Dissertation zur Erlangung des Doktorgrades

der Fakultät Chemie und Pharmazie

der Ludwig-Maximilians-Universität München

Massenspektrometrische Analyse des DNA-Metabolismus

Olesea Kosmatchev (geb. Cujeli)

aus

Chișinău (Republica Moldova)

2018

Erklärung

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

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig ohne unerlaubte Hilfe erarbeitet.

Augsburg, den 05.05.2018

Olesea Kosmatchev

Dissertation eingereicht am 29.05.2018

1. Gutachter: *Prof. Dr. Thomas Carell*
2. Gutachter: *PD Dr. Stylianos Michalakis*

Mündliche Prüfung am 20.07.2018

„Falls Gott die Welt geschaffen hat, war seine Hauptsorge sicher nicht, sie so zu machen, dass wir sie verstehen können.“

Albert Einstein.

Publikationsliste

Teile dieser Dissertation wurden bereits publiziert oder auf Konferenzen vorgestellt.

[4] K. Iwan#, R. Rahimoff#, A. Kirchner#, F. Spada#, A. S. Schröder, **O. Kosmatchev**, S. Ferizaj, J. Steinbacher, E. Parsa, M. Müller, T. Carell, *Nat. Chem. Biol.* **2018**, *14*, 72-78. *5-Formylcytosine to cytosine conversion by C-C bond cleavage in vivo.*

[3] R. Rahimoff#, **O. Kosmatchev#**, A. Kirchner#, T. Pfaffeneder, F. Spada, V. Brantl, M. Müller, T. Carell, *J. Am. Chem. Soc.* **2017**, *139*, 10359-10364. *5-Formyl- and 5-Carboxydeoxycytidines Do Not Cause Accumulation of Harmful Repair Intermediates in Stem Cells.*

[2] S. Schiffers, C. Ebert, R. Rahimoff, **O. Kosmatchev**, J. Steinbacher, A.-V. Bohne, F. Spada, J. Nickelsen, S. Michalakis, M. Müller, T. Carell, *Angew. Chem. Int. Ed.* **2017**, *56*, 11268-11271. *Quantitative LC-MS Provides No Evidence for m6dA or m4dC in the Genome of Mouse Embryonic Stem Cells and Tissues*

[1] T. Pfaffeneder#, F. Spada#, M. Wagner#, C. Brandmayr, S. Laube, D. Eisen, M. Truss, J. Steinbacher, B. Hackner, O. Kotljarova, D. Schuermann, S. Michalakis, **O. Kosmatchev**, S. Schiesser, B. Steigenberger, N. Raddaoui, U. Müller, H. Leonhardt, P. Schär, M. Müller, T. Carell, *Nat. Chem. Biol.* **2014,** *10, 574-581. Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA.*

Posterpräsentation:
O. Kosmatchev, K. Iwan, R. Rahimoff, A. Kirchner, T. Pfaffeneder, T. Carell, *"Quantification of Abasic Sites and β-Elimination Products using Tandem Mass Spectrometry"*, International Round Table on Nucleosides, Nucleotides and Nucleic Acids (XXII IRT), Paris, Juli 2016.

Patentanmeldung:
EP patent application no. 17153895

geteilte Erstautorenschaft

Inhaltsverzeichnis

Zusammenfassung

Die vorgelegte Promotionsarbeit widmet sich der Frage nach möglichen Mechanismen und Intermediaten der aktiven DNA-Demethylierung. Während der Mechanismus der DNA-Methylierung bereits sehr gut erforscht und verstanden ist, wirft die aktive Entfernung dieser Modifikation aus der DNA immer noch Fragen auf. In den vergangenen Jahren wurden bereits verschiedene Hypothesen zum Ablauf einer aktiven DNA-Demethylierung aufgestellt. Neben einer eher unwahrscheinlichen, direkten Entfernung von 5-Methyl-2'-desoxycytidin (mdC) aus der DNA, scheint der Weg über eine stufenweise Oxidation dieses Cytidinderivats mit anschließender Basenexzisionsreparatur (BER) eine plausible Alternative zu sein. Ein wichtiger Teil der vorgelegten Arbeit beschäftigt sich mit der Analyse eines alternativen Weges der Demethylierung genomischen mdCs. Statt einer stufenweisen Oxidation von mdC, ist auch eine Desaminierung von mdC zu dT als Initialschritt einer Demethylierung denkbar. Um diese Fragestellung näher zu betrachten, wurden Isotopenverfolgungsexperimente in unterschiedlichen embryonalen Stammzelllinien der Maus (mESCs) durchgeführt und mittels Massenspektrometrie untersucht. Dazu wurde zunächst, ausgehend von einem im Arbeitskreis *Carell* etablierten Verfahren, eine neue massenspektrometrische Methode zum Nachweis von isotopenmarkierten [^{13}C, D_3]-dT, das im Zuge der Desaminierung aus metabolisch markiertem mdC entsteht, entwickelt. Es konnte gezeigt werden, dass tatsächlich eine Desaminierung von mdC zu dT im Genom von mESCs stattfindet. Der Anteil dieser Reaktion ist allerdings sehr gering, da die hauptsächliche Umsetzung von mdC zu dT nachweislich im löslichen Nukleosid-/Nukleotidpool der Zellen stattfindet. Die punktuelle Desaminierung könnte jedoch auch als Signal für eine globale genomweite Reparatur von mdC in der Zelle dienen. Um eine genauere Aussage treffen zu können, welche Rolle die Desaminierung bei der aktiven Demethylierung der DNA tatsächlich spielt, bedarf es weiterer Untersuchungen.

Ein weiteres Projekt vorgelegter Doktorarbeit beschäftigt sich mit dem Nachweis und der Quantifizierung abasischer (apurinischer / apyrimidinischer (AP)) Stellen in der genomischen DNA muriner embryonaler Stammzellen und verschiedener somatischer Zelllinien. AP-Stellen entstehen im Zuge der Basenexzisionsreparatur (BER) und könnten als Nachweis einer aktiven

DNA-Demethylierung über BER dienen. Da AP-Stellen selbst massenspektrometrisch nicht nachweisbar sind, wurde im Arbeitskreis *Carell* ein neues Derivatisierungsreagenz entwickelt. Basierend auf Vorarbeiten von *Toni Pfaffeneder* wurde eine neue massenspektrometrische Methode zum Nachweis und zur Quantifizierung von AP-Stellen im sub-femtomolaren Bereich entwickelt. Diese Methode ist wesentlich sensitiver als bisher bekannte Quantifizierungsmethoden. Darüber hinaus konnte in Zusammenarbeit mit *René Rahimoff* und *Angie Kirchner* zum ersten Mal das β-Eliminierungsprodukt (β-EP), dass mittels bifunktioneller Glykosylasen im Laufe der BER entsteht, nicht nur nachgewiesen, sondern auch quantifiziert werden. Mittels Isotopenverfolgungsexperimenten mit Zucker-markierten Nukleosiden in genomischer DNA aus mESCs, konnte der Ursprung der nachgewiesenen schwer-markierten AP- oder β-EP-Stellen identifiziert werden. Eine Akkumulation von schädlichen abasischen Stellen, die beispielweise aus der Entfernung der höher oxidierten Cytidinderivaten 5-Formyl-2'-desoxycytidin (fdC) oder 5-Carboxy-2'-desoxycytidin (cadC) stammen, konnte hierbei nicht beobachtet werden, wodurch eine Beteiligung der BER an der Demethylierung der DNA eher gering erscheint.

In einem weiteren Projekt, das in Kooperation mit *Corinna L. Kufner* (AK Zinth) erfolgte, wurden unter Zuhilfenahme einer im AK Carell etablierten massenspektrometrischen Standardmethode die 8-Oxo-dG-Werte in synthetischen Oligodesoxynukleotiden unterschiedlicher Länge und Zusammensetzung nach der UVC-Bestrahlung bestimmt. Es wurde gezeigt, dass ein Zusammenhang zwischen der Lebensdauer der ladungsgetrennten Zustände zwischen benachbarten Basen und der, nach der UVC-Belichtung, entstandenen 8-Oxo-dG-Menge besteht.

Abstract

The PhD thesis submitted addresses the issue of possible mechanisms and intermediates of the active DNA demethylation. While the mechanism of DNA methylation has already been studied in detail and is well understood, the active removal of this modification from DNA is still under debate. In recent years, various hypotheses have already been postulated for the process of active DNA demethylation. In addition to a rather improbable direct removal of 5-methyl-2'-deoxycytidine (mdC) from the DNA, the path over a stepwise oxidation of these Cytidine derivatives with subsequent base excision repair (BER) seems to be a plausible alternative. An important part of the presented work is focused on the analysis of an alternative demethylation pathway of genomic mdCs. Instead of a stepwise oxidation of mdC, a deamination of mdC to dT is conceivable as an initial step of demethylation. In order to study this question in more detail, isotope tracing experiments were performed in different murine embryonic stem cell lines (mESCs) and analyzed by mass spectrometry. For this project scope, a new mass spectrometric method for the detection of isotope labelled [^{13}C, D_3]-dT, arising in the course of a deamination of metabolically marked mdC, was developed and is based on procedures already established in the Carell group. It could be shown that a deamination of mdC to dT is actually taking place in the genome of mESCs. However, the proportion of this reaction is very low, since the main conversion of mdC to dT traceably takes place in the soluble nucleoside/nucleotide-pool of the cell. Nevertheless, the selective deamination could also serve as signal for a global genome-wide repair of mdC in the cell. For a precise evidence, which role that deamination plays in the active demethylation of the DNA, further investigations are required.

Another project of the submitted PhD thesis is devoted to detection and quantification of abasic (apurinic/apyrimidinic (AP)) sites in the genomic DNA of murine embryonic stem cells and various somatic cell lines. AP-sites are produced in the course of base excision repair (BER) and could serve as proof of an active DNA demethylation next to BER. Since AP-sites themselves are not detectable via mass spectrometry, a new derivatisation reagent was developed in the *Carell* group. Based on the work of *Toni Pfaffeneder*, a new mass spectrometric method for the

detection and quantification of AP-sites in the sub-femtomolar range was developed. This method is much more sensitive than previously known quantification methods. Moreover, in collaboration with *René Rahimoff* and *Angie Kirchner*, for the first time, the β-elimination product (β-EP), that is created by bifunctional glycosylases in the course of BER, could not only be detected but also quantified. Through sugar-labelled nucleosides, isotope labelling experiments in mESCs were performed, that allowed to identify the origin of the detected heavy-labelled AP-or β-EP-sites. An accumulation of harmful abasic sites, e.g. from the higher-oxidized cytidine derivatives 5-formyl-2 '-deoxycytidine (fdC) or 5-carboxy-2 '-deoxycytidine (cadC), could not be observed, thereby the participation of BER in the DNA demethylation appears to be rather low.

A further project of submitted PhD Thesis, which was carried out in collaboration with *Corinna L. Kufner* (*Zinth* group), investigates the levels of 8-oxo-dG after UVC irradiation in synthetic oligonucleotides of different lengths. The nucleoside composition was quantified through a mass spectrometric standard method established in *Carell* group. A correlation between the lifetime of the charge-separated states between neighbouring bases and the 8-oxo-dG content, created after the UVC irradiation, was shown.

1 Einleitung

1.1 Epigenetik

Der menschliche Körper verfügt über mehr als 200 verschiedene Zelltypen, wie beispielweise Haut-, Nerven- oder Muskelzellen, die eine unterschiedliche Erscheinung und Funktionen aufweisen können.[1-2] Paradoxerweise besitzen alle diese unterschiedlichen Zellen eines Organismus jedoch den gleichen genetischen Code. Wie konnten dann so viele unterschiedliche Zelltypen entstehen? In einer heranreifenden Zelle müssen Regulationsmechanismen vorhanden sein, welche die Genexpression steuern können. Spezifische Gene einer Zelle müssen zu einem bestimmten Zeitpunkt entweder aktiviert oder stillgelegt werden. Diese sogenannten Genaktivitätsmuster werden anschließend an die Tochterzellen weitervererbt. Nur so können unterschiedliche Zelltypen in einem multizellulären Organismus entwickelt werden. Das heißt, nicht nur Gene steuern die Genexpression, sondern die Gene selbst werden durch bestimmte Strukturen, die ihre Aktivität „fixieren", gesteuert. Mit diesen komplexen Regulationsmechanismen befasst sich die Wissenschaft der Epigenetik.[3-4] Die epigenetischen Mechanismen (zusätzlich zur DNA-Basen-Sequenz) dienen der Stabilisierung der Genexpressionsmaschinerie, dabei werden spezifische Zelltypen in bestimmte Bahnen der zellulären Entwicklung gelenkt.

Der Begriff „Epigenetik" stammt aus dem Jahre 1942.[5] Um bestimmte, damals noch wenig verstandene, Aspekte in der zellulären Entwicklung eines Organismus zu erklären, definierte *Waddington* alle Änderungen im Phänotyp ohne Änderungen des Genotyps als epigenetisch.[5-6] Epigenetik im weiteren Sinne wurde als Bindeglied zwischen Genotyp und Phänotyp angesehen. Die zelluläre Differenzierung wurde als ein epigenetisches Phänomen verstanden und von *Waddington* mit Hilfe seiner „epigenetischen Landschaft" aus dem Jahr 1957 (Abbildung 1.1) erklärt.[3, 7-8]

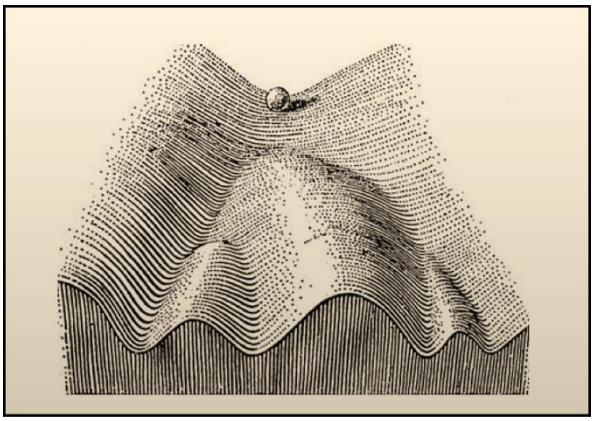


Abbildung 1.1: Epigenetische Landschaft von Waddington (1957).[7] Epigenetische Landschaft repräsentiert mögliche Wege der zellulären „Entscheidung" während der Entwicklung. Die Kugel repräsentiert eine totipotente Zelle, die verschiedenen Entwicklungspfaden folgen könnte und dementsprechend unterschiedliche Formen annehmen könnte.

Doch welche Strukturen genau werden verändert und welche Mechanismen liegen diesen Veränderungen zugrunde?

Um diese hoch komplexen Fragen zu beantworten, stehen der modernen Forschung embryonale Stammzellen (ES-Zellen) zur Verfügung. ES-Zellen sind pluripotente Zellen, die aus der inneren Zellmasse eines Präimplantationsembryos im Blastozystenstadium gewonnen werden. Ihre bemerkenswerten Eigenschaften sind die Fähigkeit zur Differenzierung in unterschiedliche Zelltypen der drei Keimbahnen: Endoderm, Mesoderm und Ektoderm und die nahezu unbegrenzte Selbsterneuerung *in vitro*.[9-10] Im Laufe der Differenzierung verlieren die Stammzellen ihre Pluripotenz. Die Schlüsselereignisse dabei sind die Stilllegung und die Aktivierung bestimmter Untergruppen von Genen.[11-14] Zusätzlich findet eine epigenetische Regulierung statt, dabei durchläuft das Genom ausdifferenzierter Zellen eine globale Reorganisation des Chromatins.[15-18]

Chromatin ist aus „transkriptionsfreundlichem", weniger kondensiertem Euchromatin und aus still gelegtem, hochkondensiertem Heterochromatin zusammengesetzt.[19] Die zentrale Untereinheit des Chromatins, das Nukleosom, besteht aus einer doppelsträngigen DNA (147 Basenpaare), die um ein Histonoktamer gewickelt ist.[20] Das Histonoktamer ist ein Komplex aus jeweils zwei Molekülen von H2A, H2B, H3 und H4.[21-23] Die benachbarten Nukleosomen werden durch die Linker-DNA und das Linker-Histon H1 verknüpft.[20, 24-25] Die Organisation

dieser Grundbausteine definiert letztendlich die Chromatinstruktur und die Zugänglichkeit der DNA für die Transkriptions- und Replikationsmaschinerie. In diesem Zusammenhang fokussiert sich die moderne Forschung unter anderem auf die Untersuchung post-translationaler Histon-Modifikationen, wie zum Beispiel Acetylierung, Methylierung, Phosphorylierung, Ubiquitinierung, ADP-Ribosylierung, Desaminierung, Propionylierung und Butyrylierung.[26-28] Der Gegenstand dieser Modifizierungen sind N- oder C-terminale Histonfortsätze, die aus dem Nukleosom herausragen.[26] Die Abfolge dieser Modifikationen bestimmt die Bindungsaffinität der Nukleosomen zur DNA und reguliert somit auch die Chromatinstruktur. Folglich kann die DNA entweder für andere Proteine, wie Transkriptionsfaktoren, zugänglich gemacht werden, oder deren Zugang zu den DNA-Bindungsstellen blockiert werden.[26, 29-31]

In der epigenetischen Regulation spielen nicht nur die Histon-Modifikationen, sondern auch die kovalenten Modifikationen des Cytosins in der DNA eine wichtige Rolle. Dabei können bestimmte DNA-Abschnitte, zum Beispiel Promotorregionen, methyliert werden, was zu einer Stilllegung der Genexpression führt. Dabei wird das 2'-Desoxycytidin (dC) zu einer der essentiellen DNA-Modifikationen, dem mdC umgewandelt. Diese Modifikation wird im nächsten Kapitel ausführlich erläutert. Auch wenn die Rolle der DNA-Methylierung gut erforscht ist, bleibt es immer noch unklar, ob alle Histon-Modifikationen zur epigenetischen Regulation beitragen oder nur einzelne von epigenetischer Natur sind.[32]

1.1.1 DNA-Methylierung

Die DNA-Methylierung ist eine der am besten erforschten epigenetischen Modifikationen und ist essentiell für die Entwicklung der Säugetiere. Es ist eine vererbbare epigenetische Markierung, die durch einen kovalenten Transfer einer Methylgruppe von einem Methylgruppen-Donor *S*-Adenosylmethionin (SAM) auf die C-5 Position eines Cytosin-Ringes zustande kommt. Dies geschieht mit Hilfe der DNA-Methyltransferasen (Dnmt).[33] In Pflanzen findet die Methylierung des Cytidins sowohl in symmetrischen dC-dG (CpG)-Dinukleotiden wie auch in asymmetrischen CpX-Kontexten statt (X= 2'-Desoxyadenosin (dA), 2'-Desoxythymidin (dT), dC). In Säugetieren weisen Stammzellen und somatische Zellen jedoch ein unterschiedliches Methylierungsverhalten auf.[34] In somatischen Zellen findet mehr als 98% der Methylierung fast ausschließlich in CpG-Dinukleotiden statt. Nur ein sehr CpG-armen Genomabschnitten findet eine CpX-Methylierung statt.[34-37] Im Gegensatz dazu finden sich in Oozyten, adulten Neuronen und embryonalen Stammzellen (ES) nur ca. drei Viertel aller DNA-Methylierungen in CpG-Kontext,[34] vor allem in CpG-armen Genomabschnitten [34, 38-39] Hier wird vermutlich der Methylierungs-Mangel durch methylierte CpX-Dinukleotide kompensiert.[38] Das menschliche Genom weist bei ca. 60% seiner Promotorregionen eine Anhäufung von CpG-Dinukleotiden (CpG-Inseln) auf.[40-41] Diese Regionen sind in der Regel hypomethyliert und spielen eine wichtige Rolle in der Kontrolle der Transkription.[42-43]

Üblicherweise wird die DNA-Methylierung während der Zygotenentwicklung entfernt und während der Präimplantation des Embryos wieder neu hergestellt.[44] DNA-Methylierung spielt eine sehr wichtige Rolle in einer Reihe von Schlüsselprozessen wie der genomischen Prägung, Inaktivierung des X-Chromosoms, Transkription und Austausch repetitiver Elemente. Bei einer Fehlregulierung trägt sie zur Entstehung von Krankheiten wie Krebs bei. [33, 45-47]

DNA-Methylierung wird von der Dnmt-Enzymfamilie katalysiert: Dnmt1, Dnmt2, Dnmt3a, Dnmt3b und Dnmt3l.[48-52] Dnmt1 ist eine *maintenance* Methyltransferase, die während der Replikation für das Kopieren des DNA-Methylierungsmusters auf den Tochterstrang verantwortlich ist. Sie ist somit für die Vererbung des mdC verantwortlich und essentiell für die

Entwicklung des Organismus. *Dnmt1-knockout*-Mäuse sterben während der frühen embryonalen Entwicklung.[53-54] Dnmt2 ist homolog zu Dnmt1 und Dnmt3a/b, aber anstatt DNA, methyliert sie Cytosin-38 im Anticodon-Loop von Asparaginsäure-tRNA (*transfer ribonucleic acid*).[55] Dnmt3a und Dnmt3b sind im Gegensatz zur Dnmt1, *de novo* Methyltransferasen und methylieren bevorzugt CpG-Dinukleotide während der embryonalen Entwicklung. Mäuse mit fehlender Dnmt3a sterben ca. vier Wochen nach der Geburt. *Dnmt3b*-Knockout-Individuen sterben bereits zwischen Tag E14,5 und E 18,5.[53, 56] Im Gegensatz zu anderen Methyltransferasen besitzt Dnmt3l keine katalytische Aktivität. Sie unterstützt die *de novo* Methyltransferasen Dnmt3a und Dnmt3b durch Steigerung von deren Bindungsfähigkeit zum SAM, und somit durch Steigerung deren Aktivität *in vivo*.[57] Die Kooperation zwischen verschiedenen Dnmts ist für die Methylierung mancher Genabschnitte, insbesondere repetitiver Sequenzen, erforderlich. Darüber hinaus wurde aber gezeigt, dass die DNA-Methyltransferasen auch andere Funktionen in der DNA-Methylierung annehmen können. So sammeln sich Hinweise, dass zum Beispiel Dnmt1 für die *de novo* Methylierung genomischer DNA benötigt wird, wobei die Dnmt3a und Dnmt3b zur *maintenance* Methylierung während der Replikation beitragen können.[58-59]

Obwohl es keine Hinweise auf einen direkten Zusammenhang zwischen der Tumor-Entwicklung und einer Mutation oder einem Defekt in einer der Dnmts festgestellt wurden, hängen Defekte in der DNA-Methylierung eng mit der Krebsentstehung zusammen. Epigenetische Auffälligkeiten von Krebs beinhalten globale DNA-Hypomethylierung und *locus*-spezifische Hypermethylierung von CpG-Inseln (CGIs).[60] Die Hypomethylierung kommt hauptsächlich durch den Verlust der Methylierung von normalerweise stark methylierten repetitiven Elementen einschließlich Satelliten und Retrotransposons zustande, und führt zu genomischer Instabilität und Aktivierung von Onkogenen. *Locus*-spezifische Hypermethylierung kommt bei CGIs von Tumorsuppressorgenen vor und führt zum vererbbaren transkriptionalen Gen-Silencing. Dabei wird die Bindung von Genregulatoren an das Gen verhindert .[61]

Außerdem nimmt die methylierte DNA an der Chromatin-Bildung Teil, in dem sie mit unterschiedlichen anderen epigenetischen Modifikationen, wie Histoncode, Polycomb-Komplexen, Nukleosom-Positionierung, nicht-kodierender RNA und Adenosintriphosphat (ATP)-abhängigen Chromatin umformenden Proteine interagiert.[60]

1.2 Wege der aktiven DNA-Demethylierung

Im Gegensatz zur DNA-Methylierung, bleibt der Prozess der aktiven DNA-Demethylierung weitgehend unklar, da viele unterschiedliche Mechanismen dazu beitragen können.[62] Wegen der hohen Stabilität der C5-CH_3-Bindung wurde die Methylierung früher als eine irreversible Markierung angesehen. Diese könne nur durch „Verdünnung" oder durch eine *de novo* DNA-Synthese entfernt werden.[63-64] Die Ergebnisse aktueller Studien haben allerdings gezeigt, dass die Methylierung nicht zwingend durch DNA Replikation passiv „verdünnt" wird, sondern auch aktiv entfernt werden kann.[65] Man unterscheidet zwischen genomweiter und *locus*-spezifischer Demethylierung. Die aktive genomweite DNA-Demethylierung findet im aus Sperma stammenden paternalen Pronukleus einer Zygote statt und dient zur schnellen Aktivierung der zygotischen Transkription.[66-68] Die *locus*-spezifische aktive DNA-Demethylierung wurde zum Beispiel in post-mitotischen Zellen eines adulten Gehirns[69-70] und während der Zellreprogrammierung beobachtet.[71]

Doch was für ein Mechanismus steht hinter der aktiven DNA-Demethylierung? Eine direkte Reaktion von mdC zu dC in einem Schritt ist ohne die Aktivierung von stabilen C5-CH_3-Bindungen unwahrscheinlich und wurde in Säugetieren auch nie beobachtet. Alle bekannten Umwandlungen von mdC zu dC *in vivo* umfassen mehrere Schritte und führen zu Einzelstrangbrüchen. Verallgemeinernd lassen sich folgende Schritte in solch einem Umwandlungsmechanismus beschreiben. Zuerst wird die Base modifiziert, entweder durch eine Desaminierung oder durch eine Oxidation des mdC. Anschließend folgt ein Austausch des modifizierten Nukleotids (Abbildung 1.2).[72] Viele Hinweise deuten auf eine Beteiligung der Basenexzisionsreparatur (BER) in der aktiven DNA-Demethylierung hin.[73-76] Nach dem die modifizierte Base durch eine der elf bekannten DNA-Glykosylasen erkannt und ausgeschnitten wird, entsteht eine abasische (apurinische/apyrimidinische (AP)) Stelle, die durch eine AP-Endonuklease (Ape1), DNA-Polymerase (Pol β) und –Ligase (I oder IIIα) in mehreren weiteren Schritten prozessiert wird.[77] Es gibt zwei Arten von BER Glykosylasen: monofunktionelle und bifunktionelle. Monofunktionelle Glykosylasen sind nur in der Lage die modifizierten Basen

auszuschneiden. Es entsteht eine AP-Stelle. Die bifunktionellen Glykosylasen besitzen zusätzlich zu der Glykosylase-Aktivität noch β- oder β,δ-Lyase Aktivität (Fähigkeit die AP-Stellen zu spalten). Sie spalten den DNA Strang entweder durch β- oder β,δ-Eliminierung.[78-79] Die BER wird im Abschnitt 1.3.1 ausführlich betrachtet.

Neben der BER können auch Nukleotidexzisionsreparatur (NER)[80-81] und nicht-kanonische Mismatch-Reparatur (ncMMR)[82] als abschließender Schritt der aktiven DNA-Demethylierung erfolgen.

Darüber hinaus wurde auch ein anderer Mechanismus der aktiven DNA-Demethylierung postuliert, bei dem die C-C-Bindung zwischen dem C5-Kohlenstoffatom und der substituierten Methylgruppe des modifizierten Cytosins gespalten wird. Es entsteht direkt das unmodifizierte dC (Abbildung 1.2). Dieser Weg der aktiven DNA-Demethylierung konnte vor kurzem von *Carell* und Mitarbeitern massenspektrometrisch nachgewiesen werden (Abschnitt 3.1).[83]

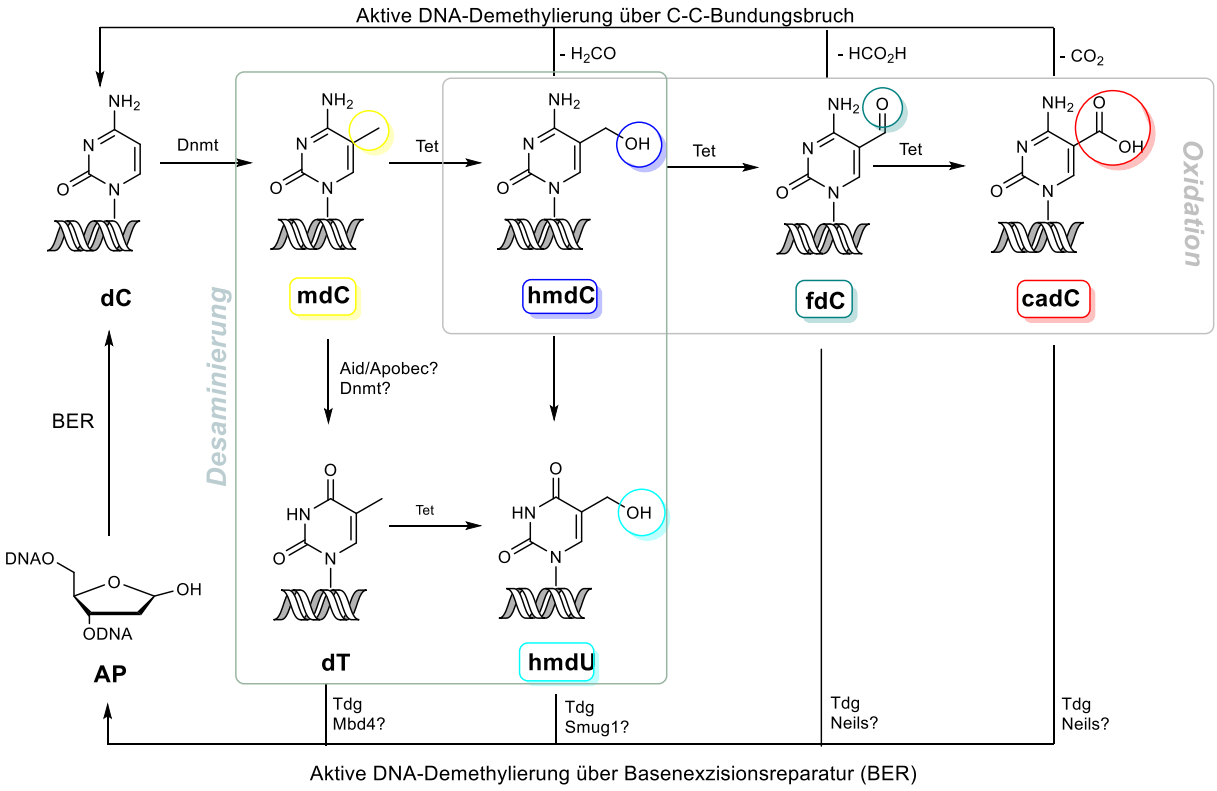


Abbildung 1.2: Mögliche Wege der aktiven DNA-Demethylierung.

Die unterschiedlichen möglichen Wege der aktiven DNA-Methylierung werden in den folgenden Kapiteln ausführlich erläutert.

1.2.1 Aktive DNA-Demethylierung über Oxidation mit BER

Der zurzeit am besten biologisch und biochemisch verifizierte Mechanismus der aktiven DNA-Demethylierung wird über die *ten-eleven-translocation* (Tet)-Enzyme initiiert.[84-89] Tet Enzyme sind Fe (II)-abhängige Dioxygenasen, die die DNA mit Hilfe von molekularem Sauerstoff und α-Ketoglutarat oxidieren. Dabei entstehen Succinat und Kohlenstoffdioxid (CO_2) als Nebenprodukte. Tet-Enzyme katalysieren die Umwandlung von mdC zu 5-Hydroxymethyl-2'-desoxycytidin (hmdC),[88, 90] hmdC zum fdC,[87] fdC zum cadC (Abbildung 1.3).[84, 86]

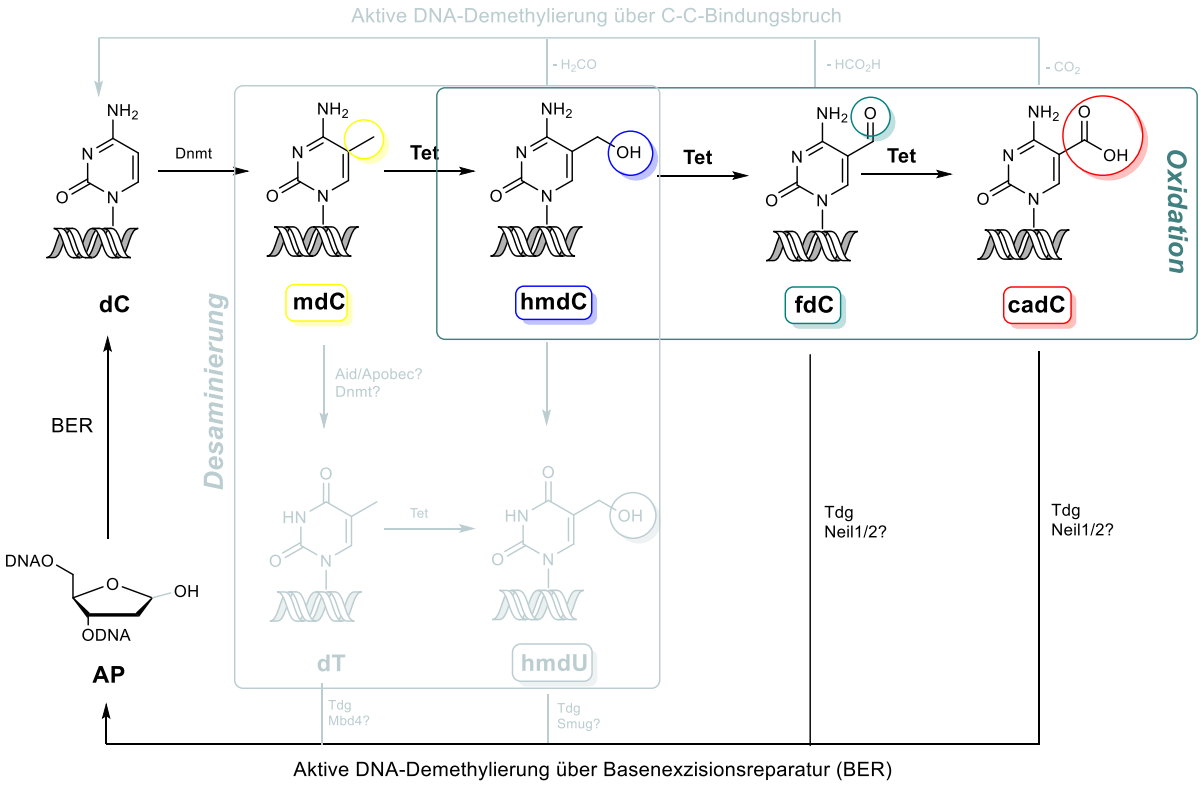


Abbildung 1.3: Aktive DNA-Demethylierung über Oxidation von mdC.

Die Verbindung zwischen den Tet-Enzymen und der aktiven Demethylierung wurde in zahlreichen Studien bestätigt.[91] Bei der Tet-vermittelten Oxidation von 5-Methylcytosin (mC) verringert sich der Gehalt an dieser Nukleobase und umgekehrt führt die Abwesenheit von Tet zur Hypermethylierung in Promotor-Regionen.[92] Dies bedeutet aber nicht, dass auch alle

oxidierten Basen sofort in das unmodifizierte Cytosin konvertiert werden. Es wurde gezeigt, dass in frühen Embryonen 5-Hydroxymethylcytosin (hmC) die ersten Zellteilungszyklen übersteht.[68] Mit Hilfe einer sensitiven massenspektrometrischen Methode (HPLC-MS/MS), die 2010 von *Carell* und Mitarbeitern vorgestellt wurde, konnten die hmC-Mengen in unterschiedlichen Geweben der Maus ermittelt werden.[36, 93] Die aktuellen UPLC-MS/MS Quantifizierungsstudien zeigen einen besonders hohen Gehalt dieser Cytosin-Modifikation in humanen Hirngeweben. Mit zunehmendem Alter steigt der mdC-Gehalt zunächst an, bis es seinen höchsten Wert von ca. 1.2% im vollentwickelten Gehirn erreicht, um anschließend konstant zu bleiben.[94] Im Vergleich zum adulten Hirngewebe der Maus, ist dieser Wert etwa doppelt so hoch.[95] Außerdem deutet eine Vielzahl von hmdC Reader-Proteinen, die in keinem Zusammenhang mit DNA-Reparatur stehen, darauf hin, dass diese Cytidin-Modifikation eine eigenständige epigenetische Markierung darstellt.[94, 96]

Im Gegensatz dazu konnte keine Anreicherung für fdC und cadC im adulten Hirngewebe festgestellt werden.[94,97] Im Gegensatz vermindert sich der fdC-Gehalt mit zunehmendem Alter. Diese Beobachtung deutet darauf hin, dass fdC eher als Intermediat bei dem Prozess der aktiven DNA-Demethylierung eine Rolle spielt.

Nach der Tet-initiierten Oxidation von mC erfolgt ein Nukleotid-Austausch durch BER. Die oxidierten dC Modifikationen, fdC und cadC, werden dabei von bestimmten DNA-Glykosylasen ausgeschnitten. Eine wichtige Glykosylase in diesem Zusammenhang ist die Thymin-DNA Glykosylase (Tdg). Es ist eine monofunktionale DNA Glykosylase. Sie erhielt ihren Namen für ihre Aktivität gegen das Produkt der Desaminierung von mC, die Base Thymin (T), die mit Guanin (G) fehlgepaart ist (T:G). Außerdem war es das erste entdeckte Enzym, das die Basen fC und caC ausschneiden konnte.[75, 84] Die bevorzugten Wechselwirkungen zwischen Tdg und caC Basen und die schnellere Exzision von fC:G als von T:G führten zur Annahme, dass die aktive DNA-Demethylierung eher über Oxidation von mC und nicht über Desaminierung von mC stattfindet.[75, 98]

Zuvor wurde gezeigt, dass Tdg für die Entwicklung der Maus unverzichtbar ist. Die Tdg-*doppelknockout* Mausembryonen sterben ca. am 10. Tag nach der Befruchtung.[76, 99] Darüber hinaus konnte bis jetzt noch keine andere Glykosylase in Säugetieren gefunden werden, die essentiell für die embryonale Entwicklung wäre und eine signifikante Glykosylase Aktivität gegenüber fC oder caC zeigen würde.[75, 100-101] Tdg ist interessanterweise nur sehr wenig in Oocyten und Zygoten exprimiert und nicht für die Demethylierung des paternalen Pronukleus erforderlich.[102]

Auch bifunktionelle Glykosylasen *Nei-like 1* (Neil1) und *Nei-like 2* (Neil2) kommen bei der aktiven DNA-Demethylierung für das Entfernen von fC und caC in Frage. Es ist bekannt, dass diese Enzyme oxidativ geschädigte Pyrimidine (zum Beispiel 5-Hydroxycytosin, 5-Hydroxyuracil, Thyminglykol) ausschneiden können.[77-78] Außerdem tragen Neil1 und Neil2 zur Reaktivierung von epigenetisch stillgelegten Reporter-Plasmiden bei, allerdings in geringerem Maße als Tdg.[74, 103] *Leonhardt* und Mitarbeiter zeigten, dass Neil1 und Neil2 die Abwesenheit von Tdg teilweise kompensieren können.[74] In weiteren Studien beobachteten *Niehrs* und Mitarbeiter eine Glykosylase-Aktivität an einer DNA, die fC und caC enthielt. Sie schlugen vor, dass Tdg Neil1 und Neil2 anstatt von Ape1 rekrutiert, um die entstandene 2'-Desoxyribose auszuschneiden.[103] Ferner wurde eine heterozygoter Neil1-*knockout* Maus untersucht. Diese Maus entwickelte ein Stoffwechselsyndrom, welches eine Kombination aus schwerer Fettleibigkeit, Stoffwechselstörung, Fettleberkrankheit und Veranlagung zur Hyperinsulinämia darstellt. Jedoch steht keiner dieser Phänotypen in direktem Zusammenhang mit einer defekten Demethylierung.[104] Außerdem wurde gezeigt, dass Neil2 für die frühe Entwicklung in *Xenopus* Embryonen, aber nicht in Embryonen der Maus erforderlich ist.[103, 105] Eine physiologisch relevante aktive DNA-Demethylierung, die entweder Neil1 oder Neil2 einbezieht, ist bisher noch nicht entdeckt worden. Es wäre aber denkbar, dass es nach der Aktivität von Tdg zwei alternative Wege in der aktiven DNA-Demethylierung geben könnte. Es könnten entweder Ape1 oder Neil Enzyme die abasische Stelle, die durch Tdg generiert wurde, weiter prozessieren.[106]

Eine weitere DNA-Glykosylase, die für das Ausschneiden der modifizierten Basen in Frage käme, ist die Uracil-DNA-Glykosylase (Ung). Ung kommt in einer mitochondrialen (Ung1) und einer Kern-Isoform (Ung2) vor und schneidet Uracil, das durch die Desaminierung von Cytosin entsteht, aus. Ung2 wurde vor kurzem bei der Suche nach Glykosylasen, die an Tet-abhängiger Genreaktivierung beteiligt sind, identifiziert. In kultivierten Zellen wurde gezeigt, dass Ung2 und nicht die Ung1 Form das caC in genomischer DNA ausschneiden kann.[73] Im Gegensatz dazu wurde aber keine Aktivität gegen caC *in vitro* beobachtet.[84] Ein Ung-Defizit in Zygoten beeinträchtigt tatsächlich die Demethylierung an manchen Positionen (*Nanog* und *Line1* Elemente), aber die globalen Mengen an oxidiertem mdC werden nicht verändert.[73] Es wurde über das Zusammenwirken von Ung2 und aktivierungsinduzierten Cytidin-Desaminasen (Aid) in Zygoten berichtet, aber es bleibt unklar, ob das Fehlen von Ung die Oxidation- oder Desaminierung-vermittelte Demethylierung beeinträchtigt.[107] Die Ung-defiziente Maus weist ein erhöhtes Level an Uracil in DNA auf, aber entwickelt sich ganz normal bis ins Erwachsenenalter.[108]

1.2.2 Aktive DNA-Demethylierung über Desaminierung mit BER

Wie oben beschrieben, beinhaltet einer der möglichen Wege der aktiven DNA-Demethylierung als ersten Schritt eine Desaminierung von mdC zu dT bzw. hmdC zu hmdU durch noch ein unbekanntes Desaminase-Enzym (Abbildung 1.4). Daraus resultierende dT:dG- bzw. hmdU:dG-Fehlpaarungen werden anschließend im Rahmen der BER prozessiert.[109-114] Gegen eine hmdC zu hmdU Desaminierung als der Schlüsselschritt der aktiven Demethylierung sprechen Isotopenverfolgungsstudien, die von *Carell* und Mitarbeitern durchgeführt wurden. Mit Hilfe einer neu entwickelten sensitiven Methode, die auf Tandemmassenspektrometrie (MS/MS) in Verbindung mit Ultra-Hochdruck-Flüssigkeitschromatographie (UHPLC) basiert, wurde gezeigt, dass das in mESC gefundene hmdU nicht von hmdC stammt, sondern durch eine Tet-vermittelte Oxidation von dT (Abschnitt 3.4).[115]

Neben der enzymatischen mdC zu dT Desaminierung wurde auch eine spontane mdC zu dT Desaminierung (C → T Transition) beobachtet. Diese Punktmutation führte vermutlich zu den CpG-armen Abschnitten im menschlichen Genom und wird von den Reparaturmechanismen teilweise toleriert, was zu einer dauerhaften Mutation führen kann.[41, 116]

Bezüglich der Desaminierung von mdC zu dT könnte ein Mitglied der aktivierungsinduzierten Cytidin-Desaminasen (Aid)/*apolipoprotein B mRNA editing enzyme, catalytic polypeptide* (Apobec) Familie die benötigte Desaminase-Aktivität liefern. Gegen die Aid/Apobec-vermittelte Demethylierung spricht allerdings die Tatsache, dass diese Enzyme hauptsächlich einzelsträngige DNA und bestimmte DNA Sequenzen (*hotspots*) bevorzugen.[117] Weitere Studien zeigten, dass sich mit steigender sterischen Hinderung an der C5-Position die Desaminaseaktivität des Aid Enzyms verringert.[118-119] Das Enzym kann Cytosin viel leichter desaminieren als Methylcytosin.[117, 119] Der menschliche Organismus verfügt über acht Apobec Enzyme, eine Apobec1 und sieben Apobec3 Varianten.[120]

Ferner könnten die *de novo* DNA Methyltransferasen Dnmt3a/b die Rolle der Desaminase übernehmen.[82, 110] Diese können dC zu 2'-Desoxyuridin (dU) [121] in Abwesenheit von SAM

desaminieren, allerdings nur *in vitro*. Da aber dieser Cofaktor vermutlich überall in der Zelle vorhanden ist, scheint dieser Prozess eher unwahrscheinlich.

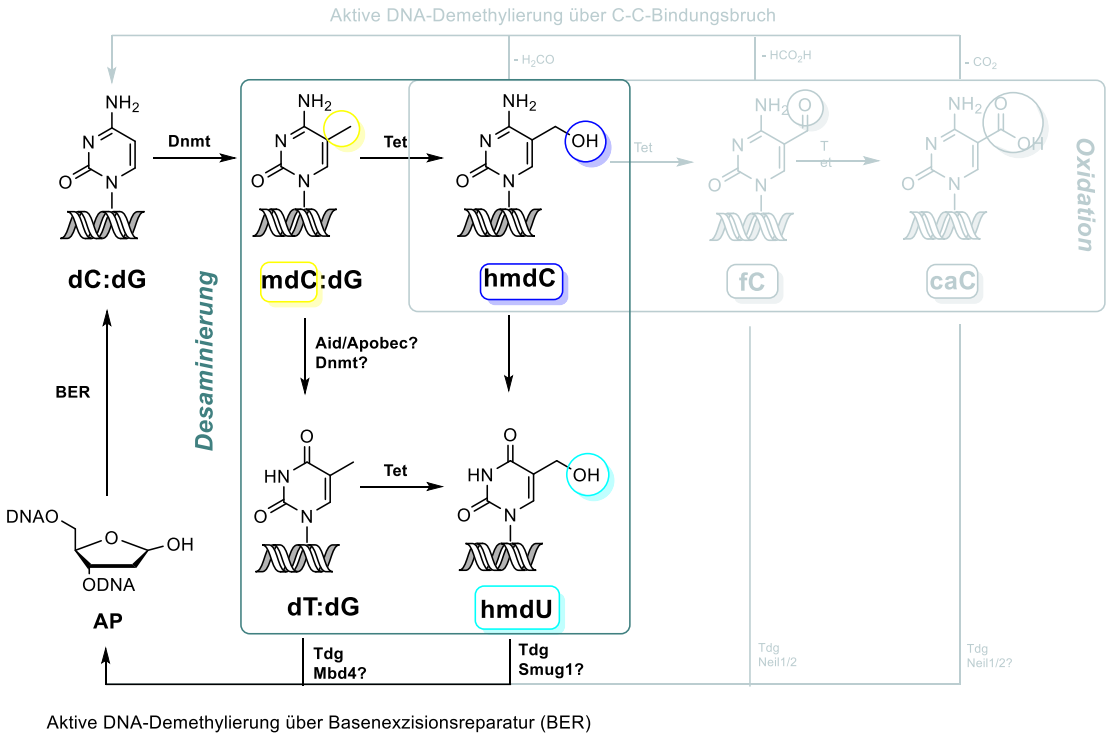


Abbildung 1.4: Desaminierung von mdC bzw. hmdC.

Im Gegensatz zu Oxidationsprodukten sind die Desaminierungsprodukte von dC und mdC immer fehlgepaart. Durch die Desaminierung von dC entsteht Uracil und kann durch Ung2, Smug1 (*single-strand-selective monofunctional uracil-DNA glycosylase*), Tdg und Mbd4 (*Methyl-CpG-binding domain protein 4*) ausgeschnitten werden.[122] Thymidin entsteht bei der Desaminierung von mdC und wird entweder durch Tdg oder Mbd4 ausgeschnitten. Diese Reaktionen sind im Zusammenhang mit somatischer Hypermutation sehr gut untersucht worden. Außerdem wurde gezeigt, dass Ung2 und Tdg an der Entfernung von zueinander nahegelegenen mC Basen, entweder durch *long-patch* BER oder nicht-kanonische Mismatch-Reparatur, beteiligt sind.[107, 123]

1.3 Abasische Stellen

1.3.1 Entstehung und Prozessierung von abasischen Stellen

Abasische Stellen (apurinische/apyrimidinische (AP)) sind die am häufigsten auftretenden nicht-kodierenden DNA-Schäden. Sie können zum Beispiel spontan durch Hydrolyse der *N*-glykosidischen Bindung entstehen. Dadurch kommt es zu mehr als 10 000 AP-Stellen pro Säugetierzelle an einem Tag unter normalen physiologischen Bedingungen. Dabei geschieht die Depurinierung ~100 Mal schneller als die Depyrimidinierung.[124-126] Abasische Stellen können auch dann entstehen, wenn die 2'-Desoxyribose freien Radikalen in der Zelle ausgesetzt ist.[127-130] Des Weiteren treten AP-Stellen in Folge einer ineffizienten Basenexzisionsreparatur geschädigter, falsch eingebauter oder modifizierter DNA-Basen auf, die zum Beispiel im Laufe der aktiven DNA-Demethylierung entstehen können (Abschnitt 1.2.2.).[78, 131] Es wird zwischen der am häufigsten vorkommenden *short-patch* BER (nur ein Nukleotid wird ersetzt) und *long-patch* BER (zwei bis zehn Nukleotide werden ersetzt) unterschieden.[77, 132] Während der *short-patch* BER schneidet die schadenspezifische monofunktionale DNA-Glykosylase die geschädigte oder modifizierte Base aus und generiert somit eine AP-Stelle, die im Gleichgewicht zwischen offenkettiger und ringgeschlossener Form vorliegt. Anschließend wird diese am 5'-Ende durch eine AP-Endonuklease (Ape1) prozessiert. Es kommt zu einem DNA-Einzelstrangbruch mit 3'-OH- und 5'-Desoxyribose-phosphat (5'-dRP)-Enden. Das 5'-dRP wird durch die DNA-Polymerase β (Pol β) entfernt und das richtige Nukleotid eingefügt. Alternativ kann die geschädigte Base durch eine bifunktionelle Glykosylase mit einer β-Lyase Aktivität (AP Lyase) entfernt werden. Es entsteht ein DNA-Einzelstrangbruch mit einem ungesättigten Hydroxyaldehyd (β-Eliminierungsprodukt (βE)) am 3'-Ende (3'-dRP) und einem Phosphat am 5'-Ende. Das βE wird anschließend durch Ape1 entfernt und von Pol β prozessiert. Wenn die Base durch eine Glykosylase mit β,δ-Lyase Aktivität erkannt wird, werden beide Seiten um die abasische Stelle gespalten. Es entsteht ein *trans*-4-Hydroxy-2,4-pentadienal. Dabei kommt es zu einer Nukleotid-Lücke umgeben von 3'- und 5'-Phosphat-Enden. Das 3'-Phosphat wird durch das bifunktionelle Enzym Polynukleotid

Kinase 3'-Phosphatase (PNKP) entfernt und das 3'-OH wird durch Pol β prozessiert. Anschließend findet der letzte BER Schritt statt, in dem die DNA-Ligase die freistehenden DNA-Enden zusammenfügt (Abbildung 1.5).[71,72]

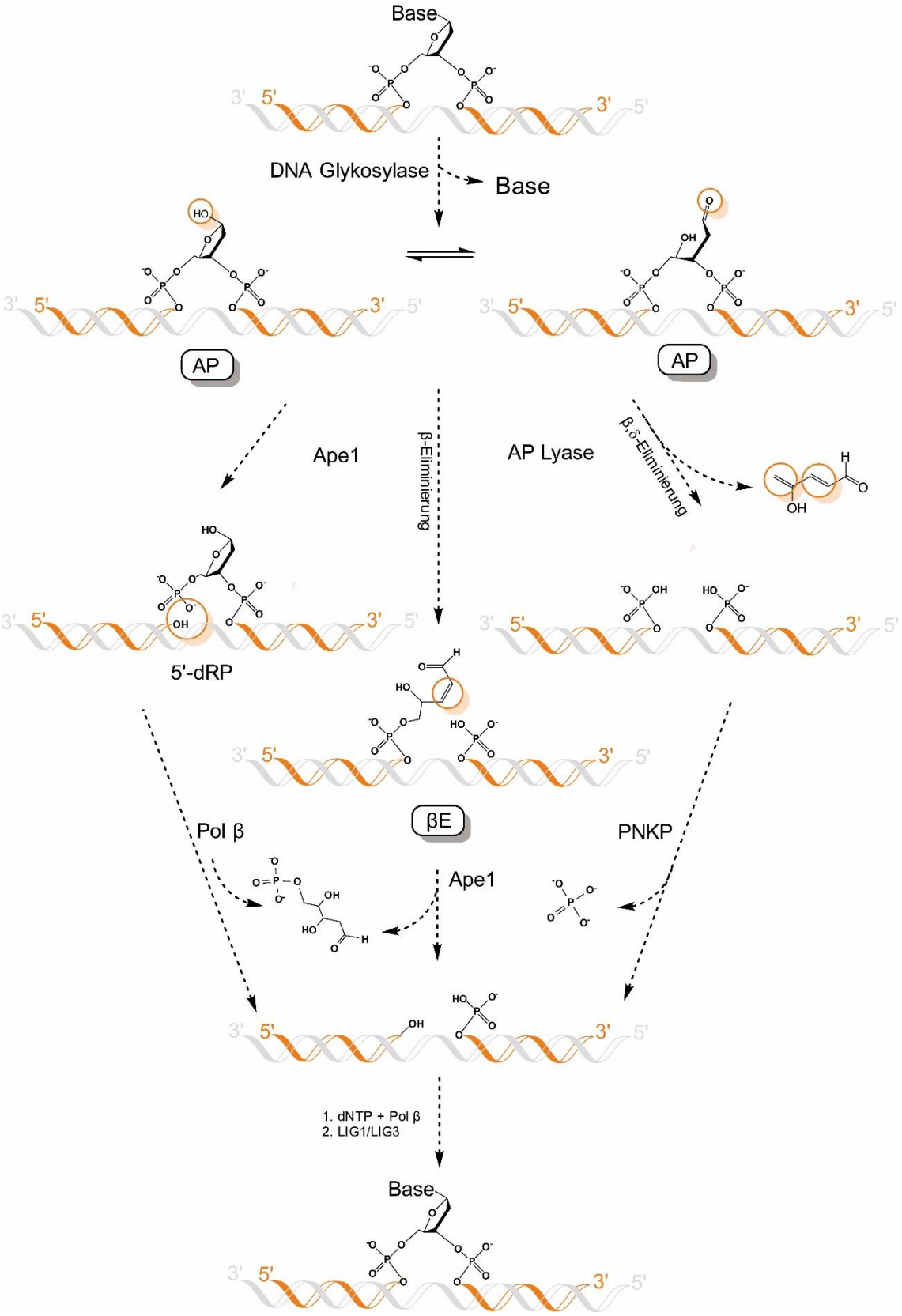


Abbildung 1.5: *short-patch* Basenexzisionsreparatur (BER).[77]

Wie bereits erwähnt, treten abasische Stellen als wichtige Intermediate während der Basenexzisionsreparatur auf. Doch welche Rolle spielen diese in epigenetisch relevanten Prozessen, vor allem bei der aktiven DNA-Demethylierung? Wie im Abschnitt 1.2. ausführlich erläutert, könnte ein möglicher Weg der aktiven DNA-Demethylierung via Oxidation oder Desaminierung von mdC mit anschließender BER erfolgen. Wenn die aktive DNA-Demethylierung tatsächlich über BER ablaufen würde, dann müssten abasische Stellen als Intermediate von der Zelle toleriert werden. Das würde bedeuten, dass Einzel- und Doppelstrangbrüche[72, 78, 107, 133-136] sowie ein erhöhtes Risiko von Mutationen[137-141], die von abasischen Stellen ausgeht, geduldet werden müssten. An dieser Stelle stellt sich die Frage, wie schnell die promutagenen abasischen Stellen weiter prozessiert werden? Welche Mengen werden von der Zelle in Kauf genommen? Welcher Anteil aller abasischen Stellen stammt tatsächlich von epigenetisch modifizierten Cytosinen? In welchen Mengen wird neben den gesättigten abasischen Stellen das β-Eliminierungsprodukt gebildet? Um alle diese Fragen beantworten zu können bedarf es exakter, hochempfindlicher Messmethoden zur Detektion, Bestimmung und Unterscheidung abasischer Stellen in der genomischen DNA.

1.3.2 Quantifizierung von abasischen Stellen

Für die Detektion und Quantifizierung von abasischen Stellen werden ihre chemischen Eigenschaften genutzt. Wie oben beschrieben, liegen die AP-Stellen in einem Gleichgewicht zwischen offenkettiger und geschlossener Form vor. Da die Aldehydgruppe der offenkettigen Form von AP und der βE nicht geladen und damit vor allem der Massenspektrometrie unzugänglich ist, bedarf es Derivatisierungsreagenzien, die selektiv mit dieser reagieren können, um deren Detektion zu ermöglichen. Die Reagenzien, die heutzutage verwendet werden, beruhen hauptsächlich auf Hydroxylamin-Derivaten.[128, 142-147] Diese reagieren selektiv mit der Aldehydgruppe der abasischen Stelle unter Ausbildung von stabilen Oximen. Die analytischen Methoden, die für die quantitativen Studien von AP-Stellen in genomischer DNA eingesetzt wurden, basieren auf ^{32}P-Postlabeling,[148] ^{14}C-[147, 149] und Fluoreszenz-Markierung,[150-152] immunologischen oder ELISA (*enzyme-linked Immunosorbent Assay*)-ähnlichen Assays[128, 142-145, 153-154] und Massenspektrometrie.[130, 146] Allerdings bringen die aufgelisteten Techniken mehrere Nachteile mit sich. Obwohl sie sensitiv zu sein scheinen, fehlt es entweder an eindeutiger Selektivität gegenüber den abasischen Stellen, oder es bedarf einer besonderen und umständlichen Handhabung von radioaktiven Substanzen, beziehungsweise werden große Mengen an biologischem Material benötigt. Außerdem kann es beim Arbeitsablauf (DNA-Isolation, Derivatisierung mit dem Reagenz) zur Bildung von artifiziellen abasischen Stellen kommen, die das Messergebnis verfälschen können.[142, 149, 155] Ferner kann keine von den aufgelisteten Methoden zwischen AP und βE unterscheiden.

Als gängigstes Reagenz für die Detektion und Quantifizierung von abasischen Stellen hat sich die *Aldehyde Reactive Probe* (ARP) durchgesetzt (Abbildung 1.6)[128, 154]. Das Molekül besitzt eine Biotin Einheit, die über einen Linker mit der Hydroxylamin-Gruppe verbunden ist. Das Biotin bildet einen Avidin-Biotin-Meerrettich-Peroxidase-Komplex in einem ELISA-ähnlichen Assay und kann so kolorimetrisch quantifiziert werden. Die ARP wurde außerdem auch in einer *in vivo* Analyse eingesetzt, um die abasischen Stellen noch vor der DNA-Isolation aus der Zelle zu maskieren. So konnte das Risiko einer Derivatisierung von ARP mit künstlich

erzeugten abasischen Stellen minimiert werden.[142] Neben ARP wurden auch andere Hydroxylamine für die AP-Stellen Quantifizierung verwendet.

Das neueste Derivatisierungsreagenz in diesem Zusammenhang ist ein Alkoxyamin AA3 (Abbildung 1.6).[151] Es besitzt neben der Hydroxylamin- auch eine Alkin-Gruppe, die in einer kupferkatalysierten Klick-Rektion mit unterschiedlichen biochemischen Azid-Tags reagieren kann. So kann das AA3 für Fluoreszenzmarkierung eingesetzt werden, da viele Fluoreszenz-Farbstoffe als Azide auf dem Markt angeboten werden. In dieser Studie wurde allerdings gezeigt, dass die ARP unter Standardbedingungen weniger als 20% aller abasischen Stellen markieren kann. Auch bei AA3 in Verbindung mit Biotinazid konnte eine volle Maskierung von abasischen Stellen nicht erreicht werden. Es wurden lediglich ca. 60% Derivatisierung erreicht.[151]

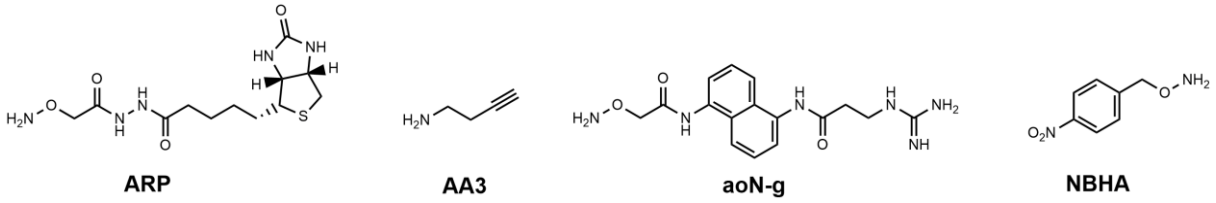


Abbildung 1.6: Derivatisierungsreagenzien für Quantifizierung von abasischen Stellen.

Um die Reaktionsgeschwindigkeit und Selektivität der Hydroxylamine gegenüber abasischen Stellen zu optimieren, wurde im Rahmen einer Studie der Einfluss unterschiedlicher Struktur-Elemente auf die Geschwindigkeit der Oximligation mit abasischen Stellen untersucht. Das effektivste Hydroxylamin in dieser Studie war das ao-Ng (Abbildung 1.6), das um Einiges schneller reagierte als die ARP. Diese Beobachtung wurde durch aromatische *stacking*-Effekte der Naphtalin-Gruppe, die die Interkalation des Reagenzes in die DNA erleichtert, erklärt.[144]

Da Hydroxylamine nicht nur mit abasischen Stellen, sondern auch mit anderen Aldehyden und Ketonen in der Zelle reagieren können, wird die Zuverlässigkeit von Messergebnissen, die mit oben beschriebenen Analysemethoden erzielt wurden, in Frage gestellt. Um das Oximligationsprodukt der abasischen Stellen von allen anderen Oximligationsprodukten zu

unterscheiden, bedarf es sensitiver und selektiver Quantifizierungsmethoden, die eine eindeutige Bestimmung der abasischen Stellen im Spurenbereich ermöglichen. Für diese Anwendung eignet sich die Umkehrphasenflüssigkeitschromatographie (rp-LC) in Verbindung mit Massenspektrometern, die mit ESI *(electrospray ionisation*) als Ionenquelle ausgestattet sind (LC/ESI-MS/MS). So eine Methode bringt mehrere Vorteile mit sich. Erstens benötigt ESI niedrigere Temperaturen im Vergleich zu anderen Ionisationsmethoden. Zweitens wird durch die Tandem-Massenspektrometrie (MS/MS) das Signal-Rausch Verhältnis verbessert und somit die sensitive Detektion des Analyts gewährleistet. Wenn aber die Ionisationseffizienz des Analyts sehr niedrig ist, kann dieses nicht mehr sensitiv detektiert werden. Für die Verbesserung der Detektion solcher Analyten wird oft die chemische Derivatisierung angewendet.[156] So benutzte *Roberts et al.* das Derivatisierungsreagenz NBHA für die Entwicklung einer ersten LC-MS/MS-Methode für die Detektion und Quantifizierung von abasischen Stellen (Abbildung 1.6). Seine Methode erfordert sehr hohe Mengen an DNA (100 µg) und ist somit ungeeignet für die routinierte Laboranalyse.[146]

Zusammenfassend sollte ein für die LC/ESI-MS/MS Analyse von abasischen Stellen geeignetes Derivatisierungsreagenz folgende Eigenschaften aufweisen. Für die Derivatisierung mit Aldehyd-Gruppe von abasischen Stellen sollte es eine Hydroxylamin-Einheit besitzen. Ferner sollte es in geladener Form im Lösungsmittel vorliegen. Vorzugweise sollte ein geeignetes Hydroxylamin ein nicht-polares Fragment aufweisen, damit es von Salzen und störenden Substanzen, die die Ionisationseffizienz negativ beeinflussen, abgetrennt werden kann. Ionen mit nicht-polaren Fragmenten bevorzugen Tröpfchen-Luft-Grenzen und lagern sich somit auf der Tröpfchen-Oberfläche an. Dementsprechend gehen solche Ionen schneller in die Gasphase über und zeigen deutlich höhere ESI-Effizienz. Zusätzlich werden die hydrophoben Substanzen durch eine mobile Phase mit einem hohen Anteil an organischem Lösungsmittel eluiert. Ein hoher Anteil an organischem Lösungsmittel begünstigt die Erzeugung geladener Teilchen im Elektrospray und liefert somit eine höhere ESI-Effizienz. Der Target-Analyt soll schließlich unter *collision-induced dissociation* (CID) effizient fragmentieren und intensive Tochterionen für eine sensitive massenspektrometrische Detektion liefern.[156] So ein Derivatisierungsreagenz

wurde im AK Carell von *Toni Pfaffeneder* in Rahmen seiner Doktorarbeit entwickelt (Abbildung 1.7).[157] Dieses Reagenz wurde für die massenspektrometrische Studien, die im Abschnitt 3.2 beschrieben wurden verwendet.

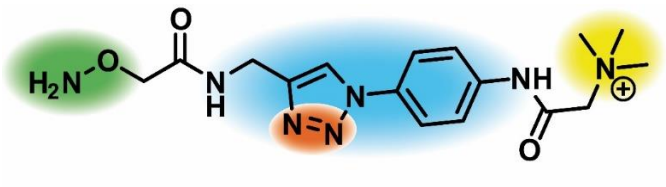


Abbildung 1.7: Das Derivatisierungsreagenz, das für die massenspektrometrische Studie im Abschnitt 3.2 verwendet wurde.

Dieses Derivatisierungsreagenz besitzt eine Hydroxylamin-Einheit (grün), die mit der Aldehyd-Gruppe der abasischen Stelle unter Ausbildung eines stabilen Oximligationsprodukts reagiert. Das positiv geladene quaternäre Alkylamin (gelb) ermöglicht eine Signalintensivierung bei der LC-MS/MS-Analyse und eine Interaktion mit dem negativ geladenen Phosphat-Rückgrat der DNA. Die Phenyl-Triazol-Kernstruktur (blau) begünstigt die Einlagerung in die DNA-Doppelhelix. Die Triazol-Einheit (rot) kann unter Verlust von molekularem Stickstoff leicht fragmentieren und eine hohe Ausbeute an intensiven Tochterionen liefern.

1.4 8-Oxo-dG – Oxidationsmarker

Eine maßgebliche Quelle für AP-Stellen ist die Reparatur oxidativer DNA-Schäden. Für die Entstehung solcher DNA-Schäden sind endogene und exogene Substanzen, sowie elektromagnetische Strahlung, denen die DNA täglich ausgesetzt ist, verantwortlich. Dies kann unter anderem zum Zelltod, Krebs und zu Erbkrankheiten führen. Das 8-Oxo-7,8-dihydro-2'-desoxyguanosin (8-Oxo-dG) ist der am besten untersuchte oxidative DNA-Schaden. Das 8-Oxo-dG dient als Marker für zellulären Stress jeden Ursprungs.[158] Die Mengen an 8-Oxo-dG in genomischer DNA sind stets höher als die anderer Oxidationsprodukten, wie fdU und hmdU, und liegen in einem Bereich von 1-4 x 10^{-6} pro dN.[159-161] Somit repräsentiert dieser Wert in gewisser Weise ein Gleichgewicht zwischen Bildung und Reparatur in der DNA. Für den 8-Oxo-dG-Gehalt in der DNA ist ein Richtwert vereinbart, der nicht mehr als ~ 1x 10^{-6} pro dN betragen darf.[159] Die dafür zuständige Behörde ist das Europäische Standardkomitee für oxidative Schäden (ESCODD). Höhere 8-Oxo-dG-Werte deuten auf Probenpräparationsfehler beim Zellaufschluss hin. Dies wird durch den Zusatz verschiedener Antioxidantien bei der Probenpräparation vermieden.[162-164] Der 8-Oxo-dG-Gehalt hängt mit dem Alterungsprozess in Säugetieren zusammen. Mit fortschreitendem Alter erhöhen sich die 8-Oxo-dG-Werte in verschiedenen Organen und Geweben. Beispielweise wurde so ein 8-Oxo-dG-Anstieg sowohl in der nuklearen, sowie auch in der mitochondrialen DNA von Mäusen beobachtet.[165-166] Außerdem spielt das 8-Oxo-dG eine wichtige Rolle bei der Krebsentstehung. Wie oben erwähnt, kann die DNA-Methylierung der CpG-Inseln in Promotorregionen eines Genes zu der Repression der Genexpression führen. Eine Studie von *Nishida et al.* zeigte eine direkte Korrelation zwischen dem 8-Oxo-dG-Gehalt und dem Methylierungsgrad des Promotors eines Tumor-Supressor Genes (TSG). Der oxidative Stress führt zur Veränderung von aktivem zum gehemmten Chromatinzustand, was zu ungewöhnlich hoher Methylierung von TSG führt.[167] Darüber hinaus wurde berichtet, dass die BER-Enzyme, die an der 8-Oxo-dG-Reparatur beteiligt sind, auch an der DNA-Methylierung, der Demethylierung und der Histon-Modifizierung teilnehmen.[168] Das 8-Oxo-dG wird vor allem durch reaktive Sauerstoffspezies (*reactive oxygen species*; ROS), wie

das Hyperoxid-Anion ($O_2^{\bullet -}$), Wasserstoffperoxid (H_2O_2) und das hochreaktive Hydroxyl-Radikal ($OH^{\bullet}$) gebildet.[169] ROS können sowohl das Genom als auch das Epigenom oxidativ schädigen. Exogene ROS können beispielweise aus Umweltgiften, wie aus unterschiedlichen Redox-Verbindungen, Schwermetallen und durch hochfrequente Strahlung entstehen. Der P450-Metabolismus, Entzündungsprozesse sowie oxidative Phosphorylierung in Mitochondrien, stellen die endogenen Quellen für ROS dar.[158, 170-171] Das 8-Oxo-dG kann neben ROS auch durch einen Ein-Elektron-Mechanismus gebildet werden (Abschnitt 4.2.1). Das Enzym, das für das Ausschneiden von 8-Oxo-dG verantwortlich ist, ist die bifunktionelle 8-Oxoguanin-Glykosylase (Ogg1), die im Zuge einer effizienten BER 8-Oxo-dG aus der DNA entfernt.[77]

2 Aufgabenstellung

Das erste Ziel dieser Doktorarbeit bestand darin, die Desaminierung von mdC zu dT als ein möglicher Initiationsschritt des langgesuchten Prozesses der aktiven DNA-Demethylierung zu untersuchen. Dazu sollte eine [^{13}C, D_3]-Markierung der Methylgruppe des mdCs mittels [^{13}C, D_3-*methyl*]-Methionin metabolisch eingeführt werden. Auf der Grundlage einer in AK *Carell* etablierten massenspektrometrischen Methode sollte eine neue Messmethode in Verbindung mit Isotopenverfolgung für den Nachweis und Quantifizierung von [^{13}C, D_3]-markierten dT, welches durch eine mögliche Desaminierung von [^{13}C, D_3]-mdC entstehen könnte, entwickelt werden. Mit dieser neuen massenspektrometrischer Methode sollten unterschiedliche mESCs untersucht werden.

Ein weiteres Thema, das in der vorgelegten Doktorarbeit untersucht werden sollte, ist die Frage nach der Rolle von BER in dem Prozess der aktiven DNA-Demethylierung. Dazu sollten die abasischen Stellen, die ein Intermediat während BER darstellen, massenspektrometrisch nachgewiesen und wenn möglich auch quantifiziert werden. Der massenspektrometrische Nachweis und die Quantifizierung von abasischen Stellen sollten mittels eines neuen Derivatisierungsreagenzies, das im AK *Carell* entwickelt und synthetisiert wurde, erfolgen. Zunächst sollten möglichst milde Bedingungen für die Derivatisierung ausgearbeitet werden, damit keine artifiziellen abasischen Stellen während dieser Prozedur entstehen und somit die Messergebnisse verfälschen könnten. Parallel dazu sollte eine neue Messmethode, die auf Vorarbeiten von *Toni Pfaffeneder* basiert, entwickelt und für einen möglichst sensitiven Nachweis optimiert werden. Anschließend sollte diese massenspektrometrische Methode für die epigenetisch relevante Fragenstellungen, in erster Linie aber nach der Beteiligung von BER in der aktiven DNA-Demethylierung, angewendet werden.

3 Veröffentlichte Arbeiten

3.1 Umwandlung von 5-Formylcytosin zum Cytosin durch Spaltung von C-C-Bindung *in vivo*

K. Iwan#, R. Rahimoff#, A. Kirchner#, F. Spada#, A. S. Schröder, **O. Kosmatchev**, S. Ferizaj, J. Steinbacher, E. Parsa, M. Müller, T. Carell, *Nat. Chem. Biol.* **2018**, *14*, 72-78. *5-Formylcytosine to cytosine conversion by C-C bond cleavage in vivo.*

(# geteilte Erstautorenschaft)

Zusammenfassung

Wie im Abschnitt 1.2.1 beschrieben, bleibt der Prozess der aktiven DNA-Demethylierung weitgehend ungeklärt. Ein möglicher Mechanismus über den die aktive DNA-Demethylierung verlaufen könnte, ist die C-C-Bindungsspaltung zwischen dem C5-Kohlenstoffatom und der substituierten Methylgruppe des modifizierten Cytosins. Es entsteht direkt das unmodifizierte dC (Abbildung 1.2). Dieses Manuskript beschreibt den Einbau synthetischer isotopenmarkierter und (*R*)-2'-fluorirter dC- und fdC-Derivaten in das Genom von kultivierten Stammzellen. Mittels einer sensitiver UHPLC-MS/MS Methode konnte die Umwandlung von diesen Molekülen in Rahmen einer Isotopenverfolgungsstudie verifiziert werden. Es wurde gezeigt, dass das markierte fdC-Derivat in das korrespondierende markierte dC-Derivat, effizient umgewandelt wird. Höchstwahrscheinlich geschieht diese Umwandlung nach dem Einbau in das Genom. Hier wird das modifizierte fdC über ein C-C-Bindungsbruch in das dC umgesetzt.

Autorenbeitrag

Die Hauptarbeit bei diesem Projekt lag bei *Katharina Iwan*, *René Rahimoff* und *Angie Kirchner*. Für dieses Projekt führte ich die enzymatische Hydrolyse der genomischen DNA aus embryonalen murinen Stammzellen durch, die mit schwermarkiertem dC gefüttert wurden. Die massenspektrometrische Quantifizierung und die Auswertung von Messergebnissen dieser

Proben führte ich ebenfalls durch. Dies geschah in enger Zusammenarbeit mit *Jessica Steinbacher*.

Lizenz

ARTICLE
PUBLISHED ONLINE: 27 NOVEMBER 2017 | DOI: 10.1038/NCHEMBIO.2531
nature chemical biology

5-Formylcytosine to cytosine conversion by C–C bond cleavage *in vivo*

Katharina Iwan[1,2], René Rahimoff[1,2], Angie Kirchner[1,2], Fabio Spada[1,2], Arne S Schröder[1], Olesea Kosmatchev[1], Shqiponja Ferizaj[1], Jessica Steinbacher[1], Edris Parsa[1], Markus Müller[1] & Thomas Carell[1]*

Tet enzymes oxidize 5-methyl-deoxycytidine (mdC) to 5-hydroxymethyl-dC (hmdC), 5-formyl-dC (fdC) and 5-carboxy-dC (cadC) in DNA. It was proposed that fdC and cadC deformylate and decarboxylate, respectively, to dC over the course of an active demethylation process. This would re-install canonical dC bases at previously methylated sites. However, whether such direct C–C bond cleavage reactions at fdC and cadC occur *in vivo* remains an unanswered question. Here we report the incorporation of synthetic isotope- and (*R*)-2′-fluorine-labeled dC and fdC derivatives into the genome of cultured mammalian cells. Following the fate of these probe molecules using UHPLC–MS/MS provided quantitative data about the formed reaction products. The data show that the labeled fdC probe is efficiently converted into the corresponding labeled dC, most likely after its incorporation into the genome. Therefore, we conclude that fdC undergoes C–C bond cleavage in stem cells, leading to the direct re-installation of unmodified dC.

Modification of genomic cytosine modulates the interaction of DNA-binding factors with the genome, thus affecting gene expression and chromatin structure[1,2]. The primary and most prevalent modification is methylation to mdC, which in mammals is catalyzed by the DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b, at least partly in co-operation with the catalytically inactive Dnmt3l[3]. Because Dnmt1 is a maintenance methyltransferase that copies the methylation pattern during replication, the information that this pattern conveys is inherited through cell division. Genomic mdC can be iteratively oxidized to hmdC[4,5], fdC[6,7] and cadC[7,8] by the Ten-eleven translocation (Tet) family of α-ketoglutarate-dependent dioxygenases (**Fig. 1a**). These oxidized cytidine derivatives are prominently detected in DNA isolated from neuronal tissues[5,9,10] and in the genome of embryonic stem cells (**Fig. 1b**), in which their levels change during differentiation[4,7,11]. For example, hmdC can reach levels of up to 1.3% per deoxyguanosine (dG) in DNA isolated from brain[12]. Although the presence of mdC and hmdC is believed to influence the transcriptional activity of genes[13,14], no clear function has yet been assigned to the higher oxidized modifications fdC and cadC. Recent reports, however, show that fdC is a stable[15], or at least semi-stable[16], base in the genome. These discoveries and the identification of specific reader proteins that recognize fdC and cadC suggest that they might have regulatory purposes as well[17–20]. So far, however, fdC and cadC are mainly considered to be intermediates of an active demethylation process that allows cells to replace mdC by a canonical dC nucleotide[20–22]. One such scenario involves fdC and cadC as substrates of the thymine-DNA glycosylase (Tdg), which cleaves the corresponding glycosidic bond. This converts fdC and cadC into abasic sites, which are further processed through base excision repair (BER) as depicted in **Figure 1a**. This Tdg-initiated process establishes an active demethylation pathway, ultimately incorporating unmodified dC nucleotides at former fdC and cadC sites[8,23]. A problem associated with this mechanism is that the removal of every mdC creates a potentially harmful single-strand break intermediate. If an mdC is close to the first in the opposite DNA strand, even double strand breaks may be generated. In addition to these concerns, it was shown that both maternal and paternal genomes of mouse zygotes undergo active demethylation independently of Tdg[24]. To explain such an excision-independent demethylation and provide an alternative to generating harmful repair intermediates, it was suggested that fdC and cadC may directly deformylate and decarboxylate, respectively, under C–C bond cleavage (**Fig. 1a**)[9,25]. Indeed, chemical pathways that allow such a direct deformylation and decarboxylation of fdC and cadC have been described[26,27]. These pathways involve addition of a helper nucleophile to the C6 position of fdC and cadC in a Michael-addition-type reaction followed by deformylation or decarboxylation and final elimination of the helper nucleophile[26]. The chemistry is therefore quite similar to the known reaction mechanisms employed by the Dnmt proteins[3]. Although chemically plausible, it is unclear whether such direct C–C bond cleavage reactions occur within the genome[28]. This process would provide a new and harmless way to convert mdC back into dC in the genome without forming potentially harmful abasic site intermediates.

Here we report a sensitive MS-based isotope tracing study investigating whether a C–C bond cleavage reaction occurs on fdC bases (**Fig. 1c**). We supplemented the medium of cultured mammalian cells with synthetic isotope- and fluorine-labeled fdC derivatives to metabolically integrate the nucleosides as reporter molecules into their genome. After isolation of the genomic DNA, the levels of the modified dC derivatives were measured by ultra-high performance liquid chromatography coupled to tandem mass spectrometry (UHPLC–MS/MS), thereby tracing isotopically or fluorine-labeled dC derivatives.

RESULTS

Isotopically labeled fdC is directly converted into dC

We started the study with a [$^{13}C_5$][$^{15}N_2$]fdC (**1**; **Fig. 2a**), in which all five C-atoms of the ribose ring were exchanged with ^{13}C and the two in-ring nitrogen atoms replaced with ^{15}N (**Supplementary Note**). This provides compound **1**, which is seven mass units heavier than naturally occurring fdC and hence easily distinguishable by MS. The large mass difference allows exact tracing of all transformations that may take place on this base with high sensitivity, because the natural

[1]Center for Integrated Protein Science Munich CiPSM at the Department of Chemistry, Ludwig-Maximilians-Universität München, Munich, Germany. [2]These authors contributed equally to this work. *e-mail: Thomas.carell@lmu.de

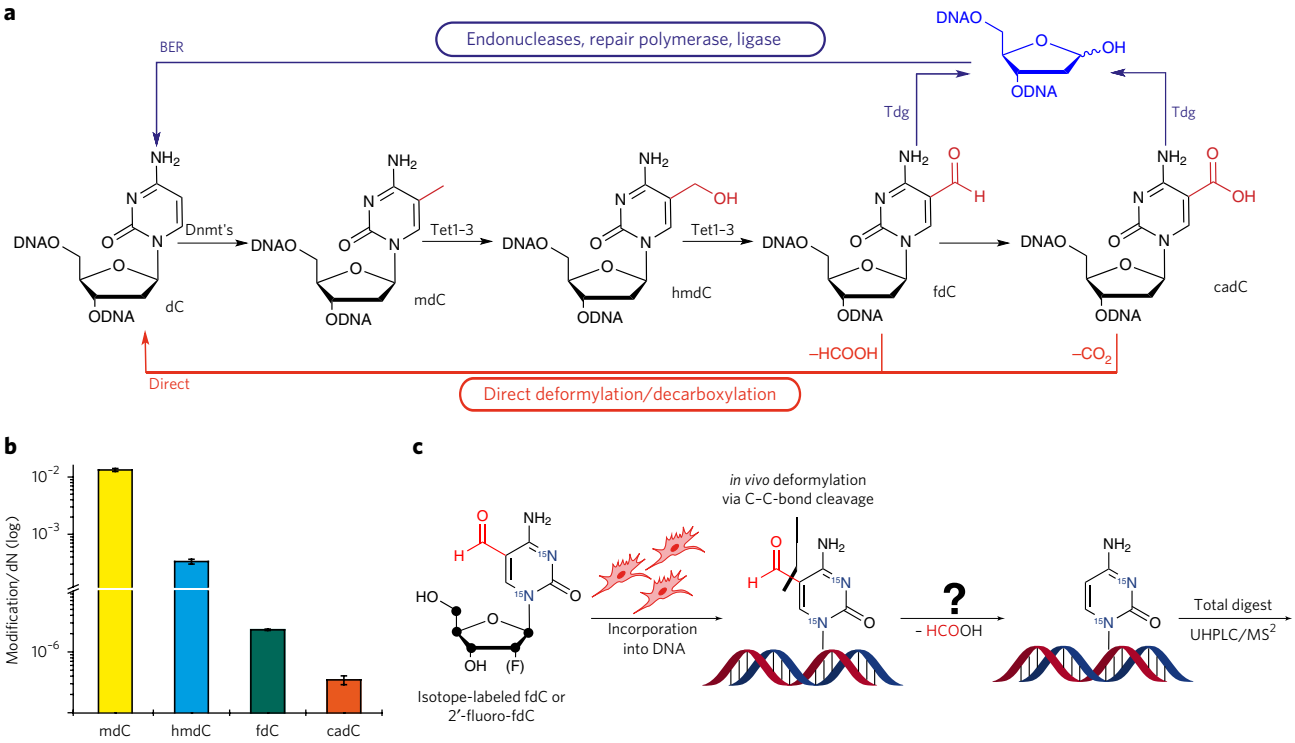


Figure 1 | Isotope tracing studies. (**a**) Suggested pathways of active demethylation. Thymine-DNA glycosylase (Tdg)-based cleavage of the glycosidic bond of 5-formyl-dC (fdC) and 5-carboxy-dC (cadC) results in an abasic site, which initiates a BER process that leads to the replacement of fdC and cadC by canonical dC (blue). Deformylation of fdC (−HCOOH) and decarboxylation ($-CO_2$) of cadC provides dC directly (red). (**b**) UHPLC-coupled MS/MS experiments allow exact quantification of various dC derivatives in mESCs. Mean and s.d. of technical triplicates from two independent cultures are shown. (**c**) Schematic depiction of the feeding experiment using synthetic isotope and fluorine-labeled fdC derivatives, which are metabolically integrated into the genome. ● = ^{13}C atoms.

abundance of such highly isotopically modified dC derivatives is essentially null. Possible transformations are the deformylation of **1** to $[^{13}C_5][^{15}N_2]$dC (**2**) and its deamination to $[^{13}C_5][^{15}N_2]$dU (**3**), followed by methylation of **3** to $[^{13}C_5][^{15}N_2]$dT (**4**). Alternatively, compound **1** can deaminate to $[^{13}C_5][^{15}N_2]$fdU (**5**) and, finally, the deformylated compound **2** can be methylated to $[^{13}C_5][^{15}N_2]$mdC (**6**). Analysis of the MS pattern of **1** showed that cleavage of the glycosidic bond is the dominant fragmentation pathway. This leads to a clearly detectable fingerprint mass transition of m/z = 263.1 to m/z = 142.1 (**Fig. 2b**). Detection of the demodified product dC **2** would be possible on the basis of its mass transition from m/z = 235.1 to m/z = 114.0. For the experiment, we added **1** to the medium of mouse embryonic stem cells (mESCs) under priming conditions. After 3 d, the genomic DNA was isolated using a standard protocol and digested to the individual nucleosides. The obtained nucleoside mixture was analyzed by UHPLC coupled to a triple quadrupole mass spectrometer. We noted that **1** was indeed metabolically incorporated into the genome of mESCs. The mass transition of **1** (m/z = 263.1 to m/z = 142.1) was clearly detectable at a retention time of 5.50 min under our conditions (**Fig. 2b**). By using the mass transitions specific for all the expected natural dC derivatives, we were also able to detect next to **1**, natural mdC, hmdC and fdC (**Fig. 1b**).

Analysis of the nucleoside mixtures revealed the presence of a new isotope-labeled dC derivative at a retention time of 1.95 min displaying the expected mass transition (m/z = 235.1 to m/z = 114.0) for **2**, showing that $[^{13}C_5][^{15}N_2]$fdC is indeed demodified (**Fig. 2c**). To unequivocally prove the identity of **2**, an even heavier isotopically modified dC isotopologue, $[^{13}C_9][^{15}N_3]$dC (**7**), with a characteristic MS transition of m/z = 240.1 to m/z = 119.1 (**Supplementary Results**, **Supplementary Fig. 1**), was used as an internal standard. Compound **7** was added to the nucleoside mixture and it eluted at the same retention time as **2** (**Fig. 2c**), confirming that the UHPLC–MS/MS signal at 1.95 min is caused by **2**. Exact quantification of the conversion showed that when **1** was supplied to mESC cultures at 50 μM for 3 d, steady state incorporation levels of about $3–5 \times 10^{-7}$ of $[^{13}C_5][^{15}N_2]$fdC per deoxynucleoside (dN) in genomic DNA were reached (**Fig. 2d**). We observed higher levels of product **2** (up to a factor of 10), as shown in **Figure 2d**.

Whereas we can exclude spontaneous deformylation of **1** based on previous stability studies[26], **2** can in principle form either by C–C bond cleavage in the genome or at the level of the soluble nucleoside/nucleotide pool. Conversion of **1** in the soluble pool to **2** would then be followed by metabolic incorporation of the **2**-triphosphate into the genome. It is known that soluble **2** is the substrate for other metabolic processes such as deamination to **3** (catalyzed by cytidine deaminase and deoxycytidylate deaminase), which is followed by methylation by thymidylate synthase to give **4** (refs. 29,30). To distinguish the two possible conversion scenarios (genomic DNA or soluble pool), we reasoned that if **1** is converted into **2** in the soluble pool, we would find compounds **3** and, particularly, **4** in the genome.

