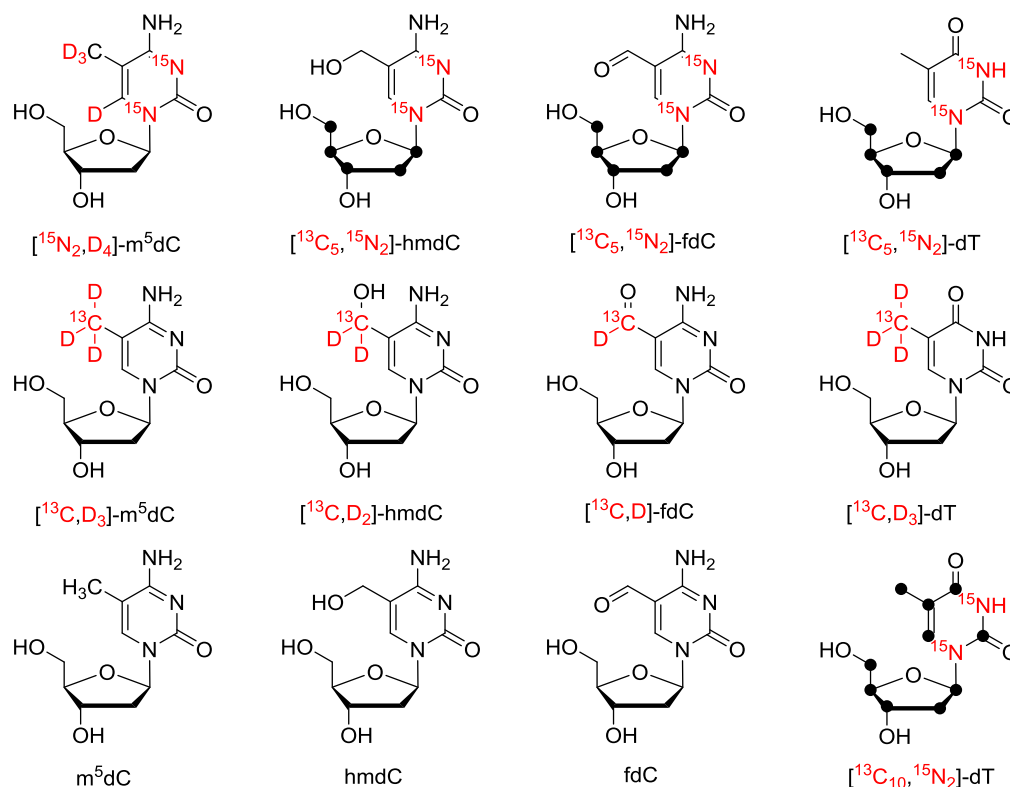
Figure 24: Overview over modified nucleosides and their heavy labeled analogues derived from administration of $[^{13}C,D_3]$-methionine on mESCs, as well as added isotopologues that enable absolute quantitative analysis of the data. The black dots indicate ^{13}C atoms.

To this end, we first had to verify the mass transitions for the standards and developed the method for the UHPLC-QQQ-MS. This method was now able to monitor:

I)   all unlabeled, naturally occurring modifications (see **Figure 24** lower row)

II)   the labeled modifications that result from the addition of $[^{13}C,D_3]$ onto dC with its deamination and oxidation products (see **Figure 24** middle row)

III)   the amount of co-fed $[^{13}C_{10},^{15}N_2]$-dT (see **Figure 24** lower row) and $[^{13}C_{10},^{15}N_5]$-dG (not depicted)

IV)   as well as all isotope standards including $[^{15}N_5]$-dG (not depicted) simultaneously (see **Figure 24** upper row).

Furthermore, our experiments were slightly modified from the culturing conditions used prior. *In vitro* mESCs are derived from pre-implantation blastocysts (see section **1.1.2.2.2**). These cells are therefore naïve pluripotent and require addition of specific factors to the medium. Feeder independent culturing of mESCs can utilize serum with two inhibitors (2i),[380] capturing the homogeneous[388] and very naïve pluripotent state. CHIR 99021, one of the two inhibitors, acts on *glycogen synthase kinase 3* (GSK3) α and β and promotes *wingless/Int1* (WNT)/β-catenin signaling. This inhibitor is used for several culturing conditions of naïve as well as primed pluripotent mESCs. The second inhibitor, PD 0325901, inhibits *mitogen-activated protein kinases* (MAPK) *kinases* (MEK1/2), which phosphorylate MAPK. Additional supplementation with *leukemia inhibitory factor* (LIF)[389] is possible, forming 2i/L conditions. This protein is an interleukin 6 class cytokine, which affects cell growth by inhibiting differentiation, because it binds the LIF receptor and thereby activates *Janus kinase/signal transducer and activator of transcription* (JAK/STAT) and MAPK signaling cascades.[390] Inhibition of MEK blocks transduction of most of the fibroblast growth factor (FGF) signaling activity. MEK inhibition[391-392] furthermore leads to DNA hypomethylation[393-395] due to low expression of DNMT1,[396] 3a and 3b,[397-398] as well as associated cofactors DNMT3l and UHRF1, and higher oxidation of m^5dC to hmdC by TET activation. Impaired maintenance of methylation includes methylation at ICRs[399] that is not reestablished during differentiation[400-402] and results in bi-allelic expression of imprinted genes.[403] Studies of imprinted gene expression *in vivo* should therefore avoid such culturing conditions.

Instead of PD 0325901, the proto-oncogene tyrosine-protein kinases SRC inhibitor CGP77675[404] can be used to culture naïve mESCs. In these so-called 'alternative 2i' (a2i) conditions, first described by *Shimizu et al.*,[405] MAPK signaling is suppressed in a more moderate fashion[405] than through inhibition of MEK1/2. Cells in these conditions exhibit reduced genome-wide DNA methylation due to expression of DNMT3 enzymes and gamete-derived methylation is retained.[392, 399] These culturing conditions are therefore more suitable to study events that involve DNA methylation, especially in regards to differential methylation.

Serum-based medium can be supplemented only with LIF (serum/L). ES cells in serum/L display heterogeneity and fluctuate between the naïve inner cell mass (ICM)-like state and the primed epiblast-like state,[406] therefore representing a wide variety of states of the pluripotency spectrum.

More homogeneous *priming* of mESCs can be initiated by supplementation with FGF or Activin A, or molecules activating the corresponding signaling pathways. Addition of IWR1 to serum-based medium containing CHIR 99021, hereafter called C/R *priming*, is one commonly used *priming* condition. The IWR1 molecule is a potent inhibitor for the WNT signaling pathway leading to a loss of multipotency. The molecule induces β-catenin degradation by abrogation of the Axin protein turnover which is part of the degradation complex.[407] The molecular cause for the effect is inhibition of tankyrases (TNKS-1

and -2), which transfer poly-(ADP ribose) (PAR) chains to targeted proteins[408-409] and which are involved in the nuclear transport of β-catenin.

In our experiments, naïve mESCs were mostly maintained in a2i/L conditions, unless stated otherwise. Additionally, mESCs were primed in serum/L conditions as well as C/R *priming* conditions.

3.2.2.1 Time course of labeling wt mESCs with [^{13}C,D$_3$]-methionine

In a first experiment, we wanted to investigate the changes in labeling of m^{5}dC and dT in naïve and primed conditions in a labeling time course and a pulse and chase experiment on Kindlin$^{+/+}$ (K3$^{+/+}$) wt mESCs.

For the labeling time course, we cultured an a2i/L or C/R *priming* culture for 3 d. Subsequently, the medium was supplemented with [^{13}C,D$_3$]-methionine and with [^{13}C$_{10}$,^{15}N$_2$]-dT for up to additional 48 h with cell harvests at every 12 h.

The complementary experiment, in which we were interested in monitoring the decrease in labeling of the modifications (pulse and chase), was performed by labeling the cells with [^{13}C,D$_3$]-methionine for 2 d in naïve conditions and additional 3 d in naïve or C/R conditions, respectively. Additionally we added [^{13}C$_{10}$,^{15}N$_2$]-dT as a marker for replication for the last 24 h of the labeling phase. We then changed the medium to normal methionine and collected the cells at 0 h, 12 h, 24 h, 36 h and 48 h.
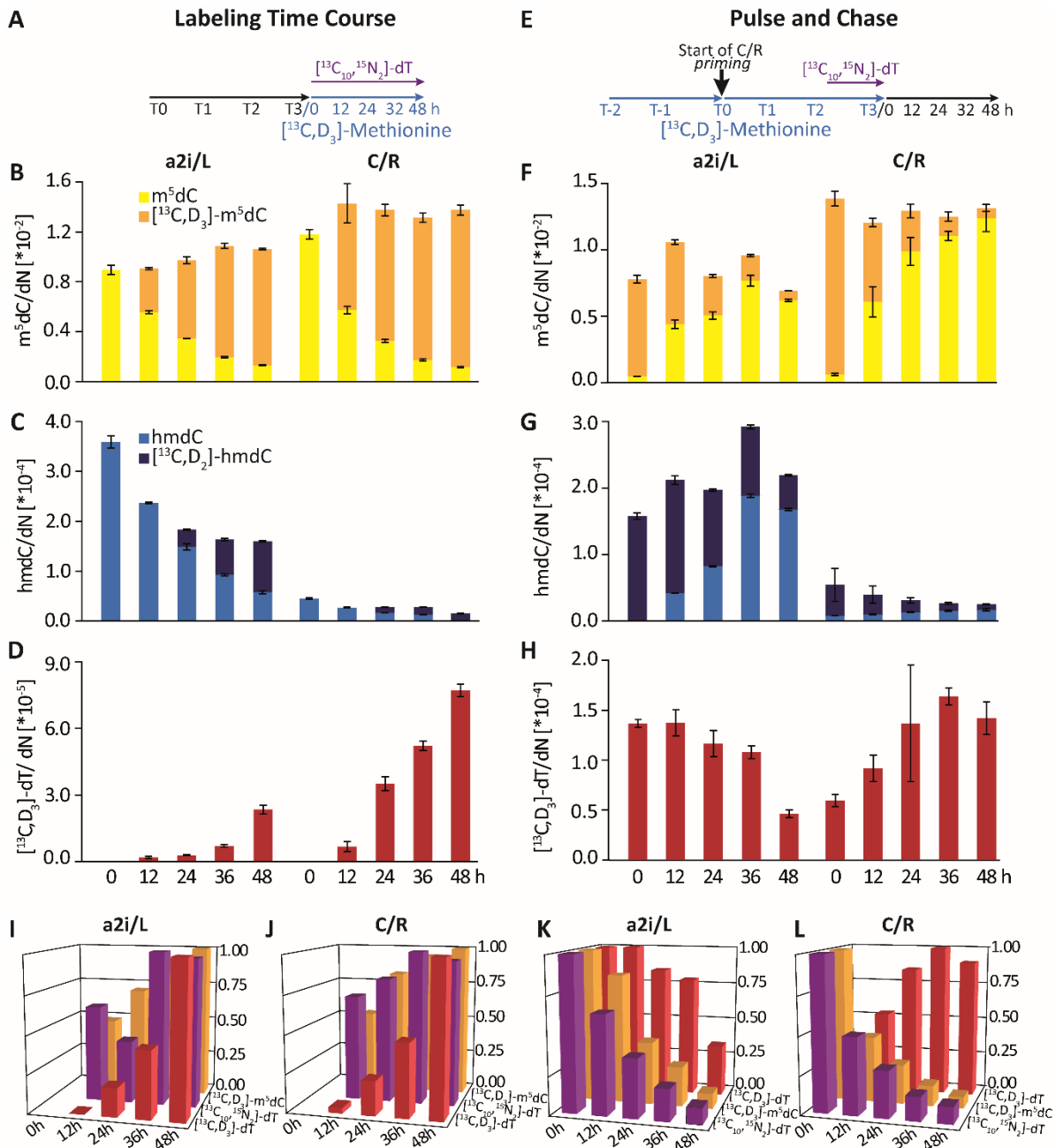
Figure 25: Level changes of $[^{13}C,D_3]$-m^{5}dC, $[^{13}C,D_2]$-hmdC, $[^{13}C,D_3]$-dT and $[^{13}C_{10},^{15}N_2]$-dT in K3$^{+/+}$ mESCs in naïve (a2i/L) and primed conditions (C/R). A and E show the labeling strategy (d: day). Depicted in B, C and D are absolute results of a labeling time course experiment, in which the increase of labeled material is monitored every 12 h after addition of $[^{13}C,D_3]$-methionine and $[^{13}C_{10},^{15}N_2]$-dT. Depicted in F, G and H are the absolute results of the corresponding pulse and chase experiment, where the labeling was performed to almost 100% and then the decrease of labeled material was observed after medium change to unlabeled methionine. I and J show the corresponding relative changes in labeling for the labeling time course experiment, whereas K and L visualize the relative changes in labeling for the pulse and chase. The right half of F, G and H, as well as L show results for two combined biological replicates, whereas all other parts of the figure only represent data from one biological replicate measured in technical triplicates.

In the labeling time course experiment of naïve (a2i/L) and primed (C/R) mESCs (see **Figure 25B and F**), the gDNA shows increasing levels of $[^{13}C,D_3]$-m^{5}dC at 12 h and subsequent time points after the medium change. The $[^{13}C,D_2]$-hmdC nucleoside, as a product of m^{5}dC oxidation, is not detected until

after 24 h of labeling with [^{13}C,D$_3$]-methionine (see **Figure 25C and G**), but also increases over the following time points. In C/R *priming* conditions, hmdC levels are however generally low and division of an already low global amount of hmdC into two species renders both of them difficult to detect above background level. Like we expected, we also find the deaminated derivative of [^{13}C,D$_3$]-m^{5}dC, [^{13}C,D$_3$]-dT, in naïve (up to 3•10^{-5} per dN) and *priming* (up to 9•10^{-5} per dN) conditions. The modification accumulates nicely in a time-dependent manner. Interestingly, whereas hmdC is rather low during the *priming* period using CHIR 99021 and IWR1, [^{13}C,D$_3$]-dT is three times more abundant than in naïve conditions. This result indicates that during *priming* of mESCs the process that generates dT from m^{5}dC takes place more frequently than in naïve conditions.

In the pulse and chase experiment, we follow the decrease of the labeled nucleosides over time. Indeed, [^{13}C,D$_3$]-m^{5}dC and [^{13}C,D$_2$]-hmdC show a reduction over time. However, since we performed the a2i/LIF experiment only performed once, it seems like the global m^{5}dC and hmdC levels (see **Figure 25F** and **G** left side) behave rather unexpected and fluctuate a lot, although the labeled fraction shows a steady decrease over time (see **Figure 25K**). The reduction in [^{13}C,D$_3$]-m^{5}dC seems similar to that of the replication marker [^{13}C$_{10}$,^{15}N$_2$]-dT, suggesting dilution through replication. In the C/R conditions (**Figure 25L**), the modification decreases faster than the replication marker, which indicates active turnover. For [^{13}C,D$_3$]-dT, we also see a clear decrease of the modification in naïve conditions. Surprising are the results for the priming conditions. We performed the experiment twice with different seeding densities, then normalized and combined the data. It seems like both replicates of the labeling time course in C/R conditions show a decrease in labeled [^{13}C,D$_3$]-m^{5}dC, [^{13}C,D$_2$]-hmdC and [^{13}C$_{10}$,^{15}N$_2$]-dT, but not in [^{13}C,D$_3$]-dT. We conclude that deamination of m^{5}dC to dT is performed at a high rate in this cell state, overcoming dilution through replication and finally leading to an overall increase of the modification during the first 36 h after the medium change. A repair process involving deamination of m^{5}dC to dT is therefore likely to occur throughout those developmental stages of mESCs. The opposite results for the oxidized nucleosides indicate that the two pathways for active demethylation of m^{5}dC might complement each other in different embryonic states.

3.2.2.2 Investigation of TET TKO mESCs

Next, we wanted to evaluate whether the strong decline of [^{13}C,D$_3$]-m^{5}dC can be explained through oxidation events, and whether there are changes in the formation of [^{13}C,D$_3$]-dT in the absence of m^{5}dC oxidation. Therefore we decided to investigate the labeling with [^{13}C,D$_3$]-methionine in TET triple *knockout* (TET TKO) mESCs, in which no TET enzymes are expressed and therefore generation of hmdC, fdC or cadC is not possible. To get a more detailed picture, we performed the labeling procedure over a 3 d C/R *priming* period and subsequently conducted a chase experiment with cell harvests at every 12 h for a total of 48 h. [^{13}C$_{10}$,^{15}N$_2$]-dT was again used as a marker for replication.
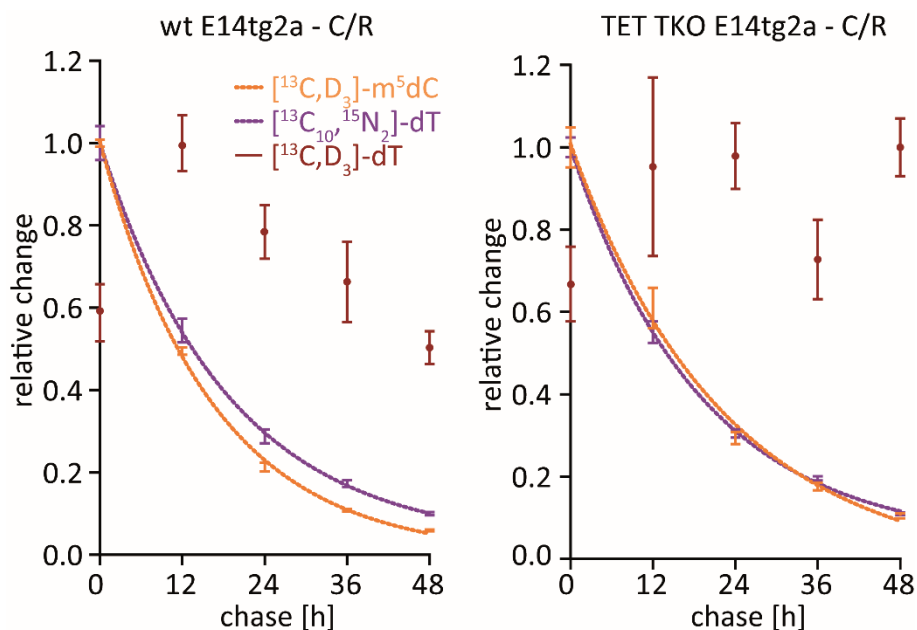
Figure 26: Relative changes of nucleoside labeling after administration of [^{13}C,D$_3$]-methionine to E14tg2a mESCs (left) and TET TKO mESCs (right) in 5 d of C/R *priming* conditions and subsequent medium change to unlabeled methionine. Depicted are data for several time points (0 h, 12 h, 24 h, 36 h and 48 h) after the medium change in one biological and three technical replicates.

Indeed, the results (see **Figure 26**) of the chase experiment with E14tg2a cells are very similar to those obtained from wt K3$^{+/+}$ cells (see **Figure 25J**). In this experiment, labeled m^{5}dC turns over quicker than [^{13}C$_{10}$,^{15}N$_2$]-dT, which indicates active turnover rather than dilution by replication. Furthermore, as seen in wt K3$^{+/+}$ cells, [^{13}C,D$_3$]-dT also counteracts dilution through replication and increases over the first time points of the chase. The highest level of this modification is, however, reached already at the 12 h time point and not – as with wt K3$^{+/+}$ cells – at 36 h. In contrast to this, the TET TKO cells do not show a stronger decrease of labeled m^{5}dC than [^{13}C$_{10}$,^{15}N$_2$]-dT. This result suggests that the major part of m^{5}dC turnover stems indeed from oxidation of the nucleotide to hmdC, fdC and cadC. In the TET TKO cells, we find a more or less constant level of [^{13}C,D$_3$]-dT at all time points. This nucleotide is therefore actively generated in the course of the experiment. Since the [^{13}C,D$_3$]-m^{5}dC levels are three orders of magnitude higher than those of [^{13}C,D$_3$]-dT, a lack of a difference between the relative changes of [^{13}C,D$_3$]-m^{5}dC and [^{13}C$_{10}$,^{15}N$_2$]-dT is expected. Our results might indicate that in absence of m^{5}dC oxidation a deamination reaction of m^{5}dC to dT takes place more frequently and functions as a back-up mechanism for m^{5}dC turnover.

3.2.2.3 Determination of the effect of soluble deaminases on formation of [^{13}C,D$_3$]-dT

In a further experiment, we wanted to investigate possible enzymes that could perform the transition of m^{5}dC to dT. First, we were interested in the deaminases that perform deamination in the soluble nucleoside and nucleotide pool, CDA and DCTD. *Angie Kirchner* previously generated a double *knockout* cell line (from now on termed CD DKO), in which the two enzymes are not expressed. To this

end, she applied the CRISPR/Cas genome editing methodology, which utilizes guide RNA to direct the editing process, but additional undesired mutations are possible. We analyzed three clones #3, #40 and #41 for our investigation to make sure observed effects stem from lack of the enzymes and not from potential off-target events.

Since all three KO clones did not show any [^{13}C,D$_3$]-dT in the gDNA (data not shown), the observation suggests that the nucleoside is mainly generated from excised labeled m^{5}dC in the soluble pool of the cells and *knockout* of the enzymes leads to a complete loss of the nucleoside.

Due to the high abundance of natural dT, which co-elutes with [^{13}C,D$_3$]-dT, a detection of the labeled nucleoside might be impaired by competition of the nucleosides for charge, so-called ion suppression. This circumstance creates a rather high limit of detection (LOD) for [^{13}C,D$_3$]-dT of approximately 10^{-5} per dN. In comparison, the LOD of epigenetically relevant fdC lies in the range of 10^{-7} per dN and enables quantification of this nucleoside at low levels. The accumulation of [^{13}C,D$_3$]-dT at levels of 10^{-7}–10^{-6} per dN in the CD DKO cells might therefore be obscured by the background signal.

3.2.2.3.1 *Administration of [^{13}C$_9$,^{15}N$_3$]-dC and [^{13}C,D$_3$]-methionine on the CD DKO mESCs*

In an attempt to overcome the problems with the LOD, we thought of elevating the exact mass of our target nucleosides by applying [^{13}C$_9$,^{15}N$_3$]-dC to the culturing medium and additionally adding [^{13}C,D$_3$]-methionine. This dC isotopologue should be incorporated into the gDNA just like the unlabeled nucleoside[179, 410] and subjected to all enzymatic processes that take place on the gDNA (methylation, oxidation and deamination).

With 100% labeling of dC and 100% labeling with [^{13}C,D$_3$]-methionine, we would expect only [^{13}C$_9$,^{15}N$_3$]-dC, [^{13}C$_{10}$,D$_3$,^{15}N$_3$]-m^{5}dC and subsequently [^{13}C$_{10}$,D$_3$,^{15}N$_2$]-dT. It is, however, known[410] that labeling of dC will not exceed 30-40%. We must therefore take into consideration that there will still be unlabeled dC, which will lead to formation of [^{13}C,D$_3$]-m^{5}dC and [^{13}C,D$_3$]-dT. Furthermore, labeling with [^{13}C,D$_3$]-methionine is indeed almost 100% efficient, but our cultures are supplemented with serum and it contains residual unlabeled methionine from this biological source. As a result, also completely unlabeled m^{5}dC and dT, as well as [^{13}C$_9$,^{15}N$_3$]-m^{5}dC and [^{13}C$_9$,^{15}N$_2$]-dT are possible nucleosides in the gDNA of the mESCs (see **Figure 27**). Nevertheless, these nucleosides might still not reflect the whole picture of all dT derivatives. Formation of dT opposite of dC as a result of genomic deamination could finally lead to BER by TDG and MBD4, which will release thymine. This thymine could potentially be unlabeled or labeled with [^{13}C,D$_3$], [^{13}C$_4$,^{15}N$_2$] or [^{13}C$_5$,D$_3$,^{15}N$_2$]. Subsequent salvage, (ribosylation and phosphorylation to the triphosphate) could finally lead to reincorporation into the gDNA as [^{13}C$_4$,^{15}N$_2$]-dT or [^{13}C$_5$,D$_3$,^{15}N$_2$]-dT.

All of these possible dC, m^{5}dC and dT isotopologues have to be considered, when establishing an UHPLC-QQQ-MS/MS method to make sure that all potential processes in the cell and their outcome

are monitored. Due to the selection for specific precursor and product ions, other molecules are not recorded and the information from the experiment would be incomplete. As a marker for replication in this experiment, we chose not to apply a dT isotopologue because of the already high number of possible dT isotopologues potentially arising in the course of the experiment, but we supplemented the cells with $[^{13}C_{10},^{15}N_5]$-dG (7.5 μM).

This experiment can only shed light on the existence and action of the soluble deaminases, when we additionally investigate the nucleosides in the soluble pool. The soluble pool of wt mESCs is expected to contain dC and $[^{13}C_9,^{15}N_3]$-dC. Deamination of these nucleosides to the respective dU compounds could subsequently generate dT and $[^{13}C_9,^{15}N_2]$-dT. Since methylation of dU to dT by TS depends on 5,10-CH_2-THF, which is independent of methionine as determined in DNMT TKO mESCs, a modification with $[^{13}C,D_3]$-methionine in the soluble pool can be excluded. On top of the two dT isotopologues directly generated in the soluble pool, DNA repair might also release other dT isotopologues from the gDNA, which could be salvaged in the soluble pool (see **Figure 27**). The CD DKO cells are expected to show significantly lower amounts of $[^{13}C_9,^{15}N_2]$-dT, since soluble deamination is impaired. Traces from spontaneous deamination or as a result of DNA repair are however possible. Comparison of the nucleoside content of the gDNA and the soluble pool might therefore give indication as to the occurrence of genomic deamination.
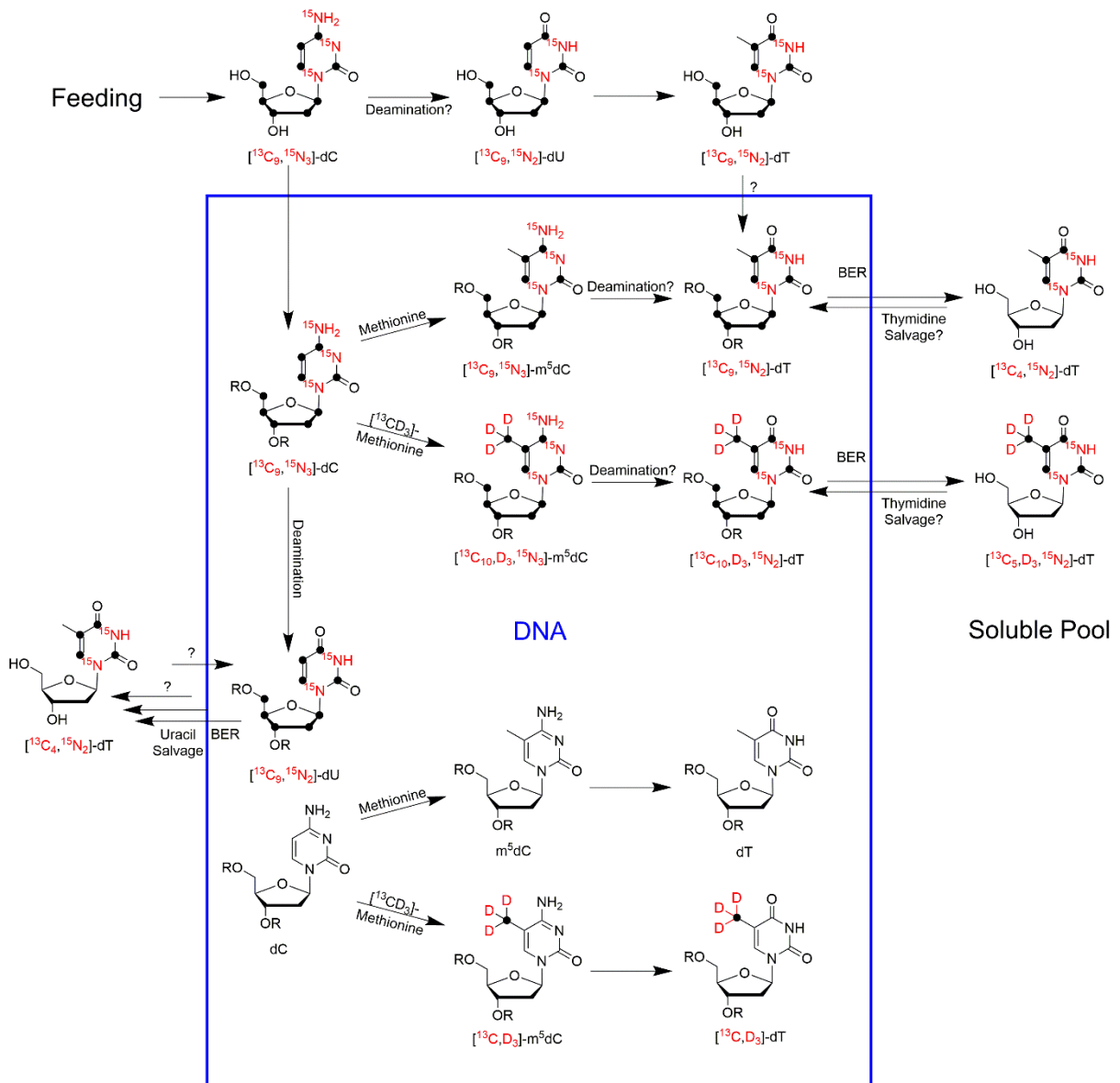
Figure 27: Possible nucleosides in the soluble pool and gDNA upon administration of $[^{13}C_9,^{15}N_3]$-dC on mESCs supplemented with $[^{13}C,D_3]$-methionine. Black dots indicate ^{13}C atoms; R marks the DNA strand.
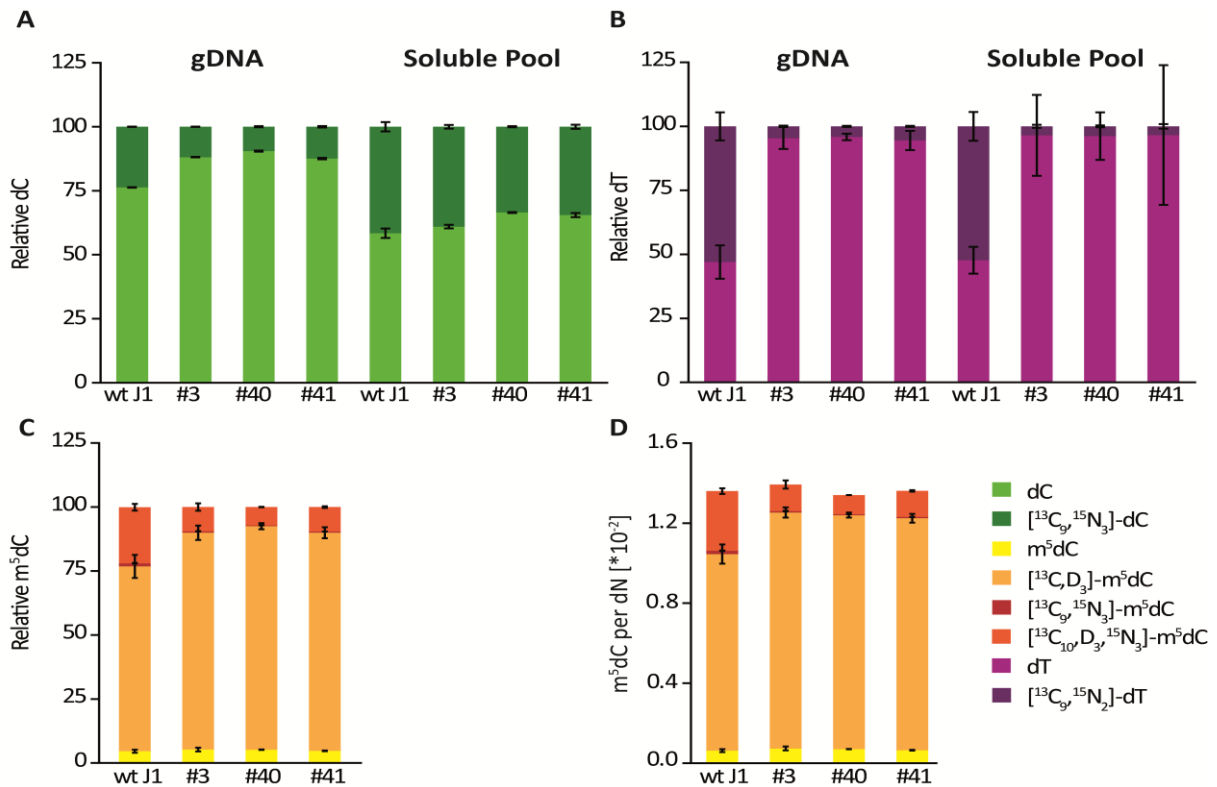
Figure 28: Isotope levels of modifications in the gDNA and soluble pool of wt and CD DKO mESCs upon administration of $[^{13}C_9,^{15}N_3]$-dC and $[^{13}C,D_3]$-methionine after 5 d of *priming* under C/R conditions. A) shows relative distribution of dC isotopologues, B) depicts relative distribution of dT isotopologues, C) illustrates relative distribution of m5dC isotopologues and D) shows absolute levels of m5dC isotopologues per dN. The data represent technical triplicates of one biological sample.

Our results show that the gDNA (see **Figure 28**) exhibited up to 25% $[^{13}C_9,^{15}N_3]$-dC in the wt, and 10-12% $[^{13}C_9,^{15}N_3]$-dC in the CD DKO. The reduced presence of the $[^{13}C_9,^{15}N_3]$-dC nucleoside in the CD DKO probably stems from high intracellular levels of dCMP due to the absence of DCTD. Since phosphorylation of nucleosides is an equilibrium reaction, most of the administered free $[^{13}C_9,^{15}N_3]$-dC does not get phosphorylated and the labeled material is incorporated to smaller extent than in the wt. Furthermore, up to 53% $[^{13}C_9,^{15}N_2]$-dT was found in the gDNA of the wt, and 4-5.5% $[^{13}C_9,^{15}N_2]$-dT in the gDNA of the CD DKO. In the wt gDNA, we additionally found traces of $[^{13}C,D_3]$-dT. The other dT isotopologues were not detected. The m5dC species as the precursors for labeled dT interestingly comprise 72% $[^{13}C,D_3]$-m5dC, up to 22% $[^{13}C_{10},D_3,^{15}N_3]$-m5dC, 5% of unlabeled m5dC and around 1% of $[^{13}C_9,^{15}N_3]$-m5dC in the wt gDNA. In contrast, the CD DKO show 85-88% $[^{13}C,D_3]$-m5dC, 7-10% $[^{13}C_{10},^{15}N_3,D_3]$-m5dC, 5% of unlabeled m5dC and only traces <0.5% of $[^{13}C_9,^{15}N_3]$-m5dC. The overall percentage of nucleosides containing labeled dC both in the results for the dC and m5dC species is comparable. The observation, that next to unlabeled dT, only $[^{13}C_9,^{15}N_2]$-dT, but no $[^{13}C_{10},D_3,^{15}N_2]$-dT is detected, suggests again that the major part of labeled dT is produced by deamination in the soluble nucleoside/-tide pool. Taking into consideration, that upon administration of $[^{13}C,D_3]$-methionine on cultures without further labeling we only detect small amounts of $[^{13}C,D_3]$-dT, the cause for the

absence of $[^{13}C_{10},D_3,^{15}N_2]$-dT might still be ion suppression. An isotopic effect can also be present, since the $[^{13}C,D_3]$-methyl group might be too bulky or alter the electronic properties of the nucleoside so that the action of deaminases is affected. Our finding furthermore hints at generally low repair of m^{5}dC involving deamination.

The soluble pool of the same experiment contained 45% $[^{13}C_9,^{15}N_3]$-dC in the wt, and 40-45% $[^{13}C_9,^{15}N_3]$-dC in the CD DKO. The difference in labeling between wt and CD DKO is not as pronounced as in DNA, because the overall dC pool is at an equilibrium and we analyze the free nucleoside after enzymatic dephosphorylation of all free and phosphorylated dC species present in the cytosol. In comparison, the dT isotopologues comprise 52% $[^{13}C_9,^{15}N_2]$-dT in the wt, and 3-4% $[^{13}C_9,^{15}N_2]$-dT in the CD DKO. Due to the low abundance of repaired m^{5}dC nucleoside, it is not surprising that the nucleoside was not detected in the soluble pool. In contrast to the results for dC, the results for dT are comparable to those of the gDNA. This might suggest an isotopic effect in the generation of the dCTP from $[^{13}C_9,^{15}N_3]$-dC and incorporation of the corresponding dCMP derivative into gDNA by DNA polymerases, since we would expect the isotopologues to be at an equilibrium after 5 d of labeling with $[^{13}C_9,^{15}N_3]$-dC. The fact that the percentage of $[^{13}C_9,^{15}N_2]$-dT is the same in the soluble pool and the gDNA might subsequently mean, that generation of dTTP, its incorporation into gDNA as dTMP or DNA repair is not affected by the isotopes. The latter is logical since repair processes address a great variety of nucleotides, of which some are bulky, and must therefore be performed by rather unspecific enzymes.