To investigate the behavior of soluble dC, we cultured mESCs in the presence of an isotopically labeled dC derivative, $[^{13}C_9][^{15}N_3]$dC, and indeed detected the expected presence of the corresponding isotopically labeled deamination products $[^{13}C_9][^{15}N_2]$dT and $[^{13}C_9][^{15}N_2]$dU (**8** and **9**, respectively) in the genome (**Fig. 2e**). In contrast, when **1** was supplied to mESC cultures, we detected next to **2** only the direct deamination product **5** in the genome, but not **3** and **4** (**Fig. 2d**). Even after three consecutive days of feeding compound **1** to mESCs, we were unable to detect even traces of **4** in the genome. This argues against formation of **2** in the soluble pool. We next

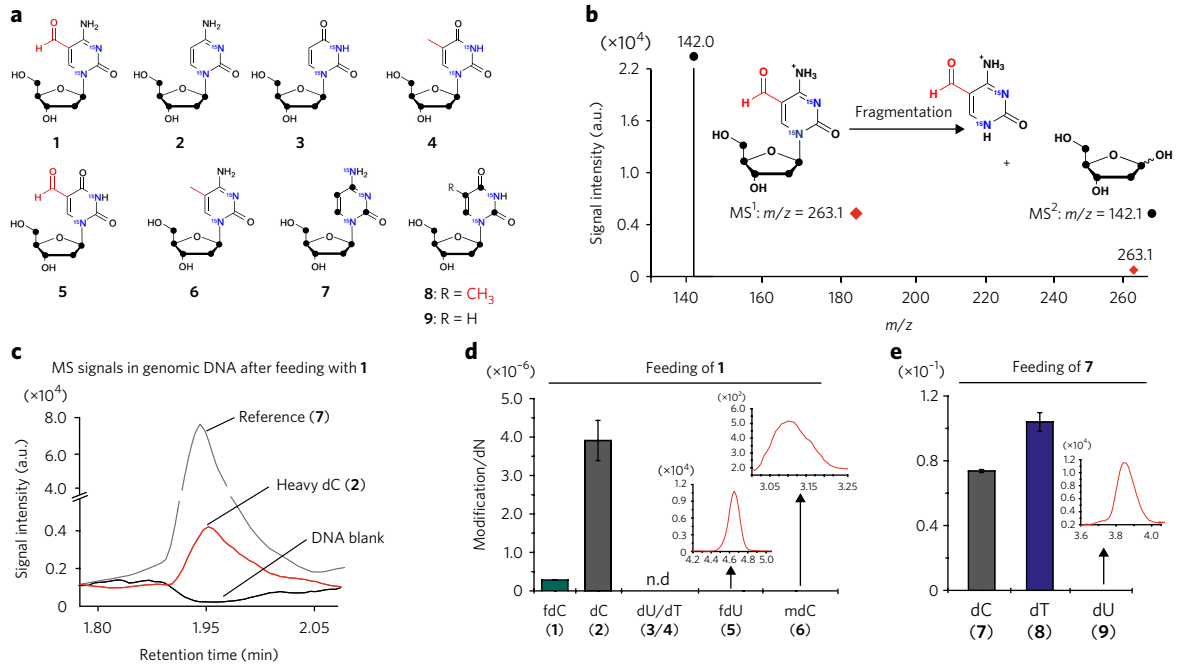


Figure 2 | Conversion of isotopically labeled fdC into dC in mESCs. (**a**) Overview of the compounds that may be detected after feeding of **1** or **7** to mESCs. (**b**) Feeding of **1** to mESCs results in incorporation of the isotopologue into the genomic DNA, as proven by its fingerprint MS transition. a.u., arbitrary units. (**c**) Analysis of gDNA after feeding of **1** shows the presence of labeled dC **2**. (**d**) Quantitative data obtained upon feeding **1** to mESCs. Mean and s.d. of technical triplicate measurements from a single culture are shown. (**e**) Quantitative data obtained upon feeding of **7** to mESCs. Mean and s.d. of technical triplicates from two independent cultures are shown. For a schematic overview of the dC or fdC metabolic pathways see **Supplementary Figure 3**.

analyzed the soluble nucleoside/nucleotide pool directly for the content of **2** after feeding of **1** (**Supplementary Fig. 2**). To this end, we fed **1** to mESCs over 3 d. The cells were washed extensively and finally resuspended in 50% (v/v) MeCN to extract soluble metabolites. After further purification by solid-phase extraction, the nucleotides were dephosphorylated to nucleosides. Analysis of this solution by UHPLC–MS/MS did not give any signal for **2**. All these control experiments suggest that **1** undergoes C–C bond cleavage to **2** directly in the genome and not in the soluble pool, although the latter scenario cannot be fully ruled out because of the complexity of the metabolic pathways (**Supplementary Fig. 3**). Interestingly, we also noted the presence of the remethylated product **6** in the genome of mESCs fed with **1**, but because of the low signal intensity we were unable to obtain quantitative data (**Fig. 2d**).

2′-fluorinated cytosines detect biochemical conversions

To reach higher sensitivity, we experimented with various other isotope-labeled fdC derivatives, and finally found that 2′-fluorinated dC derivatives **10**–**17** are excellent probe molecules (**Fig. 3a** and **Supplementary Fig. 4**). The exchange to an F-atom makes the compound 18 atom units heavier. The compounds also have a slightly shifted retention time (**Fig. 3b**) and give sharp signals in the UHPLC–MS/MS analysis because of a glycosidic bond that is more labile in the MS-fragmentation step. Furthermore, the 2′-(*R*)-configured compounds are well tolerated by the cells used for this study. The F-substituent does affect the ability of the molecule to undergo further biochemical conversions, but the effect is small. (*R*)-2′-F-dC (**10**), for example, is efficiently methylated by DNA methyltransferases[31] and (*R*)-2′-F-mdC (**11**) is also oxidized to (*R*)-2′-F-hmdC (**12**) by the Tet enzymes (**Fig. 3a**,**c**), although a reduced speed of oxidation is observed[32].

To show that the fluorinated compounds are valid probe molecules, we first added **10** to the mESC culture at 0.5 µM, 1.0 µM or 2.5 µM for 3 d. Under these conditions, UHPLC–MS/MS analysis of the isolated genomic DNA showed a clear dose-dependent integration of **10** into the genome, up to 1×10^{-3} per dN. We next searched for other 2′-fluorinated pyrimidine nucleosides and detected a dose-dependent presence of (*R*)-2′-F-dU (**13**) and (*R*)-2′-F-dT (**14**), which was formed by deamination of **10** to **13** followed by methylation to **14** (**Fig. 3c**). In addition, we detected a dose-dependent formation of (*R*)-2′-F-mdC and (*R*)-2′-F-hmdC, confirming that compound **10** is biochemically converted, as expected (**Fig. 3c**).

To quantify the levels of methylation, we synthesized the isotope-labeled compounds (*R*)-2′-[D_3]F-mdC (**18**), (*R*)-2′-[$^{15}N_2$]F-dC (**19**) and (*R*)-2′-[$^{15}N_2$]F-fdC (**20**) and used them as internal standards for quantification (**Fig. 3a**). Upon feeding mESCs with 1 µM **10** for 3 d, we detected around 3% (±0.5%) of **11** relative to **10**, which is similar to proportions observed for the natural bases (**Supplementary Fig. 5**).

(*R*)-2′-F-fdC is converted into (*R*)-2′-F-dC in mESCs

To study the direct C–C bond cleavage process, we again cultured mESCs in the presence of (*R*)-2′-F-fdC (**15**; 350 µM for 3 d), isolated the DNA and analyzed the nucleoside composition. Next to genomic (*R*)-2′-F-fdC (5.7×10^{-7}/dN), we detected (*R*)-2′-F-dC at a level of 7.3×10^{-6}/dN (**Fig. 3d**). Because the nucleosides **13** and **14** were not detected, we suspected again that the observed reaction of **15** to **10** occurs directly within the genome. The detection limit of **13** and **14** is, however, around 5 fmol, and so the compounds may just escape observation.

To substantiate the conclusion that genomic **15** undergoes intragenomic C–C bond cleavage to **10**, we first assessed whether **15** can spontaneously deformylate. To investigate this possibility, we heated an aqueous solution of **15** to 60 °C for 3 d; however, we did not detect compound **10**. We also incubated **15** in culture medium for 3 d and were unable to detect any **10**. Finally, we added a 28-mer oligonucleotide containing a single **15** to culture medium for 3 d, then re-isolated the DNA strand and searched for **10**. Formation

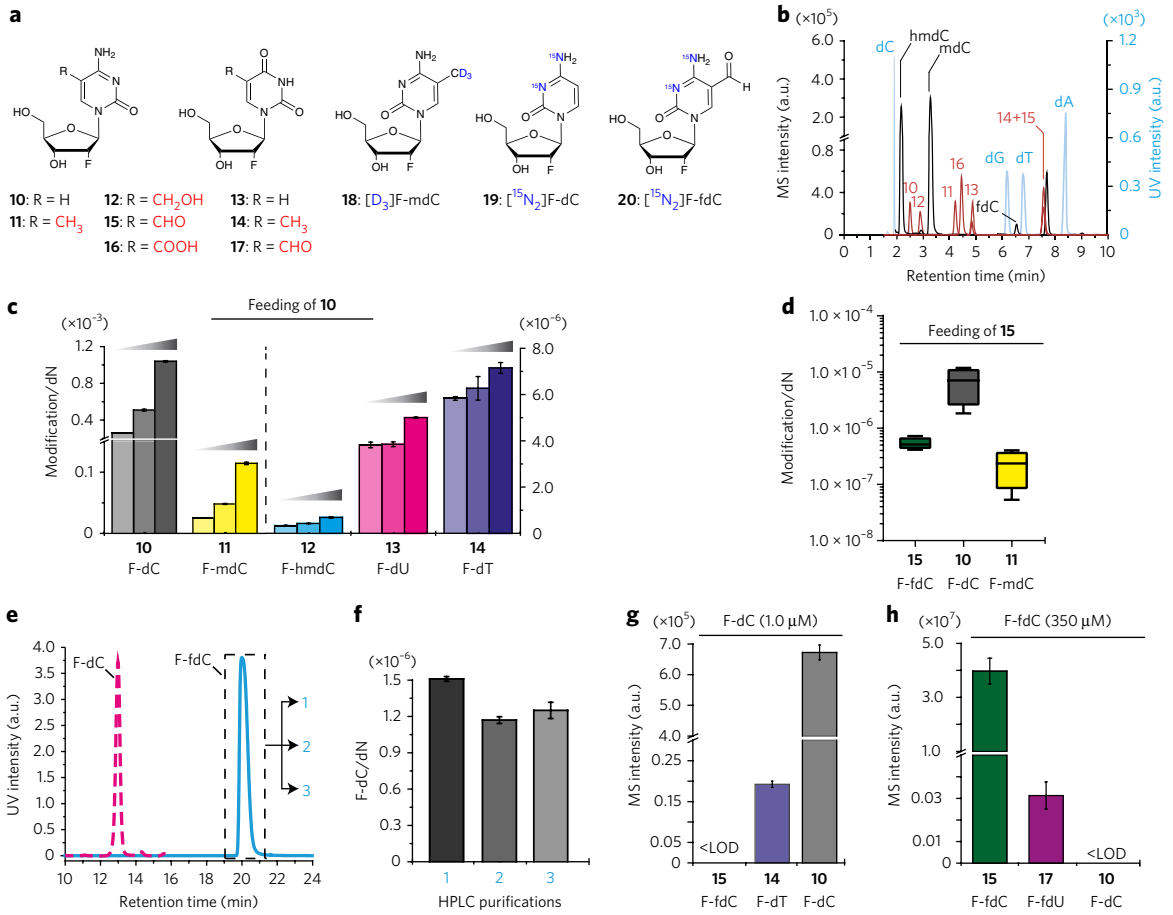


Figure 3 | F-fdC is converted into F-dC within the genome. (**a**) Chemical structures of 2′-fluorinated dC and dU derivatives that were investigated and internal standards used. (**b**) Resulting UV and MS traces of the nucleosides under investigation. Light blue, canonical bases; black, natural dC derivatives; red, fluorinated bases. (**c**) Quantitative data of fluorinated pyrimidine derivatives after feeding with **10** for 3 d at different concentrations (0.5 μM, 1.0 μM and 2.5 μM). A DNA sample from a single culture was measured as technical triplicates. (**d**) Quantitative data of fluorinated pyrimidine derivatives after feeding with **15** at 350 μM for 3 d. Technical triplicates from four independent cultures were measured. (**e**) HPLC of three consecutive purifications of **15** (1–3, blue line). Material from the dashed-line box was collected. The purple dashed line marks the position where a peak of contaminating F-dC would be expected. (**f**) Quantitative data after feeding of **15** that has been purified three consecutive times (from **e**). The levels of the deformylation product **10** remain the same, ruling out any contribution from a possible contamination. (**g**,**h**) Quantitative data of fluorinated pyrimidine derivatives in the soluble pool after feeding **10** at 1 μM (**g**) and **15** at 350 μM (**h**) for 3 d. Technical triplicates from single cultures were measured. In **c**, **e**–**h** mean values with s.d. are shown.

of **10** was again not detected. Together, these experiments exclude background deformylation.

We next analyzed the possibility that **15** is contaminated with traces of **10**. To this end, the purity of **15** was checked by MS, and indeed, **10** was not found. To exclude the presence of even traces of **10** below the detection limit, we performed three consecutive HPLC purifications of **15**. This compound **15** is well separable from **10** because **10** elutes 7.50 minutes earlier than **15** during the HPLC purification (**Fig. 3e**). Feeding of the material **15** obtained from three consecutive purifications resulted in unchanged values of genomic **10**, arguing against the possibility that the detected **10** is an accumulated impurity (**Fig. 3f**).

To further substantiate that the C–C bond cleavage does not occur in the soluble nucleoside/nucleotide pool, we added **10** to the mESC culture for 3 d. UHPLC–MS/MS analysis of the soluble pool allowed us to detect **10**, **14** and, in traces, **13** (**Fig. 3g**). However, when we repeated the study with **15** (**Fig. 3h**), we detected just **15** in the soluble pool plus the deaminated compound (*R*)-2′-F-fdU (**17**), but not **10**. We next determined the medium concentration of **10** that would be needed to reach the measured value for genome-integrated **10** (7.3×10^{-6}/dN; **Fig. 3d**) and found that a concentration of 5–10 nM would be required (**Supplementary Fig. 6**) in the soluble pool. With a detection limit of 30 amol for **10** (40 μL injection), this is a concentration at which **10** is unambiguously detectable.

All of these control experiments support the idea that the C–C bond cleavage to F-dC occurs within the genome. Interestingly, upon feeding of **15** to mESCs, we also detected the methylated derivative **11**, demonstrating that the demodified product **10** is methylated to **11** in the genome. Using the isotopically labeled internal standards (*R*)-2′-[D_3]F-mdC, (*R*)-2′-[$^{15}N_2$]F-dC and (*R*)-2′-[$^{15}N_2$]F-fdC (**Fig. 3a**), we found that remethylation of **10** results in levels of 2.8% (±0.3%) F-mdC (**Fig. 3d**), which is only slightly lower compared to the level observed with direct feeding of **10** (**Supplementary Fig. 5**).

To study the time dependence of the C–C bond cleavage process, we fed compound **15** and measured the genome-integrated levels of **15**, **10** and **11**. Already at 0.5 h, we detected a stable incorporation of **15** (**Fig. 4a**). The C–C bond cleaved product **10** appeared after about 1 h and the levels increased steadily (**Fig. 4b**). After about 4 h, we saw the first remethylated product **11** (**Fig. 4c**). If **10** was a contamination in the preparation of **15**, we would expect faster incorporation kinetics. When we fed both **15** and **10** simultaneously, a steady increase in the level of **10** was already observed after 5 min

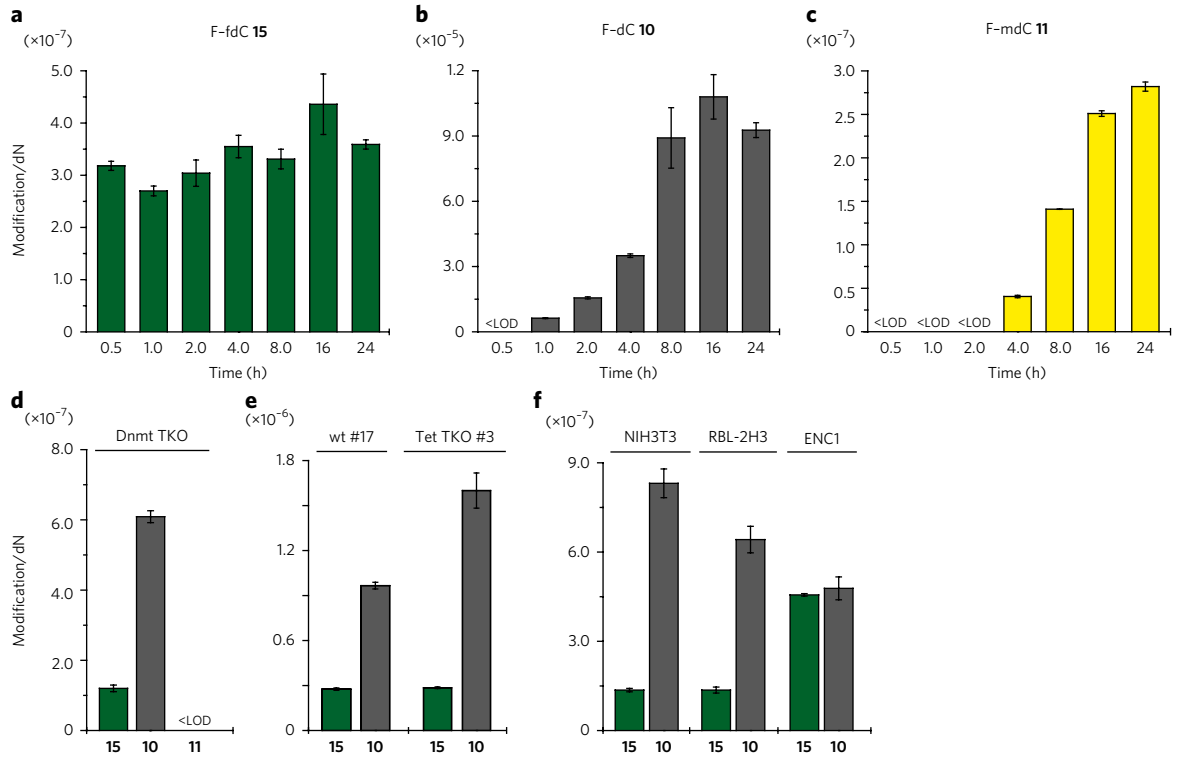


Figure 4 | Demodification of 2′-fluorinated fdC is a rapid process, does not require Dnmt or Tet enzymes and occurs also in somatic cell types. (**a**–**c**) Time-course study showing the genomic build-up of F-fdC **15** (**a**), F-dC **10** (**b**) and F-mdC **11** (**c**) upon metabolic labeling with **15**. LOD, limit of detection. (**d**-**f**) Genomic levels of **15** and **10** upon metabolic labeling of Dnmt triple knockout (TKO) mESCs (**d**), Tet TKO and corresponding wild-type mESC lines (**e**) and various somatic cell lines (**f**) with **15**. Genomic levels of **11** are also shown in **d**. In all panels, mean values and s.d. of technical triplicate measurements from single representative experiments are shown. Data from two additional independent experiments are shown in **Supplementary Figures 8** and **9**.

(**Supplementary Fig. 7**), confirming that our probe nucleosides are quickly incorporated into the genome. These data show that the C–C bond cleavage is a rapid process.

Dnmt or Tet enzymes are not required for demodification

To investigate whether the remethylation of **10** is driven by the known DNA methyltransferases and whether these are involved in demodification, we added **15** (350 μM for 3 d) to mESCs deficient in all active DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b (Dnmt triple knockout (TKO)) and analyzed the DNA. In this experiment, the demodified product **10** is again detected, but the methylated product **11** is not seen (**Fig. 4d** and **Supplementary Fig. 8**), showing that DNA methyltransferases are responsible for methylation of **10** and are not required for demodification of **15**. We finally investigated whether the Tet enzymes are involved in demodification of **15**. However, repeating the feeding experiment with mESCs lacking all three members of the Tet family (Tet TKO) resulted in full demodification activity (**Fig. 4e** and **Supplementary Fig. 9**). The fact that the conversion of **15** to **10** does not change in the absence of Tet proteins is particularly noteworthy. Because Tet enzymes were shown to accept **11** as a substrate and convert it into **12**, **15** and (*R*)-2′-F-cadC **16** (ref. 32), we can exclude the possibility that the observed C–C bond cleavage is in fact a Tet-dependent decarboxylation of **16**. Indeed, this result implies either that **15** is directly deformylated to **10** or that factors other than Tet enzymes can oxidize **15** to **16**, which would then be decarboxylated to **10**. Ascorbic acid has been shown to increase Tet enzymatic activity *in vitro* and the levels of oxidized mdC derivatives *in vivo*[33–35]. Ascorbic acid treatment of mESC cultures fed with **15** indeed resulted in increased levels of naturally occurring fdC and cadC, but had no effect on conversion of **15** into unmodified product **10** (**Supplementary Fig. 10**). This further supports that demodification of **15** to **10** does not depend on the enzymatic activity of the Tet enzymes.

Finally, we tested whether conversion of **15** into **10** occurs in nonpluripotent cells by feeding **15** to cell lines representing a variety of cell types (**Fig. 4f** and **Supplementary Fig. 11**). Albeit to various degrees, we detected the conversion of **15** to **10** in all these cell lines, arguing that this is rather widespread in mammalian cell types. In summary, our data prove that fdC is converted into dC *in vivo* through C–C bond cleavage and strongly suggest that this conversion is an intragenomic process.

DISCUSSION

In recent years, several mechanisms for active erasure of cytosine methylation from the genome have been proposed[25]. Among these, the best-established mechanism entails Tet-mediated iterative oxidation of mdC to fdC or cadC followed by the replacement of these higher oxidized derivatives with unmodified dC through BER[36]. Considering the frequent occurrence of mdC in high-density clusters and prevalent symmetrical configuration at CpG sites in vertebrate genomes, a BER-based erasure mechanism poses a substantial risk of creating clustered single and double strand breaks with potentially deleterious consequences. It is possible that excision of fdC and cadC by Tdg, processing of the abasic site and insertion of unmodified dC are orchestrated by a single multimolecular complex, thus allowing tight control of strand breaks. Alternatively, it is also conceivable that to minimize the potentially deleterious consequences of BER, complementary mechanisms are in place to remove fdC and cadC that do not involve DNA repair. In this context, it should be kept in mind that Tet3-dependent active demethylation of maternal and paternal genomes in the mouse zygote may not require

Tdg[24]. With an isotope-tracing experiment using labeled dC-derived nucleosides in combination with highly sensitive UHPLC–MS/MS detection, we show here that in mammalian cells fdC is converted to dC while the glycosidic bond is kept intact. Evidence is provided to support the theory that the C–C bond cleavage reaction happens when fdC is located inside the genome. This establishes an intragenomic demodification process independent of DNA repair. We also show that this process does not require any of the Tet-family enzymes. Therefore, unless other factors are able to oxidize fdC to cadC, this demodification process is likely a direct deformylation of fdC. Although we have firmly established the occurrence of a C–C bond cleavage of fdC or cadC to dC, the mechanism of this process remains to be defined, including the identification of the factors that mediate the demodification reactions.

We would like to emphasize that in our approach the probe nucleosides are randomly incorporated into the genome through DNA replication. Consequently, we cannot determine the sequence and genomic context wherein the demodification of fdC or cadC to dC takes place. Obviously, this would require a sequencing approach that allows identification of the converted dC bases. In addition, we detected conversion of 2′-fluorinated fdC to 2′-fluorinated dC in mESCs and different somatic cell types. This indicates that the ability to carry out the demodification reaction may be widespread in mammalian cells and tissues rather than being restricted to events of active genomic demethylation known to occur in specific developmental and tissue contexts. Assuming that the deformylation of fdC to dC establishes an active demethylation pathway, we need to emphasize that deformylation reactions and deformylases are widespread in nature. A prominent example is the enzyme lanosterol 14α-demethylase (CYP51A1), which oxidizes the C14α methyl group of lanosterol to a formyl group to achieve deformylation under concomitant introduction of a double bond (dehydrating deformylation). This enzyme, a P450-type monooxygenase, contains a heme cofactor that seems to utilize a nucleophilic Fe peroxyl anion species to attack the substrate[37,38]. Another well studied enzyme that catalyzes deformylation is the aldehyde-deformylating deoxygenase, which again uses a nucleophilic metal bond peroxyl anion radical as the attacking species. This enzyme shortens fatty acid chains by oxidizing the terminal methyl group to a formyl group, which is followed by deformylation[39,40]. In contrast to fdC, these deformylation reactions take place on formyl groups attached to saturated C atoms, whereas in fdC the formyl group is linked to an aromatic heterocycle. Such structures are known for decarboxylations, and they are catalyzed by the enzymes orotate[41] and isoorotate[42] decarboxylase. Indeed, it was suggested that the isoorotate decarboxylase could be a blueprint for a putatively existing cadC decarboxylase[43]. Our data now support the idea that fdC and possibly also cadC are converted to dC by a direct C–C bond cleavage. The questions as to when and where these reactions occur *in vivo* now require the identification of putative catalytic factors.

Received 8 May 2017; accepted 31 October 2017;
published online 27 November 2017

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

References

1. Smith, Z.D. & Meissner, A. DNA methylation: roles in mammalian development.. *Nat. Rev. Genet* **14**, 204–220 (2013).
2. Schübeler, D. Function and information content of DNA methylation. *Nature* **517**, 321–326 (2015).
3. Jeltsch, A. & Jurkowska, R.Z. Allosteric control of mammalian DNA methyltransferases – a new regulatory paradigm *Nucleic Acids Res.* **44**, 8556–8575 (2016).
4. Tahiliani, M. *et al.* Conversion of 5-methylcytosine to -hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930–935 (2009).
5. Kriaucionis, S. & Heintz, N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **324**, 929–930 (2009).
6. Pfaffeneder, T. *et al.* The discovery of 5-formylcytosine in embryonic stem cell DNA. *Angew. Chem. Int. Ed. Engl.* **50**, 7008–7012 (2011).
7. Ito, S. *et al.* Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333**, 1300–1303 (2011).
8. He, Y.F. *et al.* Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333**, 1303–1307 (2011).
9. Globisch, D. *et al.* Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS One* **5**, e15367 (2010).
10. Münzel, M., Globisch, D. & Carell, T. 5-Hydroxymethylcytosine, the sixth base of the genome. *Angew. Chem. Int. Ed. Engl.* **50**, 6460–6468 (2011).
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. Wagner, M. *et al.* Age-dependent levels of 5-methyl-, 5-hydroxymethyl-, and 5-formylcytosine in human and mouse brain tissues. *Angew. Chem. Int. Edn Engl.* **54**, 12511–12514 (2015).
13. Branco, M.R., Ficz, G. & Reik, W. Uncovering the role of 5-hydroxymethylcytosine in the epigenome. *Nat. Rev. Genet.* **13**, 7–13 (2011).
14. Wu, H. & Zhang, Y. Mechanisms and functions of Tet protein-mediated 5-methylcytosine oxidation. *Genes Dev.* **25**, 2436–2452 (2011).
15. Bachman, M. *et al.* 5-Formylcytosine can be a stable DNA modification in mammals. *Nat. Chem. Biol.* **11**, 555–557 (2015).
16. Su, M. *et al.* 5-Formylcytosine could be a semipermanent base in specific genome sites. *Angew. Chem. Int. Ed. Engl.* **55**, 11797–11800 (2016).
17. Raiber, E.A. *et al.* 5-Formylcytosine alters the structure of the DNA double helix. *Nat. Struct. Mol. Biol.* **22**, 44–49 (2015).
18. Song, C.X. *et al.* Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. *Cell* **153**, 678–691 (2013).
19. 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).
20. 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).
21. Hill, P.W., Amouroux, R. & Hajkova, P. DNA demethylation, Tet proteins and 5-hydroxymethylcytosine in epigenetic reprogramming: an emerging complex story. *Genomics* **104**, 324–333 (2014).
22. Wu, X., Inoue, A., Suzuki, T. & Zhang, Y. Simultaneous mapping of active DNA demethylation and sister chromatid exchange in single cells. *Genes Dev.* **31**, 511–523 (2017).
23. Maiti, A. & Drohat, A.C. Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. *J. Biol. Chem.* **286**, 35334–35338 (2011).
24. Guo, F. *et al.* Active and passive demethylation of male and female pronuclear DNA in the mammalian zygote. *Cell Stem Cell* **15**, 447–459 (2014).
25. Wu, S.C. & Zhang, Y. Active DNA demethylation: many roads lead to Rome. *Nat. Rev. Mol. Cell Biol.* **11**, 607–620 (2010).
26. 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).
27. Liutkevičiūtė, Z. *et al.* Direct decarboxylation of 5-carboxylcytosine by DNA C5-methyltransferases. *J. Am. Chem. Soc.* **136**, 5884–5887 (2014).
28. Schiesser, S. *et al.* Mechanism and stem-cell activity of 5-carboxycytosine decarboxylation determined by isotope tracing. *Angew. Chem. Int. Ed. Engl.* **51**, 6516–6520 (2012).
29. Jekunen, A. & Vilpo, J.A. 5-Methyl-2′-deoxycytidine. Metabolism and effects on cell lethality studied with human leukemic cells in vitro. *Mol. Pharmacol.* **25**, 431–435 (1984).
30. Vilpo, J.A. & Vilpo, L.M. Biochemical mechanisms by which reutilization of DNA 5-methylcytosine is prevented in human cells. *Mutat. Res.* **256**, 29–35 (1991).
31. Schröder, A.S. *et al.* Synthesis of (*R*)-configured 2′-fluorinated mC, hmC, fC, and caC phosphoramidites and oligonucleotides. *Org. Lett.* **18**, 4368–4371 (2016).
32. Schröder, A.S. *et al.* 2′-(*R*)-Fluorinated mC, hmC, fC and caC triphosphates are substrates for DNA polymerases and TET-enzymes. *Chem. Commun. (Camb.)* **52**, 14361–14364 (2016).
33. Blaschke, K. *et al.* Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells. *Nature* **500**, 222–226 (2013).

34. Minor, E.A., Court, B.L., Young, J.I. & Wang, G. Ascorbate induces ten-eleven translocation (Tet) methylcytosine dioxygenase-mediated generation of 5-hydroxymethylcytosine. *J. Biol. Chem.* **288**, 13669–13674 (2013).
35. Yin, R. *et al.* Ascorbic acid enhances Tet-mediated 5-methylcytosine oxidation and promotes DNA demethylation in mammals. *J. Am. Chem. Soc.* **135**, 10396–10403 (2013).
36. Wu, X. & Zhang, Y. TET-mediated active DNA demethylation: mechanism, function and beyond. *Nat. Rev. Genet.* **18**, 517–534 (2017).
37. Hargrove, T.Y. *et al.* Substrate preferences and catalytic parameters determined by structural characteristics of sterol 14alpha-demethylase (CYP51) from Leishmania infantum. *J. Biol. Chem.* **286**, 26838–26848 (2011).
38. Lepesheva, G.I. *et al.* CYP51: A major drug target in the cytochrome P450 superfamily. *Lipids* **43**, 1117–1125 (2008).
39. Aukema, K.G. *et al.* Cyanobacterial aldehyde deformylase oxygenation of aldehydes yields n-1 aldehydes and alcohols in addition to alkanes. *ACS Catal.* **3**, 2228–2238 (2013).
40. Jia, C. *et al.* Structural insights into the catalytic mechanism of aldehyde-deformylating oxygenases. *Protein Cell* **6**, 55–67 (2015).
41. Fujihashi, M., Mnpotra, J.S., Mishra, R.K., Pai, E.F. & Kotra, L.P. Orotidine monophosphate decarboxylase--a fascinating workhorse enzyme with therapeutic potential. *J. Genet. Genomics* **42**, 221–234 (2015).
42. Smiley, J.A., Angelot, J.M., Cannon, R.C., Marshall, E.M. & Asch, D.K. Radioactivity-based and spectrophotometric assays for isoorotate decarboxylase: identification of the thymidine salvage pathway in lower eukaryotes. *Anal. Biochem.* **266**, 85–92 (1999).
43. Xu, S. *et al.* Crystal structures of isoorotate decarboxylases reveal a novel catalytic mechanism of 5-carboxyl-uracil decarboxylation and shed light on the search for DNA decarboxylase. *Cell Res.* **23**, 1296–1309 (2013).

Acknowledgments
Tet TKO mESC lines were kindly provided by G.-L. Xu (Shanghai Institutes for Biological Sciences) and R. Jaenisch (Whitehead Institute, MIT, Cambridge). We are grateful to M. Okano and H. Niwa (both at Kumamoto University, Japan) for providing the Dnmt TKO mESC line and the Oct4-YFP reporter cell line, respectively. A.S.S. is supported by a fellowship from the Fonds der Chemischen Industrie. We thank the Deutsche Forschungsgemeinschaft for financial support through the programs: SFB749 (TP A4), SFB1032 (TP A5), SPP1784 and CA275-11/1. We thank the European Union Horizon 2020 program for funding the ERC Advanced project EPiR (741912). Further support is acknowledged from the Excellence Cluster CiPSM (Center for Integrated Protein Science).

Author contributions
K.I. developed and performed the UHPLC–MS/MS studies. R.R. and A.S.S. synthesized the fluorinated and isotopically labeled nucleosides. A.K. designed and performed cell culture work. F.S. designed, supervised and performed cell culture work. O.K. and J.S. analyzed feeding studies of isotopically labeled dC. S.F. contributed to experiments for the analysis of soluble nucleoside pools. M.M. supervised the biochemical work, interpreted and discussed results. T.C. designed and supervised the study. All members discussed results, interpreted data and wrote the manuscript.

Competing financial interests
The authors declare no competing financial interests.

Additional information
Any supplementary information, chemical compound information and source data are available in the online version of the paper. Reprints and permissions information is available online at http://www.nature.com/reprints/index.html. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to T.C.

ONLINE METHODS

Chemical synthesis. Synthetic schemes, detailed procedures and characterization of synthesized products can be found in the **Supplementary Note**. Unless noted otherwise, all reactions were performed using flame- or oven-dried glassware under an atmosphere of nitrogen. Compounds **7** (B.A.C.H. UG) and **10** (Carbosynth) were commercially available. **15** was synthesized as previously described in the literature[32]. Identities of these compounds were confirmed by NMR and LC–MS/MS. Molsieve-dried solvents were used from Sigma-Aldrich, and chemicals were bought from Sigma-Aldrich, TCI, Carbolution and Carbosynth. Technical grade solvents were distilled before extraction or chromatography of compounds. Reaction controls were performed using TLC Plates from Merck (Merck 60 F254), flash-column chromatography purifications were performed on a Merck Geduran Si 60 (40–63 μM). Visualization of the developed TLC plates was achieved through UV absorption or through staining with Hanessian's stain. NMR spectra were recorded in deuterated solvents on Varian VXR400S, Varian Inova 400, Bruker AMX 600, Bruker Ascend 400 and Bruker Avance III HD. HR–ESI–MS spectra were obtained from a Thermo Finnigan LTQ FT-ICR. Infrared (IR) spectroscopic measurements were performed on a PerkinElmer Spectrum BX FT-IR spectrometer with a diamond ATR (attenuated total reflection) unit. HPLC purifications were performed on a Waters Breeze system (2487 dual array detector; 1525 binary HPLC pump) using a Nucleosil VP 250/10 C18 column from Macherey Nagel, HPLC-grade MeCN was purchased from VWR.

Cell culture. Basal medium for mESC culture was DMEM high glucose containing 10% FBS, 2 mM L-glutamine, 100 U/mL penicillin, 100 μg/mL streptomycin, 1× MEM Non-essential Amino Acid Solution and 0.1 mM β-mercaptoethanol (all from Sigma). All mESC lines were maintained in their naïve state on gelatin-coated plates by supplementing basal medium with 1,000 U/mL LIF (ORF Genetics), GSK3 inhibitor CHIR99021 at 3 μM and Mek inhibitor PD0325901 1 μM ("2i"). Metabolic labeling experiments with fluorine- or isotope-labeled nucleosides were performed by plating mESCs in priming conditions consisting of basal mESC medium supplemented with 3 μM CHIR99021 and Wnt pathway inhibitor IWR1-endo at 2.5 μM as previously reported[44]. Under these conditions primed cells remained pluripotent for at least seven days as determined by epifluorescence with an Oct4-YFP knock-in cell line[45]. Priming and labeling was performed for 3 d. Over this period naturally occurring genomic mdC and hmdC (**Fig. 1b**) reached levels similar to those recently reported for epiblast-like cells, which are regarded as the closest *in vitro* counterpart to noncommitted post-implantation epiblast[46,47]. All inhibitors were purchased from Selleckchem. Dnmt TKO J1 mESCs were described in ref. 48. Two independent sets of Tet TKO and respective wt mESC lines were used: wt #17 and Tet TKO #3 were reported in ref. 49 and wt #4 and Tet TKO #29 were described in ref. 50. J1 mESCs are from the 129/Sv/Jae strain, while all Tet TKO and corresponding wt mESC lines are from mixed genetic backgrounds.

The time-course experiment was performed by culturing J1 mESCs under priming conditions for 48 h. The medium was exchanged to priming medium containing 350 μM F-fdC and cells were harvested after 0.5, 1.0, 2.0, 4.0, 8.0, 16 and 24 h, as described.

RBL-2H3, HeLa, NIH3T3 and Neuro-2a cells were cultured in DMEM high glucose containing 10% FBS, 2 mM L-glutamine, 100 U/mL penicillin, 100 μg/mL streptomycin. CHOK1 cells were maintained in DMEM/F12 supplemented as reported above for the other somatic cell lines. ENC1 neural stem cells were cultured as previously described[51]. Cells were exposed to labeled nucleosides for 4 (RBL-2H3 and NIH3T3), 5 (CHO-K1), 6 (Neuro-2a) and 7 d (HeLa and ENC1).

Labeled nucleosides were added to the culture medium at the following concentrations: F-fdC (**15**), 350 μM; F-dC (**10**), 0.5, 1.0 and 2.5 μM; [$^{13}C_5$][$^{15}N_2$]fdC (**1**), 50 μM; [$^{13}C_9$][$^{15}N_3$]dC (**7**), purchased from B.A.C.H. UG), 100 μM.

Isolation of genomic DNA. Cultures were washed with PBS and lysed by adding RLT buffer (Qiagen) containing 400 μM each of 2,6-di-tert-butyl-4-methylphenol (BHT) and desferoxamine mesylate (DM) directly to the plates. Isolation of genomic DNA was performed with Zymo-Spin V, V-E or IIC-XL columns according to the instruction of the ZR-Duet DNA/RNA MiniPrep Kit (Zymo Research) with the following modifications. DNA was sheered by bead milling in 2 mL microfuge tubes using one 5-mm diameter stainless steel bead per tube and a MM400 bead mill (Retsch) set at 30 Hz for 1 min. Lysates were then loaded onto spin columns and the bound material was first incubated for 15 min with Genomic Lysis Buffer (Zymo Research) supplemented with 0.2 mg/mL RNase A (Qiagen). After washing genomic DNA fragments were eluted with water containing 0.4 μM of each BHT and DM.

DNA digestion. 0.5–10 μg of genomic DNA in 35 μL H_2O were digested as follows: An aqueous solution (7.5 μL) of 480 μM $ZnSO_4$, containing 42 U nuclease S1 (*Aspergillus oryzae*, Sigma-Aldrich), 5 U Antarctic phosphatase (New England BioLabs) and specific amounts of labeled internal standards were 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 0.2 U snake venom phosphodiesterase I (*Crotalus adamanteus*, USB corporation), the sample was incubated for 3 h at 37 °C or overnight and then stored at −20 °C. Prior to UHPLC–MS/MS analysis, samples were filtered by using an AcroPrep Advance 96 filter plate 0.2 μm Supor (Pall Life Sciences).

LC/MS-MS analysis of DNA samples. Quantitative UHPLC–MS/MS analysis of digested DNA samples was performed using an Agilent 1290 UHPLC system equipped with a UV detector and an Agilent 6490 triple quadrupole mass spectrometer. Prior to every measurement series, external calibration curves were measured to quantify the levels of the F-nucleosides (**Supplementary Fig. 12**). Additionally, (*R*)-2′-[$^{15}N_2$]F-dC (**19**), (*R*)-2′-[$^{15}N_2$]F-hmdC (**37**) and (*R*)-2′-[$^{15}N_2$]F-fdC (**20**) were used to validate the resulting peaks by co-injection. For exact quantification of fluorinated nucleosides also internal quantification with stable isotope dilution techniques for F-fdC, F-dC and F-mdC were developed (**Supplementary Fig. 13**). Natural nucleosides were quantified with the stable isotope dilution technique. An improved method, based on earlier published work[26,28,52–54] was developed, which allowed the concurrent analysis of all nucleosides in one single analytical run[11]. The source-dependent parameters were as follows: gas temperature 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 2,500 V in the positive ion mode, capillary voltage −2,250 V in the negative ion mode and nozzle voltage 500 V. The fragmentor voltage was 380 V/ 250 V. Delta EMV was set to 500 (positive mode) and 800 (negative mode). Compound-dependent parameters are summarized in **Supplementary Tables 1–4**. Chromatography was performed by a Poroshell 120 SB-C8 column (Agilent, 2.7 μm, 2.1 mm × 150 mm) at 35 °C using a gradient of water and MeCN, each containing 0.0085% (v/v) formic acid, at a flow rate of 0.35 mL/min: 0–4 min; 0– 3.5% (v/v) MeCN; 4–7.9 min; 3.5–5% MeCN; 7.9–8.2 min; 5–80% MeCN; 8.2–11.5 min; 80% MeCN; 11.5–12 min; 80–0% MeCN; 12–14 min; 0% MeCN. The effluent up to 1.5 min and after 12 min was diverted to waste by a Valco valve. The autosampler was cooled to 4 °C. The injection volume was 39 μL.

Quantification of nucleosides. Prior to every sample set, calibration curves to quantify all fluorine labeled nucleosides were measured under the same conditions and settings. All calibration curves are valid within the range of 1-500 fmol with five measuring points and measured as technical triplicates. **Supplementary Fig. 12** shows representative calibration curves for all Fluoronucleosides used for the quantification.

To obtain the internal calibration curves for exact quantification, we analyzed each standard, namely (*R*)-2′-[$^{15}N_2$]F-fdC (**20**; n = 205 fmol), (*R*)-2′-[$^{15}N_2$]F-dC (**19**; n = 793 fmol) and (*R*)-2′-[D_3]F-mdC (**18**; n = 461 fmol), in comparison to the corresponding nonlabeled nucleoside with constant concentration. Technical triplicates were measured, and the linear regression was applied using Origin 6.0 (Microcal). Therefore, the ratio of the area under the curve of unlabeled nucleoside (A) to the labeled standard (A*) was plotted against the ratio of the amount of unlabeled nucleoside (n) to the labeled one (n*) (see **Supplementary Fig. 13**). Acceptable precision (<20% relative s.d. within each triplicate) and accuracy (80%–120%) was achieved for all three calibration curves. The accuracy is calculated as the ratio of the measured to the calculated ratios of the areas (A/A*) under the curves in percent. The ratios of the areas (A/A*) can be calculated by using the linear equations for the corresponding ratio of amount (n/n*).

The lower limit of detection was defined as the detected amount, which is three times higher than the blank response (LOD). The lower limit of detection (LLOQ) and the upper limit of detection (ULOQ) are the lowest and the highest amounts (n) and the ratio of the amounts (A/A*) fulfilling the requirements of the corresponding linear equation, respectively.

Nucleoside stability test. Compounds **1** and **15** were incubated at 100 μM in mESC culture medium at 37 °C and 5% CO_2 for 3 d. For the recovery of the nucleosides Supel-Select SPE HLB cartridges from Sigma-Aldrich were used. Prior to use, the cartridges were equilibrated with MeOH, followed by acidified H_2O (with HCl to pH = 4). The pH of the samples was adjusted to 4, and the acidic solution was loaded on the cartridges. After washing with 10 mL of H_2O, the cartridges were dried *in vacuo.* The nucleosides were eluted with MeOH/MeCN (1:1), evaporated to dryness via speedvac and resuspended in H_2O.

Oligonucleotide stability test. An oligonucleotide (6.9 pmol) containing one F-fdC (28-mer) was incubated in mESC culture medium at 37 °C and 5% CO_2 for 3 d. For the recovery of the oligonucleotide, Oligo Clean & Concentrator from Zymo Research was used according to the manual. The resulting oligonucleotide was dissolved in H_2O and digested as described for genomic DNA.

Extraction of nucleoside/nucleotide soluble pools. J1 mESCs were plated under priming conditions (as described above) for 3 d. The culture medium was supplemented with 1.0 μM F-dC (**10**), 50 μM [$^{13}C_5$][$^{15}N_2$]fdC (**1**) or 350 μM F-fdC (**15**). Cells were washed twice with PBS (Sigma-Aldrich), harvested by trypsinization and pelleted by centrifugation for 3 min at 300*g*. 500 μL ice-cold 50% (v/v) acetonitrile was added dropwise to the pellet and vortexed[55]. The mixture was incubated on ice for 10 min. The insoluble fraction was then separated from the soluble pool by centrifugation for 10 min at 21,000 × *g* at 0 °C. The supernatant was removed and used for nucleoside isolation. The soluble fraction containing the nucleosides was dried by lyophilization and metabolites were purified using Supel-Select SPE HLB cartridges (as described in the nucleoside stability test) before UHPLC–MS/MS analysis.

Life Sciences Reporting Summary. Further information on experimental design and reagents is available in the **Life Sciences Reporting Summary.**

Data availability. The data that support the findings of this study are available from the corresponding author upon reasonable request.

44. Kim, H. *et al.* Modulation of β-catenin function maintains mouse epiblast stem cell and human embryonic stem cell self-renewal. *Nat. Commun.* **4**, 2403 (2013).
45. Toyooka, Y., Shimosato, D., Murakami, K., Takahashi, K. & Niwa, H. Identification and characterization of subpopulations in undifferentiated ES cell culture. *Development* **135**, 909–918 (2008).
46. Shirane, K. *et al.* Global landscape and regulatory principles of DNA methylation reprogramming for germ cell specification by mouse pluripotent stem cells. *Dev. Cell* **39**, 87–103 (2016).
47. Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S. & Saitou, M. Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. *Cell* **146**, 519–532 (2011).
48. Tsumura, A. *et al.* Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. *Genes Cells* **11**, 805–814 (2006).
49. Hu, X. *et al.* Tet and TDG mediate DNA demethylation essential for mesenchymal-to-epithelial transition in somatic cell reprogramming. *Cell Stem Cell* **14**, 512–522 (2014).
50. Dawlaty, M.M. *et al.* Loss of Tet enzymes compromises proper differentiation of embryonic stem cells. *Dev. Cell* **29**, 102–111 (2014).
51. Liu, N. *et al.* Intrinsic and extrinsic connections of Tet3 dioxygenase with CXXC zinc finger modules. *PLoS One* **8**, e62755 (2013).
52. Cao, H. & Wang, Y. Collisionally activated dissociation of protonated 2′-deoxycytidine, 2′-deoxyuridine, and their oxidatively damaged derivatives. *J. Am. Soc. Mass Spectrom.* **17**, 1335–1341 (2006).
53. Spruijt, C.G. *et al.* Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* **152**, 1146–1159 (2013).
54. Wang, J. *et al.* Quantification of oxidative DNA lesions in tissues of Long-Evans Cinnamon rats by capillary high-performance liquid chromatography-tandem mass spectrometry coupled with stable isotope-dilution method. *Anal. Chem.* **83**, 2201–2209 (2011).
55. Dietmair, S., Timmins, N.E., Gray, P.P., Nielsen, L.K. & Krömer, J.O. Towards quantitative metabolomics of mammalian cells: development of a metabolite extraction protocol. *Anal. Biochem.* **404**, 155–164 (2010).

doi:10.1038/nchembio.2531

nature research

Corresponding author(s): Thomas Carell

☐ Initial submission ☐ Revised version ☒ Final submission

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

Describe how sample size was determined.

In most cases cell culture experiments were independently repeated two to four times and samples from each experiment were measured as technical triplicate. Exceptions: Figures 2d, 3c, 3e-h, 4a-c and 4f. These were single experiments measured as technical triplicate. In these cases accurate (repeated) quantification was not considered crucial as it would not affect the fundaments of our story and Fig. 4f just points to deformylation occurring in other types of cells as well.

2. Data exclusions

Describe any data exclusions.

No data was excluded.

3. Replication

Describe whether the experimental findings were reliably reproduced.

The data was reliably reproduced.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

Describe how samples were allocated to groups. If allocation was not random, describe how covariates were controlled. If this is not relevant to your study, explain why.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

Describe the extent of blinding used during data acquisition and analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

Nature Chemical Biology: doi:10.1038/nchembio.2531

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒	A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	A statement indicating how many times each experiment was replicated
☒	☐	The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☒	☐	A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☒	☐	The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted
☐	☒	A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒	Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.	Microsoft Excel, Origin, Agilent MassHunter Analysis (Quant)

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.	No restrictions.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	No antibodies were used.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	Stated for mESCs and NSCs (ENC1) in the Methods and Acknowledgments sections. CHO-K1, HeLa, Neuro2a, RBL-2H3, NIH3T3 were provided from stock centers (ATCC and its distributors).
b. Describe the method of cell line authentication used.	*Describe the authentication procedures for each cell line used OR declare that none of the cell lines used have been authenticated OR state that no eukaryotic cell lines were used.*
c. Report whether the cell lines were tested for mycoplasma contamination.	Not systematically, only some of the cultures were tested.
d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	No commonly misidentified cell lines were used.

Nature Chemical Biology: doi:10.1038/nchembio.2531

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

For laboratory animals, report species, strain, sex and age OR for animals observed in or captured from the field, report species, sex and age where possible OR state that no animals were used.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

Provide all relevant information on human research participants, such as age, gender, genotypic information, past and current diagnosis and treatment categories, etc. OR state that the study did not involve human research participants.

Nature Chemical Biology: doi:10.1038/nchembio.2531

3.2 5-Formyl- und 5-Carboxydesoxycytidin verursachen keine Akkumulation von schädlichen Reparaturintermediaten in Stammzellen.

R. Rahimoff#, **O. Kosmatchev#**, A. Kirchner#, T. Pfaffeneder, F. Spada, V. Brantl, M. Müller, T. Carell, *J. Am. Chem. Soc.* **2017**, *139*, 10359-10364. *5-Formyl- and 5-Carboxydeoxycytidines Do Not Cause Accumulation of Harmful Repair Intermediates in Stem Cells.*

(# geteilte Erstautorenschaft)

Zusammenfassung

Wie im Abschnitt 1.2.1 bereits beschrieben, verläuft ein möglicher Weg der aktiven DNA-Demethylierung über eine stufenweise Oxidation von mdC über hmdC zu fdC und cadC. Letztere werden im Laufe einer effektiven Basenexzisionsreparatur, vermutlich mittels Tdg, aus Genom entfernt. Es entstehen dabei potentiell schädliche abasische Stellen. Dieses Manuskript beschreibt die Entwicklung einer neuen, auf Massenspektrometrie-basierenden Methode. Diese Methode, in Verbindung mit einem neuen Hydroxylaminreagenz, ermöglicht die Detektion und eine exakte Quantifizierung von abasischen Stellen parallel zu epigenetisch relevanten und kanonischen Nukleosiden (mdC, hmdC, cadC, fdC, dC etc.) im Genom embryonaler Stammzellen der Mäuse, sowie HEK (*human embryoid kidney*) und ENC1 (*embryonic neural stem cells*). Mit der vorgestellten Methode ist es außerdem zum ersten Mal möglich, nicht nur abasische Stellen, sondern auch β-eliminierten Stellen, die ein Produkt der bifunktionellen Glykosylasen darstellen (Abschnitt 1.3.1), zu quantifizieren. Durch Fütterungsexperimente mit isotopenmarkierten Nukleosiden konnten die ursprünglichen Nukleobasen der detektierten abasischen Stellen nachverfolgt werden. Es wurde gezeigt, dass die Mengen abasischer Stellen, die von caC und fC stammen könnten, 1 bis 2 Mal niedriger sind, als der fdC- oder cadC-Gehalt dieser Zellen. Das bedeutet, dass die beiden oxidierten mdC-Spezies keine Akkumulation von schädlichen abasischen Stellen im Genom von Stammzellen verursachen.

Autorenbeitrag

Basierend auf Vorarbeiten von *Toni Pfaffeneder*[157] entwickelte ich für dieses Manuskript eine neue massenspektrometrische Methode für die Detektion und Quantifizierung von abasischen Stellen und β-Eliminierungsprodukten in genomischer DNA. Enzymatische Hydrolyse, Messungen und Auswertungen der Messergebnisse aller Proben führte ich ebenfalls durch. Die Methodenvalidierung geschah in enger Zusammenarbeit mit *Renè Rahimoff*. Außerdem isolierte ich die genomische DNA in enger Zusammenarbeit mit *Angie Kirchner*. Die Planung der Experimente, sowie die Interpretation der Messergebnisse erfolgte in enger Zusammenarbeit mit *Renè Rahimoff* und *Angie Kirchner*.

Lizenz

JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Article
pubs.acs.org/JACS

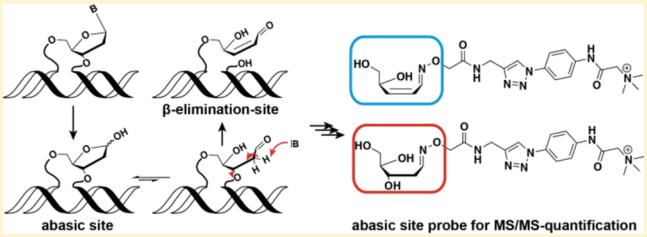
5-Formyl- and 5-Carboxydeoxycytidines Do Not Cause Accumulation of Harmful Repair Intermediates in Stem Cells

René Rahimoff,[†] Olesea Kosmatchev,[†] Angie Kirchner,[†] Toni Pfaffeneder, Fabio Spada, Victor Brantl, Markus Müller, and Thomas Carell*

Center for Integrated Protein Science at the Department of Chemistry, LMU Munich, Butenandtstrasse 5-13, Munich 81377, Germany

Supporting Information

ABSTRACT: 5-Formyl-dC (fdC) and 5-carboxy-dC (cadC) are newly discovered bases in the mammalian genome that are supposed to be substrates for base excision repair (BER) in the framework of active demethylation. The bases are recognized by the monofunctional thymine DNA glycosylase (Tdg), which cleaves the glycosidic bond of the bases to give potentially harmful abasic sites (AP-sites). Because of the turnover of fdC and cadC during cell state transitions, it is an open question to what extent such harmful AP-sites may accumulate during these processes. Here, we report the development of a new reagent that in combination with mass spectrometry (MS) allows us to quantify the levels of AP-sites. This combination also allowed the quantification of β-elimination (βE) products, which are repair intermediates of bifunctional DNA glycosylases. In combination with feeding of isotopically labeled nucleosides, we were able to trace the intermediates back to their original nucleobases. We show that, while the steady-state levels of fdC and cadC are substantially increased in Tdg-deficient cells, those of both AP- and βE-sites are unaltered. The levels of the detected BER intermediates are 1 and 2 orders of magnitude lower than those of cadC and fdC, respectively. Thus, neither the presence of fdC nor that of cadC in stem cells leads to the accumulation of harmful AP- and βE-site intermediates.



INTRODUCTION

The loss of a base is a major damaging event in the genome that leads to the formation of an abasic site (AP-site). Abasic sites are formed spontaneously or during base excision repair, when cells remove damaged bases from the genome.[1] This repair process requires the action of specific DNA glycosylases, which recognize the damaged base (Figure 1A).[2−4] Two different types of repair glycosylases are known. Monofunctional glycosylases produce AP-sites (Figure 1A, step a) that are subsequently removed by the action of endonucleases such as the apurinic/apyrimidinic endonuclease (Ape) 1 (Figure 1A, step b),[5,6] which introduces a free 3′-OH-group and a 5′-phosphate ester (5′-dRP) that is further processed by the action of polymerase β (Pol β). Bifunctional glycosylases catalyze a β-elimination (βE) reaction (Figure 1A, step c), which gives transient βE-intermediates. These can then be converted into a single nucleotide gap by a subsequent δ-elimination reaction (Figure 1A, step d). Both steps (c and d) are carried out by the bifunctional glycosylases.[7]

Recently, it was discovered that the monofunctional glycosylase Tdg (Thymine DNA glycosylase)[8,9] removes the newly discovered bases 5-formyl-cytosine (fdC)[10] and 5-carboxy-cytosine (cadC).[9,11] Pronounced amounts of these bases can be found particularly in stem cells, and fdC was recently also discovered in the genome of brain cells.[12−14] Because fdC and cadC are supposed to be intermediates of an active demethylation process, the data suggest that Tdg-mediated base excision repair contributes to epigenetic reprogramming. One of the main questions associated with the Tdg-mediated repair process is to what extent potentially harmful AP- and βE-site intermediates are generated in the genome in response to the presence of fdC and cadC.

Here, we report the development of reagent **1**[15] (Figure 1B) that in combination with highly sensitive UHPLC-triple quadrupole mass spectrometry (MS)[16] enables quantification of both AP-sites and βE intermediates with unprecedented sensitivity (~100 intermediates per genome). The method was used to quantify the steady-state levels of AP- and βE-sites during BER of fdC and cadC.

RESULTS AND DISCUSSION

Design and Synthesis of a Probe for MS-Based Abasic Site Quantification. Reagent **1** contains a reactive hydroxylamine unit, able to form stable and defined reaction products with the aldehydic form of both AP-sites and βE-intermediates.[17,18] We discovered that the formed adducts (Figure 1B) do not disturb the action of hydrolytic enzymes (Supporting Information, section 3.8), so that the formed reaction products can be excised quantitatively from the genome. To achieve ultrahigh sensitivity, we developed the reagent for triple-quadrupole MS detection. In this method, not only is the

Received: April 27, 2017
Published: July 17, 2017

DOI: 10.1021/jacs.7b04131
J. Am. Chem. Soc. 2017, 139, 10359−10364

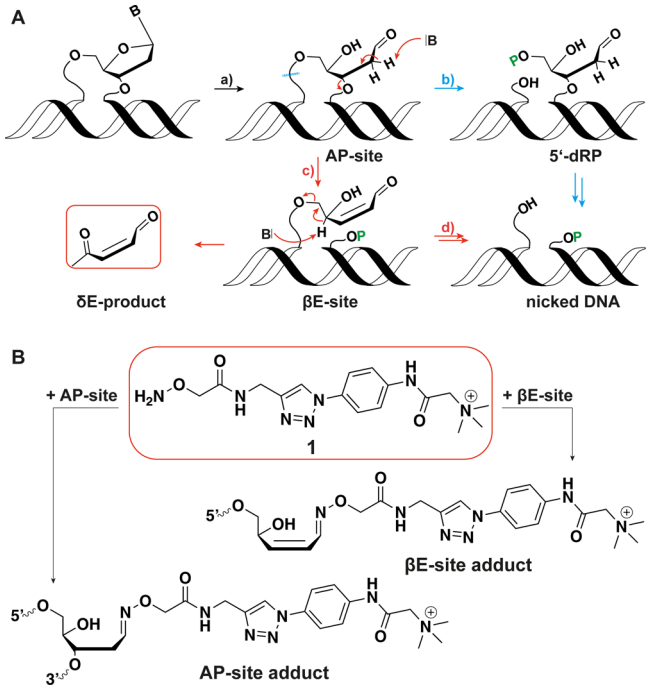


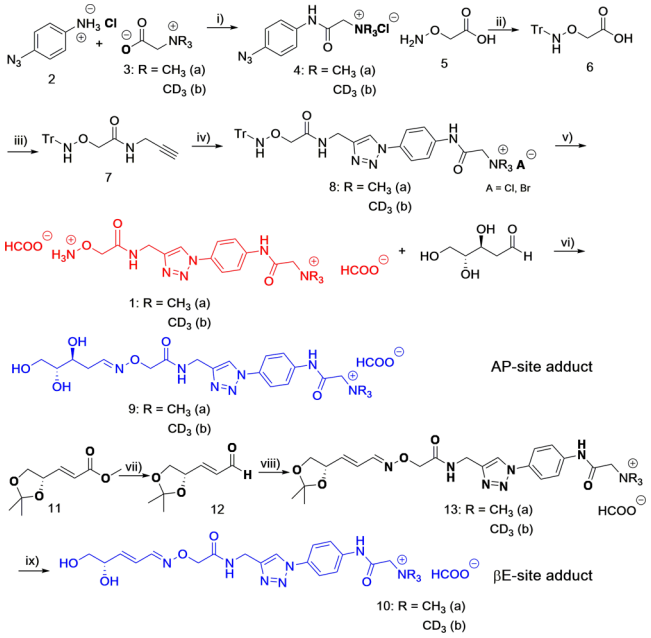
Figure 1. Mechanism of the base excision repair process. (A) Chemistry of BER with formation of AP- and βE-sites. Blue pathway: Monofunctional glycosylases act in concert with Ape1 and Pol β. Red pathway: Bifunctional glycosylases. (B) Depiction of reagent **1** and of the reaction products that are formed when **1** reacts with AP- and βE-sites.

molecular mass of the analyte observed, but also a specific mass transition that arises when the parent ion fragments. **1** therefore contains a triazole heterocycle that fragments in the triple quadrupole MS via collision-induced dissociation (CID) to give a loss of N_2. The permanent positive charge that was designed into **1** ensures highest possible sensitivity in the mass spectrometer and was found to be beneficial, assuring fast reaction with the AP- and βE-sites that are embedded in the negatively charged DNA duplex. The coupling to mass spectrometry is essential to distinguish AP- and βE-intermediates, which has not been possible to date with previously established methods.[17,19,20]

The synthesis of reagent **1** is shown in Scheme 1. It starts with *p*-azidoaniline **2**, which is reacted with trimethylamino glycine **3** using TBTU as the coupling reagent to give the azido-amide **4**. A stable isotopologue of reagent **1a**, needed for exact MS-based quantification (vide infra), is at this step accessed via a $(CD_3)_3$-glycine derivative, which introduces nine D atoms making **1b** 9 au heavier (**1b**). *O*-(Carboxymethyl)-hydroxylamine **5** was in parallel Trt-protected to **6**,[21] and **6** was reacted with propargyl amine using again TBTU as the coupling reagent to give the alkyne **7**. Reaction of **4** with **7** via a Cu(I)-catalyzed azide−alkyne click reaction furnished the triazole **8**. Cleavage of the Trt-group under harsh acidic conditions provided the reagents **1** in a light (CH_3, **a**) and a heavy version (CD_3, **b**), with a $\Delta m/z = 9$ in just five steps with a total yield of 47%.

The reagent was purified twice by reversed phase HPLC to establish a purity of >99.9%. For exact mass spectrometry-based quantification, we next needed to prepare the AP- and βE-site reaction products with **1** in unlabeled (**a**) and isotopically modified (**b**) versions to be used as internal standards. We

Scheme 1. Synthesis of 1 (Red) in a Light (1a) and Heavy (1b) Form and of the Light and Heavy Internal Standards 9a/b and 10a/b (Blue)[a]



[a](i) TBTU, DIPEA, DMF, rt, 16 h, 90%; (ii) Trt-Cl, NEt_3, pyridine, rt, 22 h, 74%; (iii) propargylamine, TBTU, DIPEA, DCM, rt, 15 h, 92%; (iv) 7 + **4a**/**b**, CuBr·SMe_2, H_2O/DCM (1:1), rt, 16 h, 77%; (v) 6 M HCl, DCM/H_2O (1:1), rt, 1 h, quant.; (vi) H_2O, 30 °C, 16 h, HPLC, 15%; (vii) (1) DIBAL-H, DCM, −78 °C to rt, (2) DMP, DCM, 0 °C to rt, o/n, 47% over two steps; (viii) (1) **1a**/**b**, $CHCl_3/H_2O$ (1:1), rt, 16 h, 68%, (2) pTSA·H_2O, H_2O, 25 °C, o/n, HPLC (2×), 14%.

reacted both isotopologues **1a**/**b** with 2′-deoxyribose, which provided the AP-site reaction products **9a**/**b**. To prepare the βE products **10a**/**b**, we reduced the acetonide-protected methyl-ester **11** with DIBAL-H to the allyl alcohol, which was subsequently oxidized selectively to the aldehyde **12** using the Dess−Martin reagent. Reaction of **12** with reagent **1a**/**b** and final cleavage of the acetonide protecting group furnished the desired compounds **10a**/**b**, in a light and heavy version, respectively. Compounds **9a**/**b** and **10a**/**b** were finally purified by reversed phase HPLC to purities >99.9%.

We next developed the MS method (Figure 2A). Analysis of the AP-site reaction product **9a** provided a symmetric signal (at t = 9.5 min, for gradient see Supporting Information, section 3.3) for the MS transitions m/z = 478.2 → 450.2 (quantifier) and m/z = 478.2 → 192.1 (qualifier) caused by the two molecular fragments formed after CID fragmentation. One fragmentation is caused by the loss of N_2, and the second fragmentation involves the formation of an aryl radical.

The first MS transition (quantifier) was used for exact quantification, while the second MS transition (qualifier) is needed for additional structure validation. The isotopologue **9b** showed the expected mass transitions m/z = 487.3 → 459.3 and m/z = 487.3 → 201.2. Similar data were obtained for the βE-reaction product **10a** and its isotopologue **10b** (Figure 2A). We next performed a dilution experiment with **9a** and **10a** and found that we could detect the compounds within a limit of detection of only 110 amol. This is due to the fact that **1** was designed so that its AP- and βE-adducts **9** and **10** have advantageous ionization (permanent positive charge) and fragmentation (triazole that leads to a neutral loss of N_2)

DOI: 10.1021/jacs.7b04131
J. Am. Chem. Soc. 2017, 139, 10359−10364

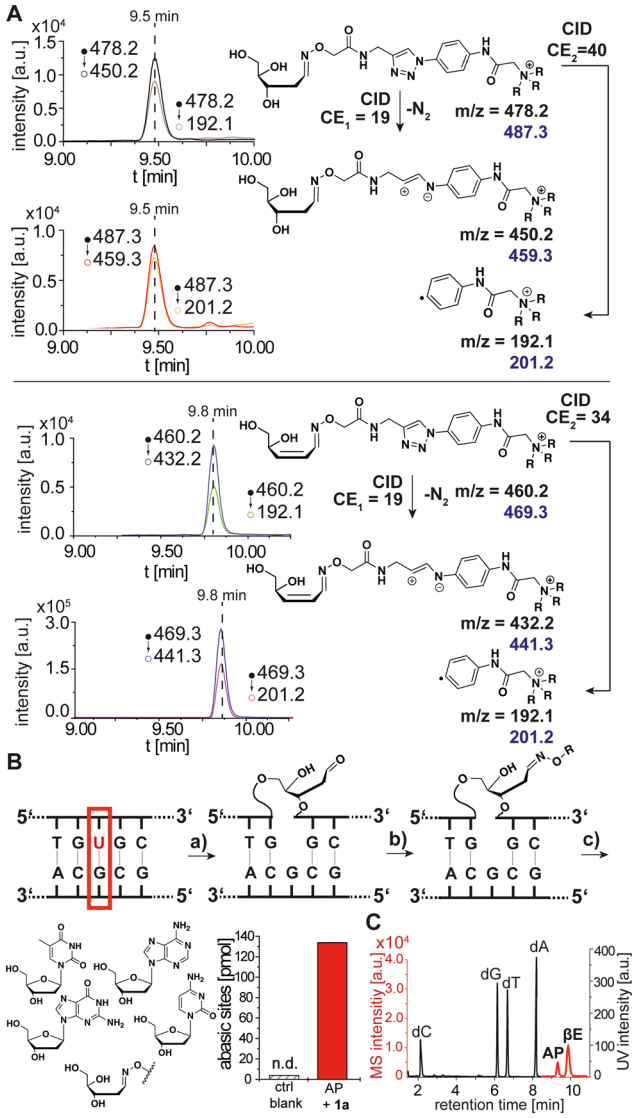


Figure 2. (A) Fragmentation pattern of **9a**/**b** and **10a**/**b**, which give defined daughter ions after loss of N_2 or an aryl radical. (B) Reaction of reagent **1** with a defined AP-site created inside a DNA duplex. Quantification of the reaction product **9a** after the treatment of the oligodeoxynucleotide with **1a**. (C) UHPLC-chromatogram of the canonical bases plus the AP (**9a**)- and βE-site (**10a**) reaction products formed with reagent **1a**.

properties that allow for ultrasensitive detection in a triple quadrupole mass spectrometer.

To investigate if the reagent is able to detect AP-sites directly in DNA, we prepared oligodeoxynucleotides with a single dU base (Figure 2B). To this DNA strand we added the U-cleaving monofunctional glycosylase UDG (uracil DNA glycosylase) to generate a defined AP-site (Supporting Information, section 3.6). We added **1a**, digested the DNA with a mixture of the enzymes nuclease S1, Antarctic phosphatase, and snake venom phosphodiesterase, added a defined amount of the internal standard **9b**, and quantified generated AP-sites using our developed method (Figure 2C; Supporting Information, section 3.8). Indeed, we clearly detected the AP-site adduct **9a** at the expected level, showing that **1a** reacts quantitatively under mild conditions and that the reaction product is excised quantitatively. Furthermore, we treated an 8oxo-dG containing oligodeoxynucleotide with human 8-oxoguanine DNA glycosylase 1 (OGG1) and observed the formation of both an AP- and a βE-site (Supporting Information, section 3.7).

Quantification of Abasic- and βE-Sites in Mouse Embryonic Stem Cells. To assess the utility of the method (derivatization plus mass spectrometry), we next measured the steady-state levels of AP- and βE-sites in mouse embryonic stem cells (mESC). Here, the removal of fdC and cadC by BER may lead to a particular increase of the levels of BER intermediates.[22−24] Alternatively, it was shown that Tdg forms a complex with the ten-eleven translocation dioxygenase (Tet) 1 and Ape1 and possibly also with the nei endonuclease VIII-like (Neil) 1 and 2 enzymes. In these scenarios, BER takes place in a tight complex that does not allow accumulation of AP-sites.[25,26] Naïve cultures of mESCs were primed for 5 days.[27] To measure the steady-state level of AP- and βE-sites, we treated the isolated genomic DNA before digestion with **1a**. To the digestion mixture were added the internal standards **9b** and **10b** in defined amounts, and the mixture was analyzed and quantified by UHPLC-MS.

It should be noted that this method does not allow the distinction between AP- and 5′-dRP-sites. As shown in Figure 3A, we measured a total of 8.8×10^{-7} AP- and 1.7×10^{-6} βE-

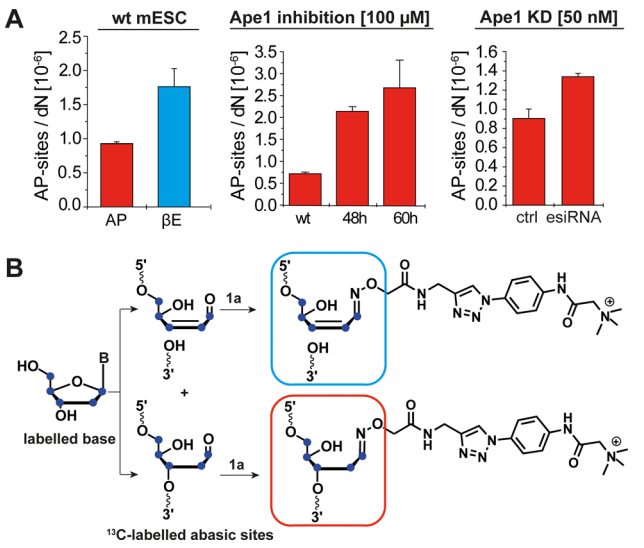


Figure 3. Quantification of BER intermediates and isotope tracing studies. (A) Quantitative data of global levels of AP- and βE-sites from mESCs. Inhibition and knockdown of Ape1 yields an increased amount of AP-sites. (B) Feeding of mESC cultures with labeled nucleosides results in the formation of ribose-labeled AP- and βE-site products **9*** and **10***, which are five mass units heavier than unlabeled products.

sites per dN, which is comparable to the levels in other cell types (Figure S4). These quantitative data show that in mESCs the steady-state levels of total AP- and βE-sites are quite substantial. Our data are slightly lower as compared to data obtained previously.[17,19,20] While AP-sites also form by spontaneous depurination, this is not possible for βE-sites, showing that DNA glycosylases generate quite a number of harmful intermediates that need to be further processed. We measure a steady-state level of around ~9200 βE-sites per genome in mESCs. Several control experiments show that these intermediates are not generated during DNA isolation, handling, or derivatization (Supporting Information, sections 2.5 and 3.9). For example, to exclude that the measured AP-

DOI: 10.1021/jacs.7b04131
J. Am. Chem. Soc. 2017, 139, 10359−10364

and βE-intermediates are formed during the treatment of genomic DNA with **1a**, we incubated genomic DNA over different time points with **1a** and quantified the levels of AP- and βE-sites (Figure S9). Increasing levels of AP- and βE-sites were not detected even after a prolonged reaction time (up to 60 min), excluding this possibility. To exclude that the AP-sites are generated during the DNA isolation procedure, we studied how the major AP-endonuclease Ape1[28] influences the levels of AP-sites (Figure 3A).

This enzyme is known to operate together with monofunctional glycosylases to convert AP-sites into single nucleotide gaps. It is known that Ape1 and Tdg interact and that the release of Tdg from the AP-site depends on the presence of Ape1.[29] We therefore added the Ape1 inhibitor CRT0044867[30] to mESC cultures and measured the levels of AP-sites after treatment for 48 and 60 h. As shown in Figure 3A, we see a clear increase of AP-sites by a factor of 3−4, confirming that AP-sites accumulated when the activity of Ape1 is inhibited. On the basis of similar signal intensities, we concluded that the amount of βE-sites stayed unchanged in this experiment. We also down-regulated Ape1 expression using esiRNA in the last 2 days of priming. qPCR studies confirmed that the esiRNA approach reduces the expression level of Ape1 by 50%. This led to a 30% increased AP-site level (Figure 3A, right panel). These data show that our method is able to report the differential formation of BER intermediates in vivo.

Tracing Studies with Isotopically Labeled Nucleosides. We next analyzed the formation of AP- and βE-sites at the different types of DNA bases with a particular emphasis on dC and its derivatives mdC, hmdC, fdC, and cadC. To this end, we prepared four different mESC cultures and added either isotopically labeled dA*, dC*, dG*, or dT* nucleosides, in which all of the ^{12}C atoms were exchanged against ^{13}C. Correspondingly, the AP- and βE-sites formed from these bases are five atom units heavier than those obtained from unlabeled material (Figure 3B).

We first determined the steady-state levels of mdC, hmdC, fdC, and cadC. For this we isolated and digested the genomic DNA using a standard protocol (Supporting Information, sections 2.3 and 3.2). The obtained levels are as expected with fdC = 1.5 × 10^{-6}/dN and cadC = 8.9 × 10^{-7}/dN (Figure 4). Next, we monitored the efficiency of incorporation. We saw for

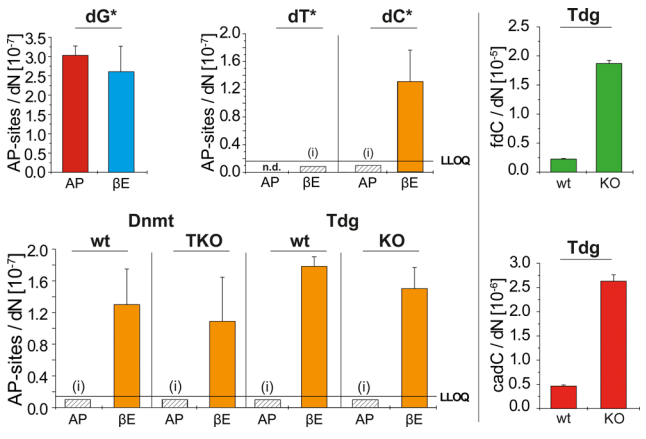


Figure 4. Quantification of labeled AP- and βE-intermediates of the isotope tracing studies and quantification of the fdC and cadC content in Tdg wt and KO cells. n.d., not detectable. (i) Levels were over the LOD, but under LLOQ.

$^{13}C_{10}$-dG an incorporation of 93%. $^{13}C_9$-dC was incorporated to 40%, and the $^{13}C_9$-dT almost fully replaced dT (97% incorporation). Interestingly, no significant incorporation of dA was observed, making any study of this base impossible. We next performed exact quantification of the AP- and βE-sites at dG*, dT*, and dC* and normalized the obtained values for direct comparison. For BER at dG, which is known to be prone to oxidative damage and spontaneous depurination,[31] we saw comparable amounts of AP- and βE-sites. We measured 2.7 × 10^{-7} sites per dN, which corresponds to a steady-state level of around ∼1500 AP- and βE-sites derived from dG (Figure 4). The high level of βE-sites is consistent with the known BER repair pathway of the main oxidative lesion 8-oxoG that involves the action of the bifunctional DNA glycosylase Ogg1.[32]

At dT, the levels of the βE-sites are below 10^{-8} per dN, which amounts to less than ∼100 βE-intermediates per genome. This level is below the lower limit of quantification (LLOQ) of our reagent and mass spectrometry-based method (Figure 4). AP-sites deriving from dT were not detectable at all. The data show that spontaneous cleavage of the glycosidic bond of dT and BER at this base is a minor process.

Most interesting are the very low levels of AP-sites detected at dC* (below LLOQ). βE-sites are in contrast clearly measurable, and with 1.2 × 10^{-7} per dN they are about one-half of those measured for dG*. dC and its derivative mdC are prone to both spontaneous deamination[33,34] and oxidation.[35] It has been recently reported that deamination preponderantly contributes to mutational load in tissues. Under cell culture conditions, however, oxidation seems to be the major source of DNA damage.[36] All known DNA glycosylases that process mismatches generated by deamination of dC-bases are monofunctional (uracil-N glycosylase (Ung) 2, single-strand specific monofunctional uracil DNA glycosylase (Smug) 1, Tdg, and methyl-binding domain glycosylase (Mbd) 4). Oxidized dC-bases are in contrast repaired by bifunctional Neil family DNA glycosylases.[37] The occurrence of oxidative damage in cultured cells is therefore consistent with higher levels of βE-products relative to AP-sites stemming from dC-type bases.

To study how mdC and its oxidized derivatives hmdC, fdC, and cadC contribute to the level of βE-sites, we repeated the dC* feeding study with mESCs lacking all three DNA methyltransferases (Dnmts). Indeed, these cells lack any genomic mdC, and, consequently, the oxidized bases hmdC, fdC, and cadC are not detected anymore. Despite this, we saw no AP- and βE-site changes, arguing that at the steady state the overwhelming majority of βE-sites are generated from canonical dC, for example, by oxidative damage. Obviously, if BER is low, the small amounts of fdC and cadC that are present in the wild type may provide only small amounts of AP- and βE-sites that are negligible. If BER is however a major process, the fact that we do not see changing AP- and βE-site levels could alternatively imply that the formed AP-sites are immediately processed.

To study this phenomenon in more detail, we next fed dC* to mESCs lacking the Tdg enzyme to interrogate the putative BER complex. We first measured the levels of fdC and cadC and saw the expected increase of fdC and cadC levels (Figure 4). Despite this, we again see unchanged levels of AP- and βE-sites. This inability to detect any significant change despite higher levels of fdC (2 × 10^{-5} per dN) and cadC (2.5 × 10^{-6} per dN) is puzzling. It may suggest that single nucleotide BER is no longer occurring or that fdC or cadC is removed by alternative pathways such as long patch BER or noncanonical

DOI: 10.1021/jacs.7b04131
J. Am. Chem. Soc. 2017, 139, 10359−10364

mismatch repair.[38,39] Interestingly, in an experiment where we used the Ape1 inhibitor CRT0044867 and fed dC*, analysis of the labeled AP-sites did not show an increase as well. The measured levels stayed below the LLOQ (Figure S1A). We did detect, however, a small increase of the global AP-sites (Figure S1B). Because Ape1 is supposed to be in complex with Tdg, to quickly convert AP-sites, this lack of increase is surprising.

Our data are consistent with either a scenario in which the Tdg generated BER intermediates at fdC and cadC are quickly turned over without Ape1 or in which other processes remove fdC and cadC.

CONCLUSION

In summary, we report the development of reagent **1** that reacts quickly with aldehydes such as those formed during BER in the genome. The reagent **1** undergoes an efficient CID derived neutral molecule loss in the MS, which allows the ultrasensitive detection of AP- and βE-BER intermediates with 100 amol sensitivity. This, together with the feeding of isotopically labeled nucleosides, allowed us to show that removal of the epigenetically relevant bases fdC and cadC proceeds without the accumulation of significant amounts of these potentially harmful BER intermediates in the genome of stem cells. Alternatively, it may be that the BER-based removal of fdC and cadC is less important than thought. Further studies are now needed to fully decipher the BER processes that operate on fdC and cadC during the enigmatic process of active epigenetic demethylation.[40,41]

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.7b04131.

Chemical synthesis, cell culture and transfection procedures, quantification of abasic sites, and additional references (PDF)

AUTHOR INFORMATION

Corresponding Author

*thomas.carell@cup.uni-muenchen.de

ORCID

Markus Müller: 0000-0002-3579-3317

Thomas Carell: 0000-0001-7898-2831

Author Contributions

†R.R., O.K., and A.K. contributed equally.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The Tdg KO mESC line was kindly provided by Primo Schär (University of Basel, Switzerland). We are also grateful to Masaki Okano (Kumamoto University, Japan) for providing the Dnmt TKO mESC line. We thank the Deutsche Forschungsgemeinschaft for financial support via SFB1032, SFB749, SPP1784, CA275, and the Excellence Cluster CiPSM.

REFERENCES

(1) Lindahl, T. *Angew. Chem., Int. Ed.* **2016**, *55*, 8528−8534.
(2) Fromme, J. C.; Banerjee, A.; Verdine, G. L. *Curr. Opin. Struct. Biol.* **2004**, *14*, 43−49.
(3) Serre, L.; Pereira de Jésus, K.; Boiteux, S.; Zelwer, C.; Castaing, B. *EMBO J.* **2002**, *21*, 2854−2865.
(4) Schneider, S.; Schorr, S.; Carell, T. *Curr. Opin. Struct. Biol.* **2009**, *19*, 87−95.
(5) Stivers, J. T.; Jiang, Y. L. *Chem. Rev.* **2003**, *103*, 2729−2760.
(6) Drohat, A. C.; Maiti, A. *Org. Biomol. Chem.* **2014**, *12*, 8367−8378.
(7) Drohat, A. C.; Coey, C. T. *Chem. Rev.* **2016**, *116*, 12711−12729.
(8) Maiti, A.; Drohat, A. C. *J. Biol. Chem.* **2011**, *286*, 35334−35338.
(9) He, Y. F.; Li, B. Z.; Li, Z.; Liu, P.; Wang, Y.; Tang, Q.; Ding, J.; Jia, Y.; Chen, Z.; Li, L.; Sun, Y.; Li, X.; Dai, Q.; Song, C. X.; Zhang, K.; He, C.; Xu, G. L. *Science* **2011**, *333*, 1303−1307.
(10) Pfaffeneder, T.; Hackner, B.; Truss, M.; Münzel, M.; Müller, M.; Deiml, C. A.; Hagemeier, C.; Carell, T. *Angew. Chem., Int. Ed.* **2011**, *50*, 7008−7012.
(11) Ito, S.; Shen, L.; Dai, Q.; Wu, S. C.; Collins, L. B.; Swenberg, J. A.; He, C.; Zhang, Y. *Science* **2011**, *333*, 1300−1303.
(12) Globisch, D.; Münzel, M.; Müller, M.; Michalakis, S.; Wagner, M.; Koch, S.; Brückl, T.; Biel, M.; Carell, T. *PLoS One* **2010**, *5*, e15367.
(13) Münzel, M.; Globisch, D.; Carell, T. *Angew. Chem., Int. Ed.* **2011**, *50*, 6460−6468.
(14) Wagner, M.; Steinbacher, J.; Kraus, T. F. J.; Michalakis, S.; Hackner, B.; Pfaffeneder, T.; Perera, A.; Müller, M.; Giese, A.; Kretzschmar, H. A.; Carell, T. *Angew. Chem., Int. Ed.* **2015**, *54*, 12511−12514.
(15) Carell, T. LMU Munich, Patent Application EP17153895, 2017.
(16) Tretyakova, N.; Villalta, P. W.; Kotapati, S. *Chem. Rev.* **2013**, *113*, 2395−2436.
(17) Kubo, K.; Ide, H.; Wallace, S. S.; Kow, Y. W. *Biochemistry* **1992**, *31*, 3703−8.
(18) Ide, H.; Akamatsu, K.; Kimura, Y.; Michiue, K.; Makino, K.; Asaeda, A.; Takamori, Y.; Kubo, K. *Biochemistry* **1993**, *32*, 8276−83.
(19) Wei, S.; Shalhout, S.; Ahn, Y. H.; Bhagwat, A. S. *DNA Repair* **2015**, *27*, 9−18.
(20) Roberts, K. P.; Sobrino, J. A.; Payton, J.; Mason, L. B.; Turesky, R. J. *Chem. Res. Toxicol.* **2006**, *19*, 300−309.
(21) Kojima, N.; Takebayashi, T.; Mikami, A.; Ohtsuka, E.; Komatsu, Y. *J. Am. Chem. Soc.* **2009**, *131*, 13208−13209.
(22) Maiti, A.; Drohat, A. C. *J. Biol. Chem.* **2011**, *286*, 35334−35338.
(23) Cortazar, D.; Kunz, C.; Selfridge, J.; Lettieri, T.; Saito, Y.; MacDougall, E.; Wirz, A.; Schuermann, D.; Jacobs, A. L.; Siegrist, F.; Steinacher, R.; Jiricny, J.; Bird, A.; Schär, P. *Nature* **2011**, *470*, 419−423.
(24) Cortellino, S.; Xu, J.; Sannai, M.; Moore, R.; Caretti, E.; Cigliano, A.; Le Coz, M.; Devarajan, K.; Wessels, A.; Soprano, D.; Abramowitz, L. K.; Bartolomei, M. S.; Rambow, F.; Bassi, M. R.; Bruno, T.; Fanciulli, M.; Renner, C.; Klein-Szanto, A. J.; Matsumoto, Y.; Kobi, D.; Davidson, I.; Alberti, C.; Larue, L.; Bellacosa, A. *Cell* **2011**, *146*, 67−79.
(25) Weber, A. R.; Krawczyk, C.; Robertson, A. B.; Kusnierczyk, A.; Vagbo, C. B.; Schuermann, D.; Klungland, A.; Schär, P. *Nat. Commun.* **2016**, *7*, 10806.
(26) Schomacher, L.; Han, D.; Musheev, M. U.; Arab, K.; Kienhofer, S.; von Seggern, A.; Niehrs, C. *Nat. Struct. Mol. Biol.* **2016**, *23*, 116−24.
(27) Nichols, J.; Smith, A. *Cell Stem Cell* **2009**, *4*, 487−492.
(28) Marenstein, D. R.; Wilson Iii, D. M.; Teebor, G. W. *DNA Repair* **2004**, *3*, 527−533.
(29) Fitzgerald, M. E.; Drohat, A. C. *J. Biol. Chem.* **2008**, *283*, 32680−90.
(30) Madhusudan, S.; Smart, F.; Shrimpton, P.; Parsons, J. L.; Gardiner, L.; Houlbrook, S.; Talbot, D. C.; Hammonds, T.; Freemont, P. A.; Sternberg, M. J.; Dianov, G. L.; Hickson, I. D. *Nucleic Acids Res.* **2005**, *33*, 4711−4724.
(31) Lindahl, T. *Nature* **1993**, *362*, 709−15.
(32) Hill, J. W.; Hazra, T. K.; Izumi, T.; Mitra, S. *Nucleic Acids Res.* **2001**, *29*, 430−438.
(33) Duncan, B. K.; Miller, J. H. *Nature* **1980**, *287*, 560−561.
(34) Shen, J. C.; Rideout, W. M., 3rd; Jones, P. A. *Nucleic Acids Res.* **1994**, *22*, 972−976.

DOI: 10.1021/jacs.7b04131
J. Am. Chem. Soc. 2017, 139, 10359−10364

(35) Madugundu, G. S.; Cadet, J.; Wagner, J. R. *Nucleic Acids Res.* **2014**, *42*, 7450−7460.
(36) Rouhani, F. J.; Nik-Zainal, S.; Wuster, A.; Li, Y.; Conte, N.; Koike-Yusa, H.; Kumasaka, N.; Vallier, L.; Yusa, K.; Bradley, A. *PLoS Genet.* **2016**, *12*, e1005932.
(37) Jacobs, A. L.; Schär, P. *Chromosoma* **2012**, *121*, 1−20.
(38) Bochtler, M.; Kolano, A.; Xu, G. L. *BioEssays* **2017**, *39*, 1−13.
(39) Santos, F.; Peat, J.; Burgess, H.; Rada, C.; Reik, W.; Dean, W. *Epigenet. Chromatin* **2013**, *6*, 39.
(40) Ooi, S. K. T.; Bestor, T. H. *Cell* **2008**, *133*, 1145−1148.
(41) Kohli, R. M.; Zhang, Y. *Nature* **2013**, *502*, 472−9.

DOI: 10.1021/jacs.7b04131
J. Am. Chem. Soc. 2017, 139, 10359−10364

3.3 Quantitative LC-MS Analyse liefert keinen Hinweis auf m6dA oder m4dC in Genom von embryonalen Stammzellen oder Gewebe der Maus.

S. Schiffers, C. Ebert, R. Rahimoff, **O. Kosmatchev**, J. Steinbacher, A.-V. Bohne, F. Spada, J. Nickelsen, S. Michalakis, M. Müller, T. Carell, *Angew. Chem. Int. Ed.* **2017**, *56*, 11268-11271. *Quantitative LC-MS Provides No Evidence for m6dA or m4dC in the Genome of Mouse Embryonic Stem Cells and Tissues*

Zusammenfassung

Neben den im Abschnitt 1 beschriebenen epigenetisch relevanten DNA-Modifikationen wurden in den letzten Jahren noch weitere DNA-Nukleoside entdeckt. Es wurde berichtet, dass das N^6-Metyldesoxyandenosin (m^6dA) in Genom verschiedener eukaryotischer Organismen entdeckt wurde. Zusammen mit N^4-Methyldesoxycytidin (m^4dC) wurde das m^6dA außerdem in bakteriellem Genom gefunden. In diesem Manuskript wurde eine neue hochsensitive UHPLC-MS/MS Methode für die Detektion und Quantifizierung von m^6dA und m^4dC entwickelt. Außerdem konnten mit dieser Methode mdC, hmdC, fdC und cadC in DNA embryonaler Stammzellen, im Gehirn oder in der Leber der Maus quantifiziert werden. Es konnte aber weder m^6dA noch m^4dC in dieser DNA detektiert werden. Somit wird die epigenetische Relevanz dieser Modifikationen in Frage gestellt.

Autorenbeitrag

Die Hauptarbeit bei diesem Projekt lag bei *Sarah Schiffers*. Für dieses Projekt führte ich die Quantifizierung von mdC, hmdC, fdC, cadC und 8-Oxo-dG mit der in AK *Carell* etablierten massenspektrometrischen Methode durch.

Lizenz

Epigenetics

International Edition: DOI: 10.1002/anie.201700424
German Edition: DOI: 10.1002/ange.201700424

Quantitative LC–MS Provides No Evidence for m^6dA or m^4dC 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 (m^5dC) 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. N^6-methyldeoxyadenosine (m^6dA) is the most recent base reported to be present in the genome of various eukaryotic organisms. This base, together with N^4-methyldeoxycytidine (m^4dC), 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 m^5dC, hmdC, fdC, and cadC, but we were unable to detect either m^4dC or m^6dA in DNA isolated from mouse embryonic stem cells or brain and liver tissue, which calls into question their epigenetic relevance.*

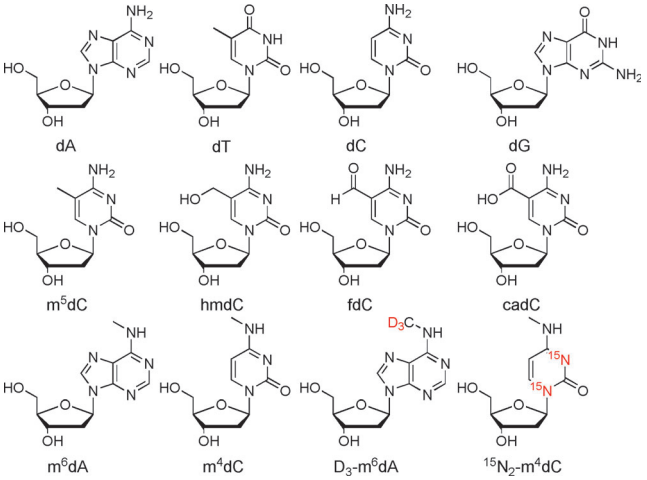


Figure 1. Depiction of the four canonical DNA bases and the epigenetic DNA marks m^5dC, hmdC, fdC, and cadC, as well as the bases m^6dA and m^4dC together with the synthesized isotopologues.

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 (m^5dC) 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 m^5dC 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 N^4-methyldeoxycytidine (m^4dC)[10] and N^6-methyldeoxyadenosine (m^6dA),[11] attempts to detect these bases in the DNA of higher organisms have failed until recently.[12–15] m^6dA has now been found in algae (0.4 mol% m^6dA/A),[12] fruit flies (0.001%-0.07% m^6dA/A),[14] and *C. elegans* (0.01%-0.4% m^6dA/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 m^4dC and m^6dA to be searched for and quantified in parallel to the more established new epigenetic DNA marks m^5dC, 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
CiPS[M], 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 m^6dA and m^4dC shown in Figure 1 as internal standards for the analytical method. The prepared compounds D_3-m^6dA and $^{15}N_2$-m^4dC 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 D_3-m^6dA and $^{15}N_2$-m^4dC 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 m^6dA and $m/z = 269.14 \rightarrow 153.10$ for its isotopologue D_3-m^6dA. For m^4dC, 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 m^4dC and $m/z = 244.11 \rightarrow 128.07$ for its isotopologue $^{15}N_2$-m^4dC (Figure 2 A).

We next modified the reported UHPLC-QQQ method[10] for the simultaneous quantification of m^4dC and m^6dA, together with the other epigenetically relevant bases m^5dC, 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 m^4dC and m^6dA are well known in bacteria, we first analysed the cyanobacterium *Synechocystis* (PCC6803), and we indeed found both bases (Figure 2 B). The base m^6dA was detected at a level of 8.4×10^{-3} per dN and for m^4dC we measured a value of 5.9×10^{-3} per dN. The constitutional isomer m^5dC 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 m^6dA has just recently been discovered,[12] and the levels of m^6dA 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 m^6dA 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, m^4dC was not detected, thus showing that this base is unlikely to be a component of the genetic material of *Chlamydomonas*. This is interesting because *Synechocystis* is considered a relative of the chloroplasts present in *Chlamydomonas*.

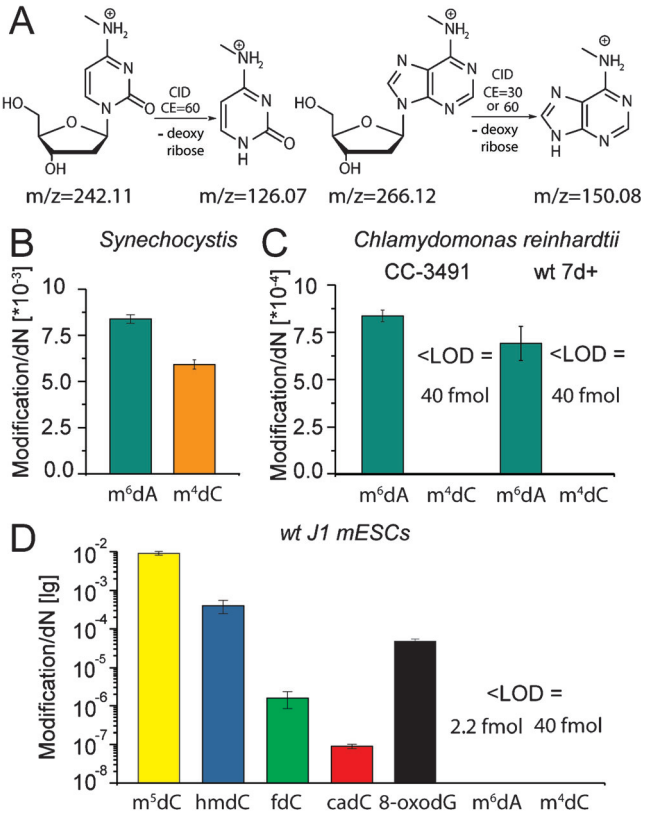


Figure 2. A) Fragmentation patterns of m^4dC and m^6dA. B–D) Quantitative data of the bases m^4dC and m^6dA in *Synechocystis* (B), *Chlamydomonas* (C) and of these bases and the other epigenetic DNA marks hmdC, fdC, cadC, and m^5dC in mouse embryonic stem cells (D).

With these positive results in hand, we extended our study to mouse embryonic stem cells (ESCs; wt J1, Figure 2 D). m^6dA 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 m^6dA 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 m^6dA, in contrast, a signal did not appear. We also re-measured wt TT2 cells as described and still did not detect m^6dA over background levels (see Figure S4).[17] To obtain unequivocal proof that m^6dA is not present in stem cells, we added $^{13}CD_3$-methionine to the mESC culture. Methionine provides the methyl group for the biosynthesis of m^6dA. With $^{13}CD_3$-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 $^{13}CD_3$-m^6dA (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

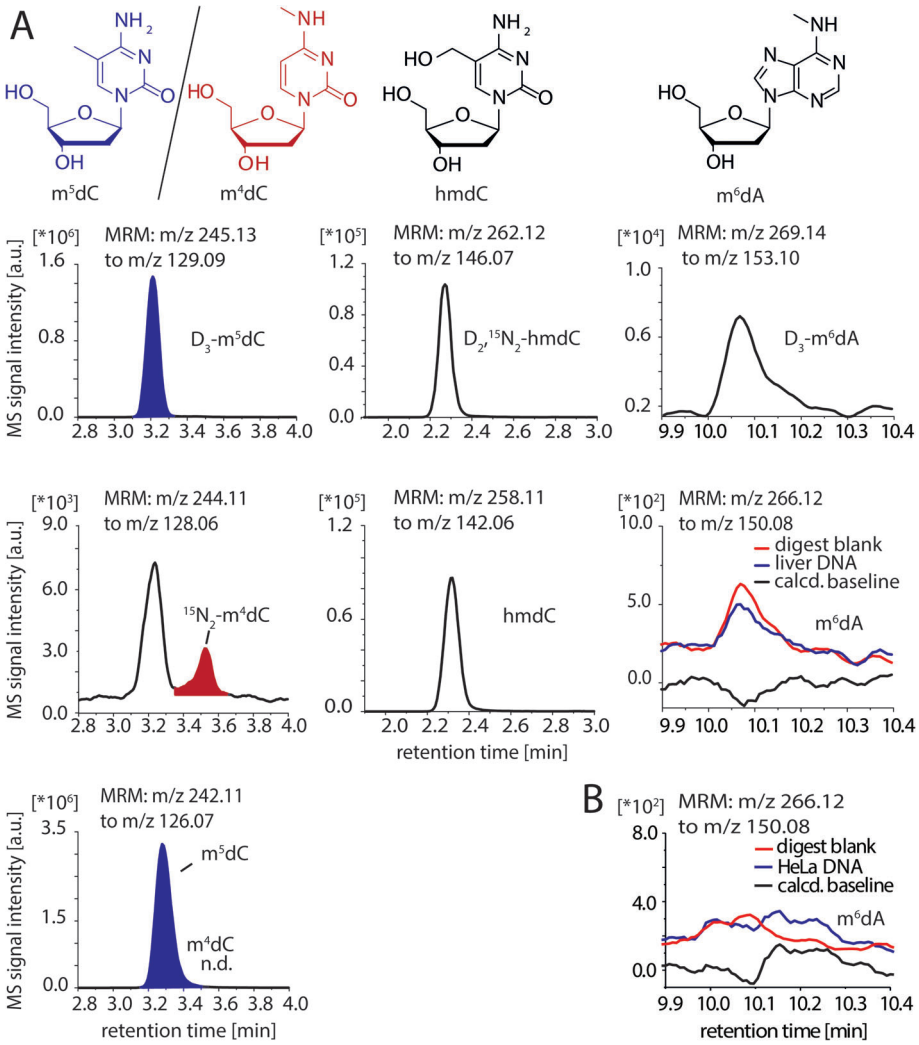


Figure 3. A) Chromatograms of the mass signal of mouse liver DNA. UHPLC-QQQ data obtained for m^5dC and m^4dC (left), hmdC (middle), and m^6dA (right, blue line) and their corresponding isotopic standards are shown. Additionally, for m^6dA, the chromatogram for the digest blank is shown (red line) and 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^6dA 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^5dC and m^4dC (left column), the epigenetic DNA mark m^5dC is clearly detected at a retention time of 3.2 min, but for m^4dC with a retention time of 3.5 min, there is obviously no signal present.

The m^6dA data are highly interesting (Figure 3 right column). While the D_3-m^6dA standard was clearly detectable at a retention time of 10.1 min, the unlabelled m^6dA 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^6dA 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^6dA. To support the evidence that m^6dA is not present in vertebrate DNA, we performed the same experiment with HeLa cells and also observed no signal for m^6dA (Figure 3B).

We then determined our limit of detection for m^6dA to be 3.5×10^{-7} per dN, which corresponds to 170 m^6dA 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^6dA that could be theoretically present and would not be detectable by our method is less than 170 m^6dA bases per genome, which led us to conclude that m^6dA 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^6dA in mESC DNA that could explain previous sequencing data,[17] we thought that m^6dA-containing bacterial DNA that gets degraded could provide the m^6dA nucleoside, which then might get incorporated into mESC DNA. This is indeed a possibility. When we added the m^6dA nucleoside to a mESCs culture, we indeed saw incorporation of some m^6dA into the genome (Figure S7).

We also cannot fully exclude the possibility that the presence of few m^6dA 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 m^6dA levels at specific time points that may have escaped our detection.[18] Our data, however, show clearly that the maximum possible levels of m^6dA in the analysed organisms and mESCs under normal conditions are far lower than so far believed.

Acknowledgements

We thank the Deutsche Forschungsgemeinschaft for financial support via SFB1032 (TP-A5), SFB749 (TP-A4), SPP1784, CA275 and the Excellence Cluster CiPS^M.

Conflict of interest

The authors declare no conflict of interest.

Keywords: DNA · epigenetics · methyldeoxyadenosine · methyldeoxycytidine · mass spectrometry

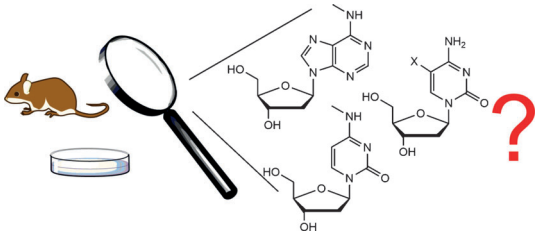
[1] J. A. Law, S. E. Jacobsen, *Nat. Rev. Genet.* **2010**, *11*, 204 – 220.
[2] M. Tahiliani, K. P. Koh, Y. Shen, W. A. Pastor, H. Bandukwala, Y. Brudno, S. Agarwal, L. M. Iyer, D. R. Liu, L. Aravind, A. Rao, *Science* **2009**, *324*, 930 – 935.
[3] S. Kriaucionis, N. Heintz, *Science* **2009**, *324*, 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, Y. 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.

Manuscript received: January 13, 2017
Revised: February 15, 2017
Final Article published: ■■ ■■, ■■■■

Communications

Epigenetics

S. Schiffers, C. Ebert, R. Rahimoff, O. Kosmatchev, J. Steinbacher, A.-V. Bohne, F. Spada, S. Michalakis, J. Nickelsen, M. Müller, T. Carell* ——— ▮▮▮▮–▮▮▮▮



Quantitative LC–MS Provides No Evidence for m^6dA or m^4dC in the Genome of Mouse Embryonic Stem Cells and Tissues

Previous results challenged: Highly sensitive mass spectrometry reveals that m^6dA is most likely not an epigenetic base in the mouse genome. It appears that in vertebrates, the levels of this modified base, and that of N^4-methyldeoxycytidine, are much lower than previously thought.

3.4 Tet-Enzyme oxidieren Thymin zu 5-Hydroxymethyluracil in murinen embryonalen Stammzellen

T. Pfaffeneder#, F. Spada#, M. Wagner#, C. Brandmayr, S. Laube, D. Eisen, M. Truss, J. Steinbacher, B. Hackner, O. Kotljarova, D. Schuermann, S. Michalakis, **O. Kosmatchev**, S. Schiesser, B. Steigenberger, N. Raddaoui, U. Müller, H. Leonhardt, P. Schär, M. Müller, T. Carell, *Nat. Chem. Biol.* **2014,** *10, 574-581. Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA.*

Zusammenfassung

Wie im Anschnitt 1.2.2 bereits erwähnt könnte ein möglicher Schlüsselschritt der aktiven DNA-Demethylierung die Desaminierung von hmdC zu hmdU beinhalten. In diesem Manuskript wurde im Rahmen einer massenspektrometrischen Isotopenverfolgungsstudie gezeigt, dass das hmdU über eine Tet-vermittelte Oxidation von dT und nicht über eine Desaminierung von hmdC im Genom der embryonalen Stammzellen der Maus entsteht.

Autorenbeitrag

Die Hauptarbeit bei diesem Projekt lag bei *Toni Pfaffeneder*, *Fabio Spada* und *Mirko Wagner*. Für dieses Projekt führte ich die enzymatische Hydrolyse, sowie die massenspektrometrische Analyse der genomischen DNA aus murinen embryonalen Stammzellen durch.

Lizenz

nature chemical biology

ARTICLE

PUBLISHED ONLINE: 18 MAY 2014 | DOI: 10.1038/NCHEMBIO.1532

Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA

Toni Pfaffeneder[1,8], Fabio Spada[1,8], Mirko Wagner[1,8], Caterina Brandmayr[1], Silvia K Laube[1], David Eisen[1], Matthias Truss[2], Jessica Steinbacher[1], Benjamin Hackner[1], Olga Kotljarova[1], David Schuermann[3], Stylianos Michalakis[4], Olesea Kosmatchev[1], Stefan Schiesser[1], Barbara Steigenberger[1], Nada Raddaoui[1], Gengo Kashiwazaki[1], Udo Müller[5], Cornelia G Spruijt[6], Michiel Vermeulen[6,7], Heinrich Leonhardt[5], Primo Schär[3], Markus Müller[1]* & Thomas Carell[1]*

Ten eleven translocation (Tet) enzymes oxidize the epigenetically important DNA base 5-methylcytosine (mC) stepwise to 5-hydroxymethylcytosine (hmC), 5-formylcytosine and 5-carboxycytosine. It is currently unknown whether Tet-induced oxidation is limited to cytosine-derived nucleobases or whether other nucleobases are oxidized as well. We synthesized isotopologs of all major oxidized pyrimidine and purine bases and performed quantitative MS to show that Tet-induced oxidation is not limited to mC but that thymine is also a substrate that gives 5-hydroxymethyluracil (hmU) in mouse embryonic stem cells (mESCs). Using MS-based isotope tracing, we show that deamination of hmC does not contribute to the steady-state levels of hmU in mESCs. Protein pull-down experiments in combination with peptide tracing identifies hmU as a base that influences binding of chromatin remodeling proteins and transcription factors, suggesting that hmU has a specific function in stem cells besides triggering DNA repair.

Methylcytosine is an epigenetically important nucleobase associated with the control of transcriptional activity, genomic imprinting, X-chromosome inactivation and suppression of transposable elements[1]. Controlled formation and removal of mC at specific genomic loci is critical for correct genome programming or reprogramming during cellular differentiation[2]. Recently, it was discovered that Tet proteins (Tet1–3) oxidize mC to give the oxidized C-derived nucleobases hmC[3], 5-formylcytosine (fC)[4,5] and 5-carboxycytosine (caC)[5,6], whose biological functions are still yet unclear (**Fig. 1a**)[7]. As fC and caC are both removed by thymine DNA glycosylase (Tdg)[6,8], it is currently assumed that they serve as intermediates of an active DNA demethylation process involving base excision repair. In addition to these oxidized C derivatives, cells also contain oxidized T nucleobases such as hmU and fU. These compounds are currently known as oxidative lesions that are thought to form upon the reaction of T with reactive oxygen species (ROS)[9,10]. It was recently suggested that hmU might also be produced by deamination of hmC, a hypothesis that remains controversial[11–14]. Deamination of hmC, situated in a base pair with G (hmC:G), would give rise to hmU:G mismatches, which are known substrates for the DNA glycosylases Tdg, Smug1, Mbd4, Ung2 (ref. 15), Neil1 and Nthl1 (ref. 16). Deamination of hmC:G to hmU:G followed by mismatch repair would therefore establish an alternative pathway to active demethylation (**Fig. 1a**).

To unravel the origin of oxidized nucleobases, and of hmU in particular, in DNA from mESCs, we performed isotope tracing and quantitative MS studies using the chemically synthesized isotopologs of mC, hmC, fC, caC, hmU and fU as internal standards (**Fig. 1b** and **Supplementary Results**, **Supplementary Fig. 1**). For the assessment of oxidation products that are formed by the action of ROS, we additionally quantified 8-oxo-G because 8-oxo-G is a well-established ROS reaction product formed from G[17,18]. We show here that hmU is generated enzymatically from thymidine during stem cell differentiation by the action of the Tet enzymes. A proteomic analysis provides new insight into how genomic hmU can influence the binding of chromatin remodeling proteins and transcription factors.

RESULTS

hmU is present at elevated levels in mESCs

We first created an inventory of the named nucleosides (**Supplementary Fig. 1**) in mESCs (**Fig. 1c**) and adult cortex tissue (**Fig. 1d**). In mESCs, we observed that hmC is, as expected, the most abundant oxidized pyrimidine (10% relative to mC), followed by fC (~1–2% of hmC). The ROS marker 8-oxo-G was detected at similar levels (~45% of fC), showing that nonenzymatic, ROS-induced oxidations of nucleobases are important processes, as expected. We also found relatively high levels of fU (22% of fC). Clearly detectable were also caC and hmU. hmU, which is at the center of this study, was unequivocally detected, as shown by its retention time and its specific fragmentation pattern, which were found to be identical with that of the internal standard [D_2]hmU (**Fig. 1e**). Notably, both caC and hmU were present in comparable amounts (5% of fC). In adult mouse cortex DNA (**Fig. 1d**; for other tissues see

[1]Center for Integrated Protein Science at the Department of Chemistry, Ludwig-Maximilians-Universität München, München, Germany. [2]Charité Universitätsklinikum, Otto-Heubner-Centrum für Kinder und Jugendmedizin, Klinik für Allgemeine Pädiatrie, Labor für Pädiatrische Molekularbiologie, Berlin, Germany. [3]Department of Biomedicine, University of Basel, Basel, Switzerland. [4]Center for Integrated Protein Science at the Department of Pharmacy—Center for Drug Research, Ludwig-Maximilians-Universität München, München, Germany. [5]Center for Integrated Protein Science at the Department of Biology, Ludwig-Maximilians-Universität München, Planegg-Martinsried, Germany. [6]Department of Molecular Cancer Research, Cancer Genomics Netherlands, Utrecht, The Netherlands. [7]Present address: Department of Molecular Biology, Faculty of Science, Radboud Institute for Molecular Life Sciences, Radboud University Nijmegen, Nijmegen, The Netherlands. [8]These authors contributed equally to this work. *e-mail: markus.mueller@cup.uni-muenchen.de or thomas.carell@cup.uni-muenchen.de

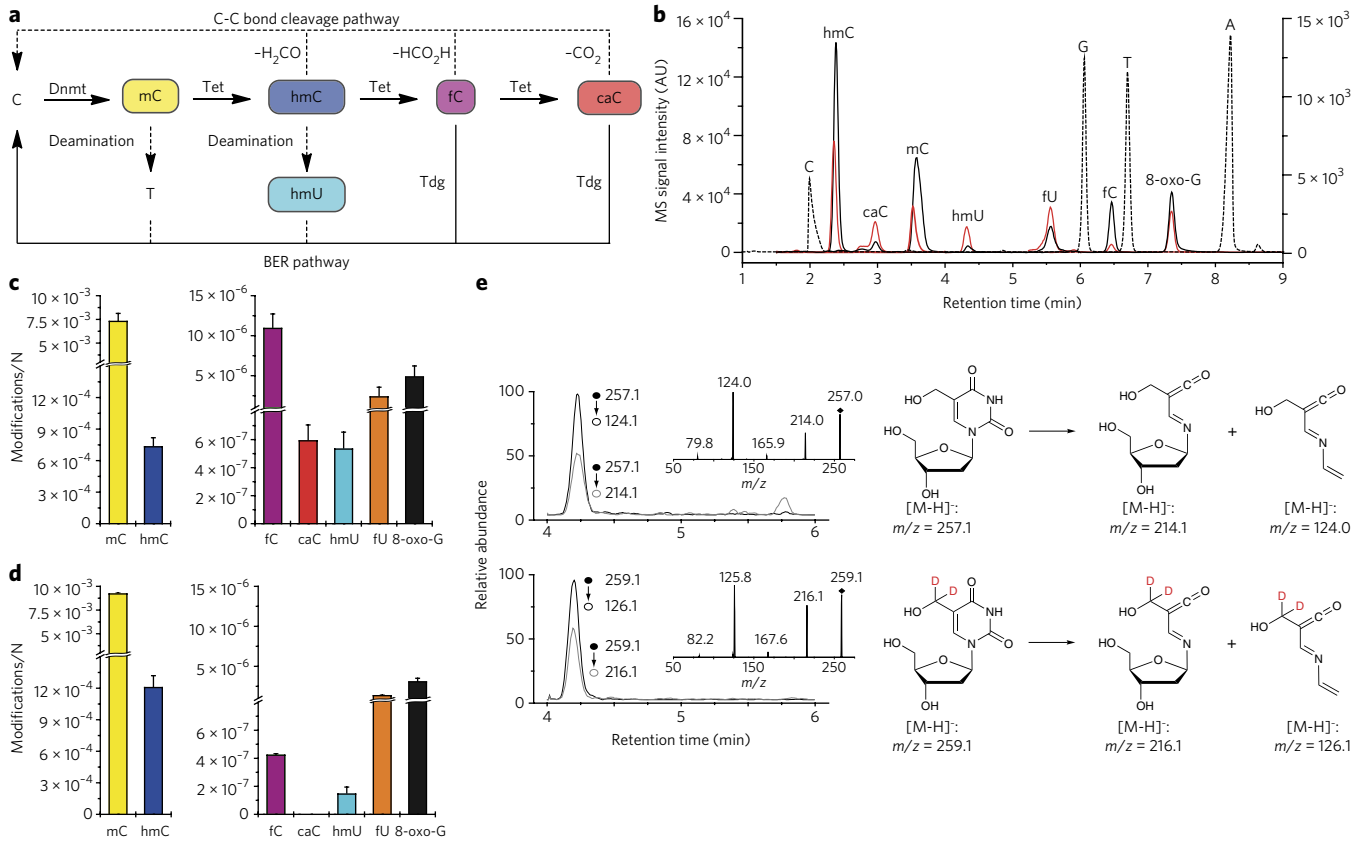


Figure 1 | Metabolism of cytosine derivatives, their detection by LC-UV-ESI-MS/MS and levels in mESCs and mouse cortex. (**a**) Potential active demethylation pathways. (**b**) Overlaid LC/UV and LC/MS/MS chromatograms of a representative DNA sample from mESCs. The dotted LC/UV chromatogram of C, G, T and A is scaled arbitrarily; the overlaid LC/MS/MS chromatograms of hmC, mC, fC and 8-oxo-G are scaled to the left *y* axis; the LC/MS/MS chromatograms of caC, hmU and fU are scaled to the right *y* axis. Red chromatograms refer to the corresponding labeled internal standards depicted in **Supplementary Figure 1**. AU, arbitrary units. (**c**,**d**) DNA modification levels per nucleoside (N) in mESCs (**c**; WT01, $n = 7$) and 3-month-old mouse cortex tissue (**d**; $n = 3$). Depicted are mean values ± s.d. (**e**) Representative LC/MS/MS chromatograms for identification and quantification of hmU. Shown are the overlaid chromatograms for the two characteristic fragment ions of hmU (top trace) and the [D_2]hmU internal standard (bottom trace) derived from a mESC DNA sample. The insets show the MS/MS full-scan spectra of synthetic hmU and [D_2]hmU matching the proposed fragmentation pathway.

Supplementary Fig. 2), the hmC levels are very high (13% relative to mC), as previously reported[19–21]. In contrast, fC was detected only in very small amounts (0.03% relative to hmC), and caC was not observed at all[22]. The detected amounts of 8-oxo-G and fU are comparable. Notably, the detected levels of hmU in the cortex were significantly lower compared to that in mESCs (27%; $P = 2.2 \times 10^{-4}$ by Student's *t*-test). Because the detected levels of 8-oxo-G were similar in the cortex (**Fig. 1d**) and in mESCs (**Fig. 1c**), the elevated hmU levels in mESCs cannot be explained by greater ROS-induced damage in mESCs and must have a different origin.

We next investigated this in more detail and quantified the levels of hmU in three different mESC lines (WT01, J1 and R1) and in a variety of tissues (**Fig. 2a**). Clearly, the hmU levels were higher in mESCs than in somatic tissues by factors of 2–15. The detected hmU levels correspond roughly to 500–1,700 hmU bases per genome in tissue and 2,900–7,800 in mESCs (**Fig. 2a**) depending on the cell type and growth conditions. We cannot explain these differences by elevated oxidative stress levels in mESCs (**Supplementary Fig. 3**). We next analyzed the levels of C, mC and the oxidized pyrimidines hmC, fC, hmU and fU as well as 8-oxo-G in somatic tissues (**Supplementary Fig. 2**) and performed a correlation and cluster analysis to reveal relationships of the modification levels (**Fig. 2b**). The data analysis confirmed that the low levels of hmU in somatic tissues correlate strongly (significant on a level <0.001) with the levels of fU and 8-oxo-G within a separated cluster. This result showed that in somatic cells, hmU and fU are ROS-induced reaction products. If we assume that in somatic tissues hmU is exclusively formed by ROS, the data showed that in mESCs, 70–80% of the detected hmU is produced by ROS-independent processes (**Supplementary Fig. 3**).

hmU formation by oxidation of T

To analyze the origin of the oxidized nucleobases in mESCs, we performed isotope tracing experiments (**Fig. 3a–c** and **Supplementary Fig. 4**). Substitution of L-methionine with [*methyl*-$^{13}CD_3$]L-methionine ([$^{13}CD_3$]L-Met) in the growth medium is known to give the labeled *S*-adenosylmethionine cofactor, which is needed for the conversion of C to mC. Replacement of L-Met by [$^{13}CD_3$]L-Met for 5 d (2 passages) in the medium at a concentration of 0.2 mM furnished 89% labeled [$^{13}CD_3$]mC, 88% labeled [*hydroxymethyl*-$^{13}CD_2$]hmC and 93% labeled [*formyl*-^{13}CD]fC (**Fig. 3b** and **Supplementary Fig. 4**). Within the detection limit (7 and 50 molecules per 10^8 nucleosides, respectively), no incorporation of the isotopes ^{13}C and D into hmU and fU was observed, showing that hmC is not the precursor of hmU. We next added isotope-labeled [^{13}C,$^{15}N_2$]thymidine ([^{13}C,$^{15}N_2$]T) to the growth medium and observed ~76% of label incorporation into T, hmU and fU (**Fig. 3a**, **Supplementary Fig. 4** and **Supplementary Table 1**). The combined data showed that hmU is not generated by deamination of hmC but by oxidation of T (T → hmU; **Fig. 3c**). Consequently, all of the detected hmU

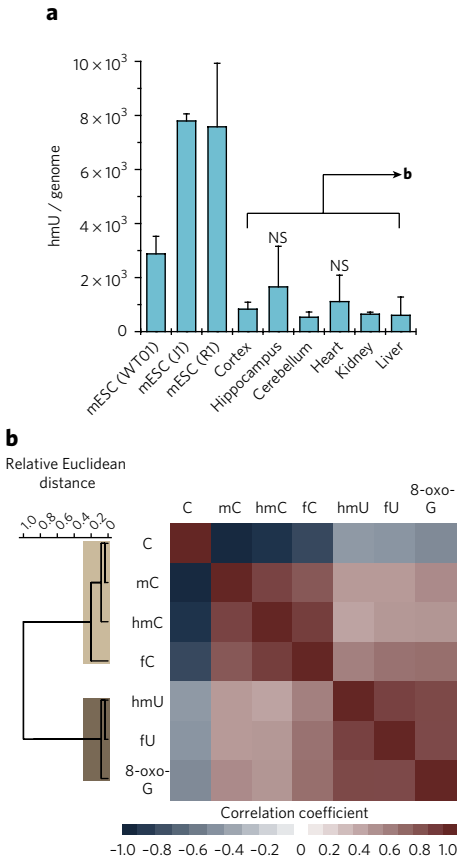


Figure 2 | hmU is present at elevated levels in mESCs compared to tissue. (**a**) hmU levels per genome in mESCs (n_{WT01} = 7 replicates, n_{J1} = 2, n_{R1} = 3) and mouse tissue (3-month-old individuals, n = 3). Levels per genome were obtained considering a mouse genome size of 2.7 × 10^9 base pairs. Depicted are mean values ± s.d. The differences between mESCs and mouse tissues are significant (P = 2.9 × 10^{-5} to 3.6 × 10^{-2}; unpaired two-tailed t-test) except for WT01 and hippocampus (P = 0.292) or heart (P = 0.069). These exceptions are due to higher hmU levels in the hippocampus and heart caused by higher background oxidation (higher 8-oxo-G levels). hmU levels normalized to oxidative background (8-oxo-G) levels are in **Supplementary Figure 3**. NS, not significant. (**b**) Unsupervised clustering analysis of Pearson correlation coefficients of 24 data sets of selected mouse organs at a defined time point (3-month-old individuals: cortex, hippocampus, cerebellum, heart, liver and kidney). Discussed correlations are strong to very strong (Pearson coefficient >0.7) and significant on a level <0.001.

(and also fU) resides in an A base pair context (hmU:A). In this base pair, hmU is repaired by Smug1 but not Tdg[12]. This was confirmed by siRNA-mediated knockdown of Smug1 in mESC cells and HEK-293T cells. Indeed, the hmU level increased in these cells (**Supplementary Fig. 5**).

The fact that we were unable to detect hmU derived from hmC deamination, which would be situated in a base pair with G (hmU:G), could also be explained by very fast repair. If repair of the hmU:G base pair is extremely efficient, for example, because deamination and glycosylase-based repair occurs in a tight complex of the involved enzymes, we would be unable to detect this type of hmU because of low steady state levels. Indeed, it was proposed that hmC deamination and hmU excision requires a complex of the cytidine deaminase Aid and the glycosylase Tdg[12]. To assess this possibility, we performed isotope tracing experiments using [$^{13}CD_3$]L-Met in *Tdg*$^{-/-}$ mESCs stably complemented with either empty vector (control) or a minigene expressing a catalytically incompetent Tdg at near-endogenous levels. These cells are able to form the Aid–Tdg complex, but the Tdg is inactive, which should give elevated hmU levels if deamination occurs. In both cell lines, we detected high levels of labeled hmC. In the control cells, [*hydroxymethyl*-$^{13}CD_2$] hmU was not detected. However, in the cell line complemented with inactive Tdg, some labeled [*hydroxymethyl*-$^{13}CD_2$]hmU was indeed observed, albeit only in small amounts (~7% of total hmU and ~0.06% of total hmC; **Supplementary Fig. 6**). Thus, the Tdg protein is required for deamination of hmC to hmU. In wild-type (WT) mESCs, the hmU:G mismatches are obviously repaired so quickly that they do not contribute to steady state levels of hmU. All of the detected hmU was derived from T oxidation and resided in hmU:A base pairs.

Tet enzymes form hmU in correlation with mC oxidation

We next investigated whether enzymatic oxidation or ROS-dependent processes are responsible for the observed oxidation of T to hmU. To this end, we studied mESCs with genetic or functional depletions of Tet or DNA methyltransferase (Dnmt) enzymes (**Fig. 4**). Although hmU levels were maintained at normal levels in the severely hypomethylated, Dnmt-depleted cell lines, they were substantially reduced upon knockdown of Tet1 and Tet2. As the levels of fU and 8-oxo-G remained unaffected, the data establish that, though both hmU and fU are generated by T oxidation, their formation must occur by two independent processes in mESCs. The data supported the idea that fU is a ROS-created lesion similar to 8-oxo-G[9,10,17,18], whereas most hmU is generated by Tet-induced oxidation of T.

In support of enzymatic T oxidation by Tet proteins, ectopic expression of the catalytic domain of Tet1 (Tet1cd) in HEK-293T cells led to a 65-fold increase for both hmC and hmU levels. This was not observed when a catalytic mutant of Tet1 (Tet1cm) was expressed (**Supplementary Fig. 7a**). To determine whether under these conditions hmU is generated by deamination, as previously suggested[11,12], we again replaced natural L-Met with [$^{13}CD_3$]L-Met in the medium, but we did not detect incorporation of heavy isotopes into hmU, providing evidence that the elevated hmU levels do not originate from hmC deamination (**Supplementary Table 1**). Finally, we confirmed enzymatic generation of hmU *in vitro* by incubating recombinant Tet1cd with a plasmid that was premethylated by bacterial methyltransferase M.SssI. In addition to oxidation of mC to hmC, fC and caC, we detected hmU at a prominent level of 9% relative to hmC (**Supplementary Fig. 7b**), showing that the catalytic center of the Tet enzymes clearly has the capacity to oxidize T to hmU.

To further confirm that hmU is formed in mESCs in the process of epigenetic reprogramming, we analyzed the dynamic changes of mC and all of the oxidized pyrimidines plus 8-oxo-G during differentiation. It was recently shown that mC and hmC levels sharply increase when mESCs maintained in the naive state are shifted to a primed state in serum-containing medium[23–25]. To investigate global kinetics of all of the Tet-generated oxidation products under more physiologically relevant priming conditions, we used established protocols based on serum-free N2B27 medium for differentiation of naive mESCs into states resembling that of post-implantation epiblasts[26,27]. mESCs were first grown for several passages in the presence of MEK and GSK3 inhibitors (dual inhibition or 2i conditions) and LIF to induce a hypomethylated state resembling that of the naive epiblast[23–25]. The data in **Figure 5a** are averaged from three independent differentiation experiments, each performed with two cell lines in the absence of growth factors. First, we observed that the fU and 8-oxo-G levels stayed constant (**Fig. 5a**), in line with the idea that both are ROS-derived products. In contrast, fC and caC levels peaked at about 8 h. Both hmU and hmC also peak between 8 h and 16 h. Isotope tracing with [$^{13}CD_3$]L-Met under these conditions provided no evidence for switched-on deamination of hmC to hmU (**Supplementary Table 1**). At these peak levels, we estimated that mESCs contain roughly 110,000 fC bases, 4,400 caC bases and

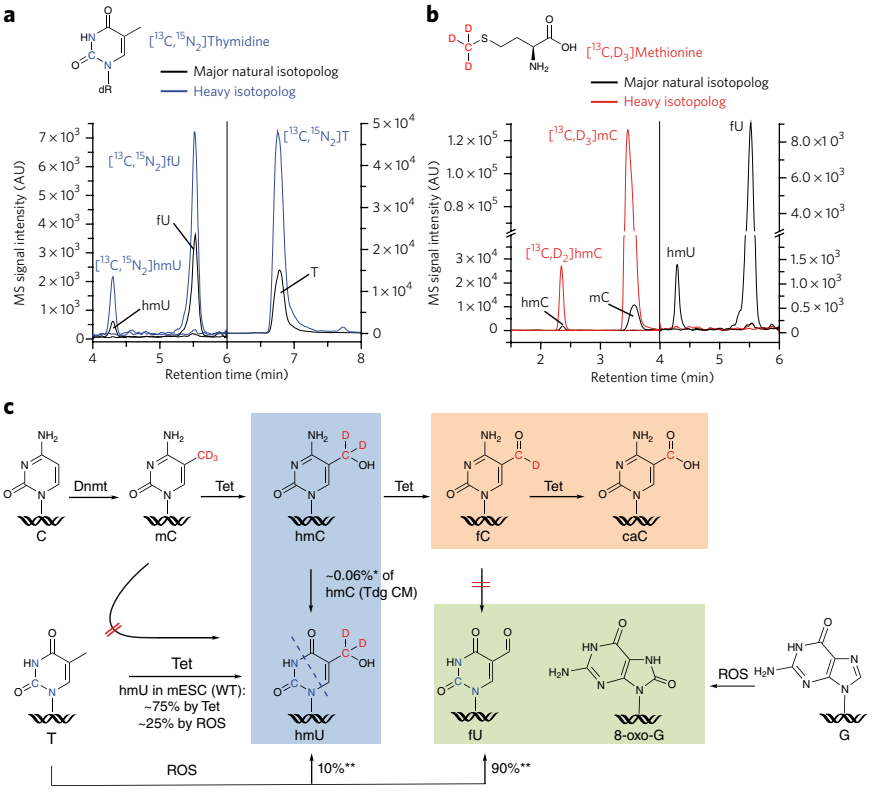


Figure 3 | hmU and fU are thymine oxidation products in WT mESCs with no detectable contribution from hmC or fC deamination. (**a**) Overlaid LC/MS/MS chromatograms of heavy ([^{13}C,^{15}N$_2$]hmU, [^{13}C,^{15}N$_2$]fU and [^{13}C,^{15}N$_2$]T; blue) and major natural (black) isotopologs of hmU, fU and T. AU, arbitrary units. (**b**) Overlaid chromatograms of heavy ([^{13}C,D$_3$]mC, [^{13}C,D$_2$]hmC, [^{13}C,D$_2$]hmU and [^{13}C,D]fU; red) and the major natural (black) isotopologs of mC, hmC, hmU and fU. (**c**) Enzymatic and ROS-dependent pathways leading to the formation of hmC, fC, caC, hmU and fU. Single asterisks denote labeled hmU generated by deamination of labeled hmC, which was observed only in Tdg catalytic mutant (CM) cells (**Supplementary Fig. 6**), representing ~7% of the total hmU content and corresponding to deamination of ~0.06% hmC. Double asterisks denote basal rates of ROS-dependent T oxidation, which were determined in HEK-293T cells, where Tet activity is lowest (3.6 hmC per 10^5 nucleosides; **Supplementary Table 2**). Here, hmU and fU roughly represent 10% and 90% of T oxidation products relative to the sum of each other.

14,000 hmU bases per genome, showing that, at its peak level, hmU is three times more abundant than caC. The dynamic peaking data allowed us to estimate half-life times for fC (7 h), caC (5 h) and hmU (4 h) during the differentiation process (**Supplementary Fig. 8**). Using the dynamic quantitative data, we performed a correlation and cluster analysis of the DNA modification levels (**Fig. 5b**). Three independent clusters were obtained. One cluster involves members of cytosine methylation (and demethylation) dynamics, including C, mC, fC and caC, but, to our surprise, not hmC. Instead, hmC groups with hmU. We therefore concluded that the formation of these modifications is tightly coupled by the action of the Tet enzymes. The ROS-induced lesions fU and 8-oxo-G form the third, well-separated cluster. Notably, hmU does not correlate with the ROS lesions.

Recently, it was shown that exposure of naive mESCs to fibroblast growth factor 2 (FGF-2) and activin A (ActA) for 48 h under similar conditions as those described above gives a homogeneous cell population whose transcriptome closely resembles that of the post-implantation epiblast[26]. Under these conditions, mC levels increased more rapidly, approaching somatic levels within a time frame closely reflecting that observed during embryonic development (**Supplementary Fig. 9a**). Levels of hmC rose steadily throughout the 48-h time course. Whereas fC peaked at 36 h, hmU reached its maximum at 24 h (**Supplementary Fig. 9b**). qPCR data showed a robust induction of Dnmt3b that is most likely responsible for the increasing mC levels (**Supplementary Fig. 9a**). A transient peak of Tet1 expression at 24 h went in hand with rising hmU and fC levels, whereas Tet3 expression started rising slowly from the same time point and may drive further accumulation of hmC and fC (**Supplementary Fig. 9b**). Tdg was progressively upregulated together with the onset of methylation activity, whereas Smug1 showed little, if any, change (**Supplementary Fig. 10**). These data supported our view that hmU peaking is not caused by downregulation of the repair pathway.

Thus, the kinetic data showed a complex interplay between methylation and two oxidation reactions during differentiation, which depends on the exact conditions. Most important is the fact that hmU showed a time-dependent occurrence similar to the other oxidized bases hmC, fC and caC independently of the conditions investigated, confirming its formation during epigenetic reprogramming.

hmU attracts specific readers

To obtain initial insight into a potential biological function of hmU in comparison to hmC, we screened for specific readers associating with hmU:A as well as with hmC:G and hmU:G using protein pulldown and relative quantification by LC/MS/MS. In previous stable isotope labeling by amino acids in cell culture (SILAC)-based proteomics studies with hmC, fC and caC containing oligonucleotides, we and others observed a high number of specific protein readers, arguing that the new bases influence a variety of different processes[28,29]. For this hmU study, we further developed this approach for the detection of proteins that directly interact with the modified bases hmU (as well as hmC for comparison) to get a more direct insight into their function (**Fig. 6a**). For the study, we not only included into the biotinylated DNA duplexes (24mers) an hmU (hmC) base but also equipped the counter strand with a polyethylene glycol–based linker carrying a reactive *N*-hydroxysuccinimide (NHS) ester moiety and a reductively cleavable disulfide bond in the middle. This reactive linker cannot be inserted into the DNA strand using solid phase synthesis. In addition, the NHS ester does not survive hybridization conditions. We therefore attached the linker as its azide derivative, using Cu(I)-catalyzed click chemistry, to an alkyne-bearing base present in the DNA duplex[30,31]. These DNA duplexes were subsequently incubated with nuclear extracts from mESCs. The DNA-bound protein complexes were isolated using streptavidin-coated magnetic beads (**Fig. 6a**)[32]. DNA duplexes with the canonical base pairs A:T and C:G at the respective positions served as reference strands. The NHS linker has two functions in the experiment: First, it will covalently trap the reader proteins specifically at the ε-amino groups of lysines, which allows the identification of transiently binding proteins as well. Second, because the linker is cleavable, it will tag the trapped lysine residues, leaving a defined label on those proteins that bind in close proximity to the modified bases hmC and hmU. This allowed us to distinguish protein readers that bound close to hmC and hmU from proteins that are secondary members of the complexes. The covalently trapped proteins were next tryptically digested and labeled with tandem mass tagging (TMT) isobaric tags to allow protein

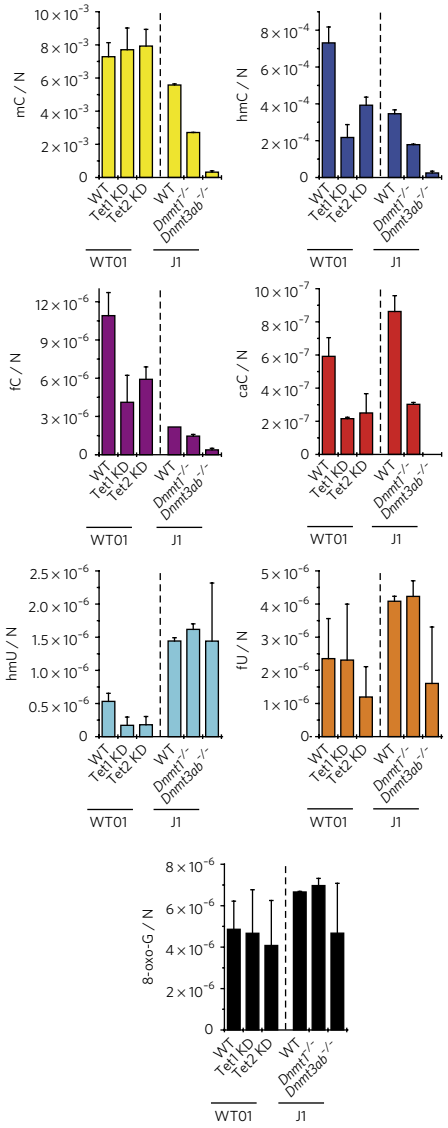


Figure 4 | Tet1 and Tet2 generate hmU in mESCs. Effect of Tet and Dnmt depletion on DNA modification levels in mESCs. J1 cell lines bearing homozygous *Dnmt*-null mutations and WT01 mESCs expressing Tet1 or Tet2 shRNAs (knockdown (KD)) were analyzed. Depicted are mean values per nucleoside (N) ± s.d. of biological replicates as follows: wild type (WT01, $n = 7$), Tet1 KD (WT01, $n = 3$) and Tet2 KD (WT01, $n = 3$); WT (J1, $n = 2$), *Dnmt1*$^{-/-}$ (J1, $n = 2$) and *Dnmt3a/b*$^{-/-}$ (J1, $n = 6$).

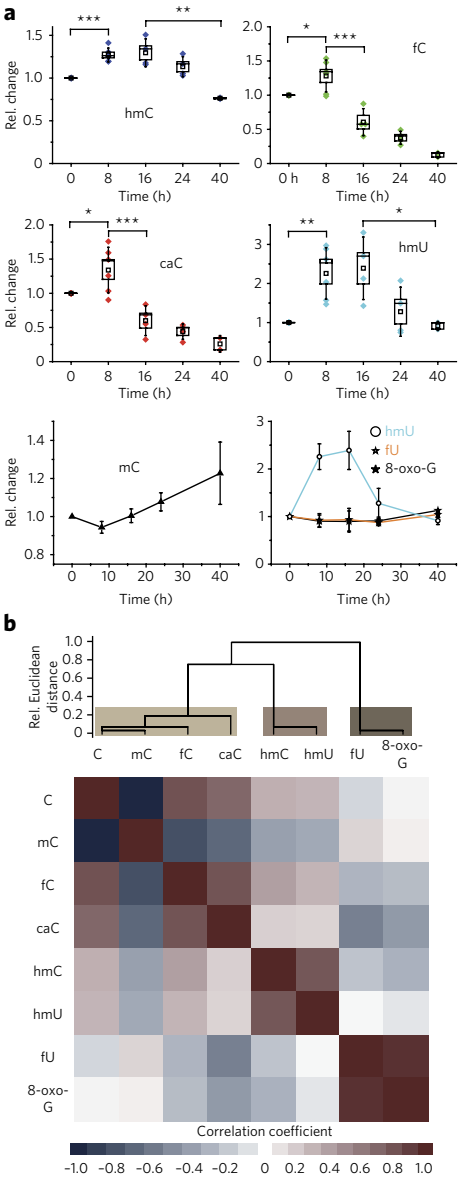


Figure 5 | hmU is produced during mESC differentiation. (**a**) Time course of DNA modification levels during early differentiation (0–40 h) of naive mESCs in the absence of growth factors. Box plot of relative modification changes for hmC, fC, caC and hmU averaged from six independent experiments using two different mESC lines. In the four upper plots, colored diamonds reflect mean values of technical triplicate measurements, open squares reflect mean values of biological replicates, boxes represent the s.e.m., and whiskers represent the the s.d. of the biological replicates; *$P < 0.05$, **$P < 0.01$ and ***$P < 0.005$ by unpaired two-tailed *t*-test. In the two lower plots, the relative modification changes of mC, hmU, fU and 8-oxo-G are depicted as biological mean values ± s.e.m. (**b**) Unsupervised clustering analysis of Pearson correlation coefficients of 22 data sets obtained from the differentiation experiments depicted in **a**. Discussed correlations are strong to very strong (Pearson coefficient >0.7) and significant on a level <0.001.

identification and quantification by MS[32,33]. Only proteins that were enriched in both the forward and the reverse TMT experiment were considered to be specific hmU (hmC) readers (**Fig. 6**). All of the readers were subsequently divided into two groups. Proteins that were enriched relative to the control strands are termed specific readers (sRs). Of those, the proteins that were identified with a peptide containing the tag are termed direct-specific readers (dsRs).

The scatter plot (**Fig. 6b** and **Supplementary Fig. 11a**) revealed over 250 sRs recognizing the hmU:A base pair (relative to T:A). Of those proteins, 99 were identified with tagged peptides. These proteins were therefore designated dsRs. A gene ontology analysis showed that 64 proteins of the 250 sRs are nucleotide binders, and 25 proteins are involved in chromatin organization (**Supplementary Fig. 11b**).

Among the dsRs we identified were the regulatory proteins Uhrf2, transcription factor HIVEP3 or the poly [ADP-ribose] polymerase 14 (Parp14), which showed that these proteins interacted directly with hmU. We also detected several H2A and H2B histone proteins among the dsRs and the sRs. Notably, in the group of the sRs, we observed also Dnmt3a and Dnmt3b, which are involved in gene expression regulation. Methyl CpG-binding protein 2 (MeCP2) was identified as well, but it showed only low enrichment. We also detected several transcriptional regulators, such as bromodomain PHD finger

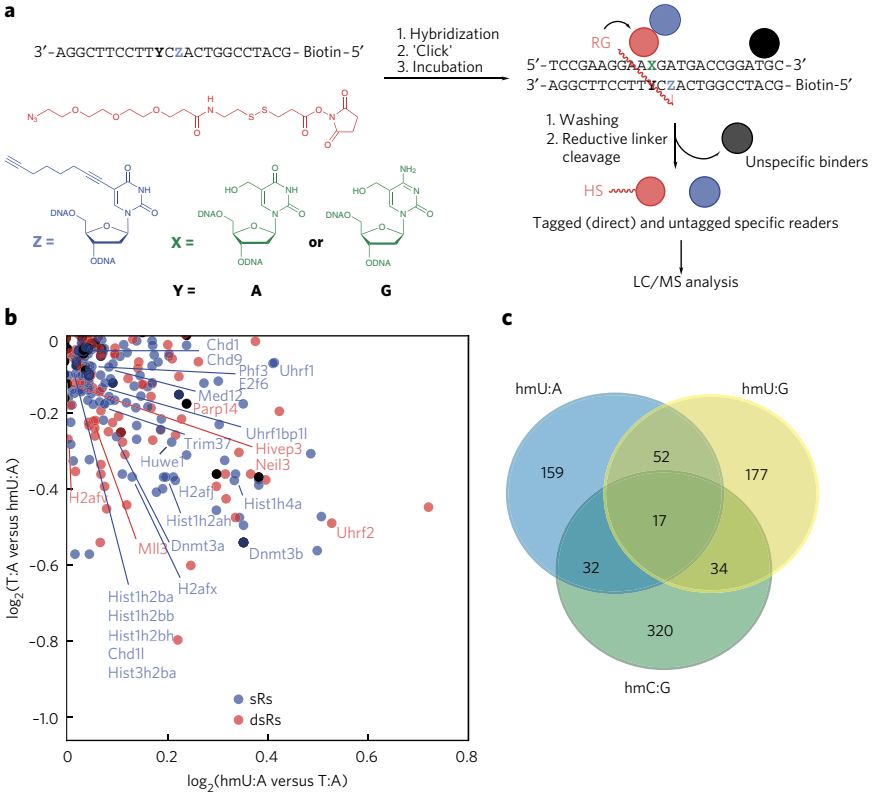


Figure 6 | Identification of hmU:A readers. (**a**) Workflow of pulldown experiments with hmU:A-containing oligonucleotides. DNA oligomers with the modified base Z containing an alkyne group for click reaction and biotin for the pulldown are hybridized with DNA oligomers containing hmU (X = hmU and Y = A). After the click reaction with the DNA-protein cross-linker (red), which has an NHS ester as a lysine-specific reactive group (RG), the dsDNA oligomers are incubated with a nuclear extract from mESCs, and specific binders are covalently linked to the DNA. After three washing steps, the linker is reductively cleaved, and the proteins are enzymatically digested and labeled with TMT2plex reagents. The same dsDNA oligomer without hmU modification (X = T and Y = A) was used as a control sample. Specifically enriched proteins are mixed with the control sample before LC/MS analysis. The same workflow was used for experiments with hmU:G or hmC:G and C:G as a control. (**b**) Magnified view of proteins enriched with hmU:A-containing oligomer. Specific readers in the forward and reverse experiment are marked in blue. Direct-specific readers are identified by the presence of the DNA-protein cross-linker and marked in red. Black dots are considered unspecific binders. A full-scale representation is shown in **Supplementary Figure 11**. (**c**) Venn diagram showing overlap of specific binders for hmU:A, hmU:G and hmC:G.

transcription factor (Bptf), transcription factor E2F6 and mediator of RNA polymerase II transcription subunit 12 (Med12). Moreover, we identified hmU-binding E3 ubiquitin-protein ligases such as Uhrf1, Trim37 and Huwe1 and finally the chromodomain helicase DNA-binding proteins 1 and 9 (Chd1 and Chd9) among the hmU readers, which are chromatin remodeling factors and regulate polymerase I and II transcription. To exclude bias by sequence-specific binding, we repeated the experiment with hmU in a different sequence context. Indeed, 78% of the identified protein readers are also observed in this control experiment, and included among them were all of the proteins discussed above, with the notable exception of Uhrf1, which was not observed in this experiment (**Supplementary Data Set 1**). The data show that hmU:A recruits specific proteins that are involved in chromatin biochemistry in a broad sense.

As a first validation of the MS-identified readers, we focused on Uhrf1 and Uhrf2. Uhrf1 is a known mC and hmC binder[34], whereas we recently identified the homolog Uhrf2 as a specific binder of hmC[29]. The MS data for hmU obtained in this study showed an enrichment of Uhrf1 and Uhrf2. In addition, Uhrf2 was identified as a direct binder. To validate the MS findings, we overexpressed Uhrf1 or Uhrf2 together with Tet1cd in HEK-293T[29] and measured the levels of mC, hmC, fC, caC and also hmU. Indeed, we observed increased levels of hmU, proving a functional relation between hmU levels and Uhrf1 and Uhrf2 (**Supplementary Fig. 12**). As Uhrf2 is known to interact with Dnmt3a and Dnmt3b[35] and these two proteins are also identified as sRs of hmU, our initial data provide a consistent picture.

Analysis of the specific hmU readers showed that 49 of the identified proteins were also enriched with an hmC-modified strand. Among these proteins were, for instance, bromodomain PHD finger transcription factor (Bptf), MeCP2 and 19 other nucleotide-binding proteins. The obtained list of specifically binding proteins is in good agreement with our recently published data set[29], which further supports the validity of the covalent trapping method. Finally, we compared the results of the hmU:A pulldown with a pulldown of hmU:G, which would be the product of hmC deamination. We found 69 of 280 sRs in both experiments. We identified transcription regulators as sRs as well in the hmU:G pulldown. However, we also found several DNA repair proteins, such as DNA repair protein complementing XP-C cells homolog (XPC), DNA repair protein Rad50 and five helicases, among others (**Supplementary Data Set 1**).

DISCUSSION

This work aimed to analyze the origin of hmU in mESCs. Although hmU as well as fU are both well-characterized oxidation products of thymidine[9,10], hmU in particular was recently discussed as an intermediate in active demethylation by deamination of hmC[11,12]. We used stable isotopologs of the main oxidation products hmC, fC, caC, hmU, fU and 8-oxo-G in combination with quantitative MS to show that the levels of hmU are strongly elevated in stem cells in comparison to somatic tissue. Notably, we did not observe equally elevated levels of fU and 8-oxo-G, both of which are formed by the reaction of DNA with ROS. This analysis suggests that hmU in stem cells is not only formed by ROS. The quantitative data allowed us to estimate that in WT stem cells, about 70–80% of the found hmU is not formed by ROS but is derived from a different process. Isotope tracing studies with labeled [$^{13}CD_3$]L-Met and [^{13}C,$^{15}N_2$] thymidine showed that the detected hmU originates exclusively from T. Previous studies already proposed a connection between Tet activity and hmU levels, but these studies did not address the origin of hmU, which led to the hypothesis that hmU could form by deamination of hmC[11,22]. To investigate the involvement of deamination[11–14], we studied *Tdg*$^{-/-}$ stem cells reconstituted with a catalytically inactive Tdg. In these cells, the putative deamination–repair complex[12] of Tdg and Aid could form, and we expected to see elevated hmU levels. Indeed, in this experiment, we were able to detect labeled hmU, showing that deamination of hmC does occur, but the levels were low. Most notably, special conditions were required to detect this deamination. In WT cells, all of the detected hmU is clearly derived from T oxidation. As such, the detected hmU is situated in a base pair with A (hmU:A). In conclusion, the majority of hmU in mESCs is produced independently of both ROS and deamination.

Using a combination of functional knockdown in mESCs, ectopic expression of Tet1 catalytic domain in HEK-293T cells and *in vitro* studies with recombinant Tet protein, we showed that the Tet enzymes, known to oxidize mC to hmC, are also responsible for oxidation from T to hmU. A kinetic study in which the temporal changes of all of the oxidized pyrimidines were observed under two different differentiation kinetics showed finally that hmU is formed in processes that also generate hmC and fC. Together with the observed peak levels of hmU of about 14,000 bases per genome, this suggests that hmU could have an epigenetic function similar to hmC. It is conceivable that the Tet enzymes introduce a small amount of hmU during the hmC manufacturing process to trigger faster demethylation by induction of DNA repair. It is known that hmC is not a substrate for repair[8,12], whereas hmU is efficiently recognized and repaired[36], for example, by the repair glycosylase Smug1 (refs. 37,38). Sporadic introduction of hmU could therefore allow recruitment of repair factors, for example, for long patch repair, as recently suggested[39]. Alternatively, hmU might influence transcription factor binding, which is known to be a hallmark of epigenetic reprogramming[40]. To study the proteins that closely interact with hmU, a new pulldown experiment was devised in which we inserted a trapping linker into the oligonucleotide. This linker is able to react covalently with proteins that assemble on the DNA duplex close to the hmU base. Because the linker is reductively cleavable, it leaves a tag on the respective lysine residues, which is detected in subsequent MS-based proteomics measurements. Using this technology, we observed that hmU:A recruits transcription factors and proteins that are involved in chromatin biochemistry, including Uhrf1 and Uhrf2. This result is in line with the recent observation that oxidative lesions in DNA can change the binding affinity of the transcription factor CREB when they are present in the respective cognate sequence[41]. A further analysis in HEK-293T revealed that the hmU readers Uhrf1 and Uhrf2 are able to modulate the levels of all of the oxidized mC bases and of hmU. It is interesting to note that the observed level increases of hmC were small, whereas they were substantial for hmU and also for the further oxidized bases fC and caC. As Uhrf1 and Uhrf2 are proteins known to be involved in numerous epigenetic processes, our data supported the hypothesis that Tet-induced oxidation of T to hmU may have an epigenetic function. Even in the case that hmU is formed just as a side product of Tet-induced hmC formation, it is now clear that the oxidation chemistry performed by the Tet enzymes has to go in hand with effective DNA repair.

Received 15 December 2013; accepted 17 April 2014; published online 18 May 2014

METHODS

Methods and any associated references are available in the online version of the paper.

References

1. Smith, Z.D. & Meissner, A. DNA methylation: roles in mammalian development. *Nat. Rev. Genet.* **14**, 204–220 (2013).
2. Franchini, D.M., Schmitz, K.M. & Petersen-Mahrt, S.K. 5-Methylcytosine DNA demethylation: more than losing a methyl group. *Annu. Rev. Genet.* **46**, 419–441 (2012).
3. Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930–935 (2009).
4. Pfaffeneder, T. *et al.* The discovery of 5-formylcytosine in embryonic stem cell DNA. *Angew. Chem. Int. Ed. Engl.* **50**, 7008–7012 (2011).
5. Ito, S. *et al.* Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333**, 1300–1303 (2011).
6. He, Y.F. *et al.* Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333**, 1303–1307 (2011).
7. Tan, L. & Shi, Y.G. Tet family proteins and 5-hydroxymethylcytosine in development and disease. *Development* **139**, 1895–1902 (2012).
8. Maiti, A. & Drohat, A.C. Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. *J. Biol. Chem.* **286**, 35334–35338 (2011).
9. Bjelland, S. *et al.* Oxidation of thymine to 5-formyluracil in DNA: mechanisms of formation, structural implications, and base excision by human cell free extracts. *Biochemistry* **34**, 14758–14764 (1995).
10. Mouret, J.F., Polverelli, M., Sarrazini, F. & Cadet, J. Ionic and radical oxidations of DNA by hydrogen peroxide. *Chem. Biol. Interact.* **77**, 187–201 (1991).
11. Guo, J.U., Su, Y., Zhong, C., Ming, G.L. & Song, H. Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell* **145**, 423–434 (2011).
12. Cortellino, S. *et al.* Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell* **146**, 67–79 (2011).
13. Nabel, C.S. *et al.* AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. *Nat. Chem. Biol.* **8**, 751–758 (2012).
14. Rangam, G., Schmitz, K.M., Cobb, A.J. & Petersen-Mahrt, S.K. AID enzymatic activity is inversely proportional to the size of cytosine C5 orbital cloud. *PLoS ONE* **7**, e43279 (2012).
15. Jacobs, A.L. & Schär, P. DNA glycosylases: in DNA repair and beyond. *Chromosoma* **121**, 1–20 (2012).
16. Zhang, Q.M. *et al.* DNA glycosylase activities for thymine residues oxidized in the methyl group are functions of the hNEIL1 and hNTH1 enzymes in human cells. *DNA Repair (Amst.)* **4**, 71–79 (2005).
17. Burrows, C.J. Surviving an oxygen atmosphere: DNA damage and repair. *ACS Symp. Ser. Am. Chem. Soc.* **2009**, 147–156 (2009).
18. 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).
19. Globisch, D. *et al.* Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS ONE* **5**, e15367 (2010).
20. Kriaucionis, S. & Heintz, N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **324**, 929–930 (2009).
21. Münzel, M. *et al.* Quantification of the sixth DNA base hydroxymethylcytosine in the brain. *Angew. Chem. Int. Ed. Engl.* **49**, 5375–5377 (2010).
22. Liu, S. *et al.* Quantitative assessment of Tet-induced oxidation products of 5-methylcytosine in cellular and tissue DNA. *Nucleic Acids Res.* **41**, 6421–6429 (2013).
23. Ficz, G. *et al.* FGF signaling inhibition in ESCs drives rapid genome-wide demethylation to the epigenetic ground state of pluripotency. *Cell Stem Cell* **13**, 351–359 (2013).
24. Habibi, E. *et al.* Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells. *Cell Stem Cell* **13**, 360–369 (2013).
25. Leitch, H.G. *et al.* Naive pluripotency is associated with global DNA hypomethylation. *Nat. Struct. Mol. Biol.* **20**, 311–316 (2013).
26. Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S. & Saitou, M. Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. *Cell* **146**, 519–532 (2011).
27. Ying, Q.L., Stavridis, M., Griffiths, D., Li, M. & Smith, A. Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. *Nat. Biotechnol.* **21**, 183–186 (2003).
28. Iurlaro, M. *et al.* A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation. *Genome Biol.* **14**, R119 (2013).
29. Spruijt, C.G. *et al.* Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* **152**, 1146–1159 (2013).
30. Burley, G.A. *et al.* Directed DNA metallization. *J. Am. Chem. Soc.* **128**, 1398–1399 (2006).
31. Rostovtsev, V.V., Green, L.G., Fokin, V.V. & Sharpless, K.B. A stepwise huisgen cycloaddition process: copper(I)-catalyzed regioselective "ligation" of azides and terminal alkynes. *Angew. Chem. Int. Ed.* **41**, 2596–2599 (2002).
32. Thompson, A. *et al.* Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal. Chem.* **75**, 1895–1904 (2003); erratum **75**, 4942 (2003); erratum **78**, 4235 (2006).
33. Gygi, S.P. *et al.* Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* **17**, 994–999 (1999).
34. Frauer, C. *et al.* Recognition of 5-hydroxymethylcytosine by the Uhrf1 SRA domain. *PLoS ONE* **6**, e21306 (2011).
35. Pichler, G. *et al.* Cooperative DNA and histone binding by Uhrf2 links the two major repressive epigenetic pathways. *J. Cell. Biochem.* **112**, 2585–2593 (2011).
36. Lewis, H.L., Muhleman, D.R. & Ward, J.F. Serologic assay of DNA base damage. I. 5-Hydroxymethyldeoxyuridine, a radiation product of thymidine. *Radiat. Res.* **75**, 305–316 (1978).
37. Boorstein, R.J. *et al.* Definitive identification of mammalian 5-hydroxymethyluracil DNA N-glycosylase activity as SMUG1. *J. Biol. Chem.* **276**, 41991–41997 (2001).

38. Kavli, B., Otterlei, M., Slupphaug, G. & Krokan, H.E. Uracil in DNA—general mutagen, but normal intermediate in acquired immunity. *DNA Repair (Amst.)* **6**, 505–516 (2007).
39. Santos, F. *et al.* Active demethylation in mouse zygotes involves cytosine deamination and base excision repair. *Epigenetics Chromatin* **6**, 39 (2013).
40. Silva, J. & Smith, A. Capturing pluripotency. *Cell* **132**, 532–536 (2008).
41. Moore, S.P.G., Toomire, K.J. & Strauss, P.R. DNA modifications repaired by base excision repair are epigenetic. *DNA Repair (Amst.)* **12**, 1152–1158 (2013).

Acknowledgments

We thank the Excellence Cluster Center for Integrated Protein Science Munich (CiPS^{M}) and the collaborative research centers SFB749, SFB646 and SFB1032 as well as German Research Foundation (DFG) grant CA275/8-4, the Volkswagen foundation, NGFNplus (01GS0870) and the Netherlands Organization for Scientific Research (NWO-VIDI) for financial support. T.P. and S.S. thank the Fonds der Chemischen Industrie for predoctoral fellowships. C.B. thanks the Boehringer Ingelheim Fonds for a predoctoral fellowship. G.K. thanks the Japan Society for the Promotion of Science (JSPS) for a postdoctoral fellowship for research abroad. We thank M. Moser (Max Planck Institute for Biochemistry) for providing R1- and C57Bl6/129–derived mESCs, G. Höfner and K.T. Wanner for their initial help with MS as well as M. Wirsing, L. Belzner and P. Laube for providing bioinformatic tools for data processing.

Author contributions

T.P. synthesized MS standards, performed the sample preparation and ultra high-performance LC/MS/MS method development, did LC/MS analysis, interpreted data and performed statistical analysis. F.S. and N.R. performed mESC differentiation and isotope tracing experiments. M.W. and C.B. performed the HEK-293T experiments, did LC/MS analysis and interpreted data. C.B. performed qPCR and analyzed the data. S.K.L. and D.E. performed the protein pulldown studies and interpreted data. M.T. performed mESC knockdown and knockout experiments. J.S. and O. Kosmatchev did sample preparation and LC/MS analysis. B.H., S.S. and J.S. prepared MS standards. O. Kotljarova performed *in vitro* assays. B.S. synthesized oligonucleotides for protein capture. G.K. synthesized tandem mass tags, and S.M. provided mouse tissue samples. U.M. and H.L. constructed Tet expression plasmids. C.G.S. and M.V. performed Uhrf1/2 over-expression in HEK-293T cells. P.S. and D.S. provided plasmids and cell lines. M.M. and T.C. conceived and supervised the project, interpreted data and wrote the manuscript.

Competing financial interests

The authors declare no competing financial interests.

Additional information

Supplementary information is available in the online version of the paper. Reprints and permissions information is available online at http://www.nature.com/reprints/index.html. Correspondence and requests for materials should be addressed to M.M. and T.C.

ONLINE METHODS

General materials and methods. Chemicals were purchased from Sigma-Aldrich, Fluka, ABCR or Acros Organics and used without further purification. Acetonitrile of LC/MS grade was purchased from Carl Roth GmbH + Co., KG. Formic acid, p.a. for MS, was purchased from Fluka, and water was purified with a Milli-Q Plus system from Merck Millipore.

The MS standards 5-methyl-2′-deoxycytidine (mC), 5-trideuteromethyl-2′-deoxycytidine ([D_3]mC), 5-hydroxymethyl-2′-deoxycytidine (hmC), 5-dideuterohydroxymethyl-2′-deoxy-(N^1,N^3–$^{15}N_2$)-cytidine ([$^{15}N_2$,D_2]hmC), 5-formyl-2′-deoxycytidine (fC), 5-formyl-2′-deoxy-(N^1,N^3–$^{15}N_2$)-cytidine ([$^{15}N_2$]fC), 5-carboxy-2′-deoxycytidine (caC), 5-carboxy-2′-deoxy-(N^1,N^3–$^{15}N_2$)-cytidine ([$^{15}N_2$]caC), 5-hydroxymethyl-2′-deoxyuridine (hmU), 5-(dideuterohydroxymethyl)-2′-deoxyuridine ([D_2]hmU), 5-formyl-2′-deoxyuridine (fU) and 5-formyl-2′-deoxy-(N^1,N^3–$^{15}N_2$)-uridine ([$^{15}N_2$]fU) were synthesized according to earlier published work[19,21,42]. All of the synthesized compounds were characterized and purity confirmed by ^{1}H-NMR, ^{13}C-NMR and ESI-MS, and some were additionally validated by ^{15}N-NMR. 8-hydroxy-2′-deoxy-($^{15}N_5$)-guanosine ([$^{15}N_5$]8-oxo-G) (99 atom% ^{15}N) was purchased from Cambridge Isotope Laboratories; 8-hydroxy-2′-deoxyguanosine (8-oxo-G) was from BIOLOG; 2′-deoxyguanosine (G) and 2′-deoxycytidine (C) were from ChemGenes. (*methyl*-^{13}C,D_3)-L-Met (99 atom% D and ^{13}C) was purchased from Sigma-Aldrich, and 2′-deoxy-(C^2–^{13}C, N^1,N^3–$^{15}N_2$)-thymidine (99 atom% ^{13}C and ^{15}N) from Hartmann Analytic. Aqueous stock solutions of these compounds were stored at −20 °C and warmed up to RT before usage.

Oligonucleotide synthesis. Oligonucleotide synthesis was performed on an ABI 394 DNA/RNA synthesizer (Applied Biosystems) using standard DNA synthesis conditions (DMT off) and acetyl-protected dC. Phosphoramidites (including the 5-hydroxymethyl-dU-CE phosphoramidite and the 5′-biotin phosphoramidite) and polystyrene carriers were obtained from Glen Research. 5-octadiynyl-dU (Z) phosphoramidite was synthesized according to literature and characterized by ^{1}H-NMR, ^{13}C-NMR, ^{31}P-NMR and ESI-MS[43]. The crude oligonucleotide was cleaved from the resin and deprotected in 30% (v/v) ammonium hydroxide and 40% (v/v) methylamine (1:1) at 65 °C for 10 min and purified by preparative and analytical HPLC (Waters Breeze and Alliance, respectively). Separation was performed by applying a VP 250/10 Nucleosil 100-7 C18 column (flow: 5 ml/min) from Macherey-Nagel with a gradient of buffer A (0.1 M $NHEt_3OAc$ in water) and buffer B (0.1 M $NHEt_3OAc$ in 80% MeCN). DNA-containing fractions were characterized by MALDI-TOF (Bruker Autoflex II) and analytical HPLC, combined, and desalted by C18-Sep-Pak cartridges (Waters). For analytical HPLC, separation was performed by applying a CC 250/4 Nucleosil 120-3 C18 column from Macherey-Nagel (flow: 0.5 ml/min) with the aforementioned buffer system.

For the protein pulldown studies, an hmU-containing oligonucleotide was hybridized with a DNA strand modified with a 5′-biotin and a 5-octadienyl-dU. The sequences of the DNA strands are summarized in **Supplementary Table 3**. To 10 nmol of the dsDNA, 0.5 µl of a 200 mM solution of the cross-linking azide-PEG_3-S-S-NHS ester ($C_{18}H_{29}N_5O_8S_2$, Jena Bioscience, Jena, Germany) in DMSO was added. In a separate tube, CuBr was dissolved in a TBTA solution (DMSO/*t*BuOH 3:1, 100 mM) resulting in a 1:1 Cu(I):TBTA ratio. This solution was immediately added to the DNA/azide mixture. Furthermore, 45 µl of DMSO/*t*BuOH (3:1) were added, the mixture was shaken at 37 °C for 3 h, and the resulting product was purified by ethanol precipitation.

LC/MS analysis of DNA samples. Quantitative LC/UV-ESI-MS/MS analysis of digested DNA samples was performed using an Agilent 1290 UHPLC system equipped with a UV detector and an Agilent 6490 triple quadrupole mass spectrometer coupled with the stable isotope dilution technique. An improved method, based on earlier published work[29,42,44,45], was developed, which allowed the concurrent analysis of all nucleosides in one single analytical run. The source-dependent parameters were as follows: gas temperature 50 °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 2,500 V in the positive ion mode, capillary voltage −2,250 V in the negative ion mode and nozzle voltage 500 V. The fragmentor voltage was 380 V. Delta EMV was set to 500 (positive mode) and 800 (negative mode). Compound-dependent parameters are summarized in **Supplementary Tables 4** and **5**. Chromatography was performed by a Poroshell 120 SB-C8 column (Agilent, 2.7 µm, 2.1 mm × 150 mm) at 30 °C using a gradient of water and MeCN, each containing 0.0085% (v/v) formic acid, at a flow rate of 0.35 ml/min: 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. The effluent up to 1.5 min and after 9 min was diverted to waste by a Valco valve. The autosampler was cooled to 10 °C. The injection volume was amounted to 29 µl. Calibration curves, method validation and data processing are in **Supplementary Note 2**. A complete compilation of LC/MS quantifications results see **Supplementary Note 3**.

DNA digestion. 5–25 µg of genomic DNA in 25 µl H_2O were digested as follows: An aqueous solution (7.5 µl) of 480 µM $ZnSO_4$, containing 42 U nuclease S1 (*Aspergillus oryzae*, Sigma-Aldrich), 5 U Antarctic phosphatase (New England BioLabs) and specific amounts of labeled internal standards (**Supplementary Note 2**) were 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 0.2 U snake venom phosphodiesterase I (*Crotalus adamanteus*, USB corporation), the sample was incubated for another 3 h at 37 °C and then stored at −20 °C. Prior to LC/MS/MS analysis, samples with up to 15 µg DNA, for which the quantification of low amounts of caC was aspired, were filtered by using an AcroPrep Advance 96 filter plate 0.2 µm Supor (Pall Life Sciences). In contrast, samples with 15–25 µg DNA (isotope-tracing experiments) were filtered by using an AcroPrep Advance 96 filter plate 10K Omega (Pall Life Sciences).

Genomic DNA isolation. Tissues of female WT mice (C57-BL6/N) were dissected at postnatal day 90 and prepared as earlier described[19,21]. Genomic DNA was extracted using the Qiagen Blood and Cell Culture DNA Midi Kit except for mESC samples differentiated in the presence of growth factors (see below). Extraction was performed following the manufacturer's instructions for genomic DNA isolation from cell culture samples or tissue samples, respectively. All buffers until loading of the sample on Genomic-tip 100/G were additionally supplemented with antioxidants 3,5-di-*tert*-butyl-4-hydroxytoluene. (BHT, 200 µM) and deferoxamine mesylate salt (desferal, 200 µM) as well as the deaminase inhibitor tetrahydrouridine (THU, 200 µM), according to published methods, to reduce background oxidation or deamination[18]. Elution buffer QF was supplemented with 200 µM BHT. Following elution, all steps were performed on ice. DNA was then precipitated with NaOAc (0.3 M final) and 0.7 volumes *i*PrOH. DNA pellets from cultured cells were washed twice with ice-cold 70% EtOH and resuspended in H_2O containing 20 µM BHT using a Qiagen TissueLyser (30 Hz, 2 min). DNA pellets from mouse tissues were resuspended in PBS buffer and additionally extracted with phenol/$CHCl_3$, precipitated, washed and resuspended as described above.

R1 mESC samples differentiated in the presence of growth factors or transfected with Smug1 esiRNAs were lysed directly in the plates with RLT buffer (Qiagen) supplemented with BHT and desferal as described above. DNA was isolated using the Zymo Quick gDNA Midi Kit according to the manufacturer's instruction, except that elution was repeated four times with 100 µl of elution buffer supplemented with BHT (200 µM). Eluted DNA was precipitated with 2 M ammonium acetate and two volumes of absolute ethanol and finally resuspended in H_2O containing 20 µM BHT. The flow-through from the spin columns was used to isolate RNA (see real-time PCR analysis).

mESC cell culture. Feeder independent WT01 mESCs (C57BL/6 strain)[46] were cultured in the presence of serum and LIF as previously described[4]. *Tdg*$^{+/-}$ and *Tdg*$^{-/-}$ mESCs were described previously[47]. *Tdg*$^{-/-}$ mESCs were complemented by random integration of either empty vector (hereafter referred to as *Tdg*$^{-/-}$ mESCs) or a minigene expressing catalytically inactive Tdg (N151A)[47]. Clonal mESC lines with targeted Tdg alleles, R1 cells (strain 129/Sv)[48], J1 cell lines (strain 129S4/SvJae)[49] and a mESC line derived from C57Bl/6/129 mixed background[50] were routinely maintained on gelatinized plates in DMEM (PAA or Sigma) supplemented with 10% FBS, 1× MEM-nonessential amino acids (NEAA), 0.2 mM L-alanyl-L-glutamine, 100 U/ml penicillin, 100 µg/ml streptomycin (all from PAA), 0.1 mM β-mercaptoethanol, 20 ng/ml (≥1 ×10^3 U/ml) mouse recombinant LIF (ORF Genetics), 1 µM PD 0325901 and 3 µM CHIR 99021 (2i; both from Axon Medchem). In these conditions, the global levels of genomic mC were very low (and, as a consequence, the levels of its oxidized derivatives were even lower; data not shown). Before DNA isolation, 2i cultures were passaged twice (over 5 d) in DMEM supplemented with FBS and LIF as above but lacking 2i. With this strategy, primed mESC cultures were obtained with no sign of overt differentiation and modified genomic cytosines reached reproducibly higher and stable levels. For isotope tracing with heavy

thymidine in serum-primed mESCs 2i cultures of R1, cells were passaged twice (5 d) in the same serum-containing medium lacking 2i and simultaneously supplemented with 100 μM [^{13}C,$^{15}N_2$]T. For isotope tracing with heavy methionine in serum-primed mESCs, 2i cultures of R1 cells and mESC lines with targeted Tdg alleles were passaged twice (over 5 d) without 2i in L-Met-free DMEM (Life Technology) supplemented as above and with 0.2 mM of either [*methyl*-^{13}C,D_3]L-Met or natural L-Met. For mESC differentiation without growth factors, R1 cells and the C57Bl/6/129 mixed background cell line were first plated at 1×10^5 cells/cm^2 on gelatin-coated plates in N2B27 medium containing 1,000 U/ml LIF to favor attachment and initial survival[51]. After 12 h, the medium was replaced without addition of LIF (defined as time point 0 h). The medium was replaced once more at 24 h. For isotope tracing with [*methyl*-^{13}C,D_3]L-Met during mESC differentiation, R1 cells were cultured for two passages in L-Met-free N2B27 medium supplemented with LIF, 2i and 0.2 mM of either [*methyl*-^{13}C,D_3]L-Met or natural L-Met. Differentiation of R1 cells in the presence of FGF-2 and ActA was as described[26], with minor modifications. Briefly, mESCs were cultured in N2B27 medium containing 2i and 1,000 U/ml LIF for several passages and then seeded at 2.2×10^5 cells/cm^2 in N2B27 medium containing 1% KnockOut Serum Replacement (Life Technologies), 12 ng/ml FGF-2 (PeproTech) and 20 ng/ml ActA (ORF Genetics) on plates coated with a thin layer of Geltrex extracellular matrix preparation (Life Technologies). The medium was exchanged after 24 h.

Knockdown experiments in mESCs. shRNA expression vectors targeting Tet1 and Tet2 were generated by cloning synthetic oligonucleotides in pLKO.1 (ref. 52). Recombinant lentiviruses were produced by cotransfecting pLKO.1 shRNA expression vectors and packaging plasmids in HEK-293 cells. 48 h after transduction in the presence of 8 μg/ml polybrene, shRNA-expressing mESCs were selected with 4 μg/ml puromycin. Cell pools were continuously cultured in the presence of puromycin. shRNA target sequences were as follows: SCR (control), 5′-CCT AAG GTT AAG TCG CCC TCG-3′ (ref. 52); Tet1, 5′-TGT AGA CCA TCA CTG TTC GAC-3′ (see ref. 52), Tet2: 5′-TTC GGA GGA GAA GGG TCA TAA-3′. esiRNAs for Smug1 knockdown were generated as described[53]. The cDNA template for *in vitro* transcription was generated by PCR using following primers: forward, 5′- CGT AAT ACG ACT CAC TAT AGG GAG CCC GTG GGT G-3′, and reverse, 5′-CGT AAT ACG ACT CAC TAT AGG GGT TTC GTC CAC TGG G-3′. R1 mESCs were weaned from 2i for two passages in FBS- and LIF-containing medium as described above. Upon plating the second passage, the cells were transfected in a p60 plate with 6 μg of Smug1 esiRNAs (34.5 nM) and 20 μl of Lipofectamine RNAi MAX (Life Technologies) according to the manufacturer's instructions and were lysed 72 h after transfection.

Culture and transfection procedures for HEK-293T cells. All transfections were performed using jetPRIME transfection reagent (PEQLAB Biotechnologie GmbH) according to the manufacturer's instructions. HEK-293T cells were seeded 24 h before transfection at a density of 2.5×10^6 cells per 75 cm^2 flask and incubated in 10 ml of medium. The transfection solution (500 μl of jetPRIME buffer, a specific amount of plasmid DNA (**Supplementary Note 4**) and 20 μl of jetPRIME reagent) was added to the medium, and the cells were incubated for 48 h, with an additional medium exchange 24 h after transfection. When cotransfection of esiRNA was performed, a second transfection step (500 μl of jetPRIME buffer, 5 μg of esiRNA and 20 μl of jetPRIME reagent) was carried out 4 h after transfection of plasmid DNA. esiRNAs were purchased from Sigma (human TDG esiRNA EHU038971; human SMUG1 esiRNA EHU098861; human CDK5RAP1 esiRNAEHU079221). **Supplementary Note 4** summarizes the overexpression and knockdown procedures.

Isotope tracing with [^{13}C,$^{15}N_2$]thymidine or [^{13}C,D_3]L-Met in HEK-293T cells transfected with Tet1cd. 24 h before transfection, 2.5×10^6 cells were seeded in a 75-cm^2 flask containing 10 ml either of (for [^{13}C,$^{15}N_2$]thymidine) DMEM medium supplemented with 50 μM [^{13}C,$^{15}N_2$]T or (for [^{13}C,D_3]L-Met) DMEM medium lacking L-Met, L-cystine and pyruvate, which was supplemented with 10% dialyzed FBS, 2 mM [^{13}C,D_3]L-Met and 0.2 mM L-cystine. Transfection was performed as described above using labeled medium.

Real-time PCR analysis of mRNA expression. For analysis of Tet1 and Tet2 knockdown in mESC total RNA was prepared with Trizol (Invitrogen), cDNA synthesis was performed with Quantitect reverse transcription kit from Qiagen, and real-time PCR was performed with the Power Sybr Green PCR master mix from Applied Biosystems on an Applied Biosystems 7500 Fast system. Knockdown efficiencies relative to control samples transfected with SCR esiRNAs were 79% and 70% for Tet1 and Tet2, respectively. The primers used to estimate them are listed in **Supplementary Note 4**. For analysis of Smug1 knockdown in mESCs and EpiLC differentiation samples, total RNA was prepared with RNeasy spin columns (Qiagen), followed by DNase treatment using TURBO DNA-free (Ambion, Life Technologies); cDNA synthesis was carried out using iScript cDNA Synthesis kit (Bio-Rad); real-time PCR was performed with SsoFast EvaGreen Supermix (Bio-Rad). Smug1 knockdown efficiency relative to control samples transfected with esiRNAs targeting GFP was estimated to be 60%. Quantification of Tet, Dnmt, Tdg and Smug1 transcripts during EpiLC differentiation and Smug1 knockdown samples was performed using the primers listed in **Supplementary Note 4**. Expression levels were quantified with respect to the housekeeping gene *Gapdh* and normalized to time point 0 h.

Tet *in vitro* assay. A plasmid was prepared from *dam*$^-$/*dcm*$^-$ competent *E. coli* strain (New England BioLabs) and methylated with M.SssI (New England BioLabs). 1.5 μg of plasmid DNA were then treated with recombinant Tet1 from the 5hmC TAB-Seq Kit (Wisegene) corresponding to ref. 54. After 3 h incubation at 37 °C and proteinase K treatment, the oxidized plasmids were purified with GeneJET PCR Purification Kit from Thermo Scientific and eluted in 25 μl water. Samples were then subjected to LC/MS/MS analysis as described[42]. The results are compiled in **Supplementary Note 3**.

Correlation and cluster analyses. Statistical data analysis was performed using IBM SPSS Statistics 19. Results of bivariate correlation analyses are summarized in **Supplementary Note 5**. Unsupervised clustering of species with respect to its correlation coefficients was applied by average linkage hierarchical clustering using a squared Euclidean distance measure.

Pulldown assay. For the pulldown assay, 250 μg (50 μl) of the crude nuclear protein extracts were filled up to 500 μl with 50 mM TEAB and 1 mM $MgCl_2$. The binding conditions were 45 mM TEAB, 1.1 mM $MgCl_2$, 2 mM HEPES, 42 mM NaCl and 20 μM EDTA. Complete Protease Inhibitor Cocktail Tablets were used from Roche Diagnostics (Indianapolis, IN, USA). The DNA oligomers with DNA-protein cross-linker were dissolved in neat DMSO, and 1 nmol was added to the protein lysate and incubated for 20 min at room temperature. Streptavidin-coated magnetic particles (Roche Diagnostics, Indianapolis, IN, USA) were washed three times with binding buffer (100 mM NaCl, 10 mM Tris, 1 mM EDTA, pH 7.4) before 200 μl of the bead slurry (equal to 2 mg beads) were added to the sample. Following 2 h incubation at room temperature under constant rotating, the beads were washed three times with 50 mM TEAB and 1 mM $MgCl_2$. The beads were reconstituted in 50 mM TEAB and 1 mM $MgCl_2$. Disulfide bonds of the cross linker were cleaved and alkylated in the process of enzymatic digestion, and the magnetic particles were removed before adding trypsin (described below).