Overall, the levels of $[^{13}C_9,^{15}N_2]$-dT in the CD DKO are still rather high and not likely to stem from spontaneous deamination. It is however possible, that the labeled nucleoside is generated after deamination and repair of labeled dC in the DNA and that subsequent methylation of the generated labeled dU(MP) produces the considerable amounts of $[^{13}C_9,^{15}N_2]$-dT. Due to the possibility that this isotopologue is generated both by processes in the gDNA as well as the soluble pool, a determination of its origin is not possible. In respect to the background signal due to natural dT, we do not observe any improvement. CDA/DCTD still have to be considered as the major contributors to formation of labeled dT upon repair of m^{5}dC.

3.2.2.3.2   Administration of F-dC and $[^{13}C,D_3]$-methionine on the CD DKO mESCs

Due to the complexity of the experiment reported above, we thought of a simplified set-up, in which we administered F-dC to wt mESCs and CD DKO cells under labeling with $[^{13}C,D_3]$-methionine. The nucleosides that could potentially be detected in this experiment are shown in **Figure 29**.
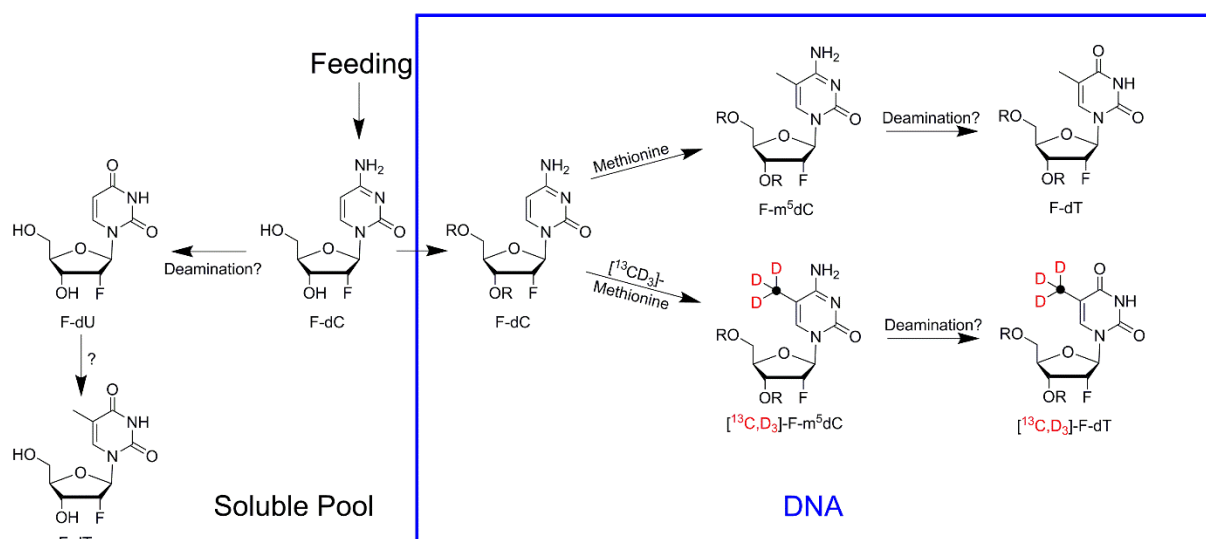
Figure 29: Possible nucleosides in the soluble pool and gDNA upon feeding of F-dC with [^{13}C,D$_3$]-methionine. Black dots indicate ^{13}C atoms.

The data (not shown) suggest that F-dC is incorporated into the gDNA of all mESCs, but wt cells contain with only approximately 0.2% of dN 100-fold more of this nucleoside. Furthermore, we observed that upon spiking of an isotope standard, the background levels, increase and the KO mESCs do not possess amounts above LOD. The measurements for F-m^{5}dC and [^{13}C,D$_3$]-F-m^{5}dC unfortunately had a technical problem, since parts of the peak eluted earlier than expected and this fraction was not recorded. We were however able to detect small amounts of the unlabeled F-m^{5}dC in wt mESCs and we also found labeled F-m^{5}dC. The gDNA of CD DKO mESCs, however, does not contain any F-m^{5}dC derivative exceeding LOD. The gDNA of wt mESCs additionally exhibited small amounts of F-dT and F-dU, whereas these nucleosides were not found in the gDNA of CD DKO mESCs. [^{13}C,D$_3$]-F-dT was not detected in any of the cell lines.

We then investigated the soluble pool for its nucleoside content. Here, the amounts of F-dC were comparable between the cell lines, but slightly higher in the wt. The area of the signal was, however, only about 1/25 of those in the gDNA of wt mESCs. This finding is interesting, since this means that most of the administered nucleoside is incorporated into the DNA. F-dU and F-dT were also only found in traces and only in the wt. Surprisingly, we furthermore found traces of [^{13}C,D$_3$]-F-m^{5}dC in the soluble pool of wt mESCs, but not in the KO mESCs. This nucleoside can only be generated and found in the soluble pool after incorporation of F-dC, its substitution with a [^{13}C,D$_3$] methyl group and finally excision through DNA repair. If this is not an artefact of the measurement, this result might hint at an inability of CDA and DCTD to process the F-nucleoside. Analysis of the gDNA and soluble pool both did not detect any [^{13}C,D$_3$]-F-dT in any of the cell lines and the formation of F-dT seems in general only little in mESCs. This experiment can therefore not determine the involvement of the deaminases CDA and DCTD in formation of dT from m^{5}dC.

3.2.2.4 *Investigation of DNMT3 enzymes in a deamination process*

Subsequently, we wanted to investigate the role of the DNMT3 enzymes in the deamination process. The previous experiments hinted at a reduction of [^{13}C,D$_3$]-labeled dT in the KO cells, especially in case of DNMT3b. Since in the Dnmt3b$^{-/-}$ cells the global m^{5}dC levels are also decreased, the effects might correlate. DNMT3b occurs in multiple splice variants, of which only one isoform (DNMT3b1) is catalytically active.[411] Complementation with the stable and constitutively expressed DNMT3b1 isoform could however only restore m^{5}dC levels, but not labeled dT. This suggested an occurrence of m^{5}dC to dT transition upon specific genomic events depending on DNMT3b, which cannot be restored after DNMT3b has been absent from the cell. The investigated mESC lines apart from the Dnmt3b1 complemented mESCs were first cultured on feeder cells and then cultured feeder independent under 2i/L conditions for two passages prior to any investigations in our laboratory. The Dnmt3b1 complemented mESCs were generated in 2i/L conditions. The parental Dnmt3b$^{-/-}$ cells had also long been cultured in 2i/L conditions. These cells are therefore expected to have lost all methylation that cannot be maintained in the 2i/l conditions, where the global methylation is strongly reduced. For our investigation, we only used cells that had been cultured in a2i/L conditions upon their receipt unless stated otherwise.
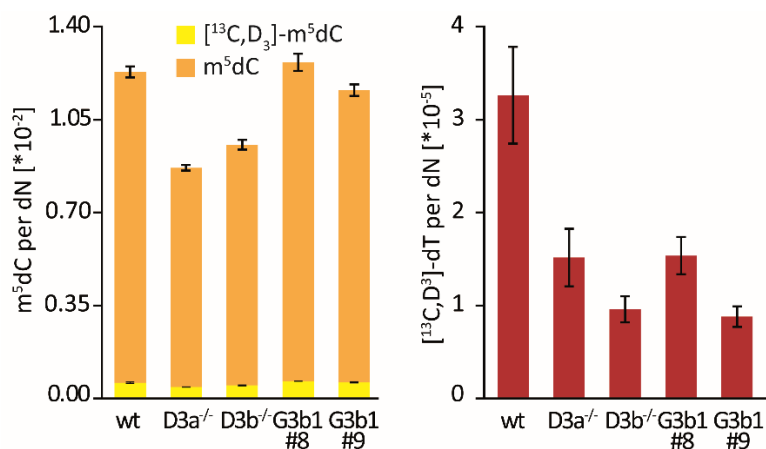
Figure 30: Levels for unlabeled and [^{13}C,D$_3$]-labeled m^{5}dC and [^{13}C,D$_3$]-dT in wt, Dnmt3a$^{-/-}$ (D3a$^{-/-}$), Dnmt3b$^{-/-}$ (D3b$^{-/-}$), as well as two clones (#8 and #9) of complemented cells expressing Dnmt3b1 (G3b1#8/9) upon administration of [^{13}C,D$_3$]-methionine after 4.5 d of *priming* with serum/L conditions (starting from 2i/L). The data represent one biological replicate, measured as technical triplicate.

The new results of labeling mESCs, that are routinely maintained in 2i/L conditions, with [^{13}C,D$_3$]-methionine during serum/L *priming* for 4.5 d (**Figure 30**) show that we can verify the previous observations with the absolute quantification method. The reduction of methylation and dT formation in Dnmt3b$^{-/-}$ cells is, however, slightly less and in Dnmt3a$^{-/-}$ slightly more pronounced. When we complemented Dnmt3b$^{-/-}$ cells with constitutively expressed DNMT3b1 (the only catalytically active isoform of DNMT3b) in two separate clones (#8 and #9), we found that the methylation levels are restored to those of the wt. However, rescue of m^{5}dC to dT transition is not possible and the levels

stay at *knockout* levels. We therefore keep the hypothesis that the DNMT3 enzymes are likely to be involved in the formation of dT from m^5dC.

Next, we investigated the levels of $[^{13}C,D_3]$-labeled m^5dC and subsequent formation of labeled dT at different time points of a C/R *priming* procedure (starting from 2i/L as the naïve culture) during the administration of $[^{13}C,D_3]$-methionine. Depicted in **Figure 31** are data from three different experiments with wt, Dnmt3a$^{-/-}$ (D3a$^{-/-}$), Dnmt3b$^{-/-}$ (D3b$^{-/-}$) mESCs, as well as two clones (#8 and #9) of complemented mESCs expressing Dnmt3b1 (G3b1#8/9). In the first experiment, we performed a time course with cell harvests at 3 d, 4 d and 5 d of the *priming*/labeling procedure. In comparison, the second experiment was only conducted a total of 3 d, whereas the third experiment was performed with 4 d of *priming*/labeling.
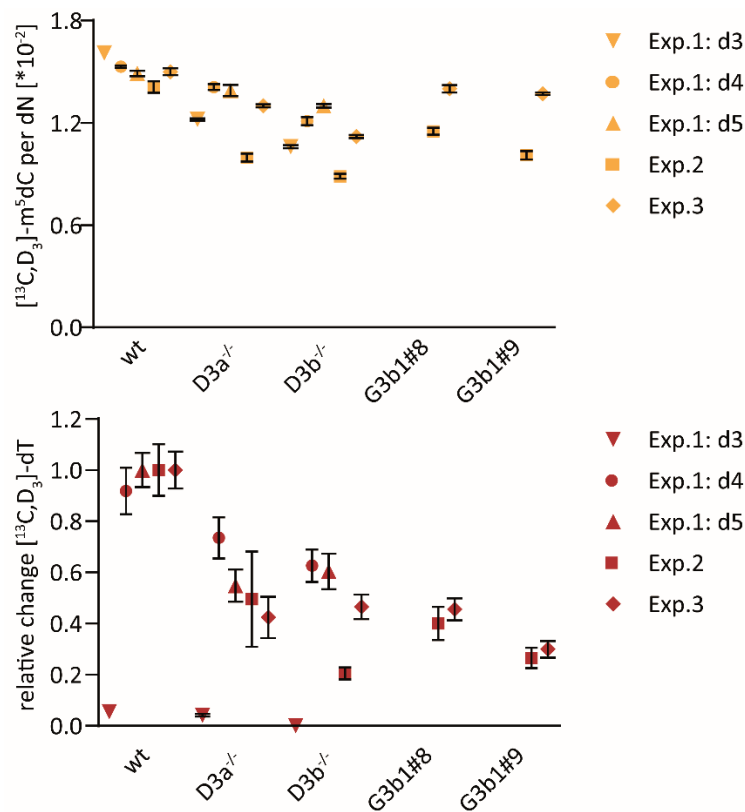
Figure 31: Levels for $[^{13}C,D_3]$-labeled m^5dC and $[^{13}C,D_3]$-dT in wt, Dnmt3a$^{-/-}$ (D3a$^{-/-}$), Dnmt3b$^{-/-}$ (D3b$^{-/-}$), as well as two clones (#8 and #9) of complemented cells expressing Dnmt3b1 (G3b1#8/9) upon administration of $[^{13}C,D_3]$-methionine after different lengths of *priming* with C/R conditions (starting from 2i/L). The data represent three separate experiments (1, 2 and 3), of which 1 is a time course after culturing the cells under *priming* conditions for a different time. All samples were measured as technical triplicate.

The time course experiment shows clear differences in the methylation levels between wt, D3a$^{-/-}$ and D3b$^{-/-}$ mESCs after 3 d of *priming* with C/R. The levels of $[^{13}C,D_3]$-m^5dC at this time point are reduced by 30-40% in the KO mESCs. At the 4 d time point, the wt shows slightly lower levels of $[^{13}C,D_3]$-m^5dC, whereas both KO mESCs exhibit increased levels. The overall reduction in comparison to the wt mESCs is therefore decreased to 10-25%. At the last time point, the levels between the wt and KO mESCs are

even more similar and the [13C,D$_3$]-m^{5}dC reduction only amounts to 5-15% compared with the wt. In the second experiment of a total of 3 d, the levels for [13C,D$_3$]-m^{5}dC are a little lower than in the first experiment, but the relative reduction of the KO mESCs towards the wt mESCs mirrors the 3 d time point of the first experiment. The third experiment with a total of 4 d of *priming*/labeling also exhibits [13C,D$_3$]-m^{5}dC levels that are very similar to the 4 d time point of the first experiment. Similar to the results for serum/L *priming*, the formation of [13C,D$_3$]-dT seems to be significantly reduced in the KO mESCs by up to 50% of wt levels. Variations of the absolute levels between the experiments might be explained by the different durations, since the seeding density of the cells was adjusted for maximum harvesting amounts. Cell density might affect the speed of *priming* and therefore the turnover and subsequently the amount of specific nucleotides in the gDNA.

The two complemented cell lines were only investigated in the second and third experiment. The respective methylation levels are elevated in comparison with the D3b$^{-/-}$ mESCs, but do not reach wt levels as observed in the serum/L *priming* conditions. The [13C,D$_3$]-dT levels, however, show the same effect as in the serum/L conditions and also do not exceed the levels of the D3b$^{-/-}$ mESCs. This strengthens the hypothesis that loss of DNMT3b leads to irreversible erasure of processes in the cell, in which m^{5}dC gets deaminated to dT. Prolonged culture in 2i/L conditions will add to this effect, because in these conditions DNMT3a and b are expressed at very low levels.

As an additional experiment, we were interested in determining the effect of m^{5}dC to dT transition after a culturing procedure in alternative naïve (a2i/L) conditions.[392, 399, 405] We made sure that the mESCs used in these experiments had been cultured in 2i/L for the shortest possible lengths. In case of the complemented Dnmt3b1 cells with their different culturing history, we expect significant differences due to the effect of long culture in 2i/L. The reason for our investigation was the hypothesis, that in these naïve conditions methylation occurs already and that an early direct repair process could subsequently take place and be observed as well.
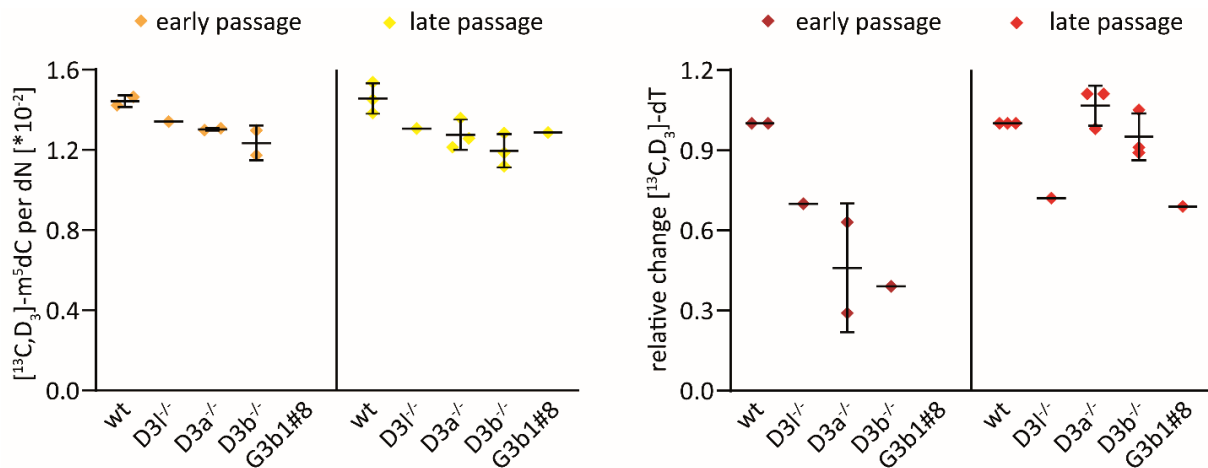
Figure 32: Levels for $[^{13}C,D_3]$-labeled m^5dC and $[^{13}C,D_3]$-dT in wt, Dnmt3l$^{-/-}$ (D3l$^{-/-}$), Dnmt3a$^{-/-}$ (D3a$^{-/-}$), Dnmt3b$^{-/-}$ (D3b$^{-/-}$), as well as one clone (#8) of complemented cells expressing Dnmt3b1 (G3b1#8) upon administration of $[^{13}C,D_3]$-methionine after different culturing lengths in naïve conditions (a2i/L) and subsequent *priming* in C/R conditions. The data represent five separate experiments, of which up to two have been performed with mESCs at early passages of culturing in a2i/L and up to three with later passages. All samples were measured as technical triplicate.

Our data (see **Figure 32**) indeed show reproducible $[^{13}C,D_3]$-m^5dC levels in the a2i/L conditions that are comparable to those found in primed cultures. The three *knockout* cell lines D3l$^{-/-}$, D3a$^{-/-}$ and D3b$^{-/-}$ exhibit reproducibly lower methylation than the wt mESCs, but the decrease is smaller than 20%. Interestingly, although the cell line was only analyzed once and therefore cannot be considered biologically significant, we did not detect an increase of $[^{13}C,D_3]$-m^5dC in the mESCs complemented with DNMT3b1.

The results for $[^{13}C,D_3]$-dT are quite interesting. In some experiments, the levels in the D3a$^{-/-}$ and the D3b$^{-/-}$ mESCs do not differ from the wt, but in other experiments the effect is the same as observed in *priming* experiments. When we looked at the culturing history of those cultures, we realized that the KO mESCs with a different result from the wt were at an early passage after thawing. At a later passage in contrast, the results did not differ from those of the wt. Since with a2i/L conditions DNMT3a and b are expressed at relatively high levels, these results suggest that upon prolonged culture under these conditions DNMT3a and b cross-complement each other in methylating the sites that are turned over by direct DNA repair. Interestingly, the D3l$^{-/-}$ mESCs were only analyzed in one early and one late passage, but show the same relative levels compared to the wt. This observation might hint at a loss of specific methylation that requires DNMT3l as a cofactor for action of the DNMT3a or b enzymes, which cannot be cross-complemented. The complemented cell line was only investigated once at a later passage and exhibits decreased $[^{13}C,D_3]$-dT levels compared with all other cell lines. This result might stem from the overall different culturing history of these cells.

To sum up the results, our data support an involvement of the DNMT3 enzymes in an active demethylation process *via* direct DNA repair. If the DNMT3b enzyme is ablated from the cells, the

ability to convert m^{5}dC to dT is not restored even after reintroduction of the enzyme. We presume however that the DNMT3 enzymes do not perform m^{5}dC deamination reaction themselves, but that they are rather responsible for dC methylation at sites that are turned over by direct DNA repair. Prime candidates for such sites[96] are genomic imprints, since maintenance of imprinted methylation in the embryo depends on all three catalytically active DNMT enzymes (1, 3a and 3b). Once imprints are lost in mESCs deficient for any of these enzymes they cannot be restored even after re-expression of the missing enzyme as their establishment requires (unknown) factor(s) present only in the germ line. Our observation of elevated formation of dT from m^{5}dC during *priming*, which recapitulates the *priming* of the pluripotent epiblast during the embryonic days E3.5-E7.5, and the fact that sDMRs are set starting from the same period on, strengthen our hypothesis.

3.2.2.5 Analysis of uni-parental mESCs for m^{5}dC to dT transition

To further analyze genomic imprints – genomic regions that are differentially methylated depending on the origin of the allele – for their contribution to the conversion of m^{5}dC into dT we acquired uni-parental mESCs. These were provided by *Tristan Bouschet/Robert Feil* at the University of Montpellier and were generated in the labs of *K. John McLaughlin* and *Azim Surani*. These cells are derived from embryos that contain only paternal (androgenetic) or maternal (parthenogenetic) genomes and therefore lack methylation that is inherited exclusively from the female and male germ line, such as maternal and paternal methylation imprints, respectively. To ensure biological significance of the results, we analyzed two androgenetic (paternal; Ag B6 and AK2) and three parthenogenetic (maternal; Pg8-021, Pg BT6 and PR8) mESC lines in comparison with a bi-parental (wt-B1) mES cell line. In this experiment, we quantified the absolute amounts of [^{13}C,D$_3$]-dT upon [^{13}C,D$_3$]-methionine labeling.
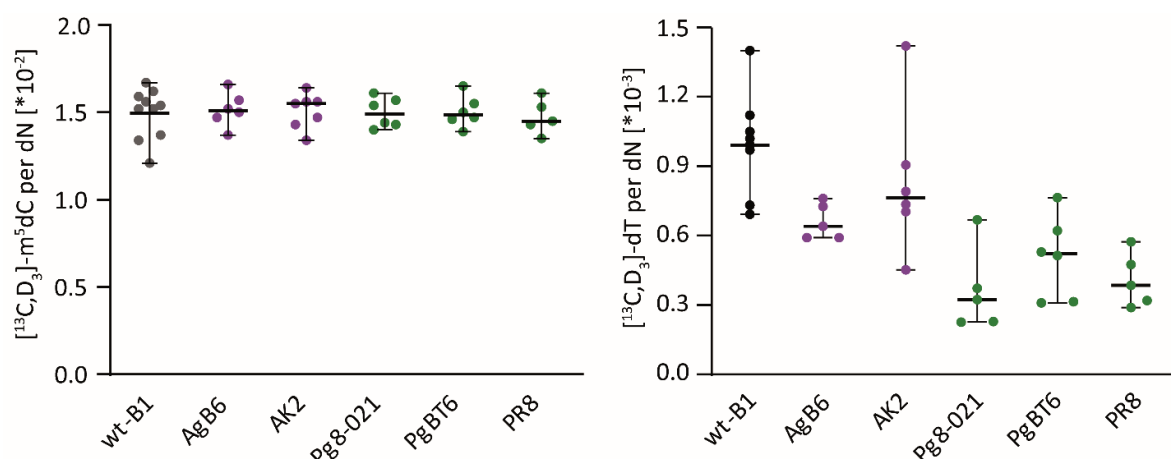
Figure 33: Levels of [^{13}C,D$_3$]-m^{5}dC and [^{13}C,D$_3$]-dT in uni-parental mESCs upon administration of [^{13}C,D$_3$]-methionine cultured for 5 d in C/R *priming* conditions; black: bi-parental control cells; purple: androgenetic cells; green: parthenogenetic cells. Each biological data point represents technical triplicates.

Our results (see **Figure 33**) show, that the levels of [^{13}C,D$_3$]-m^{5}dC do not differ between the different cell lines and are reproducible in several biological replicates. The levels for [^{13}C,D$_3$]-dT are, however, higher in bi-parental cells than in androgenetic and parthenogenetic cells. Curiously, addition of the [^{13}C,D$_3$]-dT levels of the androgenetic to the parthenogenetic cell lines equals the levels of the bi-parental cells. Whether this has a biological cause or is a coincidence needs further investigation. Furthermore, both androgenetic cell lines contain comparable levels of [^{13}C,D$_3$]-dT with respect to each other. The same result is also found for the three parthenogenetic cell lines when comparing them directly with each other. These findings strongly support the conclusion from the experiments on DNMT3 deficient cells and complementation of DNMT3b deficiency that methylation imprints represent a source for the conversion of m^{5}dC into dT.

3.2.2.6 Evaluation of haploid APOBEC3A KO mESCs

Recent literature[299-301] describes the potential of APOBEC3A to deaminate m^{5}dC in DNA. We found a haploid, parthenogenetic (maternally derived) cell line offered by the *HaploBank ES Cell Resource* and decided to investigate the formation of [^{13}C,D$_3$]-dT in these cells upon administration of [^{13}C,D$_3$]-methionine.

The results were rather inconclusive due to the low levels of the modification and the high background level from natural dT and are therefore not shown. This result matches the finding that parthenogenetic cells (see section **3.2.2.5**) generally exhibit less formation of [^{13}C,D$_3$]-dT in C/R conditions. We conclude that for an investigation of APOBEC3A as the potential DNA deaminase for m^{5}dC the generation of a respective CRISPR/Cas9-mediated diploid *knockout* mESC line would be useful. If the single *knockout* mESC line does not show an effect – maybe due to complementary action of AID –, the additional establishment of an Aid$^{-/-}$/Apobec3a$^{-/-}$ double *knockout* mESC line could shed light on the existence of genomic deamination.

3.2.3 Investigation of the base excision repair pathway

Base excision repair is a process to remove mispaired and modified bases from DNA through the action of glycosylases. The potentially mutagenic AP sites and β-elimination products that are subsequently formed cannot be quantified directly *via* UHPLC-QQQ-MS/MS. Underlying reason for this is the lack of a weak bond that will fragment in the mass spectrometer. In previous work from our group, a derivatization protocol for the analysis of AP sites and β-elimination products was established by *Toni Pfaffeneder, Olesea Kosmatchev* and *René Rahimoff*[410] using a hydroxylamine reagent (**Figure 34**) that reacts with the aldehyde moiety of the targets. The reagent furthermore contains a phenyl ring to intercalate with the DNA, a tetrazol ring, which releases nitrogen upon collision induced dissociation (CID) with N_2 gas, and a permanently charged quarternary ammonium ion, which increases the solubility and ionizability of the adduct.
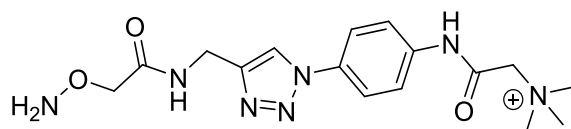
Figure 34: Structure of the derivatization reagent in use for AP sites and β-elimination products.

In cooperation with *René Rahimoff* and *Angie Kirchner*, I conducted further analysis of the base excision repair on various cell lines. *Angie Kirchner* performed the majority of the cell culture and feeding experiments for this investigation, before we both applied the DNA isolation protocol. Derivatization of the AP sites and β-elimination products was conducted as a team, and the digest of genomic DNA and mass spectrometric evaluation of the samples was performed by myself.

3.2.3.1 Global AP sites and β-elimination products in Tdg$^{-/-}$ and Smug1$^{-/-}$ cells

First, we were interested in analyzing the impact of the two DNA glycosylases TDG and SMUG1 on global AP site and β-elimination product formation. TDG was found to excise dT from dT:dG mismatches preferentially in CpG islands, as well as fdC, cadC, hmdU and fdU opposite of dG. For SMUG1, it has been shown that it excises XdU derivatives opposite of dG or dA on ds- and ssDNA. The experiments were performed in technical triplicates of one (SMUG1) or two (TDG) biological replicates.
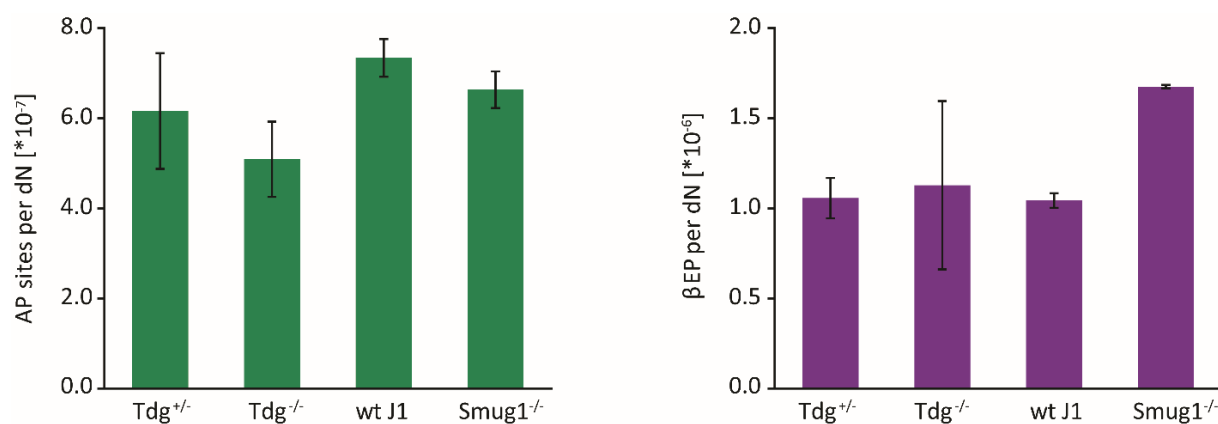
Figure 35: Levels of global AP Sites and β-elimination products (βEP) in TDG and SMUG1 KO cell lines and their respective wildtypes.

For both enzymes (**Figure 35**), the AP sites are slightly, but insignificantly decreased in the *knockout* cell lines. This result is not unexpected, since *knockout* of one of the several glycosylases in cells might not have a big contribution to the overall levels. The decreased levels of the *knockout* mESCs suggest however that both enzymes contribute to the generation of AP sites in gDNA. The experiments were only performed once (Smug1$^{-/-}$) or twice (Tdg$^{-/-}$) and have to be repeated to estimate the significance. In case of the β-elimination products, TDG also does not show any effect, but the high error bars of the Tdg$^{-/-}$ cells might obscure a difference. The Smug1$^{-/-}$ cell line, in contrast, shows 50% higher levels of β-elimination products. This observation might be explained by compensation for the loss of SMUG1 through TDG or MBD4, which could potentially recruit the bifunctional NEIL enzymes. These enzymes would subsequently generate more β-elimination products. Our observation is rather interesting, but again repetition of the experiment is necessary to verify our findings.

3.2.3.2 [$^{13}C_5$]-labeled AP sites and β-elimination products after administration of [$^{13}C_9$,N_3]-dC on Tdg$^{-/-}$ cells

Next, we wanted to investigate base excision repair specifically on dC. Therefore, *Angie Kirchner* supplemented Tdg$^{+/-}$ and Tdg$^{-/-}$ mESCs with [$^{13}C_9$,N_3]-dC. In this molecule, five ^{13}C atoms are present in the sugar moiety and should therefore still be present after excision of the base from the gDNA. All detected [$^{13}C_5$]-labeled AP sites and β-elimination products must stem from excision of the labeled and potentially modified cytosine derivatives. It is however possible that a deamination reaction on the nucleoside in the soluble pool, which could be incorporated into the gDNA, or direct deamination of the nucleotide in the gDNA can be processed by additional glycosylases. A small fraction of the labeled BER intermediates might therefore stem from these dU derivatives. Off note, previous data from our group[410] however showed no increase [$^{13}C_5$]-labeled AP sites and β-elimination products upon administration of [$^{13}C_{10}$,$^{15}N_2$]-dT and we can therefore exclude this possibility. In the KO cells, [$^{13}C_5$]-labeled AP sites and β-elimination products should only be generated from dC, m^{5}dC and hmdC

or the deaminated dU derivatives, because fdC and cadC cannot be excised anymore in the absence of TDG. The bases of hmdU and fdU will still be excised by SMUG1, and derivatives of dU can additionally be processed by UNG or MBD4.
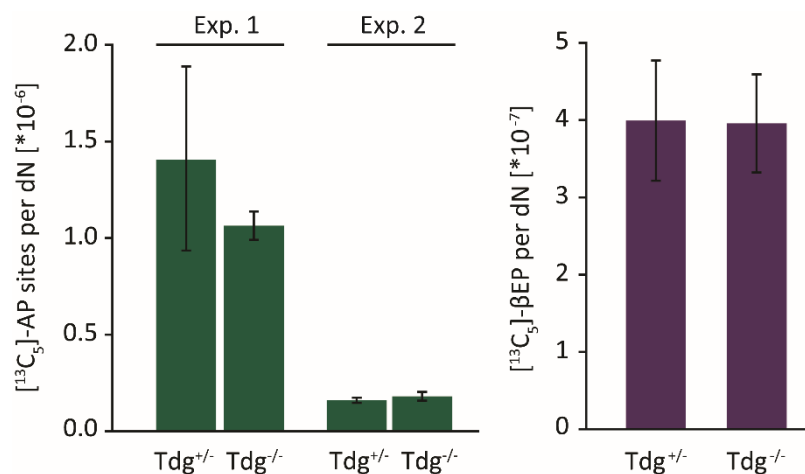
Figure 36: Levels of labeled AP sites and β-elimination products (βEP) in Tdg$^{-/-}$ cells and their respective wildtype after feeding of [$^{13}C_9$,$^{15}N_3$]-dC. The two different bar graphs for the AP sites represent two separate biological replicates (Exp.: experiment).

Upon administration of [$^{13}C_9$,$^{15}N_3$]-dC, the levels of [$^{13}C_5$]-labeled AP sites and β-elimination products did not differ in Tdg$^{+/-}$ and Tdg$^{-/-}$ cells (see **Figure 36**). The values of the [$^{13}C_5$]-labeled AP sites unfortunately show high variability between the two biological replicates and are therefore not combined. Although TDG acts on five substrates - dT in a dT:dG mismatch, fdC, cadC, hmdU and fdU –, their levels might however be rather low in comparison with the total amount of [$^{13}C_5$]-labeled AP sites and β-elimination products. Excised dU, which can be generated through deamination of dC could for example be a major contributor to the labeled BER products and is known to be processed by the UNG enzymes and MBD4. The two substrates hmdU and fdU can furthermore still be processed by SMUG1 and MBD4 and should not affect the levels. One needs to additionally consider the fact, that a substantial amount of the administered nucleoside might already be deaminated in the soluble pool leading to formation of [$^{13}C_9$,$^{15}N_2$]-dT. This dT isotopologue is incorporated into the gDNA in big quantities and might then be oxidized to hmdU and fdU by the TET enzymes. Excision of the respective bases can also be performed by SMUG1, since it accepts both dG and dA as nucleotides on the DNA strand opposite of its target.

Overall, the great variety of glycosylases and their overlap in substrate specificity establishes a highly developed network of back-up repair processes, which seems to compensate for the loss and effect of one absent glycosylase. In the mixed cell population, the overall kinetics of the BER products might be evened out and therefore locus-specific changes might not be visible.

3.2.3.3 $[^{13}C_5]$-labeled AP sites and β-elimination products after administration of $[^{13}C_5,^{15}N_2]$-fdC on Smug1$^{-/-}$ cells

Subsequently, we thought of narrowing down the potential sources for $[^{13}C_5]$-labeled AP sites and β-elimination products by applying $[^{13}C_5,^{15}N_2]$-fdC to Smug1$^{-/-}$ cells and their respective wildtype. SMUG1 shows no activity towards fdC, but in previous unpublished feeding experiments in our group (data not shown) of fdC we observed substantial amounts of fdU both in the DNA and the soluble pool.
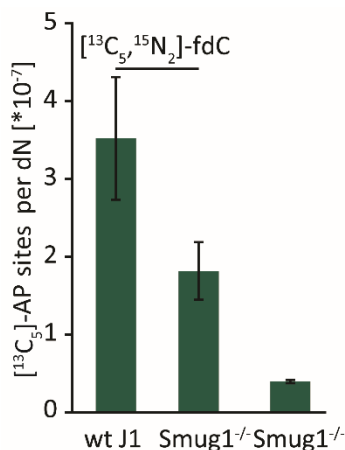
Figure 37: Levels of labeled AP sites in Smug1$^{-/-}$ cells and corresponding wt J1 cells after administration of $[^{13}C_5,^{15}N_2]$-fdC.

We found out that upon administration of $[^{13}C_5,^{15}N_2]$-fdC to wt and Smug1$^{-/-}$ cells no labeled β-elimination products are detected. This finding is reasonable, since fdC is processed by TDG, MBD4 and the NEIL enzymes. The latter are known to be bifunctional and to perform β,δ–elimination on the AP sites. After δ-elimination, the DNA does not contain any ^{13}C atoms anymore. Residual labeled β-elimination products might be present in amounts that do not exceed the LOD.

The Smug1$^{-/-}$ cells furthermore possess smaller amounts of $[^{13}C_5]$-labeled AP sites than the respective wt J1 cells (see **Figure 37**). This result matches our expectations, because SMUG1 excises fdU from DNA. In the *knockout* cells, the base excision repair of this nucleoside is reduced, but not absent due to the complementary action of TDG and MBD4. Our results however identify SMUG1 as the major glycosylase acting on this lesion. It is important to investigate the levels of the untreated control Smug1$^{-/-}$ control sample, because it shows a small signal for the labeled AP sites. This background signal is not extremely high, but has to be considered in the determination of the real amount of labeled AP sites.