Protein sample preparation. Cell lysis of mouse embryonic stem cells was performed as described in ref. 29. For each lysis, approximately 7.5×10^7 cells were used. Protein concentration was determined by Bradford assay. For each experiment 250 μg (50 μl) of the crude nuclear protein extract were used. Protein samples for MS analysis were reduced by adding 100 mM TCEP and by incubating on a shaker at 650 r.p.m. for 1 h at 60 °C and subsequently alkylated with 200 mM iodoacetamide in the dark for 30 min at 25 °C. Protein samples were digested with 0.5 μg trypsin (Promega, Madison, MA, USA) for 16 h at 37 °C. The reaction was stopped using 1 mM phenylmethylsulphonylfluoride. After tryptic digestion, peptide labeling with the TMT2plex reagents (Thermo Fisher Scientific, Waltham, MA, USA) was performed according to the manufacturer's instructions. TMT2plex reagents 126 and 127 were used to label the samples. When the sample (proteins enriched with hmU- or hmC-containing DNA strands) was labeled with TMT126, the control sample (proteins enriched with no modified DNA-bases) was labeled with the TMT127 reagent and vice versa. Subsequent to the labeling, both sample and control, were combined. This way, each experiment was performed twice as a so-called label swap experiment. Organic solvent was removed by vacuum centrifugation, and the sample was finally reconstituted in 1% (v/v) formic acid for MS analysis.

LC/MS analysis of protein samples. The samples were analyzed using an UltiMate 3000 nano liquid chromatography system (Dionex, Fisher Scientific, Waltham, MA, USA) coupled to an LTQ-Orbitrap XL (Thermo Fisher

doi:10.1038/nchembio.1532

Scientific, Waltham, MA, USA). Of each eluate, 15 µl were injected for the analysis. The samples were desalted and concentrated on a µ-precolumn cartridge (PepMap100, C18, 5 µM, 100 Å, size 300 µm i.d. x 5 mm) and further processed on a custom-made analytical column (ReproSil-Pur, C18, 3 µM, 120 Å, packed into a 75 µm i.d. x 150 mm and 8 µm picotip emitter). A 57-min multistep analytical separation was performed at a flow rate of 300 nl/min. In the first 50 min, a linear gradient was ramped up from 5% (v/v) solvent B (acetonitrile containing 0.1% formic acid and 5% DMSO) and 95% solvent A (water containing 0.1% formic acid and 5% DMSO) to 95% solvent B. This level was held for 5 min and then ramped down again to 5% solvent B within 2 min. Mass spectrometric analyses were performed starting with a full mass scan in the mass range between m/z 300 and m/z 1,650. This survey scan was followed by three MS/MS scans using the FTMS mass analyzer and high normalized collision energy of 70 in the HCD cell and three additional scans using the ion trap mass analyzer and a normalized collision energy of 35.

Protein identification and relative quantification method. The Thermo Proteome Discoverer 1.1 software (Thermo Fisher Scientific, Waltham, MA, USA) was used for protein identification and for relative quantification. The Sequest (Thermo Fisher Scientific, Waltham, MA, USA) search engine was used in combination with a Uniprot database (*Mus musculus*; date of download, 04/2013). As a limit of detection, a ratio of threefold signal over the noise filter was applied. A maximum of two missed cleavage sites was allowed. The mass tolerances were 10 p.p.m. for the precursor mass and 0.5 Da for the fragment ion mass. Carbamidocysteine was set as static modification. Dynamic modifications were: cation, Na (D, E); the residue of the DNA-protein crosslinker (+146.028 Da; K, Y); Oxidation (M) as well as TMT2plex (N-term. and K). Identified, nonredundant peptides, which were labeled with the TMT2 reagent, were used for relative quantification. The integration window tolerance was 20 p.p.m., and the integration method was set to 'most confident centroid'. The signals of the TMT2 reporter ions 126 and 127 were used to calculate ratios and monitor either preferred or nonpreferred binding of the identified proteins to the modified DNA bases in comparison to the control strand. From the identified proteins, the only proteins considered as 'specific readers' were enriched in both the forward and the reverse experiment.

42. 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).
43. Gierlich, J. *et al.* Click chemistry as a reliable method for the high-density postsynthetic functionalization of alkyne-modified DNA. *Org. Lett.* **8**, 3639–3642 (2006).
44. Cao, H. & Wang, Y. Collisionally activated dissociation of protonated 2′-deoxycytidine, 2′-deoxyuridine, and their oxidatively damaged derivatives. *J. Am. Soc. Mass Spectrom.* **17**, 1335–1341 (2006).
45. Wang, J. *et al.* Quantification of oxidative DNA lesions in tissues of Long-Evans Cinnamon rats by capillary high-performance liquid chromatography-tandem mass spectrometry coupled with stable isotope-dilution method. *Anal. Chem.* **83**, 2201–2209 (2011).
46. Chen, T., Ueda, Y., Dodge, J.E., Wang, Z. & Li, E. Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. *Mol. Cell. Biol.* **23**, 5594–5605 (2003).
47. Cortazar, D. *et al.* Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. *Nature* **470**, 419–423 (2011).
48. Nagy, A., Rossant, J., Nagy, R., Abramow-Newerly, W. & Roder, J.C. Derivation of completely cell culture-derived mice from early-passage embryonic stem cells. *Proc. Natl. Acad. Sci. USA* **90**, 8424–8428 (1993).
49. Li, E., Bestor, T.H. & Jaenisch, R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915–926 (1992).
50. Montanez, E. *et al.* Kindlin-2 controls bidirectional signaling of integrins. *Genes Dev.* **22**, 1325–1330 (2008).
51. Ying, Q.L. & Smith, A.G. Defined conditions for neural commitment and differentiation. *Methods Enzymol.* **365**, 327–341 (2003).
52. Williams, K. *et al.* TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* **473**, 343–348 (2011).
53. Kittler, R., Heninger, A.K., Franke, K., Habermann, B. & Buchholz, F. Production of endoribonuclease-prepared short interfering RNAs for gene silencing in mammalian cells. *Nat. Methods* **2**, 779–784 (2005).
54. Yu, M. *et al.* Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell* **149**, 1368–1380 (2012).

4 Unveröffentlichte Arbeiten

4.1 Analyse der Umsetzung von mdC zu dT

4.1.1 Einleitung

Wie im Abschnitt 1.2.2 beschrieben, kommt als der Schlüsselschritt der aktiven DNA-Demethylierung die Desaminierung von mdC zu dT in Frage. Dieser Weg und seine Rolle in der aktiven DNA-Demethylierung werden zurzeit kontrovers diskutiert. Im Rahmen vorliegender Dissertation wurden massenspektrometrische Studien durchgeführt, um tiefere Einblicke in dieses Thema zu gewinnen.

Für die Bestimmung unterschiedlicher Mengen an DNA-Metaboliten im Spurenbereich ist die massenspektrometrische Analyse in Verbindung mit Isotopenverfolgung die Methode der Wahl. So können nach der Fütterung von embryonalen Stammzellen von Mäusen (mESC) mit isotopenmarkierten Metaboliten, diese Markierung im Zuge des Stoffwechsels in die DNA eingebaut und mit Hilfe der massenspektrometrischen Methoden verfolgt werden. Durch die Zugabe von [^{13}C, D_3]-markierten L-Methionins ([^{13}C, D_3-*methyl*]-Methionin) zum Zellkulturmedium findet ein gezielter Einbau der [^{13}C, D_3]-Markierung auf den Kofaktor der DNA-Methyltransferasen, *S*-Adenosylmethionin (SAM) statt. Anschließend wird das [^{13}C, D_3]-SAM durch die Dnmts für die Methylierung von Cytidin rekrutiert und folglich das [^{13}C, D_3]-Methylcytidin erzeugt (Abbildung 4.1).[115] Die so markierte Methylgruppe des Cytosins wird im Laufe der aktiven DNA-Demethylierung abgebaut. Falls eine Umwandlung durch eine Desaminierung von [^{13}C, D_3]-markierten mdC zum [^{13}C, D_3]-markierten dT passiert, kann dies massenspektrometrisch verfolgt werden. Nach der DNA-Isolation und enzymatischer Hydrolyse bis zum Nukleosid-Level kann durch die Zugabe einer bestimmten Menge von isotopenmarkiertem internen Standard [$^{13}C_{10}$,$^{15}N_2$]-dT der Gehalt des entstandenen [^{13}C, D_3]-markierten dT ([^{13}C, D_3]-dT) massenspektrometrisch bestimmt werden. Der zugegebene markierte Standard besitzt das gleiche chromatographische Verhalten und denselben Ionisationscharakter wie das zu quantifizierende [^{13}C, D_3]-dT, weist aber einen

Massenunterschied von m/z = 8 zum gesuchten [^{13}C, D_3]-dT auf. Nach der Integration der Massensignale beider Nukleoside, kann mit Hilfe der für die exakte Quantifizierung benötigten Eichgeraden, der genaue Gehalt an [^{13}C, D_3]-dT bestimmt werden. Folglich können neue Erkenntnisse über den Anteil der Desaminierung von mdC während der aktiven DNA-Demethylierung erzielt werden. Ein Weg über eine spontane Desaminierung von dC zu dU mit anschließender Methylierung zu dT kann ausgeschlossen werden, da in diesem Fall die [^{13}C, D_3]-Markierung fehlen würde (Abbildung 4.1).

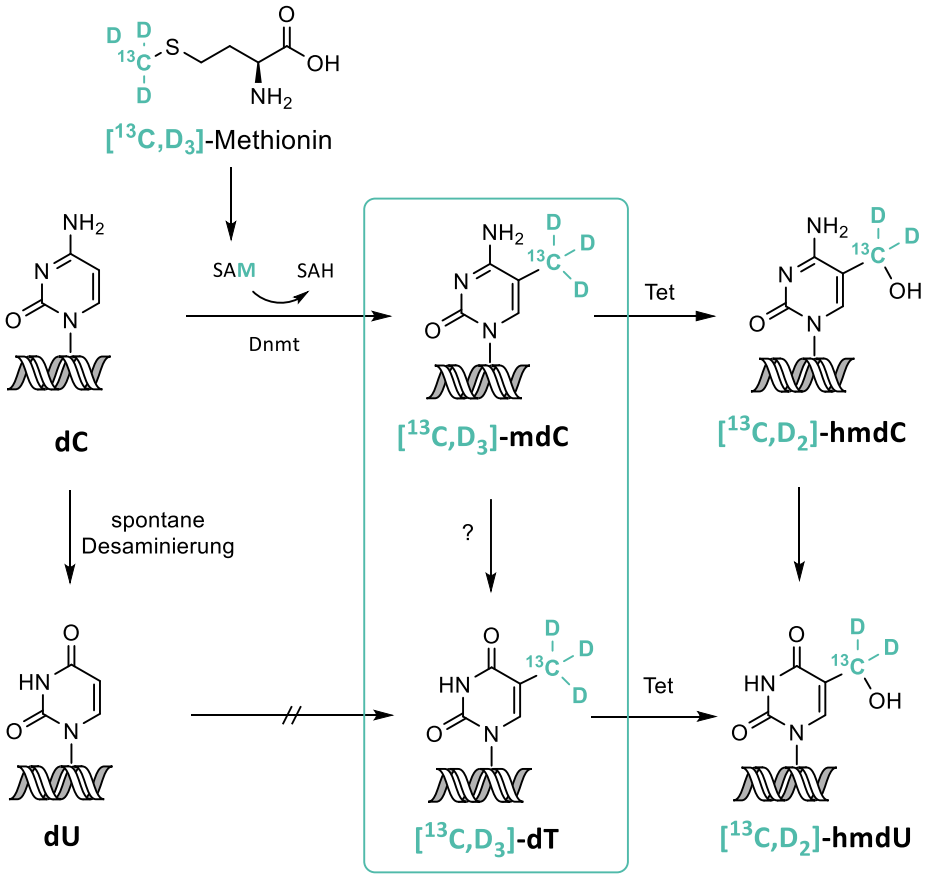


Abbildung 4.1: Desaminierung von mdC unter metabolischer Markierung mit [^{13}C,D_3-*methyl*]-Methionin.

4.1.2 Ergebnisse und Diskussion

Massenspektrometrische Aspekte

Im Rahmen vorliegender Dissertation wurde eine neue sensitive massenspektrometrische Methode für die Identifikation und Quantifizierung von schwermarkierten [^{13}C, D_3]-dT, das aus [^{13}C, D_3]-mdC stammt, entwickelt. Somit konnte der umstrittene Prozess der Desaminierung vom mdC zu dT als ein möglicher Schritt bei der aktiven DNA-Demethylierung im Detail studiert werden. Diese Methode basiert auf einer veröffentlichen Quantifizierungsmethode für die Gehaltsbestimmung von [^{13}C, D_3]-hmdU, das über Fütterungsexperimente mit [^{13}C, D_3-*methyl*]-Methionin ([^{13}C, D_3]-Met) metabolisch markiert wurde.[115] Eine erhebliche Schwierigkeit bei der Entwicklung einer sensitiven Quantifizierung von schwermarkierten [^{13}C, D_3]-dT besteht in den Ionensuppressionseffekten, die von natürlich vorkommenden dT in der DNA ausgehen. Es findet eine konkurrierende Ionisierung statt, so tragen hohe Mengen an dem in der DNA natürlich vorkommenden dT zu einem hohen Messhintergrund bei. Zudem wird die Ionensuppression durch eine späte Elution von dT bei der chromatographischen Auftrennung des Substanzgemisches an der UHPLC verstärkt. Mit steigendem Anteil an organischem Lösungsmittel im Laufpuffer erhöht sich außerdem die Menge an eluierenden Verunreinigungen. Sie werden im Massenspektrometer ionisiert und tragen zusätzlich zu einem hohen Hintergrund bei. Auch der Einsatz einer großen Menge an Verdaukomponenten, wie Enzymen, Puffer und Inhibitoren, verschlechtert das Signal-zu-Rausch Verhältnis des zu analysierenden Peaks. Außerdem wurde unter Berücksichtigung des Isotopenverteilungsmusters ein 0,01%iger natürlicher Anteil an um 4 Masseneinheiten schwererer dT Spezies berechnet. Dies entspricht genau der Masse des zu quantifizierenden [^{13}C, D_3]-dT und muss von diesem unterschieden werden. Das Gleiche betrifft auch den isotopenmarkierten Standard [$^{13}C_{10}$, $^{15}N_2$]-dT, der für die Quantifizierung von [^{13}C, D_3]-dT eingesetzt wurde. Durch einen natürlichen Anteil von (dT+4)-Spezies im isotopenmarkierten Standard könnte es ebenfalls zur Verfälschung der Quantifizierungsergebnisse kommen. Unter Berücksichtigung oben aufgeführter Punkte,

wurde die enzymatische Hydrolyse der genomischen DNA für die Quantifizierung [^{13}C, D_3]-dT so optimiert, dass ein solides Signal mit einem guten Signal-zu-Rausch-Verhältnis erzielt werden konnte. Zusammenfassend trugen eine geringe Menge an DNA und folglich geringere Mengen an Verdaubestandteilen zu einer erfolgreichen und robusten Quantifizierung von [^{13}C, D_3]-dT mittels Massenspektrometrie bei. Ein Cytidin-Desaminase-Inhibitor Tetrahydrouridin (THU), der zu dem DNA-Verdau zugegeben wurde, sollte die Desaminierung während des enzymatischen Verdaus verhindern.

Übergang vom *naïven* in den *geprimten* Zustand

Alle in diesem Projekt durchgeführten Studien wurden in *geprimten* pluripotenten Stammzellen der Maus (mPPCs) durchgeführt. Mit dem *Priming* wird eine Entwicklung vom präimplantation (*naïver* Zusatand) Epiblasten zum postimplantation Epiblasten simuliert. Um eine möglichst homogene und undifferenzierte mPPC Zellpopulation mit einem hohen Markierungsanteil zu ehrhalten, wurden mESCs erst unter *naïven* Bedingungen (2i/LIF (*leukemia inhibitory factor*)) kultiviert und anschließend in ein Medium mit Serum/LIF oder Serum/CHIR99021/IWR1-endo (Serum/CR)[172] zum *Priming* überführt. Der höchste mdC-Gehalt bei Serum/CR *Priming* wird bereits nach 3 Tagen erreicht. Im Vergleich dazu wird der gleiche mdC-Level beim Serum/LIF *Priming* wird erst nach 8 Tagen erzielt (Abbildung 4.2).
Parallel wurde der Gehalt von [^{13}C, D_3]-dT beim Übergang in den *geprimten* Zustand beim Serum/CR *Priming* ermittelt. Im Laufe der Fütterung der Zellkultur mit [^{13}C, D_3]-Met über 5 Tage steigt der [^{13}C, D_3]-dT-Level sehr schnell an (Abbildung 4.2).

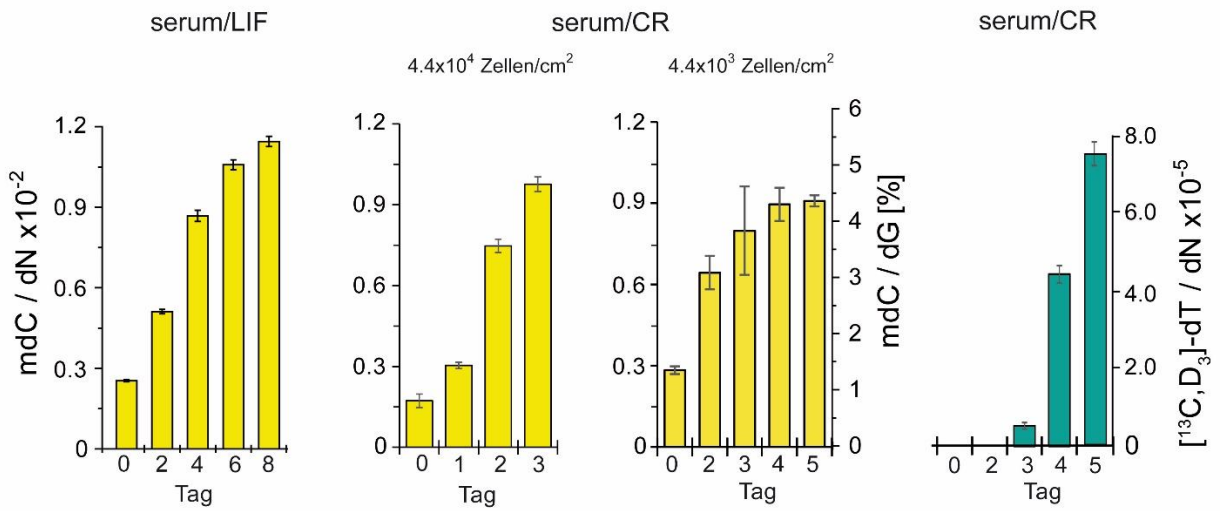


Abbildung 4.2: Absolute Quantifizierung des Gehalts von mdC bei unterschiedlichen *Priming* Bedingungen und von [^{13}C, D_3]-dT unter Serum/CR Bedingungen.

Umsetzung/Abbau von mdC ist ein aktiver Prozess

Um die globale Umsetzung von [^{13}C, D_3]-mdC zu verfolgen, wurden *Pulse-Chase*-Experimente durchgeführt. Bei einem *Pulse-Chase*-Experiment wird das isotopenmarkierte Molekül zu der Zellkultur für eine bestimmte Zeit zugegeben, um eine möglichst hohe Markierungsrate zu erzielen (*Pulse*). Anschließend wird es durch nicht markiertes Material ersetzt und die Veränderungen des isotopenmarkierten Materials beobachtet (*Chase*). Um die Änderungen von isotopenmarkierten, sowie nicht-markierten Molekülen zu verfolgen, werden die Proben zu unterschiedlichen Zeitpunkten entnommen und analysiert. Nach 5 Tagen wurde die Fütterung der Zellkultur mit [^{13}C, D_3]-Met eingestellt. Am letzten Tag der Fütterung wurde das [$^{13}C_{10}$, $^{15}N_2$]-dT zur der Zellkultur zugegeben (Abbildung 4.3A). In diesem Experiment wurde das [$^{13}C_{10}$, $^{15}N_2$]-dT nicht als interner Standard für die Quantifizierung verwendet, sondern diente als Referenz für die passive „Verdünnung“ markierter Nukleoside während der Zellproliferation im Laufe der *Chase*-Periode. Die Veränderung der nach der Fütterung entstandenen Mengen an [^{13}C, D_3]-mdC wurde zu unterschiedlichen Zeitpunkten über 50 h hinweg bestimmt. Außerdem wurde die Umsetzung von [^{13}C, D_3]-mdC mit der des [^{13}C, D_2]-hmdC verglichen (Abbildung 4.3B). [^{13}C, D_2]-hmdC entsteht durch die Oxidation von [^{13}C, D_3]-mdC.

Die Abnahme von [^{13}C, D$_3$]-mdC-Werten war auffallend höher als die von [^{13}C, D$_2$]-hmdC und [^{13}C$_{10}$, ^{15}N$_2$]-dT. Der [^{13}C, D$_2$]-hmdC-Gehalt nimmt geringfügig schneller ab, als der Gehalt von [^{13}C$_{10}$, ^{15}N$_2$]-dT (Abbildung 4.3B). Eine ähnliche Abnahme wurde beim *Priming* unter Serum/CR Bedingungen beobachtet (Abbildung 4.3C/D). Diese Befunde deuten darauf hin, dass es sich bei der Umsetzung von [^{13}C, D$_3$]-mdC um einen aktiven Prozess handelt.

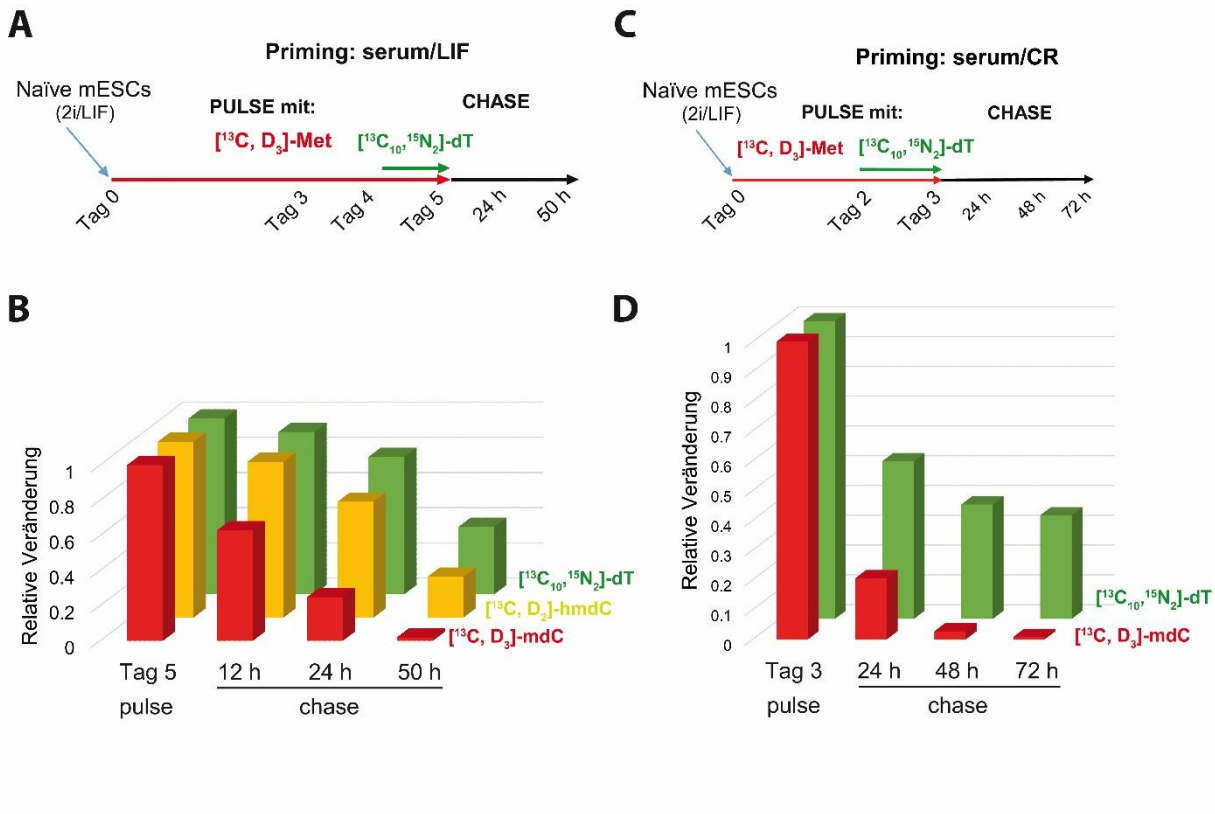


Abbildung 4.3: A Eine schematische Darstellung des *Pulse-Chase*-Experiments während des *Primings* unter Serum/LIF Bedingungen; **B** Relative Veränderung von [^{13}C, D$_3$]-mdC, [^{13}C, D$_2$]-hmdC und [^{13}C$_{10}$, ^{15}N$_2$]-dT bei dem unter **A** beschrieben Experiment. Die abgebildeten Werte wurden wie folgt ermittelt: jedes Massenintegral von [^{13}C, D$_3$]-mdC wurde durch die Summe der Massenintegrale von ([^{13}C, D$_3$]-mdC + mdC) geteilt und auf den höchsten Wert (Tag 5) normiert. Das Gleiche gilt für die Werte von[^{13}C, D$_2$]-hmdC und [^{13}C$_{10}$, ^{15}N$_2$]-dT. **C** Eine schematische Darstellung des *Pulse-Chase*-Experiments während des *Primings* unter Serum/CR Bedingungen; **D** Relative Veränderung von [^{13}C, D$_3$]-mdC und [^{13}C$_{10}$, ^{15}N$_2$]-dT bei dem unter **C** beschrieben Experiment. Die abgebildeten Werte wurden wie unter **B** aufgezeigt, ermittelt.

Der mdC Umsatz ist zu einem großen Teil oxidationsunabhängig

Um die Rolle der Tet-vermittelten Oxidation in dem globalen Umsatz von mdC zu ermitteln, wurde ein *Pulse-Chase*-Experiment in mESCs ohne Tet Proteinen (Tet TKO) durchgeführt. In Tet TKO Zellen fällt der Gehalt an genomischem [^{13}C, D$_3$]-mdC etwas langsamer ab, als in wt Zellen (Abbildung 4.4A), jedoch deutlich schneller als der Gehalt von [$^{13}C_{10}$, $^{15}N_2$]-dT (Abbildung 4.4B). Das zeigt eindeutig, dass die aktive Umsetzung von [^{13}C, D$_3$]-mdC zu einem Teil von einer Tet-vermittelten Oxidation unabhängig ist.

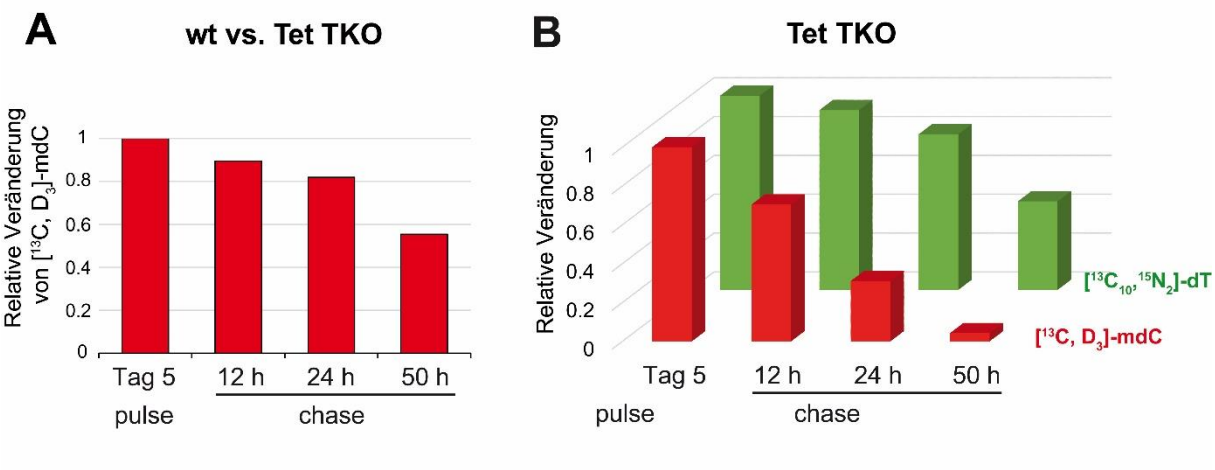


Abbildung 4.4: A Unterschied zwischen wt und Tet TKO in relativen Mengen von[^{13}C, D$_3$]-mdC im Laufe eines *Pulse-Chase*-Experiments während des *Primings* unter Serum/LIF Bedingungen (Abbildung 4.3A). Um den Unterschied zwischen dem wt und Tet TKO anschaulicher darzustellen wurden die ermittelten [^{13}C, D$_3$]-mdC-Werte von wt durch die [^{13}C, D$_3$]-mdC-Werte von Tet TKO dividiert. **B** Relative Veränderungen von [^{13}C, D$_3$]-mdC und [$^{13}C_{10}$, $^{15}N_2$]-dT in Tet TKO mPPCs bei dem in der Abbildung 4.3 A gezeigten *Pulse-Chase*-Experiment.

Die Umsetzung von mdC schließt die Desaminierung von mdC zu dT ein

Die Desaminierung von mdC zu dT wird als erster möglicher Schritt der aktiven DNA-Demethylierung diskutiert (Abschnitt 1.2.2). Folglich könnte ein Teil der sinkenden [^{13}C, D$_3$]-mdC-Werte, die bei dem *Pulse-Chase*-Experiment beobachtet wurden (Abbildung 4.3B), durch die genomische Desaminierung von [^{13}C, D$_3$]-mdC zu [^{13}C, D$_3$]-dT zustande kommen. Um diese These zu überprüfen wurde bei denselben DNA-Proben, die für vorheriges Experiment analysiert wurden, zusätzlich der relative [^{13}C, D$_3$]-dT-Gehalt ermittelt. In

Übereinstimmung mit der Zunahme der Umsetzung von [^{13}C, D$_3$]-mdC, wurde eine Zunahme von [^{13}C, D$_3$]-dT während der ersten 24 Stunden ohne [^{13}C, D$_3$]-Met Fütterung beobachtet (Abbildung 4.5). Dies zeigt einen direkten Zusammenhang zwischen [^{13}C, D$_3$]-mdC und[^{13}C, D$_3$]-dT.

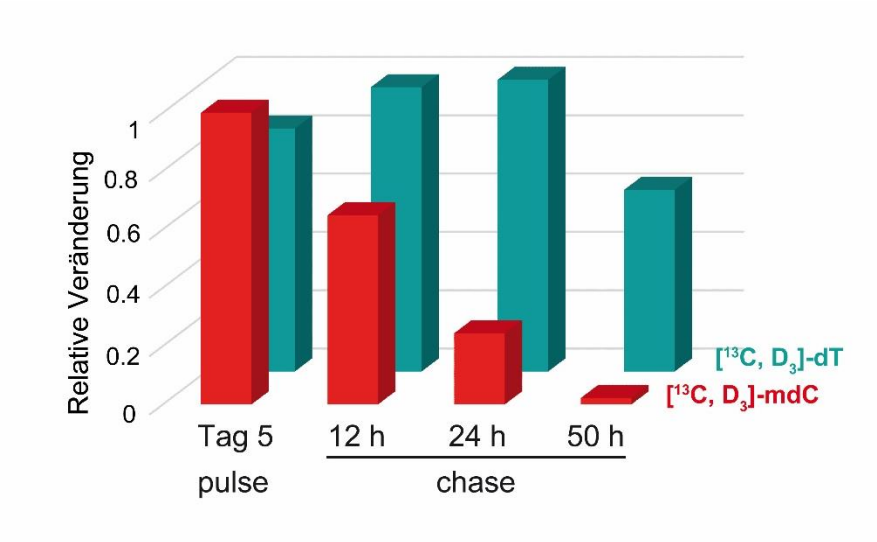


Abbildung 4.5: Relative Veränderung von [^{13}C, D$_3$]-mdC und [^{13}C, D$_3$]-dT. Die abgebildeten Werte wurden wie folgt ermittelt: jedes Massenintegral von [^{13}C, D$_3$]-mdC wurde durch die Summe der Massenintegrale von ([^{13}C, D$_3$]-mdC + mdC) geteilt und auf den höchsten Wert (Tag 5) normiert. Das Gleiche gilt für die Werte [^{13}C, D$_3$]-dT. Es wurde auf 24 h *Chase* normiert.

Als Nächstes wurde überprüft, ob das [^{13}C, D$_3$]-dT tatsächlich aus dem genomischen [^{13}C, D$_3$]-mdC und nicht auf einem anderen Weg entstehen könnte. Hierfür wurden mESCs mit fehlenden katalytisch aktiven Dnmts (Dnmt TKO) untersucht und während des *Primings* mit [^{13}C, D$_3$]-Met gefüttert und analysiert (Abbildung 4.6). Durch das Fehlen von katalytisch aktiven Dnmts sind diese Zellen nicht in der Lage das genomische [^{13}C, D$_3$]-mdC zu generieren. Folglich kann auch kein [^{13}C, D$_3$]-dT gebildet werden.

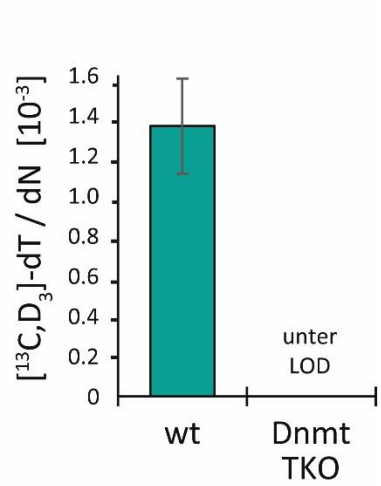


Abbildung 4.6: Absolute Quantifizierung von [^{13}C, D_3]-dT in wildtyp und Dnmt TKO Zellen. LOD = *Lower Limit of Detection,* n = 3.

Wie erwartet, konnte kein Sinal für [^{13}C, D_3]-dT detektiert werden (Abbildung 4.6). Demnach kann ein alternativer Mechanismus, der unabhängig von Dnmts verläuft, für die Generierung von [^{13}C, D_3]-dT ausgeschlossen werden.

Die überwiegende Desaminierung von mdC zu dT geschieht im löslichen Nukleosid/-tid Pool

Schon bei den ersten Messungen zu diesem Projekt wurden nach Fütterung mit isotopenmarkiertem Methionin ungewöhnlich hohe Mengen an [^{13}C, D_3]-dT in der DNA von mESCs quantifiziert. Diese Beobachtung ist dahingehend überraschend, da eine direkte Desaminierung von genomischem mdC zur Entstehung von T:G-Fehlpaarungen führen würde. Diese Fehlpaarungen sind für die Zelle schädlich und werden effizient von der DNA-Reparatursystemen ausgeschnitten. Demensprechend wurden zwei Annahmen für den mdC Metabolismus im löslichen Nukleosid/-tid Pool ausgearbeitet. Bei der ersten Annahme wird nach der Desaminierung von mdC die entstandene Base Thymin sofort von Tdg oder Mbd4 ausgeschnitten und auf eine Desoxyribose durch das Enzym Thymin Phosphorylase (TP) übertragen. Es entsteht das dT (Salvage-Mechanismus)[173-174]. Bei der zweiten Annahme wird ein Strang, der ein mdC enthält, direkt repariert. Es werden lösliche mdC und mdCMP

(Methyldesoxycytidin-Monophosphate) generiert, die anschließend von der Cytidin- bzw. der Desoxycytidylat-Desaminasen (Cda und Dctd) zu dU und dUMP desaminiert und über Thymidylat-Synthase (TS) zu [^{13}C, D_3]-dT und [^{13}C, D_3]-dTMP prozessiert werden. Nach der anschließenden Phosphorylierung wird [^{13}C, D_3]-dTTP durch die DNA-Synthese in das Genom eingebaut (Abbildung 4.7).

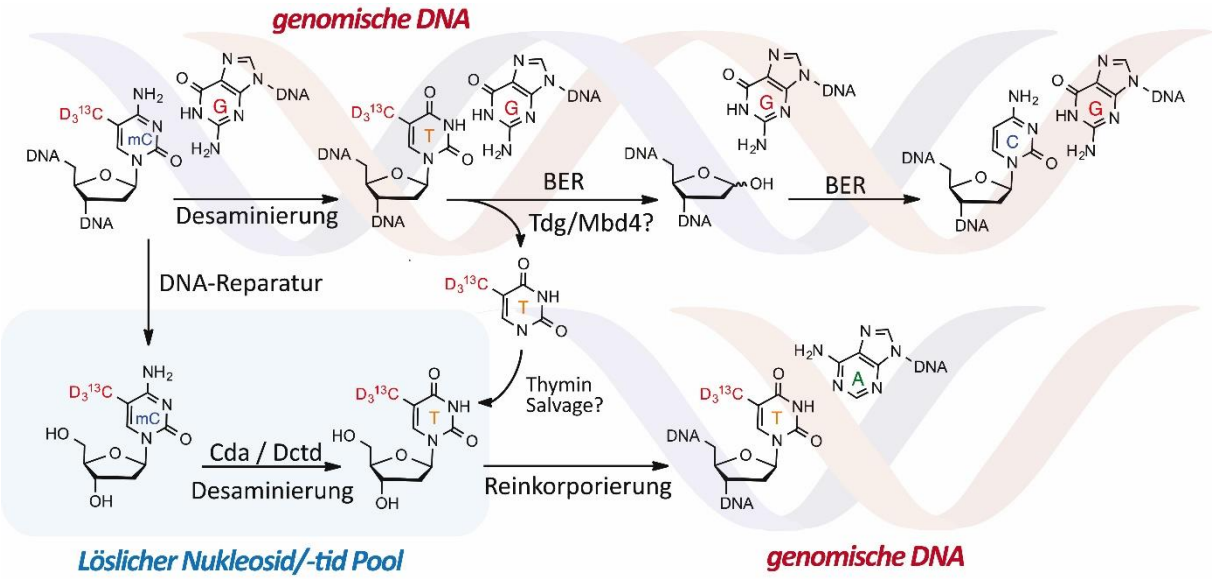


Abbildung 4.7: Mögliche Wege der mdC zu dT Umsetzung. Eine genomische Desaminierung von mdC zu dT führt zu T:G-Fehlpaarungen, die durch BER repariert werden. Anschließend entsteht das unmarkierte C:G-Basenpaar. Eine Desaminierung im löslichen Nukleosid/-tid Pool führt zu einer Markierung in T:A-Basenpaaren.

Um die beiden Annahmen zu testen wurden zwei Ansätze unternommen. Wenn die DNA–Reparatur für das Entstehen von freiem [^{13}C, D_3]-dT und [^{13}C, D_3]-dTMP verantwortlich ist, sollte der Einbau dieser Nukleoside/-tide durch einen Überschuss von unmarkiertem dT im Medium gehemmt werden. Während des *Primings* von mESCs unter Zugabe von [^{13}C, D_3]-Met und einer zunehmenden Konzentration von unmarkiertem dT, konnte eine Abnahme an [^{13}C, D_3]-dT-Werten beobachtet werten (Abbildung 4.8). Nach der Fütterung der Zellkultur mit 20 µM dT wurde etwa 2/5 des Ausgangswertes quantifiziert. Ab einer Konzentration von 150 µM von dT konnte kein Signal für [^{13}C, D_3]-dT mehr detektiert werden. Dieses Ergebnis zeigt eindeutig, dass das genomische [^{13}C, D_3]-dT zu einem großen Teil durch DNA-Synthese aus dem löslichen Pool von [^{13}C, D_3]-dT/dTMP eingebaut wurde.

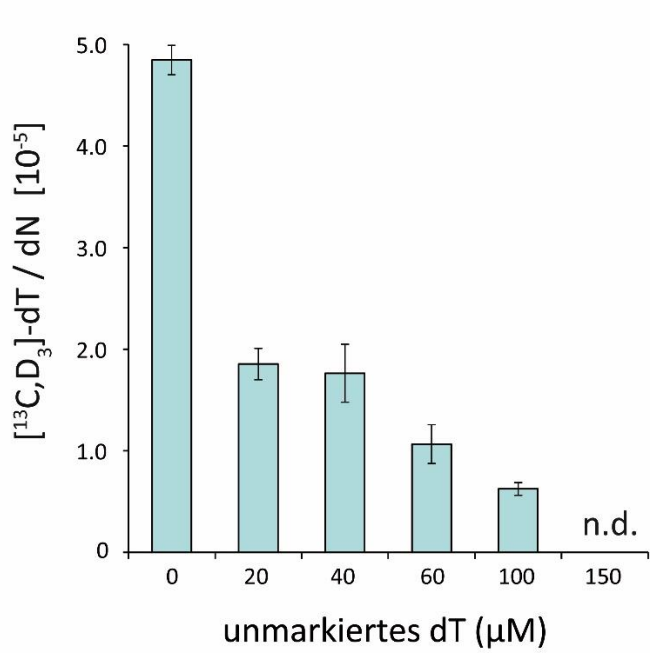


Abbildung 4.8: Absolute Quantifizierung von [^{13}C, D_3]-dT in Anwesenheit von unterschiedlichen Mengen von unmarkierten dT im Cytosol (n.d. = nicht detektierbar).

Um zu untersuchen, ob das [^{13}C, D_3]-mdC in genomischer DNA oder im löslichen Nukleotid/-sid Pool desaminiert wird, wurde mittels CRISPR/Cas9 Ansatzes eine mESC Zelllinie mit einer simultanen Deletion von *Cda* und *Dctd* (CD-DKO (*double knockout*)) erzeugt. Unter der Zugabe von [^{13}C, D_3]-Met beim *Priming* der CD-DKO Zellen, ist der [^{13}C, D_3]-dT-Gehalt im Vergleich zum Wildtyp Zellen deutlich erniedrigt (Abbildung 4.9).

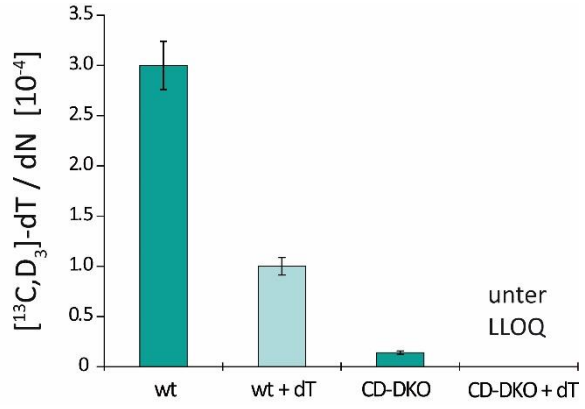


Abbildung 4.9: Absolute Quantifizierung von [^{13}C, D_3]-dT-Gehalt in Cytidin- und Desoxycytidylat-Desaminase-defizienten Zellen (CD-DKO) und Wildtypzellen (wt) mit und ohne Zugabe von dT (LLOQ = *Lower Limit of Quantification*).

Dies zeigt, dass der Großteil von [^{13}C, D_3]-dT durch folgenden Mechanismus entstehen könnte. Nach der Reparatur von genomischer DNA, die das [^{13}C, D_3]-mdC enthält, werden die resultierenden löslichen [^{13}C, D_3]-mdC / [^{13}C, D_3]-mdCMP von cytosolischen Desaminasen Cda und Dctd und TS zu [^{13}C, D_3]-mdT / [^{13}C, D_3]-mdTMP prozessiert, und nach einer Phosphorylierung in das Genom mittels DNA-Synthese wieder eingebaut (Abbildung 4.7). Zudem konnte jedoch bei einer simultanen Fütterung von [^{13}C, D_3]-Met und dT beim *Priming* von CD-DKO Zellen [^{13}C, D_3]-dT nicht mehr quantifiziert werden. Der Wert lag unter der Bestimmungsgrenze (LLOQ = *Lower Limit of Quantification*) (Abbildung 4.9). Dies würde bedeuten, dass möglicherweise nur ein Bruchteil von genomischem [^{13}C, D_3]-mdC zu [^{13}C, D_3]-dT-desaminiert. Anschließend wird die fehlgepaarte Base [^{13}C, D_3]-T ausgeschnitten und auf eine Desoxyribose mittels TP übertragen, phosphoryliert und in das Genom während der DNA-Synthese wieder eingebaut (Abbildung 4.7).

Tet-vermittelte oxidative Schäden lösen keine DNA-Reparatur aus, die für die Umwandlung von mdC zu dT verantwortlich ist.

Wie oben beschrieben ist die Umwandlung von mdC zu dT von der Tet-vermittelten Oxidation unabhängig. Allerdings könnte bereits bewiesen werden, dass die Tet Proteine dT zu hmdU oxidieren können.[115] Deswegen wurde der Einfluss von Tet Proteinen auf die Umsetzung von mdC zu dT in Tet TKO mESCs und dem entsprechenden Wildtyp durch *Priming* unter Fütterung von [^{13}C, D_3]-Met untersucht. Es konnte keinen Unterschied im [^{13}C, D_3]-dT-Gehalt zwischen beiden Zelltypen festgestellt werden (Abbildung 4.10). Dies spricht für eine Tet-unabhängige mdC zu dT Umsetzung.

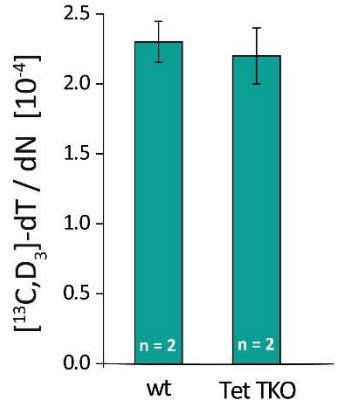


Abbildung 4.10: Absolute Quantifizierung von [^{13}C, D_3]-dT-Gehalt in wt und Tet TKO Zellen (n = Anzahl biologischer Replikate).

Aid ist nicht an der Umsetzung von genomischem mdC zu dT beteiligt

Als nächstes stellte sich die Frage, ob eine geringe genomische Desaminierung von mdC zu dT eine DNA Reparatur, die eine globale Umsetzung von mdC zu dT im löslichen Pool bewirkt, auslösen kann. Falls so eine genomische Desaminierung tatsächlich stattfindet, würde sich die Frage nach dem dafür zuständigen Enzym stellen. Wie im Abschnitt 1.2.2 beleuchtet wurde, könnte die Cytosin-Desaminase Aid diese Rolle übernehmen. Demzufolge wurden in diesem Experiment die wt- und Aid-defizienten Zellen (Aid KO) auf den Gehalt von [^{13}C, D_3]-dT, sowie auf dU-, mdC-, hmdC- und fdC-Mengen massenspektrometrisch untersucht. Unter Zugabe von [^{13}C, D_3]-Met beim *Priming* von wt und Aid KO konnte kein Unterschied im [^{13}C, D_3]-dT-, mdC- und dU-Gehalt quantifiziert werden (Abbildung 4.11). Es scheint, als würde eine genomische mdC zu dT Desaminierung, ohne die Beteiligung von Aid erfolgen. Dies schließt aber die Beteiligung anderer Desaminasen der Aid / Apobec-Familie aber nicht aus. Bei diesem Experiment fällt außerdem auf, das die globalen Werte von hmdC und fdC jeweils 2 und 2.5 Mal höher in Aid KO, im Vergleich zum entsprechenden Wildtyp, waren. Dies deutet auf eine Auswirkung auf die oxidierten mdC-Modifikationen, nicht aber auf mdC und dU.

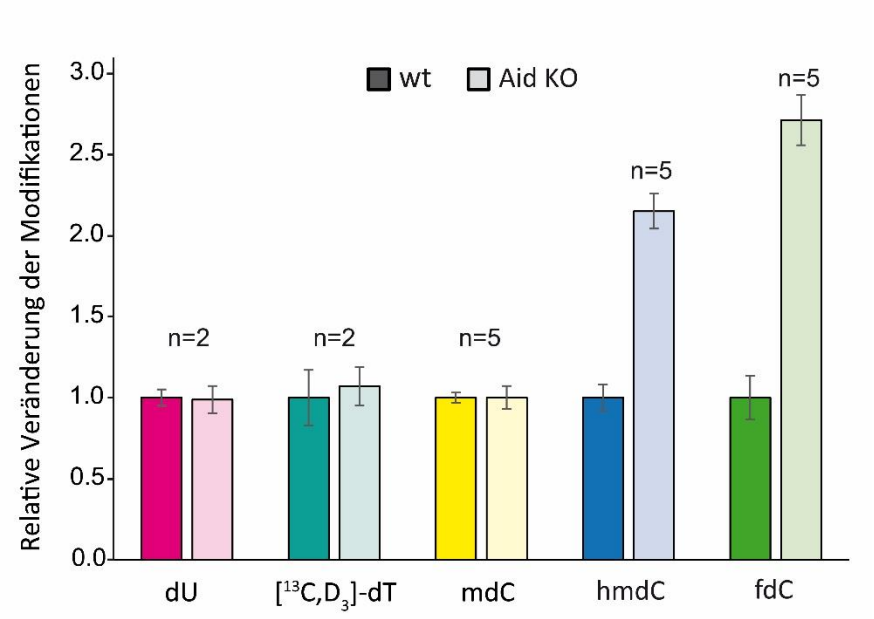


Abbildung 4.11: Darstellung der relativen Veränderung von Mengen an dU, [^{13}C, D_3]-dT, mdC, hmdC, fdC-Modifikationen im wt und Aid KO Zellen. Jeweils dunklere Farbe der Balken für jede Modifikation steht für Gehalt dieser Modifikation im wt. Jeweils hellere Farbe der Balken für jede Modifikation steht für Gehalt dieser Modifikation im Aid KO. n = Anzahl biologischer Replikate.

Die Dnmt3b ist möglicherweise an der Umsetzung von mdC zu dT beteiligt

Weitere Enzyme, die für die Desaminierung von mdC zu dT in Frage kommen, sind die *de novo* DNA-Methyltrasferasen Dnmt3a und Dnmt3b (Abschnitt 1.1.1). Zusätzlich zu den methylierenden Eigenschaften, könnten sie in die Entfernung von dC Methylierung im Genom involviert sein.[110] Außerdem wurde gezeigt, dass Dnmt3a und Dnmt3b mit Tdg und Mbd4 physisch interagieren.[175-177] Aus diesem Grund, wurde im nächsten Experiment der Einfluss von Dnmt3a und Dnmt3b auf die Umsetzung von mdC zu dT getestet. Es wurde eine geringfügige Abnahme von mdC-Mengen in Dnmt3a KO und im Dnmt3b KO beobachtet (Abbildung 4.12). Außerdem konnte in den Zellen mit Dnmt3b KO eine signifikante Abnahme der [^{13}C, D_3]-dT-Werte im Vergleich zum Wildtyp (wt) beobachtet werden. Im Vergleich dazu war die Abnahme der [^{13}C, D_3]-dT-Werte im Dnmt3a KO etwas weniger ausgeprägt (Abbildung 4.12). Die Abnahme des [^{13}C, D_3]-dT-Gehalts in beiden KOs kann mit der Abnahme des mdC-Gehalts in Verbindung gebracht werden. Dies erklärt aber nicht den großen Unterschied der [^{13}C, D_3]-dT-Werte zwischen dem Dnmt3a KO und Dnmt3b KO.

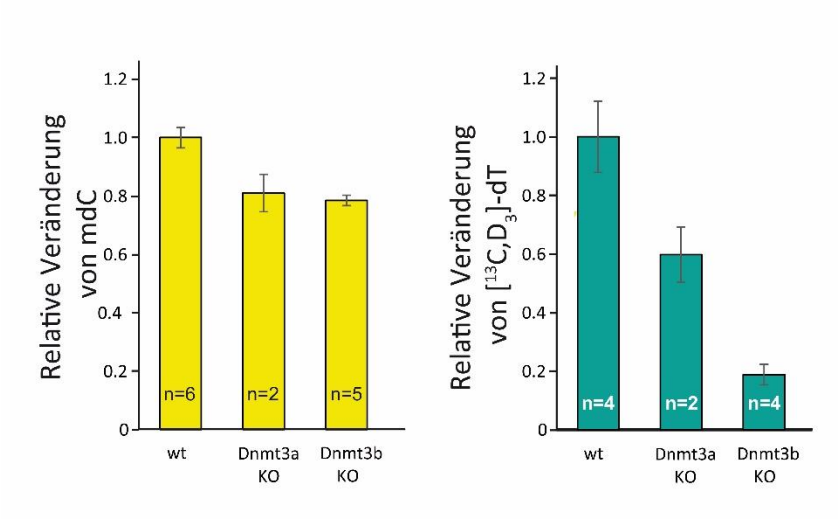


Abbildung 4.12: Relative Veränderung des mdC- und [^{13}C, D_3]-dT-Gehalts in wt, Dnmt3a KO und Dnmt3b KO.

Es scheint, als ob das Dnmt3b an der Desaminierung von mdC zu dT beteiligt ist. Um dieses Ergebnis zu verifizieren, wurden die Dnmt3b KO mESCs mit unterschiedlichen Konstrukten, die entweder zur Expression von Dnmt3b6 (G3b6) oder Dnmt3b1 (G3b1) Isoformen führen, komplementiert. Diese Isoformen wurden mit dem GFP-Protein fusioniert. Wie erwartet führte die Komplementierung mit G3b1 (G3b1-8 und G3b1-9) zu mdC-Levels auf dem wt-Niveau. Das Gegenteil wurde bei den katalytisch inaktiven Dnmt3b6 Isoformen (G3b6-4 und G3b6-8)[178] beobachtet (Abbildung 4.13). Etwas überraschend sind die [^{13}C, D_3]-dT-Werte, die in den komplementierten Dnmt3b KO Zellen auf dem gleichen Niveau, wie in den nicht komplementieren Dnmt3b KO bleiben(Abbildung 4.13).

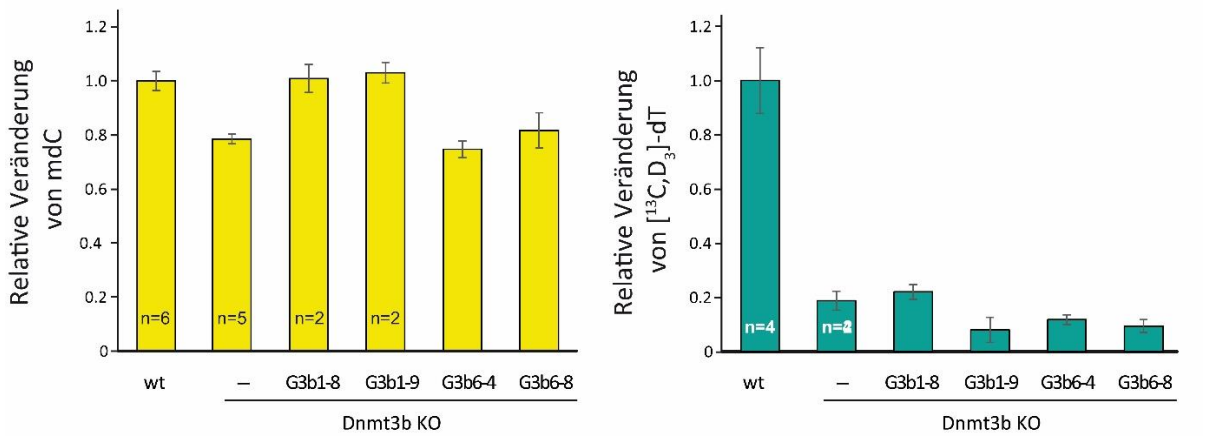


Abbildung 4.13: Relative Veränderung des mdC- und [^{13}C, D_3]-dT-Gehalts in wt und komplementierten Dnmt3b KO Zellen.

Eine mögliche Erklärung könnte sein, dass für die Methylierung von denjenigen DNA-Abschnitten, die durch die mdC zu dT Umsetzung ausgelöst wurde, Dnmt3b benötigt wird. Darüber hinaus könnte bei der längeren Abwesenheit von Dnmt3b die Methylierung an diesen Stellen verloren gehen und durch die Komplementierung nicht mehr wiederhergestellt werden. Es könnte daher sein, dass die genomische mdC zu dT Desaminierung in Folge eines Dnmt3b-assoziierten Reparaturprozesses auftritt. Das Fehlen von Methylierung in der Abwesenheit von Dnmt3b würde bedeuten, dass Dnmt1 alleine den Erhalt der Methylierung an diesen Stellen nicht ausreicht. Dies ähnelt stark dem Phänomen der genomischen Prägung (*genomic imprints*). Die unterschiedlich methylierte Regionen an den geprägten *loci* in Gameten (*gametic Differentially Methylates Regions = gDMRs*) fehlen in mESCs ohne Dnmt1 oder Dnmt3a und Dnmt3b (Dnmt3a/b DKO). Diese können mittels Komplementation mit Dnmt3a2 in Dnmt3a/b DKO mESCs, aber nicht in Dnmt1 KO mESCs, wiederhergestellt werden.[179-181] Somit sind die Dnmt3a und Dnmt3b für die Etablierung von gDMRs in der Keimbahn verantwortlich. Alle drei Dnmts sind notwendig, aber einzeln nicht ausreichend, die gDMRs aufrecht zu erhalten. Interessant an dieser Stelle ist, dass Dnmt3b spezifisch für die somatischen oder sekundären Imprints (sDMRs) verantwortlich ist.[178] Diese werden nur nach der Implantation in Embryonen angelegt und hängen von gDMRs an den jeweiligen Kontrollregionen (*Imprintig Control Regions = ICRs*) ab. [182-184] Folglich könnten die sDMRs der mdC zu dT Umsetzung unterliegen. Diese Möglichkeit wurde in mESCs untersucht, wo bestimmte Konstrukte, die entweder Dnmt3b1 oder Dnmt3b6 exprimieren können, in endogene *Dnmt3b* Allele eingebettet wurden. Diese Konstrukte waren mit GFP- oder mCherry-Proteinen fusioniert und hatten entweder homozygote ($Dnmt3b^{G1/G1}$ und $Dnmt3b^{G6/G6}$) oder heterozygote ($Dnmt3b^{Ch1/G6}$) Konfigurationen.[185] In diesen Zelllinien können nur die getaggten Dnmt3b Konstrukte exprimiert werden. Der mdC-Gehalt in diesen drei Zelllinien war dem mdC-Gehalt im Dnmt3b KO und nicht den Werten des Wildtyps ähnlich (Abbildung 4.14). Dennoch erreichten die [^{13}C, D_3]-dT Werte, beim *Priming* in Anwesenheit von [^{13}C, D_3]-Met, den Level des Wildtyps (Abbildung 4.14). Dieses Ergebnis würde bedeuten, dass ein niedriger

Expressionslevel von Dnmt3b1 oder Dnmt3b6 für mdC zu dT Umwandlung erforderlich ist. Darüber hinaus ist für diese Aufrechterhaltung keine katalytische Aktivität von Dnmt3b notwendig.

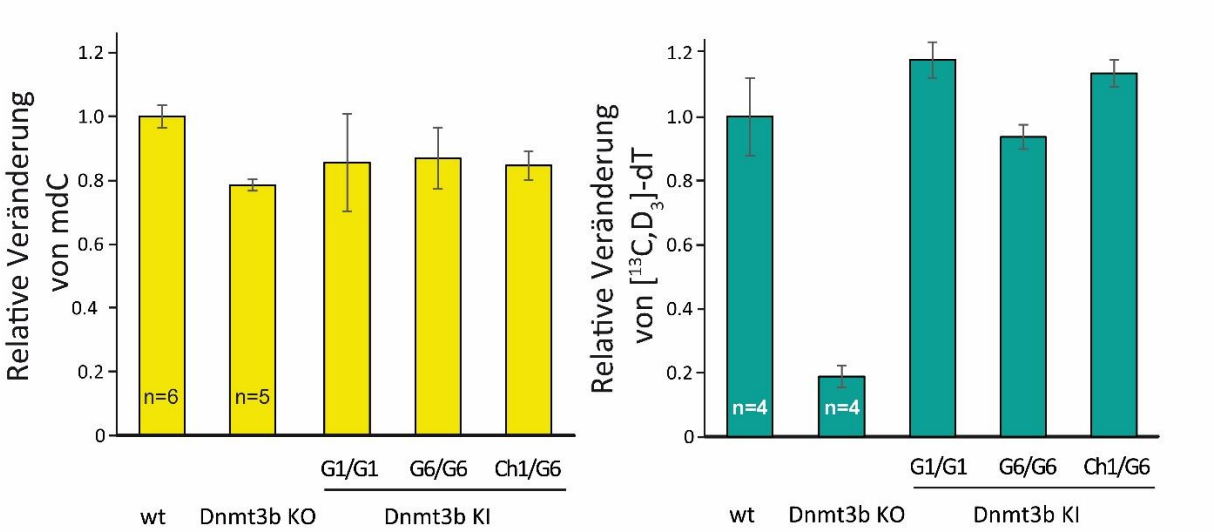


Abbildung 4.14: Relative Veränderung des mdC- und [^{13}C, D_3]-dT-Gehalts in wt, Dnmt3b KO und Dnmt3b KI.

Anschließend wurde der Einfluss von Dnmt1 auf die mdC zu dT Umsetzung untersucht. Die Dnmt1-defizienten mESCs wurden mit wt (G1wt) und katalytisch inaktiver Mutante C1229W Dnmt1 (G1cm) komplementiert[186-187]. wt und C1229W Dnmt1waren mit GFP fusioniert. Der globale mdC-Gehalt in Dnmt1 KO Zellen war erheblich reduziert im Vergleich zum wt und stieg durch die Komplementierung mit G1wt wieder an, jedoch nicht mit G1cm (Abbildung 4.15). Im Gegensatz dazu wurde das [^{13}C, D_3]-dT-Level im wt durch die Komplementierung mit G1wt nicht vollständig erreicht (Abbildung 4.15). Dieses Ergebnis ist schlüssig mit der Notwenigkeit von Dnmt3b für die Etablierung und beider Dnmt1 und Dnmt3b für die Aufrechterhaltung der Methylierung an den Stellen, die einer mdC zu dT Umsetzung unterliegen.

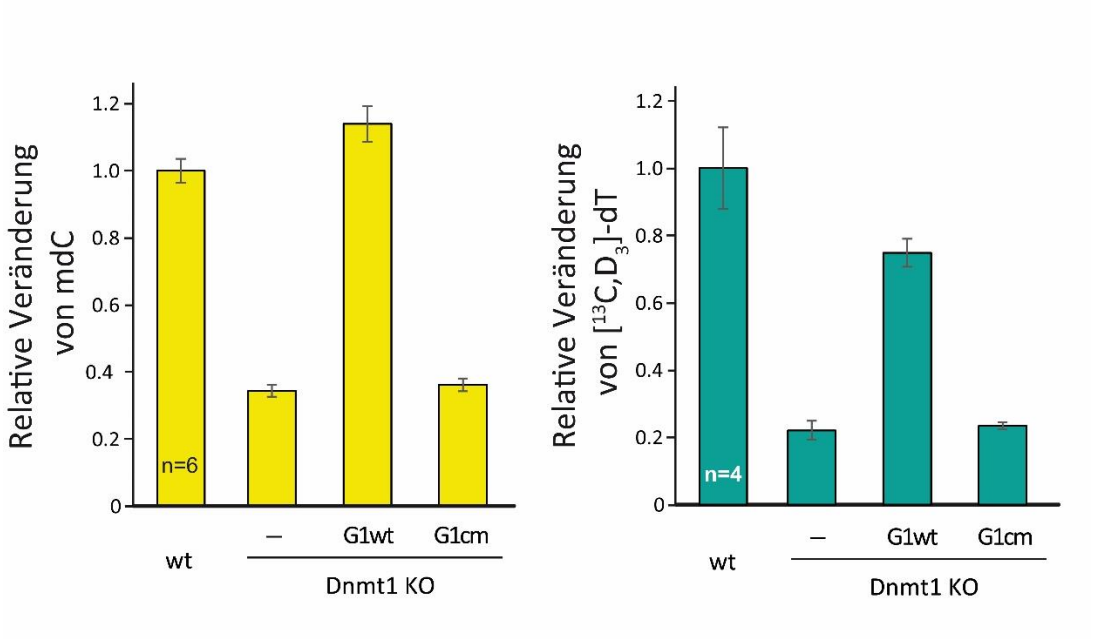


Abbildung 4.15: Relative Veränderung des mdC- und [^{13}C, D_3]-dT-Gehalts in wt, Dnmt1 KO.

Tdg und Mbd4 sind an der Umsetzung von mdC zu dT beteiligt

Bei der Untersuchung der CD-DKO Zellen konnte nur eine sehr geringe Menge an Desaminierung von genomischem mdC beobachtet werden (Abbildung 4.9). Dies zeigt, dass diese Umsetzung entweder nur sehr limitiert auftritt oder sie als Auslöser für die DNA-Reparatur dient, die mdC/mdCMP aus größeren DNA-Fragmenten freisetzt. Als nächstes sollte der Frage nachgegangen werden, welche Enzyme an dieser genomischen Desaminierung beteiligt sein könnten. Nach der Desaminierung von mdC entstandenes dT, ist mit dG fehlgepaart und kann von Tdg und Mbd4 ausgeschnitten werden.[188-190] Außerdem gibt es Hinweise, dass diese beiden Enzyme auch an der Entfernung von genomischem mC beteiligt sind.[76, 110] Zunächst wurde Tdg KO mESCs-Linie näher untersucht. Unter *Priming* in Anwesenheit von [^{13}C, D_3]-Met sinken die [^{13}C, D_3]-dT-Mengen im Vergleich zum wt sehr stark ab (Abbildung 4.16A). Dieses Ergebnis konnte 8 Mal reproduziert werden und wurde in einem weiteren unabhängigen Experiment bestätigt. Es wurden Tdg KO und Mbd4 KO mESCs an zwei Zeitpunkten während des *priming* untersucht. Die [^{13}C, D_3]-dT-Levels der beiden KO Zelllinien waren niedriger als im Tdg wt (Abbildung 4.16B).

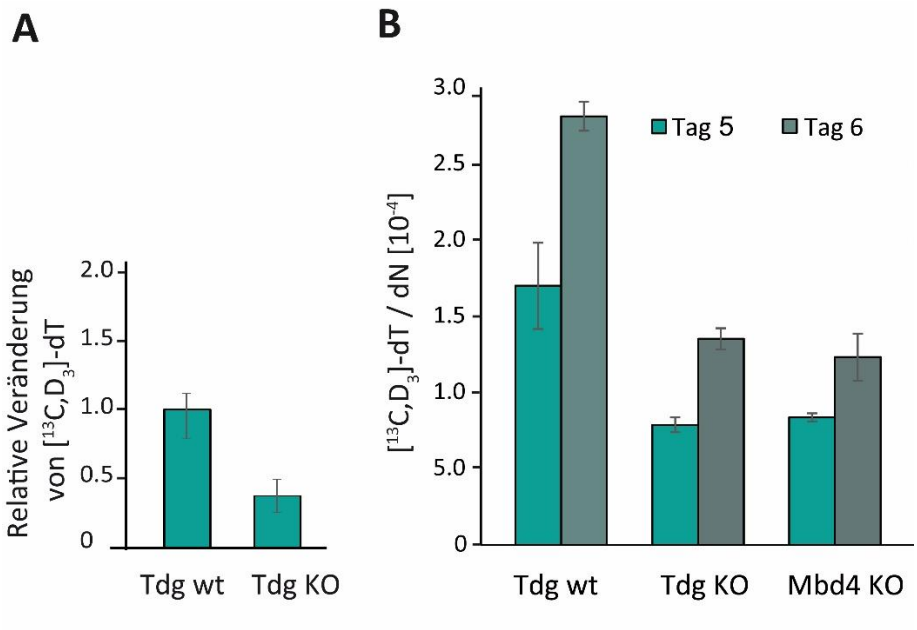


Abbildung 4.16: A Relative Veränderung von [^{13}C, D_3]-dT-Mengen in Tdg wt und Tdg KO mESCs (n = 8). **B** Absolute Quantifizierung von [^{13}C, D_3]-dT-Mengen in Tdg wt, Tdg KOund Mbd4 KO mESCs am 5. und 6. Tag des *priming* in Anwesenheit von [^{13}C, D_3]-Met.

Auf den ersten Blick scheint dieses Ergebnis überraschend zu sein. Beim Fehlen eines Enzyms, das für das Ausschneiden eines Schadens zuständig ist, würde man einen Anstieg bei den Werten für diesen Schaden erwarten. Allerdings stehen diese Ergebnisse in Übereinstimmung mit einer sehr schnellen und effizienten Reparatur von dT:dG-Fehlpaarungen, außerdem wird der größte Teil von [^{13}C, D_3]-mdC im löslichen Nukleoside-/tid Pool desaminiert. Somit könnte eine Tdg- oder Mbd4-Exzision von im Genom gebildetem fehlgepaarten [^{13}C, D_3]-dT als Auslöser für ein Reparaturprozess, der größere DNA-Fragmente ausschneidet, dienen. Somit würden die benachbarten [^{13}C, D_3]-mdC /[^{13}C, D_3]-mdCMPs in den löslichen Nukleosidpool gelangen, um dort von cytosolischen Desaminasen desaminiert werden.

MMR trägt zu Entfernung von T:G-Paarungen bei

Im Folgenden sollte der Frage nachgegangen werden, welcher Reparaturprozess für die Freisetzung größerer Menger von aus der Desaminierung von mdC stammendem dT verantwortlich ist. Ein möglicher Kandidat, der diese Aufgabe übernehmen könnte, ist die Mismatch-Reparatur (MMR). Bisher wurde angenommen, dass die Reparatur von Fehlpaarungen, die Insertion und das Entfernen von DNA-Loops, die während der Replikation entstanden, als die wichtigsten Funktionen der MMR angesehen.[191] In aktuelleren Studien wurde aber gezeigt, dass MMR auch post-replikative Funktionen, insbesondere bei der Reparatur von Fehlpaarungen, die durch die Aid-vermittelte Desaminierung zustande kommen, eine Rolle spielt.[192-195] Die Auswirkung von MMR auf die globale Umsetzung von mdC zu dT in mPPCs wurde anhand von der Mlh1 KO mESC Zelllinie, untersucht. Mlh1 ist eines von den mehreren Enzymen, die an der MMR beteiligt sind, und schneidet zusammen mit Pms2 den Schaden aus.[196-197] Die Mlh1KO Zellen wurden dem gleichen Pulse-Chase-Experiment unterzogen, das in der Abbildung 4.3A gezeigt ist. Es wurde eine langsamere Umsetzung von [^{13}C, D_3]-mdC in Mlh1 KO Zellen als in wt Zellen beobachtet (Abbildung 4.17A). Diese war mit der Umsetzung von [^{13}C, D_3]-mdC in Tet TKO Zellen vergleichbar (Abbildung 4.4). Dieses Ergebnis zeigt, dass Mlh1 an der mdC zu dT Umsetzung zumindest teilweise beteiligt ist und für die Freisetzung von löslichem [^{13}C, D_3]-mdC /[^{13}C, D_3]-mdCMP aus der DNA verantwortlich sein könnte. Deswegen stellte sich die Frage, wie sich das Fehlen der MMR auf die Akkumulation von [^{13}C, D_3]-dT in mESCs unter *Priming* in Anwesenheit von [^{13}C, D_3]-Met auswirkt. Dafür wurden neben Mlh1 auch KO Msh2 KO Zellinien untersucht. Das Msh2 Enzym ist ein weiteres MMR Enzym, das bei der Erkennung des Schadens eine Rolle spielt.[196-197] In beiden Zelllinien konnte eine Abnahme des [^{13}C, D_3]-dT-Gehaltes im Vergleich zur wt Zelllinie festgestellt werden. Erste Experimente in Msh2 KO mESCs zeigen hierbei noch geringere dT-Werte (Abbildung 4.17B). Die restlichen [^{13}C, D_3]-dT Mengen im Genom von Mlh1 und Msh2 Mutanten lassen vermuten, dass an der mdC zu dT Umsetzung noch weitere Reparaturprozesse beteiligt sein könnten.

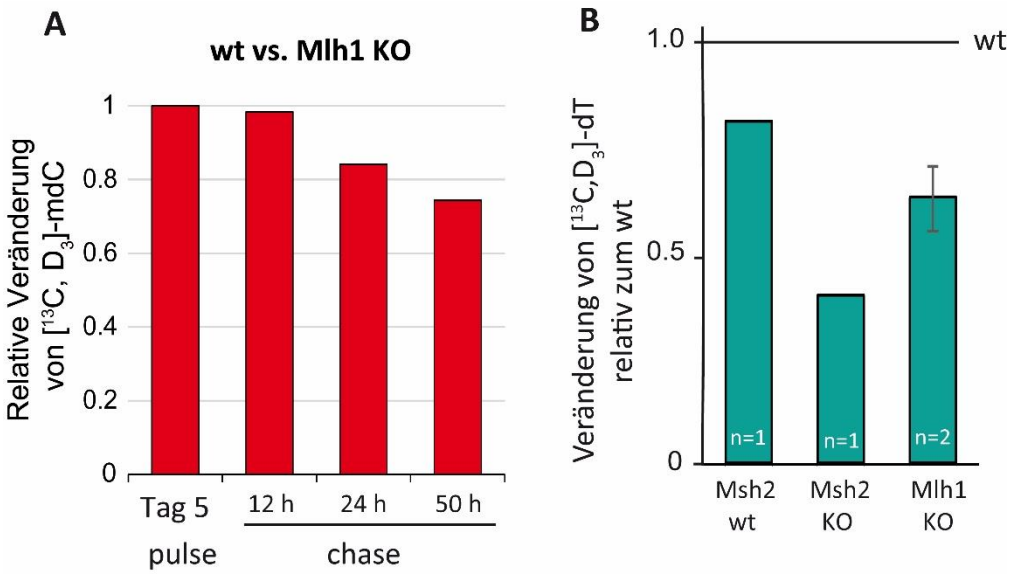


Abbildung 4.17: A Unterschied zwischen wt und Mlh1 KO in den relativen [^{13}C, D_3]-mdC-Mengen bei einem *Pulse-Chase*-Experiment. **B** Die relative Veränderung von [^{13}C, D_3]-dT-Mengen in Msh2 wt, Msh2 KO und Mlh1 KO mPPCs.

NER kann an der Entfernung von genomischem mdC beteiligt sein.

Als ein weiterer möglicher Reparaturmechanismus, der neben MMR bei der mdC Umsetzung eine Rolle spielen könnte, wurde die Nukleotidexzisionsreparatur (NER) untersucht. Die NER ist an der aktiven Entfernung von C-Methylierungen beteiligt.[81, 198-199] NER entfernt ca. 30-35 Basenpaare lange Fragmente, und könnte so zu dem Pool an löslichen [^{13}C, D_3]-mdC /[^{13}C, D_3]-mdCMP beitragen. Diese werden durch cytosolische Cda und Dctd desaminiert und in das Genom in Form von [^{13}C, D_3]-dT wiedereingebaut. In Eukaryonten kommen zwei Arten von NER vor. Die globale Genomreparatur (GG-NER) kann an jeder Stelle des Genoms stattfinden. Die transkriptionsgekoppelte Reparatur (TC-NER) ist für eine schnelle Schadensreparatur auf einem transkribierenden DNA-Strang eines aktiven Gens verantwortlich. TC-NER wird durch die RNA Polymerase II, die durch den Schaden blockiert wird, aktiviert. Dabei spielen spezifische TC-NER Faktoren Csa, Csb und Xab2 eine wichtige Rolle.[200] Nach dem der Schaden erkannt und NER eingeleitet wurde, kommt es zur Entwindung der DNA und zur Prozessierung des Schadens. Der Xpa Faktor ist an der GG-NER, als auch an der TC-NER beteiligt. Um zu ermitteln, ob und in welchem Ausmaß NER für die

mdC zu dT Umsetzung wichtig ist, wurden Xpa und Csb KO mESCs untersucht. Der [^{13}C, D_3]-dT-Level ist in beiden Mutanten ca. 3.5 höher als im wt mESCs (Abbildung 3.18). Mit der Ausschaltung des Reparaturprozesses könnte kein deaminiertes mdC in den löslichen Nukleotid/-sid Pool mehr gelangen, in einem Xpa KO bzw. Csb KO könnte dieses nicht mehr ausgeschnitten werden und somit kam es zur Akkumulation von genomischen [^{13}C, D_3]-dT. Es würde bedeuten, dass NER neben anderen Reparaturprozessen an der Entfernung von genomischen mdC beteiligt sein könnte.

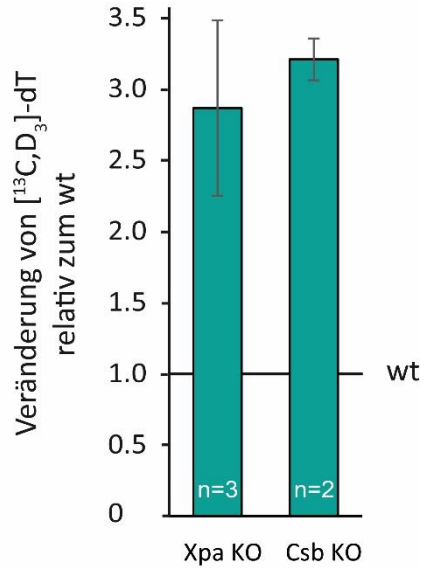


Abbildung 4.18: Relative Veränderung von [^{13}C, D_3]-dT-Mengen in Xpa KO und Csb KO mPPCs.

4.1.3 Zusammenfassung

Basierend auf einer im Arbeitskreis *Carell* etablierten massenspektrometrischen Methode wurde eine neue Methode für eine sensitive Detektion und exakte Quantifizierung von [^{13}C, D_3]-dT in genomischer DNA von embryonalen Stammzellen der Maus entwickelt. Unter Verwendung dieser neu entwickelten Methode konnte die Umsetzung von mdC zu dT in mESCs im Detail untersucht und neue Erkenntnisse gewonnen werden. Dafür wurden mESCs mit Hilfe von *priming* in die mPPCs überführt. In den mPPCs wurde eine sehr hohe Umsatzrate von genomischem mdC beobachtet, die mit der passiven „Verdünnung" während der DNA-Replikation und –Reparatur nicht zu erklären ist. Dieser Hinweis zeigt, dass ein erheblicher Teil der globalen Umsetzung von methyliertem Cytosin im Laufe eines aktiven Prozesses stattfinden muss. In Tet-defizienten mPPCs war die globale Umsetzung von mdC nur etwas niedriger im Vergleich zur Wildtyp-Zelllinie, jedoch noch höher verglichen mit der Zellproliferation. Außerdem die Umsatzrate von hmdC im Wildtyp niedriger, als diejenige von mdC. Diese Ergebnisse zeigen, dass die Oxidation von mdC nicht der einziger Mechanismus sein kann, über den die aktive Demethylierung in mPPCs abläuft. Darüber hinaus konnte durch eine massenspektrometrische Verfolgung von der [^{13}C, D_3]-mdC zu [^{13}C, D_3]-dT Umsetzung gezeigt werden, dass ein Teil der genomischen Cytosin-Methylierung durch DNA-Reparatur umgesetzt wird. Diese DNA-Reparatur erfordert eine direkte Entfernung von mdC vom Genom. Eine ausführliche Analyse von unterschiedlichen Dnmt-Linien und ihre Komplementationen wies darauf hin, dass der Reparatur-Prozess möglicherweise gezielt an spezifischen Stellen, vor Allem an sDMRs, stattfinden könnte. Das Dnmt3b könnte an der genomischen Desaminierung von mdC zu dT beteiligt sein. Im Gegensatz dazu scheint Aid an diesem Prozess nicht beteiligt zu sein, aber möglicherweise an der Desaminierung von hmdC und fdC. Ferner wurden unterschiedliche DNA-Reparaturmechanismen, die in die oxidationsunabhängige genomische mdC Umsetzung involviert sein könnten, untersucht. Die Tdg KO und Mbd4 KO Zelllinien, sowie auch Msh2 KO und Mlh1 KO Zelllinien zeigen eine Abnahme der Reparaturrate von genomischem mdC. Diese ersten Hinweise deuten auf eine Beteiligung von mehreren DNA-Reparaturmechanismen, wie MMR, BER und möglicherweise

auch NER, hin. Zusammenfassend ergibt sich ein folgendes Modell für die mdC zu dT Umsetzung. Ein enzymatischer Prozess generiert DNA Läsionen an spezifischen Stellen im Genom und löst dadurch einen Reparatur-Prozess aus, der sich auch auf benachbarte mdCs ausbreitet. Die Beteiligung von Tdg und Mbd4 weist auf Basen-Fehlpaarung hin. Eine direkte genomische Desaminierung von mdC zu dT würde dT: dG-Fehlpaarungen ergeben, die durch Tdg und Mbd4 ausgeschnitten werden können. Diese Fehlpaarung könnte neben spBER auch das Ausschneiden von größeren DNA-Fragmenten mit benachbarten mdCs, zum Beispiel durch lpBER oder MMR, auslösen. Nach anschließender Desaminierung im cytosolischen Pool wird das gebildete dT wieder in das Genom eingebaut.

4.1.4 Ausblick

Um eine exakte und parallele Quantifizierung von mdC und [^{13}C, D_3]-mdC in einer Probe zu ermöglichen, bedarf es eines neuen internen Standards, der beispielweise 6 Masseneinheiten schwerer ist als mdC. Dies würde die Notwendigkeit einer parallelen Zellkultur für eine Bestimmung von mdC, die nicht mit [^{13}C, D_3]-Met gefüttert wird, für eine Bestimmung von mdC in diesen Proben, überflüssig machen. Es wäre somit möglich in den mit [^{13}C, D_3]-Met gefütterten Zellen, parallel mdC und [^{13}C, D_3]-mdC quantifizieren.

Ein zusätzlicher Hinweis für die Desaminierung von [^{13}C, D_3]-mdC im löslichen Nukleosid/-tid Pool könnte durch einen direkten Nachweis des [^{13}C, D_3]-dT in diesem Pool erfolgen. Dazu wurde eine Methode zur Anreicherung und Extraktion von löslichen DNA Metaboliten von *Jessica Steinbacher* und mir entwickelt.

Ferner könnte der Einbau von löslichem [^{13}C, D_3]-dT durch eine Kompetition mit unmarkiertem Thymin an Stelle von unmarkiertem Thymidin unterbunden werden. Falls Thymin die Akkumulation von [^{13}C, D_3]-dT verhindert, ist es sehr wahrscheinlich, dass die Thymin Phosphorylase das [^{13}C, D_3]-T auf einen Desoxyribose überträgt, nachdem diese markierte Base von Tdg oder Mbd4 von Genom ausgeschnitten wird.

Bezüglich der Rolle von Aid bei der globalen Umsetzung von mdC zu dT könnten weitere *Pulse-Chase*-Experimente zusätzliche Hinweise auf die Beteiligung dieses Enzyms geben. Auch andere Apobec Familienmitglieder, wie Apobec1, die in den mPPCs exprimiert sind, können die Rolle der Desaminase übernehmen. Dies könnte mit einem simultanen *Knockout* in einer Aid KO Zelllinie untersucht werden. Außerdem könnte der Einfluss der Desaminase Ung2 auf die globale Umsetzung von mdC getestet werden.
Weiterhin könnte die Rolle von Tdg und Mbd4 beim Ausschneiden von [^{13}C, D_3]-dT genauer ermittelt werden. Im Tdg/Mbd4 *Doppelknockout* (DKO) würden die detektierten Mengen an [^{13}C, D_3]-dT im Vergleich zu den Tdg KO und Mbd4 KO weiter sinken. Auch hier, würde ein *Pulse-Chase*-Experiment mit dem Tdg/Mbd4 DKO, weitere Hinweise auf die Beteiligung dieser Enzyme bei der mdC zu dT Umsetzung geben.

4.1.5 Projektbeiträge

Die massenspektrometrische Methodenentwicklung, Verdauoptimierung, enzymatische Hydrolyse von Proben, Messungen sowie die Auswertungen massenspektrometrischer Daten erfolgte in enger Zusammenarbeit mit *Jessica Steinbacher*. Die Eichgerade für eine exakte Quantifizierung von [^{13}C, D_3]-dT wurde von mir aufgenommen. Enzymatischer Verdau von Proben, die mit Serum/CR für *Priming* behandelt wurden, von Proben aus allen *Pulse-Chase*-Experimenten, sowie Messungen und Auswertungen von diesen Proben wurden von mir vorgenommen. Die Arbeiten in der Zellkultur und die DNA-Isolation wurden von *Angie Kirchner* und *Fabio Spada* durchgeführt. Die Interpretation von Ergebnissen erfolgte in Zusammenarbeit mit *Fabio Spada, Angie Kirchner* und *Jessica Steinbacher*. Das Projekt wird von *Fabio Spada, Angie Kirchner* und *Sarah Schiffers* fortgeführt.

4.1.6 Materialien und Methoden

Festphasenextraktion von Nukleosiden/tiden aus dem Medium

Das Medium (3 mL) wurde auf pH = 4 mittels HCl eingestellt und auf das *Supel™-Select HLB SPE* Röhrchen (bed wt. 30 mg, *Sigma-Aldrich*) geladen. Das *Supel™-Select HLB SPE* Röhrchen wurde zuvor mit jeweils 3 mL von MeOH, H_2O und HCl (pH = 4) konditioniert. Nach dem Waschen mit 10 mL H_2O wurde das Röhrchen 1h *in vacuo* getrocknet. Die angereicherte Nukleoside/-tide aus dem Medium wurden mit 2 mL MeOH/MeCN (1:1)) eluiert, bis zur Trockne via SpeedVac eingeengt und in 50 mL H_2O aufgenommen.

Enzymatischer Verdau genomischer DNA für die UHPLC-ESI-MS/MS-Analyse

Die DNA Proben (4-7 µg in 35 µL H_2O) wurden wie folgt enzymatische verdaut. Im ersten Verdauschritt wurden jeweils 7.5 µL einer wässrigen Lösung von $ZnSO_4$ (480 µM), 5 U Antarctic phosphotase *(New England BioLabs*), 42 U nuclease S1 (*Aspergillius oryzae, Sigma-Aldrich*), 100 µM THU und einer spezifischen Menge von isotopenmarkierten Standards zu der DNA-Verdau-Lösung zugegeben und 3 h bei 37 °C in einem Thermomixer (*eppendorf comfort*) inkubiert. Für eine exakte Quantifizierung von DNA-Modifikationen wurden folgende isotopenmarkierte interne Standards zu der DNA-Verdau-Lösung zugegeben: 51.1 pmol [D_3]-mdC, 7.66 pmol [D_2,$^{15}N_2$]-hmdC, 52.0 fmol [$^{15}N_2$]-fdC, 52.0 fmol [$^{15}N_2$]-cadC, 122.0 fmol [$^{15}N_5$]-8oxodG, 160 fmol [D_2]-hmdU. [D_3]-mdC, [D_2,$^{15}N_2$]-hmdC, [$^{15}N_2$]-fdC, [$^{15}N_2$]-cadC und [D_2]-hmdU wurden im Arbeitskreis *Carell* synthetisiert. [$^{15}N_5$]8oxodG wurde von *Cambridge Isotope Laboratories* erworben. Für den zweiten Verdauschritt wurden 0.2 U snake venom phosphodiesetrase I (*Crotalus adamanteus, USB corporation*) in 7.5 µL von [Na_2]-EDTA Lösung eingesetzt und weitere 3 h bei 37 °C inkubiert. Anschließend wurden die verdauten DNA-Proben bei -20 °C gelagert. Unmittelbar vor der UHPLC-ESI-MS/MS-Analyse wurden die DNA-Proben mit AcroPrep™ Advance 96 Filterplatte (0.2 µm Supor®, *Pall Life Sciences*) für 35 min bei 4 °C 4000 rpm abfiltriert (*Eppendorf Centrifuge 5810 R*).

UHPLC-ESI-MS/MS Analysemethode

Die massenspektrometrische Analyse von enzymatisch hydrolysierten DNA-Proben erfolgte mit einem *Agilent 6490* Triple-Quadrupol Massenspektrometer, welches mit einem *Agilent 1290* UHPLC-System und einem UV-Detektor gekoppelt war. Für die Chromatographie wurde eine Poroshell SB-C_8-Säule (2.7 µm, 2.1 mm x 150 mm, *Agilent*) verwendet. Die Trennung erfolgte mittels Elutionspuffer aus Wasser und Acetonitril, die jeweils mit 0,0085% (v/v) Ameisensäure versetzt wurden. Die Flussrate betrug 0.35 mL/min bei 30 °C. Das Injektionsvolmen war 39 µL bei 4 °C. Für die Analyse genomischer DNA wurde folgender Gradient angewendet: 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. Das Eluat bis 1.5 min und nach 8.0 min wurde per Schaltventil in den Lösemittelabfall abgeleitet.