3.2.3.4 Global AP sites and β-elimination products in Neil KO cells

Next, we wanted to determine the extent to which dC derivatives are processed by the base excision repair mechanism in Neil KO cells. To this end, *Angie Kirchner* and myself cultured Neil1$^{-/-}$, Neil2$^{-/-}$ or Neil1,2$^{-/-}$ mESCs and the respective wt cell line.
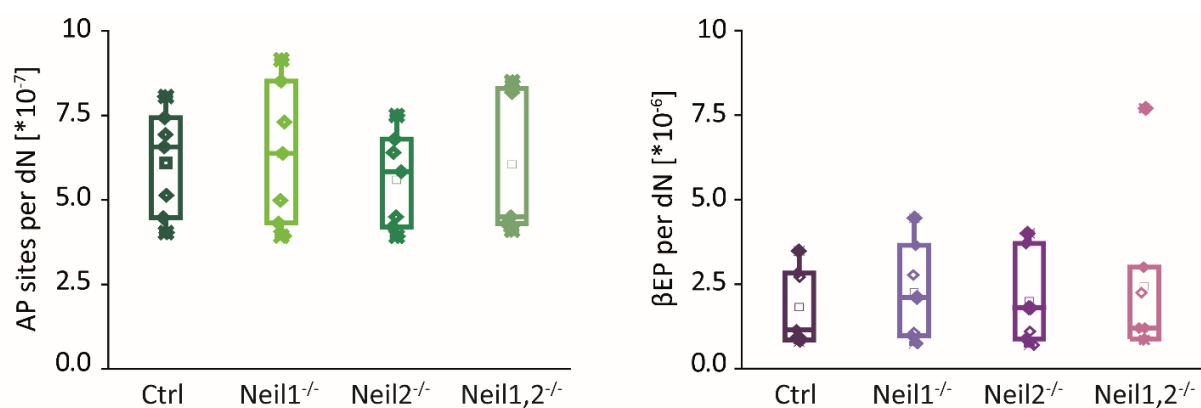
Figure 38: Statistical evaluation of the levels for the β-elimination products (βEP) and AP sites in gDNA of wt and Neil KO cells from seven different biological replicates.

After measurement of several biological replicates and statistical analysis of the data (see **Figure 38**), there did not seem to be a difference between the wt and Neil KO cells. All four cell lines show levels for AP sites in the range of $5\text{-}7.5 \bullet 10^{-7}$ per dN and levels for the β-elimination products in the range of $0.5\text{-}5 \bullet 10^{-6}$ per dN. As one outlier, we found higher levels of β-elimination products in the Neil1,2$^{-/-}$. Since the NEIL enzymes are thought to facilitate the release of other glycosylases from the AP sites they generate and to be part of a so-called BERosome,[35-36] double *knockout* of the enzymes might destabilize the machinery and not process all lesions efficiently. Nevertheless, a single biological deviation can only establish a hypothesis and not indicate biological significance. From the overall lack of an effect of the Neil KOs on global BER intermediates, we conclude, however, that the harmful AP sites and β-elimination products are still processed by APE1.

3.2.3.5 [$^{13}C_5$]-labeled AP sites and β-elimination products after administration of [$^{13}C_9,N_3$]-dC or [$^{13}C_9,N_3$]-C on Neil KO cells

In an additional experiment, we were interested in determining the amount of labeled AP sites and β-elimination products in the Neil KO cells upon administration of [$^{13}C_9,^{15}N_3$]-labeled dC or C. This experiment was especially interesting, since administration of [$^{13}C_9,^{15}N_3$]-C introduces the nucleoside in a part of the nucleotide biosynthesis pathway, where a deamination reaction through action of soluble deaminases can mostly be excluded. In the first step, RNA nucleoside specific kinases phosphorylate the nucleoside (see **Figure 39B**). Finally, the ribonucleotide reductase (RNR) generates the [$^{13}C_9,^{15}N_3$]-dCDP from [$^{13}C_9,^{15}N_3$]-CDP. Equilibrium reactions might still generate labeled dCMP from the labeled dCDP, which could subsequently be deaminated by DCTD. Nevertheless, the overall generation of labeled dU species should be reduced compared with direct administration of labeled dC (see **Figure 39A**).

A

CDA = Cytidine deaminase
DCTD = Deoxycytidylate deaminase
RNR = Ribonucleotide reductase
TK = Thymidine kinase
TS = Thymidylate synthase
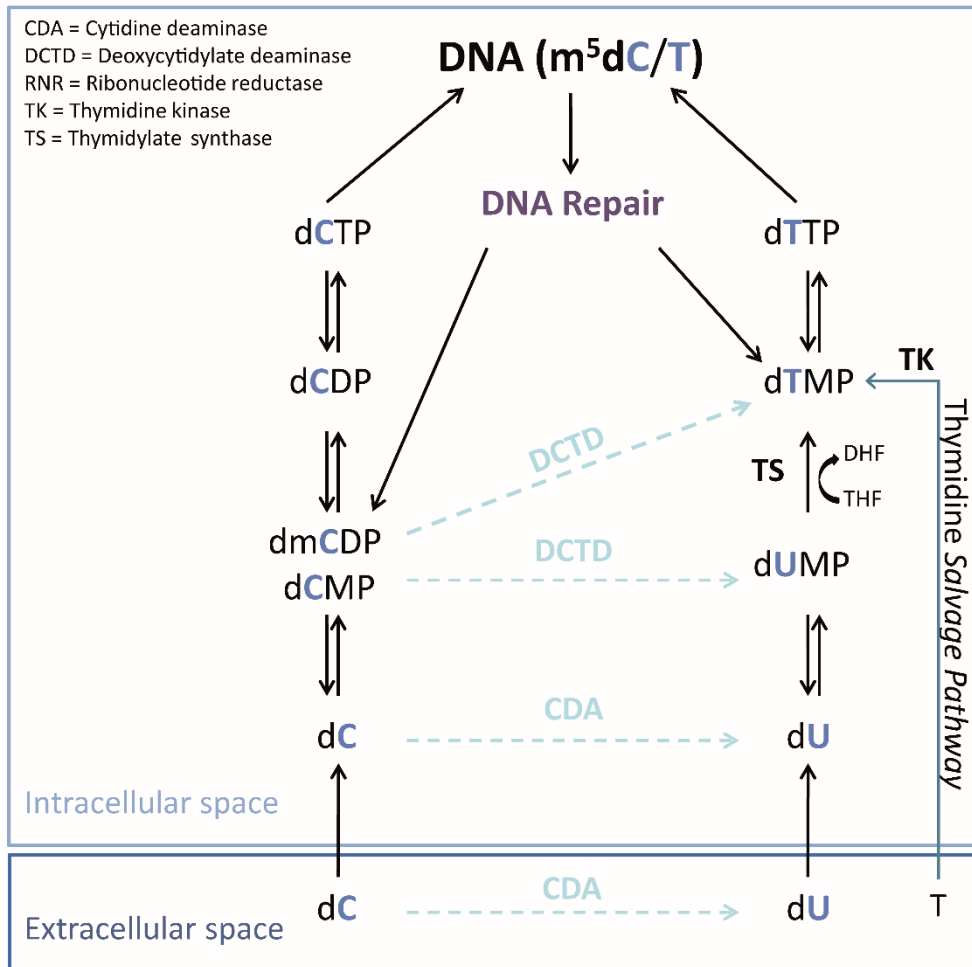
B
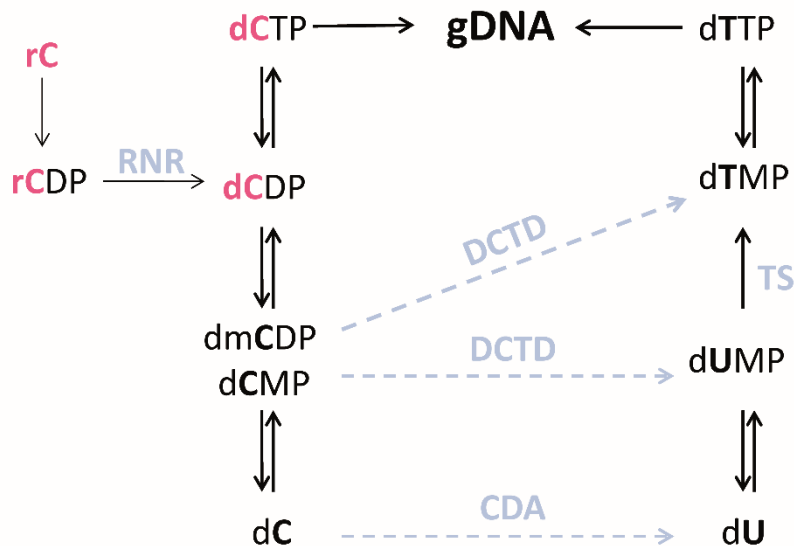
Figure 39: Metabolism of nucleosides and nucleotides and the corresponding enzymes. A) Depiction of the generation of DNA nucleotides with all responsible enzymes and in relation to RNA nucleotides. B) Generation of phosphorylated dC species upon administration of C.

Unfortunately, neither upon administration of $[^{13}C_9,^{15}N_3]$-labeled dC nor C, we were able to detect any $[^{13}C_5]$-labeled AP sites in the control and KO mESCs (data not shown). Analysis of unsupplemented

139

control cells showed high background levels for the signal of the labeled AP site adduct, which might suppress the real signal.
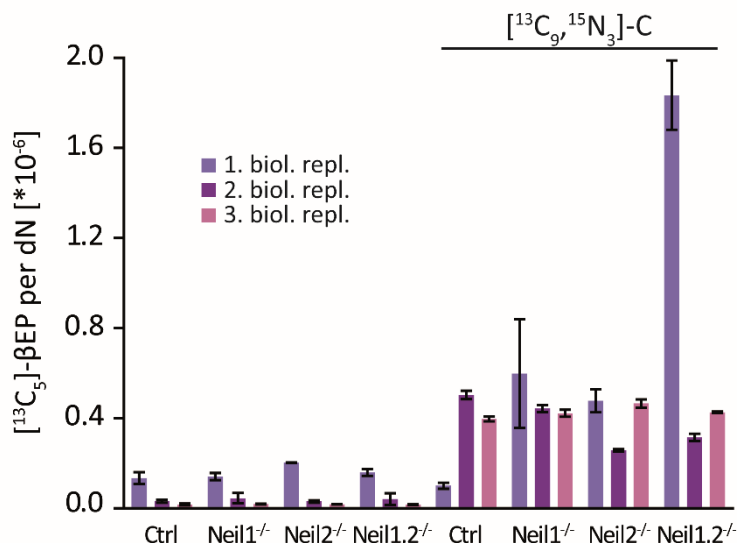
Figure 40: Levels of labeled β-elimination products ($[^{13}C_5]$-βEP) in Neil KO mESCs and corresponding control (Ctrl) cells after administration of $[^{13}C_9,N_3]$-C. On the left side the background levels of the labeled β-elimination products without administration of the nucleoside are depicted.

Furthermore, the results for $[^{13}C_5]$-labeled β-elimination products (**Figure 40**) are very diverse between the biological replicates (data for administration of $[^{13}C_9,^{15}N_3]$-dC are not shown). In one replicate, the Neil single *knockouts* show higher levels than the wt and the double *knockout* has a level that matches addition of both single *knockouts*. When we achieved this result, we were intrigued, but repetition of this experiment in two additional replicates did not lead to the same results. However, in the results for the Neil1,2$^{-/-}$ of those replicates, the levels of $[^{13}C_9,^{15}N_3]$-labeled β-elimination products are about one order of magnitude lower than those of the first replicate. It is therefore possible, that the derivatization reaction did not work to completion and the data do not reflect the real amount of the BER intermediates.

3.2.3.6 Quantification of formylcytosine as a product of BER

The high variability within the biological replicates might be due to the generally fast turnover of the BER intermediates, which is difficult to capture with our method. Subsequently, we decided to focus our investigation on BER of fdC and to analyze its other repair product, which is the excised base formylcytosine (see **Figure 41**).
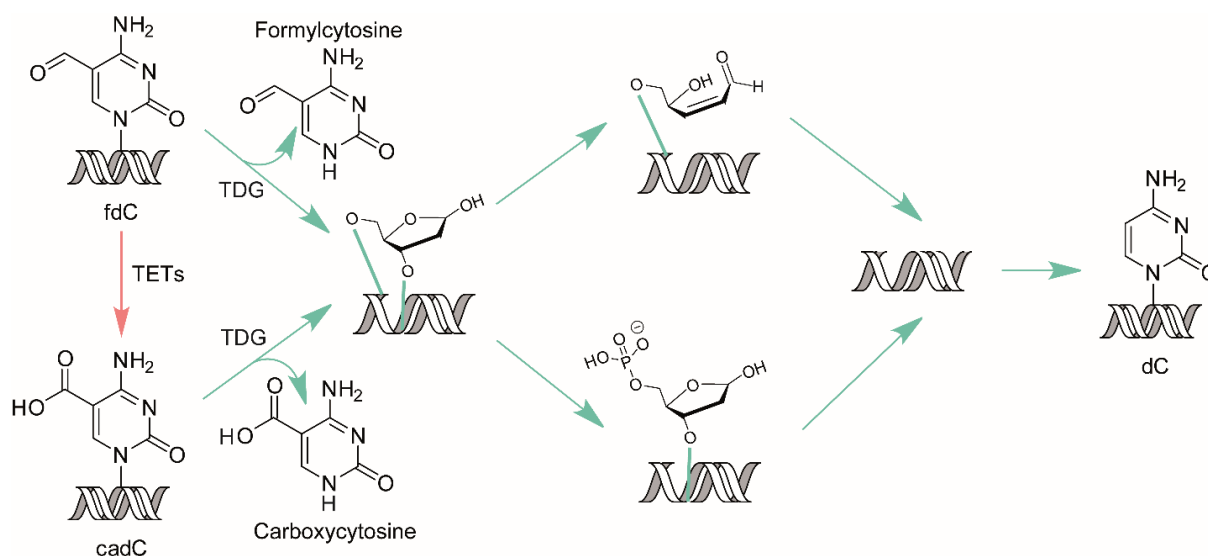
Figure 41: Schematic overview over base excision repair on fdC and cadC by TDG.

Previous unpublished research by *Toni Pfaffeneder* and *Matthias Q. Kurz* uncovered the reaction of the formyl group of fdC with the hydroxylamine moiety of the AP site reagent[410] under catalysis with *p*-methoxy-aniline. For the investigation of the excised base formylcytosine that is released in the BER of fdC, *Eva Korytiaková* synthesized the free formylcytosine and its $^{15}N_2$-isotopologue. We subsequently studied and optimized the reaction of this free base with the reagent and came to the conclusion, that addition of a catalyst is not necessary when we performed the reaction in a water/acetonitrile mixture at pH=10. After 1 h, the free base reached full conversion to the oxime adduct (see **Figure 42**) and the reaction was neutralized with formic acid and concentrated. Subsequently, excess of the derivatization reagent was removed from the sample by capturing it with an aldehyde resin to reduce ion suppression in the mass spectrometer.
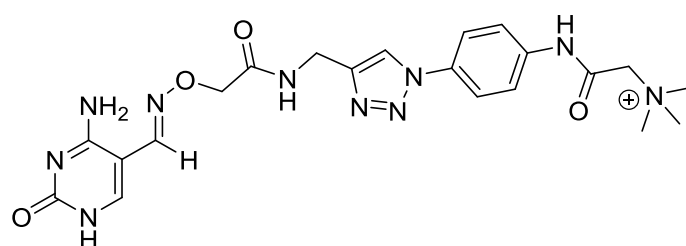
Figure 42: Structure of the formylcytosine adduct with the derivatization reagent.

Next, we wanted to investigate the soluble pool of naïve mESCs for the presence of formylcytosine. To this end, we cultured wt K3$^{+/+}$ cells under a2i/L conditions, because these conditions are known to lead to formation of fdC in the DNA. In case of base excision repair of this nucleotide, we would expect to find the free base in the cytoplasm. The extraction of the soluble pool was performed as previously described.[179] Due to the poor solubility of the free base, we decided however to perform the screening for the optimized derivatization reactions with soluble pool prior and after purification *via* SUPEL™

SELECT HLB SPE 60 mg/3 mL SPE tubes (*Supelco*). First analyses of the soluble pool did not lead to detection of derivatized formylcytosine. The reaction needs to be further optimized for smaller scale as well as same reactivity in the complex soluble pool sample creating high background signals.

3.2.4   Administration of azacytidine nucleoside analogues to study epigenetic processes

3.2.4.1 Investigation of epigenetic modification level changes upon administration of Aza(d)C to different cell culture systems

Azacytidines are antimetabolites used in the treatment of AML and MDS. They are considered to be epigenetic drugs, because they are known to inhibit DNMT enzymes. The compounds are labile towards nucleophilic attacks and therefore hydrolyze quickly. Multiple groups[159-160] and our own[161] developed methods for chemically stabilizing the azacytidines in DNA and RNA by hydrogenation, and thereby reduction of the double bond between N5 and C6. This reaction was achieved on the nucleoside (**Scheme 3**) with NaBH$_4$ and work-up with acidic aqueous buffer and the reaction was furthermore optimized on gDNA.
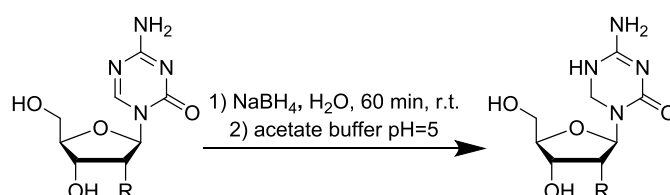
Scheme 3: Chemical stabilization of the azacytidines with NaBH$_4$ gives the hydrogenated nucleoside.

The chemically stabilized gDNA was then subjected to digestion conditions that had previously been used in our group.[412] High resolution MS fragmentation performed on an *Orbitrap XL* mass spectrometer (*Thermo Fisher Scientific*), equipped with a *HESI-II-ESI* source (*Thermo Fisher Scientific*), verified the correct fragmentation of the nucleoside with the mass transitions for *m/z* 231.1 → 115.1, as well as further fragmentation of *m/z* 231.1 → 72.1 and *m/z* 231.1 → 86.0 (see **Scheme 4**).
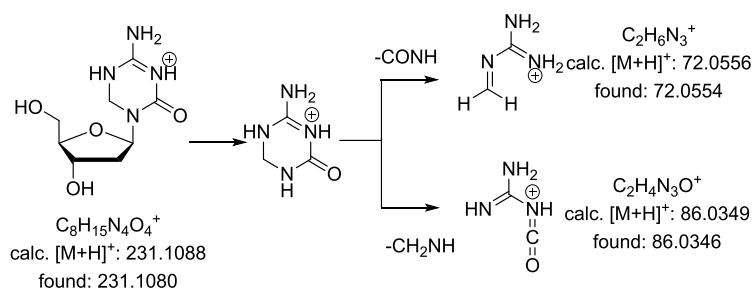
Scheme 4: Fragmentation pattern of the H$_2$-AzadC nucleoside with high resolution mass spectrometry.

With the verified mass transitions in hand, *Thomas M. Wildenhof* developed the method on the triple quadrupole mass spectrometer and was able to generate time- and dose-dependent data for the incorporation of the pharmaceutical in DNA from various cell lines. For absolute quantification, the reduced H$_2$-AzadC was utilized as an external standard, since no heavy labeled isotope was available.

When he subsequently wanted to apply the method to RNA, I joined the project. Together, we verified the fragmentations *m/z* 247.1 → 115.1, *m/z* 247.1 → 72.1 and *m/z* 247.1 → 86.0 for H$_2$-AzaC *via* high resolution mass spectrometry (as described above). Subsequently, we developed the triple quadrupole

method for quantification of AzaC-containing RNA based on a method developed by *Katharina Iwan*. Then we isolated RNA from previous samples and performed the enzymatic digestion. Analysis of the RNA samples was nevertheless not possible using the tested conditions, since just a weak signal was detected for the H_2-AzaC nucleoside. We attributed this effect to the early retention time of the nucleoside and the resulting high ion suppression by early eluting molecules and salts.

Therefore, we optimized the digestion conditions by testing different enzymes, salts and filtration methods and applied the tested conditions also to the external standard. We found that several factors were responsible for the loss of signal intensity, like EDTA and TRIS salts in the enzyme solution, and we were able to find a suitable protocol for the analysis of the samples. Subsequently, we optimized the previously established method for the DNA sample preparation in a similar manner. The final protocols for analysis of all samples and verification of previous measurements are given in **Table 1**.

Table 1: Overview over digestion conditions from the previous protocol and the optimized conditions for quantification of H_2-AzaC and H_2-AzadC in RNA and DNA, respectively.

	Old protocol	Optimized protocol DNA	Optimized protocol RNA
Amount of nucleic acid [µg]	1-5	1	1
Digestion length	1. step: 3 h 2. step: 3 h	1. step: 3 h 2. step: 3 h	1. step: 3 h 2. step: over night
Salts	1. step: 480 µM $ZnSO_4$ in 7.5 µL 2. step: 520 µM Na_2-[EDTA] in 7.5 µL	1. step: 480 µM $ZnSO_4$ in 7.5 µL 2. step: ---	1. step: 1.6 mM $ZnSO_4$ and 2.7 mM $MgCl_2$ in 7.5 µL 2. step: 1.6 mM $ZnSO_4$ and 2.7 mM $MgCl_2$ in 7.5 µL
Enzymes	1. step: 42 U Nuclease S1 and 5 U Antarctic phosphatase in 7.5 µL 2. step: 0.2 U Snake Venom Phosphodiesterase in TRIS-buffer	1. step:21 U Nuclease S1 and 2.5 U Antarctic phosphatase in 7.5 µL 2. step: 0.1 U Snake Venom Phosphodiesterase in TRIS-free buffer	1. step:21 U Nuclease S1 and 2.5 U Antarctic phosphatase in 7.5 µL 2. step:21 U Nuclease S1 and 2.5 U Antarctic phosphatase in 7.5 µL
Filtration	0.2 µm Supor	0.45 µm Supor	0.45 µm Supor
Quantification	Internal standard	External standard in digestion conditions	External standard in digestion conditions

3.2.4.1.1   Analysis of wt J1 and Dnmt KO cell lines treated with AzaC and AzadC

Thomas M. Wildenhof's experiments analyzed the time- and dose-dependent integration of AzaC and AzadC into leukemia cell lines and mESCs. In a next experiment, we wanted to investigate the incorporation of AzaC and AzadC into the DNA of wt J1 and different Dnmt KOs[161] (see **Figure 43**), and the resulting methylation levels to uncover the effects of the drugs on the selected DNA methyltransferases. To this end, the *knockout* cell lines were supplemented with either 1 µM AzaC or 1 µM AzadC in 0.0001% DMSO. As a control, only DMSO was added. Although most of the data have previously been published,[161] the full data set is displayed and analyzed in **Figure 43**-**Figure 45**.
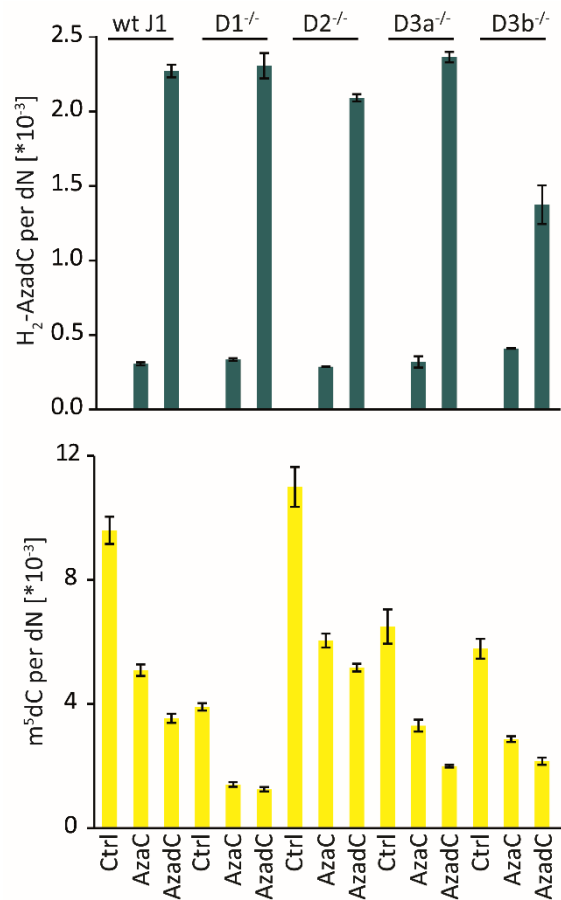
Figure 43: Results of Aza(d)C-treatment of Dnmt KO mESCs. The upper row shows a comparison of H_2-AzadC levels after treatment with DMSO (Ctrl), 1 µM AzaC or 1 µM AzadC in DNA of wt J1 mESCs and the respective Dnmt1$^{-/-}$ (D1$^{-/-}$), Dnmt2$^{-/-}$ (D2$^{-/-}$), Dnmt3a$^{-/-}$ (D3a$^{-/-}$) and Dnmt3b$^{-/-}$ (D3b$^{-/-}$) mESCs. The lower row shows the corresponding m^5dC levels.

The control samples show no H_2-AzadC and the expected levels for m^5dC in a primed mES cell line. When comparing AzaC-treated cells with AzadC-treated cells, one can clearly see that the nucleoside is integrated to an approximately 10-fold smaller extent after treatment with AzaC with about 0.25-0.4•10^{-3} H_2-AzadC per dN. After AzaC-treatment, the AzadC incorporation seems slightly higher in the Dnmt1$^{-/-}$ and Dnmt3b$^{-/-}$ cells than in the wt and the other Dnmt KOs.

After treatment with AzadC, the levels are at 1.3-2.4•10^{-3} H_2-AzadC per dN. Interestingly, the lowest levels for H_2-AzadC are observed in Dnmt3b$^{-/-}$ cells at about half of the wt levels and second-lowest in Dnmt2$^{-/-}$ at about 80%.

Analyzing the m^{5}dC levels, the Dnmt3a$^{-/-}$, Dnmt3b$^{-/-}$ and Dnmt1$^{-/-}$ cells have a 1.3-2.5 fold lower basic level, whereas the Dnmt2$^{-/-}$ cells show even slightly higher methylation. After treatment with AzaC, the methylation levels of the Dnmt3a$^{-/-}$ and Dnmt3b$^{-/-}$ cells are reduced to 50% of the controls, Dnmt2$^{-/-}$ levels are reduced to 60% and Dnmt1$^{-/-}$ levels are reduced to 40%. Treatment with AzadC, in comparison, leads to reduction of dC methylation to 50% of the controls in Dnmt2$^{-/-}$, to 40% in Dnmt3a$^{-/-}$ and Dnmt3b$^{-/-}$, and to 30% in Dnmt1$^{-/-}$. The drug therefore does not seem to show a direct correlation between the amount of its incorporation and the reduction in the m^{5}dC levels.

In conclusion, the Dnmt2$^{-/-}$ cells show comparable results for incorporation of AzaC and methylation changes as the wt, although the m^{5}dC levels are slightly higher. The incorporation of AzadC is not as high as for wt mESCs and therefore the decrease in m^{5}dC is less pronounced. The Dnmt3a$^{-/-}$ cells do not possess the same basic methylation due to the missing *de novo* methylation during *priming*, but the trends of the methylation levels are comparable to those of the wt. The most interesting results are those for Dnmt1$^{-/-}$ and Dnmt3b$^{-/-}$ cells. The former generally exhibit the largest drop of the m^{5}dC levels, but comparable incorporation of the drug. The latter incorporates a 50% smaller amount of AzadC into the DNA than the other cell lines and still shows a decrease in the methylation level that is almost as strong as for Dnmt3a$^{-/-}$. So far, it is unclear why those cells integrate less of the drug and whether the reduction is smaller due to the decreased incorporation or if other processes are involved, *e.g.* repair or reduced cell growth. The observation however suggests a better therapeutic effect in absence of DNMT3b, although *Oka et al.*[173] describe higher resistance of Dnmt3b$^{-/-}$ mESCs towards AzadC.
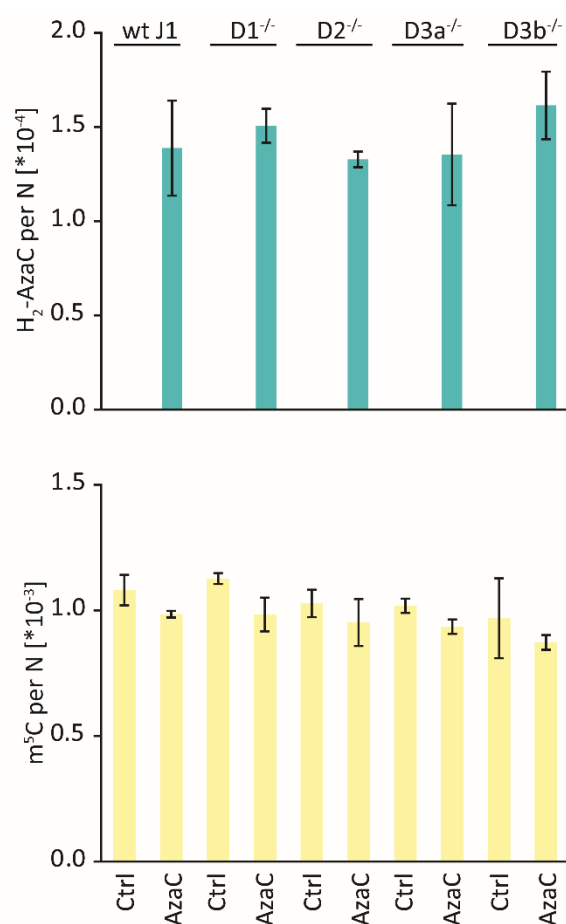
Figure 44: Comparison of H_2-AzaC and m^5C levels in isolated RNA of different Dnmt KOs. The upper row shows a comparison of H_2-AzaC levels after treatment with DMSO (Ctrl) and 1 µM AzaC in RNA of wt J1 mESCs and the respective Dnmt1$^{-/-}$, Dnmt2$^{-/-}$, Dnmt3a$^{-/-}$ and Dnmt3b$^{-/-}$ KO cells. The lower row shows the corresponding m^5C levels.

Analysis of the RNA of the Dnmt KOs (**Figure 44**) reveals a different situation. Surprisingly, the general levels of H_2-AzaC are by one order of magnitude lower than those of H_2-AzadC in gDNA. This might be explainable by the higher turnover rate of RNA. As key players for protein synthesis, RNA with covalently bound and therefore inhibited methyltransferases could be recognized to be malfunctioned even earlier and be subjected to RNA decay or repair. Overall, the incorporated AzaC levels are similar in all cell lines. Looking at the m^5C levels, a significant change of methylation due to the *knockout* of any DNMT enzyme or administration of the drug is not observed. Dnmt1$^{-/-}$, Dnmt3a$^{-/-}$ and Dnmt3b$^{-/-}$ mESCs are not expected to show any effect, since the enzymes are not active on RNA. Dnmt2$^{-/-}$ mESCs, however, are also not affected. This might be explained by the high variety of RNA methyltransferases, of which some might generally not react with azacytidines, but also by the higher turnover of RNA.

Later on, we realized that the isolated RNA was actually not total RNA, since the utilized protocol apparently depleted a major fraction of the tRNA. We decided to investigate the total RNA of the same samples using an isolation protocol optimized for the recovery of tRNA.
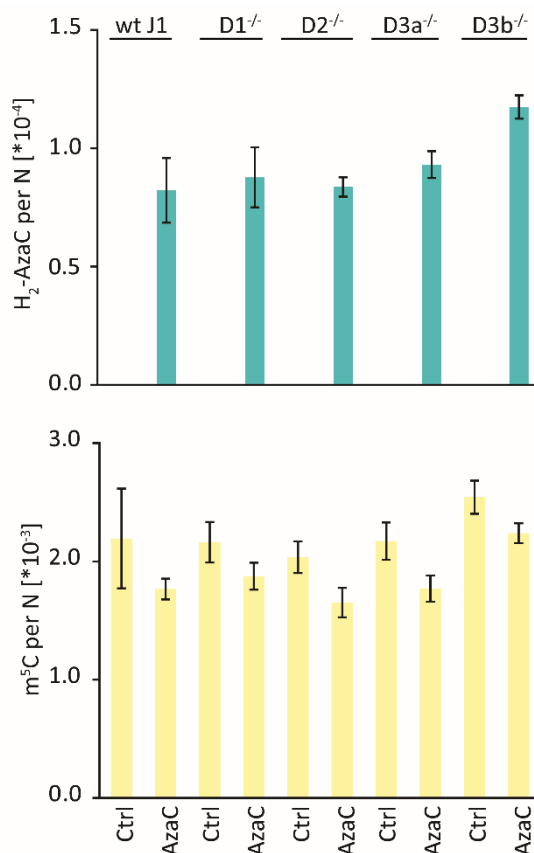
Figure 45: Comparison of H_2-AzaC and m^5C levels in total RNA of different Dnmt KOs. The upper row shows a comparison of H_2-AzaC levels after treatment with DMSO (Ctrl) and 1 µM AzaC in RNA of wt J1 mESCs and the respective Dnmt1$^{-/-}$, Dnmt2$^{-/-}$, Dnmt3a$^{-/-}$ and Dnmt3b$^{-/-}$ KO cells. The lower row shows the corresponding m^5C levels.

The results of total RNA analysis (**Figure 45**) show a significantly lower incorporation of AzaC per nucleoside than the results without tRNA. This is, however, not surprising since mRNA with their high turnover rates and thereby constant synthesis has a higher probability of incorporating AzaC. If this type of RNA is enriched over a much more abundant RNA that turns over slower - incorporating less AzaC - the average incorporation of the nucleoside increases. Interestingly, however, the levels of m^5C seem to be negatively affected when tRNA is present in the RNA preparation. Since the results are average numbers for the whole RNA content of the cell, we can assume that pure tRNA will show a rather large decrease of the m^5C levels. To take the evaluation even further, we hypothesize that the drug inhibits mostly methyltransferases modifying tRNA.

3.2.4.1.2   Analysis of treating cancer cells with AzaC and AzadC

Further studies of leukemia cell lines were performed in cooperation with *Laura Bocci* from the *Spiekermann* laboratory. In their group, they wanted to study the effects of the two pharmaceuticals *in vivo* and determine differences on a genomic and transcriptomic level.

First, a variety of cell lines was analyzed with the established method (details for those cell lines are given in **Table SI-2**). To this end, *Laura Bocci* performed *in vivo* treatment of various leukemia cell lines and provided me with the treated cells for further analysis. According to the literature, the plasma levels of the drugs in patient blood samples amount to 0.3-1.6 µM for AzadC[413-414] and 3-11 µM for AzaC.[415] We therefore decided to consistently apply 0.5 µM AzadC and 3 µM AzaC to all of the following experiments of this collaboration.
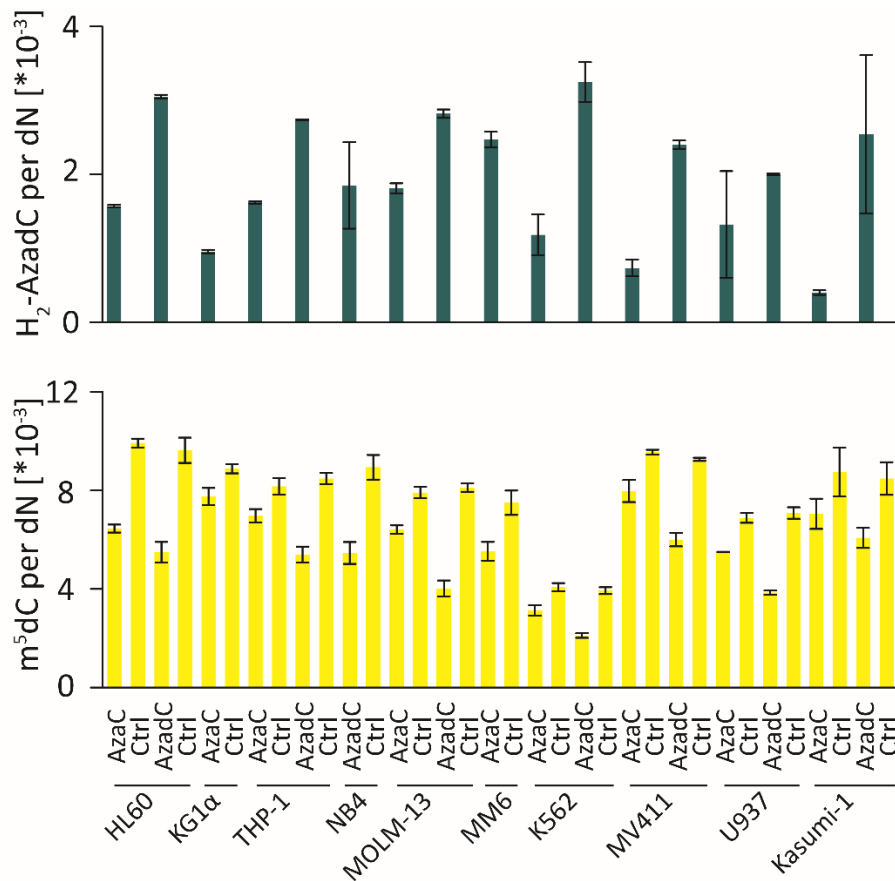
Figure 46: Screening of various AML cell lines for AzadC incorporation into DNA and the effect on m^{5}dC; all data represent one biological replicate measured in technical triplicates; dark green bars: H$_2$-AzadC, dark yellow bars: m^{5}dC.

Analyzing the DNA of the leukemia cell lines, we achieved valid results for most of the cell lines and conditions, but due to a technical problem, some samples had to be dismissed. All measurements were performed in technical triplicates, but only of one biological replicate.

In general, it seems like all cell lines incorporate more AzadC into DNA when they are treated with AzadC directly (see **Figure 46**) amounting to $3\text{-}3.5 \bullet 10^{-3}$ per dN for administration of AzadC and $0.25\text{-}2 \bullet 10^{-3}$ per dN for administration of AzaC. This matches our previous findings in the experiments of *Thomas M. Wildenhof* (see section **3.1.4**), although overall the tested cell lines incorporate more AzadC. In samples with only one treatment condition shown (KG1α, NB4 and MM6) a comparison was not possible. The cell line MM6 is an exception, because it incorporates about $3 \bullet 10^{-3}$ H_2-AzadC per dN upon administration of AzaC.

The m^5dC levels appear to be rather inhomogeneous between the different leukemia cell lines with levels ranging from 40-100% compared to naïve pluripotent mESCs. The cell line K562 shows especially low levels of m^5dC already in the untreated control and might lack or express very small amounts of DNMT enzymes. Generally, the m^5dC levels verify that a reduction upon treatment is observed in all cell lines and conditions. It seems like the decrease of the m^5dC level is bigger after treatment with AzadC than with AzaC. This is not surprising due to the higher AzadC content in the DNA, but the effect appears not proportional to the incorporated amount of AzadC between the two drugs. MM6 is again surprising, since the relatively high levels of H_2-AzadC after administration of AzaC only lead to comparably low changes in the m^5dC levels of 25%. In contrast, upon administration of AzadC leading to comparable H_2-AzadC levels in other cell lines, the reduction of m^5dC is about 30-50%. This finding hints at additional effects of AzadC on m^5dC levels, *e.g.* indirectly by interaction of the nucleoside with the DNMT enzymes.
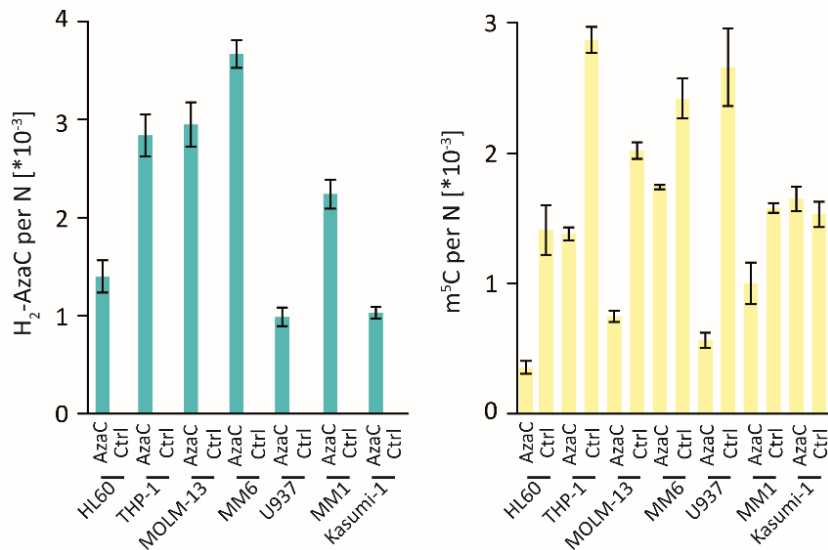
Figure 47: Screening of various AML cell lines for AzaC incorporation into RNA and the effect on m^5C; all data represent one biological replicate measured in technical triplicates; light green bars: H_2-AzaC, light yellow bars: m^5C.

We then analyzed the RNA of these cells (see **Figure 47**). Due to technical problems, data for KG1α, NB4 and MV411 are not available. As in the DNA, the cell lines express rather inhomogeneous

integration of AzaC into the RNA. The highest incorporation is observed in MM6 and amounts four-fold that of the lesser-incorporating cells. Lowest levels occur in HL60, U937 and Kasumi-1 cells. This is rather interesting, since HL60 exhibits highest sensitivity towards the drug, whereas Kasumi-1 cells seem to be most resistant (see next section **3.2.4.1.3**).

Addressing the impact on m^5C levels, however, suggests an explanation for this circumstance, since HL60 and U937 cells are affected very strongly by the drug (decrease to ~25%), whereas Kasumi-1 cells do not show any difference in the m^5C levels. The sister cell lines MM1 and MM6 exhibit different levels of AzaC - MM6 integrates 1.7-fold more than MM1 -, but show only moderate reduction of their m^5C levels (by ~30% and ~40% respectively) in comparison with strongly affected cells. THP-1 and MOLM-13 cells both exhibit higher incorporation of AzaC, but moderate reductions of the m^5C levels to 40-50% of the untreated control.

These results are comprised of only one biological replicate, in which *Laura Bocci* unfortunately performed the cell harvest in a way that was probably not optimal for cells treated with azacytidines. In detail, she froze the cells slowly in vials containing freezing medium and then placed them in an isopropanol-containing storage box in the –80 °C freezer rather than shock freezing the pelleted cells. This procedure unfortunately keeps the cells longer in a liquid environment leading to a higher probability for hydrolysis of the compound. Although the high levels for H_2-AzadC and H_2-AzaC might indicate that the results were not compromised, further experiments needed to be performed to verify our findings.

3.2.4.1.3 IC_{50} studies on the treatment of AML cell lines with AzaC and AzadC

Next, *Laura Bocci* performed IC_{50} studies with all AML cell lines that were available and provided me with the results from this experiment. The IC_{50} is defined as the concentration at which a biological process is inhibited by half. Therefore, cells were treated in three biological replicates with various concentrations of the drug and the cells were counted after 72 h. Linear regression of all data points then gave the IC_{50} values, which are depicted in **Figure 48**.
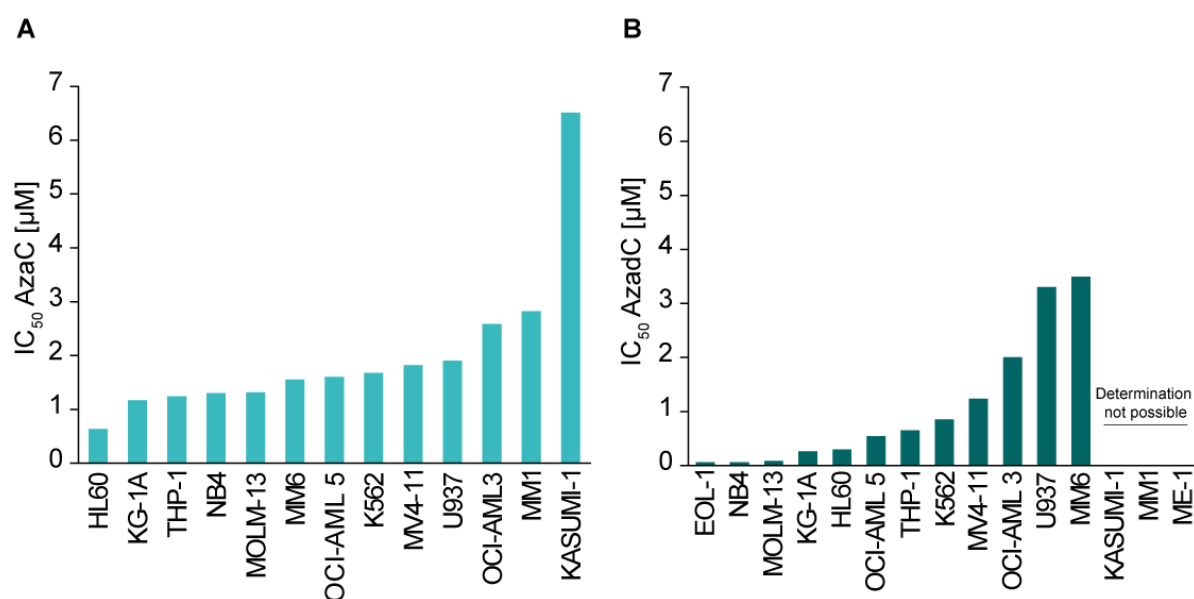
Figure 48: IC$_{50}$ values [µM] determined for treatment of various AML cell lines with AzaC (A) or AzadC (B); determination not possible means that the tested concentrations never lead to a reduction of the cell count by half after 72 h.

The results of the IC$_{50}$ determination suggest, that of the tested cancer cells none show hypersensitivity towards AzaC, but three cell lines (EOL-1, NB4 and MOLM-13) exhibit strongly reduced viability already when low concentrations of AzadC (0.1-0.2 µM) are administered. Furthermore, the response towards AzaC does not differ much between most of the cell lines (IC$_{50}$ values of 1-2 µM), whereas with AzadC the cells show IC$_{50}$ values in a broad range of concentrations. The highest IC$_{50}$ value for AzaC is found for the Kasumi-1 cells. Similarly, these cells do not even reach 50% reduction of the viability with AzadC. Comparable results were also observed for the MM1 cells. In contrast, HL60, NB4 and KG-1α cells show low IC$_{50}$ values for both drugs. Of all cell lines, MM6 seems to have the most different IC$_{50}$ values between the two azacytidines. This effect might be caused by differential expression of the DNA and RNA cytidine kinases DCK[170] and UCK,[171] two enzymes that have been shown to promote resistance towards azacytidines.

In a recent paper,[143] IC$_{50}$ values were determined for several cell lines. Of those, only two cell lines were the same (HL60 and U937) as in our experiments. The results for HL60 cells are comparable with our results. The sensitivity of U937 cells towards AzadC, however, is reported to be very high with a value between 0.05 and 0.4 µM, whereas our data show an IC$_{50}$ of ~3.5 µM. Laboratory-specific differences of cell lines in regards to *e.g.* speed of cell growth or resistances are not uncommon and their consideration is necessary to enable full interpretation of biological effects.

Based off these data, we concluded that the cell lines HL60 and MOLM-13 are generally more sensitive towards azacytidines, because they both exhibit significant cell death already with small doses of both drugs. In comparison, the cell lines MM1 and Kasumi-1 are generally more resistant towards the drugs.

This effect is so pronounced that for AzadC even with high doses above 8 µM the cell count could not be reduced to 50%.

3.2.4.1.4 Comparison of more sensitive and more resistant cell lines towards treatment with AzaC and AzadC in a time course experiment

We then decided to proceed with our analysis only with the four cell lines MOLM-13, HL60, Kasumi-1 and MM1. To achieve comparable results to the previous screening experiment and reduce dose-dependent effects, the cells were again treated with 0.5 µM AzadC and 3 µM AzaC, respectively, for 24 h. After incubation for 24 h and additionally after a total of 72 h, we harvested and analyzed the cells for their incorporation of AzadC and AzaC into the gDNA and RNA (see **Figure 49** and **Figure 50**).
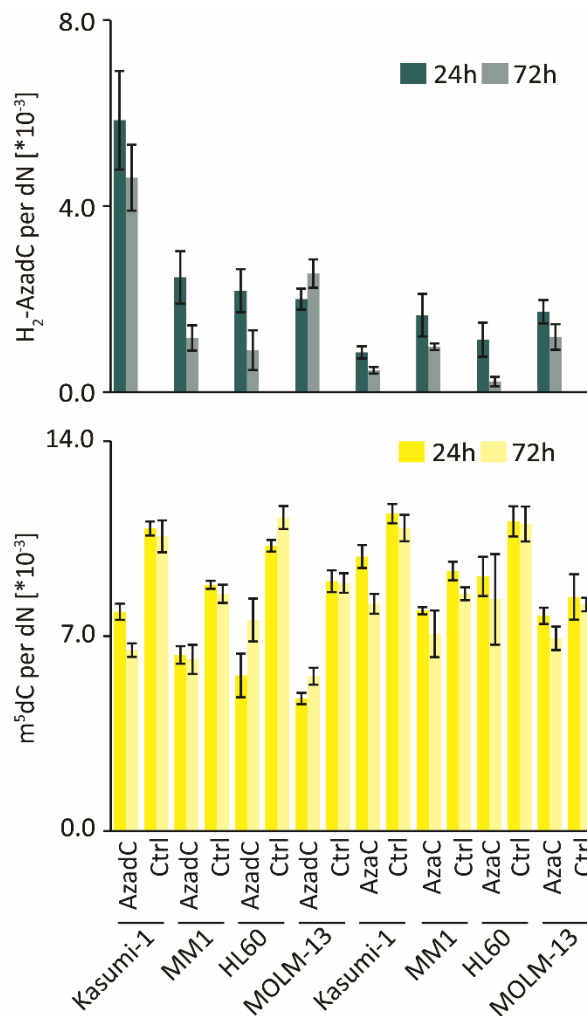
Figure 49: Overview over incorporation of AzadC into the gDNA of different cell lines. The upper part shows the levels of incorporated H_2-AzadC after treatment with 0.5 µM AzadC (left) or 3 µM AzaC (right) after 24 h (dark green) or 72 h (light green). The lower part of the figure pictures the resulting changes in the m^5dC levels at the same time points (dark yellow and light yellow, respectively).

Generally, it seems like treatment of the different cell lines with 0.5 µM AzadC leads to a higher incorporation of AzadC in comparison with treatment of six times this amount of AzaC (3 µM)

(see **Figure 49**). Due to the error bars however, the levels are only in Kasumi-1 cells significantly different. This effect is even more obvious, since the H_2-AzadC levels after treatment of AzadC are about double the H_2-AzadC levels in the other cell lines. In comparison, treatment with AzaC results only in about half the H_2-AzadC levels. Kasumi-1 therefore seems to incorporate AzadC easily upon administration of the DNA analogue and fails to incorporate high amounts of the RNA analogue. This observation suggests a reduced function of the ribonucleotide reductase enzyme or other enzymes in the process of generating AzadCTP in these cells.

The m^5dC levels are generally reduced upon treatment with the drug. This effect is more pronounced in the samples that incorporate more AzadC. Subsequently, it is not surprising that AzaC-treated samples show a smaller decrease of the m^5dC levels than the AzadC-treated cells. It is however remarkable that Kasumi-1 with the highest incorporation of AzadC after AzadC-treatment shows relatively lower impact on m^5dC (reduction to ~80%) than the cell lines MOLM-13 and HL60 (reduction to ~60%). The same reduction to ~80% of the levels in the untreated control is observed in MM1 cells. The higher resistance of these cells might therefore be explained by mechanisms to prevent the cells from losing too many of their methylation marks. These mechanisms might be higher expression/activity of DNMTs, more repair of the resulting lesions or a generally lower rate of replication and establishment of methylation on the newly-synthesized strand.

Analysis of the H_2-AzadC levels 72 h after AzadC-treatment, shows insignificant reduction in Kasumi-1 cells and insignificant elevation in MOLM-13 cells, whereas the levels drop by 50% in HL60 and MM1 cells. The m^5dC levels, which also marginally decrease in Kasumi-1 cells, are constant in MM1 cells and increase in HL60 cells and MOLM-13 cells. There does not seem to be a correlation to the changes in the AzadC content of the DNA. It is important to note the slight increase of the m^5dC levels in the MOLM-13 control sample, indicating that the levels might be affected from the long culturing of the cells rather than recovery from the treatment. However, in none of the cells significant differences were observed. This result suggests that more time is needed for the cells to recover from the loss of methylation after treatment. One possible reason for this might be the cytostatic effect of the drug, inhibiting cell proliferation and therefore maintenance methylation.

The data for the H_2-AzadC levels 72 h after AzaC-treatment generally show a reduction of the drug content. This observation is not surprising since the cells have a limited pool of the drug. Due to the sensitivity of the drug towards water, the azacytidine concentration in the medium will decrease over time. Furthermore, the DNA replication could be impaired by the cytostatic effects of the drugs leading to no further incorporation of the drug and DNA repair mechanisms then decrease the levels of the incorporated drug. As observed with the treatment of AzadC, the methylation levels do not show significant differences to the levels after 24 h. This suggests that re-establishment of the methylation is not directly affected by the choice of treatment. The relatively greater reduction of the H_2-AzadC

levels after 72 h in MM1 and HL60 cells could be explained by a more effective mechanism for repair of the lesion. In HL60 cells, the significant loss of the drug to ~20% of the initial levels after AzaC-treatment indicates additional biological pathways. One further explanation could be a reduction or loss of the RNR activity that has previously been described,[166] limiting the availability of the drug in addition to degradation by hydrolysis.
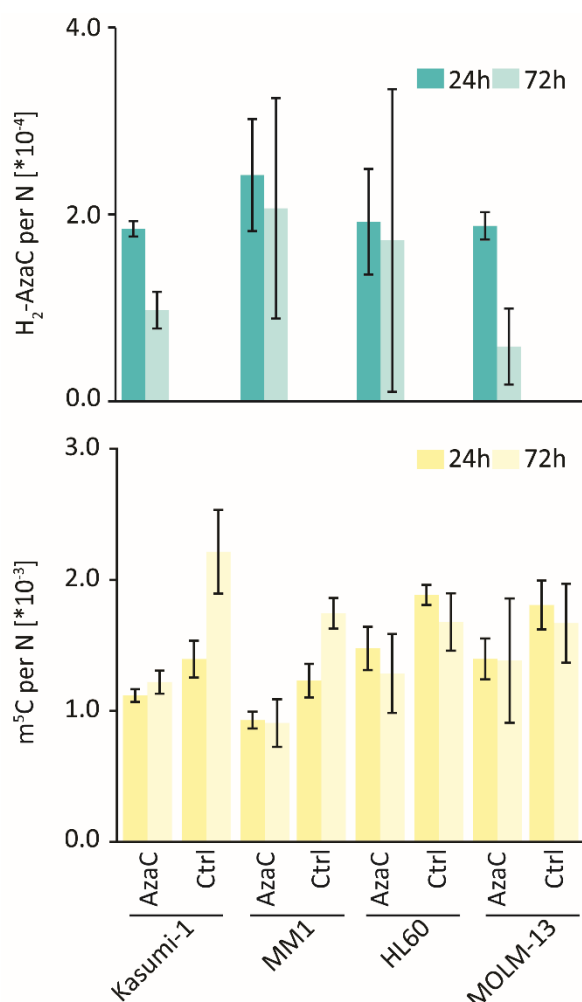
Figure 50: Overview over incorporation of AzaC into the RNA of different cell lines. The upper part shows the levels of incorporated H_2-AzaC after treatment with 3 µM AzaC after 24 h (blue) or 72 h (light blue). The lower part of the figure pictures the resulting changes in the m^5C levels at the same time points (yellow and light yellow, respectively).

Analysis of the RNA (see **Figure 50**) also seems to differ significantly from the results in **3.2.4.1.2**. The data exhibit roughly the same levels for H_2-AzaC in all cell lines rather than higher incorporation for MOLM-13 and MM1. The data after 72 h show very high inhomogeneity between the biological samples. We cannot exclude that this effect stems from technical problems. It is however clear, that in Kasumi-1 and MOLM-13 cells less H_2-AzaC is detected after 72 h. In these cells, some kind of removal process for RNA containing the nucleoside has to have taken place. Potential processes are normal RNA turnover or RNA quality control and decay by nucleases or the exosome, because RNA damage is most likely not repaired.

Interestingly, all cell lines show smaller effects on m^5C levels and greater error bars for the 72 h time points. In general, the m^5C levels upon AzaC-treatment do not differ between the 24 h and 72 h time point. If the loss of AzaC in RNA is a result from RNA turnover or decay, the m^5C integrity of the replaced RNA must be strictly controlled. This might hint at an installation by the NSUN methyltransferases and that these enzymes are not trapped by the incorporated nucleoside. Another explanation might be higher expression levels of these enzymes that can compensate for the fraction of enzymes that stays covalently attached to the RNA and has to be degraded. It seems like the cells have higher inhomogeneity of long-term effects on the RNA between biological samples, but there is no proportionality between the inhomogeneity of the AzaC levels and those of m^5C. Kasumi-1 control cells surprisingly exhibit different levels of m^5C between the two time points. This effect is also present in MM1 cells but less pronounced. A possible explanation could be a side effect of treating those cells with DMSO.

Overall, the analysis of different cell lines at various time points after treatment with AzadC or AzaC gives very complex results and can only hint at possible effects in samples of real patients. As described in current literature, the mode of action and resistance phenomena of and towards azacytidines is an important research field that needs more intense investigation.

3.2.4.1.5   Analysis of patient derived xenograft samples

To investigate the mechanism and effects of AzaC in a more advanced experimental set-up, we decided to investigate AML cells extracted from patients that were transplanted into mice. This patient derived xenograft (PDX) mouse model enables studies on individual cancer types without unnecessarily burdening the patient. The implanted cells will grow in the bone marrow of the mouse and cause AML. Therefore, growth behavior and disease phenotypes can be observed. Furthermore, treatment options can be studied in the mouse model and enable investigation of resistance phenomena. It is possible to extract the AML cells again and subsequently culture and study them *in vivo*.

This experiment was performed in cooperation with *Dr. Binje Vick* from the *Jeremias* laboratory at the Helmholtz Center Munich. The specimens derived from patient 393 (AML-393) and 372 (AML-372), respectively, were chosen according to previous mouse experiments of this laboratory, in which AML-393 appeared more resistant and AML-372 seemed more sensitive towards AzaC-treatment. Another sample from patient 485 (AML-485) was prepared in the mouse model specifically for our research, because the patient itself showed resistance towards AzaC-treatment. Two more specimens from patient 346 and 669 (AML-346 and AML-669, respectively) were selected randomly due to availability. After isolation of the PDX cells from the spleen or bone marrow, they were cultured with AzaC (1 µM) for various incubation lengths. The drug was administered every 24 h and the cells lyzed after 24 h, 48 h or 72 h.

A complete table of all provided samples is given in **Table SI-1**. Of these samples, analysis was only possible of some, because either the treatment was performed in the wrong manner or we had problems with the sample preparation (marked in blue in **Table SI-1**). However, data for incorporation of AzadC into DNA could be obtained from seven samples (# 1, 2, 3, 4, 11; 12, 13), and for incorporation of AzaC into RNA from twelve samples (# 1, 2, 3, 4, 7, 8, 10, 11, 12, 13, 14, 15) as shown in **Figure 51A and B** (in ascending sample #). The data presented are derived from single biological samples measured in technical triplicates. In most cases, one biological specimen equals one donor mouse and the isolated cells were treated with AzaC in a single experiment with a treatment length of 24 h. In some cases, however, more material from the same mouse was available and another time point could be taken. Analysis of the samples revealed incorporation of AzadC/AzaC in both DNA and RNA.
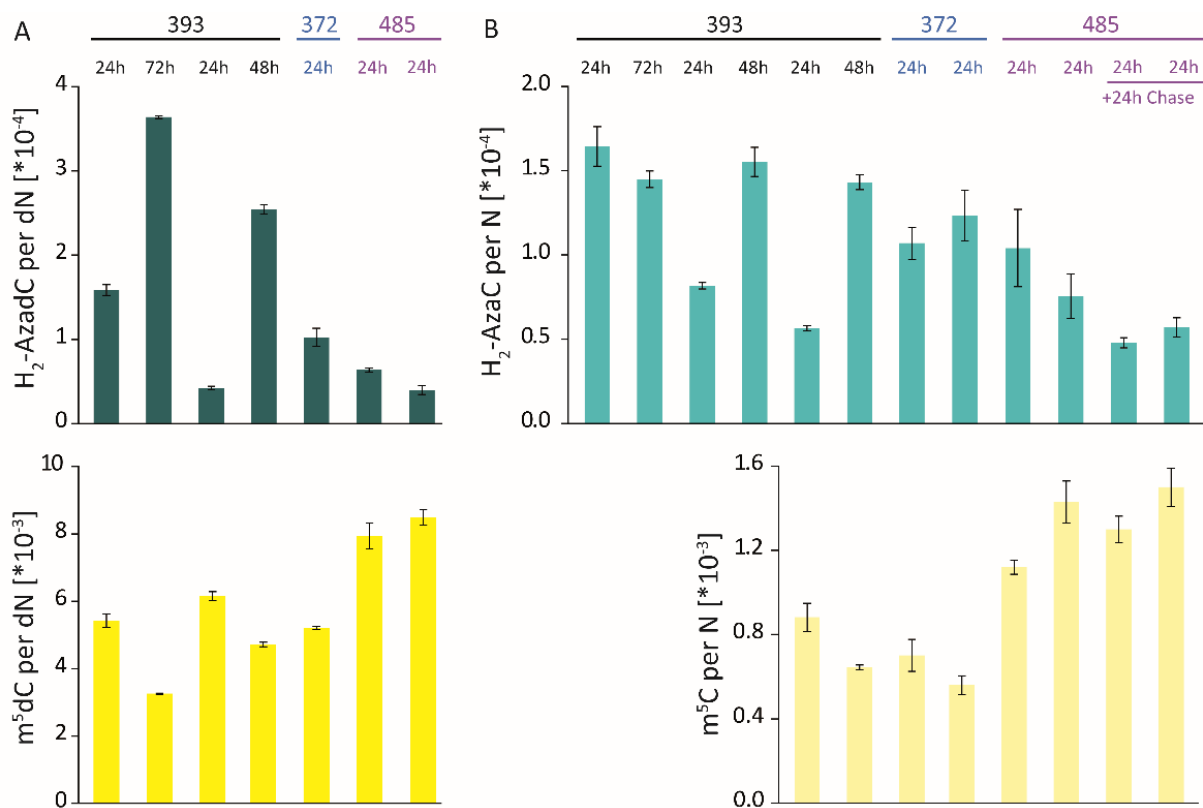
Figure 51: A) Incorporation of AzadC into DNA of AML PDX cells after AzaC-treatment with levels of H_2-AzadC (upper row) and the effect on m^5dC levels (lower row), B) Incorporation of AzaC into RNA of AML PDX cells after AzaC-treatment with levels of H_2-AzaC levels (upper row) and the effect on m^5C levels (lower row).

In general, longer treatment of PDX cells seems to lead to higher incorporation of AzadC into DNA and therefore to lower levels for m^5dC. The treatment of AML-372 only gave data for a 24 h treatment length. Compared with the mean of the respective time points for AML-393, the levels of H_2-AzadC are the same, but the m^5dC levels seem to be decreased slightly more. Samples from AML-485 showed the lowest incorporation of AzadC and also the highest levels of m^5dC. Considering the resistance of the patient towards AzaC-treatment, these results match the expectations. There must be some mechanism in these cells that prevents incorporation of the drug, *e.g.* expression levels of UCK. A

second explanation might be accelerated repair of the DNMT adducts, since generally the methylation levels appear to match levels from other somatic cells.[99] In one biological sample of AML-393 that after 24 h exhibits similar levels for administration of AzadC as those found in AML-485, the m^{5}dC levels are nevertheless reduced to a significantly greater extent. This raises the question, if AML-393 expresses generally lower methylation of the DNA – possibly due to decreased expression of the methyltransferases – or if the effect of the drug is stronger. Unfortunately, we did not analyze untreated control experiments.

Analysis of the RNA (see **Figure 51B**) gives a slightly different picture. The first four samples were only analyzed in regards to the levels for H$_2$-AzaC, because quantification of m^{5}C was not yet established. In the following analyses, m^{5}C was also addressed. In two experimental set-ups with AML-393 treated for 24 h, the levels for H$_2$-AzaC are lower than for longer treatment lengths. In one experiment they were, however, the highest compared with all other treatment lengths and cell lines. The mean of this cell line and time point is, nevertheless, comparable with AML-372 and AML-485. The two 48 h time points for AML-393 show comparable results for the incorporation of AzaC, although the extraction of the cells was performed once from the spleen and once from the bone marrow. This demonstrates reproducibility of the data, even though the extraction from the bone marrow showed significantly less homogeneous cell preparations. For AML-485, two experiments with 24 h treatment, and two experiments with 24 h treatment plus 24 h of incubation without addition of fresh AzaC (chase) were performed. We observe that the H$_2$-AzaC levels are about 25% lower for the samples with the chase. This difference might be due to the relatively high turnover rate of most RNA and reason longer treatment periods in the clinic. In total however, this specimen shows 30-60% higher levels for m^{5}C in comparison with AML-393 and AML-372 and raises the same questions about the general methylation levels of this sample as in DNA. Another explanation, like the quality of the administered pharmaceutical, can be neglected, because in AML-372, the levels of H$_2$-AzaC are comparable, but the m^{5}C levels are significantly lower. It is possible that this sample also has an overall lower methylation level, but we were not able to determine this without an untreated control.

As a conclusion, the effects of AzaC on different biological samples can differ significantly. This result is not surprising and matches various publications investigating resistance phenomena in patients. Our UHPLC-QQQ-MS method is able to determine levels of incorporation of the pharmaceuticals and the effects on the respective methylation levels in a simple protocol. This enables fast routine analysis of patient samples and can help predict treatment outcome. Monitoring of the incorporation of the drug into the (deoxy)ribonucleic acid might identify resistances correlated with DNMT enzyme, as well as DCK[170] and UCK[171] expression levels.

3.2.4.2 Investigation of deformylation via 6-Aza-2'-deoxycytidine derivatives

Active demethylation has been proposed to occur in a variety of different ways. Recent investigation of 2'-fluorinated dC derivatives and heavy labeled dC derivatives suggested deformylation as one possible pathway to eliminate m^5dC from the gDNA.[179] Further research in our group (data not shown) additionally hinted at involvement of DNMT3b and potentially the NAD^+ dependent Sirtuin histone deacetylases III (SIRT) in this process. Since SIRT proteins could attack nucleophilically in the C6-position (**Scheme 5**), we thought of a chemical way to inhibit this reaction.
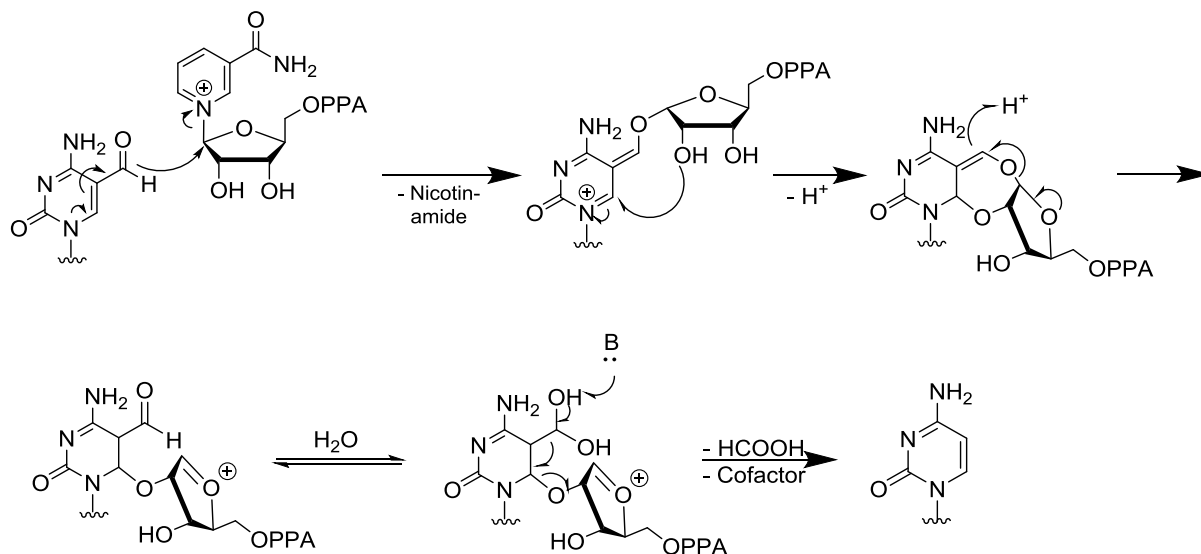
Scheme 5: Possible deformylation mechanism of fdC under catalysis of NAD^+ dependent SIRT proteins.

To this end, a special AzadC compound with a nitrogen in the 6-position, from now on abbreviated as 6-Aza-dC, and its methylated and oxidized derivatives, 6-Aza-m^5dC, 6-Aza-hmdC and 6-Aza-fdC (see **Figure 52**), were synthesized by *Alexander Schön*.
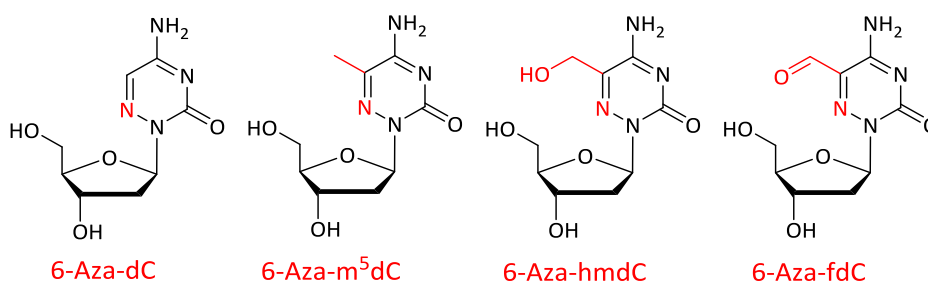
Figure 52: Structures of the 6-Aza compounds synthesized by *Alexander Schön*. Marked in red are the N atoms in position 6 and the functional groups on C5.

The compounds are thought to be taken up into cells and incorporated into their genome as dC derivatives. Administration of 6-Aza-dC to cell cultures is supposed to disable the DNMT enzymes, since their mode of action involves nucleophilic attack on the 6-position. Feeding of 6-Aza-m^5dC can give insight into the selectivity of TET enzymes, and 6-Aza-fdC might help uncover a mechanism for a

deformylation reaction. If deformylation is still detected on the compound, resulting in formation of 6-Aza-dC from 6-Aza-fdC, this provides evidence for a deformylation process that does not involve nucleophilic reactions on the 6-position and *vice versa*. Additionally, in contrast to regular fdC, this fdC derivative can easily form a hydrate in the presence of water. Since the hydrated fdC compound might be an intermediate of the deformylation reaction, a discovery of substantial amounts of 6-Aza-dC might still provide information on the mechanism of deformylation.

3.2.4.3.1 Method development for UHPLC-QQQ-MS/MS

For investigation of the compounds, we first had to develop a suitable UHPLC-QQQ-MS method. The method development included determination of the fragmentation products, *CE* optimization for the fragmentation and adjustment of the UHPLC gradient. During the fragmentation studies, a cleavage of the N-glycosidic bond, which is typical for nucleosides, was confirmed (see **Scheme 6**). For 6-Aza-fdC, an additional species – the hydrated nucleoside – was observed, of which also the free base was the most prominent fragment. Corresponding *CE* optimization was then easily achieved with the standard procedure for method development.
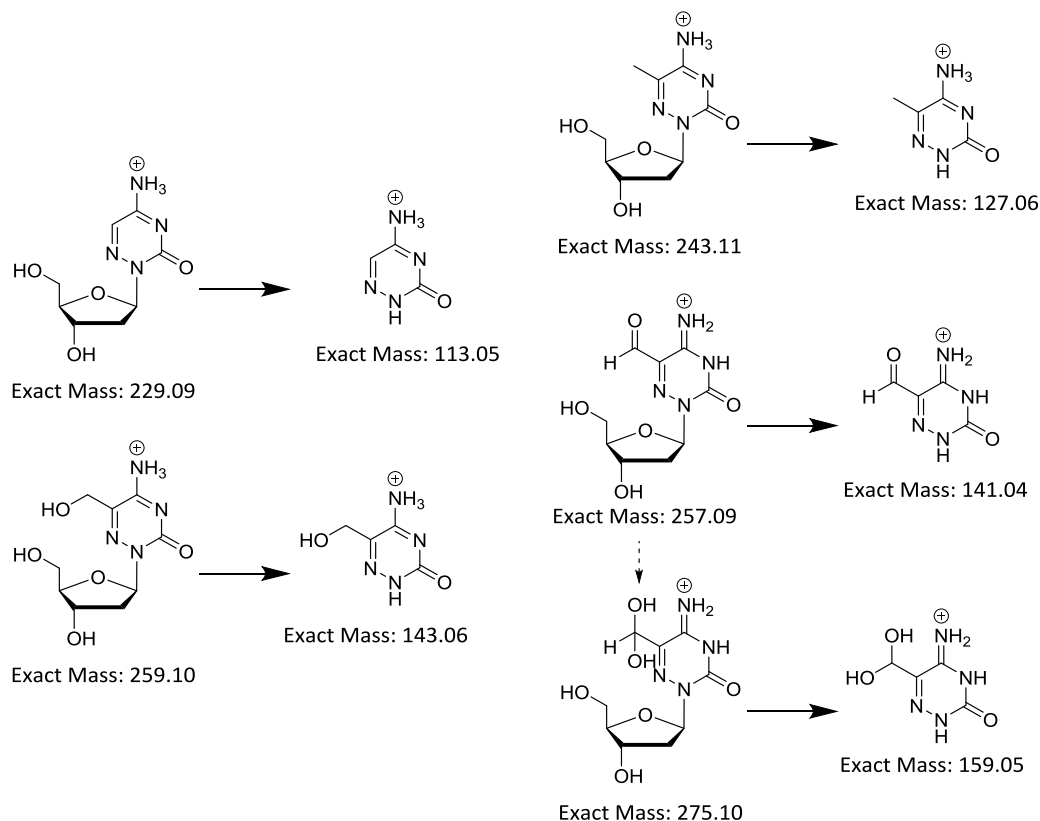
Scheme 6: Fragmentation patterns for the 6-Aza nucleosides.