Die massenspekrometrischen Ionenquell-Parameter waren wie folgt: Gastemperatur 50 °C bzw. 80 °C, Gasflussrate 15 L/min (N_2), Vernebler 30 psi, Sheathgas-Ofen 275 °C, Sheathgas-Flussrate 11 L/min (N_2), Kapillar-Spannung 2500 V (positiver Ionenmodus) und -2500 V (negativer Ionenmodus), Düsen-Spannung 500 V, Fragmentor-Spannung 380 V, Δ EMV (*electron multiplier voltage*) 500 (positiver Modus) und 800 (negativer Modus). Die ungewöhnlich niedrige Gastemperatur wurde gewählt damit unerwünschte Fragmentierungen nicht in der Quelle schon während der Ionisation passieren. Die DNA-Nukleoside wurden sowohl im positiven als auch in negativen *Selected Reaction Monitoring Mode* (*SRM*) analysiert. In dem positiven Ionenmodus wurden $[M+H]^+$ Quasimolekül-Ionen und im negativen Ionenmodus $[M-H]^-$ Quasimolekül-Ionen detektiert. Durch die Verwendung mehrere Zeitsegmente, konnte die Messzeit, die für die jeweilige Fragmentierung benötigt wird, gesteigert werden. So konnte die höchste Sensitivität der Methode erzielt werden. Die substanzabhängigen Parameter, die bei der Methodenentwicklung die höchsten Intensitäten zeigten, sind in der Tabelle 4.1 zusammengefasst.

Tabelle 4.1: Substanzabhängige MS/MS-Parameter für die Analyse von enzymatisch hydrolysierten Proben genomischer DNA. CE: *collision energy*, CAV: *collision cell accelerator voltage*.

Substanz	Vorläufer-ion (*m/z*)	MS1 Auflösung	Produkt-ion (*m/z*)	MS2 Auflösung	*Dwell time* [ms]	CE (V)	CAV (V)	Polarität
Zeitsegment 1.5–3.6 min								
[$^{15}N_2$]-cadC	274.08	*wide*	158.03	*wide*	40	5	5	Positiv
[^{13}C]-cadC	273.09	*wide*	157.04	*wide*	40	5	5	Positiv
cadC	272.09	*wide*	156.04	*wide*	40	5	5	Positiv
[D_2,$^{15}N_2$]-hmdC	262.12	*wide*	146.07	*wide*	40	27	1	Positiv
[^{13}C, D_2]-hmdC	261.12	*wide*	145.08	*wide*	40	27	1	Positiv
hmdC	258.11	*wide*	142.06	*wide*	40	27	1	Positiv
[^{13}C, D_3]-mdC	246.13	*wide*	130.09	*wide*	30	60	1	Positiv
[D_3]-mdC	245.13	*wide*	129.09	*wide*	30	60	1	Positiv
mdC	242.11	*wide*	126.07	*wide*	30	60	1	Positiv
dC	228.12	*wide*	112.05	*wide*	40	1	3	Positiv
Zeitsegment 3.6–5.6 min								
[^{13}C, D_2]-hmdU	260.09	*wide*	217.09	*wide*	50	7	5	Negativ
hmdU	257.08	*wide*	214.07	*wide*	50	7	5	Negativ
[^{13}C, D_1]-fdU	257.07	*wide*	214.07	*wide*	50	6	5	Negativ
fdU	255.06	*wide*	212.06	*wide*	50	6	5	Negativ
[$^{15}N_2$]-dU	229.06	*wide*	185.06	*wide*	50	5	5	Negativ
dU	227.07	*wide*	184.06	*wide*	50	5	5	Negativ
Zeitsegment 5.6–8.0 min								
[$^{15}N_5$]-8oxodG	289.08	*wide*	173.04	*wide*	60	9	7	Positiv
8oxodG	284.1	*wide*	168.05	*wide*	60	9	7	Positiv
[$^{13}C_{10}$, $^{15}N_5$]-dG	283.12	*wide*	162.06	*wide*	75	45	2	Positiv
dG	268.1	*wide*	152.06	*wide*	75	45	3	Positiv
[$^{15}N_2$]-fdC	258.09	*wide*	142.06	*wide*	60	5	5	Positiv
fdC	256.09	*wide*	140.05	*wide*	60	5	5	Positiv
[$^{13}C_{10}$, $^{15}N_2$]-dT	255.12	*wide*	134.06	*wide*	60	8	5	Positiv
[^{13}C, D_3]-dT	247.12	*wide*	131.07	*wide*	60	8	5	Positiv
dT	243.1	*wide*	127.05	*wide*	60	45	3	Positiv

Eichgerade

Eine Eichgerade stellt eine Beziehung zwischen Rückmeldung vom Messgerät und einer bekannter Konzentration eines Analyts dar. Für eine exakte Quantifizierung notwendige Eichgerade wurde bei 5 unterschiedlichen Konzentrationen des synthetischen unmarkierten Standards und der immer gleichen Konzentration von markiertem Standard aufgenommen. Die lineare Regression wurde unter Verwendung von Microcal Origin 6.0 (*OriginLab*) ermittelt. Dafür wurde das Verhältnis von Integralflächen von markierten (*) und unmarkierten internal Standards (A[X]/A[X*]) gegen das Verhältnis der der Stoffmengen gleicher markierten und unmarkierten Standards (n[X]/n[X*]) aufgetragen (Abbildung 4.19).

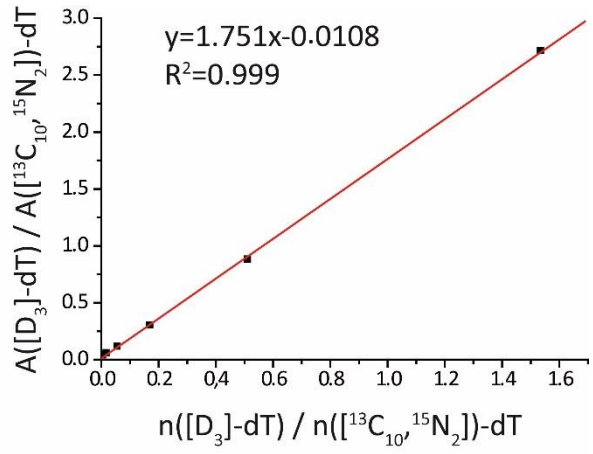


Abbildung 4.19: Ermittelte Eichgerade für die [D_3]-dT und [$^{13}C_{10}$,$^{15}N_2$]-dT.

Methodenvalidierung

Die Methodenvalidierung erfolgte auf der Grundlage der FDA (*Food and Drug Administration*) Richtlinie für die bioanalytische Methodenvalidierung aus dem Jahr 2001. Gemäß dieser Richtlinie wurde die verwendete Messmethode auf Präzision, Genauigkeit, Selektivität, Sensitivität, Reproduzierbarkeit und Stabilität geprüft.

Die Selektivität ist die Fähigkeit einer Analysemethode den Analyt von den anderen Komponenten der biologischen Matrix der zu analysierenden Probe zu unterscheiden und zu quantifizieren. Die Selektivität der Methode wurde gegen eine Blindprobe (blank) bei jeder Messung überprüft und bestätigt. Es konnten keine Interferenzen zwischen der biologischen Matrix der Probe und dem Analyt festgestellt werden. Die Blindprobe ist eine wässrige Probe ohne die zu untersuchende DNA, die sonst alle andere Bestandteile der Analyse enthält (u. a. Verdauenzyme, Puffer) und stets genauso behandelt wird, wie eine DNA-Probe.

Die Sensitivität der Messmethode wurde vor jeder Analyse durch die Messung interner markierter Standards überprüft.

Die entwickelte Messmethode sollte außerdem den Bedingungen der Genauigkeit (80-120%) und Präzision (20% RSD) erfüllen. Diese Beurteilung erfolgte mittels aufgenommener Eichgeraden, die auch für die Quantifizierung verwendet wurden.

Die Genauigkeit einer analytischen Methode beschreibt wie nah ein Mittelwert der Messergebnisse an den tatsächlichen Wert des Analyts (zum Beispiel die Konzentration) liegt. Sie wird durch das Verhältnis von theoretischen berechnetem Wert A[X]/A[X*] zu dem gemessenen A[X]/A[X*] bestimmt. Das theoretische Verhältnis von A[X]/A[X*] wird durch das einsetzten des gemessenen A[X]/A[X*] in die berechnete Eichgeradengleichung bestimmt. Die Genauigkeit der Messmethode sollte zwischen 80% und 120% liegen. Zur Beurteilung der Präzision wurde die RSD der einzelnen Eichgeradenpunkte bestimmt. Jeder Eichgeradenpunkt wurde 3 bis 5 Mal bestimmt. Und die relative Standardabweichung war nicht größer als 20%. Das galt sowohl für obere als auch für untere Bestimmungsgrenze.

Desweiterem wurde bei jeder Messung das Detektionslimit (*LOD, lower limit of detection*) bestimmt. Das Peakintegral sollte mindestens dreimal so groß sein wie die Negativkontrolle (Blindprobe oder Probe aus dem Fütterungsexperiment mit unmarkierter Verbindung).
Die Stabilität der internen Standards und Stocklösungen wurde durch das Überprüfen von Frost-Tau-, Kurzzeit- und Langzeitstabilität und postpräparative Stabilität bestätigt.

Auswertung

Für Auswertung der Messergebnisse wurde die Masshunter Quantitative Analysis Software (*Agilent*) und Excel 2013 (*Microsoft*) verwendet. Für die graphische Darstellung von Abbildungen und Diagrammen wurden Microcal Origin 6.0 (*OriginLab*) und Adobe Illustrator CS3 (*Adobe Systems*) Computerprogramme verwendet.

4.2 Untersuchungen zum Entstehungsmechanismus von 8-Oxo-dG

4.2.1 Einleitung

Das 8-Oxo-7,8-dihydro-2'-desoxyguanosin (8-Oxo-dG) ist der wichtigsten Marker für Oxidationsschäden in der DNA. Wie m Abschnitt 1.4 bereits erwähnt, kann das 8-Oxo-dG unter anderem auch durch direkte UV-Strahlung induziert werden. Es sind zwei Entstehungsmechanismen bekannt (Abbildung 4.20).

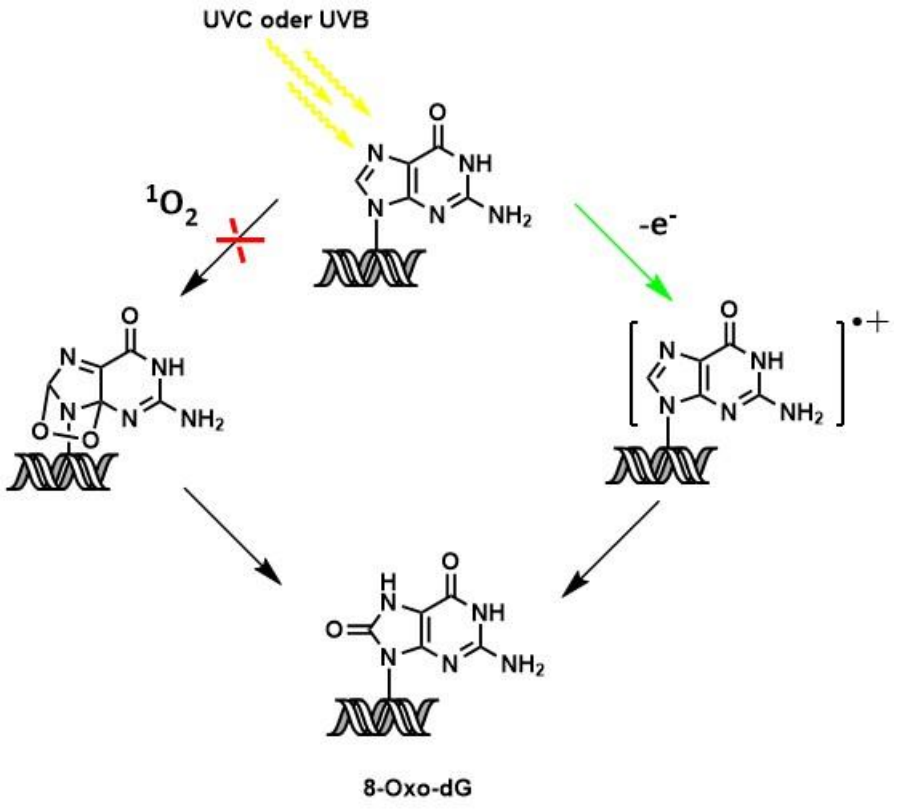


Abbildung 4.20: UV-indizierte Entstehungswege von 8-oxo-dG.

Einer der möglichen Mechanismen verläuft über eine Singulett-Sauerstoff-induzierte Bildung von 8-Oxo-dG in der doppelsträngigen DNA. Es kommt zunächst zur Bildung eines Endoperoxids, dessen Zersetzung zum 8-Oxo-dG führt. Ein weiterer Mechanismus der 8-Oxo-dG Entstehung beruht auf einer Ein-Elektron-Oxidation der DNA. Dieser Reaktionsweg verläuft über ein wichtiges Intermediat, das Guanin-Radikal-Kation ($dG^{\bullet+}$). *Gomez-Mendoza*

et al. konnten kürzlich zeigen, dass bei einer schwachen UV-Einstrahlung nur der Reaktionsweg über das Radikal-Kation möglich ist.[201] Die Frage nach dem genauen Mechanismus der Entstehung des $dG^{\bullet+}$ bleibt jedoch offen. Ein wichtiges Ziel dieses Projekts war, den gesuchten Mechanismus aufzuklären. Unser Ansatz beruhte auf ladungsgetrennten Radikalpaaren zwischen benachbarten Basen von dG.

4.2.2 Ergebnisse und Diskussion

In früheren laser-spektroskopischen Untersuchungen wurde berichtet, dass kurze Oligomere eine Ein-Photon-Ionisation im Bereich von 260 nm durchlaufen können. Die Ein-Photonen-Ionisation könnte somit als ein möglicher Mechanismus für eine direkte DNA Oxidation in Frage kommen.[202-203] Dieser Mechanismus könnte zu der Entstehung von 8-Oxo-dG führen, in dem die DNA einer UV-Belichtung bei viel niedrigerer Energie im Vergleich zum Ionisationspotential der Base ausgesetzt wird. Außerdem wurde gezeigt, dass die Ein-Photonen-Ionisation von DNA Oligonukleotiden unter Bildung von Adenin-Radikalen verläuft.[203] Im Falle der dG-Radikal-Kationen kommt auch ein Entstehungsmechanismus über ladungsgetrennte Zustände zu benachbarten Basen in Frage.[204] Dabei wird die Reaktivität des Radikalpaares durch die Lebensdauer des angeregten Zustandes und durch das Redoxpotential der Nachbarbasen bestimmt.

Die ersten Messungen wurden an einem GA-Dimer durchgeführt, da bei dieser Spezies der ladungsgetrennte Zustand eine relativ hohe Lebensdauer von 300 ps aufweist.[204-205] Im Vergleich dazu wurde ein AG-Dimer analysiert, um die Rolle des 3'- bzw. 5'- Endes zu bestimmen. Den Einfluss der Lebensdauer auf die 8-Oxo-dG Entstehung liefert ein Vergleich mit GT- und GC-Dimeren, die eine deutlich kürzere Lebensdauer aufweisen als ein GA-Dimer. Schließlich wurden auch unterschiedlich lange Oligomere analysiert, um den Einfluss der Länge auf die 8-Oxo-dG-Bildung zu bestimmen. Hierzu wurden GAGA, GAAG und ein alternierendes 20mer $(GA)_{10}$ belichtet und vermessen (Abbildung 4.21).

.

Die absorbierte Belichtungsdosis jeder Probe wurde auf die absorbierte Belichtungsdosis von dG normiert.

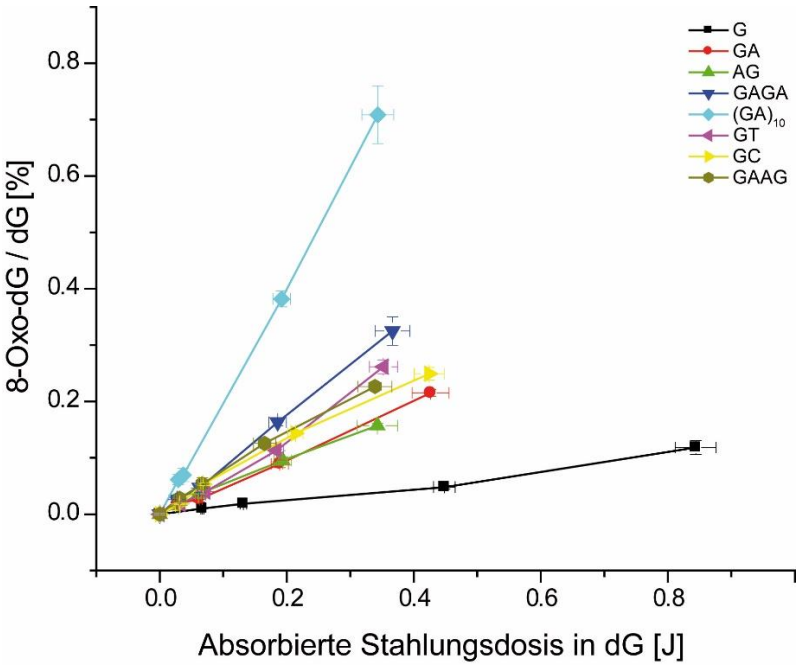


Abbildung 4.21: Relative Menge an entstandenem 8-OxoG pro dG vs. der in dG absorbierten Strahlungsdosis mit Steigungen der linearen Fits.

Als Erstes fällt auf, das bei allen Proben ein linearer Zusammenhang zwischen 8-Oxo-dG-Menge und absorbierter Belichtungsdosis besteht. Außerdem wird hier ein deutlicher Einfluss von benachbarten Basen relativ zum dG auf die Entstehung von 8-Oxo-dG beobachtet. Durch die Möglichkeit landungsgetrennte Zustände mit den Nachbarbasen auszubilden, ändert sich die detektierte 8-Oxo-dG-Menge. Diese Änderung äußert sich in unterschiedlicher Steigung für die jeweilige Sequenz des Oligonukleotids. So wird ersichtlich, dass bei allen Oligonukleotiden, bei der gleichen Absorptionsdosis in dG mehr 8-Oxo-dG entsteht als im dG-Monomer. In der Abbildung 4.21 ist außerdem deutlich zu erkennen, dass die Länge des Oligonukleotids einen nicht zu übersehenden Effekt auf die 8-Oxo-dG-Bildung hat. Wie am Vergleich zwischen GA und GAGA zu erkennen ist, führt bereits eine Kettenlänge von 4 Nukleotiden zur Erhöhung des 8-Oxo-dG-Gehalts in dieser Probe. Allerdings zeigt das Oligonukleotid mit der Sequenz GAAG eine niedrigere 8-Oxo-dG-Menge als ein Oligonukleotid mit der Sequenz GAGA. Der Grund dafür liegt möglicherweise in der Option ein ladungsgetrenntes Radikalpaar in der

jeweiligen Sequenz auszubilden. So bestehen in der Sequenz GAGA drei Möglichkeiten so ein Radikalpaar auszubilden: GA, AG und GA. Im Gegensatz dazu bestehen in der Sequenz GAAG nur zwei solche Möglichkeiten: GA und AG. Nichtdestotrotz weisen beide Tetramere eine höhere Steigung im Vergleich zu den Dimeren auf, was auf einen Einfluss der Kettenlänge auf die Bildung von 8-Oxo-dG hindeutet. Bei der UV-Belichtung von einem alternierenden 20mer $(GA)_{10}$ wird dieser Einfluss sehr deutlich. In größeren Systemen kommt es zu langlebigen angeregten Zuständen, so kann es in einem alternierenden 20mer $(GA)_{10}$ sogar mehr 8-Oxo-dG gebildet werden..[204]

.

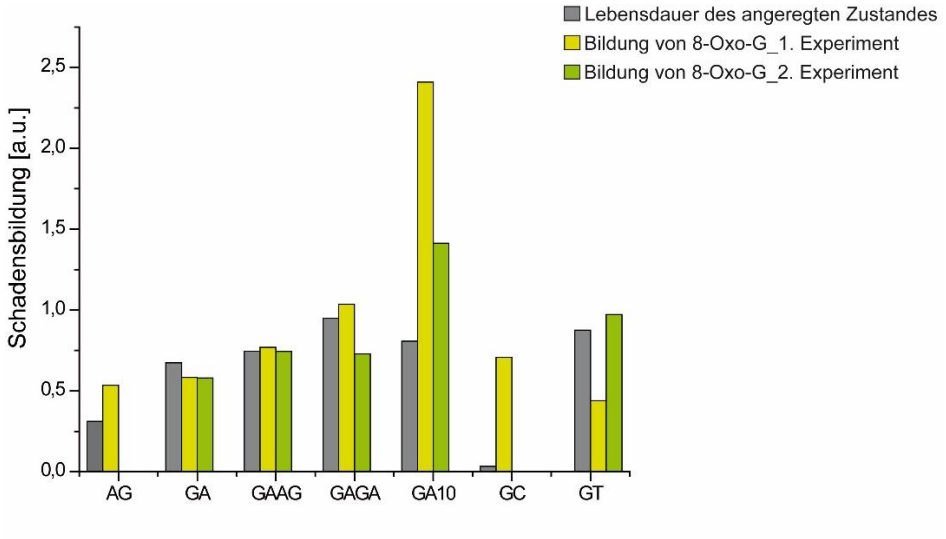


Abbildung 4.22: Lebensdauer der angeregten Zustände (graue Balken; zur Vergleichbarkeit mit Faktor 0.001 skaliert) vs. lineare Steigung der Graphen von 8-Oxo-dG-Bildung wie in Abbildung 5.2 (farbige Balken).

Um den Zusammenhang zwischen der Lebensdauer der angeregten Zustände und der Bildung von 8-Oxo-dG besser zu verdeutlichen wurden diese von allen analysierten Oligomeren in einem Balkendiagramm aufgetragen (Abbildung 4.22). Bei den Purin-Purin-Dimeren und Tetrameren wird in etwa ein direkter Zusammenhang zwischen der Bildung von 8-OxodG und der Lebensdauer der angeregten Zustände beobachtet. Im Gegensatz dazu zeigt das GC-Dimer ein abweichendes Verhalten. Eine mögliche Erklärung für diesen Unterschied liegt möglicherweise in der kurzen Lebensdauer des angeregten Zustandes von GC, so dass eine Singulett-Anregung hier unwahrscheinlich ist (Abbildung 4.20). Andererseits zeigt das GT-

Dimer eine lange Lebensdauer des angeregten Zustandes und zwei unterschiedliche Werte für die Steigung des Graphs der 8-Oxo-dG-Bildung.

4.2.3 Zusammenfassung

In diesem Teil der vorgelegten Doktorarbeit wurde gezeigt, dass möglicherweise ein Zusammenhang zwischen der bei UVC-Bestrahlung entstandenen 8-OxoG Menge und der Lebensdauer der ladungsgetrennten Zustände zwischen benachbarten Basen besteht. Als Referenz wurde dGMP gemessen, wobei sich auch hier eine kleine Menge an 8-OxoG bildet. Die Entstehung von 8-Oxo-dG bei dGMP ist jedoch deutlich kleiner als in Anwesenheit einer benachbarten Base mit der Möglichkeit zur Bildung eines ladungsgetrennten Zustandes. Für eine exaktere Verifizierung des oben beschriebenen Zusammenhangs bedarf es an weiteren Studien.

4.2.4 Projektbeiträge

Die ersten LC/MS-Messungen zu diesem Kooperationsprojekt führte *Jessica Steinbacher* durch. Die enzymatische Hydrolyse aller nachfolgenden Proben, Messungen sowie die Auswertungen massenspektrometrischer Daten führte ich durch. Die Probenvorbereitung, sowie Belichtungen erfolgte durch *Corinna L. Kufner* (AK Prof. Dr. Zinth). Die endgültige Auswertung, sowie die Interpretation der Messergebnisse erfolgte in enger Zusammenarbeit mit *Corinna L. Kufner*. Das Projekt wird von ihr und *Kerstin Kurz* fortgeführt.

4.2.5 Materialien und Methoden

Belichtung der Oligonukleotide

Alle Proben wurden bei λ = 266 nm mittels eines *EKSPLA NT 242* Nanosekunden-gepulsten Laser / parametrischer Oszillator Systems belichtet. Das System wurde bei einer Wiederholungsrate von 1 kHz und einer mittleren Leistung von 1.1 mW, gemessen von einem Strahlungs-Monitor (Nova II, *Ophir Photonics*), betrieben. Die Oligonukleotide wurden in 99.9 % gereinigtem Wasser auf 0.05 mM Basenkonzentration angelöst. Für die Belichtung wurden die angelösten Proben in die QS 10.00 UV-Küvetten (*Hellma Analytics*) überführt und während der Messung mittels eines Magnetrührers kontinuierlich durchgemischt. Die Belichtung erfolgte bei drei unterschiedlichen Zeitpunkten (2 min 52 sec, 8 min 34 sec, 17 min 7 sec).

Enzymatischer Verdau synthetischer Oligonukleotide für die UHPLC-ESI-MS/MS-Analyse

Die Oligonukleotid-Proben (4-7 µg in 35 µL H_2O) wurden wie folgt enzymatische verdaut. Im ersten Verdauschritt wurden jeweils 7.5 µL einer wässrigen Lösung von $ZnSO_4$ (480 µM), 5 U Antarctic phosphotase *(New England BioLabs*), 42 U nuclease S1 (*Aspergillius oryzae, Sigma-Aldrich*), 0.109 pmol an [$^{15}N_5$]8oxodG zu der Verdau-Lösung zugegeben und 3 h bei 37 °C in einem Thermomixer (*eppendorf comfort*) inkubiert. [$^{15}N_5$]8oxodG wurde von *Cambridge Isotope Laboratories* erworben. Für den zweiten Verdauschritt wurden 0.2 U snake venom phosphodiesetrase I (*Crotalus adamanteus, USB corporation*) in 7.5 µL von [Na_2]-EDTA Lösung eingesetzt und weitere 3 h bei 37 °C inkubiert. Anschließend wurden die verdauten Oligonukleotid-Proben bei -20 °C gelagert. Unmittelbar vor der UHPLC-ESI-MS/MS-Analyse wurden die DNA-Proben mit AcroPrep™ Advance 96 Filterplatte (0.2 µm Supor®, *Pall Life Sciences*) für 35 min bei 4 °C 4000 rpm abfiltriert (*Eppendorf Centrifuge 5810 R*).

UHPLC-ESI-MS/MS Analysemethode

Für die massenspektrometrische Analyse wurde eine im Arbeitskreis *Carell* bereits etablierte Standard-Messmethode verwendet.[115] Diese basiert auf früher veröffentlichen Arbeiten.[206-208] Die massenspektrometrische Analyse von enzymatisch hydrolysierten Oligonukleotid-Proben erfolgte mit einem *Agilent 6490* Triple-Quadrupol Massenspektrometer, welches mit einem *Agilent 1290* UHPLC-System und einem UV-Detektor gekoppelt war. Für die Chromatographie wurde eine Poroshell SB-C_8-Säule (2.7 µm, 2.1 mm x 150 mm, *Agilent*) verwendet. Die Trennung erfolgte mittels Elutionspuffer aus Wasser und Acetonitril, die jeweils mit 0,0085% (v/v) Ameisensäure versetzt wurden. Die Flussrate betrug 0.35 mL/min bei 30 °C. Das Injektionsvolumen war 39 µL bei 4 °C. Für die Analyse wurde folgender Gradient angewendet: 0 → 4 min; 0 → 3.5% (v/v) MeCN; 4 → 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. Das Eluat bis 1.5 min und nach 9.0 min wurde per Schaltventil in den Lösemittelabfall abgeleitet.

Die massenspekrometrischen Ionenquell-Parameter waren wie folgt: Gastemperatur 50 °C bzw. 80 °C, Gasflussrate 15 L/min (N_2), Vernebler 30 psi, Sheathgas-Ofen 275 °C, Sheathgas-Flussrate 11 L/min (N_2), Kapillar-Spannung 2500 V (positiver Ionenmodus) und -2500 V (negativer Ionenmodus), Düsen-Spannung 500 V, Fragmentor-Spannung 380 V, Δ EMV (*electron multiplier voltage*) 500 (positiver Modus) und 800 (negativer Modus). Die substanzabhängigen Parameter sind in der Tabelle 4.2 zusammengefasst.

Tabelle 4.2: Substanzabhängige MS/MS-Parameter für die Analyse von enzymatisch hydrolysierten Oligonikleotid-Proben. CE: *collision energy*, CAV: *collision cell accelerator voltage*.

Substanz	Vorläufer-ion (*m/z*)	MS1 Auflösung	Produkt-ion (*m/z*)	MS2 Auflösung	*Dwell time* [ms]	CE (V)	CAV (V)	Polarität
Zeitsegment 1.5–3.2 min								
[$^{15}N_2$]-cadC_1	274.08	*wide*	158.03	*wide*	60	5	5	Positiv
cadC_1	272.09	*wide*	156.04	*wide*	60	5	5	Positiv
[D_2,$^{15}N_2$]-hmdC	262.12	*wide*	146.07	*wide*	60	27	1	Positiv
hmdC	258.11	*wide*	142.06	*wide*	60	27	1	Positiv
[D_3]-mdC	245.13	*wide*	129.09	*wide*	60	60	1	Positiv
mdC	242.11	*wide*	126.07	*wide*	60	60	1	Positiv
dC	228.12	*unit*	112.05	*unit*	1	1	0	Positiv
[$^{15}N_2$]-cadC_2	158.03	*wide*	140.09	*wide*	60	13	7	Positiv
cadC_2	156.04	*wide*	138.03	*wide*	60	13	7	Positiv
Zeitsegment 3.2–4.6 min								
[D_2]-hmdU	259.09	*wide*	216.08	*wide*	48	7	5	Negativ
[D_2]-hmdU_2	259.09	*wide*	126.05	*wide*	48	7	5	Negativ
hmdU	257.08	*wide*	214.07	*wide*	48	7	5	Negativ
hmdU_2	257.08	*wide*	124.04	*wide*	48	7	5	Negativ
[$^{15}N_2$]-fdU	257.06	*wide*	213.05	*wide*	48	6	5	Negativ
[$^{15}N_2$]-fdU_2	257.06	*wide*	141.00	*wide*	48	6	5	Negativ
fdU	255.06	*wide*	212.06	*wide*	48	6	5	Negativ
fdU_2	255.06	*wide*	140.00	*wide*	48	6	5	Negativ
[$^{15}N_2$]-dU	229.06	*wide*	185.06	*wide*	48	5	5	Negativ
dU	227.07	*wide*	184.06	*wide*	48	5	5	Negativ
Zeitsegment 4.6–9.0 min								
[$^{15}N_5$]-8oxodG	289.08	*wide*	173.04	*wide*	80	9	7	Positiv
8oxodG	284.1	*wide*	168.05	*wide*	80	9	7	Positiv
[$^{13}C_{10}$, $^{15}N_5$]-dG	283.12	*wide*	162.06	*wide*	80	45	2	Positiv
[$^{15}N_2$]-fdC	258.09	*wide*	142.04	*wide*	80	5	5	Positiv
fdC	256.09	*wide*	140.05	*wide*	80	5	5	Positiv
[$^{15}N_2$]-fdC_2	142.04	*wide*	98.04	*wide*	80	13	7	Positiv
fdC_2	140.05	*wide*	97.04	*wide*	80	13	7	Positiv

Methodenvalidierung

Die Methodenvalidierung erfolgte auf der Grundlage der FDA (*Food and Drug Administration*) Richtlinie für die bioanalytische Methodenvalidierung aus dem Jahr 2001. Gemäß dieser Richtlinie wurde die verwendete Messmethode auf Präzision, Genauigkeit, Selektivität, Sensitivität, Reproduzierbarkeit und Stabilität wie im Abschnitt 3.6 geprüft.

Auswertung

Für Auswertung der Messergebnisse wurde die Masshunter Quantitative Analysis Software (*Agilent*) und Excel 2013 (*Microsoft*) verwendet. Für die graphische Darstellung von Abbildungen und Diagrammen wurden Microcal Origin 6.0 (*OriginLab*) und Adobe Illustrator CS3 (*Adobe Systems*) Computerprogramme verwendet.

5 Abkürzungsverzeichnis

5'-dRP	5'-Desoxyribose-phosphat
8-Oxo-dG	8-Oxo-2'-desoxyguanosin
β-EP	β-Eliminierungsprodukt
ADP	Adenosindiphosphat
Aid	aktivierungsinduzierte Cytidin-Desaminase
AP-Stelle	apurinische / apyrimidinische Stelle
Ape1	AP-Endonuklease 1
Apobec	*apolipoprotein B mRNA editing enzyme, catalytic polypeptide*
ARP	*Aldehyde Reactive Probe*
ATP	*Adenosintriphosphat*
BER	Basenexzisionsreparatur
caC	5-Carboxycytosin
cadC	Carboxy-2' desoxycytidin
CAV	*collision cell accelerator voltage*
Cda	Cytidin-Desaminase
CE	*collision energy*
CID	*collision-induced dissociation*
CGI	CpG-Inseln
CpG	dC-dG-Dinukleotide
DNA	Desoxyribonukleinsäure
dA	2'-Desoxyadenosin
dC	2'-Desoxycytidin
Dctd	Desoxycytidylat-Desaminase
DKO	*double knockout*

Dnmt	DNA-Methyltransferase
dT	2'-Desoxythymidin
dU	2'-Desoxyuridin
EDTA	Ethylendiamintetraessigsäure
ELISA	*enzyme-linked Immunosorbent Assay*
ES	embryonale Stammzellen
ESCODD	Europäisches Standardkomitee für oxidative Schäden
ESI	*electrospray ionisation*
fC	5-Formylcytosin
FDA	*Food and Drug Administration*
fdC	5-Formyl-2'-desoxycytidin
G	Guanin
gDMR	*gametic Differentially Methylates Region*
HEK	*human embryoid kidney*
hmC	5-Hydroxymethylcytosin
hmdC	5-Hydroxymethyl-2'-desoxycytidin
HPLC	*High Pressure Liquid Chromatography; Hochdruckflüssigkeitschromatographie*
ICR	*Imprintig Control Regions*
KO	*knockout*
LIF	*leukemia inhibitory factor*
LLOQ	*Lower Limit of Quantification*
LOD	*Lower Limit of Detection*
Mbd4	*Methyl-CpG-binding domain protein 4*
mC	5-Methylcytosin
mdC	5-Methyl-2'-desoxycytidin
mdCMP	Methyldesoxycytidin-Monophosphat

MeCN	Acetonitril
mESCs	*mouse embryonic stem cells*; embryonale Stammzelle der Maus
Met	Methionin
MeOH	Methanol
mPPC	pluripotente Stammzellen der Maus; *mouse pluripotent stem cells*
ncMMR	nicht-kanonische Mismatch-Reparatur
Neil1	Nei-like 1 Glykosylase
NER	Nukleotidexzisionsreparatur
Ogg1	8-Oxoguanin-Glykosylase
PNKP	Polynukleotid Kinase 3′-Phosphatase
Pol β	DNA-Polymerase β
ROS	*reactive oxygen species*
rp-LC	*reversed phase liquid chromatography*
SAM	S-Adenosylmethionin
Smug1	*single-strand-selective monofunctional uracil-DNA glycosylase*
T	Thymin
Tet	*ten-eleven-translocation*
Tdg	Thymin-DNA Glykosylase
THU	Tetrahydrouridin
TKO	*triple knockout*
TP	Thymin Phosphorylase
tRNA	*transfer ribonucleic acid;* transfer-Ribonukleinsäure
TS	Thymidylat-Synthase
TSG	Tumor-Supressor Gen
UHPLC	*Ultra High Pressure Liquid Chromatography;* Ultrahochdruckflüssigkeitschromatographie

Ung	Uracil-DNA-Glykosylase
UV	ultraviolett
v/v	Volumenanteil
wt	Wildtyp

6 Literaturverzeichnis

[1] W. Kuehnel, *Color Atlas of Cytology, Histology, and Microscopic Anatomy*, 4th ed., Georg Thieme Verlag, Stuttgart New York **2003**.

[2] U. Welsch, T. Deller, *Sobotta Lehrbuch Histologie*, 3. Aufl. ed., Elsevier GmbH, München, **2010**.

[3] A. D. Goldberg, C. D. Allis, E. Bernstein, *Cell* **2007**, *128*, 635-638. *Epigenetics: A Landscape Takes Shape*.

[4] R. Holliday, *Biol. Rev. Cambridge Philos. Soc.* **1990**, *65*, 431-471. *Mechanisms for the control of gene activity during development*.

[5] C. H. Waddington, *Endeavour* **1942**, *1*, 18-20. *The epigenotype*.

[6] C. H. Waddington, *Nature* **1942**, *150*, 563-656. *Canalization of development and the inheritance of acquired characters*.

[7] C. H. Waddington, *Allen & Unwin* **1957**. *The Strategy of the Genes; a Discussion of Some Aspects of Theoretical Biology*.

[8] N. Moris, C. Pina, A. M. Arias, *Nat. Rev. Genet.* **2016**, *17*, 693-703. *Transition states and cell fate decisions in epigenetic landscapes*.

[9] D. C. Kraushaar, K. Zhao, *Int. J. Biol. Sci.* **2013**, *9*, 1134-1144. *The epigenomics of embryonic stem cell differentiation*.

[10] M. Spivakov, A. G. Fisher, *Nat. Rev. Genet.* **2007**, *8*, 263-271. *Epigenetic signatures of stem-cell identity*.

[11] G. M. Keller, *Curr. Opin. Cell Biol.* **1995**, *7*, 862-869. *In vitro differentiation of embryonic stem cells*.

[12] D. A. F. Loebel, C. M. Watson, R. A. De Young, P. P. L. Tam, *Dev. Biol.* **2003**, *264*, 1-14. *Lineage choice and differentiation in mouse embryos and embryonic stem cells*.

[13] R. Eiges, N. Benvenisty, *FEBS Lett.* **2002**, *529*, 135-141. *A molecular view on pluripotent stem cells*.

[14] M. J. Weiss, S. H. Orkin, *J. Clin. Invest.* **1996**, *97*, 591-595. *In vitro differentiation of murine embryonic stem cells. New approaches to old problems*.

[15] P. Hajkova, S. Erhardt, N. Lane, T. Haaf, O. El-Maarri, W. Reik, J. Walter, M. A. Surani, *Mech. Dev.* **2002**, *117*, 15-23. *Epigenetic reprogramming in mouse primordial germ cells*.

[16] R. Jaenisch, A. Bird, *Nat. Genet.* **2003**, *33*, 245-254. *Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals*.

[17] M. A. Surani, *Nature* **2001**, *414*, 122-128. *Reprogramming of genome function through epigenetic inheritance*.

[18] T. P. Rasmussen, *Reprod. Biol. Endocrinol.* **2003**, *1*, 100-100. *Embryonic stem cell differentiation: A chromatin perspective*.

[19] D. Patterton, A. P. Wolffe, *Dev. Biol.* **1996**, *173*, 2-13. *Developmental Roles for Chromatin and Chromosomal Structure*.

[20] P. Oudet, M. Gross-Bellard, P. Chambon, *Cell* **1975**, *4*, 281-300. *Electron microscopic and biochemical evidence that chromatin structure is a repeating unit*.

[21] G. A. Bentley, A. Lewit-Bentley, J. T. Finch, A. D. Podjarny, M. Roth, *J. Mol. Biol.* **1984**, *176*, 55-75. *Crystal structure of the nucleosome core particle at 16 Å resolution*.

[22] K. Luger, A.W. Mäder, R.K. Richmond, D.F. Sargent, T. J. Richmond, *Nature* **1997**, *389*, 251-260. *Crystal structure of the nucleosome core particle at 2.8 A resolution.*

[23] R. D. Kornberg, J. O. Thonmas, *Science* **1974**, *184*, 865-868. *Chromatin Structure: Oligomers of the Histones*.

[24] H. Lodish, A. Berk, C. A. Kaiser, M. Krieger, M. P. Scott, A. Bretscher, H. Ploegh, P. Matsudiara, *Molecular cell biology*, 6 ed., W. H. Freeman, New York, **2007**.
[25] J. Allan, P. G. Hartman, C. Crane-Robinson, F. J. Aviles, *Nature* **1980**, *288*, 675-679. *The structure of histone H1 and its location in chromatin.*
[26] T. Kouzarides, *Cell* **2007**, *128*, 693-705. *Chromatin Modifications and Their Function*.
[27] M. Lawrence, S. Daujat, R. Schneider, *Trends Genet.* **2016**, *32*, 42-56. *Lateral Thinking: How Histone Modifications Regulate Gene Expression*.
[28] A. F. Kebede, R. Schneider, S. Daujat, *FEBS J.* **2015**, *282*, 1658-1674. *Novel types and sites of histone modifications emerge as players in the transcriptional regulation contest*.
[29] A. Barski, S. Cuddapah, K. Cui, T.-Y. Roh, D. E. Schones, Z. Wang, G. Wei, I. Chepelev, K. Zhao, *Cell* **2007**, *129*, 823-837. *High-Resolution Profiling of Histone Methylations in the Human Genome*.
[30] T. Jenuwein, C. D. Allis, *Science* **2001**, *293*, 1074-1080. *Translating the Histone Code*.
[31] B. D. Strahl, C. D. Allis, *Nature* **2000**, *403*, 41-45. *The language of covalent histone modifications*.
[32] M. A. Dawson, T. Kouzarides, *Cell* **2012**, *150*, 12-27. *Cancer epigenetics: from mechanism to therapy*.
[33] K. D. Robertson, *Nat. Rev. Genet.* **2005**, *6*, 597-610. *DNA methylation and human disease*.
[34] R. Lister, M. Pelizzola, R. H. Dowen, R. D. Hawkins, G. Hon, J. Tonti-Filippini, J. R. Nery, L. Lee, Z. Ye, Q.-M. Ngo, L. Edsall, J. Antosiewicz-Bourget, R. Stewart, V. Ruotti, A. H. Millar, J. A. Thomson, B. Ren, J. R. Ecker, *Nature* **2009**, *462*, 315-322. *Human DNA methylomes at base resolution show widespread epigenomic differences*.
[35] M. Ehrlich, M. A. Gama-Sosa, L. H. Huang, R. M. Midgett, K. C. Kuo, R. A. McCune, C. Gehrke, *Nucleic Acids Res.* **1982**, *10*, 2709-2721. *Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells*.
[36] D. Globisch, M. Münzel, M. Müller, S. Michalakis, M. Wagner, S. Koch, T. Brückl, M. Biel, T. Carell, *PLoS ONE* **2010**, *5*, e15367. *Tissue Distribution of 5-Hydroxymethylcytosine and Search for Active Demethylation Intermediates*.
[37] J. A. Law, S. E. Jacobsen, *Nat. Rev. Genet.* **2010**, *11*, 204-220. *Establishing, maintaining and modifying DNA methylation patterns in plants and animals*.
[38] J. U. Guo, Y. Su, J. H. Shin, J. Shin, H. Li, B. Xie, C. Zhong, S. Hu, T. Le, G. Fan, H. Zhu, Q. Chang, Y. Gao, G.-l. Ming, H. Song, *Nat. Neurosci.* **2014**, *17*, 215-222. *Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain*.
[39] R. Lister, E. A. Mukamel, J. R. Nery, M. Urich, C. A. Puddifoot, N. D. Johnson, J. Lucero, Y. Huang, A. J. Dwork, M. D. Schultz, M. Yu, J. Tonti-Filippini, H. Heyn, S. Hu, J. C. Wu, A. Rao, M. Esteller, C. He, F. G. Haghighi, T. J. Sejnowski, M. M. Behrens, J. R. Ecker, *Science* **2013**, *341*, 1237905. *Global epigenomic reconfiguration during mammalian brain development*.
[40] F. Antequera, A. Bird, *Proc. Natl. Acad. Sci. U S A* **1993**, *90*, 11995-11999. *Number of CpG Islands and Genes in Human and Mouse*.
[41] A. M. Deaton, A. Bird, *Genes. Dev.* **2011**, *25*, 1010-1022. *CpG islands and the regulation of transcription*.
[42] Z. D. Smith, A. Meissner, *Nat. Rev. Genet.* **2013**, *14*, 204-220. *DNA methylation: roles in mammalian development*.
[43] P. A. Jones, *Nat. Rev. Genet.* **2012**, *13*, 484-492. *Functions of DNA methylation: islands, start sites, gene bodies and beyond*.

[44] J.-K. Zhu, *Ann. Rev. Genet.* **2009**, *43*, 143-166. *Active DNA Demethylation Mediated by DNA Glycosylases*.
[45] S. Gopalakrishnan, B. O. V. Emburgh, K. D. Robertson, *Mutat. Res.* **2008**, *647*, 30-38. *DNA methylation in development and human disease*.
[46] B. Jin, B. Yao, J.-L. Li, C. R. Fields, A. L. Delmas, C. Liu, K. D. Robertson, *Cancer Res.* **2009**, *69*, 7412-7421. *DNMT1 and DNMT3B Modulate Distinct Polycomb-Mediated Histone Modifications in Colon Cancer*.
[47] B. Jin, Q. Tao, J. Peng, H. M. Soo, W. Wu, J. Ying, C. R. Fields, A. L. Delmas, X. Liu, J. Qiu, K. D. Robertson, *Hum. Mol. Genet.* **2008**, *17*, 690-709. *DNA methyltransferase 3B (DNMT3B) mutations in ICF syndrome lead to altered epigenetic modifications and aberrant expression of genes regulating development, neurogenesis and immune function*.
[48] T. Bestor, A. Laudano, R. Mattaliano, V. Ingram, *J. Mol. Biol.* **1988**, *203*, 971-983. *Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells*.
[49] X. Cheng, R. M. Blumenthal, *Structure (London, England : 1993)* **2008**, *16*, 341-350. *Mammalian DNA Methyltransferases: A Structural Perspective*.
[50] J. A. Yoder, T. H. Bestor, *Hum. Mol. Genet.* **1998**, *7*, 279-284. *A candidate mammalian DNA methyltransferase related to pmt1p of fission yeast*.
[51] M. Okano, S. Xie, E. Li, *Nat. Genet.* **1998**, *19*, 219-220. *Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases*.
[52] T. H. Bestor, *Hum. Mol. Genet.* **2000**, *9*, 2395-2402. *The DNA methyltransferases of mammals*.
[53] E. Li, *Nat. Rev. Genet.* **2002**, *3*, 662-673. *Chromatin modification and epigenetic reprogramming in mammalian development*.
[54] E. Li, T. H. Bestor, R. Jaenisch, *Cell*, *69*, 915-926. *Targeted mutation of the DNA methyltransferase gene results in embryonic lethality*.
[55] 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, *Science* **2006**, *311*, 395-398. *Methylation of tRNAAsp by the DNA Methyltransferase Homolog Dnmt2*.
[56] M. Okano, D. W. Bell, D. A. Haber, E. Li, *Cell* **1999**, *99*, 247-257. *DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development*.
[57] M. S. Kareta, Z. M. Botello, J. J. Ennis, C. Chou, F. Chédin, *J. Biol. Chem.* **2006**, *281*, 25893-25902. *Reconstitution and Mechanism of the Stimulation of de Novo Methylation by Human DNMT3L*.
[58] G. Egger, S. Jeong, S. G. Escobar, C. C. Cortez, T. W. H. Li, Y. Saito, C. B. Yoo, P. A. Jones, G. Liang, *Proc. Natl. Acad. Sci. U S A* **2006**, *103*, 14080-14085. *Identification of DNMT1 (DNA methyltransferase 1) hypomorphs in somatic knockouts suggests an essential role for DNMT1 in cell survival*.
[59] A. D. Riggs, Z. Xiong, *Proc. Natl. Acad. Sci. U S A* **2004**, *101*, 4-5. *Methylation and epigenetic fidelity*.
[60] M. M. Suzuki, A. Bird, *Nat. Rev. Genet.* **2008**, *9*, 465-476. *DNA methylation landscapes: provocative insights from epigenomics*.
[61] P. A. Wade, *Bioessays* **2001**, *23*, 1131-1137. *Methyl CpG-binding proteins and transcriptional repression*.
[62] S. K. T. Ooi, T. H. Bestor, *Cell* **2008**, *133*, 1145-1148. *The Colorful History of Active DNA Demethylation*.
[63] R. Ohno, M. Nakayama, C. Naruse, N. Okashita, O. Takano, M. Tachibana, M. Asano, M. Saitou, Y. Seki, *Development* **2013**, *140*, 2892-2903. *A replication-dependent passive mechanism modulates DNA demethylation in mouse primordial germ cells*.

[64] S. Kagiwada, K. Kurimoto, T. Hirota, M. Yamaji, M. Saitou, *EMBO J.* **2013**, *32*, 340-353. *Replication-coupled passive DNA demethylation for the erasure of genome imprints in mice*.
[65] S. C. Wu, Y. Zhang, *Nat. Rev. Mol. Cell Biol.* **2010**, *11*, 607-620. *Active DNA demethylation: many roads lead to Rome*.
[66] W. Mayer, A. Niveleau, J. Walter, R. Fundele, T. Haaf, *Nature* **2000**, *403*, 501-502. *Embryogenesis: Demethylation of the zygotic paternal genome*.
[67] T.-P. Gu, F. Guo, H. Yang, H.-P. Wu, G.-F. Xu, W. Liu, Z.-G. Xie, L. Shi, X. He, S.-g. Jin, K. Iqbal, Y. G. Shi, Z. Deng, P. E. Szabo, G. P. Pfeifer, J. Li, G.-L. Xu, *Nature* **2011**, *477*, 606-610. *The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes*.
[68] K. Iqbal, S.-G. Jin, G. P. Pfeifer, P. E. Szabó, *Proc. Natl. Acad. Sci.* **2011**, *108*, 3642-3647. *Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine*.
[69] D. K. Ma, M.-H. Jang, J. U. Guo, Y. Kitabatake, M.-l. Chang, N. Pow-anpongkul, R. A. Flavell, B. Lu, G.-l. Ming, H. Song, *Science* **2009**, *323*, 1074-1077. *Neuronal Activity–Induced Gadd45b Promotes Epigenetic DNA Demethylation and Adult Neurogenesis*.
[70] Junjie U. Guo, Y. Su, C. Zhong, G.-l. Ming, H. Song, *Cell* **2011**, *145*, 423-434. *Hydroxylation of 5-Methylcytosine by TET1 Promotes Active DNA Demethylation in the Adult Brain*.
[71] W. A. Pastor, L. Aravind, A. Rao, *Nat. Rev. Mol. Cell Biol.* **2013**, *14*, 341-356. *TETonic shift: biological roles of TET proteins in DNA demethylation and transcription*.
[72] M. Gehring, W. Reik, S. Henikoff, *Trends Genet.* **2009**, *25*, 82-90. *DNA demethylation by DNA repair*.
[73] J.-H. Xue, G.-F. Xu, T.-P. Gu, G.-D. Chen, B.-B. Han, Z.-M. Xu, M. Bjørås, H. E. Krokan, G.-L. Xu, Y.-R. Du, *J. Biol. Chem.* **2016**, *291*, 731-738. *Uracil-DNA Glycosylase UNG Promotes Tet-mediated DNA Demethylation*.
[74] U. Müller, C. Bauer, M. Siegl, A. Rottach, H. Leonhardt, *Nucleic Acids Res.* **2014**, *42*, 8592-8604. *TET-mediated oxidation of methylcytosine causes TDG or NEIL glycosylase dependent gene reactivation*.
[75] A. Maiti, A. C. Drohat, *J. Biol. Chem.* **2011**, *286*, 35334-35338. *Thymine DNA Glycosylase Can Rapidly Excise 5-Formylcytosine and 5-Carboxylcytosine: Potential Implications for Active Demethylation of CpG Sites*.
[76] S. Cortellino, J. Xu, M. Sannai, R. Moore, E. Caretti, A. Cigliano, M. Le Coz, K. Devarajan, A. Wessels, D. Soprano, Lara K. Abramowitz, Marisa S. Bartolomei, F. Rambow, Maria R. Bassi, T. Bruno, M. Fanciulli, C. Renner, Andres J. Klein-Szanto, Y. Matsumoto, D. Kobi, I. Davidson, C. Alberti, L. Larue, A. Bellacosa, *Cell* **2011**, *146*, 67-79. *Thymine DNA Glycosylase Is Essential for Active DNA Demethylation by Linked Deamination-Base Excision Repair*.
[77] H. E. Krokan, M. Bjoras, *Cold Spring Harb. Perspect. Biol.* **2013**, *5*, 1-22. *Base excision repair*.
[78] A. L. Jacobs, P. Schar, *Chromosoma* **2012**, *121*, 1-20. *DNA glycosylases: in DNA repair and beyond*.
[79] D. Svilar, E. M. Goellner, K. H. Almeida, R. W. Sobol, *Antioxid. Redox Signal.* **2011**, *14*, 2491-2507. *Base excision repair and lesion-dependent subpathways for repair of oxidative DNA damage*.
[80] C. Niehrs, A. Schäfer, *Trends Cell Biol.* **2012**, *22*, 220-227. *Active DNA demethylation by Gadd45 and DNA repair*.

[81] G. Barreto, A. Schafer, J. Marhold, D. Stach, S. K. Swaminathan, V. Handa, G. Doderlein, N. Maltry, W. Wu, F. Lyko, C. Niehrs, *Nature* **2007**, *445*, 671-675. *Gadd45a promotes epigenetic gene activation by repair-mediated DNA demethylation*.

[82] I. Grin, A. A. Ishchenko, *Nucleic Acids Res.* **2016**, *44*, 3713-3727. *An interplay of the base excision repair and mismatch repair pathways in active DNA demethylation*.

[83] K. Iwan, R. Rahimoff, A. Kirchner, F. Spada, A. S. Schroder, O. Kosmatchev, S. Ferizaj, J. Steinbacher, E. Parsa, M. Muller, T. Carell, *Nat. Chem. Biol.* **2018**, *14*, 72-78. *5-Formylcytosine to cytosine conversion by C-C bond cleavage in vivo*.

[84] 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. *Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA*.

[85] S. Ito, A. C. D'Alessio, O. V. Taranova, K. Hong, L. C. Sowers, Y. Zhang, *Nature* **2010**, *466*, 1129-1133. *Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification*.

[86] S. Ito, L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He, Y. Zhang, *Science* **2011**, *333*, 1300-1303. *Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine*.

[87] T. Pfaffeneder, B. Hackner, M. Truss, M. Munzel, M. Muller, C. A. Deiml, C. Hagemeier, T. Carell, *Angew. Chem. Int. Ed. Engl.* **2011**, *50*, 7008-7012. *The discovery of 5-formylcytosine in embryonic stem cell DNA*.

[88] 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. *Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1*.

[89] A. R. Weber, C. Krawczyk, A. B. Robertson, A. Kusnierczyk, C. B. Vagbo, D. Schuermann, A. Klungland, P. Schar, *Nat. Commun.* **2016**, *7*, 1-13. *Biochemical reconstitution of TET1-TDG-BER-dependent active DNA demethylation reveals a highly coordinated mechanism*.

[90] S. Kriaucionis, N. Heintz, *Science* **2009**, *324*, 929-930. *The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain*.

[91] X. Wu, Y. Zhang, *Nat. Rev. Genet.* **2017**. *TET-mediated active DNA demethylation: mechanism, function and beyond*.

[92] M. M. Dawlaty, A. Breiling, T. Le, M. I. Barrasa, G. Raddatz, Q. Gao, B. E. Powell, A. W. Cheng, K. F. Faull, F. Lyko, R. Jaenisch, *Dev. Cell* **2014**, *29*, 102-111. *Loss of Tet enzymes compromises proper differentiation of embryonic stem cells*.

[93] M. Munzel, D. Globisch, T. Carell, *Angew. Chem. Int. Ed. Engl.* **2011**, *50*, 6460-6468. *5-Hydroxymethylcytosine, the sixth base of the genome*.

[94] M. Wagner, J. Steinbacher, T. F. Kraus, S. Michalakis, B. Hackner, T. Pfaffeneder, A. Perera, M. Muller, A. Giese, H. A. Kretzschmar, T. Carell, *Angew. Chem. Int. Ed. Engl.* **2015**, *54*, 12511-12514. *Age-dependent levels of 5-methyl-, 5-hydroxymethyl-, and 5-formylcytosine in human and mouse brain tissues*.

[95] M. Münzel, D. Globisch, T. Brückl, M. Wagner, V. Welzmiller, S. Michalakis, M. Müller, M. Biel, T. Carell, *Angew. Chem. Int. Ed.* **2010**, *49*, 5375-5377. *Quantification of the sixth DNA base hydroxymethylcytosine in the brain*.

[96] C. G. Spruijt, F. Gnerlich, A. H. Smits, T. Pfaffeneder, P. W. Jansen, C. Bauer, M. Munzel, M. Wagner, M. Muller, F. Khan, H. C. Eberl, A. Mensinga, A. B. Brinkman, K. Lephikov, U. Muller, J. Walter, R. Boelens, H. van Ingen, H. Leonhardt, T. Carell, M. Vermeulen, *Cell* **2013**, *152*, 1146-1159. *Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives*.

[97] C.-X. Song, C. He, *Trends Biochem. Sci.* **2013**, *38*, 480-484. *Potential functional roles of DNA demethylation intermediates*.
[98] H. Hashimoto, X. Zhang, X. Cheng, *DNA Repair (Amst)* **2013**, *12*, 535-540. *Activity and crystal structure of human thymine DNA glycosylase mutant N140A with 5-carboxylcytosine DNA at low pH*.
[99] D. Cortazar, C. Kunz, J. Selfridge, T. Lettieri, Y. Saito, E. MacDougall, A. Wirz, D. Schuermann, A. L. Jacobs, F. Siegrist, R. Steinacher, J. Jiricny, A. Bird, P. Schar, *Nature* **2011**, *470*, 419-423. *Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability*.
[100] A. Bellacosa, A. C. Drohat, *DNA Repair (Amst)* **2015**, *32*, 33-42. *Role of base excision repair in maintaining the genetic and epigenetic integrity of CpG sites*.
[101] S. Morera, I. Grin, A. Vigouroux, S. Couve, V. Henriot, M. Saparbaev, A. A. Ishchenko, *Nucleic Acids Res.* **2012**, *40*, 9917-9926. *Biochemical and structural characterization of the glycosylase domain of MBD4 bound to thymine and 5-hydroxymethyuracil-containing DNA.*
[102] F. Guo, X. Li, D. Liang, T. Li, P. Zhu, H. Guo, X. Wu, L. Wen, T. P. Gu, B. Hu, C. P. Walsh, J. Li, F. Tang, G. L. Xu, *Cell Stem Cell* **2014**, *15*, 447-458. *Active and passive demethylation of male and female pronuclear DNA in the mammalian zygote*.
[103] L. Schomacher, D. Han, M. U. Musheev, K. Arab, S. Kienhofer, A. von Seggern, C. Niehrs, *Nat. Struct. Mol. Biol.* **2016**, *23*, 116-124. *Neil DNA glycosylases promote substrate turnover by Tdg during DNA demethylation*.
[104] V. Vartanian, B. Lowell, I. G. Minko, T. G. Wood, J. D. Ceci, S. George, S. W. Ballinger, C. L. Corless, A. K. McCullough, R. S. Lloyd, *Proc. Natl. Acad. Sci. U S A* **2006**, *103*, 1864-1869. *The metabolic syndrome resulting from a knockout of the NEIL1 DNA glycosylase*.
[105] A. Chakraborty, M. Wakamiya, T. Venkova-Canova, R. K. Pandita, L. Aguilera-Aguirre, A. H. Sarker, D. K. Singh, K. Hosoki, T. G. Wood, G. Sharma, V. Cardenas, P. S. Sarkar, S. Sur, T. K. Pandita, I. Boldogh, T. K. Hazra, *J. Biol. Chem.* **2015**, *290*, 24636-24648. *Neil2-null Mice Accumulate Oxidized DNA Bases in the Transcriptionally Active Sequences of the Genome and Are Susceptible to Innate Inflammation*.
[106] A. Slyvka, K. Mierzejewska, M. Bochtler, *Sci. Rep.* **2017**, *7*, 9001. *Nei-like 1 (NEIL1) excises 5-carboxylcytosine directly and stimulates TDG-mediated 5-formyl and 5-carboxylcytosine excision*.
[107] F. Santos, J. Peat, H. Burgess, C. Rada, W. Reik, W. Dean, *Epigenetics Chromatin* **2013**, *6*, 1-12. *Active demethylation in mouse zygotes involves cytosine deamination and base excision repair*.
[108] H. Nilsen, I. Rosewell, P. Robins, C. F. Skjelbred, S. Andersen, G. Slupphaug, G. Daly, H. E. Krokan, T. Lindahl, D. E. Barnes, *Mol. Cell* **2000**, *5*, 1059-1065. *Uracil-DNA glycosylase (UNG)-deficient mice reveal a primary role of the enzyme during DNA replication*.
[109] H. D. Morgan, W. Dean, H. A. Coker, W. Reik, S. K. Petersen-Mahrt, *J. Biol. Chem.* **2004**, *279*, 52353-52360. *Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: implications for epigenetic reprogramming*.
[110] R. Metivier, R. Gallais, C. Tiffoche, C. Le Peron, R. Z. Jurkowska, R. P. Carmouche, D. Ibberson, P. Barath, F. Demay, G. Reid, V. Benes, A. Jeltsch, F. Gannon, G. Salbert, *Nature* **2008**, *452*, 45-50. *Cyclical DNA methylation of a transcriptionally active promoter*.

[111] K. Rai, I. J. Huggins, S. R. James, A. R. Karpf, D. A. Jones, B. R. Cairns, *Cell* **2008**, *135*, 1201-1212. *DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45*.
[112] M. S. Kim, T. Kondo, I. Takada, M. Y. Youn, Y. Yamamoto, S. Takahashi, T. Matsumoto, S. Fujiyama, Y. Shirode, I. Yamaoka, H. Kitagawa, K. Takeyama, H. Shibuya, F. Ohtake, S. Kato, *Nature* **2009**, *461*, 1007-1012. *DNA demethylation in hormone-induced transcriptional derepression*.
[113] C. Popp, W. Dean, S. Feng, S. J. Cokus, S. Andrews, M. Pellegrini, S. E. Jacobsen, W. Reik, *Nature* **2010**, *463*, 1101-1105. *Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency*.
[114] N. Bhutani, J. J. Brady, M. Damian, A. Sacco, S. Y. Corbel, H. M. Blau, *Nature* **2010**, *463*, 1042-1047. *Reprogramming towards pluripotency requires AID-dependent DNA demethylation*.
[115] T. Pfaffeneder, F. Spada, M. Wagner, C. Brandmayr, S. 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, *Nat. Chem. Biol.* **2014**, *10*, 574–581. *Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA*.
[116] B. K. Duncan, J. H. Miller, *Nature* **1980**, *287*, 560-561. *Mutagenic deamination of cytosine residues in DNA*.
[117] C. S. Nabel, S. A. Manning, R. M. Kohli, *ACS Chem. Biol.* **2012**, *7*, 20-30. *The curious chemical biology of cytosine: deamination, methylation, and oxidation as modulators of genomic potential*.
[118] G. Rangam, K. M. Schmitz, A. J. Cobb, S. K. Petersen-Mahrt, *PLoS One* **2012**, *7*, e43279. *AID enzymatic activity is inversely proportional to the size of cytosine C5 orbital cloud*.
[119] C. S. Nabel, H. Jia, Y. Ye, L. Shen, H. L. Goldschmidt, J. T. Stivers, Y. Zhang, R. M. Kohli, *Nat. Chem. Biol.* **2012**, *8*, 751-758. *AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation*.
[120] S. G. Conticello, *Genome Biol.* **2008**, *9*, 229. *The AID/APOBEC family of nucleic acid mutators*.
[121] J. C. Shen, W. M. Rideout, 3rd, P. A. Jones, *Cell* **1992**, *71*, 1073-1080. *High frequency mutagenesis by a DNA methyltransferase*.
[122] H. E. Krokan, F. Drablos, G. Slupphaug, *Oncogene* **2002**, *21*, 8935-8948. *Uracil in DNA--occurrence, consequences and repair*.
[123] I. Grin, A. A. Ishchenko, *Nucleic Acids Res* **2016**, *44*, 3713-3727. *An interplay of the base excision repair and mismatch repair pathways in active DNA demethylation*.
[124] T. Lindahl, A. Andersson, *Biochemistry* **1972**, *11*, 3618-3623. *Rate of chain breakage at apurinic sites in double-stranded deoxyribonucleic acid*.
[125] T. Lindahl, B. Nyberg, *Biochemistry* **1972**, *11*, 3610-3618. *Rate of depurination of native deoxyribonucleic acid*.
[126] T. Lindahl, O. Karlstrom, *Biochemistry* **1973**, *12*, 5151-5154. *Heat-induced depyrimidination of deoxyribonucleic acid in neutral solution*.
[127] A. P. Breen, J. A. Murphy, *Free Radic. Biol. Med.* **1995**, *18*, 1033-1077. *Reactions of oxyl radicals with DNA*.
[128] K. Kubo, H. Ide, S. S. Wallace, Y. W. Kow, *Biochemistry* **1992**, *31*, 3703-3708. *A novel, sensitive, and specific assay for abasic sites, the most commonly produced DNA lesion*.
[129] M. Liuzzi, M. Talpaert-Borle, *Int. J. Radiat. Biol.* **1988**, *54*, 709-722. *Characterization of damage in gamma-irradiated and OsO4-treated DNA using methoxyamine*.

[130] X. Zhou, R. G. Liberman, P. L. Skipper, Y. Margolin, S. R. Tannenbaum, P. C. Dedon, *Anal. Biochem.* **2005**, *343*, 84-92. *Quantification of DNA strand breaks and abasic sites by oxime derivatization and accelerator mass spectrometry: application to gamma-radiation and peroxynitrite*.
[131] H. E. Krokan, R. Standal, G. Slupphaug, *Biochem. J.* **1997**, *325*, 1-16. *DNA glycosylases in the base excision repair of DNA*.
[132] P. Fortini, E. Dogliotti, *DNA Repair* **2007**, *6*, 398-409. *Base damage and single-strand break repair: Mechanisms and functional significance of short- and long-patch repair subpathways*.
[133] S. H. Feng, S. E. Jacobsen, W. Reik, *Science* **2010**, *330*, 622-627. *Epigenetic Reprogramming in Plant and Animal Development*.
[134] P. Hajkova, S. J. Jeffries, C. Lee, N. Miller, S. P. Jackson, M. A. Surani, *Science* **2010**, *329*, 78-82. *Genome-wide reprogramming in the mouse germ line entails the base excision repair pathway*.
[135] M. Wossidlo, J. Arand, V. Sebastiano, K. Lepikhov, M. Boiani, R. Reinhardt, H. Scholer, J. Walter, *EMBO J.* **2010**, *29*, 1877-1888. *Dynamic link of DNA demethylation, DNA strand breaks and repair in mouse zygotes*.
[136] H. Wu, Y. Zhang, *Cell* **2014**, *156*, 45-68. *Reversing DNA Methylation: Mechanisms, Genomics, and Biological Functions*.
[137] K. M. Kroeger, M. F. Goodman, M. M. Greenberg, *Nucleic Acids Res.* **2004**, *32*, 5480-5485. *A comprehensive comparison of DNA replication past 2-deoxyribose and its tetrahydrofuran analog in Escherichia coli*.
[138] K. M. Kroeger, Y. L. Jiang, Y. W. Kow, M. F. Goodman, M. M. Greenberg, *Biochemistry* **2004**, *43*, 6723-6733. *Mutagenic effects of 2-deoxyribonolactone in Escherichia coli. An abasic lesion that disobeys the A-rule*.
[139] K. M. Kroeger, J. Kim, M. F. Goodman, M. M. Greenberg, *Biochemistry* **2004**, *43*, 13621-13627. *Effects of the C4'-oxidized abasic site on replication in Escherichia coli. An unusually large deletion is induced by a small lesion*.
[140] K. M. Kroeger, J. Kim, M. F. Goodman, M. M. Greenberg, *Biochemistry* **2006**, *45*, 5048-5056. *Replication of an oxidized abasic site in Escherichia coli by a dNTP-stabilized misalignment mechanism that reads upstream and downstream nucleotides*.
[141] L. A. Loeb, B. D. Preston, *Annu. Rev. Genet.* **1986**, *20*, 201-230. *Mutagenesis by apurinic/apyrimidinic sites*.
[142] H. Atamna, I. Cheung, B. N. Ames, *Proc. Natl. Acad. Sci. U S A* **2000**, *97*, 686-691. *A method for detecting abasic sites in living cells: age-dependent changes in base excision repair*.
[143] B. X. Chen, K. Kubo, H. Ide, B. F. Erlanger, S. S. Wallace, Y. W. Kow, *Mutat. Res.* **1992**, *273*, 253-261. *Properties of a monoclonal antibody for the detection of abasic sites, a common DNA lesion*.
[144] N. Kojima, T. Takebayashi, A. Mikami, E. Ohtsuka, Y. Komatsu, *J. Am. Chem. Soc.* **2009**, *131*, 13208-13209. *Construction of highly reactive probes for abasic site detection by introduction of an aromatic and a guanidine residue into an aminooxy group*.
[145] J. Nakamura, V. E. Walker, P. B. Upton, S. Y. Chiang, Y. W. Kow, J. A. Swenberg, *Cancer Res.* **1998**, *58*, 222-225. *Highly sensitive apurinic/apyrimidinic site assay can detect spontaneous and chemically induced depurination under physiological conditions*.
[146] K. P. Roberts, J. A. Sobrino, J. Payton, L. B. Mason, R. J. Turesky, *Chem. Res. Toxicol.* **2006**, *19*, 300-309. *Determination of Apurinic/Apyrimidinic Lesions in DNA with High-Performance Liquid Chromatography and Tandem Mass Spectrometry*.

[147] M. Talpaert-Borle, M. Liuzzi, *Biochim. Biophys. Acta* **1983**, *740*, 410-416. *Reaction of apurinic/apyrimidinic sites with [14C]methoxyamine. A method for the quantitative assay of AP sites in DNA*.
[148] M. Weinfeld, M. Liuzzi, M. C. Paterson, *Biochemistry* **1990**, *29*, 1737-1743. *Response of phage T4 polynucleotide kinase toward dinucleotides containing apurinic sites: design of a 32P-postlabeling assay for apurinic sites in DNA*.
[149] S. E. Bennett, J. Kitner, *Nucleos. Nucleot. Nucl.* **2006**, *25*, 823-842. *Characterization of the Aldehyde Reactive Probe Reaction with AP-Sites in DNA: Influence of AP-Lyase on Adduct Stability*.
[150] E. Fundador, J. Rusling, *Anal. Bioanal. Chem.* **2007**, *387*, 1883-1890. *Detection of labeled abasic sites in damaged DNA by capillary electrophoresis with laser-induced fluorescence*.
[151] S. Wei, S. Shalhout, Y. H. Ahn, A. S. Bhagwat, *DNA Repair (Amst)* **2015**, *27*, 9-18. *A versatile new tool to quantify abasic sites in DNA and inhibit base excision repair*.
[152] G. M. Makrigiorgos, S. Chakrabarti, A. Mahmood, *Int. J. Radiat. Biol.* **1998**, *74*, 99-109. *Fluorescent labelling of abasic sites: a novel methodology to detect closely-spaced damage sites in DNA*.
[153] J. Nakamura, J. A. Swenberg, *Cancer Res.* **1999**, *59*, 2522-2526. *Endogenous apurinic/apyrimidinic sites in genomic DNA of mammalian tissues*.
[154] H. Ide, K. Akamatsu, Y. Kimura, K. Michiue, K. Makino, A. Asaeda, Y. Takamori, K. Kubo, *Biochemistry* **1993**, *32*, 8276-8283. *Synthesis and damage specificity of a novel probe for the detection of abasic sites in DNA*.
[155] D. M. Wilson, 3rd, D. Barsky, *Mutat. Res.* **2001**, *485*, 283-307. *The major human abasic endonuclease: formation, consequences and repair of abasic lesions in DNA*.
[156] T. Santa, *Biomed. Chromatogr.* **2011**, *25*, 1-10. *Derivatization reagents in liquid chromatography/electrospray ionization tandem mass spectrometry*.
[157] T. Pfaffeneder, PhD thesis, Ludwig-Maximilians-Universität München (München), **2015**.
[158] J. Cadet, S. Loft, R. Olinski, M. D. Evans, K. Bialkowski, J. Richard Wagner, P. C. Dedon, P. Moller, M. M. Greenberg, M. S. Cooke, *Free Radic. Res.* **2012**, *46*, 367-381. *Biologically relevant oxidants and terminology, classification and nomenclature of oxidatively generated damage to nucleobases and 2-deoxyribose in nucleic acids*.
[159] G. Boysen, L. B. Collins, S. Liao, A. M. Luke, B. F. Pachkowski, J. L. Watters, J. A. Swenberg, *J. Chromatogr. B. Analyt. Technol. Biomed. Life Sci.* **2010**, *878*, 375-380. *Analysis of 8-oxo-7,8-dihydro-2'-deoxyguanosine by ultra high pressure liquid chromatography-heat assisted electrospray ionization-tandem mass spectrometry*.
[160] H. Hong, H. Cao, Y. Wang, Y. Wang, *Chem. Res. Toxicol.* **2006**, *19*, 614-621. *Identification and quantification of a guanine-thymine intrastrand cross-link lesion induced by Cu(II)/H2O2/ascorbate*.
[161] F. Samson-Thibault, G. S. Madugundu, S. Gao, J. Cadet, J. R. Wagner, *Chem. Res. Toxicol.* **2012**, *25*, 1902-1911. *Profiling cytosine oxidation in DNA by LC-MS/MS*.
[162] J. L. Ravanat, T. Douki, P. Duez, E. Gremaud, K. Herbert, T. Hofer, L. Lasserre, C. Saint-Pierre, A. Favier, J. Cadet, *Carcinogenesis* **2002**, *23*, 1911-1918. *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*.
[163] K. Taghizadeh, J. L. McFaline, B. Pang, M. Sullivan, M. Dong, E. Plummer, P. C. Dedon, *Nat. Protoc.* **2008**, *3*, 1287-1298. *Quantification of DNA damage products resulting from deamination, oxidation and reaction with products of lipid peroxidation by liquid chromatography isotope dilution tandem mass spectrometry*.

[164] E. Zarakowska, D. Gackowski, M. Foksinski, R. Olinski, *Mutat. Res. Genet. Toxicol. Environ. Mutagen.* **2014**, *764-765*, 58-63. *Are 8-oxoguanine (8-oxoGua) and 5-hydroxymethyluracil (5-hmUra) oxidatively damaged DNA bases or transcription (epigenetic) marks?*
[165] M. L. Hamilton, H. Van Remmen, J. A. Drake, H. Yang, Z. M. Guo, K. Kewitt, C. A. Walter, A. Richardson, *Proc. Natl. Acad. Sci. U S A* **2001**, *98*, 10469-10474. *Does oxidative damage to DNA increase with age?*
[166] B. Nie, W. Gan, F. Shi, G. X. Hu, L. G. Chen, H. Hayakawa, M. Sekiguchi, J. P. Cai, *Oxid. Med. Cell Longev.* **2013**, *2013*, 303181. *Age-dependent accumulation of 8-oxoguanine in the DNA and RNA in various rat tissues*.
[167] N. Nishida, T. Arizumi, M. Takita, S. Kitai, N. Yada, S. Hagiwara, T. Inoue, Y. Minami, K. Ueshima, T. Sakurai, M. Kudo, *Dig. Dis.* **2013**, *31*, 459-466. *Reactive oxygen species induce epigenetic instability through the formation of 8-hydroxydeoxyguanosine in human hepatocarcinogenesis*.
[168] J. Li, A. Braganza, R. W. Sobol, *Antioxid. Redox Signal.* **2013**, *18*, 2429-2443. *Base Excision Repair Facilitates a Functional Relationship Between Guanine Oxidation and Histone Demethylation*.
[169] H. Verhagen, H. E. Poulsen, S. Loft, G. van Poppel, M. I. Willems, P. J. van Bladeren, *Carcinogenesis* **1995**, *16*, 969-970. *Reduction of oxidative DNA-damage in humans by brussels sprouts*.
[170] C. J. Burrows, *ACS Symp. Ser. Am. Chem. Soc.* **2009**, *2009*, 147-156. *Surviving an Oxygen Atmosphere: DNA Damage and Repair*.
[171] J. E. Klaunig, L. M. Kamendulis, *Annu. Rev. Pharmacol. Toxicol.* **2004**, *44*, 239-267. *The role of oxidative stress in carcinogenesis*.
[172] H. Kim, J. Wu, S. Ye, C. I. Tai, X. Zhou, H. Yan, P. Li, M. Pera, Q. L. Ying, *Nat. Commun.* **2013**, *4*, 2403. *Modulation of beta-catenin function maintains mouse epiblast stem cell and human embryonic stem cell self-renewal*.
[173] M. Cappiello, L. Mascia, C. Scolozzi, F. Giorgelli, P. L. Ipata, *Biochim. Biophys. Acta* **1998**, *1425*, 273-281. *In vitro assessment of salvage pathways for pyrimidine bases in rat liver and brain*.
[174] M. G. Tozzi, M. Camici, L. Mascia, F. Sgarrella, P. L. Ipata, *FEBS J.* **2006**, *273*, 1089-1101. *Pentose phosphates in nucleoside interconversion and catabolism*.
[175] M. J. Boland, J. K. Christman, *J. Mol. Biol.* **2008**, *379*, 492-504. *Characterization of Dnmt3b:thymine-DNA glycosylase interaction and stimulation of thymine glycosylase-mediated repair by DNA methyltransferase(s) and RNA*.
[176] R. Gallais, F. Demay, P. Barath, L. Finot, R. Jurkowska, R. Le Guevel, F. Gay, A. Jeltsch, R. Metivier, G. Salbert, *Mol. Endocrinol.* **2007**, *21*, 2085-2098. *Deoxyribonucleic acid methyl transferases 3a and 3b associate with the nuclear orphan receptor COUP-TFI during gene activation*.
[177] Y. Q. Li, P. Z. Zhou, X. D. Zheng, C. P. Walsh, G. L. Xu, *Nucleic Acids Res.* **2007**, *35*, 390-400. *Association of Dnmt3a and thymine DNA glycosylase links DNA methylation with base-excision repair*.
[178] G. Auclair, S. Guibert, A. Bender, M. Weber, *Genome Biol.* **2014**, *15*, 545. *Ontogeny of CpG island methylation and specificity of DNMT3 methyltransferases during embryonic development in the mouse*.
[179] T. Chen, Y. Ueda, J. E. Dodge, Z. Wang, E. Li, *Mol. Cell Biol.* **2003**, *23*, 5594-5605. *Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b*.

[180] A. Thakur, S. J. Mackin, R. E. Irwin, K. M. O'Neill, G. Pollin, C. Walsh, *Epigenetics Chromatin* **2016**, *9*, 53. *Widespread recovery of methylation at gametic imprints in hypomethylated mouse stem cells following rescue with DNMT3A2*.
[181] K. L. Tucker, C. Beard, J. Dausmann, L. Jackson-Grusby, P. W. Laird, H. Lei, E. Li, R. Jaenisch, *Genes. Dev.* **1996**, *10*, 1008-1020. *Germ-line passage is required for establishment of methylation and expression patterns of imprinted but not of nonimprinted genes*.
[182] J. Beygo, M. Elbracht, K. de Groot, M. Begemann, D. Kanber, K. Platzer, G. Gillessen-Kaesbach, A. Vierzig, A. Green, R. Heller, K. Buiting, T. Eggermann, *Eur. J. Hum. Genet.* **2015**, *23*, 180-188. *Novel deletions affecting the MEG3-DMR provide further evidence for a hierarchical regulation of imprinting in 14q32*.
[183] M. Kagami, M. J. O'Sullivan, A. J. Green, Y. Watabe, O. Arisaka, N. Masawa, K. Matsuoka, M. Fukami, K. Matsubara, F. Kato, A. C. Ferguson-Smith, T. Ogata, *PLoS Genet.* **2010**, *6*, e1000992. *The IG-DMR and the MEG3-DMR at human chromosome 14q32.2: hierarchical interaction and distinct functional properties as imprinting control centers*.
[184] S. Lopes, A. Lewis, P. Hajkova, W. Dean, J. Oswald, T. Forne, A. Murrell, M. Constancia, M. Bartolomei, J. Walter, W. Reik, *Hum. Mol. Genet.* **2003**, *12*, 295-305. *Epigenetic modifications in an imprinting cluster are controlled by a hierarchy of DMRs suggesting long-range chromatin interactions*.
[185] C. B. Mulholland, M. Smets, E. Schmidtmann, S. Leidescher, Y. Markaki, M. Hofweber, W. Qin, M. Manzo, E. Kremmer, K. Thanisch, C. Bauer, P. Rombaut, F. Herzog, H. Leonhardt, S. Bultmann, *Nucleic Acids Res.* **2015**, *43*, e112. *A modular open platform for systematic functional studies under physiological conditions*.
[186] L. Schermelleh, F. Spada, H. P. Easwaran, K. Zolghadr, J. B. Margot, M. C. Cardoso, H. Leonhardt, *Nat. Methods* **2005**, *2*, 751-756. *Trapped in action: direct visualization of DNA methyltransferase activity in living cells*.
[187] C. S. Schmidt, S. Bultmann, D. Meilinger, B. Zacher, A. Tresch, K. C. Maier, C. Peter, D. E. Martin, H. Leonhardt, F. Spada, *PLoS One* **2012**, *7*, e52629. *Global DNA hypomethylation prevents consolidation of differentiation programs and allows reversion to the embryonic stem cell state*.
[188] B. Hendrich, U. Hardeland, H. H. Ng, J. Jiricny, A. Bird, *Nature* **1999**, *401*, 301-304. *The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites*.
[189] P. Neddermann, J. Jiricny, *J. Biol. Chem.* **1993**, *268*, 21218-21224. *The purification of a mismatch-specific thymine-DNA glycosylase from HeLa cells*.
[190] F. Petronzelli, A. Riccio, G. D. Markham, S. H. Seeholzer, M. Genuardi, M. Karbowski, A. T. Yeung, Y. Matsumoto, A. Bellacosa, *J. Cell Physiol.* **2000**, *185*, 473-480. *Investigation of the substrate spectrum of the human mismatch-specific DNA N-glycosylase MED1 (MBD4): fundamental role of the catalytic domain*.
[191] J. Jiricny, *Cold Spring Harb. Perspect. Biol.* **2013**, *5*, a012633. *Postreplicative mismatch repair*.
[192] D. M. Franchini, C. F. Chan, H. Morgan, E. Incorvaia, G. Rangam, W. Dean, F. Santos, W. Reik, S. K. Petersen-Mahrt, *PLoS One* **2014**, *9*, e97754. *Processive DNA demethylation via DNA deaminase-induced lesion resolution*.
[193] D. M. Franchini, E. Incorvaia, G. Rangam, H. A. Coker, S. K. Petersen-Mahrt, *PLoS One* **2013**, *8*, e82097. *Simultaneous in vitro characterisation of DNA deaminase function and associated DNA repair pathways*.

[194] J. Pena-Diaz, S. Bregenhorn, M. Ghodgaonkar, C. Follonier, M. Artola-Boran, D. Castor, M. Lopes, A. A. Sartori, J. Jiricny, *Mol. Cell* **2012**, *47*, 669-680. *Noncanonical mismatch repair as a source of genomic instability in human cells*.
[195] C. E. Schrader, J. E. Guikema, E. K. Linehan, E. Selsing, J. Stavnezer, *J. Immunol.* **2007**, *179*, 6064-6071. *Activation-induced cytidine deaminase-dependent DNA breaks in class switch recombination occur during G1 phase of the cell cycle and depend upon mismatch repair*.
[196] A. B. Buermeyer, S. M. Deschenes, S. M. Baker, R. M. Liskay, *Annu. Rev. Genet.* **1999**, *33*, 533-564. *Mammalian DNA mismatch repair*.
[197] G. M. Li, *Cell Res.* **2008**, *18*, 85-98. *Mechanisms and functions of DNA mismatch repair*.
[198] N. Le May, D. Fradin, I. Iltis, P. Bougneres, J. M. Egly, *Mol. Cell* **2012**, *47*, 622-632. *XPG and XPF endonucleases trigger chromatin looping and DNA demethylation for accurate expression of activated genes*.
[199] N. Le May, D. Mota-Fernandes, R. Velez-Cruz, I. Iltis, D. Biard, J. M. Egly, *Mol. Cell* **2010**, *38*, 54-66. *NER factors are recruited to active promoters and facilitate chromatin modification for transcription in the absence of exogenous genotoxic attack*.
[200] O. D. Scharer, *Cold Spring Harb. Perspect. Biol.* **2013**, *5*, a012609. *Nucleotide excision repair in eukaryotes*.
[201] M. Gomez-Mendoza, A. Banyasz, T. Douki, D. Markovitsi, J. L. Ravanat, *J. Phys. Chem. Lett.* **2016**, *7*, 3945-3948. *Direct Oxidative Damage of Naked DNA Generated upon Absorption of UV Radiation by Nucleobases*.
[202] C. E. Crespo-Hernández, R. Arce, *J. Phys. Chem. B* **2003**, *107*, 1062-1070. *Near Threshold Photo-Oxidation of Dinucleotides Containing Purines upon 266 nm Nanosecond Laser Excitation. The Role of Base Stacking, Conformation, and Sequence*.
[203] S. Marguet, D. Markovitsi, F. Talbot, *J. Phys. Chem. B* **2006**, *110*, 11037-11039. *One- and two-photon ionization of DNA single and double helices studied by laser flash photolysis at 266 nm*.
[204] D. B. Bucher, B. M. Pilles, T. Carell, W. Zinth, *Proc. Natl. Acad. Sci.* **2014**, *111*, 4369-4374. *Charge separation and charge delocalization identified in long-living states of photoexcited DNA*.
[205] D. B. Bucher, C. L. Kufner, A. Schlueter, T. Carell, W. Zinth, *J. Am. Chem. Soc.* **2016**, *138*, 186-190. *UV-Induced Charge Transfer States in DNA Promote Sequence Selective Self-Repair*.
[206] H. Cao, Y. Wang, *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 1335-1341. *Collisionally activated dissociation of protonated 2'-deoxycytidine, 2'-deoxyuridine, and their oxidatively damaged derivatives*.
[207] S. Schiesser, T. Pfaffeneder, K. Sadeghian, B. Hackner, B. Steigenberger, A. S. Schroder, J. Steinbacher, G. Kashiwazaki, G. Hofner, K. T. Wanner, C. Ochsenfeld, T. Carell, *J. Am. Chem. Soc.* **2013**, *135*, 14593-14599. *Deamination, oxidation, and C-C bond cleavage reactivity of 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxycytosine*.
[208] J. Wang, B. Yuan, C. Guerrero, R. Bahde, S. Gupta, Y. Wang, *Anal. Chem.* **2011**, *83*, 2201-2209. *Quantification of oxidative DNA lesions in tissues of Long-Evans Cinnamon rats by capillary high-performance liquid chromatography-tandem mass spectrometry coupled with stable isotope-dilution method*.