Chromatographic analysis of the nucleosides, however, was challenging in case of 6-Aza-fdC because of its hydrate formation. The pure nucleoside and its hydrate have different chromatographic properties. Since hydrate formation is, however, an equilibrium reaction, constant shifting of the

nucleoside between its two forms resulted in elution of both precursor ions in a broad peak spanning both retention times (see **Figure 53**). Consequently, the MS signals for both species exhibit the same elution pattern.
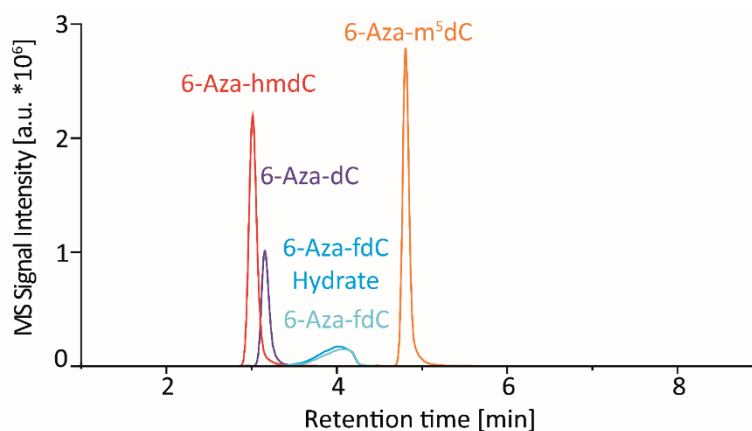
Figure 53: UHPLC-MS/MS chromatogram overlay for the fragmentation of all 6-Aza nucleosides in their respective MRM.

3.2.4.3.2   Administration of the 6-Aza-dC derivatives to wt mESCs

The purified compounds were administered to wt K3$^{+/+}$ mESCs by *Angie Kirchner* and *Ewelina Kaminska* and subsequently the DNA isolated. After total enzymatic digest, the nucleosides were analyzed with UHPLC-QQQ-MS/MS.

First results showed an incorporation of 6-Aza-fdC into the DNA upon administration of 6-Aza-fdC, but the absolute levels were below the limit of quantification. The underlying cause might be the broadly eluting peak of this nucleoside and therefore increased background. Importantly, 6-Aza-dC was not found in the same sample, which might indicate lack of a nucleophilic attack on the 6-position of the nucleoside, and therefore impaired deformylation reaction. Administration of the other 6-Aza-dC derivatives did not even result in detection of a peak and the nucleosides hence seem to not be incorporated into DNA.

We performed these experiments three times with the same outcome.

Simultaneously, we thought about analyzing the DNA in a full scan *via* high resolution MS to exclude deamination of the administered nucleosides and therefore formation of 6-Aza-dU derivatives. *Dr. Mirko Wagner* performed the analysis on the Orbitrap MS and could not confirm the incorporation of 6-Aza-fdC, which might be due to the lower sensitivity of the machine. However, the measurements reproduced the lack of incorporation of the other supplemented nucleosides or corresponding oxidation products. Only upon administration of 6-Aza-m^{5}dC we observed a yet to be validated peak

(**Figure 54**) with the mass-to-charge ratio (m/z) of 244.0938, which is likely to be 6-Aza-dT (calc. [M+H]$^+$: 244.0928). No other 6-Aza-dU derivatives were found in any of the other experiments.
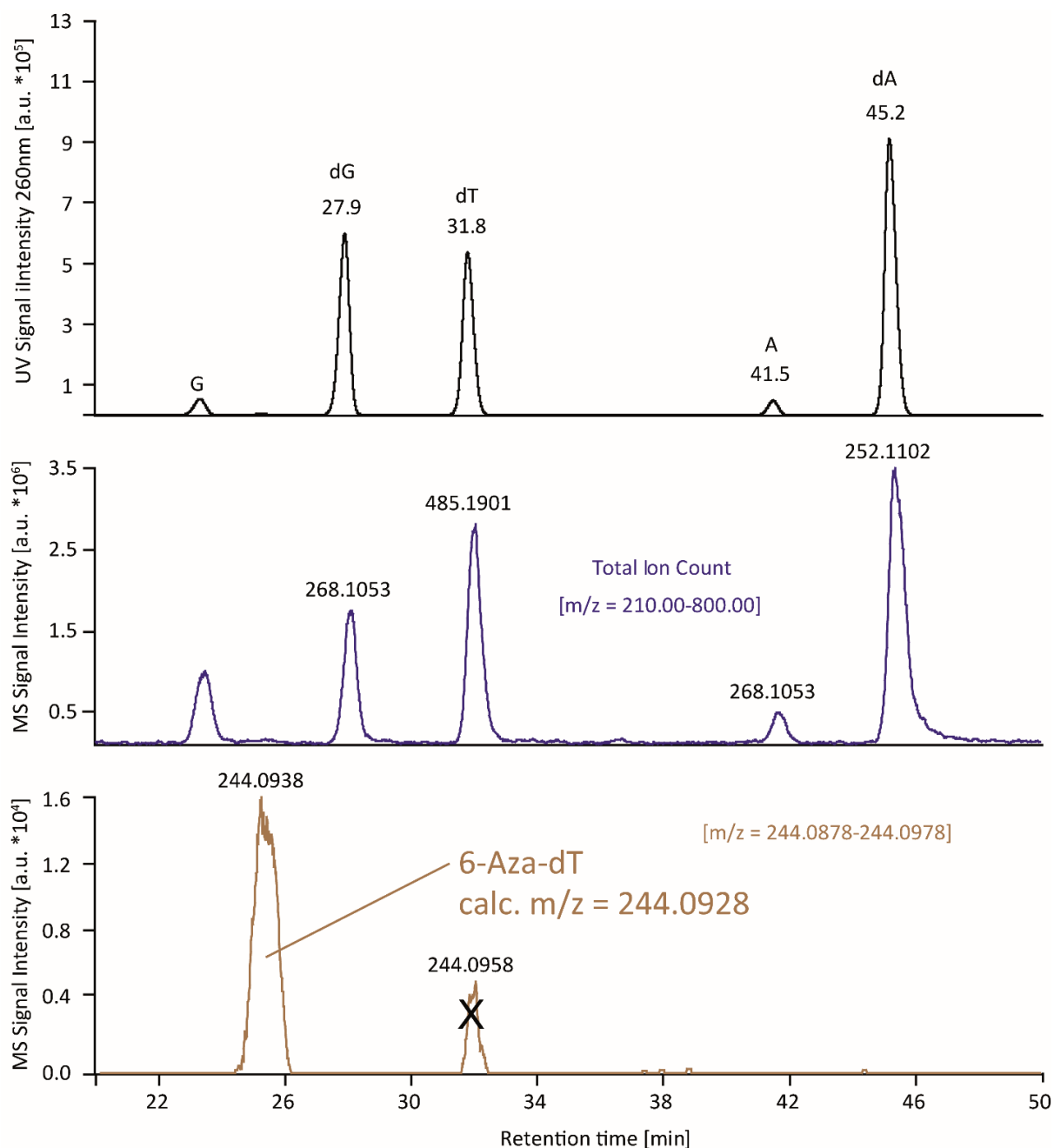
Figure 54: UV- and MS-Chromatogram of DNA analyzed on an Orbitrap mass spectrometer.

We conclude that the nucleosides are either not taken up by the cells or not as readily accepted by enzymes involved in nucleotide or DNA biosynthesis. A third explanation might be recognition of these nucleotides as lesions and an efficient repair process. It is however also possible, that the digestion protocol applied in the experiment is not suitable for efficient cleavage of the DNA into nucleosides.

3.2.4.3.3   Administration of 6-Aza-dC derivatives to somatic cells

Next, we decided to investigate other cell lines. ESCs might not incorporate a sufficient amount of the 6-Aza-fdC nucleoside to observe a deformylation reaction, but other cell lines might be able to accumulate more of the nucleoside in their DNA. To this end, *Ewelina Kaminska* administered the nucleoside to Neuro-2a, CHO-K1 and RBL-2H3 cells.

Indeed, when we analyzed the DNA, we found the 6-Aza-fdC nucleoside in two-digit fmol amounts in all three cell lines, with CHO-K1 cells incorporating the highest amount. Using an external calibration curve, an estimated quantification was possible, but calculation of the real amount was impaired due to the occurrence of the molecule in two different species. Derivatization of the incorporated nucleotide could overcome this problem. We however again did not observe any incorporation of 6-Aza-dC upon administration of the 6-Aza-dC or formation of the nucleotide upon administration of 6-Aza-fdC. Although the levels of 6-Aza-fdC are still low in these somatic cells, our finding hints at impaired C-C bond cleavage on this molecule, where nucleophilic attack at position 6 cannot occur.

The experiment needs to be repeated in more biological replicates to verify our findings.

3.2.4.3.4   Derivatization of 6-Aza-fdC with methoxyamine

To achieve peak sharpening of the broadly in two species eluting 6-Aza-fdC nucleoside, *Alexander Schön* developed a derivatization protocol to establish a Schiff-base of methoxyamine on the formyl group of the nucleoside. Indeed, method development on the UHPLC-QQQ-MS was successful with the positively charged nucleoside as the precursor and the positively charged base as the product ion (see **Figure 55A**). Chromatographic separation (**Figure 55B**) revealed two sharp peaks for the E/Z-isomers, which are furthermore shifted by several minutes in comparison with the underivatized nucleoside.
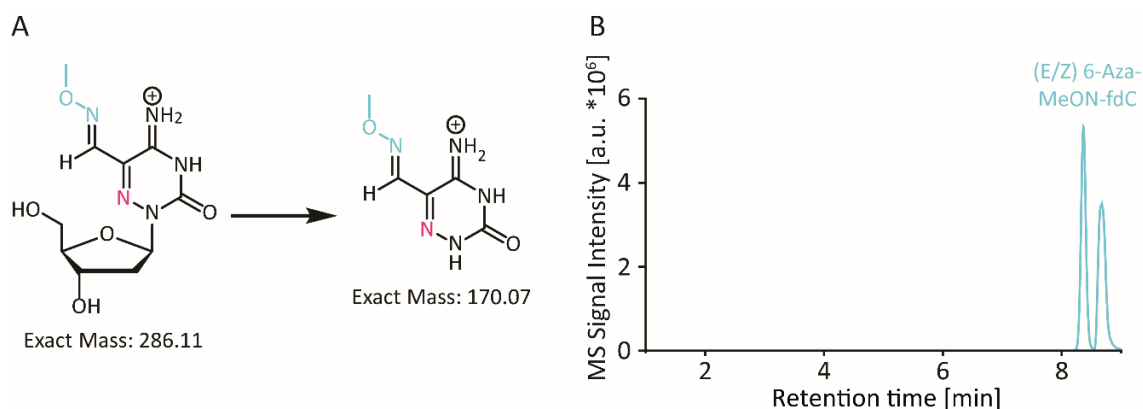
Figure 55: A) Structure of the precursor ion and product ion for the methoxyamine-derivatized 6-Aza-fdC B) UHPLC-MS/MS chromatogram for the fragmentation reaction of the methoxyamine derivatized 6-Aza-fdC nucleoside. The two peaks correspond to the E- and the Z-isomer.

In a first test, I applied these derivatization conditions to the digested and dried nucleoside mixture. To this end, I dissolved the pellet in NaOH (500 μL, pH=10) and added 100 fmol methoxyamine. After

an incubation time of 15 min at 900 rpm in a ThermoShaker (*Eppendorf*), the reaction was brought to pH=7 with formic acid (200 µL, pH=3) and the solvent was removed over three days using the lyophylle and SpeedVac. After resuspension in ultra-pure water, the samples were filtered utilizing a 0.2 µm Supor filtration plate (*Pall Corporation*) and subjected to UHPLC-QQQ-MS/MS.

Unfortunately, with the first test we neither detected free, nor derivatized 6-Aza-fdC in the samples. This result might stem from the instability of the nucleoside in basic conditions as observed by *Alexander Schön* during prolonged reaction of the free nucleoside. The long drying process therefore might have caused a degradation of the product. Other explanations might be higher LOD of the derivatized nucleoside or incomplete derivatization, which decreased both molecules to amounts below LOD. In case of incomplete derivatization, the reaction on the DNA has to be optimized. Due to significant cell death and low incorporation of the nucleoside into the DNA, optimization on DNA directly requires a large amount of input material and provides potentially limited material for a screening procedure. Therefore, optimization of the protocol on artificial strands with known amounts of the nucleotide could be an alternative strategy. The synthesis of the strands would however require production of the corresponding phosphoramidite building blocks.

As a conclusion, derivatization of the 6-Aza-fdC nucleotide is a suitable strategy to achieve elution of the nucleoside as one sharp peak and therefore decrease the loss of sensitivity due to background signals. It is however important to perform the derivatization protocol as fast as possible and to neutralize the reaction mixture in a timely manner to avoid degradation of the product. Once the derivatization procedure is established, determination of the absolute levels of 6-Aza-fdC should be possible.

3.2.5 Analysis of modifications in RNA

3.2.5.1 i^6A, ms^2i^6A, t^6A

RNA modifications are important for the stability, secondary structure and interactions of the nucleic acids with each other or with proteins, *e.g.* in processes like RNA interference or translation. In cooperation with *Dr. Noelia Fradejas-Villar* from the group of *Prof. Schweizer,* we investigated the levels of i^6A and ms^2i^6A in tRNA isolated from TRIT1 *knockout* mice. TRIT1 is the enzyme that is responsible for the transfer of an isopentenyl group onto adenosine. Both modifications are known to have a stabilizing effect in the interaction of codon and anticodon in position A37. We wanted to determine whether a *knockout* of the enzyme was really achieved, and whether isopentenylated A is still present in the tRNA. This could indicate the existence of an additional 'writer' enzyme. Furthermore, we were interested in the analysis of t^6A, because the mice showed a mitochondrial effect and this modification is known to be synthesized in the mitochondria by OSGEPL1.[416]

The method for this analysis had previously been developed by *Katharina Iwan*. However, during this experiment, problems with the quantification of ms^2i^6A arose: although [D_3]-ms^2i^6A was added to the digestion mixture as usual with the established protocol, the MS signal for the corresponding fragmentation could not be observed. After trying different digestion set-ups with more or less RNA or enzymes, I finally found the solution for this problem by adding 1% formic acid to the digestion mixture after the digest was completed. Thiomethylated modifications are very sensitive to pH changes and have to be analyzed in slightly acidic conditions to ensure exact quantification.

The isotope standard for t^6A ([$^{13}C_4^{15}N$]-t^6A) had already been synthesized previously and used for quantification with high resolution MS, but the method for the UHPLC-QQQ-MS was not fully developed. *Katharina Iwan* had however optimized the parameters for mass spectrometry. The valid internal calibration curve was then produced by myself to perform exact quantification.
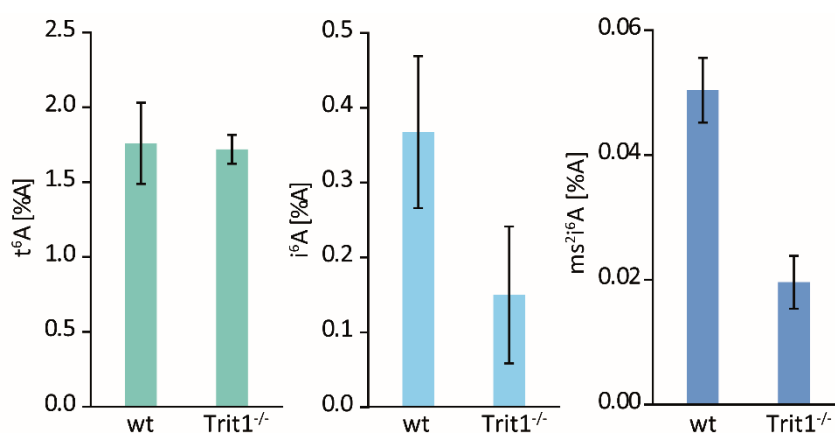
Figure 56: Quantification results of t^6A, i^6A and ms^2i^6A in wt and TRIT1 *knockdown* mice. The levels for the modifications t^6A (left), i^6A (middle) and ms^2i^6A (right) are given as percentage of A.

The results of the quantification as a percentage of A are displayed in **Figure 56**. The t^6A level does not seem to differ in the tRNA of the TRIT1 *knockout* samples in comparison with the wt, leading to the conclusion that the observation of a mitochondrial effect does not correlate with modification levels in tRNA. The levels for i^6A and ms^2i^6A are significantly reduced by more than 50%. This confirms a *knockdown* of the enzyme, but not a complete *knockout* since the levels are still clearly quantifiable. It is however possible that the detection of these nucleosides gives evidence for additional, yet unknown enzymes that install isopentenyl and thiomethyl groups on tRNA in mice.

3.2.5.2 ms^2A, ms^2m^6A, ms^2t^6A

As already described in **3.2.5.1**, isopentenylated RNA nucleosides are known to be thiomethylated. This raised the question, whether enzymes exist, which establish thiomethylation of unmodified adenosines. The formed ms^2A could then further be modified by methylation leading to formation of ms^2m^6A, or isopentenylation/threonylation generating ms^2i^6A/ms^2t^6A as an alternative biosynthetic pathway for these modifications. Thiomethylated A might furthermore occur as an intermediate of modification removal processes.

In cooperation with *Matthias Q. Kurz* and *Timm Ensfelder*, analysis of various RNA was performed in regards to these modifications and potential 'writer', 'reader' and 'eraser' enzymes. *Matthias Q. Kurz* synthesized the ms^2A and $[D_3]$-ms^2A nucleosides. Utilizing those reference molecules, I developed a suitable UHPLC-QQQ-MS method. Like most other nucleosides, ms^2A and its derivatives fragment by cleavage of the N-glycosidic bond. The mass transitions for ms^2A were *m/z* 314.1 → *m/z* 182.1 and for $[D_3]$-ms^2A *m/z* 317.1 → *m/z* 185.1.

Subsequently, *Timm Ensfelder* prepared RNA from *E. coli* strains. For digestion of this RNA, I first applied a previous RNA digest protocol from *Katharina Iwan*,[417] but the nucleosides showed weak and inconsistent signals in the MS. I attributed this observation to decreased solubility of the nucleosides. Before addition of the standards to the digestion mixture, I therefore equilibrated the respective

solutions at 37 °C to ensure a homogeneous solution. Since this enhanced the MS signal, but the signal for a specific administered amount was not stable between the samples, I performed the digestion according to a different protocol.[161]

Since we were interested in ms^2A as a potential intermediate of ms^2i^6A biosynthesis or an isopentenyl group removing process, we analyzed *E. coli* wt and MiaA and MiaB KO strains. Clear signals for ms^2A were found in the wt, but only weak signals below LOD were detected in the KO. This observation suggests formation of the nucleoside after deisopentenylation of ms^2i^6A. It is, however, not possible to exclude, that traces of the nucleoside exist which are not exceeding LOD. Therefore, ms^2A as a precursor for ms^2i^6A might exist, but this biosynthetic pathway would be minor.

Further experiments were evaluated by *Dr. Mirko Wagner*. Of note and as a final conclusion (see section **3.2.5.1**), modifications containing thiomethyl groups on A require more acidic pH. Future analyses should consider these observations.

4. Outlook

In the course of this Ph.D. thesis, several UHPLC-MS/MS methods have been developed, which could in principle be used or optimized for a variety of further investigations. Mass spectrometric evaluation of RNA, specifically, poses many options from analyzing total RNA in a variety of organisms, investigation of specific enriched RNA for selected modifications, or identification of interacting proteins.

The azacytidine derivatization and nucleoside analysis methodology has been studied intensively in cancer cell culture and some additional measurements were conducted in patient-derived xenograft samples. As a next step, this strategy could be applied to blood samples taken from Aza(d)C-treated patients to monitor the clinical effect of the drug. Naturally, combination therapies with other chemotherapeutics can also be analyzed for their drug incorporation and methylation level changes. Furthermore, a more thorough investigation of the cAzadC compound should be conducted. This compound is not used in clinical trials and might prove useful as a milder and more cost-effective alternative to AzadC.

The 6-Aza-dC compounds in combination with other experiments performed in our group, which are not described in this research, gave valuable insights into a possible deformylation mechanism of fdC in gDNA and suggested an involvement of DNMT3b and potentially SIRT proteins. In the future, the enzymes have to be verified for their role in this process and *in vitro* assays with the purified enzymes and 6-Aza-fdC and fdC containing strands in parallel could prove the deformylation mechanism over the 6-position. To this end, a synthesis of the respective phosphoramidites and DNA strands would be necessary.

The field of the repair of dC derivatives offers a great variety of further opportunities for investigation. An optimized analysis protocol for formylcytosine in the cytoplasm of cells might give valuable insights into base excision repair. This repair pathway could have a specialized role depending on the state of the cells. As a next step towards the method development, an isotope standard needs to be synthesized, which will consist of formylcytosine and a $[D_9]$-labeled Superfly reagent moiety. It will then be used to safely identify the adduct from the soluble pool and enable absolute quantification *via* the isotope dilution technique. Once the method is established on wt mESCs, investigation of TDG$^{-/-}$ and catalytically inactive TDG mutant cells will be performed.

Finally, the investigation of active demethylation of m^{5}dC *via* a pathway that involves formation of dT will continue in regards to finding the genomic deaminase. As a prime candidate, we suspect the APOBEC3A enzyme. Generation of a CRISPR/Cas-directed *knockout* cell line as a single KO, and in combination with the KO of AID will be useful to investigate the role of this enzyme. Once the cell line is established, mass spectrometric analysis of naïve and primed cultures supplemented with

[^{13}C,D$_3$]-methionine can then identify a role of this enzyme in genomic deamination. Furthermore, we want to sequence some of the previously generated and mass spectrometrically analyzed samples in respect to genomic imprints.

5. Experimental

5.1 Materials

All buffers and solutions were, if not stated otherwise, produced from ultra-pure water. This was obtained from the water filtration device arium® pro-DI from the company *Sartorius* with 18.2 MΩ/cm at 25 °C. Chemicals were purchased from the companies *Sigma*, *Merck*, *AppliChem, Fluka* and *Roth* in the commercially available quality grades *p.a., molecular biology grade, LC-MS grade* and *Chromasolv*®.

5.1.1 Devices

Agarose Gel chamber	Mini Sub-Cell GT MINI, *Bio-Rad* (Munich)
Analyzer	6490 Triple Quadrupole LC-MS system, *Agilent Technologies* (Santa Clara, CA)
Autoclave	Vakulab S3000, *Systec* (Gießen)
Visualization chamber Agarose	E-BOX VX5, *Vilber Lourmat* (Eberhardzell)
CO_2-Incubator	Hera Cell 150, *Heraeus* (Hanau)
Liquid chromatograph	1290 Infinity II LC System, *Agilent Technologies* (Santa Clara, CA)
Homogenizer	TissueLyser, *Qiagen* (Hilden) Bead mill MM 200 *Retsch* (Haan)
Microscope	EVOS FL Cell Imaging System (*Life technologies*)
NanoDrop	ND-1000 UV/VIS, *peqlab* (Erlangen)
pH-meter	MP220, *METTLER TOLEDO* (Schwerzenbach)
Rotor	SORVALL SS-34, *Thermo Electron Corporation* (Dreieich)
Thermoshaker	Thermomixer comfort, *eppendorf* (Hamburg)
Deep freezer -80 °C	VIP Series -86 °C, *SANYO* (Bad Nenndorf)
Vortexer	Vortex Schüttler, *VWR* (Göttingen)
Water filtration device	arium®pro H_2Opro-DI-D, *Sartorius* (Göttingen)
Centrifuge	Eppendorf Centrifuge Typ 5810 R, *eppendorf* (Hamburg)
Centrifuge	Eppendorf Centrifuge Typ 5424 R, *eppendorf* (Hamburg)
Centrifuge	SORVALL Evolution RC, *Thermo Electron Corporation* (Dreieich)
Centrifuge	MiniSpin*, Eppendorf* (Hamburg)

5.1.2 Buffers, Media, Solutions

RLT[+]	RLT buffer (*Qiagen*) β-mercaptoethanol (14.3 mM)) 3,5-di-*tert*-butyl-4-hydroxytoluene (BHT, 400 µM) Desferoxamine mesylate salt (Desferal (DF), 400 µM)
GTC-Citrate[+] (pH=6.9)	Guanidinium Thiocyanate (3.5 M) Sodium citrate (25 mM) DF (400 µM) BHT (400 µM) β-mercaptoethanol (14.3 mM)
G2[+]	G2 buffer (*Qiagen*) DF (400 µM) BHT (400 µM)
Serum-containing medium for culture of naïve mESCs	DMEM (high glucose, *PAA* or *Sigma*)) FBS (10%) β-mercaptoethanol (0.1 mM) 1×MEM-nonessential amino acids (NEAA, 1x, *PAA*) L-Alanyl-L-Glutamine (2 mM, *PAA*) Pen/Strep (1x, *PAA*)
For **2i/L** conditions add:	PD 0325901 (1 µM, *Axon MedChem*) CHIR 99021 (3 µM, *Axon MedChem*) Mouse recombinant LIF (1000 U/mL, *ORF Genetics*)
For **alternative 2i/L** (**a2i/L**) conditions add:	CGP 77675 (1.5 µM, *Axon MedChem*) CHIR 99021 (3 µM, *Axon MedChem*) Mouse recombinant LIF (1000 U/mL, *ORF Genetics*)
DPBS	Dulbecco's phosphate buffered saline (*Sigma*)
HBSS	Hank's balanced salt solution with Ca^{2+} and Mg^{2+}, but without $NaHCO_3$ (*Sigma*)

5.2 Biochemical Methods

5.2.1 Methods for the investigation of m[6]dA as a modification in gDNA

Cell culture of wt J1 and Kindlin3[+/+] mESCs

Feeder independent wt J1 cells (strain 129S4/SvJae)[418] or Kindlin3[+/+][387] cells were cultured in the presence of serum and LIF as previously described.[177] They were routinely maintained on gelatinized plates in 2i/L or a2i/L medium. For *priming* experiments, 2i/a2i cultures were passaged when applicable in DMEM supplemented with FBS and LIF as above but lacking the inhibitors.

Cell culture of HeLa and HEK293T cells

HeLa and HEK293T cells were cultivated at 37 °C in water saturated, CO_2-enriched (5%) atmosphere. DMEM (10% FBS) was used as growing medium. When reaching a confluence of 70% to 80% the cells were passaged or lyzed, respectively.

Administration of m⁶A

Administration of m^{6}A

HeLa and HEK cells were supplemented with 100 µM m^{6}A and 2% DMSO and incubated for 4 d. Feeder independent wt J1 cells were supplemented with 100 µM m^{6}A and 0.25% DMSO for 4 d, with splitting after 2 d, in FBS/LIF *priming* conditions. All cells were lyzed after 4 d with RLT$^+$ buffer and stored at -80 °C until further gDNA isolation.

Transfection of m^{6}A-containing strands into mESCs

The following strand containing m^{6}A was synthesized by *Matthias Kurz*:

5'-UGm^{6}ACCGm^{6}AUGGm^{6}AGGUm^{6}AGUm^{6}AUm^{6}AG-3'
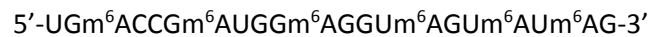
As a control, the strand containing only A instead of m^{6}A was synthesized. The counter strand contained a 6-FAM-tag to monitor the fluorescence at 509 nm to estimate the transfection efficiency. After desalting on an MF-Millipore™ cellulose acetate and cellulose nitrate membrane filter (0.025 µm VSWP, 25 mm diameter), the concentration of the strand was adjusted to 10 µM.

For transfection, six times 1x10^6 wt J1 cells were plated into gelatin-coated 6-wells in 2i/L medium and incubated at 37 °C and 5% CO$_2$ for 24 h. Then, the Lipofectamine RNAiMax (*Invitrogen*) reagent was used as described in the manufacturer's instructions. In brief, the lipid mixture was produced by adding RNAiMax to Opti-MEM medium and the RNA (containing A or m^{6}A, respectively) was diluted in Opti-MEM medium. Addition of the RNA mix to the lipid mix and incubation for 5 min at room temperature gave the complex mix that was added to the cells. An untransfected culture was used as a control. After 6 h, 12 h, 24 h and 48 h the medium was removed from the cells, they were washed with DPBS and HBSS was added. Then they were analyzed for fluorescence with an excitation filter of 470/22 nm and an emission filter at 510/42 nm using an EVOS FL Cell Imaging System (*Life technologies*). Subsequently, the cells were lyzed with GTC-Citrate$^+$ buffer and the gDNA isolated as described in a previous publication.[412]

Treatment of mESCs with all-trans retinoic acid

For treatment with *all-trans* retinoic acid (ATRA), heterozygous *knockin* mESCs derived from E14tg2a cells expressing YFP instead of the pluripotency marker OCT-4 (termed Oly2-1) were cultured in the presence of serum and LIF as previously described[177] and routinely maintained in 2i/L. Per gelatin-coated p60 or p100 plate, 9.2•10^5 cells, or 2.5•10^6 cells respectively were seeded in DMEM supplemented with FBS and LIF and either 0.1% DMSO or 100 nM ATRA. After different time points, the cells were washed with DPBS and analyzed for fluorescence with an excitation filter of 500/24 nm and an emission filter of 524/27 nm using an EVOS FL Cell Imaging System (*Life technologies*) in HBSS Subsequently, cells were lyzed in GTC-Citrate$^+$ buffer.

Treatment of wt mESCs with Trichostatin A

Wt Kindlin3[+/+] (K3[+/+]) cells were seeded in 2i/L medium in 4x gelatin-coated p100 plates and incubated at 5% CO_2 and 37 °C. After 24 h, 10 nM Trichostatin A (TSA) was added to the medium and the cells incubated for 16 h. Then, the medium was removed and the cells were washed with DPBS. One p100 was lyzed with GTC-Citrate[+] buffer and on the other plates fresh medium was added for subsequent lysis time points at 2 h, 4 h and 6 h.

In a different set-up, the cells were cultured in a2i/L medium with administration of 500 nM TSA for 16 h at 5% CO_2 and 37 °C. After washing with DPBS, an untreated control and one treated plate were lyzed with GTC-Citrate[+] and the other plate incubated with fresh medium for 6 h before lysis.

All lysates were stored at -20 °C until gDNA isolation.

Genomic DNA (gDNA) isolation

Cell culture samples were lyzed directly in the plates with GTC-Citrate[+] buffer. The gDNA was isolated using a standard protocol.[412]

Digest of gDNA

1 µg (10 µg) of gDNA in 35 µl H_2O were digested as described in a previous publication[374] using the nuclease S1 (*Aspergillus oryzae*, *Sigma-Aldrich*), Antarctic phosphatase (*New England BioLabs*) and snake venom phosphodiesterase I (*Crotalus adamanteus*, *USB corporation*) enzymes supplemented with specific amounts of labeled internal standards. The sample was incubated for two times 3 h (3 h and overnight, respectively) at 37 °C.

Digest of RNA

1.0 µg of RNA in 35 µL ultrapure H_2O were digested as follows: an aqueous solution (7.5 µL) of 480 µM $ZnSO_4$, containing 18.4 U nuclease S1 (*Aspergillus oryzae*, *Sigma-Aldrich*), 5 U Antarctic phosphatase (*NEB*) and specific amounts of the labelled internal standards [D$_3$]-m[6]A was added, and the mixture was incubated at 37 °C for 3 h. After addition of 7.5 µL of a 520 µM Na_2-[EDTA] solution, containing 0.15 U snake venom phosphodiesterase I (*Crotalus adamanteus*, *USB corporation*), the sample was incubated for 3 h at 37 °C. After the digestion the total digestion volume of 50 µL was filtered by using an AcroPrep Advance 96 filter plate 0.2 µm Supor (*Pall Corporation*) prior to UHPLC-MS/MS analysis.

Analysis of gDNA samples

Quantitative LC-ESI-MS/MS analysis of digested gDNA samples was performed as described in a previous publication.[374]

Analysis of RNA samples

The quantitative UHPLC-MS/MS analysis of digested RNA samples was performed on a 1290 UHPLC system equipped with a UV detector and an Agilent 6490 triple quadrupole mass spectrometer using

a stable isotope dilution technique. An optimized quantitative UHPLC-MS/MS method was previously published[417] to analyze multiple nucleosides in one run in parallel. The source parameters and parameters for chromatographic separation were slightly modified with the following changes: 10.8→20 min; 80% MeCN; 20→ 20.8 min, 80%→0% (v/v) MeCN, 20.8→22 min; 0% MeCN.

Table 2: Compound dependent LC-MS/MS parameters for the analysis of RNA. CE: *Collision Ene*rgy, CAV: *Collision Cell Accelerator Voltage*, EMV: *Electron Multiplier Voltage*. The nucleosides are measured in the positive mode. ([M+H]+ species in *ion selected reaction monitoring mode* (SRM))

Compound	Precursor Ion (m/z)	MS1 Resolution	Product Ion (m/z)	MS2 Resolution	Dwell time [ms]	CE (V)	CAV (V)	Polarity
Time segment 3.5-14 min								
[D$_3$]-m^{6}A	285.14	Wide	153.10	Wide	20	2	5	Positive
m^{6}A	282.12	Wide	150.08	Wide	20	2	5	Positive

5.2.2 Methods for the analysis of active demethylation of m^{5}dC *via* deamination

Cell culture of mESCs

The [^{13}C,D$_3$]-methionine feeding experiments on wt and KO mESCs were performed by *Dr. Fabio Spada* or *Yingqian Zhang* and are not described in detail. The samples were lyzed in GTC-Citrate+ buffer and stored at -80 °C until gDNA isolation.

gDNA Isolation

The gDNA isolation of mESC samples supplemented with [^{13}C,D$_3$]-methionine was performed by *Dr. Fabio Spada* or *Luis de la Osa de la Rosa* according to a standard protocol[412] and is not described in detail.

Digest of gDNA and UHPLC-MS/MS analysis

The digest of gDNA and UHPLC-MS/MS analysis was performed as described in a previous publication[412] using 4 µg of DNA and S1 Nuclease, Antarctic Phosphatase and Snake Venom Phosphodiesterase.

Extraction of the soluble pool

The extraction of the soluble pool was performed as described in a previous publication.[179] In detail, a confluent p60 plate with mESCs was extracted with 50% MeCN (ice-cold), incubated on ice for 5 min and the insoluble fraction pelleted at 21130 g and 0 °C for 5 min. The supernatant was transferred and the solvent removed by lyophilization on a *Christ Alpha L--D* plus. A SUPEL™ SELECT HLB SPE 60 mg/3 mL SPE tube (*Supelco*) was pre-equilibrated with 1.5 mL MeOH and 3 mL HCl (pH=4). The dried residue was dissolved in 1 mL of HCl (pH=4) and loaded on the equilibrated SPE tube. Subsequently, the tube was washed with 1.5 mL HCl (pH=4) and then dried with high vacuum. Elution of the soluble

pool was performed with 500 µL MeCN/MeOH (1:1) and the solvent removed by lyophilization on a *Christ Alpha L--D* plus. The residue was dissolved in 200 µL ultra-pure water and 25 µL of the solution was used per technical replicate of the digest.

Digest of the soluble pool

Of the dissolved sample obtained from the extraction of the soluble pool, 25 µL were diluted to a total volume of 35 µL. A mixture of internal standards, AP and ZnSO$_4$ in concentrations as described in a previous publication[412] was added to give a volume of 50 µL. The sample was incubated for 3 h at 37 °C, before the routine filtration procedure.[412]

5.2.3   Methods for the analysis of base excision repair

Cell culture of Neil KO cell lines

Neil wt and Neil1[-/-], Neil2[-/-] and Neil1,2[-/-] *knockout* cell lines were cultured as previously described.[177] For determination of global AP sites and β-elimination products, the cells were moved into the primed state by removing 2i from the medium. The cells were cultured under serum/LIF *priming* conditions for 3 d, passaged and expanded for another 48 h. After removal of the medium and washing the cells with DPBS, they were directly lyzed with G2[+] buffer as described in a previous publication.[410]

Cell culture of other mESCs and feeding experiments

Culture of additional cell lines with the corresponding feeding experiments was performed by *Angie Kirchner* and are not described in detail.

gDNA isolation

Isolation of gDNA was performed as described in a previous publication.[410]

Derivatization of the abasic sites and β-elimination products in gDNA

To capture AP sites and β-elimination products, the gDNA was derivatized with a hydroxylamine reagent (Superfly, SF) as described in a previous publication.[410]

Digest of gDNA

For determination of global AP sites and β-elimination products 5 µg of gDNA, and for labeled AP sites and β-elimination products 20 µg of gDNA, respectively, were digested as described in a previous publication.[410]

UHPLC-MS/MS analysis

Analysis of the digested gDNA was performed as described in a previous publication.[410]

5.2.4 Methods for the Aza(d)C project

Cell culture of mESCs for Aza(d)C-treatment

Feeder independent wt J1 cells (strain 129S4/SvJae)[418] and the respective Dnmt1-/-,[401] Dnmt2-/-,[419] Dnmt3a-/-[419] and Dnmt3b-/-[419] *knockout* cell lines were cultured as previously described.[177]

For drug treatment, cells were moved into the primed state by removing 2i from the medium. Cells were incubated 2 d in DMEM supplemented with FBS and LIF in 6-well plates (*VWR*). After splitting, 2×10^5 cells were transferred into a 6-well plate culture dish and incubated additional 2 d. Then the medium was replaced with DMEM supplemented with FBS and LIF medium containing Aza(d)C (1 µM, 0.001% DMSO) and incubated for another 24 h. After removal of the medium and washing the cells with DPBS, they were directly lyzed with RLT+ buffer as described in a previous publication[412] and subjected to the described DNA isolation.