7 Anhang

7.1 Zusatzmaterialien zu Abschnitt 3.1

K. Iwan#, R. Rahimoff#, A. Kirchner#, F. Spada#, A. S. Schröder, **O. Kosmatchev**, S. Ferizaj, J. Steinbacher, E. Parsa, M. Müller, T. Carell, *Nat. Chem. Biol.* **2018**, *14*, 72-78. *5-Formylcytosine to cytosine conversion by C-C bond cleavage in vivo.*

(# geteilte Erstautorenschaft)

Supplementary Tables

Supplementary Table 1: Compound-dependent LC-MS/MS-parameters used for the analysis of genomic DNA fed with fluorinated nucleosides. CE: collision energy, CAV: collision cell accelerator voltage, EMV: electron multiplier voltage. The nucleosides were analyzed in the positive ($[M+H]^+$ species) as well as the negative ($[M-H]^-$ species) 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 1.5-3.0 min								
F-dC	246.09	Wide	112.06	Wide	70	15	3	Positive
$[^{15}N_2]$-F-dC	248.08	Wide	114.04	Wide	70	15	3	Positive
$[^{15}N_2]$-cadC	274.08	Wide	158.03	Wide	40	5	5	Positive
cadC	272.09	Wide	156.04	Wide	40	5	5	Positive
F-hmdC	276.10	Wide	142.06	Wide	50	10	3	Positive
$[^{15}N_2]$-F-hmdC	278.09	Wide	144.06	Wide	50	10	3	Positive
$[^{15}N_2,D_2]$-hmdC	262.12	Wide	146.07	Wide	25	27	1	Positive
hmdC	258.11	Wide	142.06	Wide	25	27	1	Positive
$[D_3]$-mdC	245.13	Wide	129.09	Wide	50	60	1	Positive
mdC	242.11	Wide	126.07	Wide	50	60	1	Positive
Time segment 3.0-4.7 min								
F-hmdU	275.07	Wide	255.06	Wide	80	3	7	Negative
F-dU	245.06	Wide	225.06	Wide	80	3	5	Negative
$[D_3]$-F-mdC	263.12	Wide	129.09	Wide	80	15	3	Positive
F-mdC	260.10	Wide	126.07	Wide	80	15	3	Positive
F-cadC	290.08	Wide	156.04	Wide	80	5	5	Positive
Time segment 4.7-10 min								
F-fdC	274.08	Wide	140.05	Wide	90	15	3	Positive
$[^{15}N_2]$-F-fdC	276.08	Wide	142.04	Wide	90	15	3	Positive
F-fdU	273.05	Wide	253.05	Wide	30	3	5	Negative
F-dT	259.07	Wide	239.07	Wide	70	3	5	Negative
dT	243.10	Wide	127.05	Wide	20	5	5	Negative
$[^{15}N_2]$-fdC	258.09	Wide	142.04	Wide	30	5	5	Positive
fdC	256.09	Wide	140.05	Wide	30	5	5	Positive
$[^{15}N_2]$-fdC	142.04	Wide	98.04	Wide	20	13	7	Positive
fdC	140.05	Wide	97.04	Wide	20	13	7	Positive

Supplementary Table 2: Compound-dependent LC-MS/MS ranges of the corresponding linear equations.

compound	n (ULOQ)	n (LLOQ)	A/A* (ULOQ)	A/A* (LLOQ)
F-fdC	25.1 fmol	0.390 fmol	0.2267	0.003345
F-dC	800 fmol	3.13 fmol	4.074	0.01512
F-mdC	49.9 fmol	0.780 fmol	0.04174	0.0005833

Nature Chemical Biology: doi:10.1038/nchembio.2531

Supplementary Table 3: Compound-dependent LC-MS/MS-parameters used for the analysis of genomic DNA fed with 1. CE: collision energy, CAV: collision cell accelerator voltage, EMV: electron multiplier voltage. The nucleosides were analyzed in the positive ($[M+H)]^+$ species) as well as the negative ($[M-H]^-$ species) 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 1.5-3.3 min								
sh-cadC	279.10	Wide	158.03	Wide	50	5	5	Positive
$[^{15}N_2]$-cadC	274.08	Wide	158.03	Wide	20	5	5	Positive
cadC	272.09	Wide	156.04	Wide	20	5	5	Positive
sh-hmdC	265.12	Wide	144.06	Wide	60	27	1	Positive
$[^{15}N_2,D_2]$-hmdC	262.12	Wide	146.07	Wide	20	27	1	Positive
hmdC	258.11	Wide	142.06	Wide	20	27	1	Positive
sh-mdC	249.12	Wide	128.06	Wide	60	60	1	Positive
$[D_3]$-mdC	245.13	Wide	129.09	Wide	20	60	1	Positive
mdC	242.11	Wide	126.07	Wide	20	60	1	Positive
sh-dC	235.11	Wide	114.04	Wide	60	10	2	Positive
dC	228.10	Wide	112.05	Wide	15	1	1	Positive
$[^{13}C_9,^{15}N_3]$-dC	240.10	Unit	119.10	Unit	70	10	2	Positive
Time segment 3.3-4.15 min								
$[D_2]$-hmdU	259.09	Wide	216.08	Wide	80	7	5	Negative
hmdU	257.08	Wide	214.07	Wide	80	7	5	Negative
sh-dU	234.08	Wide	192.09	Wide	120	5	5	Negative
$[^{15}N_2]$-dU	229.06	Wide	185.06	Wide	80	5	5	Negative
dU	227.07	Wide	184.06	Wide	120	5	5	Negative
Time segment 4.15-9 min								
$[^{15}N_2]$-8-oxo-dG	289.08	Wide	173.04	Wide	10	9	7	Positive
8-oxo-dG	284.10	Wide	168.05	Wide	10	9	7	Positive
sh-fdC	263.10	Wide	142.04	Wide	80	5	5	Positive
$[^{15}N_2]$-fdC	258.09	Wide	142.04	Wide	80	5	5	Positive
fdC	256.09	Wide	140.05	Wide	10	5	5	Positive
sh-fdU	262.07	Wide	141.01	Wide	80	15	5	Negative
$[^{15}N_2]$-fdU	257.06	Wide	141.01	Wide	80	15	5	Negative
fdU	255.06	Wide	139.01	Wide	10	15	5	Negative
sh-dT	250.11	Wide	129.04	Wide	20	5	5	Positive
dT	243.10	Wide	127.05	Wide	10	45	3	Positive

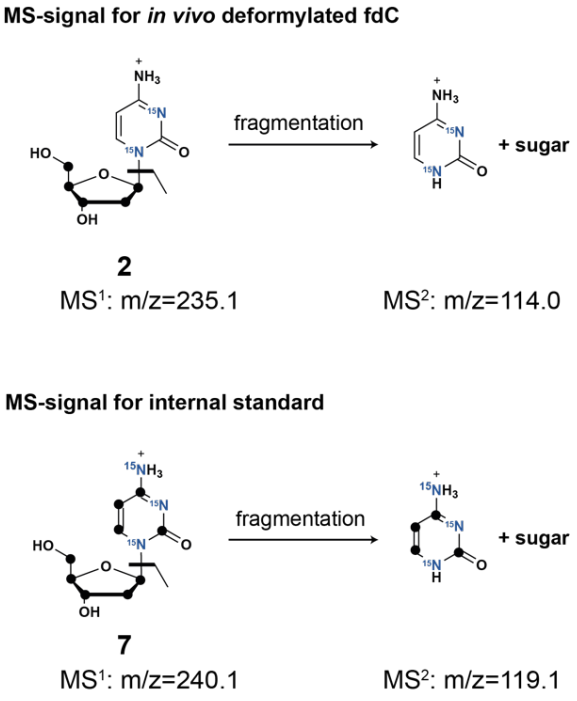
Nature Chemical Biology: doi:10.1038/nchembio.2531

Supplementary Table 4: Compound-dependent LC-MS/MS-parameters used for the analysis of genomic DNA fed with 7. CE: collision energy, CAV: collision cell accelerator voltage, EMV: electron multiplier voltage. The nucleosides were analyzed in the positive ($[M+H)]^+$ species) as well as the negative ($[M-H]^-$ species) ion selected reaction monitoring mode (SRM).

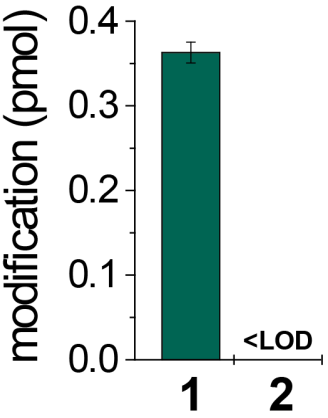
compound	Precursor ion (m/z)	MS1 Resolution	Product Ion (m/z)	MS2 Resolution	Dwell time [ms]	CE (V)	CAV (V)	Polarity
Time segment 1.5-3.3 min								
$[^{13}C_9,^{15}N_3]$-dC	240.10	Unit	119.10	Unit	70	30	5	Positive
dC	228.1	Wide	112.10	Wide	70	1	3	Positive
Time segment 3.3-4.8 min								
dU	227.07	Wide	184.06	Wide	80	5	5	Negative
$[^{13}C_9,^{15}N_2]$-dU	238.09	Wide	193.10	Wide	80	5	5	Negative
Time segment 4.8-12.0 min								
dT	243.1	Wide	127.05	Wide	70	40	3	Positive
$[^{13}C_9,^{15}N_2]$-dT	254.12	Unit	133.06	Unit	70	3	5	Positive
$[D_3]$-dT	246.10	Unit	130.05	Unit	30	3	5	Positive

Nature Chemical Biology: doi:10.1038/nchembio.2531

Supplementary Figures

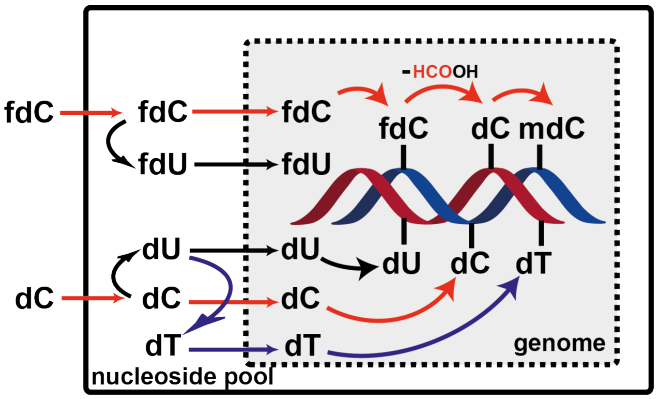


Supplementary Figure 1: Characteristic MS transition of [$^{13}C_5$][$^{15}N_2$]-dC and [$^{13}C_9$][$^{15}N_3$]-dC. Fingerprint MS-fragmentation pathways and detected MS signals of **2** and **7** used to track isotopically labelled dC and its derivatives in mESC cultures.

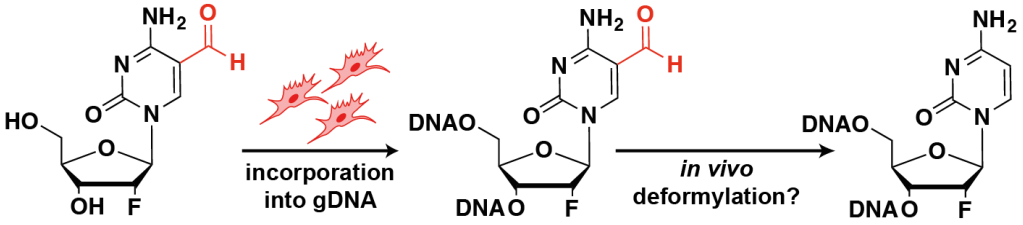


Supplementary Figure 2: Quantitative data of isotopically labelled pyrimidine derivatives in the soluble nucleoside pool of mESCs after feeding [$^{13}C_5$][$^{15}N_2$]-fdC shows its incorporation and no formation of [$^{13}C_5$][$^{15}N_2$]-dC. LC-MS/MS analysis of the soluble nucleoside pool from J1 mESCs cultured in the presence of 50 µM **1** under priming conditions for three days. Compound **1** is incorporated but no formation of **2** is detected in the soluble nucleoside pool. Mean values and s.d. of technical triplicates from a single culture are shown.

Nature Chemical Biology: doi:10.1038/nchembio.2531

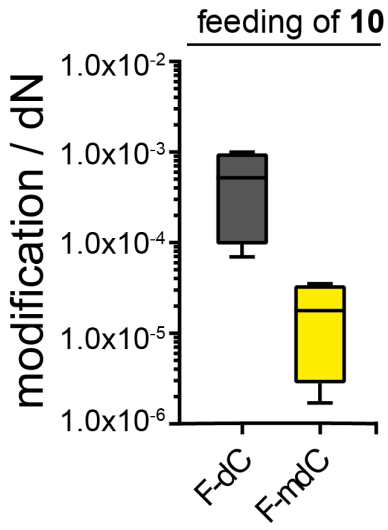


Supplementary Figure 3: Schematic overview of the possible metabolic conversions of fdC and dC.

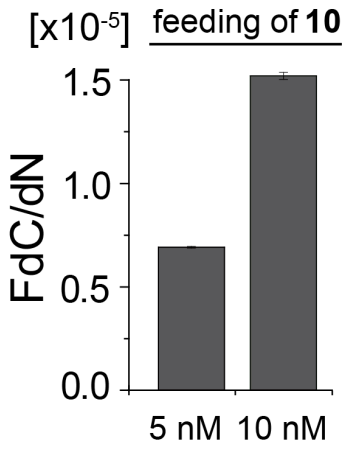


Supplementary Figure 4: **Schematic overview of the feeding experiment with (*R*)-2´-F-fdC to mESCs.** Feeding of **15** results in the incorporation of (*R*)-2´-F-fdC into genomic DNA and is converted *in vivo* to (*R*)-2´-F-dC by a C-C bond cleavage reaction.

Nature Chemical Biology: doi:10.1038/nchembio.2531

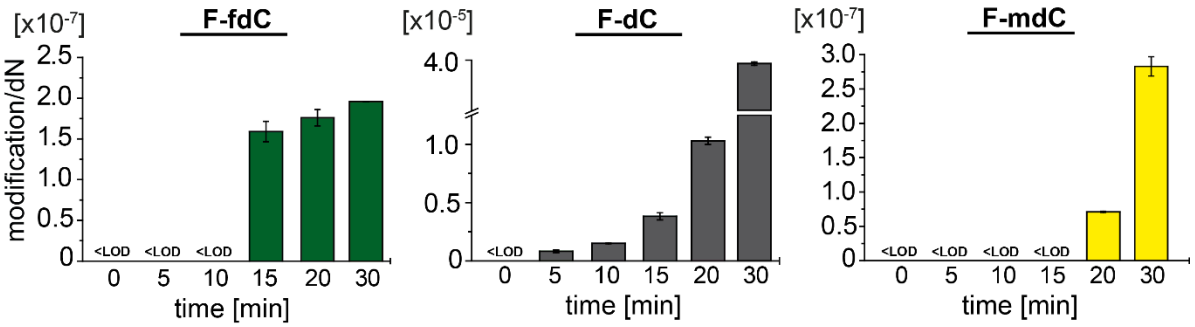


Supplementary Figure 5: Quantitative data of fluorinated pyrimidine derivatives after treatment with 10 **show incorporation of** 10 **and 3% re-methylation.**
LC-MS/MS analysis of genomic DNA from J1 mES cells cultured in the presence of 1.0 µM (*R*)-2'-F-dC under priming conditions for three days. DNA samples of four independent cultures were measured, each as technical triplicate. Mean values with s.d. are depicted.

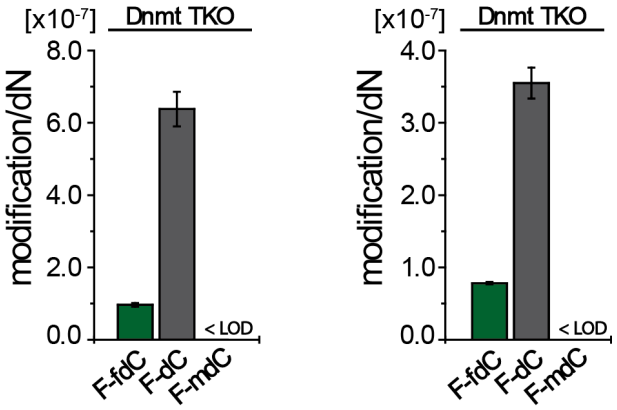


Supplementary Figure 6: Quantification data of incorporated (*R*)-2'-F-dC after feeding of 10 **at two concentrations. Levels similar to those resulting from feeding of (*R*)-2'-F-fdC were observed.**
Analysis of genomic DNA from J1 mES cells cultured for three days under priming conditions in the presence of 5 and 10 nM (*R*)-2'-F-dC. Mean values and s.d. of technical triplicates from single cultures are shown.

Nature Chemical Biology: doi:10.1038/nchembio.2531

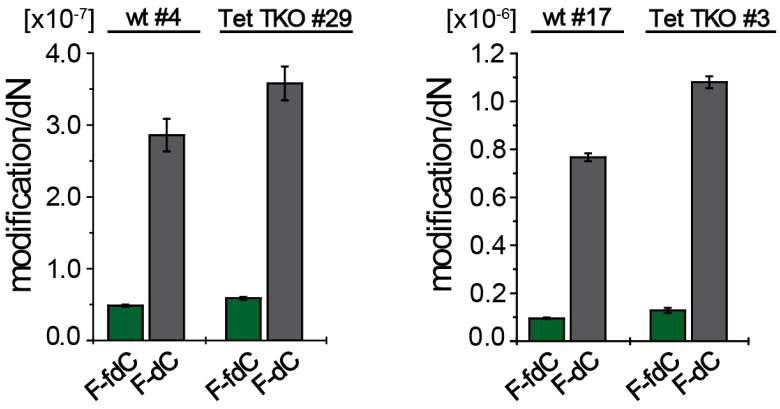


Supplementary Figure 7: Time dependent co-feeding with 15 **and** 10 **show incorporation of** 10 **already after 5 min,** 15 **after 15 min and** 11**, resulting from both compounds, after 20 min.**
LC-MS/MS analysis of the genomic DNA from J1 mESCs cultured under priming conditions and after 48 h supplemented with both (*R*)-2'-F-fdC (350 µM) and (*R*)-2'-F-dC (1.0 µM) for 0, 5, 10, 15, 20 and 30 min. Mean values and s.d. of technical triplicates from single cultures are shown.



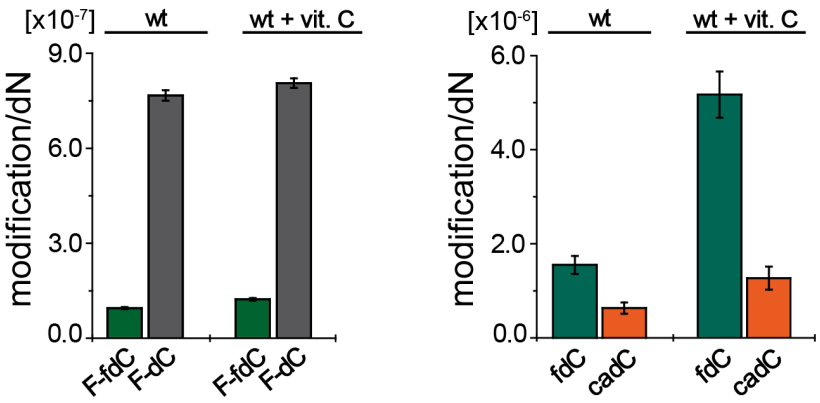
Supplementary Figure 8: Catalytically active DNA methyltransferases are required for methylation of (*R*)-2'-F-dC, but not for deformylation of (*R*)-2'-F-fdC (related to Fig. 4d).
LC-MS/MS analysis of genomic DNA from Dnmt TKO J1 mESCs cultured in the presence of (*R*)-2'-F-fdC under priming conditions. Shown are mean values and s.d. from two independent cultures, each measured as technical triplicates.

Nature Chemical Biology: doi:10.1038/nchembio.2531



Supplementary Figure 9: Tet enzymes are not required for deformylation of (*R*)-2'-F-fdC (related to Fig. 4e).

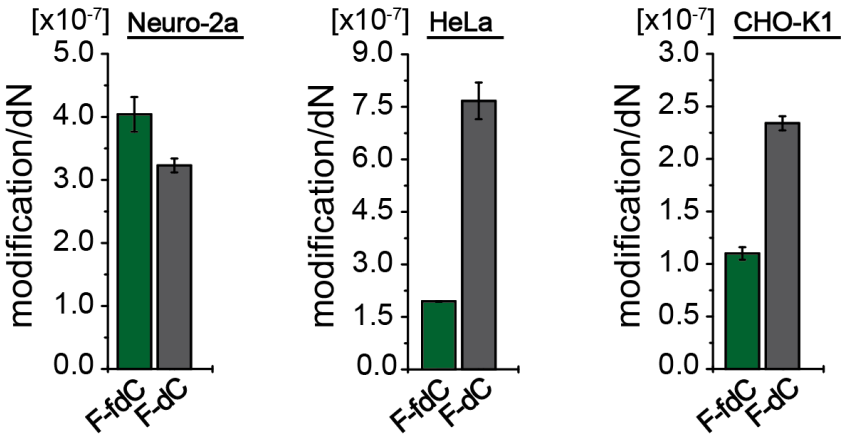
LC-MS/MS analysis of genomic DNA from wt and Tet TKO mESCs cultured in the presence of (*R*)-2'-F-fdC under priming conditions. Left panel: cell lines described in (*34*). Right panel: cell lines reported in (*33*) (same cell lines as for experiment shown in Figure 4e). Shown are mean values and s.d. from two independent cell culture experiments, each measured as technical triplicates.



Supplementary Figure 10: Deformylation of (*R*)-2'-F-fdC is not affected by treatment with ascorbic acid-2-phosphate.

LC-MS/MS analysis of genomic DNA from wt mESCs cultured in the presence of (*R*)-2'-F-fdC under priming conditions with or without ascorbic acid-2-phosphate. The cell lines used in this experiment were reported in (*33*) and are the same as used for experiment shown in Figure 4e and Supplementary Figure 9. Shown are mean values and s.d. from two independent cell culture experiments, each measured as technical triplicates.

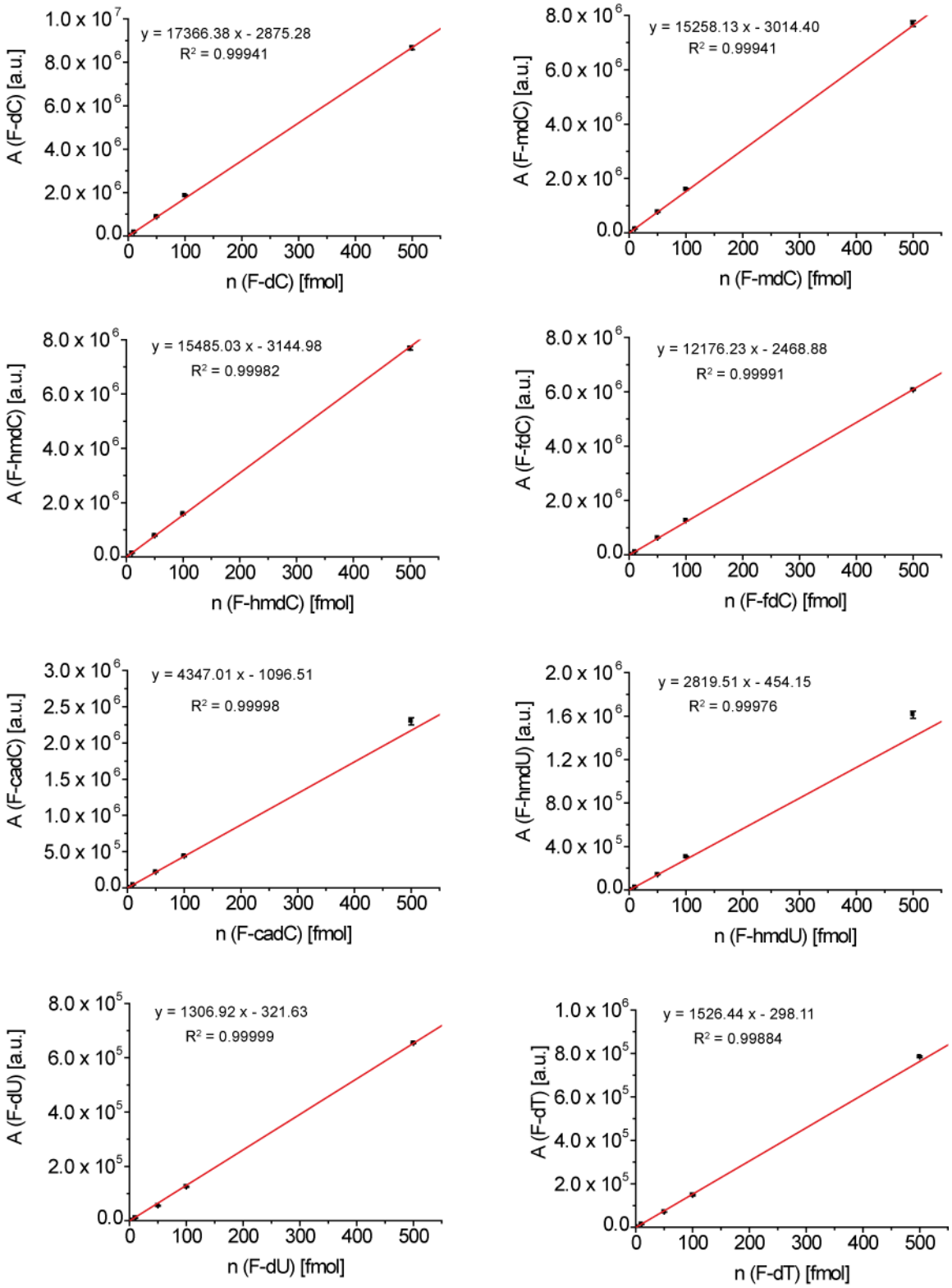
Nature Chemical Biology: doi:10.1038/nchembio.2531



Supplementary Figure 11: Deformylation of (*R*)-2'-F-fdC to (*R*)-2'-F-dC occur also in non-pluripotent mouse and human cell lines (related to Fig. 4f).

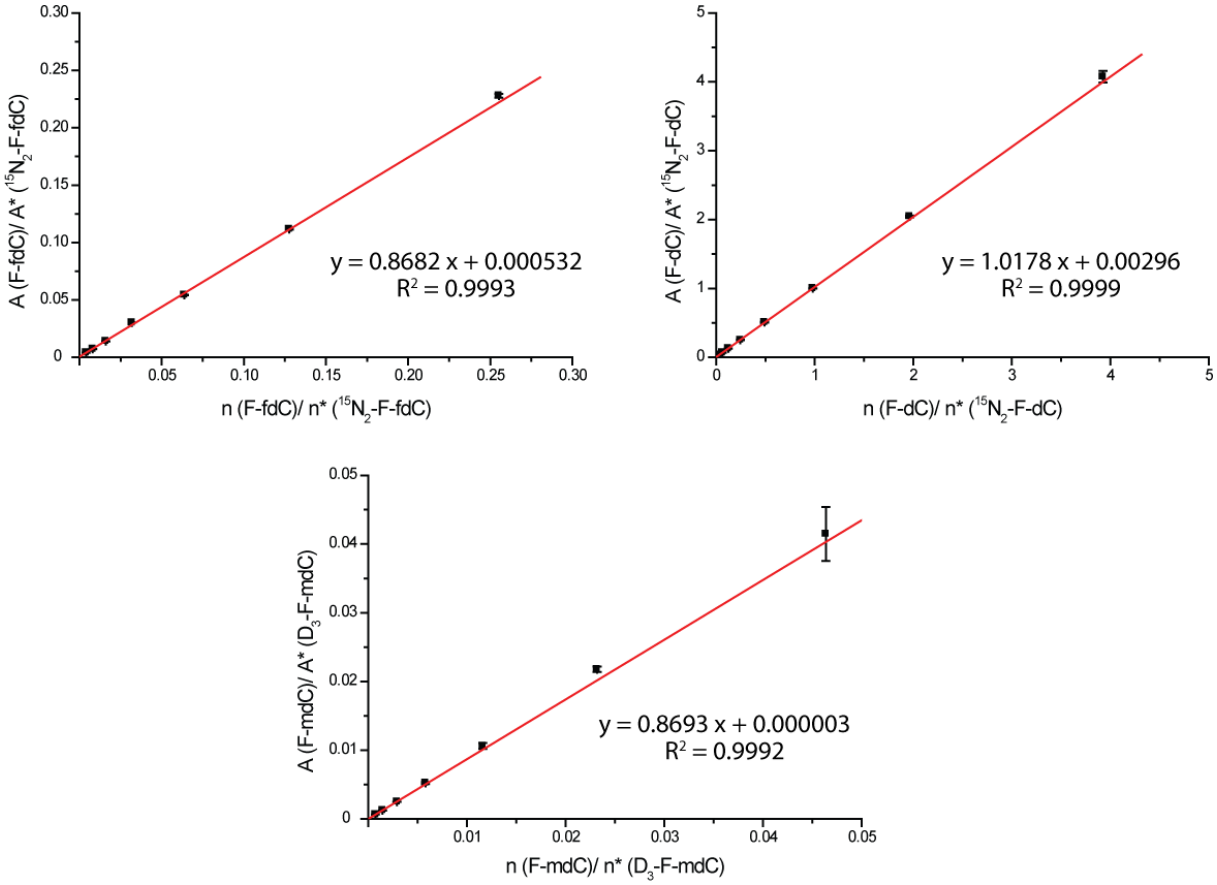
LC-MS/MS analysis of genomic DNA isolated from Neuro-2a, HeLa and CHO-K1 cells cultured in the presence of (*R*)-2'-F-fdC. Albeit at variable levels, incorporation of (*R*)-2'-F-fdC into the genome as well as its conversion to (*R*)-2'-F-dC are detected in all cell lines. Shown are mean values and s.d. from three independent cell culture experiments, each measured as technical triplicates.

Nature Chemical Biology: doi:10.1038/nchembio.2531



Supplementary Figure 12: Representative external calibration curves for all investigated nucleosides.

Nature Chemical Biology: doi:10.1038/nchembio.2531



Supplementary Figure 13: Internal calibration curves for the exact quantification of (*R*)-2'-F-fdC, (*R*)-2'-F-dC and (*R*)-2'-F-mdC with the corresponding linear equation and coefficient of determination.

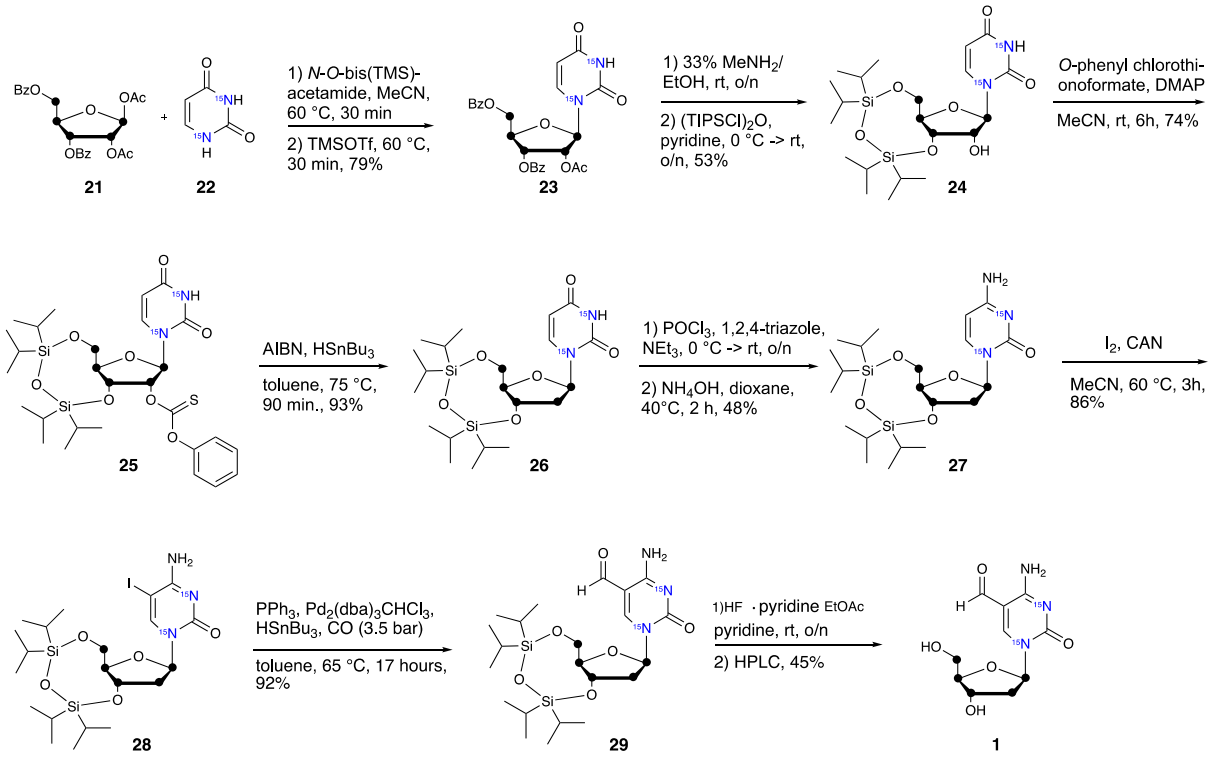
Nature Chemical Biology: doi:10.1038/nchembio.2531

Synthetic procedures

General information

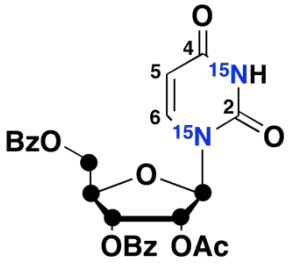
Isotopically labelled **7** was bought from *B.A.C.H. UG*, 2'-deoxy-2'-(*R*)-fluorocytidine **10** was purchased from *Carbosynth*. Compounds **8** and **9** resulted from the feeding experiment of **7**, compounds **11 - 14** resulted from the feeding of **10**. **15** and **30** were synthesized as described in the literature.[1,2] Identity of compounds used in this study were confirmed by standard spectroscopic methods such as NMR and MS.

Synthesis of 2'-deoxy-5-formyl(1',2',3',4',5'-$^{13}C_5$, $^{15}N_2$)cytidine (1)



Scheme 1: Synthesis of $^{13}C_5$-$^{15}N_2$-labelled fdC **1**.

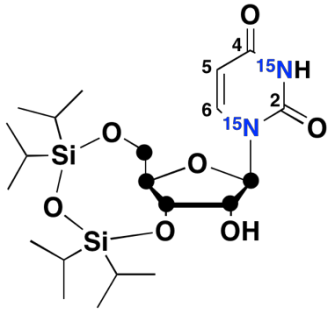
2'-*O*-acetyl-3',5'-di-*O*-benzoyl(1',2',3',4',5'-$^{13}C_5$, N^1,N^3- $^{15}N_2$)-uridine (23)



The ^{13}C ribose fragment **21** and $^{15}N_2$ urea **22** were synthesized according to the literature.[3,4] Bis-benzoyl and bis-acetyl protected ^{13}C ribose **21** (2.00 g, 4.47 mmol, 1.0 eq), $^{15}N_2$-Uracil (510 mg, 4.47 mmol, 1.0 eq) and *N,O*-bis(trimethylsilyl)acetamide (3.28 mL, 13.4 mmol, 3.0 eq) were suspended in dry MeCN (25 mL) and heated to 60 °C whereas a clear solution developed after 30 minutes of stirring. Subsequently TMS triflate (2.83 mL, 15.7 mmol, 3.5 eq) was added at 60 °C and the mixture was stirred for an additional 30 minutes. After the reaction was cooled to room temperature, MeCN was evaporated *in vacuo*, the residue was taken up in sat. aqueous $NaHCO_3$ (75 mL) and extracted with DCM (3 x 75 mL). The combined organic layers were dried over Na_2SO_4 and the solvents were removed by rotary evaporation. Purification of the crude material through column chromatography (20% EtOAc/^{i}Hex --> 50% EtOAc/^{i}Hex) yielded 1.75 g (3.49 mmol, 79%) of **23** as a white foam.

^{1}H-NMR (400 MHz, $CDCl_3$, ppm): δ = 8.08 - 8.05 (m, 4H, Bz-**H**), 7.67- 7.57 (m, 2H, Bz-**H**), 7.51 - 7.46 (m, 4H, Bz-**H**), 7.49 (td, ^3J = 7.8 Hz, $^1J_{N\text{-}H}$ = 3.2 Hz, 1H, 5-H), 6.41 (d, $^1J_{C\text{-}H}$ = 169.1 Hz, 1H, 1'-H), 5.96 (d, $^1J_{C\text{-}H}$ = 150.3 Hz, 1H, 3'-H), 5.74 (d, $^1J_{C\text{-}H}$ = 153.0 Hz, 2'-H), 5.57 5.74 (m, 1H, 6-H), 4.97 (d, $^1J_{C\text{-}H}$ = 154.6 Hz, 1H, 4'-H), 4.82 (d, $^1J_{13C\text{-}H}$ = 150.1 Hz, 1H, 5'-H), 2.06 (s, 3H, Ac-C**H$_3$**). **^{13}C-NMR (101 MHz, $CDCl_3$, ppm):** δ = 169.7 (Ac-**C**=O), 166.0 (Bz-**C**=O), 165.3 (Bz-**C**=O), 162.4 (d, $^1J_{C\text{-}N}$ = 9.5 Hz, C-4),150.1 (t, $^1J_{C\text{-}N}$ = 19.4 Hz, C-2), 139.2 (d, $^1J_{C\text{-}N}$ = 12.5 Hz, C-5), 133.9 (2 x C_{Ar}), 133.8 (2 x **C**$_{Ar}$), 129.8 (2 x **C**$_{Ar}$), 129.6 (2 x **C**$_{Ar}$), 128.8 (2 x **C**$_{Ar}$), 128.7 (2 x **C**$_{Ar}$), 103.5 (d, $^1J_{C\text{-}N}$ = 6.9 Hz, C-6) 87.2 (ddd, *J* = 44.4 Hz, 13.7 Hz, 3.7 Hz, 1'-C), 80.4 (dd, *J* = 43.1 Hz, 38.3 Hz, 4'-C), 73.0 (dd, *J* = 44.3, 39.9, 2'-C), 70.9 (td, *J* = 39.9 Hz, 39.4 Hz, 3.7 Hz, 3'-C), 63.7 (d, *J* = 43.1 Hz), 20.5 (Ac-**C**H$_3$). **^{15}N-NMR (41 MHz, $CDCl_3$, ppm):** δ = -224.4, -224.5. **HRMS (ESI$^+$):** calc. for $C_{20}{}^{13}C_5H_{26}N^{15}N_2O_9^+$ $[M+NH_4]^+$: 519.1772, found: 519.1784. **HRMS (ESI$^-$):** calc. for $C_{20}{}^{13}C_5H_{21}{}^{15}N_2O_9$ $[M\text{-}H]^-$: 500.1361, found: 500.1359. **IR (ATR):** ν (cm^{-1}) = 1712 (s), 1688 (s), 1679 (s), 1631 (w), 1601 (w), 1584 (w), 1450 (m), 1369 (m), 1315 (w), 1249 (s), 1220 (s), 1176 (m), 1107 (s), 1067 (s), 1025 (s), 806 (m), 762 (w), 707 (s), 686 (m).

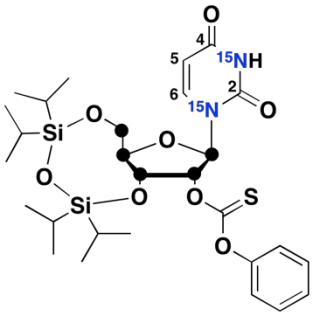
3',5'-(1,1,3,3,-tetraisopropyldisiloxane-1,3-diyl)-(1',2',3',4',5'-$^{13}C_5$, N^1,N^3-$^{15}N_2$)-uridine (24)



First, **23** (1.75 g, 3.49 mmol, 1.0 eq) was deprotected by stirring the compound in a solution of $MeNH_2$ in ethanol (10 mL, 33%) overnight. The ethanolic solution was concentrated *in vacuo* and resuspended in water (25 mL). The aqueous phase was extracted with DCM (5 x 25 mL) until no UV absorption on a TLC plate was visible anymore. Subsequently, the water was removed by rotary evaporation and the deprotected nucleoside was dried on high vac. Then, pyridine (3.0 mL) was added and the light red, clear solution was cooled to 0 °C. 1,3-dichloro-1,1,3,3-tetraisopropyldisiloxane (1.34 µL, 4.19 mmol, 1.2 eq) was added dropwise and the mixture was then warmed to room temperature and stirred for 14 hours. MeOH (3.0 mL) was added and the volatiles were removed *in vacuo*. The crude product mixture was then charged with saturated solution of aqueous NH_4Cl (50 mL), extracted twice with EtOAc (50 mL) and the organic layers dried over Na_2SO_4. After removal of the solvent *in vacuo*, pure product was isolated after column chromatography (30% EtOAc/^{i}Hex --> 40% EtOAc/^{i}Hex), which yielded 920 mg (1.86 mmol, 53%) of **24** as an off white foam.

^{1}H-NMR (599 MHz, $CDCl_3$, ppm): δ = 9.23 (d, $^2J_{N\text{-}H}$ = 90.9 Hz, 1H, N-**H**), 7.72 (td, ^3J = 8.1 Hz, $^2J_{N\text{-}H}$ = 2.2 Hz, 1H, 5-H), 5.88 (d, $^1J_{C\text{-}H}$ = 174.6 Hz, 1H, 1'-H), 5.69 (ddd, ^3J = 7.8 Hz, $^3J_{N\text{-}H}$ = 4.6 Hz, , $^3J_{N\text{-}H}$ = 2.8 Hz, 1H, 6-H), 4.47 - 3.87 (m, 2'-H, 3'-H, 4'-H, 5'-H), 1.10 - 1.01 (m, 28H, Si(C**H**$_3$)$_2$, Si-C**H**). **^{13}C-NMR (151 MHz, $CDCl_3$, ppm):** δ = 163.3 (d, $^1J_{C\text{-}N}$ = 9.5 Hz, C-4), 150.0 (t, $^1J_{C\text{-}N}$ = 18.4 Hz, 2-C), 139.9 (d, $^1J_{C\text{-}N}$ = 12.7 Hz, C-5), 105.4 (d, $^2J_{C\text{-}N}$ = 34.7 Hz, C-6) , 90.9 (dd, $^1J_{C\text{-}C}$ = 42.6 Hz, $^1J_{C\text{-}N}$ = 11.0 Hz, 1'-C), 81.9 (t, J = 42.3 Hz, 2'-C), 75.1 (t, $^1J_{C\text{-}C}$ = 40.6, 3'-C), 68.9 (t, J = 39.9 Hz, 4'-C), 60.2 (d, J = 39.9 Hz), 17.4 - 16.8 (Si-C(**C**H$_3$)), 13.4 - 12.8 (4 x Si-**C**(CH$_3$)). **^{15}N-NMR (41 MHz, $CDCl_3$, ppm):** δ = -224.4, -224.5. **HRMS (ESI$^+$):** calc. for $C_{16}{}^{13}C_5H_{39}{}^{15}N_2O_7Si_2^+$ $[M+H]^+$: 494.2399, found: 494.2412. **HRMS (ESI$^-$):** calc. for $C_{16}{}^{13}C_5H_{37}{}^{15}N_2O_7Si_2^-$ $[M-H]^-$: 492.2253, found: 492.2252. **IR (ATR):** ν (cm^{-1}) = 2944 (m), 2867 (m), 1688 (s), 1679 (s), 1650 (m), 1631 (w), 1462 (m), 1444 (m), 1383 (w), 1255 (m), 1198 (w), 1135 (m), 1096 (s), 1028 (s), 976 (s), 919 (w), 884 (s), 852 (m), 807 (m), 761 (m), 694 (s).

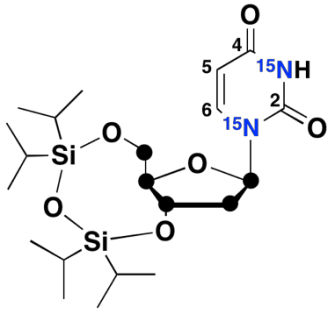
3',5'-(1,1,3,3,-tetraisopropyldisiloxane-1,3-diyl)-2'-O-phenoxythiocarbonyl-(1',2',3',4',5'-$^{13}C_5$, *N*^{1},*N*^{3}-$^{15}N_2$)-uridine (25)



Compound **25** was synthesized according to the literature with slight modifications.[5] **24** (280 mg, 0.57 mmol, 1.0 eq) was dissolved in MeCN (8 mL) and *O*-phenyl chlorothionoformate (120 µL, 0.86 mmol, 1.5 eq) and 4-(diemthylamino)pyridine (202 mg, 1.65 mmol, 2.9 eq) were added at room temperature. After the yellowish mixture, which showed a colourless precipitate after 30 minutes, was stirred for a total of six hours, the reaction was aborted through addition of MeOH (10 mL) and the volatiles were removed under reduced pressure. Subsequently, a saturated aqueous $NaHCO_3$ solution (25 mL) was given to the crude reaction mixture and extracted with DCM (3 x 25 mL). Combined organic layers were then dried over Na_2SO_4 and the solvent was removed *in vacuo*. Purification through column chromatography (25% EtOAc/^{i}Hex) yielded 262 mg (0.42 mmol, 74%) of pure product **25** as a colourless oil.

^{1}H-NMR (599 MHz, $CDCl_3$, ppm): δ = 8.27 (d, $^2J_{N\text{-}H}$ = 91.1 Hz, 1H, N-**H**), 7.71 (dt, ^3J = 8.1 Hz, $^2J_{N\text{-}H}$ = 2.2 Hz, 1H, 5-H), 7.42 (t, ^3J = 8.0 Hz, 2H, $\mathbf{H}_{Ar}$), 7.30 (t, ^3J = 7.6 Hz, 1H, $\mathbf{H}_{Ar}$), 7.12 (d, ^3J = 8.3 Hz, 2H, $\mathbf{H}_{Ar}$), 6.15 (td, $^1J_{C\text{-}H}$ = 164.2 Hz, ^3J = 5.5 Hz, 1H, 2'-H), 6.07 (d, $^1J_{C\text{-}H}$ = 174.8 Hz, 1H, 1'-H), 5.72 (ddd, ^3J = 7.7 Hz, $^3J_{N\text{-}H}$ = 4.9 Hz, , $^3J_{N\text{-}H}$ = 2.5 Hz, 1H, 6-H), 4.55 (ddt, $^1J_{C\text{-}H}$ = 142.4 Hz, ^3J = 8.4 Hz, 4.0 Hz,1H, 4'-H), 4.45 - 3.90 (m, 3H, 3'-H, 5'-H), 1.12 - 1.03 (m, 28H, Si(C$\mathbf{H}_3$)$_2$, Si-C**H**). **^{13}C-NMR (151 MHz, $CDCl_3$, ppm):** δ = 193.7 ((O)$_2$**C**=S), 162.5 (d, $^1J_{C\text{-}N}$ = 9.5 Hz, C-4), 153.4 ($\mathbf{C}_{Ar}$) 149.3 (t, $^1J_{C\text{-}N}$ = 18.5 Hz, 2-C), 139.4 (d, $^1J_{C\text{-}N}$ = 12.8 Hz, C-5), 129.6 (2 x $\mathbf{C}_{Ar}$), 126.7 ($\mathbf{C}_{Ar}$), 121.7 (2 x $\mathbf{C}_{Ar}$), 102.3 (d, $^2J_{C\text{-}N}$ = 7.5 Hz, C-6), 88.6 (dd, $^1J_{C\text{-}C}$ = 53.5 Hz, $^1J_{C\text{-}N}$ = 12.2 Hz, 1'-C), 83.7 (t, *J* = 41.9 Hz, 2'-C), 82.2 (t, $^1J_{C\text{-}C}$ = 42.3, 3'-C), 68.1 (t, *J* = 41.0 Hz, 4'-C), 59.4 (d, *J* = 43.1 Hz), 17.4 - 16.8 (Si-C($\mathbf{C}H_3$)), 13.4 - 12.4 (4 x Si-**C**(CH_3)). **^{15}N-NMR (41 MHz, $CDCl_3$, ppm):** δ = -223.5, -223.6. **HRMS (ESI$^+$):** calc. for $C_{23}{}^{13}C_5H_{43}{}^{15}N_2O_8SSi_2$+ $[M+H]^+$: 630.2382, found: 630.2393. **IR (ATR):** ν (cm^{-1}) = 2944 (w), 2866 (w), 2359 (w), 1722 (w), 1709 (m), 1691 (m), 1679 (m), 1649 (w), 1641 (w), 1631 (w), 1591 (w), 1462 (w), 1443 (w), 1382 (w), 1274 (m), 1202 (s), 1137 (w), 1098 (s), 1030 (s), 925 (m), 883 (m), 853 (m), 808 (m), 766 (m), 689 (s).

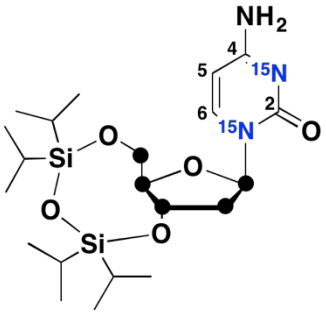
3',5'-(1,1,3,3,-tetraisopropyldisiloxane-1,3-diyl)-2'-deoxy-(1',2',3',4',5'-$^{13}C_5$, N^1,N^3-$^{15}N_2$)-uridine (26)



The deoxygenation of **25** was carried out as described in the literature with slight modifications.[5] First, **25** (260 mg, 0.41 mmol, 1.0 eq), AIBN (14 mg, 0.08 mmol, 0.2 eq) and tributyltin hydride (165 µL, 0.62 mmol, 1.5 eq) were dissolved in toluene (8 mL). The mixture was freeze-pump-thaw degassed (3 x) and then heated to 75 °C for 90 minutes. The solvent was removed *in vacuo* and the residue was taken up in water (25 mL). The aqueous layer was then extracted with DCM (3 x 25 mL) and the combined organic layers were dried over Na_2SO_4. Flash-chromatography of the crude (25% EtOAc/^{i}Hex --> 30% EtOAc/^{i}Hex) yielded **26** as a white foam (180 mg, 0.38 mmol, 93%).

^{1}H-NMR (599 MHz, $CDCl_3$, ppm): δ = 8.43 (d, $^2J_{N\text{-}H}$ = 91.1 Hz, 1H, N-**H**), 7.71 (dt, ^3J = 8.1 Hz, $^2J_{N\text{-}H}$ = 2.2 Hz, 1H, 5-H), 6.19 (dd, $^1J_{C\text{-}H}$ = 175 Hz, ^3J = 6.5 Hz, 1H, 1'-H), 5.70 - 5.67 (m, 1H, 6-H), 4.57 (d, $^1J_{C\text{-}H}$ = 142.1 Hz, 3'-H), 4.27 - 3.64 (m, 3H, 4'-H, 5'-H), 2.63 - 2.14 (m, 2H, 2'H), 1.10 - 1.01 (m, 28H, Si(C**H**$_3$)$_2$, Si-C**H**). **^{13}C-NMR (151 MHz, $CDCl_3$, ppm):** δ = 162.9 (d, $^1J_{C\text{-}N}$ = 9.4 Hz, C-4), 149.8 (t, $^1J_{C\text{-}N}$ = 18.5 Hz, 2-C), 139.6 (d, $^1J_{C\text{-}N}$ = 12.9 Hz, C-5), 101.6 (d, $^2J_{C\text{-}N}$ = 7.0 Hz, C-6), 85.1 (t, $^1J_{C\text{-}C}$ = 41.9 Hz, 4'-C), 84.3 (dd, $^1J_{C\text{-}C}$ = 36.4 Hz, $^1J_{C\text{-}N}$ = 10.6 Hz, 1'-C), 67.0 (t, $^1J_{C\text{-}C}$ = 39.2 Hz, 3'-C), 60.0 (d, J = 37.0 Hz, 5'-C), 39.9 (t, $^1J_{C\text{-}C}$ = 37 Hz, 2'-C), 17.5 - 16.7 (Si-C(**C**H$_3$)), 13.4 - 12.4 (4 x Si-**C**(CH$_3$)). **^{15}N-NMR (41 MHz, $CDCl_3$, ppm):** δ = -224.1, -224.2. **HRMS (ESI$^+$):** calc. for $C_{16}{}^{13}C_5H_{39}{}^{15}N_2O_6Si_2^+$ $[M+H]^+$: 478.2450, found: 478.2451. **HRMS (ESI$^-$):** calc. for $C_{16}{}^{13}C_5H_{37}{}^{15}N_2O_6Si_2^-$ $[M-H]^-$: 476.2304, found: 476.2306. **IR (ATR):** ν (cm^{-1}) = 2944 (w), 2866 (w), 1680 (s), 1446 (w), 1384 (w), 1246 (m), 1094 (m), 1065 (m), 1029 (s), 967 (m), 919 (w), 883 (s), 853 (m), 806 (m), 761 (m), 693 (s).

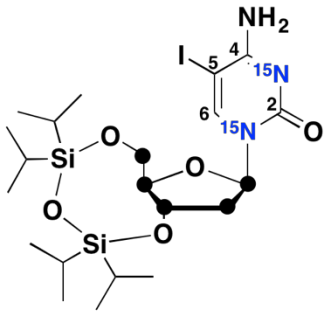
3',5'-(1,1,3,3,-tetraisopropyldisiloxane-1,3-diyl)-2'-deoxy-(1',2',3',4',5'-$^{13}C_5$, N^1,N^3-$^{15}N_2$)-cytidine (27)



For the amination reaction, triazole (398 mg, 5.76 mmol, 9.0 eq) was dissolved in MeCN (12 mL) and cooled to 0°C. Then, $POCl_3$ (120 µL, 1.27 mmol, 2.0 eq) was added dropwise and the milky suspension was kept at 0 °C for ten minutes. Triethyl amine (753 µL, 5.50 mmol, 8.6 eq) was subsequently given to the slurry and stirred for an additional 20 minutes. Finally, **26** (304 mg, 0.64 mmol, 1.0 eq) was added and the reaction was allowed to warm to room temperature overnight. After TLC showed complete consumption of the starting material, the crude was suspended in DCM (25 mL) and washed with brine (25 mL). The organic phase was dried over Na_2SO_4 and the volatiles were removed *in vacuo*. This residue was then diluted with dioxane (15 mL), charged with NH_4OH (4 mL, 25%) and stirred at 40 °C for three hours. The mixture was subsequently neutralized with saturated aqueous NH_4Cl (50 mL) and extracted with EtOAc (2 x 25 mL). Column chromatography (5% MeOH/DCM) yielded 147 mg (0.31 mmol, 48%) of **27** as a colourless glass.

^{1}H-NMR (800 MHz, $CDCl_3$, ppm): δ = 7.90 (ddd, ^3J = 7.3 Hz, $^2J_{N\text{-}H}$ = 2.7 Hz, 1.1 Hz, 1H, 5-H), 6.17 (d, $^1J_{C\text{-}H}$ = 174.7 Hz, ^3J = 7.7 Hz, 1H, 1'-H), 5.64 (dd, ^3J = 7.4 Hz, $^2J_{N\text{-}H}$ = 3.4 Hz, 6-H), 4.46 (d, $^1J_{C\text{-}H}$ = 137.8 Hz, 1H, 3'-H), 4.25 (d, $^1J_{C\text{-}H}$ = 149.2 Hz, 1H, 4'-H), 4.10 (d, $^1J_{C\text{-}H}$ = 139.7 Hz, 1H, 5'-H), 3.86 (d, $^1J_{C\text{-}H}$ = 143.6 Hz, 1H, 5'-H), 2.61 (d, $^1J_{C\text{-}H}$ = 134.2 Hz, 1H, 2'-H), 2.61 (d, $^1J_{C\text{-}H}$ = 134.2 Hz, 1H, 2'-H), 2.41 (d, $^1J_{C\text{-}H}$ = 143.2 Hz, 1H, 2'-H), 1.10 - 0.98 (m, 28H, Si(C**H**$_3$)$_2$, Si-C**H**). **^{13}C-NMR (201 MHz, $CDCl_3$, ppm):** δ = 164.6 (d, $^1J_{C\text{-}N}$ = 6.2 Hz, C-4), 154.6 (t, $^1J_{C\text{-}N}$ = 12.2 Hz, 2-C), 140.3 (d, $^1J_{C\text{-}N}$ = 12.8 Hz, C-5), 92.0 (d, $^2J_{C\text{-}N}$ = 7.0 Hz, C-6), 84.1 (d, $^1J_{C\text{-}C}$ = 40.4 Hz, 1'-C), 83.9 (t, $^1J_{C\text{-}C}$ = 42.8 Hz, 4'-C), 65.7 (t, $^1J_{C\text{-}C}$ = 36.9, 3'-C), 59.9 (d, J = 43.0 Hz, 5'-C), 38.9 (t, $^1J_{C\text{-}C}$ = 36.9 Hz, 2'-C), 16.5 - 15.8 (Si-C(**C**H$_3$)), 12.4 - 11.4 (4 x Si-**C**(CH$_3$)). **^{15}N-NMR (41 MHz, $CDCl_3$, ppm):** δ = -217.7, -217.9. **HRMS (ESI$^+$):** calc. for $C_{16}{}^{13}C_5H_{40}N^{15}N_2O_5Si_2^+$ $[M+H]^+$: 477.2609, found: 477.2610. **HRMS (ESI$^-$):** calc. for $C_{16}{}^{13}C_5H_{38}N^{15}N_2O_5Si_2^-$ $[M-H]^-$: 475.2464, found: 475.2464. **IR (ATR):** ν (cm^{-1}) = 3337 (w), 2944 (w), 2867 (w), 1622 (m), 1465 (m), 1394 (w), 1267 (w), 1180 (w), 1120 (m), 1095 (s), 1030 (s), 937 (w), 883 (s), 782 (m), 694 (s).

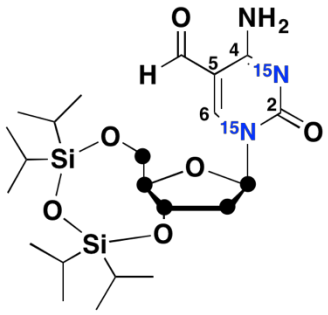
3',5'-(1,1,3,3,-tetraisopropyldisiloxane-1,3-diyl)-5-iodo-2'-deoxy-(1',2',3',4',5'-$^{13}C_5$, N^1,N^3-$^{15}N_2$)-cytidine (28**)**



Ceric ammonium nitrate (320 mg, 0.58 mmol, 2.0 eq), iodine (147 mg, 0.58 mmol, 2.0 eq) and **27** (139 mg, 0.29 mmol, 1.0 eq) were dissolved in MeCN (20 mL) and heated to 60 °C for two and a half hours. The dark brownish solution was then poured into a mixture of brine (25 mL) and saturated aqueous $Na_2S_2O_3$ solution (5 mL). After extraction with DCM (3 x 25 mL), the organic phases were dried over Na_2SO_4 and then removed *in vacuo*. Purification via column chromatography (4% MeOH/DCM) yielded **28** as an off white powder (149 mg, 0.25 mmol, 86%).

^{1}H-NMR (599 MHz, $CDCl_3$, ppm): δ = 8.07 (sd, $^2J_{N\text{-}H}$ = 2.4 Hz, 1H, 5-H), 5.98 (dd, $^1J_{C\text{-}H}$ = 175.8 Hz, ^3J = 6.8 Hz, 1H, 1'-H), 4.49 (d, $^1J_{C\text{-}H}$ = 126.2 Hz, 1H, 3'-H), 4.29 (d, $^1J_{C\text{-}H}$ = 146.0 Hz, 1H, 4'-H), 4.13 (d, $^1J_{C\text{-}H}$ = 137.9 Hz, 1H, 5'-H), 3.86 (d, $^1J_{C\text{-}H}$ = 136.7 Hz, 1H, 5'-H), 2.63 (d, $^1J_{C\text{-}H}$ = 133.4 Hz, 1H, 2'-H), 2.44 (d, $^1J_{C\text{-}H}$ = 136.0 Hz, 1H, 2'-H), 1.11 - 0.99 (m, 28H, Si(C**H**$_3$)$_2$, Si-C**H**). **^{13}C-NMR (151 MHz, $CDCl_3$, ppm):** δ = 163.7 (d, $^1J_{C\text{-}N}$ = 5.2 Hz, C-4), 154.6 (t, $^1J_{C\text{-}N}$ = 10.5 Hz, 2-C), 146.5 (d, $^1J_{C\text{-}N}$ = 12.6 Hz, C-5), 85.7 - 84.9 (1'-C + 4'-C), 66.5 (t, $^1J_{C\text{-}C}$ = 39.2, 3'-C), 59.7 (d, *J* = 43.2 Hz, 5'-C), 55.3 (C-6), 39.9 (t, $^1J_{C\text{-}C}$ = 37.0 Hz, 2'-C), 17.8 - 16.9 (8 x Si-C(**C**H$_3$)), 13.5 - 12.3 (4 x Si-**C**(CH$_3$)). **^{15}N-NMR (41 MHz, $CDCl_3$, ppm):** δ = -209.7, -209.9. **HRMS (ESI$^+$):** calc. for $C_{16}{}^{13}C_5H_{39}IN^{15}N_2O_5Si_2^+$ $[M+H]^+$: 603.1576, found: 603.1581. **HRMS (ESI$^-$):** calc. for $C_{16}{}^{13}C_5H_{37}IN^{15}N_2O_5Si_2^-$ $[M-H]^-$: 601.1430, found: 601.1427. **IR (ATR):** ν (cm^{-1}) = 3435 (w), 3328 (w), 2944 (w), 2867 (w), 1631 (m), 1600 (m), 1570 (m), 1463 (m), 1386 (w), 1262 (w), 1235 (w), 1155 (w), 1116 (w), 1094 (s), 1044 (m), 1023 (s), 696 (m), 937 (w), 919 (w), 884 (m), 871 (m), 852 (w), 787 (w), 775 (m), 697 (m).

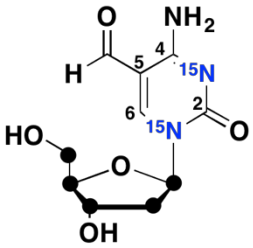
3',5'-(1,1,3,3,-tetraisopropyldisiloxane-1,3-diyl)-5-formyl-2'-deoxy-(1',2',3',4',5'-$^{13}C_5$, N^1,N^3-$^{15}N_2$)-cytidine (29)



A high pressure autoclave glass vessel was charged with PPh_3 (13 mg, 0.05 mmol, 0.6), $Pd_2(dba)_3CHCl_3$ (8 mg, 0.01 mmol, 0.1 eq) and **28** (48 mg, 0.08 mmol, 1.0 eq) and suspended in toluene (5 mL). The vessel was flushed twice with CO and then set to a final CO pressure of 50 psi. The mixture was then heated to 65 °C and tributyltin hydride (27 µL diluted with 300 µL toluene) was added through a syringe pump with an addition speed of 20 µL per hour. After the addition was finished, the volatiles were removed *in vacuo* and the yellowish crude mixture was purified by flash column chromatography (20% EtOAc/^{i}Hex --> 40% EtOAc/^{i}Hex --> 50% EtOAc/^{i}Hex), whereas 37 mg (0.073 mmol, 92%) of product **29** could be obtained as a light yellowish oil.

^{1}H-NMR (400 MHz, $CDCl_3$, ppm): δ = 9.50 (s, 1H, C**H**O), 8.53 (sd, $^2J_{N\text{-}H}$ = 2.7 Hz, 1H, 5-H), 6.01 (dd, $^1J_{C\text{-}H}$ = 177.6 Hz, ^3J = 6.7 Hz, 1H, 1'-H), 4.54 - 3.63 (m, 4H, 3'-H, 4'-H, 5'-H), 2.78 (d, $^1J_{C\text{-}H}$ = 135.1 Hz, 1H, 2'-H), 2.44 (d, $^1J_{C\text{-}H}$ = 137.5 Hz, 1H, 2'-H), 1.10 - 0.96 (m, 28H, Si(C**H**$_3$)$_2$, Si-C**H**). **^{13}C-NMR (151 MHz, $CDCl_3$, ppm):** δ = 187.5 (**C**HO), 162.9 (d, $^1J_{C\text{-}N}$ = 5.9 Hz, C-4), 153.3 (t, $^1J_{C\text{-}N}$ = 7.9 Hz, 2-C), 152.7 (d, $^1J_{C\text{-}N}$ = 14.3 Hz, C-5), 104.8 (C-6), 86.1 - 84.9 (1'-C + 4'-C), 65.9 (t, $^1J_{C\text{-}C}$ = 38.2, 3'-C), 59.4 (d, J = 42.9 Hz, 5'-C), 39.5 (t, $^1J_{C\text{-}C}$ = 36.9 Hz, 2'-C), 17.5 - 16.8 (8 x Si-C(**C**H$_3$)), 13.5 - 12.4 (4 x Si-**C**(CH$_3$)). **^{15}N-NMR (41 MHz, $CDCl_3$, ppm):** δ = -206.6, -206.8. **HRMS (ESI$^+$):** calc. for $C_{17}{}^{13}C_5H_{40}N^{15}N_2O_6Si_2^+$ $[M+H]^+$: 505.2559, found: 505.2567. **IR (ATR):** ν (cm^{-1}) = 3393 (w), 2944 (m), 2866 (m), 2361 (w), 1658 (s), 1498 (m), 1463 (m), 1437 (m), 1412 (w), 1387 (w), 1308 (w), 1228 (m), 1183 (m), 1160 (m), 1118 (s), 1092 (s), 1064 (s), 1030 (s), 970 (m), 939 (m), 919 (w), 883 (m), 778 (m), 749 (m), 720 (s), 693 (s).

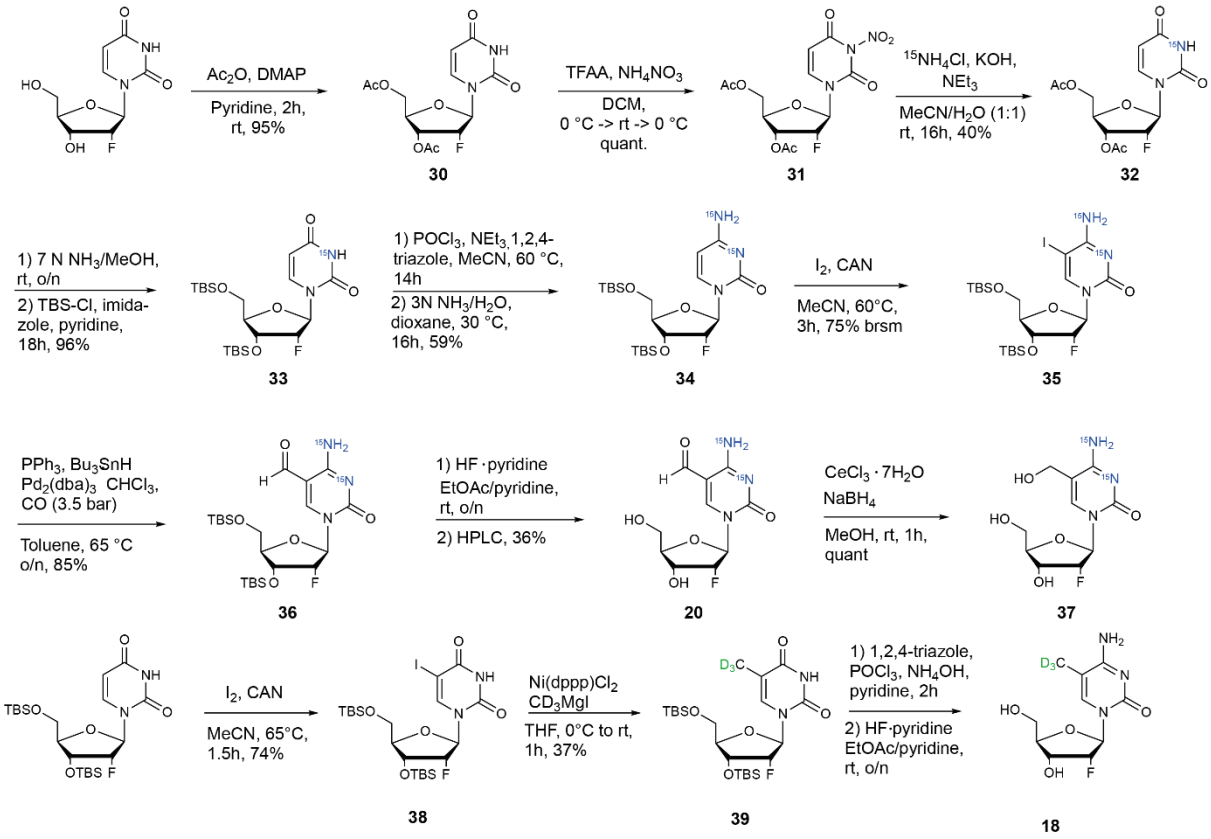
5-formyl-2'-deoxy-(1',2',3',4',5'-$^{13}C_5$, N^1,N^3-$^{15}N_2$)-cytidine (1)



29 (30 mg, 0.06 mmol, 1.0 eq) was dissolved in EtOAc (1 mL) and pyridine (24 µL) and HF•pyridine (16 µL, 0.6 mmol, 10 eq) were added. After the mixture was stirred for 12 hours at room temperature, a white precipitate formed. Excess of HF was quenched with TMSOMe (60 µL) and the colourless slurry was stirred for an additional 30 minutes. Then, the mixture was centrifuged (6000 rpm à 10 mins.), the supernatant removed and the colourless powder washed with EtOAc (5 mL) for three consecutive times. HPLC purification (0% --> 25% MeCN in 30 minutes) yielded **1** (7 mg, 0.026 mmol, 45%) as a colourless solid.

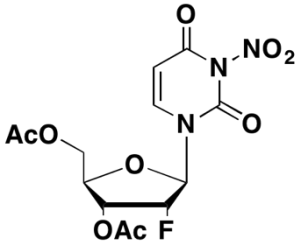
^{1}H-NMR (400 MHz, D_2O, ppm): δ = 9.50 (s, 1H, C**H**O), 8.84 (sd, $^2J_{N\text{-}H}$ = 2.9 Hz, 1H, 5-H), 6.15 (dt, $^1J_{C\text{-}H}$ = 175.4 Hz, ^3J = 5.8 Hz, 1H, 1'-H), 4.61 (d, $^1J_{C\text{-}H}$ = 151.9 Hz, 1H, 3'-H), 4.30 - 3.93 (m, 2H, 4'-H + 5'-H), 3.76 - 3.53 (m, 1H, 5'-H), 2.74 (d, $^1J_{C\text{-}H}$ = 136.0 Hz, 1H, 2'-H), 2.53 (d, $^1J_{C\text{-}H}$ = 134.5 Hz, 1H, 2'-H). **^{13}C-NMR (151 MHz, D_2O, ppm):** δ = 190.5 (**C**HO), 161.7 (d, $^1J_{C\text{-}N}$ = 7.4 Hz, C-4), 154.7 (d, $^1J_{C\text{-}N}$ = 14.2 Hz, C-5), 154.0 (t, $^1J_{C\text{-}N}$ = 11.3 Hz, 2-C), 105.4 (C-6), 87.8 - 86.5 (1'-C + 4'-C), 69.3 (t, $^1J_{C\text{-}C}$ = 36.5, 3'-C), 60.5 (d, J = 41.5 Hz, 5'-C), 40.0 (t, $^1J_{C\text{-}C}$ = 36.0 Hz, 2'-C). **^{15}N-NMR (41 MHz, D_2O, ppm):** δ = -209.9, -210.2. **HRMS (ESI$^+$):** calc. for $C_5{}^{13}C_5H_{13}N^{15}N_2NaO_5{}^+$ $[M+Na]^+$: 285.0856, found: 285.0857.

Synthesis of isotopically labelled 2'-(*R*)-fluorocytidine derivatives



Scheme 2: Synthesis of $^{15}N_2$-labelled 2'-(*R*)-F-fdC **20**, 2'-(*R*)-F-hmdC **37** and [D_3]-2'-*(R)*-F-mdC **18**.

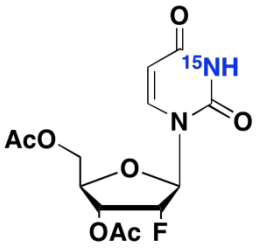
3',5'-di-*O*-acetyl-2'-deoxy-2'-(*R*)-fluoro-3-nitrouridine (31**)**



NH_4NO_3 (1.45 g, 18.1 mmol, 2.0 eq) was cautiously powdered with a mortar and pestle and subsequently suspended in DCM (50 mL). The resulting mixture was cooled to 0 °C and trifluoroacetic anhydride (5.10 g, 36.2 mmol, 4.0 eq) was added dropwise. After 30 minutes of warming up to room temperature, a brownish slurry developed which was stirred for another two hours until almost all of the remaining NH_4NO_3 was dissolved (note that complete dissolution of NH_4NO_3 could not be observed even after extended stirring). The resulting brownish-yellow solution was then again cooled to 0 °C and **30** (3.00 g, 9.06 mmol, 1.0 eq) was added in small portions. The reaction mixture was stirred for additional three hours at 0 °C and was quenched through the addition of PBS-Buffer (100 mL, pH = 7.4). The aqueous layer was extracted with DCM (3 x 75 mL), combined organic layers were dried over Na_2SO_4 and the solvent was removed *in vacuo*. The crude residue was purified by column chromatography (1% MeOH/DCM to 2.5% MeOH/DCM) and **31** (3.60 g, 9.59 mmol, quant.) was yielded as a colourless powder.

^{1}H-NMR (599 MHz, $CDCl_3$, ppm): δ = 7.47 (d, ^3J = 8.3 Hz, 1H, 6-H), 5.89 (d, ^3J = 8.3 Hz, 1H, 5-H), 5.81 (dd, $^3J_{H\text{-}F}$ = 17.2 Hz, ^3J = 1.4 Hz, 1H, 1'-H), 5.41 (ddd, $^2J_{H\text{-}F}$ = 51.6 Hz, ^3J = 4.8 Hz, ^3J = 1.4 Hz, 1H, 2'-H), 5.13 (ddd, $^3J_{H\text{-}F}$ = 18.0 Hz, ^3J = 4.8 Hz, ^3J = 3.5 Hz, 1H, 3'-H), 4.47 - 4.43 (m, 2H, 5'-H + 4'-H), 4.33 - 4.30 (m, 1H, 5'H), 2.16 (s, 3H, C3'-O-(C=O)-C**H$_3$**), 2.10 (s, 3H, C5'-O-(C=O)-C**H$_3$**). **^{13}C-NMR (151 MHz, $CDCl_3$, ppm):** δ = 170.2 (C5'-O-(**C**=O)-CH_3), 169.9 (C3'-O-(**C**=O)-CH_3), 154.9 (C-4), 145.1 (C-2), 140.1 (C-6), 102.0 (C-5), 92.0 (d, $^2J_{C\text{-}F}$ = 37.3 Hz, C-1'), 90.8 (d, $^1J_{C\text{-}F}$ = 193.1 Hz, C-2'), 79.3 (C-4'), 69.3 (d, $^2J_{C\text{-}F}$ = 16.0 Hz, C-3'), 61.8 (C-5'), 20.8 (C5'-O-(C=O)-**C**H_3), 20.5 (C3'-O-(C=O)-**C**H_3). **^{19}F-NMR (377 MHz, $CDCl_3$, ppm):** δ = −199.4 (dt, $^2J_{F\text{-}H}$ = 51.8 Hz, $^3J_{F\text{-}H}$ = 18.4 Hz). **HRMS (ESI$^+$):** calc. for $C_{13}H_{18}FN_4O_9^+$ $[M+NH_4]^+$: 393.1052, found: 393.1048. **IR (ATR):** ν (cm^{-1}) = 3096 (w), 1737 (s), 1693 (s), 1644 (s), 1617 (m), 1558 (w), 1540 (w), 1506 (w), 1456 (w), 1436 (m), 1377 (m), 1289 (m), 1252 (m), 1224 (s), 1119 (m), 1100 (s), 1055 (s), 990 (m), 960 (m), 901 (m), 884 (m), 852 (w), 828 (m), 804 (m), 748 (m), 738 (m), 705 (w), 677 (w). **Melting Range:** 132 °C - 134 °C.

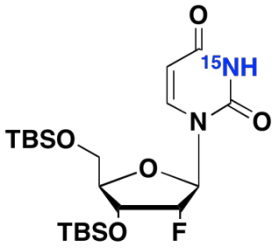
3',5'-di-*O*-acetyl-2'-deoxy-2'-(*R*)-fluoro-(N^3-^{15}N)-uridine (32)



First, $^{15}NH_4Cl$ (400 mg, 7.34 mmol, 1.2 eq) was dissolved in H_2O (15 mL) and KOH (412 mg, 7.34 mmol, 1.2 eq) and NEt_3 (1.50 mL, 10.0 mmol, 1.6 eq) were added. **31** (2.30 g, 6.13 mmol, 1.0 eq) was dissolved in MeCN (15 mL) and also given to the clear solution, which was then vigorously stirred at room temperature overnight. Subsequently, MeCN was removed *in vacuo* and the resulting aqueous phase was neutralized through the addition of sat. NH_4Cl (100 mL). The aqueous layers were extracted with DCM (3 x 100 mL), organic layers were collected and dried over Na_2SO_4. The volatiles were removed under reduced pressure and the crude mixture was purified by column chromatography (DCM --> 2% MeOH/DCM). **32** (818 mg, 2.47 mmol, 40%) was yielded as a white foam.

^{1}H-NMR (800 MHz, DMSO-d_6, ppm): δ = 7.71 (d, ^3J = 8.1 Hz, 1H, 6-H), 5.87 (dd, $^3J_{H\text{-}F}$ = 22.4 Hz, ^3J = 2.0 Hz, 1H, 1'-H), 5.69 - 5.68 (m, 1H, 5-H), 5.33 (ddd, $^2J_{H\text{-}F}$ = 52.5 Hz, ^3J = 5.3 Hz, ^3J = 2.0 Hz, 1H, 2'-H), 5.26 (ddd, $^3J_{H\text{-}F}$ = 25.2 Hz, ^3J = 5.3 Hz, ^3J = 2.8 Hz, 1H, 3'-H), 4.34 (dd, ^2J = 12.4 Hz, ^3J = 3.0 Hz, 1H, 5'-H), 4.27 (m, 1H, 4'-H), 4.16 (dd, ^2J = 12.3 Hz, ^3J = 5.8 Hz, 1H, 5'-H), 2.11 (s, 3H, C3'-O-(C=O)-C**H$_3$**), 2.04 (s, 3H, C5'-O-(C=O)-C**H$_3$**). **^{13}C-NMR (202 MHz, DMSO-d_6, ppm):** δ = 170.2 (C5'-O-(**C**=O)-CH_3), 169.5 (C3'-O-(**C**=O)-CH_3), 163.2 (sd, $^1J_{C\text{-}N}$ = 8.7 Hz, C-4), 150.1 (sd, $^1J_{C\text{-}N}$ = 17.6 Hz, C-2), 142.9 (C-6), 102.1 (C-5), 90.8 (d, $^1J_{C\text{-}F}$ = 186.3 Hz, C-2'), 90.5 (d, $^2J_{C\text{-}F}$ = 37.0 Hz, C-1'), 77.9 (C-4'), 69.8 (d, $^2J_{C\text{-}F}$ = 14.3 Hz, C-3'), 62.7 (C-5'), 20.5 (C5'-O-(C=O)-**C**H_3), 20.3 (C3'-O-(C=O)-**C**H_3). **^{19}F-NMR (377 MHz, DMSO-d_6, ppm):** δ = -198.0 (ddd, $^2J_{H\text{-}F}$ = 52.5 Hz, $^3J_{H\text{-}F}$ = 22.3, $^3J_{H\text{-}F}$ = 17.2). **^{15}N-NMR (41 MHz, DMSO-d_6, ppm):** δ = -222.0. **HRMS (ESI$^+$):** calc. for $C_{13}H_{19}FN_2{}^{15}NO_7^+$ $[M+NH_4]^+$: 349.1172, found: 349.1174. **HRMS (ESI$^-$):** calc. for $C_{13}H_{14}FN^{15}NO_7^-$ $[M-H]^-$: 330.0761, found: 330.0763. **IR (ATR):** ν (cm^{-1}) = 3092 (w), 3002 (w), 2929 (w), 2878 (w), 2817 (w), 1737 (s), 1716 (s), 1672 (s), 1464 (w), 1449 (w), 1429 (w), 1417 (m), 1380 (m), 1365 (m), 1335 (w), 1299 (w), 1266 (m), 1236 (s), 1120 (w), 1095 (m), 1073 (s), 1048 (s), 995 (m), 968 (m), 900 (s), 875 (m), 865 (m), 825 (s), 763 (m), 751 (m), 732 (w), 699 (w), 684 (w). **Melting Range:** 169 °C - 173 °C.

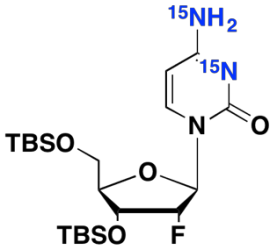
3',5'-bis-*O*-(*tert*-butyl(dimethyl)silyl)-2'-deoxy-2'-(*R*)-fluoro-(N^3-^{15}N)-uridine (33)



For the deprotection of **32** (800 mg, 2.42 mmol, 1.0 eq), the compound was dissolved in methanolic ammonia (7N, 5 mL) and stirred at room temperature overnight. The volatiles were then removed *in vacuo* and the mixture was dried on high vac. The resulting crude yellow oil was dissolved in pyridine (10 mL), imidazole (741 mg, 10.9 mmol, 4.5 eq) and TBSCl (1.09 g, 7.26 mmol, 3.0 eq) were added and the mixture was stirred at room temperature for 16 hours. After pyridine was removed by rotary evaporation, the residue was suspended in sat. $NaHCO_3$ (100 mL) and extracted with DCM (3 x 75 mL) and the combined organic layers were dried over Na_2SO_4. The obtained oily compound was then coevaporated with toluene (3 x 25 mL), whereas 1.11 g (2.33 mmol, 96%) of **33** were yielded as a colourless waxy solid.

^{1}H-NMR (599 MHz, $CDCl_3$, ppm): δ = 7.92 (d, ^3J = 8.2 Hz, 1H, 6-H), 6.06 (dd, $^3J_{H\text{-}F}$ = 15.3 Hz, ^3J = 2.0 Hz, 1H, 1'-H), 5.69 (dd, ^3J = 8.2 Hz, $^3J_{H\text{-}N}$ = 2.6 Hz, 1H, 5-H), 4.77 (ddd, $^2J_{H\text{-}F}$ = 52.3 Hz, ^3J = 4.3 Hz, ^3J = 2.0 Hz, 1H, 2'-H), 4.29 (ddd, $^3J_{H\text{-}F}$ = 18.9 Hz, ^3J = 7.2 Hz, ^3J = 4.2 Hz, 1H, 3'-H), 4.08 (dd, ^3J = 7.2 Hz, ^3J = 1.6 Hz, 1H, 4'-H), 4.05 (dd, ^2J = 11.7 Hz, ^3J = 2.1 Hz, 1H, 5'-H), 3.78 (dd, ^2J = 11.7 Hz, ^3J = 1.8 Hz, 1H, 5'-H), 0.93 - 0.90 (m, 18H, $Si(CH_3)_2(^t\mathbf{Bu})$), 0.12 - 0.10 (m, 12H, $Si(C\mathbf{H_3})_2(^t$Bu)). **^{13}C-NMR (151 MHz, $CDCl_3$, ppm):** δ = 163.1 (d, $^1J_{C\text{-}N}$ = 8.7 Hz, 4-C), 150.0 (sd, $^1J_{C\text{-}N}$ = 17.8 Hz, 2-C), 139.9 (C-6), 102.5 (C-5), 93.2 (d, $^1J_{C\text{-}F}$ = 193.3 Hz, C-2'), 87.9 (d, $^2J_{C\text{-}F}$ = 33.8 Hz, C-1'), 83.9 (C-4'), 68.7 (d, $^2J_{C\text{-}F}$ = 16.2 Hz, C-3'), 60.9 (C-5'), 26.1 ($Si(C(\mathbf{C}H_3)_3)$), 25.7 ($Si(C(\mathbf{C}H_3)_3)$), 18.5 ($Si(\mathbf{C}(CH_3)_3)$), 18.2 ($Si(\mathbf{C}(CH_3)_3)$), -4.5 ($Si(\mathbf{C}H_3)_2{}^t$Bu), -4.9 ($Si(\mathbf{C}H_3)_2{}^t$Bu), -5.3 ($Si(\mathbf{C}H_3)_2{}^t$Bu), -5.4 ($Si(\mathbf{C}H_3)_2{}^t$Bu). **^{19}F-NMR (377 MHz, $CDCl_3$, ppm):** δ = -202.5 (ddd, $^2J_{H\text{-}F}$ = 52.4 Hz, $^3J_{H\text{-}F}$ = 19.0, $^3J_{H\text{-}F}$ = 15.3). **^{15}N-NMR (41 MHz, $CDCl_3$, ppm):** δ = -224.4. **HRMS (ESI^+):** calc. for $C_{21}H_{40}FN^{15}NO_5Si_2^+$ $[M]^+$: 476.2425, found: 476.2429. **HRMS (ESI^-):** calc. for $C_{21}H_{38}FN^{15}NO_5Si_2^-$ $[M\text{-}H]^-$: 474.2279, found: 474.2291. **IR (ATR):** ν (cm^{-1}) = 3381 (w), 3254 (w), 2953 (w), 2927 (m), 2856 (m), 1749 (w), 1713 (s), 1707 (s), 1692 (s), 1681 (s), 1629 (w), 1471 (w), 1452 (m), 1398 (w), 1388 (w), 1362 (w), 1321 (w), 1362 (m), 1321 (w), 1276 (m), 1251 (m), 1230 (w), 1157 (m), 1120 (s), 1094 (m), 1072 (m), 1052 (m), 993 (m), 978 (m), 938 (w), 899 (w), 880 (m), 863 (m), 828 (s), 811 (m), 802 (s), 778 (s), 757 (s), 748 (m), 711 (m), 704 (m), 663 (m).

3',5'-bis-*O*-[*tert*-butyl(dimethyl)silyl]-2'-deoxy-2'-(*R*)-fluoro-(N^3,N^4-$^{15}N_2$)-cytidine (34)

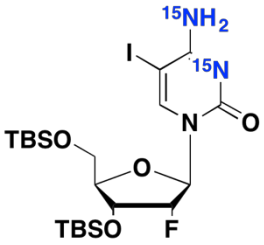


In a round bottom flask, 1,2,4-triazole (1.04 g, 15.1 mmol, 9.9 eq) was dissolved in MeCN (30 mL), cooled to 0 °C and stirred for ten minutes. To the resulting turbid solution, NEt_3 (2.00 mL, 14.4 mmol, 9.4 eq) was added dropwise and the mixture was stirred at 0 °C for another 20 minutes. Subsequently, a solution of **33** (730 mg, 1.53 mmol, 1.0 eq) in MeCN (10 mL) was given to the reaction and heated to 60 °C for 14 hours. After the starting material was consumed (as judged by TLC analysis), the mixture was quenched through addition of saturated $NaHCO_3$ (100 mL) and the resulting aqueous layer was extracted with DCM (3 x 100 mL). Combined organic layers were dried over Na_2SO_4 and the solvents were removed under reduced pressure. The resulting crude triazole derivative of **26** was then resuspended in dioxane (3.0 mL) and an aqueous solution of $^{15}NH_3$ (3N, 3.0 mL) was added. After the reaction was stirred at room temperature for 24 hours, starting material was still visible on the TLC. Thus, the heat was slightly increased to 30 °C and stirred for another 16 hours. The reaction was stopped through addition of saturated aqueous solution of NH_4Cl (50 mL), extracted with DCM (3 x 50 mL) and the combined organic layers were backwashed with brine (100 mL) and subsequently dried over Na_2SO_4. Purification through column chromatography (DCM --> 2% MeOH/DCM --> 3% MeOH/DCM --> 4% MeOH/DCM) yielded **34** (430 mg, 0.90 mmol, 59%) as a colourless foam.