Cell culture of cancer cell lines

Various cancer cell lines were cultured and treated by *Laura Bocci* from the group of *Prof. Spiekermann* and are not described in detail.

Extraction and treatment of PDX cells

Cells were routinely extracted from either the murine spleen or the bone marrow by the *Dr. Binje Vick* or technicians from the *Dr. Jeremias* laboratory and the procedure is not described in detail. In brief, AML cells were subsequently enriched *via* Ficoll gradient centrifugation. After a cell count and microscopic inspection of the enrichment, about 10 million cells were cultured in 5 mL DD-medium (supplied by the *Jeremias'* laboratory) with 1 µM AzaC in *Jeremias'* or our laboratory. Every 24 h the cells were counted and readjusted to the starting conditions with administration of fresh AzaC. After 24 h, 48 h and 72 h, respectively, the cells were washed with DPBS and lyzed with 1.6 mL RLT+ buffer. Lysates were stored at -20 °C until DNA and RNA isolation.

gDNA isolation

The isolation of gDNA was performed as described in a previous publication.[161]

RNA isolation

The isolation of RNA was performed as described in a previous publication.[161]

As a variation, the RNA was isolated with two instead of one volume of EtOH to obtain total RNA instead of tRNA-depleted RNA.

Reduction and re-isolation of gDNA and RNA

The reduction of gDNA and RNA, as well as subsequent re-isolation was performed as described in a previous publication.[161]

Digest of gDNA

The digest of gDNA and the external calibration curve – a serial dilution of pure H$_2$-AzadC – was performed as described in a previous publication.[161]

Digest of RNA

The digest of RNA and the external calibration curve – a serial dilution of pure H$_2$-AzaC – was performed as described in a previous publication.[161]

UHPLC-MS analysis of Aza(d)C

The UHPLC-MS/MS analysis of Aza(d)C samples was performed as described in a previous publication.[161]

5.2.5 Methods for the investigation of cAzadC

Cell culture of wt mESCs for treatment with cAzadC

Feeder independent wt J1 cells were in principal cultured as described in **5.2.4**. For drug treatment, cells were moved into the primed state by removing 2i from the medium. Cells were incubated 2 d in DMEM supplemented with FBS and LIF in 6-well plates (*VWR*). After splitting, 2x10^5 cells were transferred into a 6-well plate culture dish and supplemented with either 1 µM or 5 µM cAzadC, 0.01% DMSO or 0.05% DMSO, respectively. All experiments were treated for 72 h. Of the cells treated with cAzadC, one experiment each was incubated for the full 3 d without any medium change, as were the DMSO controls. One replicate of cells was supplemented every 24 h with fresh medium containing 1 µM or 5 µM cAzadC, respectively. After removal of the medium and washing the cells with DPBS, they were directly lyzed with RLT$^+$ buffer as described in a previous publication[412] and subjected to the described DNA isolation.

gDNA isolation and total enzymatic digest

Essentially, the gDNA was isolated as described previously.[412] Due to the higher stability of cAzadC, a hydrogenation procedure was not necessary and the gDNA was directly subjected to a total enzymatic digest as described in a previous publication.[161]

UHPLC-MS analysis of cAzadC and calibration curves for cAzadC

The external calibration curve was generated by serially diluting pure cAzadC and measuring in technical triplicates prior to each measurement.

5.2.6 Methods for the analysis of 6-Aza-dC derivatives

Cell culture of mESCs or somatic cells and feeding experiments

Culture of mESCs, CHO-K1, RBL-2H3 or Neuro-2a cells with the corresponding feeding experiments was performed by *Angie Kirchner* or *Ewelina Kaminska* and are not described in detail.

gDNA isolation

Isolation of gDNA was performed by *Luis de la Osa de la Rosa*, *Ewelina Kaminska* or *Angie Kirchner* as described in a previous publication.[412]

Digest and UHPLC-MS/MS analysis of gDNA

Digest of the gDNA was performed as described in a previous publication[412] using S1 Nuclease, Antarctic Phosphatase and Snake Venom Phosphodiesterase. No isotope standards were added to the digestion solution.

Quantification of the 6-Aza-dC derivatives was performed *via* external calibration curves that were measured as technical triplicates of a serial dilution of the pure nucleosides prior to each sample batch.

Derivatization of 6-Aza-fdC

To achieve a sharper peak during the HPLC of 6-Aza-fdC, the nucleoside (60 nmol, 1 eq.) was derivatized with methoxyamine (600 nmol, 10 eq.) in NaOH pH 10. After 15 min at r.t. the reaction was neutralized with water/formic acid pH 3.

The digested gDNA was lyophyllized, resuspended in NaOH (pH 10, 500 µL) and incubated with methoxyamine (100 fmol) for 15 min at r.t. Subsequently, the reaction was neutralized with water/formic acid (pH 3, 200 µL) and the solvent removed. Before UHPLC-MS analysis, the sample was resuspended in water (50 µL) and filtrated by using an AcroPrep Advance 96 filter plate 0.2 µm Supor (*Pall Corporation*) as described previously.[412]

5.2.7 Methods for analysis of RNA in regards to their i^6A, ms^2i^6A and t^6A levels

RNA digestion

0.5 µg-1.0 µg of RNA in 34 µL ultrapure H_2O were digested as follows: An aqueous solution (7.5 µL) of 480 µM $ZnSO_4$, containing 18.4 U nuclease S1 (*Aspergillus oryzae*, *Sigma-Aldrich*), 5 U Antarctic phosphatase (*NEB*) and specific amounts of the labelled internal standards $[D_3]$-ms^2i^6A was added, and the mixture was incubated at 37 °C for 3 h. After addition of 7.5 µL of a 520 µM Na_2-[EDTA] solution, containing 0.15 U snake venom phosphodiesterase I (*Crotalus adamanteus*, *USB corporation*), the sample was incubated for 3 h at 37 °C. After the digestion 1 µL 50% formic acid (LC-MS grade, S*igma-aldrich*) was added to the digestion mixture and mixed well, giving a total digestion volume of 50 µL with 1% formic acid. Subsequently, 3 µL of the reaction mixture were diluted with 47 µL ultrapure water containing specific amounts of $[^{13}C_4,^{15}N]$-t^6A and $[D_2]$-i^6A. The diluted sample, as well as the remaining digestion mixture were filtered by using an AcroPrep Advance 96 filter plate 0.2 µm Supor (*Pall Corporation*) prior to UHPLC-MS/MS analysis. All measurements were performed in technical duplicates or triplicates.

UHPLC-MS/MS analysis of RNA

The quantitative UHPLC-MS/MS analysis of digested RNA samples was performed on a 1290 UHPLC system equipped with a UV detector and an Agilent 6490 triple quadrupole mass spectrometer using a stable isotope dilution technique. An optimized quantitative UHPLC-MS/MS method based on previously published methods[274, 357, 417, 420] was developed to analyze all nucleosides in one run in parallel. The source parameters and parameters for chromatographic separation were slightly modified from a previously published method[417] with the following changes: 10.8→20 min; 80% MeCN; 20→ 20.8 min, 80%→0% (v/v) MeCN, 20.8→22 min; 0% MeCN.

Table 3: Compound dependent LC-MS/MS parameters for the analysis of RNA. CE: *Collision Ene*rgy, CAV: *Collision Cell Accelerator Voltage*, EMV: *Electron Multiplier Voltage*. The nucleosides are measured in the positive mode. ([M+H]+ species in *ion selected reaction monitoring mode* (SRM)).

Compound	Precursor Ion (m/z)	MS1 Resolution	Product Ion (m/z)	MS2 Resolution	Dwell time [ms]	CE (V)	CAV (V)	Polarity
Time segment 3.5-9 min								
$[D_2]$-i^{6}A	338.2	Wide	206.1	Wide	20	15	5	Positive
i^{6}A	336.2	Wide	204.1	Wide	20	15	5	Positive
$[^{13}C_4,^{15}N]$-t^{6}A	418.2	Wide	286.1	Wide	20	25	5	Positive
t^{6}A	413.1	Wide	281.1	Wide	20	25	5	Positive
Time segment 9-14 min								
$[D_3]$-ms^{2}i^{6}A	385.2	Wide	253.1	Wide	245	25	5	Positive
ms^{2}i^{6}A	382.2	Wide	250.1	Wide	245	25	5	Positive

Quantification of the nucleosides

In order to obtain the internal calibration curves for exact quantification, each standard, namely $[D_3]$-ms^{2}i^{6}A, $[^{13}C_4,^{15}N]$-t^{6}A and $[D_2]$-i^{6}A was analyzed with the *MassHunter Workstation Software Quantitative Analysis (Version B.07.01)* in comparison to the corresponding non-labelled nucleoside with constant concentration. This gives different ratios for the amount n of the natural nucleoside over the amount n* of the labeled nucleoside. Each sample was measured in technical triplicates to obtain the area A for the natural nucleoside and A* for the labeled nucleoside. The ratio of A/A* was then plotted against the ratio of n/n* with Origin® 6.0 (Microcal™) and linear regression was performed to achieve the calibration curve (**Figure 57/Table SI-3**). Quantification of the canonical base adenosine was performed as described previously.[417] Using the calibration curves one can determine the amount n from the sample with the values for A, A* and the known amount n*. Finally, the level of the modified nucleosides is calculated as percentage of the amount of adenosine.

6. Literature

[1]   M. Muramatsu*, et al.*, Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell* **2000**, *102*, 553-563.

[2]   T. Lindahl, R. D. Wood, Quality control by DNA repair. *Science* **1999**, *286*, 1897-1905.

[3]   D. E. Volk*, et al.*, Solution structure of a cis-opened (10R)-N6-deoxyadenosine adduct of (9S,10R)-9,10-epoxy-7,8,9,10-tetrahydrobenzo[a]pyrene in a DNA duplex. *Biochemistry* **2003**, *42*, 1410-1420.

[4]   J. E. LeClerc, A. Borden, C. W. Lawrence, The thymine-thymine pyrimidine-pyrimidone(6-4) ultraviolet light photoproduct is highly mutagenic and specifically induces 3' thymine-to-cytosine transitions in Escherichia coli. *Proc. Natl. Acad. Sci. U S A* **1991**, *88*, 9685-9689.

[5]   S. Zuo, R. J. Boorstein, G. W. Teebor, Oxidative damage to 5-methylcytosine in DNA. *Nucleic Acids Res.* **1995**, *23*, 3239-3243.

[6]   M. L. Michaels, C. Cruz, A. P. Grollman, J. H. Miller, Evidence that MutY and MutM combine to prevent mutations by an oxidatively damaged form of guanine in DNA. *Proc. Natl. Acad. Sci. U S A* **1992**, *89*, 7022-7025.

[7]   L. Alsoe*, et al.*, Uracil Accumulation and Mutagenesis Dominated by Cytosine Deamination in CpG Dinucleotides in Mice Lacking UNG and SMUG1. *Sci. Rep.* **2017**, *7*, 7199.

[8]   F. el Ghissassi, A. Barbin, J. Nair, H. Bartsch, Formation of 1,N6-ethenoadenine and 3,N4-ethenocytosine by lipid peroxidation products and nucleic acid bases. *Chem. Res. Toxicol.* **1995**, *8*, 278-283.

[9]   G. Slupphaug*, et al.*, A nucleotide-flipping mechanism from the structure of human uracil-DNA glycosylase bound to DNA. *Nature* **1996**, *384*, 87-92.

[10]   C. D. Mol*, et al.*, Crystal structure and mutational analysis of human uracil-DNA glycosylase: structural basis for specificity and catalysis. *Cell* **1995**, *80*, 869-878.

[11]   B. Sun, K. A. Latham, M. L. Dodson, R. S. Lloyd, Studies on the catalytic mechanism of five DNA glycosylases. Probing for enzyme-DNA imino intermediates. *J. Biol. Chem.* **1995**, *270*, 19501-19508.

[12]   H. M. Nash, R. Lu, W. S. Lane, G. L. Verdine, The critical active-site amine of the human 8-oxoguanine DNA glycosylase, hOgg1: direct identification, ablation and chemical reconstitution. *Chem. Biol.* **1997**, *4*, 693-702.

[13]   S. D. Bruner, D. P. Norman, G. L. Verdine, Structural basis for recognition and repair of the endogenous mutagen 8-oxoguanine in DNA. *Nature* **2000**, *403*, 859-866.

[14]   J. C. Fromme*, et al.*, Product-assisted catalysis in base-excision DNA repair. *Nat. Struct. Biol.* **2003**, *10*, 204-211.

[15]   S. S. Parikh*, et al.*, Base excision repair initiation revealed by crystal structures and binding kinetics of human uracil-DNA glycosylase with DNA. *EMBO J.* **1998**, *17*, 5214-5226.

[16]   T. R. Waters, P. Gallinari, J. Jiricny, P. F. Swann, Human thymine DNA glycosylase binds to apurinic sites in DNA but is displaced by human apurinic endonuclease 1. *J. Biol. Chem.* **1999**, *274*, 67-74.

[17]   U. Hardeland, M. Bentele, J. Jiricny, P. Schar, Separating substrate recognition from base hydrolysis in human thymine DNA glycosylase by mutational analysis. *J. Biol. Chem.* **2000**, *275*, 33449-33456.

[18]   J. W. Hill, T. K. Hazra, T. Izumi, S. Mitra, Stimulation of human 8-oxoguanine-DNA glycosylase by AP-endonuclease: potential coordination of the initial steps in base excision repair. *Nucleic Acids Res.* **2001**, *29*, 430-438.

[19]   M. A. Pope, S. L. Porello, S. S. David, Escherichia coli apurinic-apyrimidinic endonucleases enhance the turnover of the adenine glycosylase MutY with G:A substrates. *J. Biol. Chem.* **2002**, *277*, 22605-22615.

[20]   A. R. Weber*, et al.*, Biochemical reconstitution of TET1-TDG-BER-dependent active DNA demethylation reveals a highly coordinated mechanism. *Nat. Commun.* **2016**, *7*, 10806.

[21]   M. Guillet, S. Boiteux, Endogenous DNA abasic sites cause cell death in the absence of Apn1, Apn2 and Rad1/Rad10 in Saccharomyces cerevisiae. *EMBO J.* **2002**, *21*, 2833-2841.

[22]   N. Owiti*, et al.*, Def1 and Dst1 play distinct roles in repair of AP lesions in highly transcribed genomic regions. *DNA Repair (Amst)* **2017**, *55*, 31-39.

[23]   S. Boiteux, S. Jinks-Robertson, DNA repair mechanisms and the bypass of DNA damage in Saccharomyces cerevisiae. *Genetics* **2013**, *193*, 1025-1064.

[24]   T. Lindahl, Instability and decay of the primary structure of DNA. *Nature* **1993**, *362*, 709-715.

[25]   M. Goulian, B. Bleile, B. Y. Tseng, The effect of methotrexate on levels of dUTP in animal cells. *J. Biol. Chem.* **1980**, *255*, 10630-10637.

[26]   N. Kim, S. Jinks-Robertson, Abasic sites in the transcribed strand of yeast DNA are removed by transcription-coupled nucleotide excision repair. *Mol. Cell. Biol.* **2010**, *30*, 3206-3215.

[27]   S. Tornaletti, L. S. Maeda, P. C. Hanawalt, Transcription arrest at an abasic site in the transcribed strand of template DNA. *Chem. Res. Toxicol.* **2006**, *19*, 1215-1220.

[28]   N. Kozhukhar*, et al.*, The efficiency of the translesion synthesis across abasic sites by mitochondrial DNA polymerase is low in mitochondria of 3T3 cells. *Mitochondr. DNA A DNA Mapp. Seq. Anal.* **2016**, *27*, 4390-4396.

[29]   R. J. Carter, J. L. Parsons, Base Excision Repair, a Pathway Regulated by Posttranslational Modifications. *Mol. Cell. Biol.* **2016**, *36*, 1426-1437.

[30]   P. Burkovics, V. Szukacsov, I. Unk, L. Haracska, Human Ape2 protein has a 3'-5' exonuclease activity that acts preferentially on mismatched base pairs. *Nucleic Acids Res.* **2006**, *34*, 2508-2515.

[31]   S. C. Popoff, A. I. Spira, A. W. Johnson, B. Demple, Yeast structural gene (APN1) for the major apurinic endonuclease: homology to Escherichia coli endonuclease IV. *Proc. Natl. Acad. Sci. U S A* **1990**, *87*, 4193-4197.

[32]   K. L. Meadows, B. Song, P. W. Doetsch, Characterization of AP lyase activities of Saccharomyces cerevisiae Ntg1p and Ntg2p: implications for biological function. *Nucleic Acids Res.* **2003**, *31*, 5560-5567.

[33]   Y. Matsumoto, K. Kim, Excision of deoxyribose phosphate residues by DNA polymerase beta during DNA repair. *Science* **1995**, *269*, 699-702.

[34]   R. Prasad, W. A. Beard, P. R. Strauss, S. H. Wilson, Human DNA polymerase beta deoxyribose phosphate lyase. Substrate specificity and catalytic mechanism. *J. Biol. Chem.* **1998**, *273*, 15263-15270.

[35]   P. M. Hegde*, et al.*, The C-terminal Domain (CTD) of Human DNA Glycosylase NEIL1 Is Required for Forming BERosome Repair Complex with DNA Replication Proteins at the Replicating Genome: DOMINANT NEGATIVE FUNCTION OF THE CTD. *J. Biol. Chem.* **2015**, *290*, 20919-20933.

[36]   R. Steinacher*, et al.*, SUMOylation coordinates BERosome assembly in active DNA demethylation during cell differentiation. *EMBO J.* **2019**, *38*.

[37]   A. Das*, et al.*, NEIL2-initiated, APE-independent repair of oxidized bases in DNA: Evidence for a repair complex in human cells. *DNA Repair (Amst)* **2006**, *5*, 1439-1448.

[38]   L. Wiederhold*, et al.*, AP endonuclease-independent DNA base excision repair in human cells. *Mol. Cell* **2004**, *15*, 209-220.

[39]   C. J. Whitehouse*, et al.*, XRCC1 stimulates human polynucleotide kinase activity at damaged DNA termini and accelerates DNA single-strand break repair. *Cell* **2001**, *104*, 107-117.

[40]   A. E. Vidal, S. Boiteux, I. D. Hickson, J. P. Radicella, XRCC1 coordinates the initial and late stages of DNA abasic site repair through protein-protein interactions. *EMBO J.* **2001**, *20*, 6530-6539.

[41]   Y. Kubota*, et al.*, Reconstitution of DNA base excision-repair with purified human proteins: interaction between DNA polymerase beta and the XRCC1 protein. *EMBO J.* **1996**, *15*, 6662-6670.

[42]   K. W. Caldecott*, et al.*, An interaction between the mammalian DNA repair protein XRCC1 and DNA ligase III. *Mol. Cell. Biol.* **1994**, *14*, 68-76.

[43]   C. H. Waddington, The epigenotype. 1942. *Int. J. Epidemiol.* **2012**, *41*, 10-13.

[44]   K. D. Robertson, DNA methylation and human disease. *Nat. Rev. Genet.* **2005**, *6*, 597-610.

[45]   M. Tahiliani*, et al.*, Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **2009**, *324*, 930-935.

[46]   M. Ehrlich*, et al.*, DNA methylation in thermophilic bacteria: N4-methylcytosine, 5-methylcytosine, and N6-methyladenine. *Nucleic Acids Res.* **1985**, *13*, 1399-1412.

[47]   M. Yu*, et al.*, Base-resolution detection of N4-methylcytosine in genomic DNA using 4mC-Tet-assisted-bisulfite- sequencing. *Nucleic Acids Res.* **2015**, *43*, e148.

[48]   S. Kumar*, et al.*, N4-cytosine DNA methylation regulates transcription and pathogenesis in Helicobacter pylori. *Nucleic Acids Res.* **2018**, *46*, 3429-3445.

[49]   A. Lobner-Olesen, O. Skovgaard, M. G. Marinus, Dam methylation: coordinating cellular processes. *Curr. Opin. Microbiol.* **2005**, *8*, 154-160.

[50]   M. V. Rojas, N. Galanti, DNA methylation in Trypanosoma cruzi. *FEBS Lett.* **1990**, *263*, 113-116.

[51]   M. A. Gorovsky, S. Hattman, G. L. Pleger, (6 N)methyl adenine in the nuclear DNA of a eucaryote, Tetrahymena pyriformis. *J. Cell Biol.* **1973**, *56*, 697-701.

[52]   S. Hattman, C. Kenny, L. Berger, K. Pratt, Comparative study of DNA methylation in three unicellular eucaryotes. *J. Bacteriol.* **1978**, *135*, 1156-1157.

[53]   S. Bromberg, K. Pratt, S. Hattman, Sequence specificity of DNA adenine methylase in the protozoan Tetrahymena thermophila. *J. Bacteriol.* **1982**, *150*, 993-996.

[54]   E. E. Capowski, J. M. Wells, G. S. Harrison, K. M. Karrer, Molecular analysis of N6-methyladenine patterns in Tetrahymena thermophila nuclear DNA. *Mol. Cell. Biol.* **1989**, *9*, 2598-2605.

[55]   K. M. Karrer, T. A. VanNuland, Position effect takes precedence over target sequence in determination of adenine methylation patterns in the nuclear genome of a eukaryote, Tetrahymena thermophila. *Nucleic Acids Res.* **1998**, *26*, 4566-4573.

[56]   K. M. Karrer, S. Stein-Gavens, Constancy of adenine methylation in Tetrahymena macronuclear DNA. *J. Protozool.* **1990**, *37*, 409-414.

[57]   D. J. Cummings, A. Tait, J. M. Goddard, Methylated bases in DNA from Paramecium aurelia. *Biochim. Biophys. Acta* **1974**, *374*, 1-11.

[58]   P. M. Rae, B. B. Spear, Macronuclear DNA of the hypotrichous ciliate Oxytricha fallax. *Proc. Natl. Acad. Sci. U S A* **1978**, *75*, 4992-4996.

[59]   C. M. Zhu, H. R. Henney, Jr., DNA methylation pattern during the encystment of Physarum flavicomum. *Biochem. Cell Biol.* **1990**, *68*, 944-948.

[60]   Y. Fu*, et al.*, N6-methyldeoxyadenosine marks active transcription start sites in Chlamydomonas. *Cell* **2015**, *161*, 879-892.

[61]   P. Babinger, I. Kobl, W. Mages, R. Schmitt, A link between DNA methylation and epigenetic silencing in transgenic Volvox carteri. *Nucleic Acids Res.* **2001**, *29*, 1261-1271.

[62]   Y. Wang*, et al.*, N6-adenine DNA methylation is associated with the linker DNA of H2A.Z-containing well-positioned nucleosomes in Pol II-transcribed genes in Tetrahymena. *Nucleic Acids Res.* **2017**, *45*, 11594-11606.

[63]   K. Pratt, S. Hattman, Deoxyribonucleic acid methylation and chromatin organization in Tetrahymena thermophila. *Mol. Cell. Biol.* **1981**, *1*, 600-608.

[64]   G. S. Harrison, R. C. Findly, K. M. Karrer, Site-specific methylation of adenine in the nuclear genome of a eucaryote, Tetrahymena thermophila. *Mol. Cell. Biol.* **1986**, *6*, 2364-2370.

[65]   K. M. Karrer, T. A. VanNuland, Methylation of adenine in the nuclear DNA of Tetrahymena is internucleosomal and independent of histone H1. *Nucleic Acids Res.* **2002**, *30*, 1364-1370.

[66]   G. Z. Luo*, et al.*, N(6)-methyldeoxyadenosine directs nucleosome positioning in Tetrahymena DNA. *Genome Biol.* **2018**, *19*, 200.

[67]   M. Salvini, E. Barone, S. Ronca, R. Nobili, DNA Methylation in Vegetative and Conjugating Cells of a Protozoan Ciliate: *Blepharisma japonicum. Dev. Genet.* **1986**, *9*, 149-158.

[68]   R. L. Adams, E. L. McKay, L. M. Craig, R. H. Burdon, Methylation of mosquito DNA. *Biochim. Biophys. Acta* **1979**, *563*, 72-81.

[69]   E. M. Reyes, I. Camacho-Arroyo, G. Nava, M. A. Cerbon, Differential methylation in steroid 5 alpha-reductase isozyme genes in epididymis, testis, and liver of the adult rat. *J. Androl.* **1997**, *18*, 372-377.

[70]   U. Gunthert, M. Schweiger, M. Stupp, W. Doerfler, DNA methylation in adenovirus, adenovirus-transformed cells, and host cells. *Proc. Natl. Acad. Sci. U S A* **1976**, *73*, 3923-3927.

[71]   D. Ratel*, et al.*, Undetectable levels of N6-methyl adenine in mouse DNA: Cloning and analysis of PRED28, a gene coding for a putative mammalian DNA adenine methyltransferase. *FEBS Lett.* **2006**, *580*, 3179-3184.

[72]   E. L. Greer*, et al.*, DNA Methylation on N6-Adenine in C. elegans. *Cell* **2015**, *161*, 868-878.

[73]   G. Zhang*, et al.*, N6-methyladenine DNA modification in Drosophila. *Cell* **2015**, *161*, 893-906.

[74]   M. J. Koziol*, et al.*, Identification of methylated deoxyadenosines in vertebrates reveals diversity in DNA modifications. *Nat. Struct. Mol. Biol.* **2016**, *23*, 24-30.

[75]   D. Liang*, et al.*, The decreased N(6)-methyladenine DNA modification in cancer cells. *Biochem. Biophys. Res. Commun.* **2016**, *480*, 120-125.

[76]   T. P. Wu*, et al.*, DNA methylation on N(6)-adenine in mammalian embryonic stem cells. *Nature* **2016**, *532*, 329-333.

[77]   N. C. Parashar, G. Parashar, H. Nayyar, R. Sandhir, N(6)-adenine DNA methylation demystified in eukaryotic genome: From biology to pathology. *Biochimie* **2018**, *144*, 56-62.

[78]   Z. Liang*, et al.*, DNA N(6)-Adenine Methylation in Arabidopsis thaliana. *Dev. Cell* **2018**, *45*, 406-416 e403.

[79]   Q. Zhang*, et al.*, N(6)-Methyladenine DNA Methylation in Japonica and Indica Rice Genomes and Its Association with Gene Expression, Plant Development, and Stress Responses. *Mol. Plant.* **2018**, *11*, 1492-1508.

[80]   C. Ma*, et al.*, N6-methyldeoxyadenine is a transgenerational epigenetic signal for mitochondrial stress adaptation. *Nat. Cell Biol.* **2018**, *21*, 319-327.

[81]   B. Yao*, et al.*, DNA N6-methyladenine is dynamically regulated in the mouse brain following environmental stress. *Nat. Commun.* **2017**, *8*, 1122.

[82]   S. L. Kigar*, et al.*, N(6)-methyladenine is an epigenetic marker of mammalian early life stress. *Sci. Rep.* **2017**, *7*, 18078.

[83]   J. Liu*, et al.*, Abundant DNA 6mA methylation during early embryogenesis of zebrafish and pig. *Nat. Commun.* **2016**, *7*, 13052.

[84]   C. L. Xiao*, et al.*, N(6)-Methyladenine DNA Modification in the Human Genome. *Mol. Cell* **2018**, *71*, 306-318 e307.

[85]   C. Zhou*, et al.*, DNA N(6)-methyladenine demethylase ALKBH1 enhances osteogenic differentiation of human MSCs. *Bone Res.* **2016**, *4*, 16033.

[86]   Q. Xie*, et al.*, N(6)-methyladenine DNA Modification in Glioblastoma. *Cell* **2018**, *175*, 1228-1243 e1220.

[87]   W. Huang*, et al.*, Determination of DNA adenine methylation in genomes of mammals and plants by liquid chromatography/mass spectrometry. *RSC Adv.* **2015**, *5*, 64046-64054.

[88]   F. Zheng*, et al.*, Genomewide analysis of 6-methyladenine DNA in peripheral blood mononuclear cells of systemic lupus erythematosus. *Lupus* **2019**, *28*, 359-364.

[89]   C. W. Q. Koh*, et al.*, Single-nucleotide-resolution sequencing of human N6-methyldeoxyadenosine reveals strand-asymmetric clusters associated with SSBP1 on the mitochondrial genome. *Nucleic Acids Res.* **2018**, *46*, 11659-11670.

[90]   J. Xiong*, et al.*, N6-Hydroxymethyladenine: a hydroxylation derivative of N6-methyladenine in genomic DNA of mammals. *Nucleic Acids Res.* **2019**, *47*, 1268-1277.

[91]   Z. K. O'Brown*, et al.*, Sources of artifact in measurements of 6mA and 4mC abundance in eukaryotic genomic DNA. *BMC Genomics* **2019**, *20*, 445.

[92]   A. Lentini*, et al.*, A reassessment of DNA-immunoprecipitation-based genomic profiling. *Nat. Methods* **2018**, *15*, 499-504.

[93]   A. Bird, DNA methylation patterns and epigenetic memory. *Genes Dev.* **2002**, *16*, 6-21.

[94]   W. Reik, A. Lewis, Co-evolution of X-chromosome inactivation and imprinting in mammals. *Nat. Rev. Genet.* **2005**, *6*, 403-410.

[95]   M. G. Goll, T. H. Bestor, Eukaryotic cytosine methyltransferases. *Annu. Rev. Biochem.* **2005**, *74*, 481-514.

[96]   M. Kaneda*, et al.*, Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. *Nature* **2004**, *429*, 900-903.

[97]   T. H. Bestor, The DNA methyltransferases of mammals. *Hum. Mol. Genet.* **2000**, *9*, 2395-2402.

[98]   J. P. Issa, CpG island methylator phenotype in cancer. *Nat. Rev. Cancer* **2004**, *4*, 988-993.

[99]   R. Lister*, et al.*, Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **2009**, *462*, 315-322.

[100]   K. Shirane*, et al.*, Mouse oocyte methylomes at base resolution reveal genome-wide accumulation of non-CpG methylation and role of DNA methyltransferases. *PLoS Genet.* **2013**, *9*, e1003439.

[101]   R. Lister*, et al.*, Global epigenomic reconfiguration during mammalian brain development. *Science* **2013**, *341*, 1237905.

[102]   B. H. Ramsahoye*, et al.*, Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. *Proc. Natl. Acad. Sci. U S A* **2000**, *97*, 5237-5242.

[103]   W. Xie*, et al.*, Base-resolution analyses of sequence and parent-of-origin dependent DNA methylation in the mouse genome. *Cell* **2012**, *148*, 816-831.

[104]   Z. D. Smith, A. Meissner, DNA methylation: roles in mammalian development. *Nat. Rev. Genet.* **2013**, *14*, 204-220.

[105]   M. Okano, D. W. Bell, D. A. Haber, E. Li, DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **1999**, *99*, 247-257.

[106]   K. E. Bachman, M. R. Rountree, S. B. Baylin, Dnmt3a and Dnmt3b are transcriptional repressors that exhibit unique localization properties to heterochromatin. *J. Biol. Chem.* **2001**, *276*, 32282-32287.

[107]   H. Leonhardt, A. W. Page, H. U. Weier, T. H. Bestor, A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. *Cell* **1992**, *71*, 865-873.

[108]   L. S. Chuang*, et al.*, Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1. *Science* **1997**, *277*, 1996-2000.

[109]   A. D. Riggs, Z. Xiong, Methylation and epigenetic fidelity. *Proc. Natl. Acad. Sci. U S A* **2004**, *101*, 4-5.

[110]   G. Egger*, et al.*, Identification of DNMT1 (DNA methyltransferase 1) hypomorphs in somatic knockouts suggests an essential role for DNMT1 in cell survival. *Proc. Natl. Acad. Sci. U S A* **2006**, *103*, 14080-14085.

[111]   M. S. Kareta*, et al.*, Reconstitution and mechanism of the stimulation of de novo methylation by human DNMT3L. *J. Biol. Chem.* **2006**, *281*, 25893-25902.

[112]   L. Vasiliauskaite*, et al.*, Defective germline reprogramming rewires the spermatogonial transcriptome. *Nat. Struct. Mol. Biol.* **2018**, *25*, 394-404.

[113]   N. Veland*, et al.*, DNMT3L facilitates DNA methylation partly by maintaining DNMT3A stability in mouse embryonic stem cells. *Nucleic Acids Res.* **2019**, *47*, 152-167.

[114]   J. F. Costello, C. Plass, Methylation matters. *J. Med. Genet.* **2001**, *38*, 285-303.

[115]   P. A. Jones, D. Takai, The Role of DNA Methylation in Mammalian Epigenetics. *Science* **2001**, *293*, 1068-1070.

[116]   R. Jaenisch, A. Bird, Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat. Genet.* **2003**, *33 Suppl*, 245-254.

[117]   K. Agrawal, V. Das, P. Vyas, M. Hajduch, Nucleosidic DNA demethylating epigenetic drugs - A comprehensive review from discovery to clinic. *Pharmacol. Ther.* **2018**, *188*, 45-79.

[118]   K. D. Robertson, P. A. Jones, DNA methylation: past, present and future directions. *Carcinogenesis* **2000**, *21*, 461-467.

[119]   S. B. Baylin*, et al.*, Alterations in DNA methylation: a fundamental aspect of neoplasia. *Adv. Cancer Res.* **1998**, *72*, 141-196.

[120] Y. Akiyama*, et al.*, Cell-type-specific repression of the maspin gene is disrupted frequently by demethylation at the promoter region in gastric intestinal metaplasia and cancer cells*. Am. J. Pathol.* **2003**, *163*, 1911-1919.

[121] A. Gupta*, et al.*, Hypomethylation of the synuclein gamma gene CpG island promotes its aberrant expression in breast carcinoma and ovarian carcinoma*. Cancer Res.* **2003**, *63*, 664-673.

[122] M. E. Figueroa*, et al.*, MDS and secondary AML display unique patterns and abundance of aberrant DNA methylation*. Blood* **2009**, *114*, 3448-3458.

[123] Y. Jiang*, et al.*, Aberrant DNA methylation is a dominant mechanism in MDS progression to AML*. Blood* **2009**, *113*, 1315-1325.

[124] S. Agrawal*, et al.*, The C/EBPdelta tumor suppressor is silenced by hypermethylation in acute myeloid leukemia*. Blood* **2007**, *109*, 3895-3905.

[125] U. Germing, G. Kobbe, R. Haas, N. Gattermann, Myelodysplastic syndromes: diagnosis, prognosis, and treatment*. Dtsch. Arztebl. Int.* **2013**, *110*, 783-790.

[126] S. H. Lim, P. M. Dubielecka, V. M. Raghunathan, Molecular targeting in acute myeloid leukemia*. J. Transl. Med.* **2017**, *15*, 183.

[127] C. Saygin, H. E. Carraway, Emerging therapies for acute myeloid leukemia*. J. Hematol. Oncol.* **2017**, *10*, 93.

[128] H. Dohner, D. J. Weisdorf, C. D. Bloomfield, Acute Myeloid Leukemia*. N. Engl. J. Med.* **2015**, *373*, 1136-1152.

[129] National Cancer Institute: Adult Acute Myeloid Leukemia Treatment (PDQ®), Patient Version*.* Bethesda, MD: National Cancer Institute*.* Date last modified Oct 19th, 2018. https://www.ncbi.nlm.nih.gov/pubmed/26389377. Accessed June 04th, 2019.

[130] J. A. Bell*, et al.*, Economic burden of elderly patients with acute myeloid leukemia treated in routine clinical care in the United States*. Leuk. Res.* **2018**, *71*, 27-33.

[131] N. Gangat, M. M. Patnaik, A. Tefferi, Myelodysplastic syndromes: Contemporary review and how we treat*. Am. J. Hematol.* **2016**, *91*, 76-89.

[132] H. Kantarjian*, et al.*, Decitabine improves patient outcomes in myelodysplastic syndromes: results of a phase III randomized study*. Cancer* **2006**, *106*, 1794-1803.

[133] P. Fenaux*, et al.*, Efficacy of azacitidine compared with that of conventional care regimens in the treatment of higher-risk myelodysplastic syndromes: a randomised, open-label, phase III study*. Lancet Oncol.* **2009**, *10*, 223-232.

[134] P. A. Jones, S. M. Taylor, Cellular differentiation, cytidine analogs and DNA methylation*. Cell* **1980**, *20*, 85-93.

[135] P. A. Jones, Altering gene expression with 5-azacytidine*. Cell* **1985**, *40*, 485-486.

[136] P. A. Jones, S. M. Taylor, T. Mohandas, L. J. Shapiro, Cell cycle-specific reactivation of an inactive X-chromosome locus by 5-azadeoxycytidine*. Proc. Natl. Acad. Sci. U S A* **1982**, *79*, 1215-1219.