^{1}H-NMR (800 MHz, $CDCl_3$, ppm): δ = 8.07 (d, ^3J = 7.5 Hz, 1H, 6-H), 6.03 (dd, $^3J_{H\text{-}F}$ = 15.9 Hz, ^3J = 0.9 Hz, 1H, 1'-H), 5.80 (d, ^3J = 7.5 Hz, 1H, 5-H), 4.79 (dd, $^2J_{H\text{-}F}$ = 52.2 Hz, ^3J = 3.9 Hz, 1H, 2'-H), 4.23 (ddd, $^3J_{H\text{-}F}$ = 22.6 Hz, ^3J = 8.6 Hz, ^3J = 3.9 Hz, 1H, 3'-H), 4.12 - 4.07 (m, 2H, 4'-H + 5'-H), 3.79 (dd, ^2J = 11.9 Hz, ^3J = 1.8 Hz, 1H, 5'-H), 0.94 (s, 9 H, $Si(CH_3)_2(^t\mathbf{Bu})$), 0.89 (s, 9 H, $Si(CH_3)_2(^t\mathbf{Bu})$), 0.13 (s, 3H, $Si(C\mathbf{H_3})_2(^t$Bu)), 0.12 (s, 3H, $Si(C\mathbf{H_3})_2(^t$Bu)), 0.10 (s, 3H, $Si(C\mathbf{H_3})_2(^t$Bu)), 0.08 (s, 3H, $Si(C\mathbf{H_3})_2(^t$Bu)). **^{13}C-NMR (201 MHz, $CDCl_3$, ppm):** δ = 165.0 (dd, $^1J_{C\text{-}N}$ = 21.8 Hz, $^1J_{C\text{-}N}$ = 8.54 Hz 4-C), 154.9 (sd, $^1J_{C\text{-}N}$ = 9.2 Hz, 2-C), 141.5 (C-6), 94.5 (C-5), 93.4 (d, $^1J_{C\text{-}F}$ = 192.3 Hz, C-2'), 88.9 (d, $^2J_{C\text{-}F}$ = 33.4 Hz, C-1'), 83.1 (C-4'), 68.0 (d, $^2J_{C\text{-}F}$ = 17.3 Hz, C-3'), 60.4 (C-5'), 26.1 ($Si(C(\mathbf{C}H_3)_3)$), 25.8 ($Si(C(\mathbf{C}H_3)_3)$), 18.6 ($Si(\mathbf{C}(CH_3)_3)$), 18.2 ($Si(\mathbf{C}(CH_3)_3)$), -4.5 ($Si(\mathbf{C}H_3)_2{}^t$Bu), -4.9 ($Si(\mathbf{C}H_3)_2{}^t$Bu), -5.3 ($Si(\mathbf{C}H_3)_2{}^t$Bu), -5.4 ($Si(\mathbf{C}H_3)_2{}^t$Bu). **^{19}F-NMR (376 MHz, $CDCl_3$, ppm):** δ = -201.3 (ddd, $^2J_{H\text{-}F}$ = 51.8 Hz, $^3J_{H\text{-}F}$ = 20.8, $^3J_{H\text{-}F}$ = 15.6). **^{15}N-NMR (41 MHz, $CDCl_3$, ppm):** δ = -288.1, -288.3. **HRMS (ESI^+):** calc. for $C_{21}H_{41}FN^{15}N_2O_4Si_2^+$ $[M]^+$: 476.2555, found: 476.2558. **HRMS (ESI^-):** calc. for $C_{21}H_{39}FN^{15}N_2O_4Si_2^-$ $[M\text{-}H]^-$: 474.2409, found: 474.2411. **IR (ATR):** ν (cm^{-1}) = 3170 (bw), 2954 (w),

2929 (w), 2857 (w), 1717 (w), 1645 (m), 1519 (w), 1506 (m), 1394 (w), 1362 (w), 1278 (w), 1253 (m), 1157 (m), 1157 (w), 1120 (m), 1071 (m), 991 (w), 881 (m), 860 (m), 835 (s), 812 (m), 777 (s), 670 (m), 592 (m).

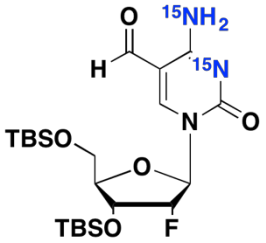
3',5'-bis-*O*-(*tert*-butyl(dimethyl)silyl)-5-iodo-2'-deoxy-2'-(*R*)-fluoro-(N^3,N^4-$^{15}N_2$)-cytidine (35)



In a round bottom flask, **34** (30 mg, 0.06 mmol, 1.0 eq) was dissolved in MeCN (5 mL) and ceric ammonium nitrate (71 mg, 0.13 mmol, 2.2 eq) and iodine (32 mg, 0.13 mmol, 2.2 eq) were added. The dark brownish mixture was then stirred at 60 °C for a total time of three hours and after the reaction was cooled to room temperature, MeCN was removed *in vacuo*. Addition of DCM (20 mL) lead to the precipitation of cerium salts, which were filtered off and the DCM was finally removed under reduced pressure. Column chromathography (5% MeOH/DCM) of the crude mixture yielded **35** (37 mg, 60 µmol, quant.) as an orange oil.

^{1}H-NMR (400 MHz, $CDCl_3$, ppm): δ = 7.99 (s, 1H, 6-H), 5.98 (dd, $^3J_{H\text{-}F}$ = 16.0 Hz, ^3J = 1.8 Hz, 1H, 1'-H), 4.81 (ddd, $^2J_{H\text{-}F}$ = 52.3 Hz, ^3J = 4.4 Hz, ^3J = 1.8 Hz, 1H, 2'-H), 4.19 (ddd, $^3J_{H\text{-}F}$ = 18.9 Hz, ^3J = 7.6 Hz, ^3J = 4.4 Hz, 1H, 3'-H), 4.09 - 4.01 (m, 2H, 4'-H + 5'-H), 3.78 (dd, ^2J = 11.9 Hz, ^3J = 2.1 Hz, 1H, 5'-H), 0.95 (s, 9 H, C3'/C5'-O-Si(CH_3)$_2$(^{t}**Bu**)), 0.89 (s, 9 H, Si(CH_3)$_2$(^{t}**Bu**)), 0.16 (s, 3H, Si(C**H$_3$**)$_2$(^{t}Bu)), 0.15 (s, 3H, Si(C**H$_3$**)$_2$(^{t}Bu)), 0.10 (s, 3H, Si(C**H$_3$**)$_2$(^{t}Bu)), 0.08 (s, 3H, Si(C**H$_3$**)$_2$(^{t}Bu)). **^{13}C-NMR (101 MHz, $CDCl_3$, ppm):** δ = 164.0 (dd, $^1J_{C\text{-}N}$ = 23.0 Hz, $^1J_{C\text{-}N}$ = 6.0 Hz, 4-C), 154.5 (sd, $^1J_{C\text{-}N}$ = 8.9 Hz, 2-C), 146.4 (C-6), 57.0 (C-5), 93.1 (d, $^1J_{C\text{-}F}$ = 193.3 Hz, C-2'), 89.2 (d, $^2J_{C\text{-}F}$ = 33.0 Hz, C-1'), 83.9 (C-4'), 68.8 (d, $^2J_{C\text{-}F}$ = 16.5 Hz, C-3'), 61.1 (C-5'), 57.4 (C-5), 26.5 (Si(C(**C**H$_3$)$_3$)), 25.8 (Si(C(**C**H$_3$)$_3$)), 18.9 (Si(**C**(CH$_3$)$_3$)), 18.2 (Si(**C**(CH$_3$)$_3$)), -4.4 (Si(**C**H$_3$)$_2$^{t}Bu), -4.7 (Si(**C**H$_3$)$_2$^{t}Bu), -4.8 (Si(**C**H$_3$)$_2$^{t}Bu), -4.9 (Si(**C**H$_3$)$_2$^{t}Bu). **^{19}F-NMR (376 MHz, $CDCl_3$, ppm):** δ = -201.7 (ddd, $^2J_{H\text{-}F}$ = 52.3 Hz, $^3J_{H\text{-}F}$ = 19.0, $^3J_{H\text{-}F}$ = 16.8). **^{15}N-NMR (41 MHz, $CDCl_3$, ppm):** δ = -280.4, -280.5. **HRMS (ESI$^+$):** calc. for $C_{21}H_{40}FIN^{15}N_2O_4Si_2^+$ [M]$^+$: 602.1521, found: 602.1530. **HRMS (ESI$^-$):** calc. for $C_{21}H_{38}FIN^{15}N_2O_4Si_2^{+-}$ [M-H]$^-$: 600.1376, found: 600.1378. **IR (ATR):** ν (cm^{-1}) = 3305 (w), 2955 (w), 2916 (s), 2850 (s), 1730 (s), 1635 (m), 1470 (s), 1418 (w), 1382 (w), 1361 (w), 1276 (m), 1253 (m), 1220 (m), 1178 (s), 1123 (m), 1105 (m), 1062 (m), 1047 (m), 992 (m), 942 (w), 889 (m), 864 (m), 836 (s), 777 (s), 740 (m), 719 (m), 670 (m), 643 (m), 622 (m), 593 (m).

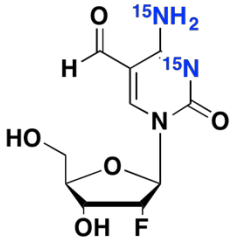
3',5'-bis-*O*-(*tert*-butyl(dimethyl)silyl)5-formyl-2'-deoxy-2'-(*R*)-fluoro-(N^3,N^4-$^{15}N_2$)-cytidine (36**)**



The reaction vessel of a high pressure CO-autoclave was charged with **35** (220 mg, 0.37 mmol, 1.0 eq), PPh_3 (58 mg, 0.2 mmol, 0.6 eq), $Pd_2(dba)_3CHCl_3$ (38 mg, 0.04 mmol 0.1 eq) and toluene (5.5 mL). The mixture was purged with CO twice and finally set under a pressurized atmosphere of CO (50 psi). After heating to 65 °C, tributyltin hydride (120 µL, 0.44 mmol, 1.2 eq) was added to the reaction through a syringe pump (17 µL/h). After the addition was finished, the mixture was cooled to room temperature and the volatiles were removed under reduced pressure. The crude was subjected to purification by column chromatography (20% EtOAc/ ^{i}Hex --> 50% EtOAc/ ^{i}Hex) and **36** (158 mg, 0.31 mmol, 85%) was obtained as a yellow foam.

^{1}H-NMR (599 MHz, $CDCl_3$, ppm): δ = 9.50 (s, 1H, C**H**O), 8.53 (s, 1H, 6-H), 8.26 (dd, $^1J_{H\text{-}N}$ = 92.9 Hz, ^1J = 6.0 Hz, 1H, $^{15}NH_2$), 7.29 (d, $^1J_{H\text{-}N}$ = 92.9 Hz, 1H, $^{15}NH_2$), 6.02 (dd, $^3J_{H\text{-}F}$ = 16.3 Hz, ^3J = 2.0 Hz, 1H, 1'-H), 4.92 (dd, $^2J_{H\text{-}F}$ = 51.9 Hz, ^3J = 4.1 Hz, 1H, 2'-H), 4.23 - 4.14 (m, 3H, 3'-H, 4'-H, 5'-H), 3.83 (dd, ^2J = 12.2 Hz, ^3J = 2.1 Hz, 1H, 5'-H), 0.94 (s, 9 H, Si(CH$_3$)$_2$(**^{t}Bu**)), 0.89 (s, 9 H, Si(CH$_3$)$_2$(**^{t}Bu**)), 0.14 (s, 3H, Si(C**H$_3$**)$_2$(^{t}Bu)), 0.11 (s, 3H, Si(C**H$_3$**)$_2$(^{t}Bu)), 0.10 (s, 3H, Si(C**H$_3$**)$_2$(^{t}Bu)), 0.08 (s, 3H, Si(C**H$_3$**)$_2$(^{t}Bu)). **^{13}C-NMR (151 MHz, $CDCl_3$, ppm):** δ = 187.3 (**C**HO), 162.9 (dd, $^1J_{C\text{-}N}$ = 20.9 Hz, $^1J_{C\text{-}N}$ = 5.1 Hz, C-4), 153.2 (C-6), 153.0 (C-2), 105.7 (C-5), 93.0 (d, $^1J_{C\text{-}F}$ = 192.9 Hz, C-2'), 89.8 (d, $^2J_{C\text{-}F}$ = 34.4 Hz, C-1'), 83.8 (C-4'), 68.2 (d, $^2J_{C\text{-}F}$ = 16.5 Hz, C-3'), 60.7 (C-5'), 26.2 (Si(C(**C**H$_3$)$_3$)), 25.7 (Si(C(**C**H$_3$)$_3$)), 18.8 (Si(**C**(CH$_3$)$_3$)), 18.2 (Si(**C**(CH$_3$)$_3$)), -4.3 (Si(**C**H$_3$)$_2$^{t}Bu), -4.8 (Si(**C**H$_3$)$_2$(C(CH$_3$)$_3$)), -4.8 (Si(**C**H$_3$)$_2$^{t}Bu), -5.2 (Si(**C**H$_3$)$_2$^{t}Bu). **^{19}F-NMR (377 MHz, $CDCl_3$, ppm):** δ = -200.9 (ddd, $^2J_{H\text{-}F}$ = 51.7 Hz, $^3J_{H\text{-}F}$ = 19.9, $^3J_{H\text{-}F}$ = 15.7). **^{15}N-NMR (41 MHz, $CDCl_3$, ppm):** δ = -175.2, -286.0. **HRMS (ESI$^+$):** calc. for $C_{22}H_{41}FN^{15}N_2O_5Si_2^+$ [M]$^+$: 504.2504, found: 504.2513. **HRMS (ESI$^-$):** calc. for $C_{22}H_{39}FN^{15}N_2O_5Si_2^-$ [M-H]$^-$: 502.2358, found: 502.2357. **IR (ATR):** ν (cm^{-1}) = 3383 (bw), 2954 (w), 2929 (w), 2857 (w), 1663 (s), 1636 (m), 1506 (m), 1472 (w), 1418 (w), 1361 (w), 1252 (m), 1237 (m), 1167 (w), 1120 (m), 1070 (s), 993 (m), 979 (w), 835 (s), 813 (m), 777 (s), 699 (w), 667 (m), 598 (m), 575 (w).

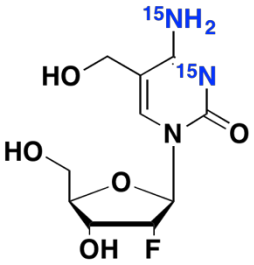
5-formyl-2'-deoxy-2'-(*R*)-fluoro-(N^3,N^4-$^{15}N_2$)-cytidine (20**)**



For the TBS deprotection, **36** (100 mg, 0.20 mmol, 1.0 eq) was dissolved in a mixture of EtOAc (2.8 mL) and pyridine (0.8 mL). HF•pyridine (52 µL, 2.00 mmol, 10 eq) was added, the reaction was stirred for 16 hours at room temperature and a colourless precipitate formed. To quench the reaction, methoxytrimethyl silane (200 µL) was given to the mixture and stirred for another two hours at room temperature, whereas more of the colourless precipitate formed. The product was isolated through repeated washing with EtOAc after centrifuging (6000 rpm, three times in total) of the suspension. **20** (32 mg, 0.12 mmol, 58%) were yielded as a slightly yellowish powder, which was finally purified twice by semipreparative HPLC (Waters Nucleodur C18 ec, 0% MeCN/H_2O --> 13% MeCN/H_2O in 30 minutes, 5.0 mL/min) for cell-feeding purposes.

^{1}H-NMR (800 MHz, D_2O, ppm): δ = 9.54 (s, 1H, C**H**O), 8.90 (s, 1H, 6-H), 6.10 (d, $^3J_{H\text{-}F}$ = 17.7 Hz, 1H, 1'-H), 5.22 (dd, $^2J_{H\text{-}F}$ = 52.3 Hz, ^3J = 4.3 Hz, 1H, 2'-H), 4.39 (ddd, $^3J_{H\text{-}F}$ = 24.2 Hz, ^3J = 9.3 Hz, ^3J = 4.3 Hz, 1H, 3'-H) 4.37 (dt, ^3J = 9.3 Hz, ^3J = 3.0 Hz, 1H, 4'-H), 4.15 (dd, ^2J = 13.1 Hz, ^3J = 2.4 Hz, 1H, 5'-H), 3.94 (dd, ^2J = 13.1 Hz, ^3J = 3.5 Hz, 1H, 5'-H). **^{13}C-NMR (201 MHz, D_2O, ppm):** δ = 190.3 (**C**HO), 162.8 (dd, $^1J_{C\text{-}N}$ = 21.8 Hz, $^1J_{C\text{-}N}$ = 6.5 Hz, C-4), 155.0 (C-6), 154.9 (C-2), 105.9 (C-5), 93.7 (d, $^1J_{C\text{-}F}$ = 186.0 Hz, C-2'), 89.9 (d, $^2J_{C\text{-}F}$ = 34.9 Hz, C-1'), 82.5 (C-4'), 67.2 (d, $^2J_{C\text{-}F}$ = 16.7 Hz, C-3'), 59.1 (C-5'). **^{19}F-NMR (376 MHz, D_2O, ppm):** δ = -201.6 (ddd, $^2J_{H\text{-}F}$ = 52.3 Hz, $^3J_{H\text{-}F}$ = 24.0, $^3J_{H\text{-}F}$ = 17.7). **^{15}N-NMR (41 MHz, D_2O, ppm):** δ = -199.9, -200.4. **HRMS (ESI^+):** calc. for $C_{10}H_{13}FN^{15}N_2O_5^+$ $[M+H]^+$: 276.0774, found: 276.0778. **HRMS (ESI^-):** calc. for $C_{10}H_{11}FN^{15}N_2O_5^-$ $[M-H]^-$: 274.0629, found: 274.0629.

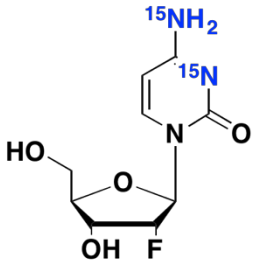
5-hydroxymethyl-2'-deoxy-2'-(*R*)-fluoro-(N^3,N^4-$^{15}N_2$)-cytidine (37)



First, **20** (1.00 mg, 3.60 µmol, 1.0 eq) was dissolved in MeOH (1.0 mL), $CeCl_3 \bullet 7\ H_2O$ (4.10 mg, 10.9 µmol, 3.0 eq) and $NaBH_4$ (0.20 mg, 3.60 µmol, 1.0 eq) were added subsequently, whereas an evolution of gas could be observed. After stirring at room temperature for one hour, complete conversion to the desired product could be observed (as judged by LC-MS). Then, a saturated aqueous solution of NH_4Cl (1.0 mL) was given to the crude and the solvent was removed by lyophillization. Purification of **37** was achieved through semipreparative HPLC (Waters Nucleodur C18 ec, 0% MeCN/H_2O --> 10% MeCN/H_2O in 45 minutes, 5.0 mL/min) and yielded 0.80 mg (2.80 µmol, 80%) of pure **37** as a white powder.

^{1}H-NMR (800 MHz, D_2O, ppm): δ = 7.85 (s, 1H, 6-H), 5.92 (dd, $^3J_{H\text{-}F}$ = 19.4 Hz, ^3J = 1.1 Hz, 1H, 1'-H), 5.07 (dd, $^2J_{H\text{-}F}$ = 52.9 Hz, ^3J = 4.5 Hz, 1H, 2'-H), 4.35 (s, 2H, C**H$_2$**OH), 4.28 (ddd, $^3J_{H\text{-}F}$ = 22.0 Hz, ^3J = 8.9 Hz, ^3J = 4.5 Hz, 1H, 3'-H) 4.07 (dt, ^3J = 8.9 Hz, ^3J = 3.3 Hz, 1H, 4'-H), 3.97 (dd, ^2J = 13.0 Hz, ^3J = 2.4 Hz, 1H, 5'-H), 3.78 (dd, ^2J = 13.0 Hz, ^3J = 4.0 Hz, 1H, 5'-H). **^{13}C-NMR (201 MHz, D_2O, ppm):** δ = 165.2 (dd, $^1J_{C\text{-}N}$ = 21.2 Hz, $^1J_{C\text{-}N}$ = 6.7 Hz, C-4), 156.9 (C-2), 140.9 (C-6), 106.5 (C-5), 93.7 (d, $^1J_{C\text{-}F}$ = 184.4 Hz, C-2'), 89.8 (d, $^2J_{C\text{-}F}$ = 35.0 Hz, C-1'), 82.1 (C-4'), 67.7 (d, $^2J_{C\text{-}F}$ = 16.1 Hz, C-3'), 59.6 (C-5'), 57.7 (**C**H_2OH). **^{19}F-NMR (376 MHz, D_2O, ppm):** δ = -200.5 (ddd, $^2J_{H\text{-}F}$ = 52.9 Hz, $^3J_{H\text{-}F}$ = 22.0, $^3J_{H\text{-}F}$ = 19.4). **^{15}N-NMR (41 MHz, D_2O, ppm):** δ = -201.9, -202.0. **HRMS (ESI$^+$):** calc. for $C_{10}H_{15}FN^{15}N_2O_5^+$ $[M+H]^+$: 278.0931, found: 278.0933. **HRMS (ESI$^-$):** calc. for $C_{10}H_{13}FN^{15}N_2O_5^-$ $[M-H]^-$: 276.0785, found: 276.0787.

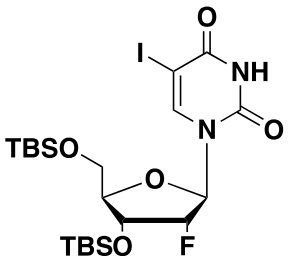
2'-deoxy-2'-(*R*)-fluoro-(N^3,N^4-$^{15}N_2$)-cytidine (19)



34 (117 mg, 0.25 mmol, 1.0 eq) was dissolved in EtOAc (3.5 mL) and pyridine (98 µL). Subsequently HF•pyridine (70% solution, 123 µL, 4.92 mmol, 20 eq) was added and the mixture stirred at room temperature for 22 hours. TMSOMe (4.9 mL) was then added and after another hour of stirring at room temperature, the solvents were removed *in vacuo*. The residue was dissolved in H_2O and then purified by HPLC which yielded 44 mg (0.18 mmol, 72%) of **19** as a colourless solid.

^{1}H-NMR (800 MHz, D_2O, ppm): δ = 7.81 (d, ^3J = 7.6 Hz, 1H, 6-H), 6.02 (d, ^3J = 7.7 Hz, 1H, 5-H), 5.97 (d, $^3J_{H\text{-}F}$ = 19.8 Hz, 1H, 1'-H), 5.15 (ddd, $^2J_{H\text{-}F}$ = 52.9 Hz, ^3J = 4.9 Hz, ^3J = 1.1 Hz, 1H, 2'-H), 4.34 (ddd, $^3J_{H\text{-}F}$ = 22.5 Hz, ^3J = 9.0 Hz, ^3J = 4.6 Hz, 1H, 3'-H), 4.16-4.08 (m, 1H, 4'-H), 4.01 (dd, ^2J = 13.0 Hz, ^3J = 2.4 Hz, 1H, 5'-H), 3.83 (dd, ^2J = 13.0 Hz, ^3J = 4.4 Hz, 1H, 5'-H). **^{13}C-NMR (201 MHz, D_2O, ppm):** δ = 162.6 (C4), 149.9 (C2), 140.6 (C6), 96.0 (C5), 93.2 (d, $^1J_{C\text{-}F}$ = 193.8 Hz, C2'), 89.9 (d, $^2J_{C\text{-}F}$ = 35.7 Hz, C1'), 82.2 (C4'), 67.8 (d, $^2J_{C\text{-}F}$ = 15.5 Hz, C3'), 59.9 (C5'). **^{19}F-NMR (376 MHz, D_2O, ppm):** δ = −200.3 (dt, $^2J_{F\text{-}H}$ = 53.1 Hz, $^3J_{F\text{-}H}$ = 19.0 Hz). **HRMS (ESI$^-$):** calc. for $C_{11}H_{15}FN^{15}N_2O_6^-$ [M+OAc]$^-$: 306.0891, found: 306.0894.

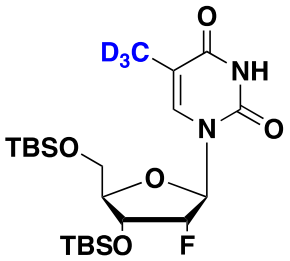
3',5'-bis-*O*-(*tert*-butyl(dimethyl)silyl)- 5-iodo-2'-deoxy-2'-(*R*)-fluoro -uridine (38)



TBS-protected 3',5'-bis-*O*-[*tert*-butyl(dimethyl)silyl]-2'-deoxy-2'-fluoro-uridine (0.50 g, 1.05 mmol, 1.0 eq) was dissolved in MeCN (100 mL) and CAN (2.30 g, 2.10 mmol, 2.0 eq) and I_2 (1.07 g, 2.10 mmol, 2.0 eq) were added to the mixture. The dark red brownish solution was heated to 65 °C for 1.5 hours and subsequently poured into saturated $NaHCO_3$ and a saturated solution of $Na_2S_2O_3$ was added until the solution became colourless. The aqueous phases were exrtracted with EtOAc (2x50 mL) and the combined organic fractions dried over Na_2SO_4. Column chromatography (20% EtOAc/^{i}Hex) yielded the iodinated compound as a brownish foam (469 mg, 0.78 mmol, 74%).

^{1}H-NMR (400 MHz, $CDCl_3$, ppm): δ = 9.34 (s, 1H, N**H**), 7.98 (s, 1H, 6-H), 6.08 (dd, $^3J_{H\text{-}F}$ = 14.3 Hz, ^3J = 3.8 Hz, 1H, 1'-H), 4.82 (dt, $^2J_{H\text{-}F}$ = 52.5 Hz, ^3J = 4.2 Hz, 1H, 2'-H), 4.27 (dt, $^3J_{H\text{-}F}$ = 12.9 Hz, ^3J = 4.9 Hz, 1H, 3'-H), 4.08 - 4.06 (m, 1H, 4'-H), 3.98 (dd, ^2J = 11.9 Hz, ^3J = 1.8 Hz, 1H, 5'-H), 3.76 (dd, ^2J = 11.9 Hz, ^3J = 2.1 Hz, 1H, 5'-H), 0.95 (s, 9 H, C3'/C5'-O-Si$(CH_3)_2$(^{t}**Bu**)), 0.90 (s, 9 H, Si$(CH_3)_2$(^{t}**Bu**)), 0.16 (s, 3H, Si(C**H$_3$**)$_2$(^{t}Bu)), 0.15 (s, 3H, Si(C**H$_3$**)$_2$(^{t}Bu)), 0.10 (s, 3H, Si(C**H$_3$**)$_2$(^{t}Bu)), 0.08 (s, 3H, Si(C**H$_3$**)$_2$(^{t}Bu)). **^{13}C-NMR (101 MHz, $CDCl_3$, ppm):** δ = 160.0 (4-C), 149.9 (2-C), 144.0 (C-6), 92.4 (d, $^1J_{C\text{-}F}$ = 195.3 Hz, C-2'), 87.4 (d, $^2J_{C\text{-}F}$ = 33.7 Hz, C-1'), 85.1 (C-4'), 69.6 (d, $^2J_{C\text{-}F}$ = 15.6 Hz, C-3'), 69.2 (C-5'), 61.7 (C-5), 26.3 (Si(C(**C**H$_3$)$_3$)), 25.7 (Si(C(**C**H$_3$)$_3$)), 18.7 (Si(**C**(CH$_3$)$_3$)), 18.2 (Si(**C**(CH$_3$)$_3$)), -4.6 (Si(**C**H$_3$)$_2$^{t}Bu), -5.0 (Si(**C**H$_3$)$_2$^{t}Bu), -5.1 (Si(**C**H$_3$)$_2$^{t}Bu), -5.1 (Si(**C**H$_3$)$_2$^{t}Bu). **HRMS (ESI$^+$):** calc. for $C_{21}H_{39}FIN_2O_5Si_2^+$ $[M+H]^+$: 601.1421, found:601.1432.

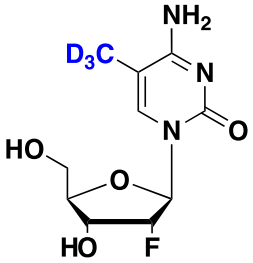
3',5'-bis-*O*-[*tert*-butyl(dimethyl)silyl]- 5-(D_3) -methyl 2'-deoxy-2'-(*R*)-fluoro-uridine (39**)**



38 (100 mg, 0.167 mmol, 1.0 eq) was dissolved in THF (4 mL) and Ni(dppp)Cl_2 (45 mg, 0.083 mmol, 0.5 eq) was added. The mixture was cooled to 0 °C and CD_3MgI (0.55 mL, 1M solution in ether, 3.3 eq) was added slowly. The clear red solution was slowly warmed to rt and stirred for one hour and subsequently quenched with sat. NH_4Cl (10 mL). The aqueous phase was extracted twice with EtOAc (15 mL), combined organic fractions dried over Na_2SO_4 and volatiles removed *in vacuo*. The crude mixture was finally purified thorugh column chromatography (20% EtOAc/^{i}Hex) to yield 30 mg (0.061 mmol, 37%) of **39**.

^{1}H-NMR (400 MHz, $CDCl_3$, ppm): δ = 8.82 (s, 1H, N**H**), 7.38 (s, 1H, 6-H), 6.07 (dd, $^3J_{H\text{-}F}$ = 15.5 Hz, ^3J = 3.6 Hz, 1H, 1'-H), 4.83 (dt, $^2J_{H\text{-}F}$ = 52.9 Hz, ^3J = 4.2 Hz, 1H, 2'-H), 4.29 (dt, $^3J_{H\text{-}F}$ = 13.4 Hz, ^3J = 5.1 Hz, 1H, 3'-H), 4.08 - 4.00 (m, 1H, 4'-H), 3.97 (dd, ^2J = 11.8 Hz, ^3J = 2.0 Hz, 1H, 5'-H), 3.76 (dd, ^2J = 11.8 Hz, ^3J = 2.8 Hz, 1H, 5'-H), 0.93 (s, 9 H, C3'/C5'-O-Si$(CH_3)_2$(^{t}**Bu**)), 0.91 (s, 9 H, Si$(CH_3)_2$(^{t}**Bu**)), 0.12 - 0.18 (m, 12H, Si(C**H$_3$**)$_2$(^{t}Bu)). **^{13}C-NMR (101 MHz, $CDCl_3$, ppm):** δ = 163.7 (4-C), 150.2 (2-C), 135.5 (C-6), 111.6 (5-C), 92.3 (d, $^1J_{C\text{-}F}$ = 194.2 Hz, C-2'), 87.4 (d, $^2J_{C\text{-}F}$ = 33.5 Hz, C-1'), 84.5 (C-4'), 69.5 (d, $^2J_{C\text{-}F}$ = 15.8 Hz, C-3'), 61.7 (C-5'), 26.0 (Si(C(**C**H$_3$)$_3$)), 25.7 (Si(C(**C**H$_3$)$_3$)), 18.5 (Si(**C**(CH$_3$)$_3$)), 18.2 (Si(**C**(CH$_3$)$_3$)), -4.7 (Si(**C**H$_3$)$_2$^{t}Bu), -5.1 (Si(**C**H$_3$)$_2$^{t}Bu), -5.3 (Si(**C**H$_3$)$_2$^{t}Bu), -5.3 (Si(**C**H$_3$)$_2$^{t}Bu). **HRMS (ESI$^+$):** calc. for $C_{22}H_{39}D_3FN_2O_5Si_2^+$ $[M+H]^+$: 492.2799, found: 492.2805.

5-(D_3)-methyl-2'-deoxy-2'-(*R*)-fluoro-5 -cytidine (18**)**



39 (30 mg, 0.06 mmol, 1.0 eq) was dissolved in pyridine (1.0 mL) and *1,2,4*-triazole (34 mg, 0.49 mmol, 8.0 eq) and $POCl_3$ (12 µL, 0.12 mmol, 1.2 eq) were added at rt. The mixture was stirred for two hours and then poured into sat. NH_4Cl (10 mL). The resulting aqueous phase was quickly extracted with EtOAc (10 mL), dried over Na_2SO_4 and the solvents removed *in vacuo*. The residue was redissolved in dioxane (3.0 mL) and NH_4OH (25%, 1.0 mL) was added. After stirring at rt overnight, volatiles were removed *in vacuo* and the crude taken up in EtOAc (0.5 mL). Subsequently, HF•pyridine (70% solution, 8 µL, 0.32 mmol, 5.4 eq) and pyridine (12 µL) were added stirred at room temperature overnight. HF•pyridine was quenched through addition of TMSOMe (100 µL) and the residue was finally purified by HPLC (0 --> 10% MeCN/H_2O in 45 mins) to give **18** as a white powder.

^{1}H-NMR (400 MHz, D_2O, ppm): δ = 7.54 (s, 1H, 6-H), 5.87 (dd, $^3J_{H\text{-}F}$ = 19.8 Hz, ^3J = 1.2 Hz, 1H, 1'-H), 5.00 (dd, $^2J_{H\text{-}F}$ = 53.0 Hz, ^3J = 4.6 Hz, 1H, 2'-H), 4.23 (ddd, $^3J_{H\text{-}F}$ = 22.3 Hz, ^3J = 8.9 Hz, ^3J = 4.6 Hz, 1H, 3'-H), 4.04-3.97 (m, 1H, 4'-H), 3.91 (dd, ^2J = 13.0 Hz, ^3J = 2.4 Hz, 1H, 5'-H), 3.72 (dd, ^2J = 13.0 Hz, ^3J = 4.3 Hz, 1H, 5'-H). **^{13}C-NMR (101 MHz, D_2O, ppm):** δ = 166.2 (C4), 157.0 (C2), 139.1 (C6), 104.4 (C5), 93.7 (d, $^1J_{C\text{-}F}$ = 184.5 Hz, C2'), 89.7 (d, $^2J_{C\text{-}F}$ = 35.1 Hz, C1'), 82.1 (C4'), 67.8 (d, $^2J_{C\text{-}F}$ = 16.6 Hz, C3'), 59.8 (C5').

HRMS (ESI$^+$): calc. for $C_{10}H_{12}D_3FN_2O_4^+$ $[M+H]^+$: 263.1229, found:263.1230.

References

1 Schröder, A. S. *et al.* 2'-(*R*)-fluorinated mC, hmC, fC and caC triphosphates are excellent substrates for DNA polymerases and TET-enzymes. *Chem. Commun.* **52**, 14361-14364, (2016).

2 Brunner, K. *et al.* Cell-penetrating and neurotargeting dendritic siRNA nanostructures. *Angew. Chem. Int. Ed.* **54**, 1946-1949, (2015).

3 Saito, Y., Zevaco, T. A. & Agrofoglio, L. A. Chemical synthesis of 13C labeled anti-HIV nucleosides as mass-internal standards. *Tetrahedron* **58**, 9593-9603, (2002).

4 Schiesser, S. *et al.* Mechanism and stem-cell activity of 5-carboxycytosine decarboxylation determined by isotope tracing. *Angew. Chem. Int. Ed.* **51**, 6516-6520, (2012).

5 Robins, M. J. & Wilson, J. S. Smooth and efficient deoxygenation of secondary alcohols. A general procedure for the conversion of ribonucleosides to 2'-deoxynucleosides. *J. Am. Soc.* **103**, 932-933, (1981).

7.2 Zusatzmaterialien zu Abschnitt 3.2

R. Rahimoff#, **O. Kosmatchev#**, A. Kirchner#, T. Pfaffeneder, F. Spada, V. Brantl, M. Müller, T. Carell, *J. Am. Chem. Soc.* **2017**, *139*, 10359-10364. *5-Formyl- and 5-Carboxydeoxycytidines Do Not Cause Accumulation of Harmful Repair Intermediates in Stem Cells.*

(# geteilte Erstautorenschaft)

Supporting Information

5-Formyl- and 5-Carboxydeoxycytidines Do Not Cause Accumulation of Harmful Repair Intermediates in Stem Cells

René Rahimoff[#], Olesea Kosmatchev[#], Angie Kirchner[#], Toni Pfaffeneder, Fabio Spada, Victor Brantl, Markus Müller, Thomas Carell*

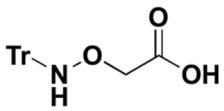
Center for Integrated Protein Science at the Department of Chemistry, LMU Munich, Butenandtstrasse 5-13, Munich 81377, Germany

1. Chemical Synthesis

Unless noted otherwise, all reactions were performed using oven dried glassware under an atmosphere of nitrogen. Molsieve-dried solvents were used from *Sigma Aldrich* and chemicals were bought from *Sigma Aldrich*, *TCI*, *Carbolution* and *Carbosynth*. Isotopically labeled trimethylamino glycine was obtained from *Eurisotop*. For extraction and chromatography purposes, technical grade solvents were distilled prior to their usage. Reaction controls were performed using TLC-Plates from *Merck* (Merck 60 F_{254}), flash column chromatography purifications were performed on *Merck* Geduran Si 60 (40-63 μM). Visualization of the TLC plates was achieved through UV-absorption or through staining with *Hanessian's stain*. NMR spectra were recorded in deuterated solvents on *Varian VXR400S*, *Varian Inova 400*, *Bruker AMX 600*, *Bruker Ascend 400* and *Bruker Avance III HD*. HR-ESI-MS spectra were obtained from a *Thermo Finnigan* LTQ FT-ICR. IR-measurements were performed on a *Perkin Elmer Spectrum BX FT-IR* spectrometer with a diamond-ATR (*Attenuated Total Reflection*) unit. HPLC purifications were performed on a *Waters Breeze* system (2487 dual array detector, 1525 binary HPLC pump) using a Nucleosil VP 250/10 C18 column from *Macherey Nagel*. HPLC-grade MeCN was purchased from *VWR*. For HPLC purifications of compounds **1a/b**, **9a/b** and **10a/b** a buffer system of 0.25 mM ammonium formate in H_2O, pH = 4.3 (referred to as buffer A) and 0.25 mM ammonium formate in 80% MeCN/H_2O (referred to as buffer B) was used.

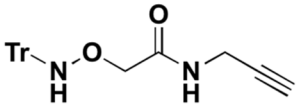
Synthesis of Hydroxylamine 1 and internal standards 9a/b and 10a/b

(*N*-Tritylaminooxy)acetic acid (6)



(*N*-Tritylaminooxy)acetic acid was synthesized according to Kojima *et al*.[1]

***N*-(prop-2-ene-1-yl)-2-((tritylamino)oxy)acetamide (7)**

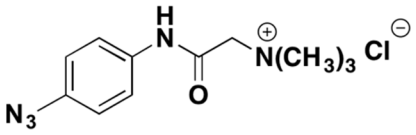


N-trityl protected aminooxyacetic acid **6** (2.50 g, 7.50 mmol, 1.0 eq) was suspended in DCM (40 mL) and was subsequently charged with TBTU (2.89 g, 9.00 mmol, 1.2 eq), DIPEA (1.60 mL, 9.00 mmol, 1.2 eq) and propargylamine (1.40 mL, 22.6 mmol, 3.0 eq). The suspension was stirred at rt, whereupon

after 15 hours a clear yellowish solution was formed. The mixture was diluted with EtOAc (300 mL), the organic phase was washed with NH_4Cl (300 mL) and $NaHCO_3$ (300 mL) and then dried over Na_2SO_4. Volatiles were finally removed *in vacuo* and the crude mixture was purified via column chromatography (10% EtOAc --> 40% EtOAc/^{i}Hex). **7** (2.57g, 6.93 mmol, 92%) was yielded as a colorless solid.

^{1}H-NMR (300 MHz, $CDCl_3$): δ/ppm = 7.37−7.22 (m, 15H, $(C_6\mathbf{H_5})_3C$), 6.59 (s, 1H, $(C_6H_5)_3C$-N**H**-O), 5.81 (bs, 1H, O=C-N**H**), 4.25 (s, 2H, O-C$\mathbf{H_2}$C=O), 3.85 (dd, ^3J = 5.5 Hz, ^4J = 2.6 Hz, 2H, HN-C$\mathbf{H_2}$), 2.15 (t, ^4J = 2.6 Hz, 1H, C≡C-**H**). **^{13}C-NMR (75 MHz, $CDCl_3$):** δ/ppm = 169.1 (**C**=O), 143.9 (3C, 3 ×O-NH-C-**C**), 129.0 (6C, $\mathbf{C_{Ar}}$-H), 128.2 (6C, $\mathbf{C_{Ar}}$-H), 127.4 (3C, $\mathbf{C_{tert}}$-H), 79.3 (C≡**C**-H), 74.6 ($C(C_6H_5)_3$), 73.4 (O-CH_2), 71.7 (C≡C- H), 28.8 (NH-CH_2). **HRMS (ESI$^+$):** calc. for $C_{24}H_{22}N_2NaO_2$ $[M+Na]^+$: 393.1573; found: 393.1571. **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3288 (w), 3222 (w), 3056 (w), 2913 (w), 2359 (w), 2339 (w), 1635 (m), 1542 (m), 1489 (m), 1065 (m), 996 (m), 763 (m), 747 (m), 707 (s), 697 (s), 685 (s), 627 (s). **Melting Range:** 157 - 158 °C.

2-((4-Azidophenyl)amino-*N*,*N*,*N*-trimethyl-2-oxoethaneaminium chloride (4a**)**

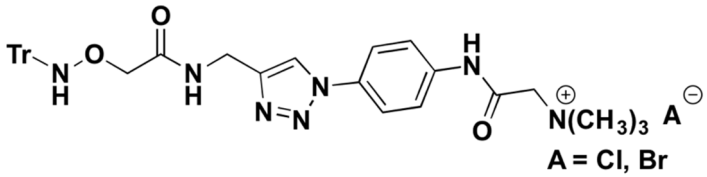


Betaine **3a** (0.30 g, 2.56 mmol, 1.0 eq) was first dried on high vac at 180 °C for 20 minutes. After cooling to rt, the colorless solid was suspended in DMF (25 mL). 4-Azidoanilin hydrochloride **2** (0.54g, 3.17 mmol, 1.2 eq), TBTU (0.99 g, 3.07 mmol, 1.2 eq) and DIPEA (1.10 mL, 6.32 mmol, 2.4 eq) were added whereupon a yellow brownish solution formed gradually. After stirring for one hour at rt all solids were dissolved and the reaction was further stirred at rt over night. DMF was then removed *in vacuo* and the crude mixture was purified by column chromatography (DCM/MeOH/H_2O/7N NH_3 in methanol = 90:10:0.6:0.6) and **4a** was yielded as the corresponding phenyl triazolate salt. The salt was then redissolved in H_2O (50 mL) and was acidified to pH = 1. The aqueous phase was extracted with Et_2O until TLC analysis of the organic phase fractions showed no UV absorption. The aqueous layer was subsequently neutralized with conc. NH_3 and the chloride salt of **4a** (0.62 g, 2.30 mmol, 90%) was yielded as a brownish powder.

^{1}H-NMR (300 MHz, dmso d^6): δ/ppm = 11.22 (s, 1H, N**H**), 7.69 (d, ^3J = 8.9 Hz, 2H, C**H**=C-NH), 7.12 (d, ^3J = 8.4 Hz, 2H, C**H**=C-N_3), 4.42 (s, 2H, C$\mathbf{H_2}$), 3.30 (s, 9H, N(C$\mathbf{H_3}$)$_3$). **^{13}C-NMR (101 MHz, dmso d^6):** δ/ppm = 162.0 (**C**=O), 135.0 (NH-**C**=CH), 134.9 (N_3-**C**=CH), 121.2 (2C, NH-C=**C**H), 119.5 (2C, N_3-**C**=CH), 64.3 (**C**H_2),

53.4 (3C, N(**C**H_3)$_3$). **HRMS (ESI$^+$):** calc. for $C_{11}H_{16}N_5O^+$ [M^+]: 234.1349; found: 234.1348. **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3348 (w), 2983 (w), 2118 (s), 2083 (m), 1692 (s), 1676 (m), 1615 (m), 1549 (m), 1508 (s), 1287 (s), 1256 (m), 1050 (s), 1038 (s), 922 (s), 833 (s). **Melting Range:** 144 - 146 °C.

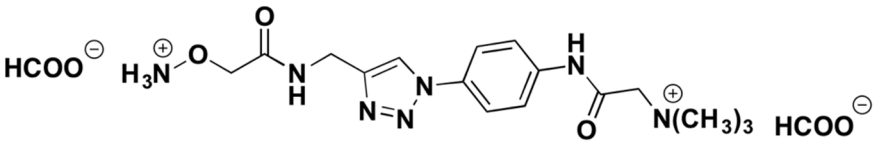
***N,N,N*-trimethyl-2-oxo-2-((4-(4-((2-((tritylamino)oxy)acetamido)methyl)-1*H*-1,2,3-triazole1-yl)phenyl)amino)ethanaminium chloride/bromide (**8a**)**



First, a mixture of DCM and H_2O (à 5 mL) was freeze-pump-thaw degassed (3x) and then azide **4a** (0.18 g, 0.65 mmol, 1.0 eq), alkyne **7** (0.24 g, 0.65 mmol, 1.0 eq) and CuBr•SMe_2 (40 mg, 0.20 mmol, 0.3 eq) were added. The suspension was stirred vigorously over night at rt whereupon a colorless emulsion formed. The mixture was then concentrated under reduced pressure and purified by column chromatography using a short plug of silica (DCM/MeOH/H_2O/7N NH_3 in methanol = 80:20:0.6:0.6). **8a** was yielded as a slightly yellow brownish solid (0.32 g, 0.50 mmol, 77%).

^{1}H-NMR (300 MHz, dmso d^6): δ/ppm = 11.67 (s, 1H, N**H**-C_6H_4), 8.55 (s, 1H, CH_2-C=C**H**-N), 8.34 (s, 1H, Ph_3C-N**H**), 8.32 (t, ^{3}J = 5.8 Hz, 1H, O=C-N**H**-CH_2), 7.91–7.84 (m, 4H, C_6**H$_4$**), 7.34–7.19 (m, 15H, C(C_6**H$_5$**)$_3$), 4.53 (s, 2H, (C**H$_2$**-N(CH_3)$_3$), 4.45 (d, ^{3}J = 5.8, 2H, NH-C**H$_2$**), 3.85 (s, 2H, N-O-CH_2), 3.33 (s, 9H, N(C**H$_3$**)$_3$). **^{13}C-NMR (101 MHz, dmso d^6):** δ/ppm = 169.7 (O=**C**-NH-CH_2), 162.4 (O=**C**-CH_2-N), 146.0 (CH_2-**C**=C), 144.1 (3C, O-NH-C-**C**), 138.1 (N-**C**=CH-CH), 132.6 (N-**C**=CH-CH), 128.9 (6C, **C**$_{Ar}$-H), 127.6 (6C, **C**$_{Ar}$-H), 126.7 (3C, **C**-H), 121.0 (CH_2-C=**C**H-N), 120.5 (4C, N-C=**C**H-**C**H=C-N), 73.7 (**C**(C_6H_6)$_3$), 73.2 (O-**C**H_2), 64.4 (**C**H_2-N(CH_3)$_3$), 53.4 (N(**C**H_3)$_3$), 33.8 (NH-**C**H_2). **HRMS (ESI$^+$):** calc. for $C_{35}H_{38}N_7O_3^+$ [M^+]: 604.3031; found: 604.3026. **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3387 (w), 3054 (w), 2923 (w), 1685 (m), 1613 (m), 1558 (m), 1519 (s), 1490 (m), 1446 (m), 1413 (m), 1312 (m), 1265 (m), 1224 (m), 1192 (m), 1085 (m), 1045 (m), 1002 (m), 990 (m), 948 (m), 922 (m), 876 (m), 838 (m), 757 (s), 698 (s), 627 (s). **Melting Range:** 142 - 152 °C.

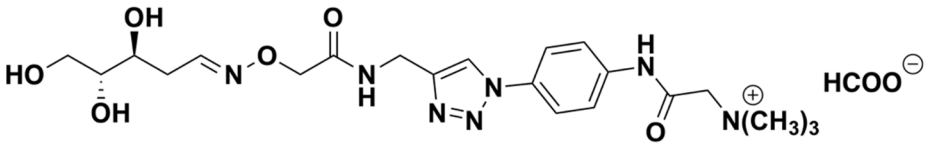
2-((4-(4-((2-(Aminooxy)acetoamido)methyl)-1*H*-1,2,3-triazole-1-yl)phenyl-amino)-*N,N,N*-trimethyl-2-oxoethanaminium formate (1a**)**



Trityl protected compound **8a** (0.24g, 0.38 mmol) was dissolved in DCM (6 mL) and then 6M HCl (6 mL) was added. The mixture was rigorously stirred at rt for one hour until a phase separation was visible. The aqueous phase was then extracted with DCM (5 x 5mL) until TLC analysis of the organic phase fractions showed no UV absorption. The pH was adjusted to 9-10 using 2M NH_3 and the aqueous phase was removed *in vacuo*. 75 mg of **1a** were then further purified by preparative HPLC (0% --> 20% buffer B) and yielded 23 mg (0.05 mmol, 26%) of **1a** as the colorless formate salt.

^{1}H-NMR (400 MHz, D_2O): δ/ppm = 8.40 (s, 1H, CH_2-C=C**H**-N), 8.21 (s, 1H, **H**COO), 7.58 (d, ^{3}J = 9.2 Hz, 2H, CH-C**H**=C-N_3), 7.52 (d, ^{3}J = 9.2 Hz, 2H, CH-C**H**=C-NH), 4.53 (s, 2H, N-O-C**H_2**), 4.27 (s, 2H, NH-C**H_2**), 4.22 (s, 2H, C**H_2**-N$(CH_3)_3$), 3.36 (s, 9H, N(C**H_3**$)_3$). **^{13}C-NMR (101 MHz, D_2O):** δ/ppm = 169.7 (O=**C**-NH-CH_2), 162.4 (O=**C**-CH_2-N), 146.0 (CH_2-**C**=C), 144.1 (3C, O-NH-C-**C**), 138.1 (N-**C**=CH-CH), 132.6 (N-**C**=CH-CH), 128.9 (6C, **C**$_{Ar}$-H), 127.6 (6C, **C**$_{Ar}$-H), 126.7 (3C, **C**-H), 121.0 (CH_2-C=**C**H-N), 120.5 (4C, N-C=**C**H-**C**H=C-N), 73.7 (**C**$(C_6H_6)_3$), 73.2 (O-**C**H_2), 64.4 (**C**H_2-N$(CH_3)_3$), 53.4 (N(**C**$H_3)_3$), 33.8 (NH-**C**H_2). **HRMS (ESI$^+$):** calc. for $C_{16}H_{24}N_7O_3^+$ [M^+]: 362.1935; found: 362.1935. **IR (ATR):** $\tilde{\nu}$ (cm^{-1}) = 3130 (m), 3037 (s), 2807 (m), 2649 (m), 2363 (w), 1684 (s), 1610 (m), 1556 (s), 1517 (s), 1487 (m), 1475 (m), 1442 (m), 1403 (s), 1312 (m), 1262 (m), 1193 (m), 1128 (w), 1083 (w), 1048 (m), 991 (m), 967 (w), 921 (s), 837 (s).

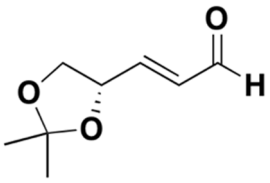
***N,N,N*-Trimethyl-2-oxo-2-((4-(4-((2-((((3*S*,4*R*)-3,4,5-trihydroxypentyliden)amino)oxy)acet-amido)-methyl)-1*H*-1,2,3-triazol-1-yl)phenyl)amino)ethanaminium formate (**9a**)**



1a (50.0 mg, 0.12 mmol, 1-0 eq) and 2'-desoxyribose (182 mg, 1.36 mmol, 11.8 eq) were dissolved in H_2O (2.7 mL) and incubated over night at 30 °C and 1400 rpm in a *Eppendorf Comfort* thermomixer. The mixture was filtered over a 0.2 μm syringefilter and was subsequently purified by HPLC twice (0 → 15% buffer B). Pure product **9a** (9.1 mg, 17 μmol, 15%) was obtained as a colorless foam. The compound was present as a mixture of *E*/*Z* isomers in aqueous solution that were not assigned.

^{1}H-NMR (600 MHz, D_2O): δ/ppm = 8.46 (s, 1H, **H**COO), 8.34 (s, 1H, CH_2-C=C**H**-N), 7.79 (d, J=9.0 Hz, 2H, CH-C**H**=C-N_3), 7.74 - 7.71 (m, 8H, CH-C**H**=C-NH, C1´-**H^A**), 7.08 (t, ^{3}J = 5.4 Hz, 1H, C1´-**H^B**), 4.67 (s, 2H, N-O-C**H_2^B**), 4.63 (s, 2H, NH-C**H_2**), 4.62 (s, 2H, N-O- C**H_2^A**), 4.35 (s, 2H, C**H_2**-N$(CH_3)_3$), 3.92–3.87 (m, 1H, C3´-**H^B**), 3.85 - 3.80 (m, 1H, C3´-**H^A**), 3.78–3.69 (m, 1H, C5´-**H**), 3.66–3.53 (m, 2H, C5´-**H**, C4´-**H**), 3.42 (s, 9H, N(C**H_3**)$_3$), 2.79–2.69 (m, 2H, C2´-**H^B**), 2.58 - 2.54 (m, 1H, C2´-**H^A**), 2.41 - 2.35 (m, 1H, C2´-**H^A**). **^{13}C-NMR (150 MHz, D_2O):** δ/ppm = 172.4 (O=**C**-NH-CH_2), 170.9 (H**C**OO), 162.7 (O=**C**-CH_2-N), 153.5 (C1´^{A}), 153.1 (C1´^{B}), 145.1 (CH_2-**C**=C), 136.8 (N- **C**=CH-CH), 133.5 (N-**C**=CH-CH), 122.5 (2C, **C**H=C-NH), 122.3 (CH_2-C=**C**H-N), 121.9 (2C, **C**H=C-N_3), 74.2 (C4´), 74.0 (C4´), 71.7 (N-O-**C**H_2^B), 71.5 (N-O-**C**H_2^A), 69.0 (C3´^{A}), 68.8 (C3´^{B}), 65.1 (**C**H_2-N$(CH_3)_3$), 62.3 (C5´), 54.3 (N(**C**H_3)$_3$), 34.1 (NH-**C**H_2), 32.4 (C2´^{A}), 29.2 (C2´^{B}). **HRMS (ESI$^+$):** calc. for $C_{21}H_{32}N_7O_6^+$ $[M]^+$: 478.2409; found: 478.2404.

(*S*,*E*)-3-(2,2-Dimethyl-1,3-dioxolan-4-yl)acrylaldehyde (12**)**



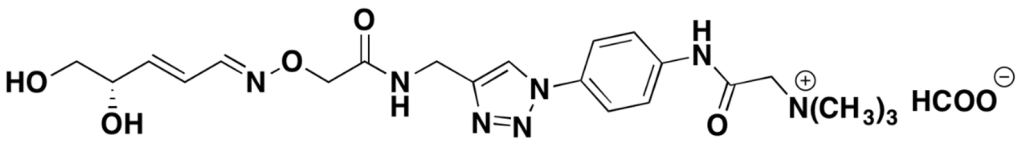
Methyl (2*E*)-3-[(4*S*)-2,2-dimethyl-1,3-dioxolan-4-yl]prop-2-enoate **11** (0.20 g, 1.08 mmol, 1.0 eq) was dissolved in DCM (2.0 mL) and cooled to –78 °C. DIBAL-H (2.20 mL, 2M in toluene, 2.1 eq) was added and the yellowish mixture was slowly warmed to rt. After 90 minutes, DCM (5.0 mL) and H_2O (4.0 mL) and NaOH (2M, 2.0 mL) were added. After stirring for an additional hour at rt, the organic phase was separated from the aqueous and dried over Na_2SO_4. Volatiles were removed under reduced pressure and the allylic alcohol was isolated in quantitative yield and used without further purification.

^{1}H-NMR (400 MHz, $CDCl_3$): δ/ppm = 5.88 (dt, ^{3}J = 15.4 Hz, ^{4}J = 5.0 Hz, 1H, 5´-H) 5.65 (dd, ^{3}J = 15.6 Hz, ^{4}J = 7.6 Hz, 1H, 5´-H), 4.47 (q, ^{3}J = 7.3 Hz, 1H, 4´-H), 4.08 (d, ^{3}J = 5.1 Hz, 2H, 1´-H), 4.30 (dd, ^{3}J = 8.2 Hz, ^{4}J = 6.1 Hz, 1H, 3´-H), 3.53 (t, ^{3}J = 7.9 Hz, 1H, 2´-H), 2.34 (br s, 1H, CH_2-O**H**), 1.36 (s, 3H, O-C(C**H_3**)(CH_3)-O), 1.32 (s, 3H, O-C(CH_3)(C**H_3**)-O).

The allylic alcohol was dissolved in DCM (2.0 mL), cooled to 0 °C and was charged with *Dess-Martin-periodinan* (0.45 g, 1.08 mmol, 1.0 eq). The milky suspension was slowly warmed to rt and stirred over night. After the addition of saturated Na_2SO_4 (10 mL) and a solution of $Na_2S_2O_3$ (171 mg, dissolved in 10 mL H_2O), the mixture was extracted with DCM (3 x 15 mL) and dried over Na_2SO_4. Organic solvents were removed *in vacuo* and the crude mixture was purified via column chromatography (2.5% MeOH/DCM). Aldehyde **12** (80 mg, 0.51 mmol, 47%) was isolated as a colorless oil.

^{1}H-NMR (400 MHz, CD_2Cl_2): δ/ppm = 9.50 (d, ^{3}J = 7.6 Hz, 1H, 1′-C**H**O), 6.70 (dd, ^{3}J = 15.6 Hz, ^{4}J = 5.3 Hz, 1H, 3′-H), 6.23 (dt, ^{3}J = 15.6 Hz, ^{4}J = 5.8 Hz, 1H, 2′-H), 4.73 (q, ^{3}J = 6.8 Hz, 1H, 4′-H), 4.18 (dd, ^{3}J = 8.4Hz, ^{4}J = 6.8Hz, 1H, 5′-H), 3.67 (dd, ^{3}J = 8.4Hz, ^{4}J = 6.8Hz, 1H, 5′-H), 1.39 (s, 3H, O-C(C**H$_3$**)(CH$_3$)-O), 1.35 O-C(CH$_3$)(C**H$_3$**)-O). **^{13}C-NMR (101 MHz, CD_2Cl_2):** δ/ppm = 193.0 (-**C**HO), 153.4 (3′-C), 132.1 (2′-C), 110.3 (**C$_{quart}$**), 74.9 (4′-C), 68.7 (5′-C), 26.2 (O-C(**C**H$_3$)(CH$_3$)-O), 25.4 (O-C(CH$_3$)(**C**H$_3$)-O). **HRMS (EI):** calc. for $C_8H_{11}O_3^{\bullet}$ [M - H]$^{\bullet}$: 155.0708; found: 155.0707.

2-((4-(4-((2-((((1*E*,2*E*)-3-((S)-2,2-Dimethyl-1,3-dioxolane-4-yl)allylidene)-amino)oxy)acetamido)methyl)-1*H* -1,2,3-triazole-1-yl)phenyl)amino)-*N* ,*N* ,*N* -trimethyl-2-oxoethane-1- aminium formate (10a**)**



Aldehyde **12** (20 mg, 0.13 mmol, 9.0 eq) and hydroxylamine **1a** were dissolved in a 1:1 mixture of H_2O and $CHCl_3$ (à 2.5 mL) and stirred at rt. The course of the reaction was monitored by HPLC (0 --> 30% buffer B) whereas it was determined that after one hour the reaction was complete. The aqueous phase was then washed with DCM (3 x 10 mL) and concentrated *in vacuo*. **13a** (5.10 mg, 9.50 μmol, 68%) was yielded as a brownish viscous oil that was used without further purification.

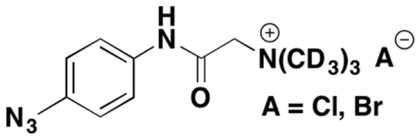
^{1}H-NMR (400 MHz, D_2O): δ/ppm = 8.31 (s, 1H, **H**COO), 8.12 (s, 1H, CH$_2$-C=C**H**-N), 7.84 (d, 1H, 1′-H), 7.59 − 7.53 (m, 4H, C**H**-C**H**=C-N$_3$, C**H**-C**H**=C-NH), 6.20 − 6.03 (m, 2H, 2′ + 3′-H's), 4.54 − 4.52 (m, 1H, 4′-H), 4.50 (s, 2H, N-O-C**H$_2$**), 4.47 (s, 2H, NH-C**H$_2$**), 4.20 (s, 2H, C**H$_2$**-N(CH$_3$)$_3$), 4.02 − 3.98 (m, 1H, 5′-H), 3.45 − 3.50 (m, 1H, 5′-H), 3.27 (s, 9H, CH$_2$-N(C**H$_3$**)$_3$), 1.26 (s, 3H, O-C(C**H$_3$**)(CH$_3$)-O), 1.24 (s, 3H, O-C(CH$_3$)(C**H$_3$**)-O). **HRMS (ESI$^+$):** calc. for $C_{24}H_{34}N_7O_5^+$ [M$^+$]: 500.2616; found: 500.2617.

For the deprotection of the acetonide, **13a** (4.00 mg, 7.50 μmol, 1.0 eq) was dissolved in MeOH and PTSA • H_2O (1.40 mg, 7.50 μmol, 1.0 eq) was added. The mixture was incubated in a *Eppendorf comfort* thermomixer (1300 rpm, 25 °C) over night and the solvent was removed *in vacuo* by lyophylization. The crude product was finally purified by semi preparative HPLC (0 --> 35% buffer B in 45 minutes) and pure **10a** was yielded as a colorless foam.

^{1}H-NMR (400 MHz, D_2O): δ/ppm = 8.53 (s, 1H, **H**COO), 8.30 (s, 1H, CH$_2$-C=C**H**-N), 8.01 (s, d, ^{3}J = 8.9 Hz, 1H, 1′-H), 7.78 − 7.70 (m, 4H, C**H**-C**H**=C-N$_3$, C**H**-C**H**=C-NH), 6.31 − 6.32 (m, 2H, 2′ + 3′-H's), 4.63 (s, 2H, N-O-C**H$_2$**), 4.62 (s, 2H, NH-C**H$_2$**), 4.35 − 4.32 (m, 3H, C**H$_2$**-N(CH$_3$)$_3$ + 4′-H), 3.61 (dd, ^{1}J = 11.7 Hz, ^{3}J = 4.4, 1H,

5′-H), 3.51 (dd, ^{1}J = 11.7 Hz, ^{3}J = 6.5, 1H, 5′-H), 3.22 (s, 9H, CH_2-N(C**H₃**)$_3$). **^{13}C-NMR (101 MHz, D_2O):** δ/ppm = 172.3 (O=**C**-NH-CH_2), 170.9 (H**C**OO), 162.7 (O=**C**-CH_2-N), 153.7 (1′-**C**), 144.5 (CH_2-**C**=C), 142.8 (3′-**C**), 136.9 (N-**C**=CH-CH), 133.6 (N-**C**=CH-CH), 123.0 (2′-**C**), 122.5 (2C, **C**H=C-NH), 122.3 (CH_2-C=**C**H-N), 122.0 (2C, **C**H=C-N_3), 72.0 (N-O-**C**H_2) 71.6 (4′-**C**), 64.4 (5′-**C**), 65.1 (**C**H_2-N(CH_3)$_3$), 54.3 (N(**C**H_3)$_3$), 34.1 (NH-**C**H_2). **HRMS (ESI$^+$):** calc. for $C_{21}H_{30}N_7O_5^+$ $[M]^+$: 460.2303; found: 460.2305.

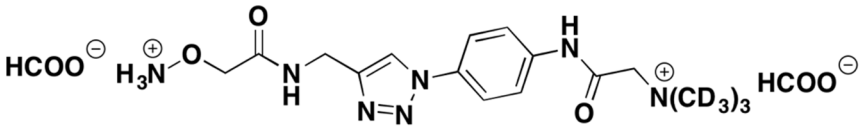
2-((4-Azidophenyl)amino-*N*,*N*,*N*-tri(methyl-*d_3*)-2-oxoethaneaminium chloride (4b**)**



4b was synthesized analogously to **4a** except that [d_{11}]-betaine (98% deuterium, *Euriso-Top GmbH*) was used to introduce isotopic labels. Deuterium labels from the methylene group were not stable under the reaction conditions and a complete D/H exchange was observed. Thus, a [d_9]-labeled product was obtained.

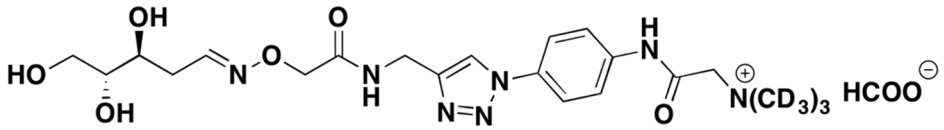
^{1}H-NMR (600 MHz, D_2O): δ/ppm = 7.39 (d, ^{3}J=8.6, 2H, CH-C**H**=C-NH), 7.04 (d, ^{3}J=8.5, 2H, CH-C**H**=C-N_3), 4.18 (s, 2H, C**H₂**). **^{13}C-NMR (150 MHz, D_2O, ppm):** δ/ppm = 162.7 (**C**=O), 137.5 (NH-**C**=CH), 132.5 (N_3-**C**=CH), 123.5 (2C, NH-C=**C**H), 119.6 (2C, N_3-C=**C**H), 65.0 (**C**H_2). **HRMS (ESI$^+$):** calc. for $C_{11}H_7D_9N_5O^+$ $[M]^+$: 243.1914; found: 243.1916.

2-((4-(4-((2-(Aminooxy)acetoamido)methyl)-1*H*-1,2,3-triazole-1-yl)phenyl-amino)-*N*,*N*,*N*-tri(methyl-*d_3*)-2-oxoethanaminium formate (1b**)**



Isotopologue **1b** was synthesized according to **1a** with the slight modification that the trityl protected intermediate **8b** was not isolated and deprotected without further purification.

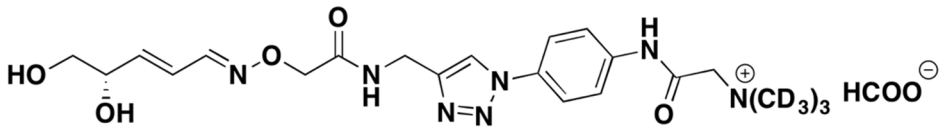
***N,N,N*-Tri(methyl-d_3)-2-oxo-2-((4-(4-((2-((((3*S*,4*R*)-3,4,5-trihydroxypentyliden)amino)oxy)acetamido)-methyl)-1*H*-1,2,3-triazol-1-yl)phenyl)amino)ethanaminium formate (**9b**)**



Internal standard **9b** was synthesized analogously to **9a**, whereupon a mixture of (*E*)/(*Z*)-isomers was obtained (depicted as **A** and **B**).

^{1}H-NMR (600 MHz, D_2O): δ/ppm = 8.46 (s, 2H, **H**COO), 8.35 (s, 1H, CH_2-C=C**H**-N), 7.81 (d, ^{3}J = 7.5 Hz, 2H, CH-C**H**=C-N_3), 7.75 - 7.73 (m, 8H, CH-C**H**=C-NH, C1´-**H^A**), 7.08 (t, ^{3}J = 5.4 Hz, 0.1H, C1´-**H^B**), 4.67 (s, 2H, N-O-C**H$_2^B$**), 4.64 (s, 2H, NH-C**H$_2$**), 4.62 (s, 2H, N-O-C**H$_2^A$**), 4.34 (s, 2H, C**H$_2$**-N$(CH_3)_3$), 3.93–3.87 (m, 1H, C3´-**H^B**), 3.86–3.79 (m, 1H, C3´-**H^A**), 3.78–3.69 (m, 1H, 1x C5´-**H$_2$**), 3.67–3.53 (m, 2H, 1x C5´-H_2, C4´-**H**), 2.80–2.68 (m, 2H, C2´-**H$_2^B$**), 2.59 - 2.54 (m, 1H, C2´-**H^A**), 2.43–2.34 (m, 1H, C2´-**H^A**). **^{13}C-NMR (150 MHz, D_2O, ppm):** δ/ppm = 172.4 (O=**C**-NH-CH_2), 170.9 (H**C**OO), 162.8 (O=**C**-CH_2-N), 153.5 (**C1´^{A}**), 153.1 (**C1´^{B}**), 145.1 (CH_2-**C**=C), 136.8 (N-**C**=CH-CH), 133.6 (N-**C**=CH-CH), 122.6 (2C, **C**H=C-NH), 122.4 (CH_2-C=**C**H-N), 122.0 (2C, **C**H=C-N_3), 74.2 (**C4´**), 74.0 (**C4´**), 71.7 (N-O-**C**H_2^B), 71.5 (N-O-**C**H_2^A), 69.0 (**C3´^{A}**), 68.8 (**C3´^{B}**), 64.9 (**C**H_2-N$(CD_3)_3$), 62.3 (**C5´**), 53.3 (N(**C**$D_3)_3$), 34.1 (NH-**C**H_2), 32.4 (**C2´^{A}**), 29.2 (**C2´^{B}**). **HRMS (ESI$^+$):** calc. for $C_{21}H_{23}D_9N_7O_6^+$ $[M]^+$: 487.2973; found: 487.2967.

2-((4-(4-((2-((((1*E*,2*E*)-3-((S)-2,2-Dimethyl-1,3-dioxolane-4-yl)allylidene)-amino)oxy)acetamido)methyl)-1*H* -1,2,3-triazole-1-yl)phenyl)amino)-*N* ,*N* ,*N* -tri(methyl-d_3)-2-oxoethane-1- aminium formate (10b**)**



10b was synthesized according to **10a**.

^{1}H-NMR (800 MHz, D_2O): δ/ppm = 8.47 (s, 1H, **H**COO), 8.32 (s, 2H, CH_2-C=C**H**-N), 8.02 (d, ^{3}J = 9.4 Hz, 1H, 1´-H), 7.80 − 7.71 (m, 4H, C**H**-C**H**=C-N_3, C**H**-C**H**=C-NH), 6.34 − 6.26 (m, 2H, 2´ + 3´-H's), 4.65 (s, 2H, N-O-C**H$_2$**), 4.63 (s, 2H, NH-C**H$_2$**), 4.35 - 4.32 (m, 3H, C**H$_2$**-N$(CH_3)_3$ + 4´-H), 3.63 (dd, ^{1}J = 11.7 Hz, ^{3}J = 4.4, 1H, 5´-H), 3.53 (dd, ^{1}J = 11.7 Hz, ^{3}J = 6.5, 1H, 5´-H). **^{13}C-NMR (150 MHz, D_2O):** δ/ppm = 172.3 (O=**C**-NH-CH_2), 170.9 (H**C**OO), 162.8 (O=**C**-CH_2-N), 153.7 (1´-**C**), 145.2 (CH_2-**C**=C), 142.8 (3´-**C**), 136.9 (N-**C**=CH-CH), 133.6 (N-**C**=CH-CH), 123.0 (2´-**C**), 122.5 (2C, **C**H=C-NH), 122.3 (CH_2-C=**C**H-N), 122.0 (2C, **C**H=C-N_3), 72.1 (N-O-

CH_2) 71.6 (4′-**C**), 64.4 (5′-**C**), 65.1 (**C**H_2-N(CH_3)$_3$), 53.3 (N(**C**D_3)$_3$), 34.1 (NH-**C**H_2). **HRMS (ESI$^+$):** calc. for $C_{21}H_{21}D_9N_7O_5^+$ $[M]^+$: 469.2868; found: 469.2874.

2. Cell Culture and Transfection Procedures

DMEM high glucose containing 10% FBS (*PAN Biotech*), 2 mM L-glutamine, 100 U/mL penicillin, 100 µg/mL streptomycin, 1x MEM Non-essential Amino Acid Solution and 0.1 mM β-mercaptoethanol (*Sigma Aldrich*) was used as basal medium for mESC cultures. The mESC lines were maintained in naïve state on gelatin coated plates by supplementing basal medium with 1000 U/mL LIF (*ORF Genetics*), 3.0 µM *GSK3* inhibitor CHIR99021 and 1.0 µM Mek inhibitor PD0325901 (*2i*; *Selleckchem*). Metabolic labelling experiments with isotope-labeled nucleosides were performed by plating mESCs in priming conditions consisting of basal mESC medium supplemented with 1000 U/mL LIF. Labeled nucleosides (*B.A.C.H. UG*) were added to the culture medium at the following concentrations: dG* [$^{13}C_{10}{}^{15}N_5$] at a concentration of 100 µM for three days, followed by treatment with 200 µM labeled dG* for two days; dC* [$^{13}C_9{}^{15}N_3$] and dT* [$^{13}C_{10}{}^{15}N_2$] was used at a concentration of 100 µM for five days. J1 wild type mESCs were obtained from the 129S4/SvJae strain and Dnmt TKO J1 mESCs were described in *Li et al.* and *Tsumura et al*.[2-3] The Tdg wild type and the Tdg KO cell lines were reported in *Cortazar et al*.[4]

HEK293T cells were cultured in RPMI-1640 medium containing 10% FBS, 2 mM L-glutamine, 100 U/ml penicillin, 100 µg/ml streptomycin. ENC1 neural stem cells were cultured as previously described.[5]

2.1 Ape1 Inhibition

a) For treatment with the Ape1 inhibitor CRT0044876 (*Sigma-Aldrich*), mESCs were cultured for three days in priming conditions, passaged and re-plated in priming medium containing 100 µM CRT0044876 for 48 or 60 h. After 24 h fresh medium containing the inhibitor was added and DMSO was used as vehicle control. **b)** Wild type and Dnmt TKO mESCs were cultured for three days in priming medium containing 100 µM dC*, passaged and re-plated under priming conditions with the addition of 100 µM CRT0044876 (*Sigma-Aldrich*) and 100 µM dC* for 48 h. After 24 h fresh medium containing the inhibitor and 100 µM dC* was added. DMSO was used as vehicle control.

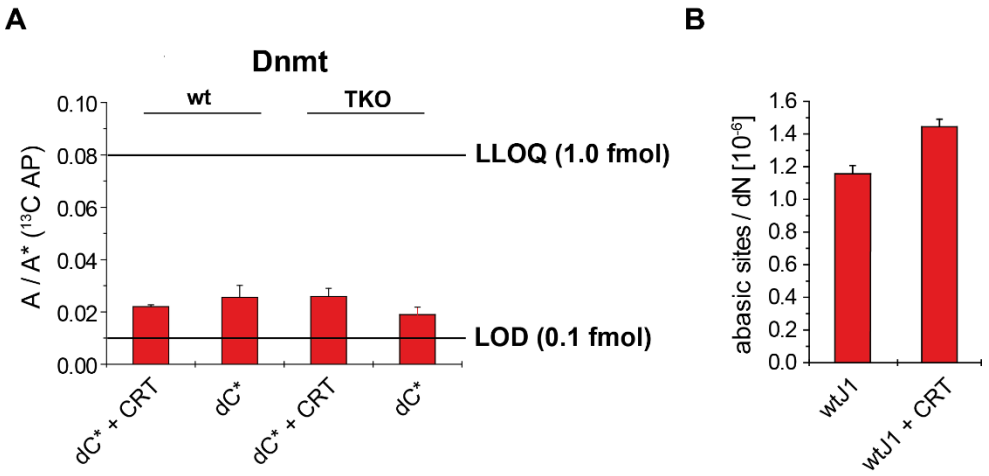


Figure S1: AP-sites after inhibition of Ape1. **(A)** A/A* ratios of labeled AP-sites in wild type and Dnmt TKO mESCs were over the LOD but below the LLOQ. **(B)** Quantification of unlabeled AP-sites of the dC* feeding experiment show an increase by treatment with the inhibitor CRT0044876.

2.2 Ape1 Knockdown

mESCs were cultured under priming conditions for three days, passaged and transfected with Apex1 Mission®esiRNA (purchased from *Sigma-Aldrich*) at 50 nM using Lipofectamine 2000 according to the manufacturer. The medium was changed after 24 h and cells were harvested 48 h post transfection. For qPCR analysis a second independent knockdown experiment was performed.

2.3 Cell lysis and DNA isolation

Isolation of genomic DNA was achieved using the *QIAamp DNA Mini Kit* from *Qiagen*. All mESC samples were washed with PBS (*Sigma*) and directly lysed in the plates by adding G2 buffer containing 400 µM of 2,6-di-tert-butyl-4-methylphenol (BHT) and desferoxamine mesylate (DM). DNA was sheered by bead milling in a microfuge tube using one 5 mm diameter stainless steel bead per tube and MM400 bead mill (*Retsch*) at 30 Hz for one minute and subsequent centrifugation at 15000 rpm for ten minutes. Depending on the amount of genomic DNA to isolate, the cell lysate was treated with proteinase K (25 µL for *genomic tips 20G* or 100 µL for *genomic tips 100G*) and RNase A (2.0 µL/20G, 10 µL/100G) at 50 °C for one hour. After 30 minutes, additional RNase A (2.0 µL or 10 µL, respectively) was added to the mixture. *Genomic tip* columns were then equilibrated with *QBT* loading buffer (1.0 mL/*20G* or 4.0 mL/*100G*) and then the lysate, which was vortexed for one minute, was applied on the columns. After the entire liquid had entered the column, washing steps were carried out with *QC* buffer (2.0 mL/*20G* or 2x7.5 mL/*100G*) and the genomic DNA was finally eluted with *QF* buffer (2.0 mL/*20G* or 5.0 mL/*100G*) supplemented with 400 µM BHT. Precipitation was then achieved through the

addition of *i*-PrOH (1.4 mL/*20G* or 3.5 mL/*100G*, 70% Vol) and the resulting genomic DNA pellet was centrifuged (15 minutes, 6000 g, 4 °C). The supernatant was discarded, washing steps were carried out using 70% EtOH (5.0 mL, 15 minutes, 6000 g, 4 °C). Finally, the pure DNA pellet was resuspended in 1.0 mL 70% EtOH and centrifuged (10 minutes, 15000 rpm, 4 °C). Next, the supernatant was removed and the pellet was re-dissolved in ddH_2O (50 - 100 μL) with 20 μM BHT. The concentration was finally determined with a *NanoDrop (ND 1000, Peqlab)*.

2.4 Artificial Generation of AP-Sites

Salmon sperm DNA (à 5.0 ug, *Sigma Aldrich*) was heated in a 10 mM sodium citrate buffer (pH = 5.0) for three and ten minutes at 70 °C, respectively. The mixture was subsequently neutralized to pH = 7.5 using 0.2 mM NaOH and then derivatized and quantified as described before. The data shows that harsh conditions need to be applied in order to generate AP-sites artificially.

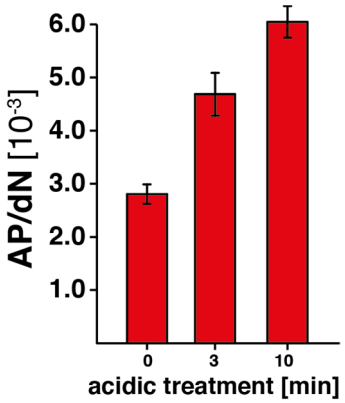


Figure S2: Artificial generation of AP-sites under harsh conditions.

2.5 Re-Isolation of wtJ1 mESC DNA

Reisolation of genomic DNA was performed by adding 30 μg of DNA to 3200 μL of G2 buffer. The DNA containing lysis buffer was applied to a pre-equilibrated 20G column and DNA isolation was performed as described (see cell lysis and DNA isolation). This experiment shows that the isolation procedure does not introduce β–elimination product artificially.

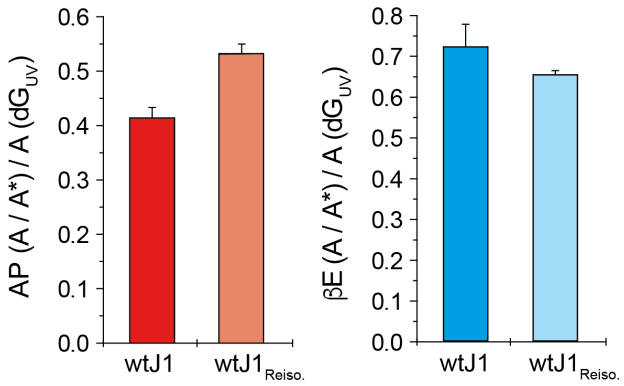


Figure S3: A/A* ratios of AP- and βE-sites normalized by dG. The DNA of wild type mESCs shows similar values of the BER intermediates before and after reisolation.

2.6 Quantification of AP- and βE-Sites in Different Cell Lines

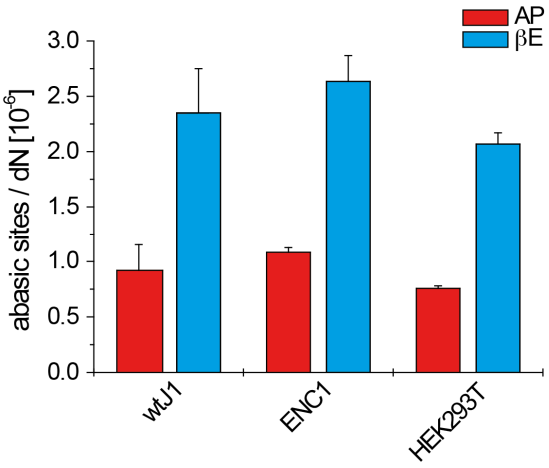


Figure S4: Quantification of AP- and βE-intermediates in wild type mESCs, ENC1 neural stem cells and somatic HEK293T cells shows similar values for the BER intermediates in all three cell lines. The quantification results of one biological replicate is depicted.

2.7 RNA Isolation and Real-Time Quantitative Polymerase Chain Reaction (qPCR)

At 48 h post-transfection, total RNA was isolated with the peqGOLD Total RNA Kit (*PeqLab*) according to the manufacturer's instructions. Prior to the qPCR, 1 µg of RNA was transcribed to 1 µg of cDNA using the iScript cDNA Synthesis Kit (*Bio-Rad*) according to the manufacturer. The final cDNA concentration was 50 ng/µL. To 10 µL of iTaq Universal SYBR Green Supermix (*Bio-Rad*) 2 µL of cDNA were added per reaction. For each primer pair, the forward and the reverse primer were mixed together in equal amounts and diluted with nuclease-free ddH_2O to a final concentration of 1.25 µM. 12 µL of cDNA/iTaq mastermix and 8 µL of primer mix were added per well. Each sample was measured in triplicates on a MasterCycler RealPlex (*Eppendorf*) with the following PCR conditions: Step 1 95 °C (2:00 minutes), Step 2 95 °C (0:15 minutes), Step 3 55 °C (0:15 minutes), Step 4 72 °C (0:20 minutes); Step 2 – 4 were repeated 40 times.

Ape1 forward primer: 5'-GGTCAGCTCCGTCAGACAAA-3'

Ape1 reverse primer: 5'-TCGGAAGGCTTCATCCCAAC-3'

Alpha-tubulin forward primer: 5'-TGTGGATTCTGTGGAAGGCG-3'

Alpha-tubulin reverse primer: 5'-AGCACACATTGCCACATACAAA-3'

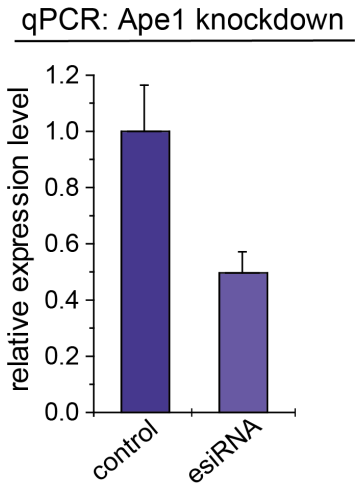


Figure S5: Relative mRNA expression levels show an estimated knockdown efficiency of about 50% using Ape1 (Apex1) esiRNA. As housekeeping reference gene alpha-tubulin was used.

3. Quantification of Abasic Sites

3.1 Derivatization of Genomic DNA with 1a

Derivatization of abasic sites (5.0 μg for unlabeled gDNA, 20 μg for labeled gDNA) with **1** was carried out in a total volume of 20 μL, whereas the solution was buffered with HEPES (20 mM, pH = 7.5) and Na_2EDTA (0.1 mM). A stock of **1a** in H_2O (23.8 mM) was added to the buffered solution (final concentration of **1a =** 1.5 mM) and the reaction was started by vortexing the mixture for 5 seconds. The gDNA was incubated for 40 minutes at 37 °C/1400 rpm in an *Eppendorf Comfort* thermomixer. The reaction was stopped via addition of 1-naphthylaldehyde (66.7 μL, 2M in *i*-PrOH) to quench excess of **1a** and incubated again for 10 minutes at 37 °C/1400 rpm. Derivatized DNA was then precipitated via addition of NaOAc (3.3 μL, 3M), vortexing and incubation at 37 °C/1400 rpm for another 5 minutes. After absolute *i*-PrOH (66.7 μL) was added, the tubes were inverted several times and then centrifuged (60 minutes, 10 °C, 15000 rpm). The supernatant was removed and washing steps were carried out (1x75% *i*-PrOH, 10 °C, 15000 rpm ,30 minutes; 2x75% cold EtOH, 4 °C, 15000 rpm, 30 minutes), whereas after each washing step the supernatant was carefully removed. The resulting DNA pellet was finally re-dissolved in 35 μL of ddH_2O and then enzymatically digested to the nucleoside level.

3.2 Enzymatic Digestion of Derivatized Genomic DNA

For enzymatic digestion, genomic DNA (5.0 µg for unlabeled gDNA or 20 µg for labeled gDNA in 35 µL H_2O) was incubated in an aqueous solution containing 480 µM $ZnSO_4$, 5 U Antarctic phosphatase (*New England BioLabs*) and 42 U nuclease S1 (*Aspergillius oryzae*, *Sigma-Aldrich*) at 37 °C for 3 h. The solution also contained specific amounts of labeled internal standards for accurate quantification of DNA-modifications and the derivatized abasic sites. In the second digestion round, 0.2 U snake venom phosphodiesterase I (*Crotalus adamanteus*, *USB corporation*) in 7.5 µl of a 520 µM $[Na]_2$-EDTA solution was added and the mixture was incubated for further 3h or overnight at 37 °C (when digesting 20 ug of gDNA). After digestion, the samples were stored at −20 °C and filtered using an *AcroPrep Advance* 96 filter plate 0.2 µm (0.20 µm *Supor*, *Pall Life Sciences*) prior to LC-MS/MS analysis (39 µL injection volume at 4 °C).

3.3 LC-ESI-MS/MS Analysis of DNA Samples

For the LC-MS/MS studies a triple quadrupole mass spectrometer (*Agilent 6490*) and an *Agilent 1290* UHPLC system with UV detector were used. Based on earlier published work[6-10], a new method was developed that was coupled to the isotope dilution technique, allowing an exact quantification of derivatized abasic sites, all canonical nucleosides and cytosine modifications in one single analytical run. The chromatographical separation was performed on a *Poroshell* 120 SB-C8 column (*Agilent*, 2.7 µm, 2.1 mm × 150 mm). Elution buffers were H_2O and MeCN, each containing 0.0085% (v/v) formic acid, at a flow rate of 0.35 ml/min at 30 °C. The gradient was: 0 → 5 min; 0 → 3.5% (v/v) MeCN; 5 → 6.9 min; 3.5 → 5% MeCN; 6.9 → 13.2 min; 5 → 80% MeCN; 13.2 → 14.8 min; 80% MeCN; 14.8 → 15.3 min; 80 → 0% MeCN; 15.3 → 17 min; 0% MeCN. The eluent up to 1.5 min and after 12.2 min was diverted to waste by a *Valco* valve.

The source-dependent parameters were as follow: gas temperature 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 (positive mode) and −2250 V (negative ion mode), nozzle voltage 500 V, fragmentor voltage 380 V, Δ EMV 500 (positive mode) and 800 (negative mode). Compound-dependent parameters which gave highest intensities during method development are summarized in **Supplementary Table 1**.

Supplementary Table 1. Compound-dependent LC-MS/MS-parameters used for the analysis of genomic DNA. CE: collision energy, CAV: collision cell accelerator voltage, EMV: electron multiplier voltage. The nucleosides were analyzed in the positive ($[M+H]^+$ species) as well as the negative ($[M-H]^-$ species) 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 1.5–3.8 minutes								
[$^{15}N_2$]5cadC	274.08	*wide*	158.03	*wide*	170	5	5	Positive
5cadC	272.09	*wide*	156.04	*wide*	170	5	5	Positive
[$^{15}N_2$,D_2]5hmdC	262.12	*enhanced*	146.07	*enhanced*	40	27	1	Positive
5hmdC	258.11	*enhanced*	142.06	*enhanced*	40	27	1	Positive
[D_3]5mdC	245.13	*enhanced*	129.09	*enhanced*	30	60	1	Positive
5mdC	242.11	*enhanced*	126.07	*enhanced*	30	60	1	Positive
dC	228.12	*enhanced*	112.05	*enhanced*	25	5	5	Positive
[$^{13}C_9$,$^{15}N_3$]dC	240.12	*enhanced*	119.06	*enhanced*	25	5	5	Positive
Time segment 3.8–5.8 minutes								
[D_2]5hmdU	259.09	*wide*	216.08	*wide*	48	7	5	Negative
5hmdU	257.08	*wide*	214.07	*wide*	48	7	5	Negative
[$^{15}N_2$]5fdU	257.06	*wide*	213.05	*wide*	48	6	5	Negative
5fdU	255.06	*wide*	212.06	*wide*	48	6	5	Negative
Time segment 5.8–8.1 minutes								
[$^{15}N_5$]8oxodG	289.08	*wide*	173.04	*wide*	90	9	7	Positive
8oxodG	284.10	*wide*	168.05	*wide*	90	9	7	Positive
dG	268.10	*wide*	152.06	*wide*	75	45	3	Positive
[$^{13}C_{10}$,$^{15}N_5$] dG	283.12	*wide*	162.06	*wide*	75	45	3	Positive
[$^{15}N_2$]5fdC	258.09	*wide*	142.04	*wide*	50	5	5	Positive
5fdC	256.09	*wide*	140.05	*wide*	50	5	5	Positive
dT	243.1	*enhanced*	127.05	*enhanced*	35	40	3	Positive
[$^{13}C_{10}$,$^{15}N_2$]dT	255.12	*wide*	130.07	*wide*	50	8	5	Positive
Time segment 8.1–12.2 minutes								
1-Naphthyl-Oxim	500.24	*wide*	472.23	*wide*	5	19	5	Positive
9b_1	487.30	*wide*	459.29	*wide*	38	19	5	Positive
9b_2	487.30	*wide*	201.18	*wide*	38	40	5	Positive
[$^{13}C_5$]9a _1	483.26	*wide*	455.25	*wide*	38	19	5	Positive
[$^{13}C_5$]9a _2	483.26	*wide*	192.13	*wide*	38	40	5	Positive
9a_1	478.24	*wide*	450.23	*wide*	38	19	5	Positive
9a_2	478.24	*wide*	192.13	*wide*	38	40	5	Positive
10b_1	469.29	*wide*	441.28	*wide*	38	19	3	Positive
10b_2	469.29	*wide*	201.18	*wide*	38	33	3	Positive
[$^{13}C_5$]10a _1	465.23	*wide*	437.22	*wide*	38	19	3	Positive
[$^{13}C_5$]10a_2	465.23	*wide*	192.13	*wide*	38	34	3	Positive
10a _1	460.23	*wide*	432.22	*wide*	38	19	3	Positive
10a_2	460.23	*wide*	192.13	*wide*	38	34	3	Positive
1b	371.25	*wide*	343.24	*wide*	5	19	5	Positive
1a	362.19	*wide*	334.19	*wide*	5	19	5	Positive

For the chromatographical separation of enzymatically hydrolysed synthetic DNA containing the adduct **9a**, a slightly different gradient was used: 0 → 5 min; 0 → 3.5% (v/v) MeCN; 5 → 6.9 min; 3.5 → 5% MeCN; 6.9 → 12.2 min; 5 → 80% MeCN; 12.2 → 13.8 min; 80% MeCN; 13.8 → 14.3 min; 80 → 0% MeCN; 14.3 → 16 min; 0% MeCN.