[137] T. Haaf, M. Schmid, 5-Azadeoxycytidine induced undercondensation in the giant X chromosomes of Microtus agrestis*. Chromosoma* **1989**, *98*, 93-98.

[138] E. Jablonka, R. Goitein, M. Marcus, H. Cedar, DNA hypomethylation causes an increase in DNase-I sensitivity and an advance in the time of replication of the entire inactive X chromosome*. Chromosoma* **1985**, *93*, 152-156.

[139] T. Murakami*, et al.*, Induction of apoptosis by 5-azacytidine: drug concentration-dependent differences in cell cycle specificity*. Cancer Res.* **1995**, *55*, 3093-3098.

[140] H. Kizaki*, et al.*, 1-beta-D-arabinosylcytosine and 5-azacytidine induce internucleosomal DNA fragmentation and cell death in thymocytes*. Immunopharmacology* **1993**, *25*, 19-27.

[141] W. Gorczyca*, et al.*, The cell cycle related differences in susceptibility of HL-60 cells to apoptosis induced by various antitumor agents*. Cancer Res.* **1993**, *53*, 3186-3192.

[142] J. P. Issa, DNA methylation as a therapeutic target in cancer*. Clin. Cancer Res.* **2007**, *13*, 1634-1637.

[143] T. Qin*, et al.*, Mechanisms of resistance to 5-aza-2'-deoxycytidine in human cancer cell lines*. Blood* **2009**, *113*, 659-667.

[144] C. E. Requena*, et al.*, The nucleotidohydrolases DCTPP1 and dUTPase are involved in the cellular response to decitabine*. Biochem. J.* **2016**, *473*, 2635-2643.

[145] M. Daskalakis*, et al.*, Demethylation of a hypermethylated P15/INK4B gene in patients with myelodysplastic syndrome by 5-Aza-2'-deoxycytidine (decitabine) treatment*. Blood* **2002**, *100*, 2957-2964.

[146] M. S. Soengas*, et al.*, Inactivation of the apoptosis effector Apaf-1 in malignant melanoma*. Nature* **2001**, *409*, 207-211.

[147] L. H. Li, E. J. Olin, H. H. Buskirk, L. M. Reineke, Cytotoxicity and mode of action of 5-azacytidine on L1210 leukemia*. Cancer Res.* **1970**, *30*, 2760-2769.

[148] E. Flatau, F. A. Gonzales, L. A. Michalowsky, P. A. Jones, DNA methylation in 5-aza-2'-deoxycytidine-resistant variants of C3H 10T1/2 C18 cells*. Mol. Cell. Biol.* **1984**, *4*, 2098-2102.

[149] S. Gabbara, A. S. Bhagwat, The mechanism of inhibition of DNA (cytosine-5-)-methyltransferases by 5-azacytosine is likely to involve methyl transfer to the inhibitor*. Biochem. J.* **1995**, *307 (Pt 1)*, 87-92.

[150] J. K. Christman, 5-Azacytidine and 5-aza-2'-deoxycytidine as inhibitors of DNA methylation: mechanistic studies and their implications for cancer therapy*. Oncogene* **2002**, *21*, 5483-5495.

[151] R. Juttermann, E. Li, R. Jaenisch, Toxicity of 5-aza-2'-deoxycytidine to mammalian cells is mediated primarily by covalent trapping of DNA methyltransferase rather than DNA demethylation*. Proc. Natl. Acad. Sci. U S A* **1994**, *91*, 11797-11801.

[152] D. Kuch*, et al.*, Synthesis of DNA dumbbell based inhibitors for the human DNA methyltransferase Dnmt1*. Angew. Chem. Int. Ed.* **2008**, *47*, 1515-1518.

[153] L. Schermelleh*, et al.*, Trapped in action: direct visualization of DNA methyltransferase activity in living cells*. Nat. Methods* **2005**, *2*, 751-756.

[154] C. B. Yoo*, et al.*, Delivery of 5-aza-2'-deoxycytidine to cells using oligodeoxynucleotides*. Cancer Res.* **2007**, *67*, 6400-6408.

[155] R. L. Momparler, Pharmacology of 5-Aza-2'-deoxycytidine (decitabine)*. Semin. Hematol.* **2005**, *42*, S9-16.

[156] K. T. Lin, R. L. Momparler, G. E. Rivard, High-performance liquid chromatographic analysis of chemical stability of 5-aza-2'-deoxycytidine*. J. Pharm. Sci.* **1981**, *70*, 1228-1232.

[157] D. K. Rogstad*, et al.*, Chemical decomposition of 5-aza-2'-deoxycytidine (Decitabine): kinetic analyses and identification of products by NMR, HPLC, and mass spectrometry*. Chem. Res. Toxicol.* **2009**, *22*, 1194-1204.

[158] S. Oz*, et al.*, Quantitative determination of decitabine incorporation into DNA and its effect on mutation rates in human cancer cells*. Nucleic Acids Res.* **2014**, *42*, e152.

[159] W. K. Anderson, P. F. Corey, Synthesis and antileukemic activity of 5-substituted 2,3-dihydro-6,7-bis(hydroxymethyl)-1H-pyrrolizine diesters*. J. Med. Chem.* **1977**, *20*, 812-818.

[160] A. Unnikrishnan*, et al.*, AZA-MS: a novel multiparameter mass spectrometry method to determine the intracellular dynamics of azacitidine therapy in vivo*. Leukemia* **2018**, *32*, 900-910.

[161] S. Schiffers*, et al.*, Label-Free Quantification of 5-Azacytidines Directly in the Genome*. Helv. Chim. Acta* **2019**, *102*, e1800229.

[162] M. G. Goll*, et al.*, Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2*. Science* **2006**, *311*, 395-398.

[163] S. Hussain*, et al.*, NSun2-mediated cytosine-5 methylation of vault noncoding RNA determines its processing into regulatory small RNAs*. Cell Rep.* **2013**, *4*, 255-261.

[164] S. Haag*, et al.*, NSUN6 is a human RNA methyltransferase that catalyzes formation of m5C72 in specific tRNAs*. RNA* **2015**, *21*, 1532-1543.

[165] Y. Camara*, et al.*, MTERF4 regulates translation by targeting the methyltransferase NSUN4 to the mammalian mitochondrial ribosome*. Cell Metab.* **2011**, *13*, 527-539.

[166]   J. Aimiuwu*, et al.*, RNA-dependent inhibition of ribonucleotide reductase is a major pathway for 5-azacytidine activity in acute myeloid leukemia. *Blood* **2012**, *119*, 5229-5238.

[167]   Q. Ebrahem, R. Z. Mahfouz, K. P. Ng, Y. Saunthararajah, High cytidine deaminase expression in the liver provides sanctuary for cancer cells from decitabine treatment effects. *Oncotarget* **2012**, *3*, 1137-1145.

[168]   C. M. Galmarini, J. R. Mackey, C. Dumontet, Nucleoside analogues: mechanisms of drug resistance and reversal strategies. *Leukemia* **2001**, *15*, 875-890.

[169]   M. Flasshove*, et al.*, Structural analysis of the deoxycytidine kinase gene in patients with acute myeloid leukemia and resistance to cytosine arabinoside. *Leukemia* **1994**, *8*, 780-785.

[170]   A. P. Stegmann*, et al.*, In vitro-induced resistance to the deoxycytidine analogues cytarabine (AraC) and 5-aza-2'-deoxycytidine (DAC) in a rat model for acute myeloid leukemia is mediated by mutations in the deoxycytidine kinase (dck) gene. *Ann. Hematol.* **1995**, *71*, 41-47.

[171]   A. R. Van Rompay*, et al.*, Phosphorylation of uridine and cytidine nucleoside analogs by two human uridine-cytidine kinases. *Mol. Pharmacol.* **2001**, *59*, 1181-1186.

[172]   G. Borthakur*, et al.*, Activity of decitabine in patients with myelodysplastic syndrome previously treated with azacitidine. *Leukemia Lymphoma* **2008**, *49*, 690-695.

[173]   M. Oka*, et al.*, De novo DNA methyltransferases Dnmt3a and Dnmt3b primarily mediate the cytotoxic effect of 5-aza-2'-deoxycytidine. *Oncogene* **2005**, *24*, 3091-3099.

[174]   B. Sozen, A. Can, N. Demir, Cell fate regulation during preimplantation development: a view of adhesion-linked molecular interactions. *Dev. Biol.* **2014**, *395*, 73-83.

[175]   S. Ito*, et al.*, Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* **2010**, *466*, 1129-1133.

[176]   S. Kriaucionis, N. Heintz, The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **2009**, *324*, 929-930.

[177]   T. Pfaffeneder*, et al.*, The discovery of 5-formylcytosine in embryonic stem cell DNA. *Angew. Chem. Int. Ed.* **2011**, *50*, 7008-7012.

[178]   S. Ito*, et al.*, Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **2011**, *333*, 1300-1303.

[179]   K. Iwan*, et al.*, 5-Formylcytosine to cytosine conversion by C-C bond cleavage in vivo. *Nat. Chem. Biol.* **2018**, *14*, 72-78.

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

[181]   G. Ficz*, et al.*, Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature* **2011**, *473*, 398-402.

[182]   H. Wu*, et al.*, Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. *Genes Dev.* **2011**, *25*, 679-684.

[183]   E. A. Raiber*, et al.*, 5-Formylcytosine alters the structure of the DNA double helix. *Nat. Struct. Mol. Biol.* **2015**, *22*, 44-49.

[184]   M. W. Kellinger*, et al.*, 5-formylcytosine and 5-carboxylcytosine reduce the rate and substrate specificity of RNA polymerase II transcription. *Nat. Struct. Mol. Biol.* **2012**, *19*, 831-833.

[185]   M. Bachman*, et al.*, 5-Formylcytosine can be a stable DNA modification in mammals. *Nat. Chem. Biol.* **2015**, *11*, 555-557.

[186]   M. Su*, et al.*, 5-Formylcytosine Could Be a Semipermanent Base in Specific Genome Sites. *Angew. Chem. Int. Ed.* **2016**, *55*, 11797-11800.

[187]   M. Iurlaro*, et al.*, In vivo genome-wide profiling reveals a tissue-specific role for 5-formylcytosine. *Genome Biol.* **2016**, *17*, 141.

[188]   M. Munzel*, et al.*, Quantification of the sixth DNA base hydroxymethylcytosine in the brain. *Angew. Chem. Int. Ed.* **2010**, *49*, 5375-5377.

[189]   K. E. Szulwach*, et al.*, 5-hmC-mediated epigenetic dynamics during postnatal neurodevelopment and aging. *Nat. Neurosci.* **2011**, *14*, 1607-1616.

[190] I. H. Lin, Y. F. Chen, M. T. Hsu, Correlated 5-Hydroxymethylcytosine (5hmC) and Gene Expression Profiles Underpin Gene and Organ-Specific Epigenetic Regulation in Adult Mouse Brain and Liver. *PLoS One* **2017**, *12*, e0170779.

[191] T. P. Gu*, et al.*, The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature* **2011**, *477*, 606-610.

[192] K. Iqbal, S. G. Jin, G. P. Pfeifer, P. E. Szabo, Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. *Proc. Natl. Acad. Sci. U S A* **2011**, *108*, 3642-3647.

[193] M. Wossidlo*, et al.*, 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. *Nat. Commun.* **2011**, *2*, 241.

[194] F. Guo*, et al.*, Active and passive demethylation of male and female pronuclear DNA in the mammalian zygote. *Cell Stem Cell* **2014**, *15*, 447-459.

[195] N. Rougier*, et al.*, Chromosome methylation patterns during mammalian preimplantation development. *Genes Dev.* **1998**, *12*, 2108-2113.

[196] T. Nakamura*, et al.*, PGC7/Stella protects against DNA demethylation in early embryogenesis. *Nat. Cell Biol.* **2007**, *9*, 64-71.

[197] M. Saitou, S. C. Barton, M. A. Surani, A molecular programme for the specification of germ cell fate in mice. *Nature* **2002**, *418*, 293-300.

[198] C. A. Edwards, A. C. Ferguson-Smith, Mechanisms regulating imprinted genes in clusters. *Curr. Opin. Cell Biol.* **2007**, *19*, 281-289.

[199] G. Ficz*, et al.*, FGF signaling inhibition in ESCs drives rapid genome-wide demethylation to the epigenetic ground state of pluripotency. *Cell Stem Cell* **2013**, *13*, 351-359.

[200] S. Seisenberger*, et al.*, The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. *Mol. Cell* **2012**, *48*, 849-862.

[201] H. J. Lee, T. A. Hore, W. Reik, Reprogramming the methylome: erasing memory and creating diversity. *Cell Stem Cell* **2014**, *14*, 710-719.

[202] Y. Okada*, et al.*, A role for the elongator complex in zygotic paternal genome demethylation. *Nature* **2010**, *463*, 554-558.

[203] M. J. Yebra, A. S. Bhagwat, A cytosine methyltransferase converts 5-methylcytosine in DNA to thymine. *Biochemistry* **1995**, *34*, 14752-14757.

[204] C. C. Chen, K. Y. Wang, C. K. Shen, The mammalian de novo DNA methyltransferases DNMT3A and DNMT3B are also DNA 5-hydroxymethylcytosine dehydroxymethylases. *J. Biol. Chem.* **2012**, *287*, 33116-33121.

[205] Z. Liutkeviciute*, et al.*, Direct decarboxylation of 5-carboxylcytosine by DNA C5-methyltransferases. *J. Am. Chem. Soc.* **2014**, *136*, 5884-5887.

[206] S. Schiesser*, et al.*, Mechanism and stem-cell activity of 5-carboxycytosine decarboxylation determined by isotope tracing. *Angew. Chem. Int. Ed.* **2012**, *51*, 6516-6520.

[207] G. Barreto*, et al.*, Gadd45a promotes epigenetic gene activation by repair-mediated DNA demethylation. *Nature* **2007**, *445*, 671-675.

[208] S. G. Jin, C. Guo, G. P. Pfeifer, GADD45A does not promote DNA demethylation. *PLoS Genet.* **2008**, *4*, e1000013.

[209] D. K. Ma*, et al.*, Neuronal activity-induced Gadd45b promotes epigenetic DNA demethylation and adult neurogenesis. *Science* **2009**, *323*, 1074-1077.

[210] D. P. Gavin*, et al.*, Growth arrest and DNA-damage-inducible, beta (GADD45b)-mediated DNA demethylation in major psychosis. *Neuropsychopharmacol.* **2012**, *37*, 531-542.

[211] F. Matrisciano*, et al.*, Activation of group II metabotropic glutamate receptors promotes DNA demethylation in the mouse brain. *Mol. Pharmacol.* **2011**, *80*, 174-182.

[212] R. P. Zhang, J. Z. Shao, L. X. Xiang, GADD45A protein plays an essential role in active DNA demethylation during terminal osteogenic differentiation of adipose-derived mesenchymal stem cells. *J. Biol. Chem.* **2011**, *286*, 41083-41094.

[213] D. P. Gavin*, et al.*, Gadd45b and N-methyl-D-aspartate induced DNA demethylation in postmitotic neurons. *Epigenomics* **2015**, *7*, 567-579.

[214] P. Rajput, V. Pandey, V. Kumar, Stimulation of ribosomal RNA gene promoter by transcription factor Sp1 involves active DNA demethylation by Gadd45-NER pathway. *Biochim. Biophys. Acta* **2016**, *1859*, 953-963.

[215] S. L. Kigar, L. Chang, A. P. Auger, Gadd45b is an epigenetic regulator of juvenile social behavior and alters local pro-inflammatory cytokine production in the rodent amygdala. *Brain Behav. Immun.* **2015**, *46*, 60-69.

[216] O. Sabag*, et al.*, Establishment of methylation patterns in ES cells. *Nat. Struct. Mol. Biol.* **2014**, *21*, 110-112.

[217] G. L. Sen*, et al.*, DNMT1 maintains progenitor function in self-renewing somatic tissue. *Nature* **2010**, *463*, 563-567.

[218] K. M. Schmitz*, et al.*, TAF12 recruits Gadd45a and the nucleotide excision repair complex to the promoter of rRNA genes leading to active DNA demethylation. *Mol. Cell* **2009**, *33*, 344-353.

[219] N. Le May*, et al.*, XPG and XPF endonucleases trigger chromatin looping and DNA demethylation for accurate expression of activated genes. *Mol. Cell* **2012**, *47*, 622-632.

[220] N. Le May*, et al.*, NER factors are recruited to active promoters and facilitate chromatin modification for transcription in the absence of exogenous genotoxic attack. *Mol. Cell* **2010**, *38*, 54-66.

[221] K. Rai*, et al.*, DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. *Cell* **2008**, *135*, 1201-1212.

[222] S. Kienhofer*, et al.*, GADD45a physically and functionally interacts with TET1. *Differentiation* **2015**, *90*, 59-68.

[223] Z. Li*, et al.*, Gadd45a promotes DNA demethylation through TDG. *Nucleic Acids Res.* **2015**, *43*, 3986-3997.

[224] F. Agius, A. Kapoor, J. K. Zhu, Role of the Arabidopsis DNA glycosylase/lyase ROS1 in active DNA demethylation. *Proc. Natl. Acad. Sci. U S A* **2006**, *103*, 11796-11801.

[225] P. Arnaud, R. Feil, MEDEA takes control of its own imprinting. *Cell* **2006**, *124*, 468-470.

[226] P. Hajkova*, et al.*, Genome-wide reprogramming in the mouse germ line entails the base excision repair pathway. *Science* **2010**, *329*, 78-82.

[227] L. Zhang*, et al.*, Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA. *Nat. Chem. Biol.* **2012**, *8*, 328-330.

[228] I. Grin, A. A. Ishchenko, An interplay of the base excision repair and mismatch repair pathways in active DNA demethylation. *Nucleic Acids Res.* **2016**, *44*, 3713-3727.

[229] A. Masaoka*, et al.*, Mammalian 5-formyluracil-DNA glycosylase. 2. Role of SMUG1 uracil-DNA glycosylase in repair of 5-formyluracil and other oxidized and deaminated base lesions. *Biochemistry* **2003**, *42*, 5003-5012.

[230] U. Hardeland, M. Bentele, J. Jiricny, P. Schar, The versatile thymine DNA-glycosylase: a comparative characterization of the human, Drosophila and fission yeast orthologs. *Nucleic Acids Res.* **2003**, *31*, 2261-2271.

[231] S. Cortellino*, et al.*, Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell* **2011**, *146*, 67-79.

[232] P. Neddermann, J. Jiricny, The purification of a mismatch-specific thymine-DNA glycosylase from HeLa cells. *J. Biol. Chem.* **1993**, *268*, 21218-21224.

[233] A. Maiti, A. C. Drohat, Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. *J. Biol. Chem.* **2011**, *286*, 35334-35338.

[234] P. Liu, A. Burdzy, L. C. Sowers, Repair of the mutagenic DNA oxidation product, 5-formyluracil. *DNA Repair (Amst)* **2003**, *2*, 199-210.

[235] I. Knaevelsrud*, et al.*, Opposite-base dependent excision of 5-formyluracil from DNA by hSMUG1. *Int. J. Radiat. Biol.* **2009**, *85*, 413-420.

[236] D. Cortazar*, et al.*, Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. *Nature* **2011**, *470*, 419-423.

[237]   F. Neri*, et al.*, Single-Base Resolution Analysis of 5-Formyl and 5-Carboxyl Cytosine Reveals Promoter DNA Methylation Dynamics*. Cell Rep.* **2015**.

[238]   X. Lu*, et al.*, Base-resolution maps of 5-formylcytosine and 5-carboxylcytosine reveal genome-wide DNA demethylation dynamics*. Cell Res.* **2015**, *25*, 386-389.

[239]   L. Shen*, et al.*, Genome-wide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics*. Cell* **2013**, *153*, 692-706.

[240]   C. X. Song*, et al.*, Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming*. Cell* **2013**, *153*, 678-691.

[241]   J. P. Jost, M. Siegmann, L. Sun, R. Leung, Mechanisms of DNA demethylation in chicken embryos. Purification and properties of a 5-methylcytosine-DNA glycosylase*. J. Biol. Chem.* **1995**, *270*, 9734-9739.

[242]   B. Zhu*, et al.*, 5-Methylcytosine DNA glycosylase activity is also present in the human MBD4 (G/T mismatch glycosylase) and in a related avian sequence*. Nucleic Acids Res.* **2000**, *28*, 4157-4165.

[243]   U. Sibghat*, et al.*, Base analog and neighboring base effects on substrate specificity of recombinant human G:T mismatch-specific thymine DNA-glycosylase*. Biochemistry* **1996**, *35*, 12926-12932.

[244]   D. Cortazar*, et al.*, The enigmatic thymine DNA glycosylase*. DNA Repair* **2007**, *6*, 489-504.

[245]   S. Cortellino*, et al.*, The base excision repair enzyme MED1 mediates DNA damage response to antitumor drugs and is associated with mismatch repair system integrity*. Proc. Natl. Acad. Sci. U S A* **2003**, *100*, 15071-15076.

[246]   F. Petronzelli*, et al.*, Investigation of the substrate spectrum of the human mismatch-specific DNA N-glycosylase MED1 (MBD4): fundamental role of the catalytic domain*. J. Cell. Physiol.* **2000**, *185*, 473-480.

[247]   C. B. Millar*, et al.*, Enhanced CpG mutability and tumorigenesis in MBD4-deficient mice*. Science* **2002**, *297*, 403-405.

[248]   E. Wong*, et al.*, Mbd4 inactivation increases Cright-arrowT transition mutations and promotes gastrointestinal tumor formation*. Proc. Natl. Acad. Sci. U S A* **2002**, *99*, 14937-14942.

[249]   B. Hendrich, A. Bird, Identification and characterization of a family of mammalian methyl-CpG binding proteins*. Mol. Cell. Biol.* **1998**, *18*, 6538-6547.

[250]   A. Bellacosa*, et al.*, MED1, a novel human methyl-CpG-binding endonuclease, interacts with DNA mismatch repair protein MLH1*. Proc. Natl. Acad. Sci. U S A* **1999**, *96*, 3969-3974.

[251]   T. K. Hazra*, et al.*, Identification and characterization of a human DNA glycosylase for repair of modified bases in oxidatively damaged DNA*. Proc. Natl. Acad. Sci. U S A* **2002**, *99*, 3523-3528.

[252]   I. Morland*, et al.*, Human DNA glycosylases of the bacterial Fpg/MutM superfamily: an alternative pathway for the repair of 8-oxoguanine and other oxidation products in DNA*. Nucleic Acids Res.* **2002**, *30*, 4926-4936.

[253]   T. A. Rosenquist*, et al.*, The novel DNA glycosylase, NEIL1, protects mammalian cells from radiation-mediated cell death*. DNA Repair (Amst)* **2003**, *2*, 581-591.

[254]   H. Dou, S. Mitra, T. K. Hazra, Repair of oxidized bases in DNA bubble structures by human DNA glycosylases NEIL1 and NEIL2*. J. Biol. Chem.* **2003**, *278*, 49679-49684.

[255]   M. Liu*, et al.*, The mouse ortholog of NEIL3 is a functional DNA glycosylase in vitro and in vivo*. Proc. Natl. Acad. Sci. U S A* **2010**, *107*, 4925-4930.

[256]   V. Vartanian*, et al.*, The metabolic syndrome resulting from a knockout of the NEIL1 DNA glycosylase*. Proc. Natl. Acad. Sci. U S A* **2006**, *103*, 1864-1869.

[257]   M. Takao*, et al.*, A back-up glycosylase in Nth1 knock-out mice is a functional Nei (endonuclease VIII) homologue*. J. Biol. Chem.* **2002**, *277*, 42205-42213.

[258]   U. Muller*, et al.*, TET-mediated oxidation of methylcytosine causes TDG or NEIL glycosylase dependent gene reactivation*. Nucleic Acids Res.* **2014**, *42*, 8592-8604.

[259]   L. Schomacher*, et al.*, Neil DNA glycosylases promote substrate turnover by Tdg during DNA demethylation*. Nat. Struct. Mol. Biol.* **2016**, *23*, 116-124.

[260] M. E. Fitzgerald, A. C. Drohat, Coordinating the initial steps of base excision repair. Apurinic/apyrimidinic endonuclease 1 actively stimulates thymine DNA glycosylase by disrupting the product complex. *J. Biol. Chem.* **2008**, *283*, 32680-32690.

[261] S. H. Wilson, T. A. Kunkel, Passing the baton in base excision repair. *Nat. Struct. Biol.* **2000**, *7*, 176-178.

[262] T. Lindahl*, et al.*, DNA N-glycosidases: properties of uracil-DNA glycosidase from Escherichia coli. *J. Biol. Chem.* **1977**, *252*, 3286-3294.

[263] H. Nilsen*, et al.*, Nuclear and mitochondrial uracil-DNA glycosylases are generated by alternative splicing and transcription from different positions in the UNG gene. *Nucleic Acids Res.* **1997**, *25*, 750-755.

[264] K. A. Haushalter, M. W. Todd Stukenberg, M. W. Kirschner, G. L. Verdine, Identification of a new uracil-DNA glycosylase family by expression cloning using synthetic inhibitors. *Curr. Biol.* **1999**, *9*, 174-185.

[265] B. Kavli*, et al.*, hUNG2 is the major repair enzyme for removal of uracil from U:A matches, U:G mismatches, and U in single-stranded DNA, with hSMUG1 as a broad specificity backup. *J. Biol. Chem.* **2002**, *277*, 39926-39936.

[266] M. Otterlei*, et al.*, Post-replicative base excision repair in replication foci. *EMBO J.* **1999**, *18*, 3834-3844.

[267] K. Kemmerich, F. A. Dingler, C. Rada, M. S. Neuberger, Germline ablation of SMUG1 DNA glycosylase causes loss of 5-hydroxymethyluracil- and UNG-backup uracil-excision activities and increases cancer predisposition of Ung-/-Msh2-/- mice. *Nucleic Acids Res.* **2012**, *40*, 6016-6025.

[268] R. J. Boorstein*, et al.*, Definitive identification of mammalian 5-hydroxymethyluracil DNA N-glycosylase activity as SMUG1. *J. Biol. Chem.* **2001**, *276*, 41991-41997.

[269] S. V. Cannon-Carlson, H. Gokhale, G. W. Teebor, Purification and characterization of 5-hydroxymethyluracil-DNA glycosylase from calf thymus. Its possible role in the maintenance of methylated cytosine residues. *J. Biol. Chem.* **1989**, *264*, 13306-13312.

[270] A. Masaoka*, et al.*, Oxidation of thymine to 5-formyluracil in DNA promotes misincorporation of dGMP and subsequent elongation of a mismatched primer terminus by DNA polymerase. *J. Biol. Chem.* **2001**, *276*, 16501-16510.

[271] A. Darwanto, A. Farrel, D. K. Rogstad, L. C. Sowers, Characterization of DNA glycosylase activity by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Anal. Biochem.* **2009**, *394*, 13-23.

[272] Q. M. Zhang*, et al.*, Replication of DNA templates containing 5-formyluracil, a major oxidative lesion of thymine in DNA. *Nucleic Acids Res.* **1997**, *25*, 3969-3973.

[273] D. Schuermann, A. R. Weber, P. Schar, Active DNA demethylation by DNA repair: Facts and uncertainties. *DNA Repair (Amst)* **2016**, *44*, 92-102.

[274] T. Pfaffeneder*, et al.*, Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA. *Nat. Chem. Biol.* **2014**, *10*, 574-581.

[275] P. Nygaard, On the role of cytidine deaminase in cellular metabolism. *Adv. Exp. Med. Biol.* **1986**, *195 Pt B*, 415-420.

[276] A. Marx, A. Alian, The first crystal structure of a dTTP-bound deoxycytidylate deaminase validates and details the allosteric-inhibitor binding site. *J. Biol. Chem.* **2015**, *290*, 682-690.

[277] S. Vincenzetti*, et al.*, Site directed mutagenesis as a tool to understand the catalytic mechanism of human cytidine deaminase. *Protein Peptide Lett.* **2013**, *20*, 538-549.

[278] A. Jekunen, J. A. Vilpo, 5-Methyl-2'-deoxycytidine. Metabolism and effects on cell lethality studied with human leukemic cells in vitro. *Mol. Pharmacol.* **1984**, *25*, 431-435.

[279] J. A. Vilpo, L. M. Vilpo, Biochemical mechanisms by which reutilization of DNA 5-methylcytosine is prevented in human cells. *Mutat. Res.* **1991**, *256*, 29-35.

[280] J. A. Vilpo, L. M. Vilpo, Nucleoside monophosphate kinase may be the key enzyme preventing salvage of DNA 5-methylcytosine. *Mutat. Res.* **1993**, *286*, 217-220.

[281] R. Holliday, T. Ho, Evidence for gene silencing by endogenous DNA methylation. *Proc. Natl. Acad. Sci. U S A* **1998**, *95*, 8727-8732.

[282] M. Zauri*, et al.*, CDA directs metabolism of epigenetic nucleosides revealing a therapeutic window in cancer. *Nature* **2015**, *524*, 114-118.

[283] M. Tallis, R. Morra, E. Barkauskaite, I. Ahel, Poly(ADP-ribosyl)ation in regulation of chromatin structure and the DNA damage response. *Chromosoma* **2014**, *123*, 79-90.

[284] S. Gemble*, et al.*, Pyrimidine Pool Disequilibrium Induced by a Cytidine Deaminase Deficiency Inhibits PARP-1 Activity, Leading to the Under Replication of DNA. *PLoS Genet.* **2015**, *11*, e1005384.

[285] P. Chabosseau*, et al.*, Pyrimidine pool imbalance induced by BLM helicase deficiency contributes to genetic instability in Bloom syndrome. *Nat. Commun.* **2011**, *2*, 368.

[286] A. Sanchez*, et al.*, Replication fork collapse and genome instability in a deoxycytidylate deaminase mutant. *Mol. Cell. Biol.* **2012**, *32*, 4445-4454.

[287] S. U. Siriwardena, K. Chen, A. S. Bhagwat, Functions and Malfunctions of Mammalian DNA-Cytosine Deaminases. *Chem. Rev.* **2016**, *116*, 12688-12710.

[288] H. Arakawa, J. Hauschild, J. M. Buerstedde, Requirement of the activation-induced deaminase (AID) gene for immunoglobulin gene conversion. *Science* **2002**, *295*, 1301-1306.

[289] H. Huthoff, M. H. Malim, Cytidine deamination and resistance to retroviral infection: towards a structural understanding of the APOBEC proteins. *Virology* **2005**, *334*, 147-153.

[290] A. P. Gerber, W. Keller, RNA editing by base deamination: more enzymes, more targets, new mysteries. *Trends Biochem. Sci.* **2001**, *26*, 376-384.

[291] S. Chen*, et al.*, APOBEC3A possesses anticancer and antiviral effects by differential inhibition of HPV E6 and E7 expression on cervical cancer. *Int. J. Clin. Exp. Med.* **2015**, *8*, 10548-10557.

[292] J. F. Krisko*, et al.*, APOBEC3G and APOBEC3F Act in Concert To Extinguish HIV-1 Replication. *J. Virol.* **2016**, *90*, 4681-4695.

[293] M. M. Ahasan*, et al.*, APOBEC3A and 3C decrease human papillomavirus 16 pseudovirion infectivity. *Biochem. Biophys. Res. Commun.* **2015**, *457*, 295-299.

[294] M. E. Olson, R. S. Harris, D. A. Harki, APOBEC Enzymes as Targets for Virus and Cancer Therapy. *Cell Chem. Biol.* **2018**, *25*, 36-49.

[295] J. I. Hoopes*, et al.*, APOBEC3A and APOBEC3B Preferentially Deaminate the Lagging Strand Template during DNA Replication. *Cell Rep.* **2016**, *14*, 1273-1282.

[296] H. Chen*, et al.*, APOBEC3A is a potent inhibitor of adeno-associated virus and retrotransposons. *Curr. Biol.* **2006**, *16*, 480-485.

[297] T. Kouno*, et al.*, Crystal structure of APOBEC3A bound to single-stranded DNA reveals structural basis for cytidine deamination and specificity. *Nat. Commun.* **2017**, *8*, 15024.

[298] H. D. Morgan*, et al.*, Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: implications for epigenetic reprogramming. *J. Biol. Chem.* **2004**, *279*, 52353-52360.

[299] M. A. Carpenter*, et al.*, Methylcytosine and normal cytosine deamination by the foreign DNA restriction enzyme APOBEC3A. *J. Biol. Chem.* **2012**, *287*, 34801-34808.

[300] E. K. Schutsky*, et al.*, APOBEC3A efficiently deaminates methylated, but not TET-oxidized, cytosine bases in DNA. *Nucleic Acids Res.* **2017**, *45*, 7655-7665.

[301] P. Wijesinghe, A. S. Bhagwat, Efficient deamination of 5-methylcytosines in DNA by human APOBEC3A, but not by AID or APOBEC3G. *Nucleic Acids Res.* **2012**, *40*, 9206-9217.

[302] J. U. Guo*, et al.*, Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell* **2011**, *145*, 423-434.

[303] G. Rangam, K. M. Schmitz, A. J. Cobb, S. K. Petersen-Mahrt, AID enzymatic activity is inversely proportional to the size of cytosine C5 orbital cloud. *PLoS One* **2012**, *7*, e43279.

[304] C. S. Nabel*, et al.*, AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. *Nat. Chem. Biol.* **2012**, *8*, 751-758.

[305] M. A. Surani, S. C. Barton, M. L. Norris, Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. *Nature* **1984**, *308*, 548-550.

[306]   J. McGrath, D. Solter, Completion of mouse embryogenesis requires both the maternal and paternal genomes. *Cell* **1984**, *37*, 179-183.

[307]   D. P. Barlow*, et al.*, The mouse insulin-like growth factor type-2 receptor is imprinted and closely linked to the Tme locus. *Nature* **1991**, *349*, 84-87.

[308]   T. M. DeChiara, E. J. Robertson, A. Efstratiadis, Parental imprinting of the mouse insulin-like growth factor II gene. *Cell* **1991**, *64*, 849-859.

[309]   A. C. Ferguson-Smith*, et al.*, Embryological and molecular investigations of parental imprinting on mouse chromosome 7. *Nature* **1991**, *351*, 667-670.

[310]   M. S. Bartolomei, S. Zemel, S. M. Tilghman, Parental imprinting of the mouse H19 gene. *Nature* **1991**, *351*, 153-155.

[311]   Y. Kato*, et al.*, Role of the Dnmt3 family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse. *Hum Mol Genet* **2007**, *16*, 2272-2280.

[312]   T. Yokomine, K. Hata, M. Tsudzuki, H. Sasaki, Evolution of the vertebrate DNMT3 gene family: a possible link between existence of DNMT3L and genomic imprinting. *Cytogenet Genome Res* **2006**, *113*, 75-80.

[313]   F. Chedin, M. R. Lieber, C. L. Hsieh, The DNA methyltransferase-like protein DNMT3L stimulates de novo methylation by Dnmt3a. *Proc Natl Acad Sci U S A* **2002**, *99*, 16916-16921.

[314]   D. Bourc'his*, et al.*, Dnmt3L and the establishment of maternal genomic imprints. *Science* **2001**, *294*, 2536-2539.

[315]   Y. Stelzer*, et al.*, Parent-of-Origin DNA Methylation Dynamics during Mouse Development. *Cell Rep.* **2016**, *16*, 3167-3180.

[316]   E. Li, C. Beard, R. Jaenisch, Role for DNA methylation in genomic imprinting. *Nature* **1993**, *366*, 362-365.

[317]   C. Y. Howell*, et al.*, Genomic imprinting disrupted by a maternal effect mutation in the Dnmt1 gene. *Cell* **2001**, *104*, 829-838.

[318]   L. B. Wan, M. S. Bartolomei, Regulation of imprinting in clusters: noncoding RNAs versus insulators. *Adv. Genet.* **2008**, *61*, 207-223.

[319]   F. Sleutels, R. Zwart, D. P. Barlow, The non-coding Air RNA is required for silencing autosomal imprinted genes. *Nature* **2002**, *415*, 810-813.

[320]   H. Kobayashi*, et al.*, Bisulfite sequencing and dinucleotide content analysis of 15 imprinted mouse differentially methylated regions (DMRs): paternally methylated DMRs contain less CpGs than maternally methylated DMRs. *Cytogenet. Genome Res.* **2006**, *113*, 130-137.

[321]   S. P. Lin*, et al.*, Asymmetric regulation of imprinting on the maternal and paternal chromosomes at the Dlk1-Gtl2 imprinted cluster on mouse chromosome 12. *Nat. Genet.* **2003**, *35*, 97-102.

[322]   H. Seitz*, et al.*, Imprinted microRNA genes transcribed antisense to a reciprocally imprinted retrotransposon-like gene. *Nat. Genet.* **2003**, *34*, 261-262.

[323]   N. E. Cockett*, et al.*, Polar overdominance at the ovine callipyge locus. *Science* **1996**, *273*, 236-238.

[324]   I. Sanli, R. Feil, Chromatin mechanisms in the developmental control of imprinted gene expression. *Int. J. Biochem. Cell Biol.* **2015**, *67*, 139-147.

[325]   R. Hirasawa, R. Feil, Genomic imprinting and human disease. *Essays Biochem.* **2010**, *48*, 187-200.

[326]   M. G. Butler, Genomic imprinting disorders in humans: a mini-review. *J. Assist. Reprod. Gen.* **2009**, *26*, 477-486.

[327]   S. B. Cassidy, S. Schwartz, J. L. Miller, D. J. Driscoll, Prader-Willi syndrome. *Genet. Med.* **2012**, *14*, 10-26.

[328]   S. S. Margolis, G. L. Sell, M. A. Zbinden, L. M. Bird, Angelman Syndrome. *Neurotherapeutics* **2015**, *12*, 641-650.

[329] T. Eggermann, K. Eggermann, N. Schonherr, Growth retardation versus overgrowth: Silver-Russell syndrome is genetically opposite to Beckwith-Wiedemann syndrome. *Trends Genet.* **2008**, *24*, 195-204.

[330] S. Manipalviratn, A. DeCherney, J. Segars, Imprinting disorders and assisted reproductive technology. *Fertil. Steril.* **2009**, *91*, 305-315.

[331] G. F. Cox*, et al.*, Intracytoplasmic sperm injection may increase the risk of imprinting defects. *Am. J. Hum. Genet.* **2002**, *71*, 162-164.

[332] M. R. DeBaun, E. L. Niemitz, A. P. Feinberg, Association of in vitro fertilization with Beckwith-Wiedemann syndrome and epigenetic alterations of LIT1 and H19. *Am. J. Hum. Genet.* **2003**, *72*, 156-160.

[333] J. A. Bokar*, et al.*, Purification and cDNA cloning of the AdoMet-binding subunit of the human mRNA (N6-adenosine)-methyltransferase. *RNA* **1997**, *3*, 1233-1247.

[334] Z. Bodi, J. D. Button, D. Grierson, R. G. Fray, Yeast targets for mRNA methylation. *Nucleic Acids Res.* **2010**, *38*, 5327-5335.

[335] G. Jia*, et al.*, N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat. Chem. Biol.* **2011**, *7*, 885-887.

[336] X. Zhang*, et al.*, Structural insights into FTO's catalytic mechanism for the demethylation of multiple RNA substrates. *Proc. Natl. Acad. Sci. U S A* **2019**.

[337] G. Zheng*, et al.*, Sprouts of RNA epigenetics: the discovery of mammalian RNA demethylases. *RNA Biol.* **2013**, *10*, 915-918.

[338] X. Wang*, et al.*, N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* **2014**, *505*, 117-120.

[339] L. Fu*, et al.*, Tet-mediated formation of 5-hydroxymethylcytosine in RNA. *J. Am. Chem. Soc.* **2014**, *136*, 11582-11585.

[340] M. Basanta-Sanchez*, et al.*, TET1-Mediated Oxidation of 5-Formylcytosine (5fC) to 5-Carboxycytosine (5caC) in RNA. *Chembiochem* **2017**, *18*, 72-76.

[341] L. Kawarada*, et al.*, ALKBH1 is an RNA dioxygenase responsible for cytoplasmic and mitochondrial tRNA modifications. *Nucleic Acids Res.* **2017**, *45*, 7401-7415.

[342] K. Biemann*, et al.*, in *Angew. Chem. Int. Ed., Vol. 78*, **1966**, pp. 600-601.

[343] H. Grosjean, M. Sprinzl, S. Steinberg, Posttranscriptionally modified nucleosides in transfer RNA: their locations and frequencies. *Biochimie* **1995**, *77*, 139-141.

[344] I. Diaz, M. Ehrenberg, ms2i6A deficiency enhances proofreading in translation. *J. Mol. Biol.* **1991**, *222*, 1161-1171.

[345] R. K. Wilson, B. A. Roe, Presence of the hypermodified nucleotide N6-(delta 2-isopentenyl)-2-methylthioadenosine prevents codon misreading by Escherichia coli phenylalanyl-transfer RNA. *Proc. Natl. Acad. Sci. U S A* **1989**, *86*, 409-413.

[346] L. B. Jenner, N. Demeshkina, G. Yusupova, M. Yusupov, Structural aspects of messenger RNA reading frame maintenance by the ribosome. *Nat. Struct. Mol. Biol.* **2010**, *17*, 555-560.

[347] M. L. Gefter, R. L. Russell, Role modifications in tyrosine transfer RNA: a modified base affecting ribosome binding. *J. Mol. Biol.* **1969**, *39*, 145-157.

[348] B. Esberg, G. R. Bjork, The methylthio group (ms2) of N6-(4-hydroxyisopentenyl)-2-methylthioadenosine (ms2io6A) present next to the anticodon contributes to the decoding efficiency of the tRNA. *J. Bacteriol.* **1995**, *177*, 1967-1975.

[349] H. C. Leung, Y. Chen, M. E. Winkler, Regulation of substrate recognition by the MiaA tRNA prenyltransferase modification enzyme of Escherichia coli K-12. *J. Biol. Chem.* **1997**, *272*, 13073-13083.

[350] T. H. Tsang, M. Buck, B. N. Ames, Sequence specificity of tRNA-modifying enzymes. An analysis of 258 tRNA sequences. *Biochim. Biophys. Acta* **1983**, *741*, 180-196.

[351] D. M. Connolly, M. E. Winkler, Genetic and physiological relationships among the miaA gene, 2-methylthio-N6-(delta 2-isopentenyl)-adenosine tRNA modification, and spontaneous mutagenesis in Escherichia coli K-12. *J. Bacteriol.* **1989**, *171*, 3233-3246.

[352] J. W. Yarham*, et al.*, Defective i6A37 modification of mitochondrial and cytosolic tRNAs results from pathogenic mutations in TRIT1 and its substrate tRNA*. PLoS Genet.* **2014**, *10*, e1004424.

[353] M. Spinola*, et al.*, Identification and functional characterization of the candidate tumor suppressor gene TRIT1 in human lung cancer*. Oncogene* **2005**, *24*, 5502-5509.

[354] M. Pratt-Hyatt*, et al.*, Mod5 protein binds to tRNA gene complexes and affects local transcriptional silencing*. Proc. Natl. Acad. Sci. U S A* **2013**, *110*, E3081-3089.

[355] T. N. Lamichhane*, et al.*, Lack of tRNA modification isopentenyl-A37 alters mRNA decoding and causes metabolic deficiencies in fission yeast*. Mol. Cell. Biol.* **2013**, *33*, 2918-2929.

[356] T. Carell*, et al.*, Structure and function of noncanonical nucleobases*. Angew. Chem. Int. Ed.* **2012**, *51*, 7110-7131.

[357] V. Reiter*, et al.*, The CDK5 repressor CDK5RAP1 is a methylthiotransferase acting on nuclear and mitochondrial RNA*. Nucleic Acids Res.* **2012**, *40*, 6235-6240.

[358] F. Y. Wei*, et al.*, Cdk5rap1-mediated 2-methylthio modification of mitochondrial tRNAs governs protein translation and contributes to myopathy in mice and humans*. Cell Metab.* **2015**, *21*, 428-442.

[359] H. Wang*, et al.*, CDK5RAP1 deficiency induces cell cycle arrest and apoptosis in human breast cancer cell line by the ROS/JNK signaling pathway*. Oncol. Rep.* **2015**, *33*, 1089-1096.

[360] M. K. Shin*, et al.*, Association between CDK5RAP1 polymorphisms and susceptibility to vitiligo in the Korean population*. Eur. J. Dermatol.* **2012**, *22*, 495-499.

[361] B. I. Kang*, et al.*, Identification of 2-methylthio cyclic N6-threonylcarbamoyladenosine (ms2ct6A) as a novel RNA modification at position 37 of tRNAs*. Nucleic Acids Res.* **2017**, *45*, 2124-2136.

[362] M. Matuszewski*, et al.*, A hydantoin isoform of cyclic N6-threonylcarbamoyladenosine (ct6A) is present in tRNAs*. Nucleic Acids Res.* **2017**, *45*, 2137-2149.

[363] S. Arragain*, et al.*, Identification of eukaryotic and prokaryotic methylthiotransferase for biosynthesis of 2-methylthio-N6-threonylcarbamoyladenosine in tRNA*. J. Biol. Chem.* **2010**, *285*, 28425-28433.

[364] S. Kimura*, et al.*, Discovery of the beta-barrel-type RNA methyltransferase responsible for N6-methylation of N6-threonylcarbamoyladenosine in tRNAs*. Nucleic Acids Res.* **2014**, *42*, 9350-9365.

[365] K. Miyauchi, S. Kimura, T. Suzuki, A cyclic form of N6-threonylcarbamoyladenosine as a widely distributed tRNA hypermodification*. Nat. Chem. Biol.* **2013**, *9*, 105-111.

[366] B. El Yacoubi, M. Bailly, V. de Crecy-Lagard, Biosynthesis and function of posttranscriptional modifications of transfer RNAs*. Annu. Rev. Genet.* **2012**, *46*, 69-95.

[367] F. V. t. Murphy, V. Ramakrishnan, A. Malkiewicz, P. F. Agris, The role of modifications in codon discrimination by tRNA(Lys)UUU*. Nat. Struct. Mol. Biol.* **2004**, *11*, 1186-1191.

[368] S. S. Phelps, A. Malkiewicz, P. F. Agris, S. Joseph, Modified nucleotides in tRNA(Lys) and tRNA(Val) are important for translocation*. J. Mol. Biol.* **2004**, *338*, 439-444.

[369] T. Niimi*, et al.*, Recognition of the Anticodon Loop of tRNA^Ile 1 by Isoleucyl-tRNA Synthetase from Escherichia coli*. Nucleos. Nucleot.* **1994**, *13*, 1231-1237.

[370] C. A. Lin, S. R. Ellis, H. L. True, The Sua5 protein is essential for normal translational regulation in yeast*. Mol Cell. Biol* **2010**, *30*, 354-363.

[371] T. Suzuki, T. Suzuki, A complete landscape of post-transcriptional modifications in mammalian mitochondrial tRNAs*. Nucleic Acids Res.* **2014**, *42*, 7346-7357.

[372] F. Y. Wei*, et al.*, Deficit of tRNA(Lys) modification by Cdkal1 causes the development of type 2 diabetes in mice*. J. Clin. Invest.* **2011**, *121*, 3598-3608.

[373] S. Brambillasca*, et al.*, CDK5 regulatory subunit-associated protein 1-like 1 (CDKAL1) is a tail-anchored protein in the endoplasmic reticulum (ER) of insulinoma cells*. J. Biol. Chem.* **2012**, *287*, 41808-41819.

[374] S. Schiffers*, et al.*, Quantitative LC-MS Provides No Evidence for m(6) dA or m(4) dC in the Genome of Mouse Embryonic Stem Cells and Tissues*. Angew. Chem. Int. Ed.* **2017**, *56*, 11268-11271.

[375] B. S. Zhao*, et al.*, m(6)A-dependent maternal mRNA clearance facilitates zebrafish maternal-to-zygotic transition*. Nature* **2017**, *542*, 475-478.

[376] M. P. Charles*, et al.*, N(6)-Methyldeoxyadenosine, a nucleoside commonly found in prokaryotes, induces C2C12 myogenic differentiation*. Biochem. Biophys. Res. Commun.* **2004**, *314*, 476-482.

[377] D. Ratel*, et al.*, The bacterial nucleoside N(6)-methyldeoxyadenosine induces the differentiation of mammalian tumor cells*. Biochem. Biophys. Res. Commun.* **2001**, *285*, 800-805.

[378] S. Gajovic, L. St-Onge, Y. Yokota, P. Gruss, Retinoic acid mediates Pax6 expression during in vitro differentiation of embryonic stem cells*. Differentiation* **1997**, *62*, 187-192.

[379] S. H. Orkin, K. Hochedlinger, Chromatin connections to pluripotency and cellular reprogramming*. Cell* **2011**, *145*, 835-850.

[380] J. Silva*, et al.*, Promotion of reprogramming to ground state pluripotency by signal inhibition*. PLoS Biol.* **2008**, *6*, e253.

[381] M. H. Rosner*, et al.*, A POU-domain transcription factor in early stem cells and germ cells of the mammalian embryo*. Nature* **1990**, *345*, 686-692.

[382] M. Yoshida, M. Kijima, M. Akita, T. Beppu, Potent and specific inhibition of mammalian histone deacetylase both in vivo and in vitro by trichostatin A*. J. Biol. Chem.* **1990**, *265*, 17174-17179.

[383] S. Imai, C. M. Armstrong, M. Kaeberlein, L. Guarente, Transcriptional silencing and longevity protein Sir2 is an NAD-dependent histone deacetylase*. Nature* **2000**, *403*, 795-800.

[384] S. Sharma*, et al.*, Acetylation-Dependent Control of Global Poly(A) RNA Degradation by CBP/p300 and HDAC1/2*. Mol. Cell* **2016**, *63*, 927-938.

[385] L. Zare, H. Baharvand, M. Javan, In vivo conversion of astrocytes to oligodendrocyte lineage cells using chemicals: targeting gliosis for myelin repair*. Regen. Med.* **2018**, *13*, 803-819.

[386] C. G. D. Silva*, et al.*, Use of trichostatin A alters the expression of HDAC3 and KAT2 and improves in vitro development of bovine embryos cloned using less methylated mesenchymal stem cells*. Reprod. Domest. Anim.* **2019**, *54*, 289-299.

[387] M. Moser*, et al.*, Kindlin-3 is essential for integrin activation and platelet aggregation*. Nat. Med.* **2008**, *14*, 325-330.

[388] Q. L. Ying*, et al.*, The ground state of embryonic stem cell self-renewal*. Nature* **2008**, *453*, 519-523.

[389] A. G. Smith*, et al.*, Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides*. Nature* **1988**, *336*, 688-690.

[390] X. Yue, L. Wu, W. Hu, The regulation of leukemia inhibitory factor*. Cancer Cell Microenviron.* **2015**, *2*.

[391] Y. J. Sim*, et al.*, 2i Maintains a Naive Ground State in ESCs through Two Distinct Epigenetic Mechanisms*. Stem Cell Rep.* **2017**, *8*, 1312-1328.

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

[393] H. G. Leitch*, et al.*, Naive pluripotency is associated with global DNA hypomethylation*. Nat. Struct. Mol. Biol.* **2013**, *20*, 311-316.

[394] F. Gaudet*, et al.*, Induction of tumors in mice by genomic hypomethylation*. Science* **2003**, *300*, 489-492.

[395] R. Z. Chen*, et al.*, DNA hypomethylation leads to elevated mutation rates*. Nature* **1998**, *395*, 89-93.

[396] F. von Meyenn*, et al.*, Impairment of DNA Methylation Maintenance Is the Main Cause of Global Demethylation in Naive Embryonic Stem Cells*. Mol. Cell* **2016**, *62*, 983.

[397] J. Choi*, et al.*, DUSP9 Modulates DNA Hypomethylation in Female Mouse Pluripotent Stem Cells*. Cell Stem Cell* **2017**, *20*, 706-719 e707.

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

[399] M. Yagi*, et al.*, Derivation of ground-state female ES cells maintaining gamete-derived DNA methylation*. Nature* **2017**, *548*, 224-227.

[400] T. M. Holm*, et al.*, Global loss of imprinting leads to widespread tumorigenesis in adult mice*. Cancer Cell* **2005**, *8*, 275-285.

[401] K. L. Tucker*, et al.*, Germ-line passage is required for establishment of methylation and expression patterns of imprinted but not of nonimprinted genes*. Genes Dev.* **1996**, *10*, 1008-1020.

[402] T. W. Theunissen*, et al.*, Molecular Criteria for Defining the Naive Human Pluripotent State*. Cell Stem Cell* **2016**, *19*, 502-515.

[403] R. N. Plasschaert, M. S. Bartolomei, Genomic imprinting in development, growth, behavior and stem cells*. Development* **2014**, *141*, 1805-1813.

[404] M. Missbach*, et al.*, A novel inhibitor of the tyrosine kinase Src suppresses phosphorylation of its major cellular substrates and reduces bone resorption in vitro and in rodent models in vivo*. Bone* **1999**, *24*, 437-449.

[405] T. Shimizu*, et al.*, Dual inhibition of Src and GSK3 maintains mouse embryonic stem cells, whose differentiation is mechanically regulated by Src signaling*. Stem Cells* **2012**, *30*, 1394-1404.

[406] Z. D. Smith*, et al.*, A unique regulatory phase of DNA methylation in the early mammalian embryo*. Nature* **2012**, *484*, 339-344.

[407] B. Chen*, et al.*, Small molecule-mediated disruption of Wnt-dependent signaling in tissue regeneration and cancer*. Nat. Chem. Biol.* **2009**, *5*, 100-107.

[408] P. Liscio*, et al.*, Design, synthesis, crystallographic studies, and preliminary biological appraisal of new substituted triazolo[4,3-b]pyridazin-8-amine derivatives as tankyrase inhibitors*. J. Med. Chem.* **2014**, *57*, 2807-2812.

[409] J. Lu*, et al.*, Structure-activity relationship studies of small-molecule inhibitors of Wnt response*. Bioorg. Med. Chem. Lett.* **2009**, *19*, 3825-3827.

[410] R. Rahimoff*, et al.*, 5-Formyl- and 5-Carboxydeoxycytidines Do Not Cause Accumulation of Harmful Repair Intermediates in Stem Cells*. J. Am. Chem. Soc.* **2017**, *139*, 10359-10364.

[411] S. Xie*, et al.*, Cloning, expression and chromosome locations of the human DNMT3 gene family*. Gene* **1999**, *236*, 87-95.

[412] F. R. Traube*, et al.*, Isotope-dilution mass spectrometry for exact quantification of noncanonical DNA nucleosides*. Nat. Protoc.* **2019**, *14*, 283-312.

[413] W. Blum*, et al.*, Phase I study of decitabine alone or in combination with valproic acid in acute myeloid leukemia*. J. Clin. Oncol.* **2007**, *25*, 3884-3891.

[414] A. F. Cashen*, et al.*, Pharmacokinetics of decitabine administered as a 3-h infusion to patients with acute myeloid leukemia (AML) or myelodysplastic syndrome (MDS)*. Cancer Chemoth. Pharm.* **2008**, *61*, 759-766.

[415] G. Marcucci*, et al.*, Bioavailability of azacitidine subcutaneous versus intravenous in patients with the myelodysplastic syndromes*. J. Clin. Pharmacol.* **2005**, *45*, 597-602.

[416] J. Oberto*, et al.*, Qri7/OSGEPL, the mitochondrial version of the universal Kae1/YgjD protein, is essential for mitochondrial genome maintenance*. Nucleic Acids Res.* **2009**, *37*, 5343-5352.

[417] T. T. Ensfelder*, et al.*, ALKBH5-induced demethylation of mono- and dimethylated adenosine*. Chem. Commun.* **2018**.

[418] E. Li, T. H. Bestor, R. Jaenisch, Targeted mutation of the DNA methyltransferase gene results in embryonic lethality*. Cell* **1992**, *69*, 915-926.

[419] M. Okano, S. Xie, E. Li, Dnmt2 is not required for de novo and maintenance methylation of viral DNA in embryonic stem cells*. Nucleic Acids Res.* **1998**, *26*, 2536-2540.

[420] D. Globisch*, et al.*, Systems-Based Analysis of Modified tRNA Bases*. Angew. Chem. Int. Ed.* **2011**, *50*, 9739-9742.

[421] K. G. Steube*, et al.*, A model system in haematology and immunology: the human monocytic cell line MONO-MAC-1*. Leuk. Res.* **1997**, *21*, 327-335.

[422] S. J. Collins, R. C. Gallo, R. E. Gallagher, Continuous growth and differentiation of human myeloid leukaemic cells in suspension culture*. Nature* **1977**, *270*, 347-349.

[423]   C. Sundstrom, K. Nilsson, Establishment and characterization of a human histiocytic lymphoma cell line (U-937). *Int. J. Cancer* **1976**, *17*, 565-577.

[424]   H. Asou*, et al.*, Establishment of a human acute myeloid leukemia cell line (Kasumi-1) with 8;21 chromosome translocation. *Blood* **1991**, *77*, 2031-2036.

[425]   C. B. Lozzio, B. B. Lozzio, Cytotoxicity of a factor isolated from human spleen. *J. Natl. Cancer Inst.* **1973**, *50*, 535-538.

[426]   S. Tsuchiya*, et al.*, Establishment and characterization of a human acute monocytic leukemia cell line (THP-1). *Int. J. Cancer* **1980**, *26*, 171-176.

[427]   H. P. Koeffler, D. W. Golde, Acute myelogenous leukemia: a human cell line responsive to colony-stimulating activity. *Science* **1978**, *200*, 1153-1154.

[428]   H. W. Ziegler-Heitbrock*, et al.*, Establishment of a human cell line (Mono Mac 6) with characteristics of mature monocytes. *Int. J. Cancer* **1988**, *41*, 456-461.

[429]   B. Lange*, et al.*, Growth factor requirements of childhood acute leukemia: establishment of GM-CSF-dependent cell lines. *Blood* **1987**, *70*, 192-199.

[430]   Y. Matsuo*, et al.*, Two acute monocytic leukemia (AML-M5a) cell lines (MOLM-13 and MOLM-14) with interclonal phenotypic heterogeneity showing MLL-AF9 fusion resulting from an occult chromosome insertion, ins(11;9)(q23;p22p23). *Leukemia* **1997**, *11*, 1469-1477.

[431]   M. Lanotte*, et al.*, NB4, a maturation inducible cell line with t(15;17) marker isolated from a human acute promyelocytic leukemia (M3). *Blood* **1991**, *77*, 1080-1086.

Appendix

Table SI-1: Overview over all specification of the PDX samples. Marked in black are all samples of which the DNA or RNA has been analyzed, marked in blue are samples that did not give any results.

Sample #	Patient #	Mouse #	Cell origin	Cell # begin-ning [Mio]	Cell # end [Mio]	Specialties	AzaC conc.	Treat-ment immediate?	Treat-ment length
1, 2	393	11036	Spleen	10	15	After 48 h grew to 15 mio --> readjusted to 10 Mio.	1 µM	Yes	24h
					N.d.				72h
3, 4	393	12393	Spleen	10	N. d.	Many erythrocytes	1 µM	Yes	24h
									48h
5	393	14131	Spleen	10	N. d.		0.1 µM	Yes	24h
6	346	12906	Spleen	10	N. d.	Cells grew attached to flask	0.1 µM	No, after 24h	24h
7	393	16127	Spleen	14.2 (live, 64%)	16.3 (live, 66%)		1 µM	Yes	24h
8	393	15869	Bone marrow	23.3 (live, 39%)	14.1 (live, 32%)	Bone marrow: cells are flaking and look like fibers; some erythrocytes	1 µM	After 4h	48h
9	669	14933	Spleen	13	13.8		1 µM	After 20h	24h
10	372	14884	Spleen	10	8.84 (live, 21%)		1 µM	Yes	24h
11	372	14885	Spleen	10	10.5 (live, 57%)		1 µM	Yes	24h
12, 13	485		Spleen	7.25	8.65 (live, 28%)	Mostly erythrocytes	1 µM	Yes	24h
					N. d.				24h
14, 15	485		Spleen	7.25	N. d.	Many erythrocytes; Treatment only at 0h, not after 24h --> chase	1 µM	Yes	24h + 24h Chase
									24h + 24h Chase

Table SI-2: Overview over the cell lines used in a screening for the incorporation and response of AzadC and AzaC.

Cell line	Origin	Tissue	Cell type	Reference
MONO-MAC-1 (MM1)	human	peripheral blood	B lymphoblast	[421]
HL-60	human	peripheral blood	promyeloblast	[422]
U-937	human	pleural effusion	hystiocytic lymphoma	[423]
Kasumi-1	human	peripheral blood	myeloblast	[424]
K562	human	bone marrow	lymphoblast	[425]
THP-1	human	peripheral blood	leukemic monocyte	[426]
KG1α	human	bone marrow	promyeloblast, macrophage	[427]
MONO-MAC-6 (MM6)	human		acute monocytic leukemia	[428]
MV4-11	human	peripheral blood	macrophage (lymphoblast)	[429]
MOLM-13	human		acute myeloid leukemia	[430]
NB4	human	bone marrow	acute promyelocytic leukemia	[431]

The parent cell line, MONO-MAC-1[421] (sister to MONO-MAC-6[428]), was established from peripheral blood of a multiple myeloma patient who had become resistant to steroid-based therapy.

HL-60[422] is a promyelocytic cell line. Peripheral blood leukocytes were obtained by leukopheresis from a 36-year-old Caucasian female with acute promyelocytic leukemia.

U-937[423] was established in in 1974 from the pleural effusion of a 37-year-old man with generalized diffuse histiocytic lymphoma.

The Kasumi-1[424] cell line was established from the peripheral blood of an acute myeloid leukemia (AML) patient.

The continuous cell line K562[425] was established from the pleural effusion of a 53-year-old female with chronic myelogenous leukemia in terminal blast crises.

The THP-1[426] cell line was derived from the peripheral blood of a 1 year old man with acute monocytic leukaemia.

The variant subline KG-1α of the human acute myelogenous leukemia cell line KG-1 was isolated by *H.P. Koeffler, et al*.[427]

The cell line MONO-MAC-6 was established from the peripheral blood of a 64-year-old male with relapsed acute monocytic leukaemia (AML FAB M5) in 1985 following myeloid metaplasia.

The cell line MV4-11[429] was established from a 10-year-old boy with acute monocytic leukemia (AML FAB M5) at diagnosis.

The cell line MOLM-13[430] was established from the peripheral blood of a 20-year-old man with acute myeloid leukemia AML FAB M5a at relapse in 1995 after initial myelodysplastic syndromes.

NB4[431] cell line was established from the bone marrow of a 23-year-old woman with acute promyelocytic leukemia (APL = AML FAB M3) in second relapse in 1989.
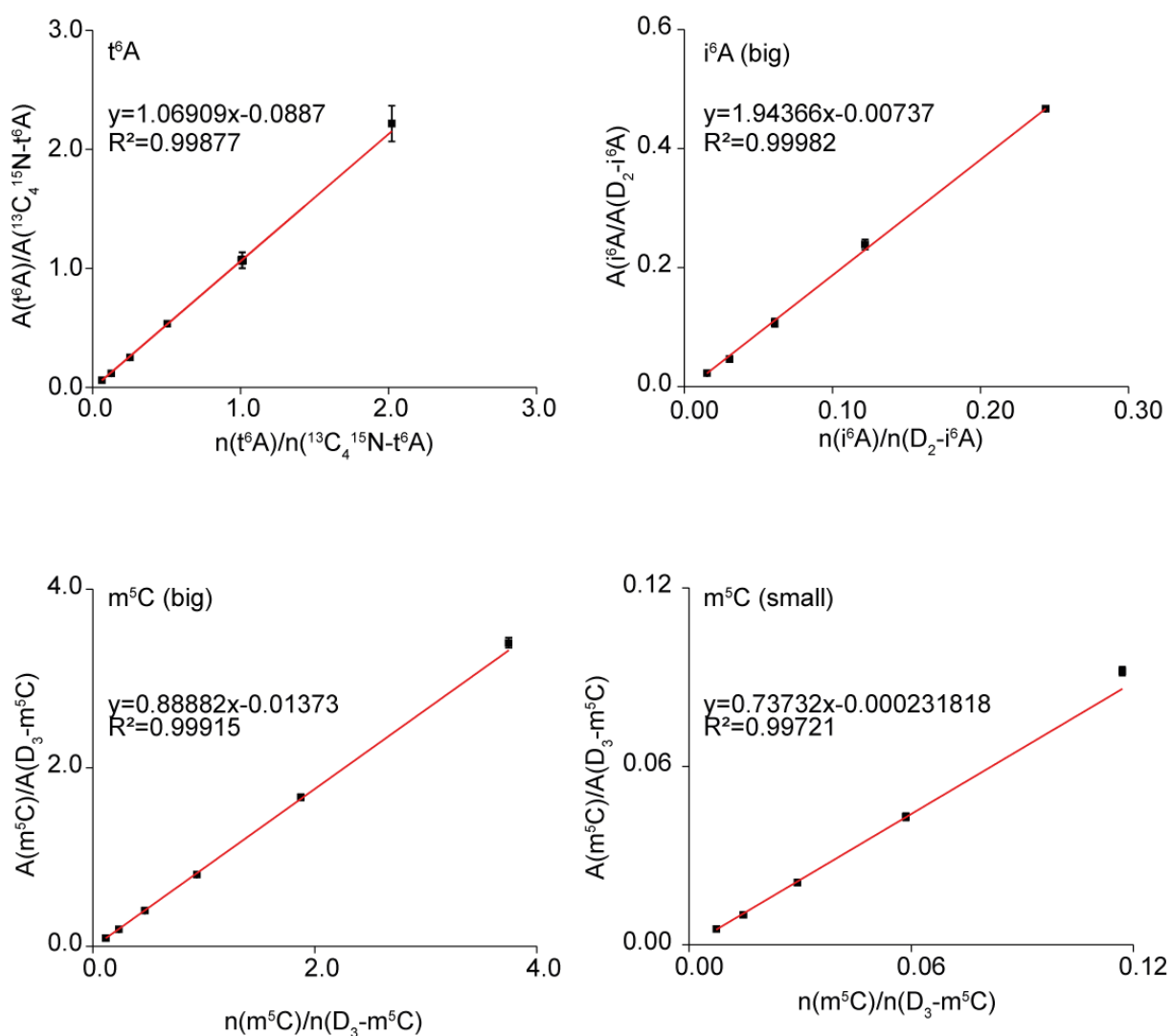
Figure 57: Internal calibration curves for the RNA nucleosides t^{6}A (upper left), i^{6}A (upper right) and m^{5}C in a wide range (lower left) and a small range (lower right).

Table SI-3: Parameters for the internal calibration curves for the RNA nucleosides t^{6}A (row row), i^{6}A (second row) and m^{5}C in bigger amounts (third row) and smaller amounts (last row).

Nucleoside	Linear Regression in H$_2$O/MeCN	LLOQ [pmol]	ULOQ [pmol]	LLOQ A/A*	ULOQ A/A*
t^{6}A	y=1.06909x-0.0887	0.019	0.611	0.0606	2.217
i^{6}A (big)	y=1.94366x-0.00737	0.062	0.997	0.0225	0.467
m^{5}C (big)	y=0.88882x-0.01373	0.323	10.332	0.0920	3.401
m^{5}C (small)	y=0.73732x-0.00023	0.020	0.323	0.0053	0.092

_____

List of abbreviations

°C	degree centigrade
5'-dRP	5'-deoxyribose phosphate
5,10-CH$_2$-THF	5,10-methylenetetrahydrofolate
8-oxodG	7,8-dihydro-8-oxoguanine
βEP	β-elimination product
μ	micro
a2i/L	alternative two inhibitors/LIF
A	area
(A/C/G/U/T) (D/M/T) P	((2'-deoxy)adenosine/cytidine/guanosine/uridine/thymidine) (di/mono/tri) phosphate
AID/AICDA	activation-induced cytidine deaminase
Ala-Gln	L-alanyl-glutamine
AML	acute myeloid leukemia
APE	AP endonuclease
APOBEC	apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like family
AP site	abasic site
AS	Angelman syndrome
(H$_2$-)Aza(d)C	(dihydrogenated) 5-Aza-(2'-deoxy)-cytidine
BER	base excision repair
bp	base pairs
BWS	Beckwith-Wiedemann syndrome
c	centi
ca(d)C	5-carboxy-(2'-deoxy)cytidine
CDA	cytidine deaminase
CDKAL1	CDK5 regulatory subunit associated protein 1-like 1
CDK5	cyclin-dependent protein kinase 5
CDK5RAP1	CDK5 regulatory subunit associated protein 1
CGP77675	inhibitor of proto-oncogene tyrosine-protein kinases SRC
CHCl$_3$	chloroform
CHIR 99021	inhibitor of GSK3α and β
CO$_2$	carbon dioxide
CpG	2'-deoxycytidine-phosphate-2'-deoxyguanine
CSR	class switch recombination
d	days
	2'-deoxy
D$_2$-	double deuterated
D$_3$-	triple deuterated
Da	Dalton
dA	2'-deoxyadenosine
dC	2'-deoxycytidine
DCK	deoxycytidine kinase
DCTD	deoxycytidylate deaminase
DCTPP1	dCTP pyrophosphatase 1
dG	2'-deoxyguanosine
DMAPP	dimethylallyldiphosphate
(g/s)DMR	(gametic/somatic) differentially methylated region
DMSO	dimethylsulfoxide
(g)DNA	(genomic) deoxyribonucleic acid
DNMT	DNA methyltransferase

DPBS	*Dulbecco's phosphate buffered saline*
Dr.	doctor
dU	2'-deoxyuridine
dT	thymidine
ELP3	elongator protein complex 3
ERK	extracellular signal-related kinases
ESI	elektrospray ionisation
et al.	and others
EtOH	ethanol
F-	2'-fluorinated
FBS	fetal calf serum
f(d)C	5-formyl-(2'-deoxy)cytidine
Fe^{II}	iron(II)
FTO	fat mass and obesity associated protein
g	gram
g	gravitation constant ($6{,}674 \bullet 10^{-11}$ m^3/(kg$\bullet$s²))
GADD45	growth arrest and DNA damage 45
GGR	global genome repair
GSK3β	glycogen synthase kinase 3 β
h	hour
hENT-1/-2	human equilibrative nucleoside transporter 1/2
hm(d)C	5-hydroxyl-(2'-deoxy)cytidine
H_2O	water
hpf	hours post fertilization
HPLC	high performance liquid chromatography
HRR	homologous recombination repair
Hz	Hertz
i^6A	N^6-isopentenyladenosine
ICE	imprinting control element
ICM	inner cell mass
ICR	imprinting control region
IGC	immunoglobulin gene conversion
Igf(r)2	insulin-like growth factor (receptor) 2
JAK/STAT	Janus kinase/signal transducer and activator of transcription
L	liter
LC-MS	liquid chromatography coupled to mass spectrometry
LIF	leukemia inhibitory factor
Lig IIIα	DNA ligase
LINE	long interspersed element
(l)ncRNA	*(long) non-coding* RNA
LOI	loss-of-imprinting
m	milli
	meter
M	molar
MΩ	mega ohm
MAPK	mitogen-activated protein kinase
m^6(d)A	N^6-methyl-(2'-deoxy)adenosine
MBD4	methyl-CpG-binding domain protein 4
m^5(d)C	5-methyl-(2'-deoxy)cytidine
MDS	myelodysplastic syndromes
MEK1/2	MAPK kinase
(m)ESC	(mouse) embryonic stem cells

METTL	methyltransferase-like
$MgCl_2$	magnesium chloride
min	minute
mio.	million
miRNA	*micro* RNA
MLH1	mutL homologue 1
mm	millimeter
MMR	mismatch repair
mRNA	*messenger* RNA
ms^2i^6A	2-thiomethyl-N^6-isopentenyladenosine
mtDNA	mitochondrial DNA
n	amount
NaCl	sodium chloride
Na_2-[EDTA]	disodium ethylendiamine tetraacetate
NEIL	Nei-like DNA glycosylase
NER	nucleotide excision repair
NHEJ	non-homologous end joining
NSUN	NOL1/NOP2/sun domain enzyme
OCT-4	octamer-binding protein 4
p	passage
p60/100	plate 60 mm/100 mm
PAR	poly-(ADP ribose)
PCNA	proliferating cell nuclear antigen
PD 0325901	inhibitor of MEK1/2
PGC	primordial germ cell
PNK	polynucleotide 5ʹ-hydroxyl kinase
Polβ	DNA polymerase β
PWS	Prader-Willi syndrome
QQQ	Triple Quadrupole mass spectrometer
R^2	coefficient of determination
RNA	ribonucleic acids
RNR	ribonucleotide reductase
ROS	reactive oxygen species
RPA	replication protein A
RPMi 1640	Roswell Park Memorial Institute medium
rRNA	ribosomal RNA
s	second
SAM	*S*-adenosyl-L-methionine
SHM	somatic hypermutation
siRNA	*small interfering* RNA
SMRT	single molecule real time
SMUG1	single-strand selective monofunctional uracil DNA glycosylase
SNP	single nucleotide polymorphism
snoRNA	*small nucleolar* RNA
snRNA	*small nuclear* RNA
SRS	Silver-Russell syndrome
SSBP1	single-stranded DNA-binding protein 1
ssDNA	single-stranded DNA
t^6A	N^6-threonyl-carbamoyl-adenosine
TCR	transcription-coupled repair
TDG	thymidine DNA glycosylase
TET enzyme	ten-eleven translocation enzyme

TK	thymidine kinase
TNKS-1/-2	tankyrase-1/-2
Tris-HCl	tris(hydroxymethyl) aminoethane hydrochloric acid
TRIT1	tRNA isopentenyltransferase 1
tRNA	*transfer* RNA
TS	thymidine synthase
u	enzyme unit
UCK	uridine-cytidine kinase
UHRF1	ubiquitin-like PHD and RING finger domain-containing protein 1
UNG	uracil DNA glycosylase
UV	ultraviolett light
V	Volt
WNT	wingless/Int1
wt	wildtype
XPG	Xeroderma pigmentosum complementation group G
XRCC1	X-ray repair complementing defective repair in Chinese hamster cells 1
ZnSO$_4$	zink sulfate