The source-dependent parameters were the same as for genomic DNA measurements except for Δ EMV which was set to 300 V. The time segments were 0 - 1.5 minutes; 15.5 - 4 minutes; 4 - 8 minutes and 8 - 12.2 minutes. Fragmentation parameters for **9a/b** and **1a/b** are summarized in **Supplementary Table 2**.

Supplementary Table 2. Compound-dependent LC-MS/MS-parameters for **9a/b** and **1a/b** used for the analysis of a synthetic DNA oligo containing **9a**. CE: collision energy, CAV: collision cell accelerator voltage, EMV: electron multiplier voltage. The nucleosides were analyzed in the positive ($[M+H]^+$ species) as well as the negative ($[M-H]^-$ species) ion selected reaction monitoring mode (SRM).

Time segment 8 –12.2 minutes								
9b_1	487.30	*wide*	459.29	*wide*	38	13	2	Positive
9b_2	487.30	*wide*	201.18	*wide*	38	25	2	Positive
9a_1	478.24	*wide*	450.23	*wide*	38	13	2	Positive
9a_2	478.24	*wide*	192.13	*wide*	38	25	2	Positive
1b	371.25	*wide*	343.24	*wide*	5	10	1	Positive
1a	362.19	*wide*	334.19	*wide*	5	10	1	Positive

3.4 Method Development

Method validation and data processing were performed as described in earlier published work.[7] In order to obtain calibration curves, each standard (5–8 standard concentrations) was analysed as technical triplicates and linear regression was applied using Origin® 6.0 (Microcal™). Therefore, the ratio of the area under the curve (A/A*) of the unlabeled derivatized abasic sites **9a** or **10a**, respectively, to the internal standard (*) was plotted against the ratio of the amount of substance (n/n*) of the unlabeled derivatized abasic site **9a** and **10a**, respectively, to the internal standard (*) (Figure S3). Calibration functions were calculated without weighing. Acceptable precision (< 20% relative s.d.) and accuracy (80–124%) was achieved. The precision was obtained when A/A* ratios measured in technical triplicates for each calibration standard had standard deviations <20%. The accuracy was the ratio of the used to the calculated amount of substance in percent for each concentration. To prove the accuracy, we used the respective calibration function for calculation of the substance amount n from A/A* ration for each calibration standard.

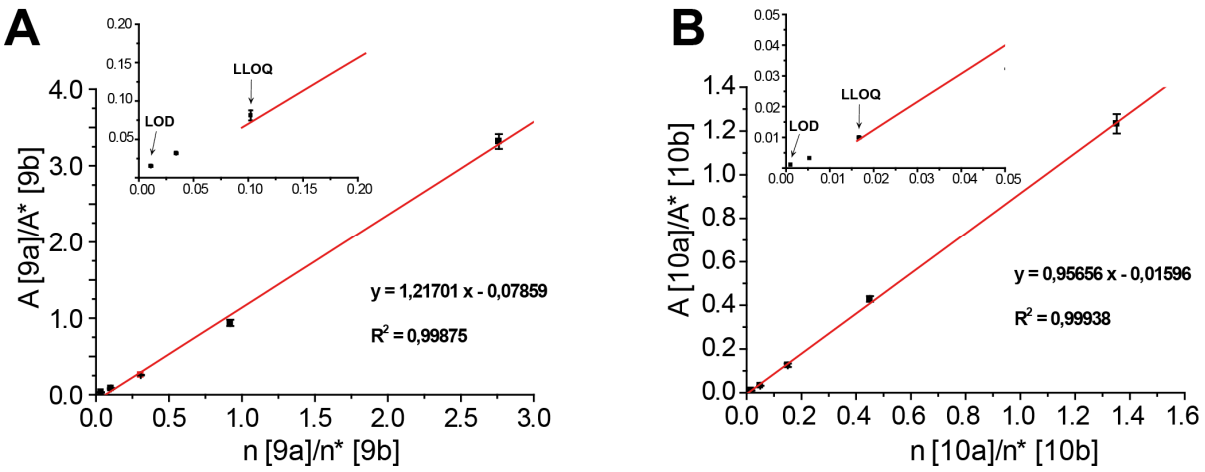


Figure S6: Calibration curves for **(A)** AP- (**9**) and **(B)** βE -site (**10**).

The lower limit of detection was defined as the detected amount that is three times higher compared to the blank response (LOD). The lower limit of quantification (LLOQ) was defined as the lowest concentration fulfilling the requirements of accuracy and precision and achieving an amount higher than LOD. A compilation of LLOQs and LODs is shown in **Supplementary Table 3**.

Supplementary Table 3. Compilation of absolute lower limits of quantification [fmol] (LLOQ and relative LLOQs [per dN] depending on the amount of digested DNA. The relative LLOQs were calculated by generating ratios of the absolute LLOQ [fmol] to the total amounts of nucleosides (N; [fmol]) in the respective amount of DNA [µg]. The total amounts of nucleosides were obtained by using the average molar mass of 308.91 g mol^{-1} for the monomeric DNA entity by taking the G-content (21% G) in mESC into account.

	Absolute LOD [fmol]	**Absolute LLOQ [fmol]**	**Relative LLOQ [per dN]**	**Relative LLOQ [per dN]**
DNA amount			**5 µg**	**20 µg**
9a	0.11	1.02	6.3E-08	1.57E-8
10a	0.11	1.01	6.3E-08	1.56E-8

3.5 Preparation of a Synthetic 13-mer with Defined Abasic Site

Oligonucleotides (5'-GTA ATG UGC TAG G-3' and 3'-CAT TAG ACG ATC C-5', à 15 nmol, *Metabion*) were incubated in *UDG*-buffer (150 μL, 20 mM Tris-HCl, pH = 8.0, 1 mM DTT, 1.0 mM EDTA, *New England Biolabs*) at 95 °C for 5 minutes and then slowly cooled to rt. UDG (5.0 μL, 25 units, *New England Biolabs*) was added, carefully mixed and the mixture was incubated for 2 hours at 37 °C. The oligonucleotide was then isolated by chloroform/phenol extraction as described in the following paragraph. A $CHCl_3$/phenol solution (200 μL, *Roti Phenol*) was added, vortexed for 30 seconds and centrifuged for 3 minutes at rt and 13400 rpm. The aqueous phase was carefully removed and $CHCl_3$/phenol treatment was repeated twice. After addition of NaOAc (20 μL, 3M), the oligonucleotide was precipitated with *i*-PrOH (600 μL). The resulting DNA pellet was centrifuged at rt for 30 minutes (15000 rpm), washed with cold EtOH (300 μL) and centrifuged at 4 °C and 15000 rpm for another 30 minutes. The washing step was repeated once more, the supernatant removed and the pellet was dried on air for five minutes before the oligonucleotide was re-dissolved in ddH_2O (150 μL). The identity was finally confirmed by MALDI-TOF analysis.

3.6 Reaction Kinetics on a Synthetic Oligo with a Defined Abasic Site

In a total reaction volume of 20 μL, the oligonucleotide (300 pmol) was buffered with a solution of HEPES (20 mM, pH = 7.5), Na_2EDTA (0.1 mM) and **1a** (1.26 μL of 23.8 mM stock). The reaction (37 °C, 800 rpm, *Eppendorf comfort* thermomixer) was started by vortexing the mixture for 5 seconds and after specific time points (t = 15 s, 30 s, 45 s, 90 s, 120 s, 150 s, 180 s, 4 min, 6 min, 8 min, 15 min, 20 min) stopped via addition of acetone (200 μL). Excess acetone was removed on a speed vac (RVC-2-33 IR, *Christ*) and was filtered on a *AcroPrep Advance 96* filter plate (0.20 μm *Supor*, *Pall Life Sciences*). 75 pmol of DNA were subsequently injected into a *Dionex micro HPLC* system and reaction products were separated using a *Zorbax SB-C_{18}* column (0.55 x 250 mm, 5.0 μm pore size) with a flow rate of 350 μL/min. The analysis was run at a column temperature of 60 °C and a gradient of 0% -> 20% buffer B in 45 minutes (whereas buffer A = 10 mM TEAB, pH = 7.5 in H_2O and buffer B = 10 mM TEAB, pH = 7.5 in 80% MeCN/H_2O). Integration of the obtained UV signals (Figure S4) finally showed that the reaction of **1a** with abasic sites on an ODN is complete after 20 minutes and that no other fragments were generated under physiological conditions.

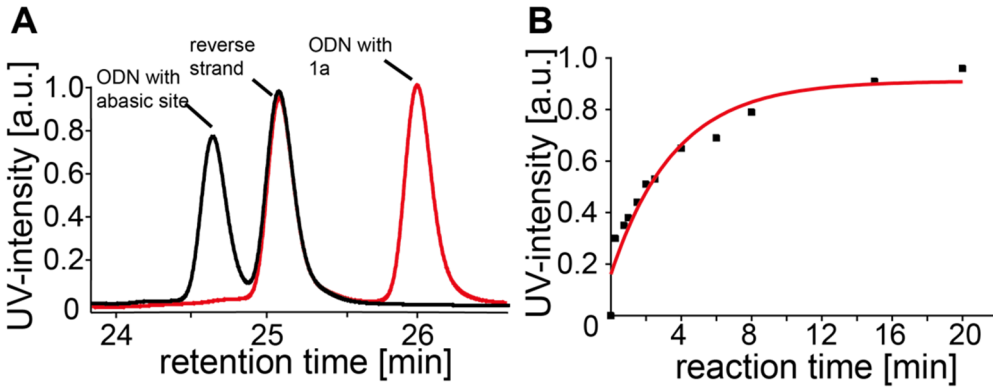


Figure S7: Reaction kinetics on oligo with defined abasic site. (A) Obtained UV-signals of ODN with abasic site and reverse strand before (black lines) and after derivatization with **1a**. **(B)** Normalized UV signals of ODN + **1** after specific time points.

3.7 *In vitro* Assay of a Synthetic Oligonucleotide with hOGG1

An 8oxoG containing synthetic 13mer 5´-GTAATG8oxoGGCTAGG-3´ and its counter strand 5´-CCTAGCCCATTAC-3´ were hybridized. hOGG1 activity was determined by treating 96 pmol of a single 8oxodG containing ds oligo in 1x buffer 2 (*New England BioLabs*), 100 µg/mL nuclease free bovine serum albumin (BSA) and 4.8 U hOGG1 (*New England BioLabs*). The reaction was incubated for 20 h at 37°C and stopped by heat inactivation at 65°C for 5 minutes. The ds oligo was recovered using the *Oligo Clean & Concentrator Kit (Zymo Research)* according to the manufacturer. The assay was performed in triplicates and diluted for exact quantification.

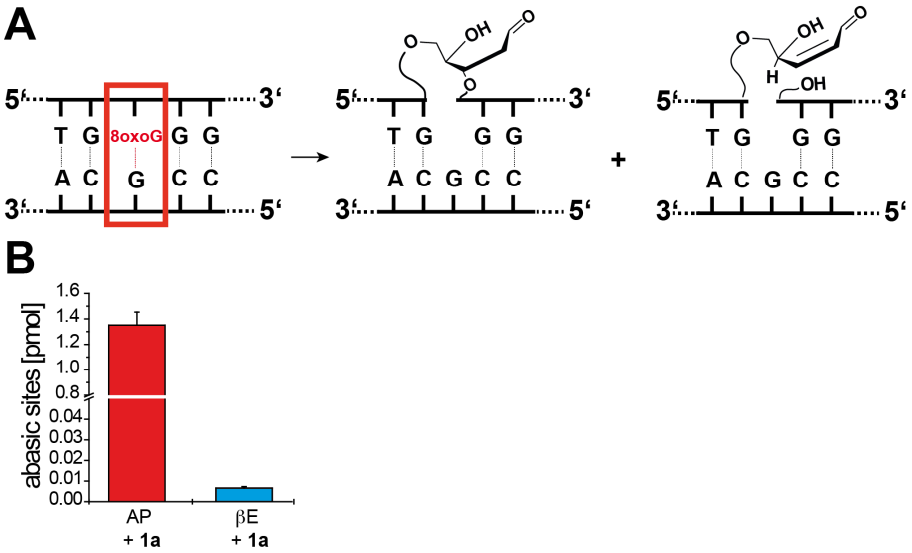


Figure S8: Reaction of an 8oxodG containing oligonucleotide with hOGG1. (A) Reaction of hOGG1 creating a defined AP- and βE-site inside a DNA duplex. **(B)** Quantification of the generated AP- and βE-sites after treatment of the oligonucleotide with **1a**.

3.8 Efficiency of Enzymatic Digestion

In order to verify that the derivatized abasic site adducts can be completely enzymatically excised from the DNA, an aliquot of the fully reacted AP-site of the ODN (see section before) was diluted 1/4000 and the amount of AP-site **9a** was quantified as described in the sections before. In this sample an amount of 133 pmol of **9a** was found (for the undiluted sample). The content of dG in this oligo was quantified to be 895 pmol and in total dG was present 7 times in the ODN. Hence, if all the derivatized abasic sites were efficiently hydrolyzed, an amount of 128 pmol of **9a** was expected which is in good agreement with the 133 pmol that were quantified.

3.9 Reaction Kinetics on Abasic Sites in Genomic DNA

Reactions were carried out by derivatizing 5 μg of gDNA with **1a** using the same conditions as mentioned above (*Derivatization of genomic DNA with **1a***). The reaction was stopped through the addition of 1-naphthylaldehyde (66.7 μL, 2M in *i*-PrOH) at specific time points (t = 1 min, 2.5 min, 5 min, 10 min, 20 min, 30 min, 60 min). Reaction aliquots were finally digested to the nucleoside level and quantified whereas it was shown that after one minute of reaction time, all abasic sites were derivatized and a prolonged incubation up to 60 minutes showed that no abasic sites are generated artificially under these conditions (Figure S5A). Since the reaction with genomic DNA was very fast, shorter time points were also investigated (t = 0 s, 10 s, 15 s, 30 s, 45 s, 1 min, 2 min, 5 min, 10 min, 15 min) and again, after one minute incubation time, all abasic- and βE- sites were derivatized (Figure S5B and C).

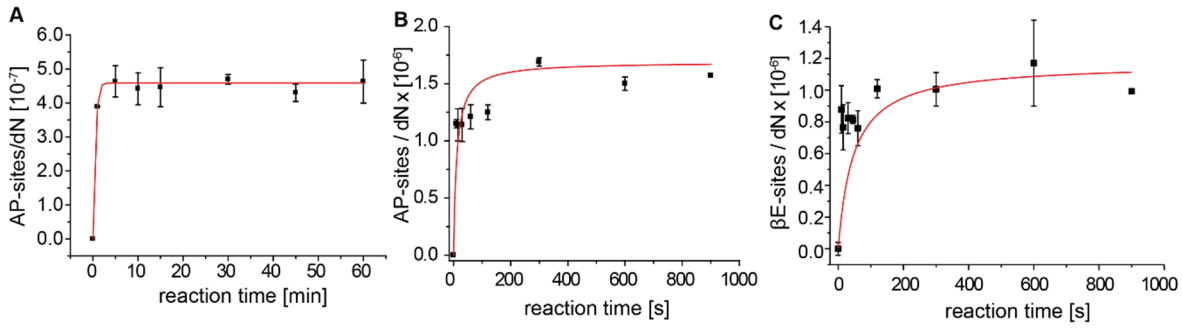


Figure S9: Derivatization of gDNA with **1** shows a fast derivatization reaction and does not artificially generate abasic sites.

Literature

1. Kojima, N.; Takebayashi, T.; Mikami, A.; Ohtsuka, E.; Komatsu, Y., *J. Am. Chem. Soc.* **2009,** *131*, 13208-13209.
2. Li, E.; Bestor, T. H.; Jaenisch, R., *Cell* **1992,** *69*, 915-26.
3. Tsumura, A.; Hayakawa, T.; Kumaki, Y.; Takebayashi, S.; Sakaue, M.; Matsuoka, C.; Shimotohno, K.; Ishikawa, F.; Li, E.; Ueda, H. R.; Nakayama, J.; Okano, M., *Genes Cells* **2006,** *11*, 805-14.
4. Cortazar, D.; Kunz, C.; Selfridge, J.; Lettieri, T.; Saito, Y.; MacDougall, E.; Wirz, A.; Schuermann, D.; Jacobs, A. L.; Siegrist, F.; Steinacher, R.; Jiricny, J.; Bird, A.; Schar, P., *Nature* **2011,** *470*, 419-423.
5. Liu, N.; Wang, M.; Deng, W.; Schmidt, C. S.; Qin, W.; Leonhardt, H.; Spada, F., *PLoS One* **2013,** *8*, e62755.
6. Cao, H.; Wang, Y., *J. Am. Soc. Mass Spectrom.* **2006,** *17*, 1335-41.
7. Pfaffeneder, T.; Spada, F.; Wagner, M.; Brandmayr, C.; Laube, S. K.; Eisen, D.; Truss, M.; Steinbacher, J.; Hackner, B.; Kotljarova, O.; Schuermann, D.; Michalakis, S.; Kosmatchev, O.; Schiesser, S.; Steigenberger, B.; Raddaoui, N.; Kashiwazaki, G.; Müller, U.; Spruijt, C. G.; Vermeulen, M., *Nature Chem. Biol.* **2014,** *10*, 574-581.
8. Schiesser, S.; Pfaffeneder, T.; Sadeghian, K.; Hackner, B.; Steigenberger, B.; Schröder, A. S.; Steinbacher, J.; Kashiwazaki, G.; Höfner, G.; Wanner, K. T.; Ochsenfeld, C.; Carell, T., *J. Am. Chem. Soc.* **2013,** *135*, 14593-14599.
9. Spruijt, C. G.; Gnerlich, F.; Smits, A. H.; Pfaffeneder, T.; Jansen, P. W.; Bauer, C.; Munzel, M.; Wagner, M.; Müller, M.; Khan, F.; Eberl, H. C.; Mensinga, A.; Brinkman, A. B.; Lephikov, K.; Müller, U.; Walter, J.; Boelens, R.; van Ingen, H.; Leonhardt, H.; Carell, T.; Vermeulen, M., *Cell* **2013,** *152*, 1146-59.
10. Wang, J.; Yuan, B.; Guerrero, C.; Bahde, R.; Gupta, S.; Wang, Y., *Anal. Chem.* **2011,** *83*, 2201-2209.

7.3 Zusatzmaterialien zu Abschnitt 3.3

S. Schiffers, C. Ebert, R. Rahimoff, **O. Kosmatchev**, J. Steinbacher, A.-V. Bohne, F. Spada, J. Nickelsen, S. Michalakis, M. Müller, T. Carell, *Angew. Chem. Int. Ed.* **2017**, *56*, 11268-11271. *Quantitative LC-MS Provides No Evidence for m6dA or m4dC in the Genome of Mouse Embryonic Stem Cells and Tissues*

Angewandte
Eine Zeitschrift der Gesellschaft Deutscher Chemiker
Chemie

Supporting Information

Quantitative LC–MS Provides No Evidence for m^6dA or m^4dC 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**

anie_201700424_sm_miscellaneous_information.pdf

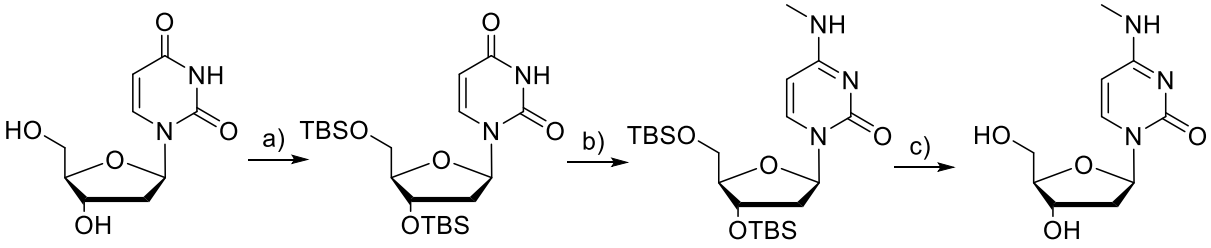
Supporting information

General Methods

Chemicals were purchased from Sigma-Aldrich, Fluka, Alfa Aesar, ChemGenes or Acros and used without further purification. Solutions were concentrated *in vacuo* on a Heidolph rotary evaporator. The solvents were of reagent grade and purified by distillation. Dry solvents were bought from Sigma-Aldrich. Acetonitrile for HPLC-purification was purchased from VWR. Water was purified by a Milli-Q Plus system from Merck Millipore. Chromatographic purification of products was accomplished using flash column chromatography on Merck Geduran Si 60 (40-63 µM) silica gel (normal phase). Thin layer chromatography (TLC) was performed on Merck 60 (silica gel F254) plates. ^{1}H and ^{13}C -NMR spectra were recorded in deuterated solvents on *Bruker ARX 300, Varian VXR400S, Varian Inova 400, Bruker AMX 600* and *Bruker Avance 800* spectrometers and calibrated to the residual solvent peak. Multiplicities are abbreviated as follows: s = singlet, d = doublet, t = triplet, q = quartet, hept = septet, m = multiplet, br. = broad. For assignment of the structures, additional 2D NMR spectra (COSY, HSQC, HMBC) were measured. High resolution electronspray ionization mass spectra (HRMS-ESI) were recorded on a Thermo Finnigan LTQ-FT (ESI-FTICR).

Synthesis of 2'-deoxy-4-methyl-cytidine

The final compounds were purified by reverse phase HPL chromatography (Macherey-Nagel, Nucleosil 100-7 C18, 10 × 250 mm, linear gradient from 0% to 10% acetonitrile in water in 45 min for 2'-deoxy-4-methyl-cytidine or 0% to 15% acetonitrile in water in 45 min for 2'-deoxy-6-methyladenosine).



Scheme S1: Synthesis of 2'-deoxy-4-methyl-cytidine. TBS-Cl, imidazole, pyridine, 0 °C to r.t., o/n, quant b) 1. NaH, THF, r.t., 45 min, 2. TPS-Cl, r.t., 15 h, 3. $MeNH_2$, r.t., 5 min, 86% c) HF in pyridine, EtOAc, r.t., o/n, quant.

Synthesis of 2'-deoxy-4-methyl-cytidine

***O*$^{3'}$,*O*$^{5'}$-TBS-2'-deoxyuridine**

2'-Deoxyuridine (3.00 g, 13,1 mmol, 1.0 eq) was dissolved in pyridine (50 mL) and cooled to 0 °C. Imidazole (4.00 g, 59.0 mmol, 4.5 eq) and TBS-Cl (5.90 g, 39.3 mmol, 3.0 eq) were then added. The mixture was allowed to warm to rt overnight and volatiles were subsequently removed *in vacuo*. The gummy residue was taken up in DCM (100 mL) and washed with sat. NH_4Cl (3x 100 mL). The organic phase was dried over Na_2SO_4 , the organic solvent removed under reduced pressure and the resulting oil was finally coevaporated with toluene (3x) to yield a white powder in quantitative yield (5.98 g, 13.1 mmol) that was used without further purification.

***O*$^{3'}$,*O*$^{5'}$-TBS-4-methyl-2'-deoxycytidine**

O$^{3'}$,*O*$^{5'}$-TBS-2'-deoxyuridine (1.00 g, 2.19 mmol, 1.0) was dissolved in THF (50 mL) and cooled to 0 °C. NaH (60% in mineral oil, 434 mg, 10.95 mmol, 5.0 eq) was added in one portion and the grey suspension was kept at 0 °C for 45 minutes. Subsequently, TPS-Cl (1.33 g, 4.38 mmol, 2.0 eq) was added and the mixture was warmed to rt and stirred for 15 hours. After all the starting material was consumed (as judged by TLC), ice cold sat. NH_4Cl (200 mL) was added cautiosly and the aqueous layer was extracted with EtOAc (2x 100 mL). After drying the organic phase over Na_2SO_4, volatiles were removed *in vacuo* and the residue was charged with $MeNH_2$ (50 mL, 33% in EtOH). After stirring at rt for 5 minutes, the reaction was aborted through addition of NH_4Cl (200 mL) and the aqueous phase was again extracted with EtOAc (2x 50 mL). After drying over Na_2SO_4 and removal of EtOAc *in vacuo*, the compound was purified by column chromatography (DCM --> 2.5% MeOH --> 4% MeOH) and the product was yielded as a colourless foam (887 mg, 1.80 mmol, 86%).

^{1}H-NMR (599 MHz, $CDCl_3$, ppm): δ = 7.87 (d, *J* = 7.8 Hz, 1H, 6-H), 6.29 (dd, *J* = 7.8 Hz, 6.3 Hz, 1H, 1'-H), 5.55 (d, *J* = 7.8 Hz, 1H, 5-H), 4.35 (q, *J* = 5.5 Hz, 1H, 3'-H), 3.90 (dd, *J* = 11.4 Hz, 2.6 Hz, 1H, 5'-H), 3.86 (m, 1H, 4'-H), 3.75 (dd, *J* = 11.1 Hz, 2.3 Hz, 1H, 5'-H), 3.0 (ds, *J* = 4.9 Hz, 3H, CH_3-NH), 2.44 - 2.36 (m, 1H, 2'-H), 2.07 (ddd, *J* = 13.5, 6.4, 4.9 Hz, 1H, 2'-H),0.96 - 0.82 (m, 18H, TBS-C(CH_3)$_3$), 0.14 - 0.01 (m, 12H, TBS-CH_3).

^{13}C-NMR (151 MHz, $CDCl_3$, ppm): δ = 164.2 (C4), 156.2 (C2), 139.5 (C6), 94.4 (C5), 87.1 (C4'), 85.7 (C3'), 70.1 (C1'), 61.9 (C5'), 42.1 (C2'), 28.0 (CH_3), 25.9 - 25.7 (TBS-C(**C**H_3)$_3$), 18.3 - 18.0 (TBS-**C**(CH_3)$_3$), -4.6 - -5.6 (TBS-**C**H_3).

HR-MS (ESI+): m/z calculated for [$C_{22}H_{44}N_3O_4Si_2^+$] ([M+H]$^+$): 470.2865, found: 470.2864.

2'-deoxy-4-methyl-cytidine

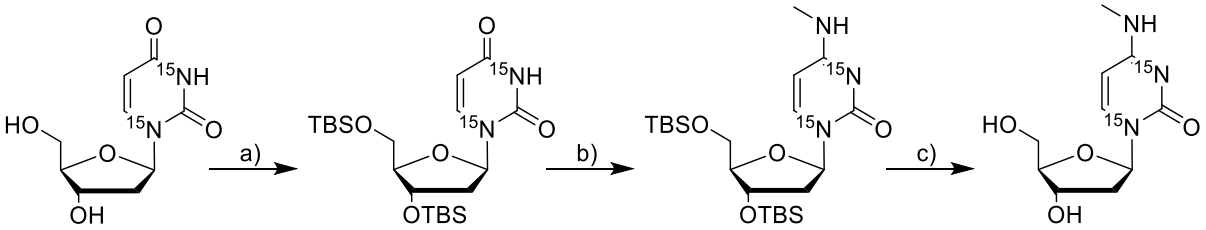
In a falcon tube, the TBS-protected nucleoside (300 mg, 0.64 mmol, 1.0 eq) was dissolved in EtOAc (5 mL). The mixture was then charged with HF•pyridine (70%, 374 µL, 14.4 mmol, 22.5 eq) and stirred at rt overnight. Then TMS•OMe (3.46 mL, 25.6 mmol, 40 eq) was added whereas a colorless precipitate was formed. The white solid was isolated through repeated centrifugation of the mixture (2x 10 minutes, 6000 rpm) and subsequent washing with EtOAc (2x 5 mL). The supernatant was discarded and the compound (154 mg, 0.64 mmol, quant.) was yielded as a white powder, that was further purified by HPLC.

^{1}H NMR (400 MHz, D_2O) δ 7.68 (d, *J* = 7.6 Hz, 1H, C6-H), 6.29 (t, *J* = 6.7 Hz, 1H, C1'-H), 5.98 (d, *J* = 7.6 Hz, 1H, C5-H), 4.45 (dt, *J* = 6.5, 3.9 Hz, 1H, C3'-H), 4.05 (dt, *J* = 5.1, 3.7 Hz, 1H, C4'-H), 3.84 (dd, *J* = 12.4, 3.7 Hz, 1H, C5'-H), 3.76 (dd, *J* = 12.4, 5.2 Hz, 1H, C5'-H), 2.89 (s, 3H, CH_3), 2.41 (ddd, *J* = 14.1, 6.4, 4.0 Hz, 1H, C2'-H), 2.29 (dt, *J* = 13.9, 6.8 Hz, 1H, C2'-H) ppm.

^{13}C NMR (101 MHz, D_2O) δ 164.5 (C4), 157.8 (C2), 139.3 (C6), 97.3 (C5), 86.4 (C4'), 85.8 (C1'), 70.6 (C3'), 61.3 (C5'), 39.1 (C2'), 27.1 (CH_3) ppm.

HR-MS (ESI+): m/z calculated for $[C_{10}H_{16}N_3O_4]^+$ ($[M+H]^+$): 242.1136, found: 242.1137.

Synthesis of 2'-deoxy-4-methyl-[*N*^{1},*N*^{3}-^{15}N$_2$]-cytidine



Scheme S2: Synthesis of 2'-deoxy-4-methyl-[*N*^{1},*N*^{3}-^{15}N$_2$]-cytidine. a) TBS-Cl, imidazole, pyridine, r.t., 72 h, 64% b) 1. NaH, THF, r.t., 45 min, 2. TPS-Cl, r.t., 17 h, 3. $MeNH_2$, r.t., 10 min, 49% c) HF in pyridine, EtOAc, r.t., 17 h, 64%.

***O*$^{3'}$,*O*$^{5'}$-TBS-2'-deoxy-[*N*^{1},*N*^{3}-^{15}N$_2$]-uridine**

2'-Deoxy-[*^{15}N,^{15}N*]-uridine (51 mg, 0.22 mmol, 1.0 eq), TBS-chloride (109 mg, 0.726 mmol, 3.3 eq) and imidazole (99 mg, 1.4 mmol, 6.6 eq) were dissolved in 1 mL pyridine and stirred at

r.t. for 72 h. The solvent was removed *in vacuo*, the residue was dissolved in DCM and washed with sat. $NaHCO_3$ (aq). *O*$^{3'}$,*O*$^{5'}$-TBS-2'-deoxy-[*N*^{1},*N*^{3}-$^{15}N_2$]-uridine (59 mg, 0.13 mmol, 64%) was obtained as a white solid.

HR-MS (ESI+): m/z calculated for $[C_{21}H_{41}{}^{15}N_2O_5Si_2]^+$ ($[M+H]^+$): 459.2489, found: 459.2493.

_O_$^{3'}$,_O_$^{5'}$-TBS-2'-deoxy-[_N_^{1},_N_^{3}-$^{15}N_2$]-cytidine

O$^{3'}$,*O*$^{5'}$-TBS-2'-deoxy-[*N*^{1},*N*^{3}-$^{15}N_2$]-uridine (59 mg, 0.13 mmol, 1.0 eq) was dissolved in THF (2 mL) and cooled to 0 °C. Sodium hydride (16 mg, 0.65 mmol, 5.0 eq) was added and the mixture was stirred at r.t. for 45 minutes. Then TPS-chloride (79 mg, 0.26 mmol, 2.0 eq) was added and the mixture was stirred at r.t. for another 17 h. The reaction was quenched with saturated $NH_4Cl_{(aq)}$, extracted with EtOAc and washed three times with saturated $NH_4Cl_{(aq)}$. The organic phase was dried with Na_2SO_4 and the solvent was removed *in vacuo*. The residue was dissolved in $MeNH_2$ (5 mL, 33% in EtOH) and stirred at r.t. for 10 minutes. The solvent was removed *in vacuo* and the residue was purified via column chromatography (DCM/MeOH 0.5% to 1%, SiO_2). *O*$^{3'}$,*O*$^{5'}$-TBS-2'-deoxy-4-methyl-[*N*^{1},*N*^{3}-$^{15}N_2$]-cytidine (30 mg, 0.064 mmol, 49%) was obtained as a white solid. Unreacted *O*$^{3'}$,*O*$^{5'}$-TBS-2'-deoxy-[*N*^{1},*N*^{3}-$^{15}N_2$]-uridine could be recovered via column chromatography.

HR-MS (ESI+): m/z calculated for $[C_{22}H_{44}N^{15}N_2O_4Si_2]^+$ ($[M+H]^+$): 472.2806, found: 472.2804.

2'-deoxy-4-methyl-[_N_^{1},_N_^{3}-$^{15}N_2$]-cytidine

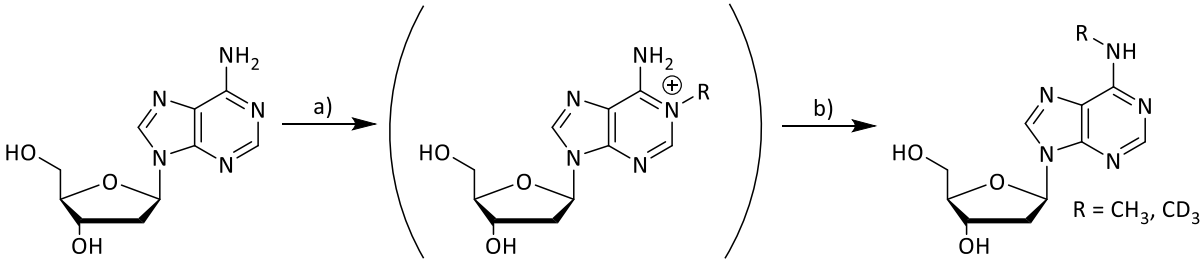
O$^{3'}$,*O*$^{5'}$-TBS-2'-deoxy-4-methyl-[*N*^{1},*N*^{3}-$^{15}N_2$]-cytidine (30 mg, 0.064 mmol, 1.0 eq) was dissolved in EtOAc (3 mL), HF in pyridine (40 µL, 1.4 mmol, 22.5 eq) was added and the reaction mixture was stirred at r.t. for 17 h. The mixture was cooled to 0 °C, quenched with TMS-OMe and stirred at r.t. for 4 h. The precipitation was centrifuged and the supernatant was removed and the solid was washed three times with EtOAc. The solvent was removed *in vacuo* and 2'-deoxy-4-methyl-[*N*^{1},*N*^{3}-$^{15}N_2$]-cytidine (10 mg, 0.041 mmol, 64%) was obtained as a white solid.

^{1}H NMR (599 MHz, D_2O) δ 7.68 (d, *J* = 7.4 Hz, 1H, C6-H), 6.29 (d, *J* = 7.7 Hz, 1H, C1'-H), 5.99 (dd, *J* = 7.8, 3.9 Hz, 1H, C5-H), 4.45 (h, *J* = 4.1, 3.5 Hz, 1H, C3'-H), 4.05 (m, 1H, C4'-H), 3.84 (m, 1H, C5'-H), 3.77 (m, 1H, C5'-H), 2.90 (d, *J* = 3.3 Hz, 3H, CH_3), 2.41 (m, 1H, C2'-H), 2.31 (q, *J* = 7.3 Hz, 1H, C2'-H) ppm.

^{13}C NMR (101 MHz, D_2O) δ 164.4 (d, *J* = 9.7 Hz, C4), 157.8 (C2), 139.2 (d, *J* = 13.2 Hz, C6), 97.3 (C5), 86.4 (C4'), 85.7 (d, *J* = 11.7 Hz, C1'), 70.6 (C3') , 61.3 (C5') , 39.0 (C2'), 27.0 (d, *J* = 2.0 Hz, CH_3) ppm.

HR-MS (ESI+): m/z calculated for $[C_{10}H_{16}N^{15}N_2O_4]^+$ ($[M+H]^+$): 244.1076, found: 244.1078.

Synthesis of 2'-deoxy-6-methyladenosine



Scheme S3: Synthesis of 2'-deoxy-6-methyladenosine and 2'-deoxy-6-[D_3]-methyladenosine. a) Methyl-p-toluenesulfonate, DMF, r.t., 20 h (R=CH_3); CD_3I, *N,N*-dimethylacetamide 28 °C, 20 h (R=CD_3), b) 2 M NaOH, 28 °C, 1 h, 13% (R=CH_3), 14% (R=CD_3).

2'-Deoxy-6-methyladenosine was synthesized as published in [1].

Methyl-*p*-toluenesulfonate (600 mg, 3.22 mmol, 4.0 eq) was dissolved in 1 mL *N,N*-dimethylformamide and 2'-Deoxyadenosine (200 mg, 0.796 mmol, 1.0 eq) was added. The solution was stirred at r.t. for 20 h. Celite was added and the solution was filtered. Acetone (30 mL) was added to the filtrate and the mixture was stirred for 1 h. The precipitation was filtered, washed with acetone and dried on high vacuum. The white solid was dissolved in 5 mL 2 M NaOH and stirred at 28 °C for 1 h. The solution was neutralized with 10% aqueous *p*-toluenesulfonic acid and the solvent was removed *in vacuo*. The residue was purified via column chromatography (DCM/MeOH 10:1 → 9:1, SiO_2) and 2'-deoxy-6-methyladenosine (27 mg, 0.10 mmol, 13%) was obtained as a white solid.

^{1}H NMR (800 MHz, DMSO-d_6) δ 8.32 (s, 1H, C8-H), 8.22 (s, 1H, C2-H), 7.77 (s, 1H, NH), 6.35 (dd, *J* = 7.9, 6.2 Hz, 1H, C1'-H), 5.30 (d, *J* = 4.0 Hz, 1H, C3'-OH), 5.24 (t, *J* = 5.9 Hz, 1H, C5'-OH), 4.41 (dq, *J* = 6.3, 3.1 Hz, 1H, C3'-H), 3.88 (q, *J* = 3.8 Hz, 1H, C4'-H), 3.62 (dq, *J* = 13.2, 4.4 Hz, 1H, C5'-H), 3.52 (ddt, *J* = 10.8, 7.7, 4.0 Hz, 1H, C5'-H), 2.95 (s, 3H, CH_3), 2.72 (ddd, *J* = 13.4, 7.7, 5.6 Hz, 1H, C2'-H), 2.26 (ddd, *J* = 13.2, 6.2, 2.8 Hz, 1H, C2'-H) ppm.

13C NMR (201 MHz, DMSO-d_6) δ 155.0 (C6), 152.4 (C2), 147.9 (C4), 139.3 (C8), 119.8 (C5), 88.0 (C4'), 84.0 (C1'), 71.0 (C3'), 61.9 (C5'), 39.4 (C2'), 27.0 (CH_3) ppm.

HR-MS (ESI+): m/z calculated for $[C_{11}H_{16}N_5O_3]^+$ ($[M+H]^+$): 266.1248, found: 266.1248.

Synthesis of 2'-deoxy-6-[*D*$_3$]-methyladenosine

2'-Deoxyadenosine (200 mg, 0.796 mmol, 1.0 eq) was dissolved in 1 mL *N,N*-dimethylacetamide and CD_3I (0.15 mL, 2.4 mmol, 3.0 eq) was added. The yellow solution was stirred at 28 °C for 20 h. Acetone was added and the precipitation was filtered, washed with acetone and dried on high vacuum. The white solid was dissolved in 5 mL 2 M NaOH and stirred at 28 °C for 1 h. The solution was neutralized with 10% aqueous *p*-toluenesulfonic acid and the solvent was removed *in vacuo*. The residue was purified via column chromatography (DCM/MeOH 10:1 → 9:1, SiO_2) and 2'-deoxy-6-[*D*$_3$]-methyladenosine (29 mg, 0.11 mmol, 14%) was obtained as a white solid.

^{1}H NMR (400 MHz, DMSO-*d*$_6$) δ 8.32 (s, 1H, C8-H), 8.22 (s, 1H, C2-H), 7.74 (s, 1H, NH), 6.35 (dd, *J* = 7.9, 6.1 Hz, 1H, C1'-H), 5.30 (d, *J* = 4.0 Hz, 1H, C3'-OH), 5.24 (s, 1H, C5'-OH), 4.41 (dq, *J* = 6.0, 2.9 Hz, 1H, C3'-H), 3.88 (td, *J* = 4.2, 2.5 Hz, 1H, C4'-H), 3.62 (m, 1H, C5'-H), 3.52 (d, *J* = 11.6 Hz, 1H, C5'-H), 2.72 (ddd, *J* = 13.4, 8.0, 5.7 Hz, 1H, C2'-H), 2.26 (ddd, *J* = 13.1, 6.1, 2.9 Hz, 1H, C2'-H) ppm.

^{13}C NMR (101 MHz, DMSO-*d*$_6$) δ 155.0 (C6), 152.3 (C2), 147.8 (C4), 139.2 (C8), 119.7 (C5), 88.0 (C4'), 83.9 (C1'), 70.9 (C3'), 61.9 (C5'), 39.4 (C2'), 26.4 (CD_3) ppm.

HR-MS (ESI+): m/z calculated for $[C_{11}H_{13}D_3N_5O_3]^+$ ($[M+H]^+$): 269.1436, found: 269.1437.

Cyanobacterial and Algal Strains, Culture Conditions and DNA Extraction

For DNA extractions, 150 mL of wild-type *Synechocystis* sp. PCC6803 cultures were grown at 30 °C in BG11 medium containing 5 mM glucose.[2] For *Chlamydomonas reinhardtii* the cell-wall-deficient strain 3491 (*cw15 mt-*), obtained from the *Chlamydomonas* Center (http://www.chlamy.org/), and the cell-walled laboratory strain 7d+ were cultured in 150-300 mL Tris-acetate-phosphate medium containing 1% sorbitol[3] at 25°C. All cultures were grown under orbital shaking and continuous irradiation with 30 μmol of photons m^{-2} s^{-1}.

For determination of N^6-methyldeoxyadenosine and N^4-methyldeoxycytidine, DNA was isolated by harvesting cells at 3000 g for 20 min, followed by resuspension of the pellets in 4-8 mL DNA extraction buffer ((2% (w/v) cetyltrimethylammonium bromide (CTAB),100 mM Tris-HCl, pH 8; 1.4 M NaCl; 20 mM EDTA pH 8; 2% (v/v) β-mercaptoethanol)). Cells were lysed upon addition of ~300 μL glass beads for 1 h at 65 °C and vortexed every 15 min. Proteins were removed by extraction with phenol:chloroform:isoamylalcohol 25:24:1 solution. RNA in the water phase was digested at 37 °C for 1 h using 0.1 mg/mL RNase A. RNase A was removed

by extraction with phenol:chloroform:isoamylalcohol 25:24:1 solution, followed by a chloroform/isoamylalcohol 24:1 extraction and isopropanol precipitation of the DNA. DNAs were dissolved in appropriate volumes of water at 4 °C overnight.

Preparation of mouse tissues

All mice used were on the wild-type C57-BL6 genetic background. All procedures concerning animals were performed with permission of the local authority (*Regierung von Oberbayern*). Brain und liver tissues of male wt mice (C57-BL6/N) were dissected at postnatal day 51 and immediately placed in 2 mL *Eppendorf*-tubes, snap frozen in liquid nitrogen and stored at -80 °C until use.

Genomic DNA isolation

Genomic DNA from murine tissues was extracted using the Qiagen Blood and Cell Culture DNA Midi Kit. Extraction was performed following the manufacturer's instructions for genomic DNA isolation from tissue samples. All buffers until loading of the sample on Genomic-tip 100/G were additionally supplemented with antioxidants 3,5-di-*tert*-butyl-4-hydroxytoluene (BHT, 200 μM) and desferoxamine mesylate salt (desferal, 200 μM). G2-buffer was additionally supplemented with the deaminase inhibitor tetrahydrouridine (THU, 200 μM), according to published methods, to reduce background oxidation or deamination.[4] Elution buffer QF was supplemented with 200 μM BHT.

mESC and HeLa samples differentiated in the presence of growth factors were lysed directly in the plates with RLT buffer (Qiagen) supplemented with BHT and desferal as described above. DNA was isolated using the Zymo Quick gDNA Midi Kit according to the manufacturer's instruction, except that elution was performed with 100 μl of bidest. water supplemented with BHT (0.2 μM).

Cell culture

Feeder independent wt J1 cells (strain 129S4/SvJae)[5] were cultured in the presence of serum and LIF as previously described[6] and routinely maintained on gelatinized plates in DMEM (PAA or Sigma) supplemented with 10% FBS, 1×MEM-nonessential amino acids (NEAA), 0.2 mM *L*-alanyl-*L*-glutamine, 100 U/ml penicillin, 100 μg/ml streptomycin (all from PAA), 0.1 mM β-mercaptoethanol, 20 ng/ml (≥1 ×10^3 U/ml) mouse recombinant LIF (ORF Genetics), 1 μM PD 0325901 and 3 μM CHIR 99021 (2i; both from Axon Medchem). Before DNA isolation, 2i cultures were passaged twice (over 4 d and 5 d, respectively) in DMEM

supplemented with FBS and LIF as above but lacking 2i. With this strategy, primed mESC cultures were obtained with no sign of overt differentiation.
As an additional experiment for determination of m^6dA as a marker of early embryogenesis, we cultured wt J1 cells in serum-free 2i/LIF on GelTrex-coated plates for more than 5 passages and harvested them with RLT buffer (Qiagen) supplemented with BHT and desferal.

Feeder independent wt TT2 cells were cultured like wt J1 cells (as described above) and primed for 4 d, 10 d, 12 d and 14 d respectively. Feeding experiments with dimethylsulfate (DMS, <0.01% DMSO) were performed on day four of the priming process and the cells were incubated with DMS for 24 h.

For isotope tracing with heavy methionine in serum-primed mESCs, 2i cultures of wt J1 cells were passaged twice (over 4 d or 5 d) without 2i in L-Metfree DMEM (Sigma) supplemented as above and with 0.2 mM of either [*methyl*-$^{13}C,D_3$]L-Met or natural L-Met. The medium was changed after 24 h. The experiment was performed in a biological duplicate.

Feeding of wt J1 cells with m^6dA was performed once by adding m^6dA nucleoside (1µM, 0.1% DMSO) to the medium on day 0 of culturing in FBS/LIF and increasing the concentration on day 2 for another two days (1 mM, 1% DMSO). DMSO controls were only treated with the respective concentration of DMSO.

HeLa 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 harvested with RLT buffer (Qiagen) supplemented with BHT and desferal, respectively.

Digest of genomic DNA

1 µg (10 µg) of genomic DNA in 35 µl H_2O were digested as follows: An aqueous solution (7.5 µl) of 480 µM $ZnSO_4$, containing 42 U (63 U) nuclease S1 (*Aspergillus oryzae*, Sigma-Aldrich), 5 U (7.5 U) Antarctic phosphatase (New England BioLabs) and specific amounts of labeled internal standards were 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 (0.2 U) snake venom phosphodiesterase I (*Crotalus adamanteus*, USB corporation), the sample was incubated for another 3 h (overnight) at 37 °C. Prior to LC-MS/MS analysis, samples were filtered by using an AcroPrep Advance 96 filter plate 0.2 µm Supor (Pall Life Sciences).

For dilution experiments, m^6dA and m^4dC were not spiked during the digest step, but after a dilution of the digest of 1:20, or 1:200 respectively.

Furthermore, a recovering experiment was performed as follows: a digest of 10 µg of murine DNA was spiked with a small amount of *Chlamydomonas* (CC-3491) DNA or synthetic m^6dA-nucleoside, which corresponds to a level of m^6dA as described earlier in mouse tissue.[7]

LC/MS analysis of DNA samples

Quantitative LC/UV-ESI-MS/MS analysis of digested DNA samples was performed using an Agilent 1290 UHPLC system equipped with a UV detector and an Agilent 6490 triple quadrupole mass spectrometer coupled with the stable isotope dilution technique. An improved method, based on earlier published work,[8] was developed, which allowed the concurrent analysis of all nucleosides in one single analytical run. The source-dependent parameters were as follows: gas temperature 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 2,500 V in the positive ion mode, capillary voltage −2,250 V in the negative ion mode and nozzle voltage 500 V. The fragmentor voltage was 380 V. Delta EMV was set to 500 (positive mode) and 800 (negative mode). Compound-dependent parameters are summarized in Table S1.

Exact quantification of the nucleosides m^5dC, hmdC, fdC, cadC and 8oxodG was performed with a previously published method.[6]

Table S1: Compound-dependent LC-MS/MS-parameters used for the analysis of genomic DNA. CE: collision energy; CAV: collision cell accelerator voltage; EMV: electron multiplier voltage. The nucleosides were analyzed in the positive ($[M+H]^+$species) as well as in the negative ($[M-H]^-$species) 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 1.5–6 min								
$[^{15}N_2]$-cadC	274.08	Wide	158.03	Wide	50	5	5	Positive
cadC	272.09	Wide	156.04	Wide	50	5	5	Positive
$[D_2$-$^{15}N_2]$-hmdC	262.12	Wide	146.07	Wide	50	27	1	Positive
hmdC	258.11	Wide	142.06	Wide	50	27	1	Positive
$[D_3]$-m^5dC	245.13	Wide	129.09	Wide	55	60	1	Positive
$[^{15}N_2]$-m^4dC	244.11	Wide	128.06	Wide	55	60	1	Positive
m^5dC/m^4dC	242.11	Wide	126.07	Wide	55	60	1	Positive
dC	228.1	Wide	112.05	Wide	1	1	0	Positive
$[^{15}N_2]$-cadC_2	158.03	Wide	140.09	Wide	50	13	7	Positive
cadC_2	156.04	Wide	138.03	Wide	50	13	7	Positive
Time segment 6-8 min								
$[D_2]$-hmdU	259.09	Wide	216.08	Wide	48	7	5	Negative
$[D_2]$-hmdU_2	259.09	Wide	126.05	Wide	48	7	5	Negative
hmdU	257.08	Wide	214.07	Wide	48	7	5	Negative
hmdU_2	257.08	Wide	124.04	Wide	48	7	5	Negative
$[^{15}N_2]$-fdU	257.06	Wide	213.05	Wide	48	6	5	Negative
$[^{15}N_2]$-fdU_2	257.06	Wide	141	Wide	48	6	5	Negative
fdU	255.06	Wide	212.06	Wide	48	6	5	Negative
fdU _2	255.06	Wide	140	Wide	48	6	5	Negative
$[^{15}N_2]$-dU	229.06	Wide	185.06	Wide	48	5	5	Negative
dU	229.06	Wide	185.06	Wide	48	5	5	Negative
Time segment 8-12 min								
$[^{15}N_5]$-8oxodG	289.09	Wide	173.04	Wide	60	9	7	Positive
8oxodG	284.1	Wide	168.05	Wide	60	9	7	Positive
$[D_3]$-m^6dA sensitive	269.14	Wide	153.1	Wide	55	30	1	Positive
m^6dA sensitive	266.12	Wide	150.08	Wide	55	30	1	Positive
$[^{13}CD_3]$-m^6dA	270.15	Wide	154.14	Wide	55	30	1	Positive
$[D_3]$-m^6dA insensitive	269.14	Wide	153.1	Wide	55	60	5	Positive
m^6dA insensitive	266.12	Wide	150.08	Wide	55	60	5	Positive
$[^{15}N_2]$-fdC	258.09	Wide	142.04	Wide	60	5	5	Positive
fdC	256.09	Wide	140.05	Wide	60	5	5	Positive
$[^{15}N_2]$-fdC_2	142.04	Wide	98.04	Wide	60	13	7	Positive
fdC _2	140.05	Wide	97.04	Wide	60	13	7	Positive

Chromatography was performed by a Poroshell 120 SB-C18 column (Agilent, 2.7 µm, 2.1 mm × 150 mm) at 30 °C using a previously published gradient[8] of water and MeOH, each containing 0.0075% (v/v) formic acid, at a flow rate of 0.35 ml/min. The effluent up to 1.5 min and after 12 min was diverted to waste by a Valco valve. The autosampler was cooled to 4 °C. The injection volume was amounted to 39 µl. Calibration curves are shown in Fig S4.

Method validation and data processing were performed as described in earlier published work. Calibration curves were obtained by analyzing each standard (5–8 standard concentrations) as technical triplicate and applying linear regression Origin® 6.0 (Microcal™). Therefore, the ratio of the area under the curve (A/A*) of the natural nucleoside to the internal standard was plotted against the ratio of the amount of substance (n/n*) of the natural nucleoside to the internal standard (see Fig. S4). Additionally, acceptable accuracy (80–120%) as well as precision (<20% RSD) was required. Accuracy was proven by calculating the amount of substance n from the obtained A/A* ratios of the calibration standards using the respective calibration function. Here, accuracy was defined as the ratio of the used amount of substance to the calculated amount of substance in percent and had to be between 80–120% for each standard concentration. Precision was determined as follows: technical replicates of A/A* ratios for each calibration standard had to have relative standard deviations (RSD) smaller than 20%. The lower limit of quantification (LLOQ) was defined as the lowest concentration fulfilling the requirements of accuracy and precision and achieving a response of at least three times the response compared with the blank response. A compilation of absolute and relative LLOQs is shown in Table S2.

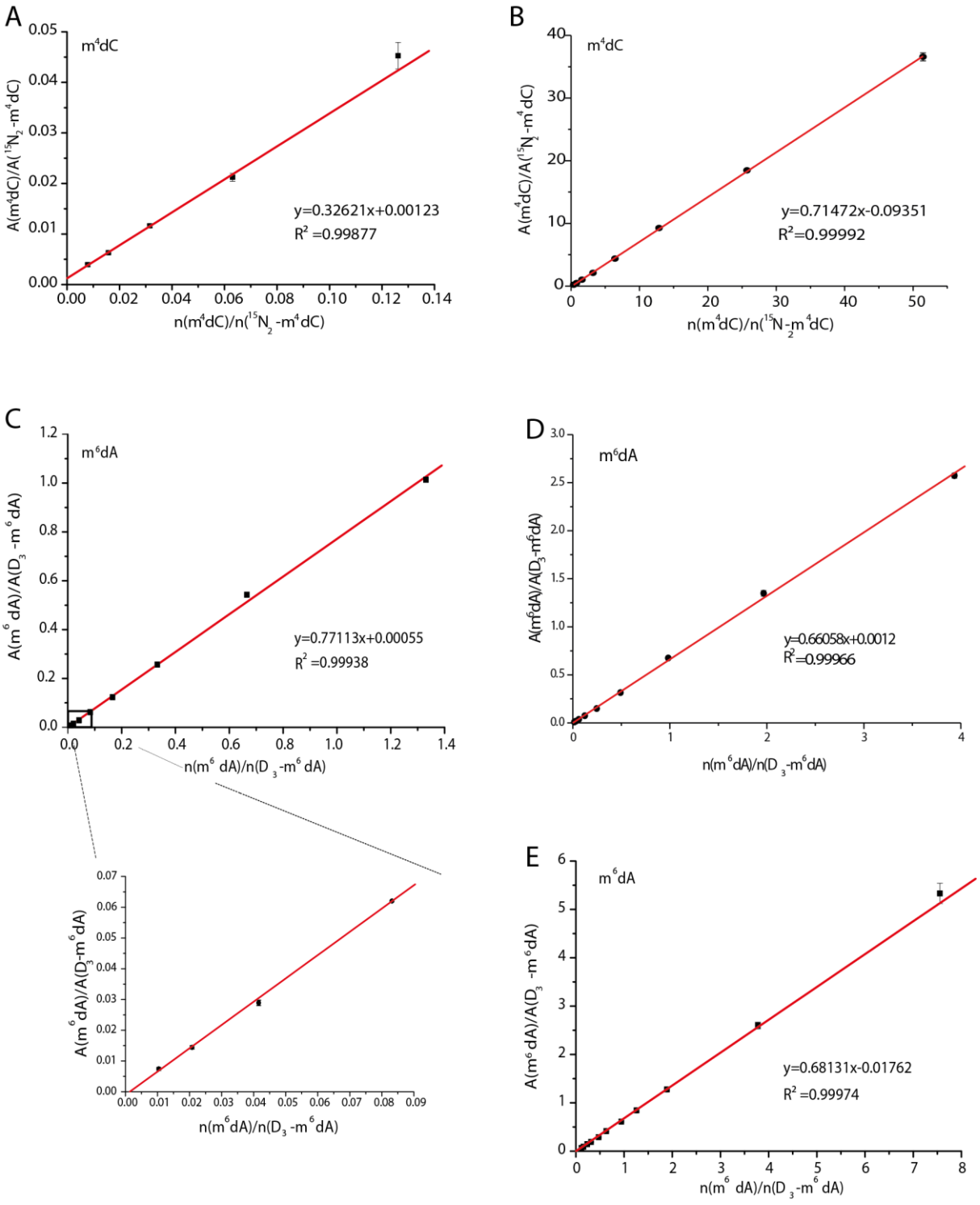


Figure S1: Calibration curves for m^4dC (upper row; A for murine, cell culture and *Chlamydomonas* DNA, B for *Synechocystis* DNA) and m^6dA (sensitive signals: C for mouse and cell culture DNA; insensitive signals: D for *Chlamydomonas* DNA and E for *Synechocystis* DNA); the ratio of the signal integral of unlabeled nucleoside towards labelled nucleoside is plotted against the ratio of the amount of natural nucleoside over the labelled nucleoside.

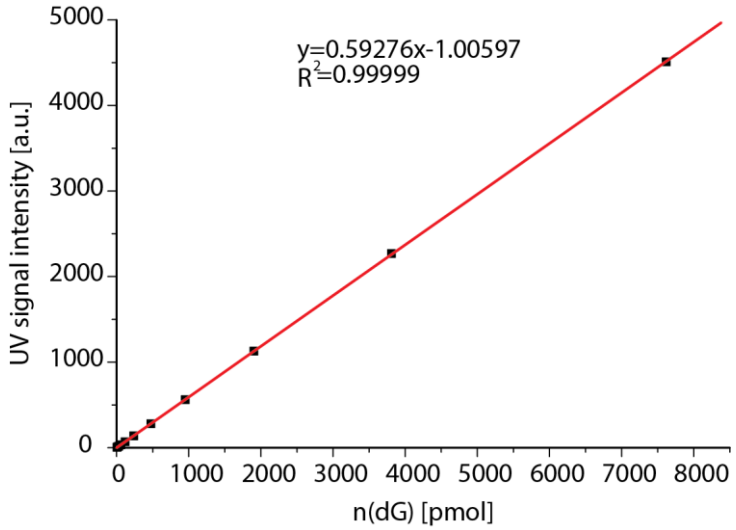


Figure S2: UV Calibration curve ford G with the UV signal intensity plotted against the amount of nucleoside [pmol].

Table S2: Compilation of absolute lower limits of quantification [fmol] (LLOQ; see Fig. S1) and relative LLOQs [per dN]. The relative LLOQs were computed by generating ratios of the absolute LLOQ [pmol] to the total amount of nucleosides (dN; [pmol]) in 10 µg of DNA (for calibration curve A and C) or 1 µg of DNA (for calibration curve B and D) or 50 ng (for calibration curve E). The total amount of nucleosides was obtained by calculating the content of dG using the respective calibration curve (see Fig. S2) and taking into account that murine DNA contains 21% dG, DNA of *Chlamydomonas* contains 32% dG and DNA of *Synechocystis* contains 23.9% dG.

	Absolute ULOQ [fmol]	Absolute LLOQ [fmol]	Relative ULOQ [per dN]	Relative LLOQ [per dN]	Relative LOD [per dN]	Absolute LOD [fmol]
m^4dC (small, A)	302.1	18.8	$1.1*10^{-5}$	$6.6*10^{-7}$	$1.4*10^{-6}$ (mouse/cell culture) $4.8*10^{-5}$ (*Chlamydomonas*)	40
m^4dC (big, B)	48100	376	$5.1*10^{-2}$	$3.9*10^{-4}$	$4.2*10^{-5}$	40
m^6dA (sensitive, C)	203.8	1.6	$7.4*10^{-6}$	$5.8*10^{-8}$	$8.0*10^{-8}$	2.2
m^6dA (insensitive,D)	1162	18.2	$1.4*10^{-3}$	$2.2*10^{-5}$	$2.0*10^{-5}$	17
m^6dA (insensitive,E)	15500	60	$2.6*10^{-1}$	$1.0*10^{-3}$	$4.3*10^{-4}$	26

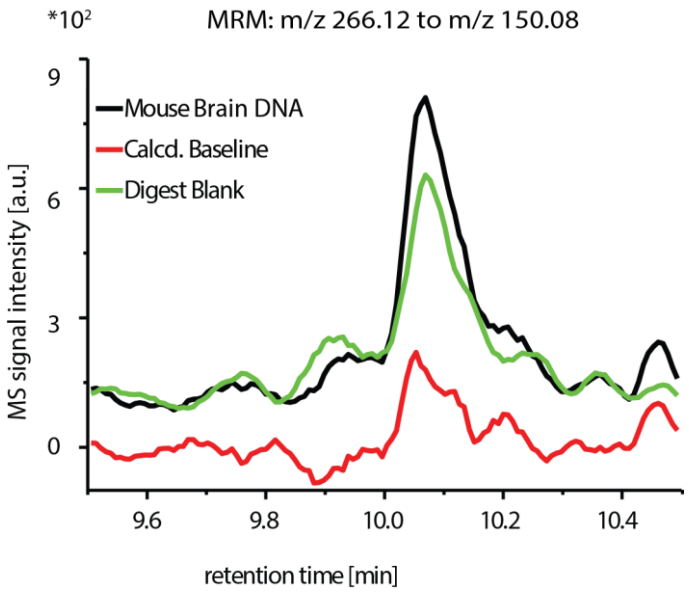


Figure S3: chromatogram of DNA from murine brain green: digest blank, in which everything was performed like in a real digest, but no DNA was added; red: calculated baseline, for which the chromatogram of the digest blank was subtracted from the chromatogram for DNA from murine brain.

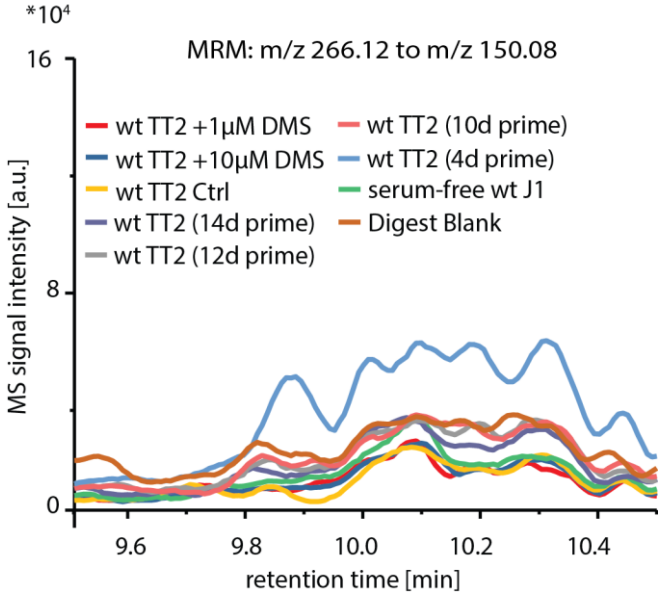


Figure S4: chromatogram of various cell culture samples; red: DNA from wt TT2 cells fed with 1 µM DMS, blue: DNA from wt TT2 cells fed with 10 µM DMS, yellow: DNA from a wt TT2 control, purple: DNA from wt TT2 cells after 14 d of priming, grey: DNA from wt TT2 cells after 12 d of priming, orange: DNA from wt TT2 cells after 10 d of priming, light blue: DNA from wt TT2 after 4 d of priming, green: DNA from wt J1 cells cultured in serum-free 2i/LIF, brown: digest blank, in which everything was performed like in a real digest, but no DNA was added

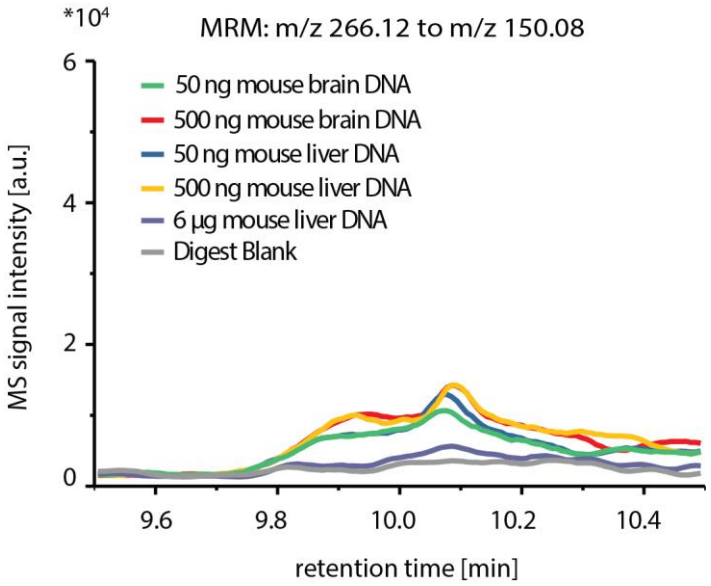


Figure S5: chromatogram of murine DNA from different tissues in different dilutions.; green: 50 ng of DNA from murine brain, red: 500 ng of DNA from murine brain, blue: 50 ng of DNA from murine liver, yellow: 500 ng of DNA from murine liver, 6 µg of DNA from murine liver, grey: digest blank, in which everything was performed like in a real digest, but no DNA was added.

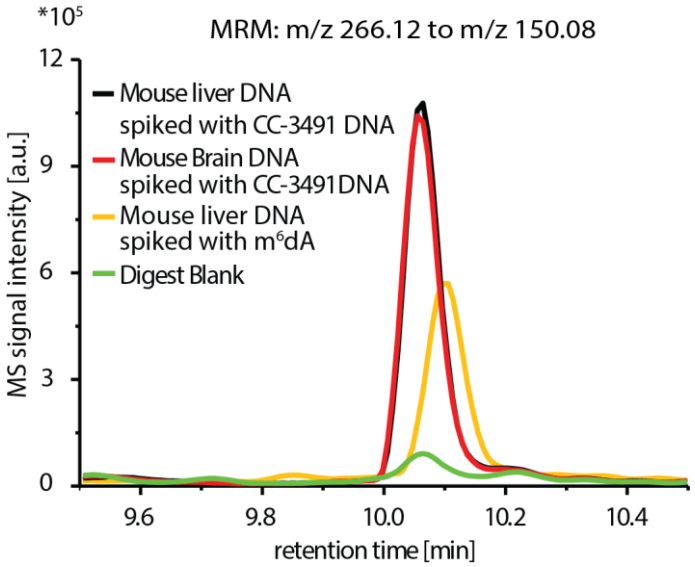


Figure S6: chromatogram of murine DNA from different tissues spiked with DNA from *Chlamydomonas reinhardtii* (CC-3491) and m^6dA nucleoside, respectively.; black: DNA from murine liver spiked with CC-3491; green: digest blank, in which everything was performed like in a real digest, but no DNA was added, red: DNA from murine brain with CC-3491, yellow: DNA from murine liver spiked with m^6dA nucleoside. The shown peak seems shifted in contrast to the other two peaks, but this is due to a measurement at a different time.

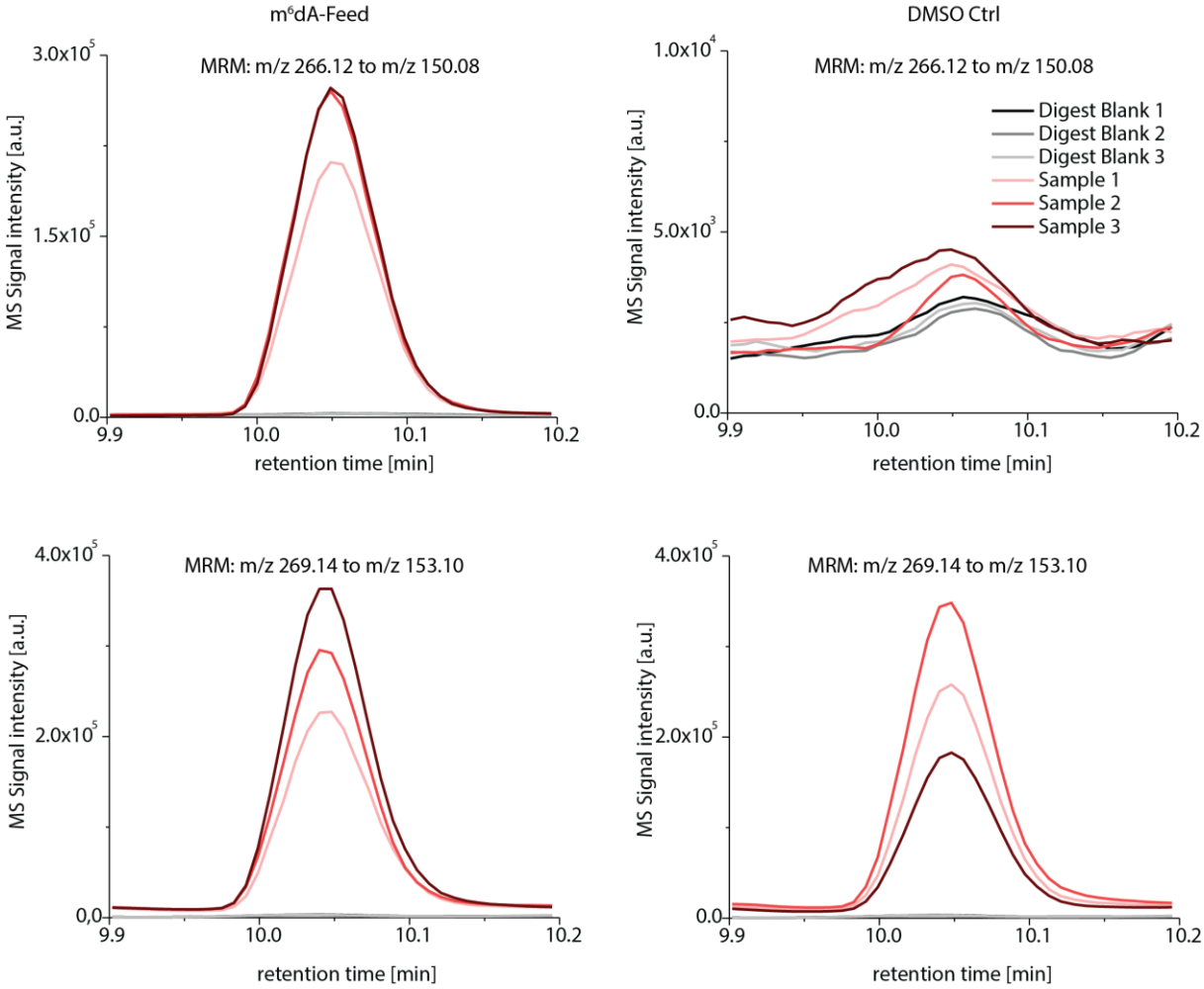


Figure S7: chromatograms for m^6dA-fed cells (left) and the respective DMSO control (right) with the transition for the natural nucleoside (m/z 266.12 to m/z 150.08) displayed in the upper row and the transition for the spiked heavy labeled nucleoside (m/z 269.14 to m/z 153.10) displayed in the lower row. The red lines indicate the chromatogram for the measured samples, the black line indicates the respective digest blank, in which everything was performed like in a real digest, but without addition of DNA or D_3-m^6dA).

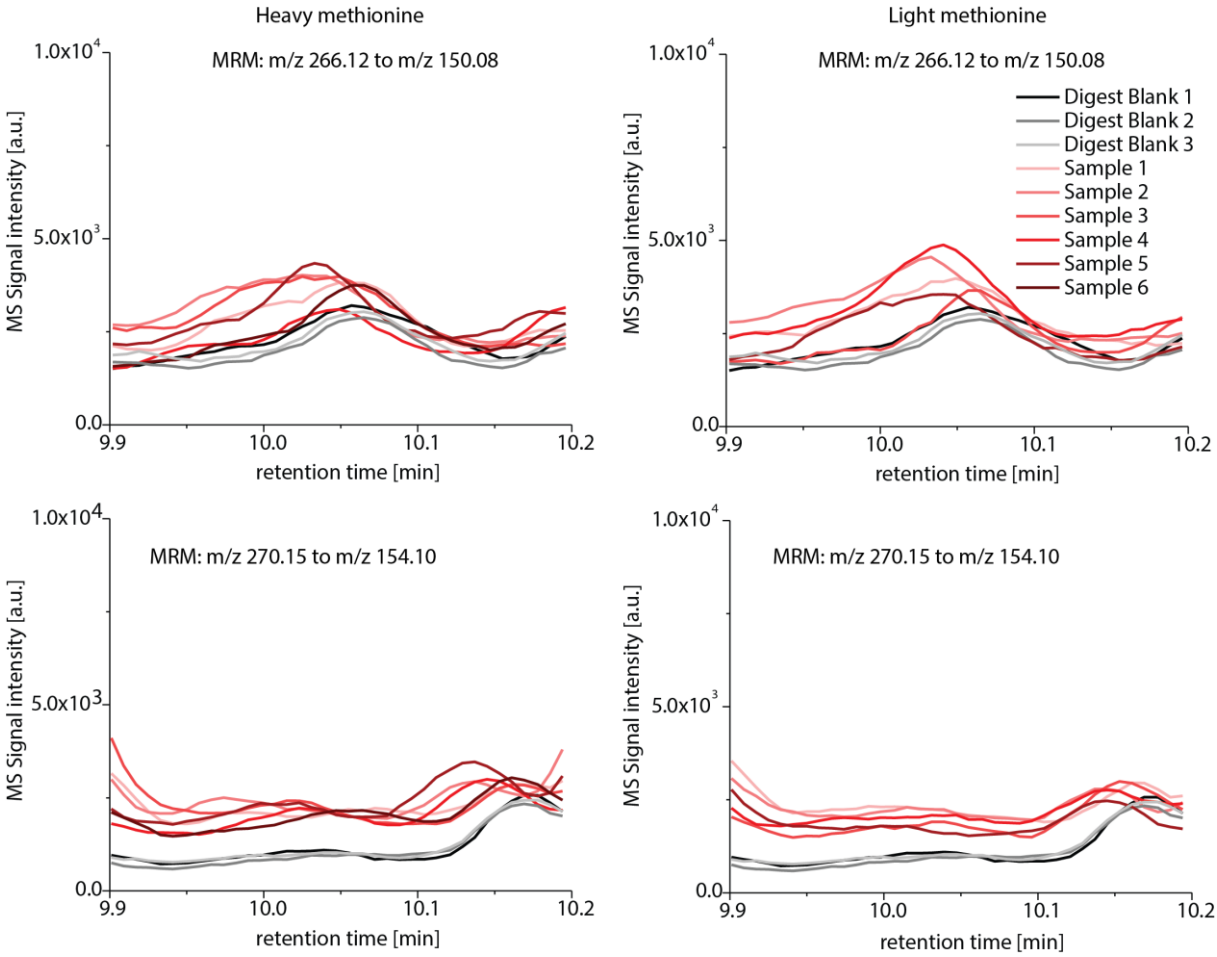


Figure S8: chromatograms for heavy methionine-fed cells (left) and the respective light methionine control (right) with the transition for the natural nucleoside (m/z 266.12 to m/z 150.08) displayed in the upper row and the transition for the heavy labeled nucleoside (m/z 270.15 to m/z 154.10; not formed) displayed in the lower row. The red lines indicate the chromatogram for the measured samples, the black line indicates the respective digest blank, in which everything was performed like in a real digest, but without addition of DNA.

[1] J. W. Jones, R. K. Robins, *J. Am. Chem. Soc.* **1963**, *85*, 193-201.
[2] J. D. Rosmarie Rippka, John B. Waterbury, Michael Herdman, Roger Y. Stanier, *J. Gen. Microbiol.* **1979**, *111*, 1-61.
[3] D. S. Gorman, R. P. Levine, *Proc. Natl. Acad. Sci. U S A* **1965**, *54*, 1665-1669.
[4] K. Taghizadeh, J. L. McFaline, B. Pang, M. Sullivan, M. Dong, E. Plummer, P. C. Dedon, *Nat. Protoc.* **2008**, *3*, 1287-1298.
[5] E. Li, T. H. Bestor, R. Jaenisch, *Cell* **1992**, *69*, 915-926.
[6] 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.
[7] 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.
[8] 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.

7.4 Zusatzmaterialien zu Abschnitt 3.4

T. Pfaffeneder#, F. Spada#, M. Wagner#, C. Brandmayr, S. Laube, D. Eisen, M. Truss, J. Steinbacher, B. Hackner, O. Kotljarova, D. Schuermann, S. Michalakis, **O. Kosmatchev**, S. Schiesser, B. Steigenberger, N. Raddaoui, U. Müller, H. Leonhardt, P. Schär, M. Müller, T. Carell, *Nat. Chem. Biol.* **2014,** *10, 574-581. Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA.*

Supplementary Information

Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA

Toni Pfaffeneder[1#], Fabio Spada[1#], Mirko Wagner[1#], Caterina Brandmayr[1], Silvia Laube[1], David Eisen[1], Matthias Truss[2], Jessica Steinbacher[1], Benjamin Hackner[1], Olga Kotljarova[1], David Schuermann[5], Stylianos Michalakis[3], Olesea Kosmatchev[1], Stefan Schiesser[1], Barbara Steigenberger[1], Nada Raddaoui[1], Gengo Kashiwazaki[1], Udo Müller[4], Cornelia G. Spruijt[6], Michiel Vermeulen[6≠], Heinrich Leonhardt[4], Primo Schär[5], Markus Müller[1]* and Thomas Carell[1]*

[1] Center for Integrated Protein Science at the Department of Chemistry, Ludwig-Maximilians-Universität München, Butenandtstr. 5-13, 81377 München, Germany.

[2] Charité Universitätsklinikum, Otto-Heubner-Centrum für Kinder und Jugendmedizin, Klinik für Allgemeine Pädiatrie, Labor für Pädiatrische Molekularbiologie, Ziegelstr. 5-9, 10098 Berlin, Germany.

[3] Center for Integrated Protein Science at the Department of Pharmacy – Center for Drug Research, Ludwig-Maximilians-Universität München, Butenandtstr. 5-13, 81377 München, Germany.

[4] Center for Integrated Protein Science at the Department of Biology, Ludwig-Maximilians-Universität München, Grosshaderner Str. 2, 82152 Planegg-Martinsried, Germany.

[5] Department of Biomedicine, University of Basel, Mattenstrasse 28, 4058 Basel, Switzerland.

[6] Department of Molecular Cancer Research, Cancer Genomics Netherlands, UMC Utrecht, 3584 CG Utrecht, Netherlands.

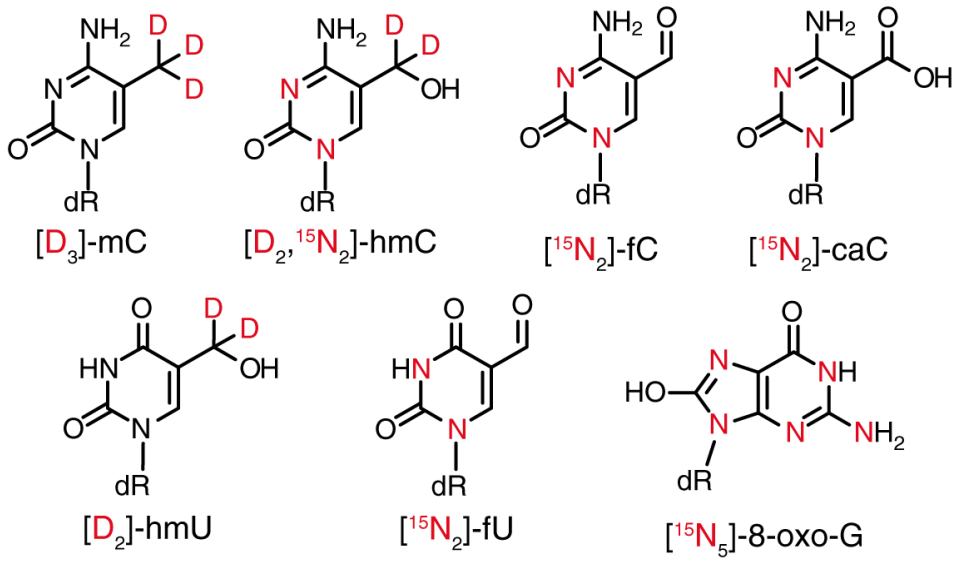
[≠] Present Address: Department of Molecular Biology, Faculty of Science, Radboud Institute for Molecular Life Sciences, Radboud University Nijmegen, 6525 GA, Nijmegen, The Netherlands

* Correspondence to: markus.mueller@cup.uni-muenchen.de and thomas.carell@cup.uni-muenchen.de

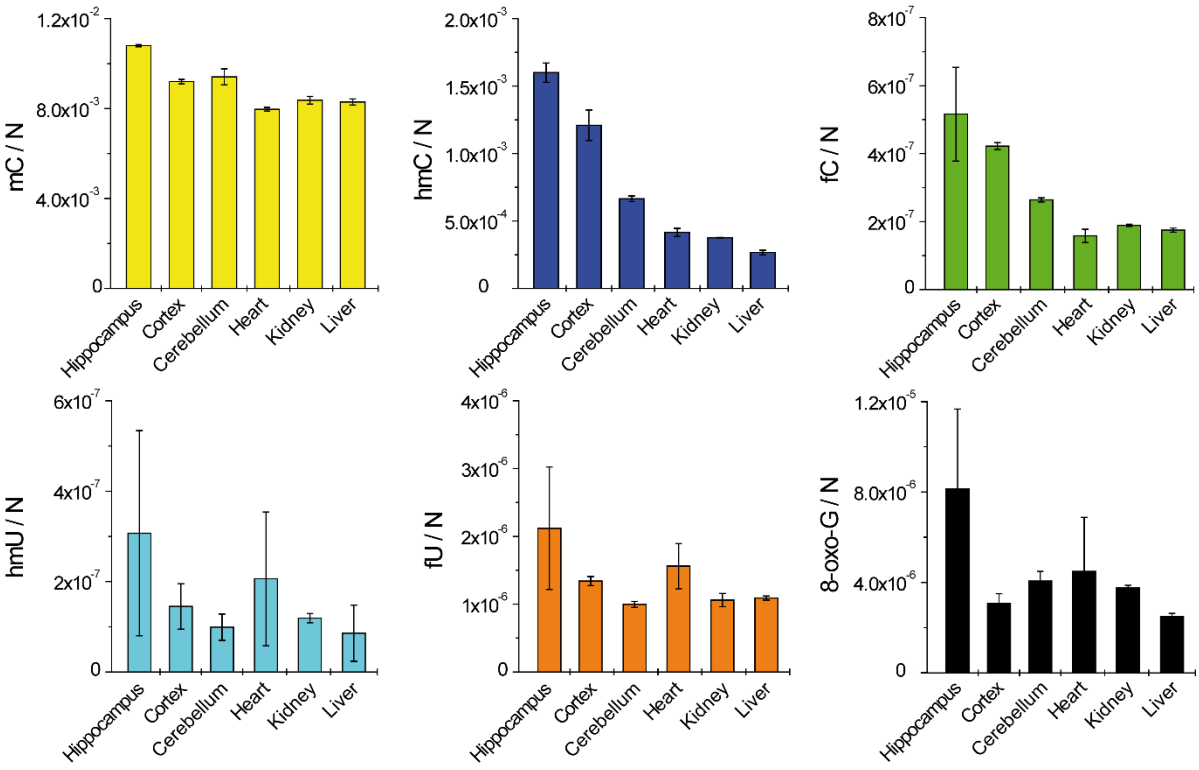
[#] These authors contributed equally

Nature Chemical Biology: doi:10.1038/nchembio.1532

Supplementary Results

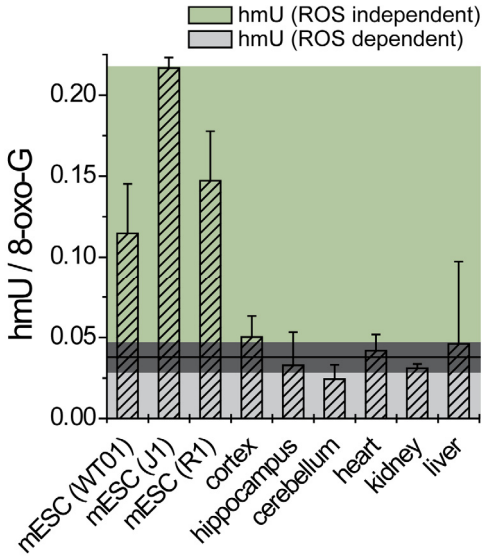


Supplementary Figure 1. Isotopically labeled nucleosides used as internal standards for quantitative LC-MS/MS analysis (dR = -2´-deoxyribose).



Supplementary Figure 2. DNA modification levels per nucleoside (N) of different murine tissues from 3 month old individuals (n = 3). Depicted are biological mean values ±SD.

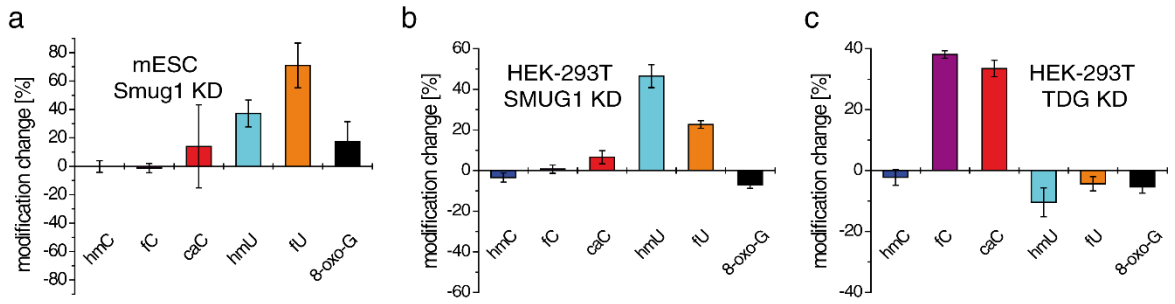
Nature Chemical Biology: doi:10.1038/nchembio.1532



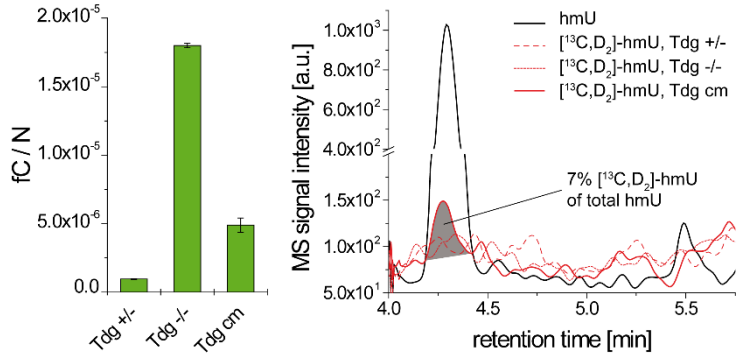
Supplementary Figure 3. hmU-levels normalized to the oxidative background marker 8-oxo-G in mESCs (n_{WT01} = 7; n_{J1} = 2, n_{R1} = 2) and murine tissues (n = 3, 3 month old individuals) in order to dissect ROS dependent and ROS independent processes. Normalization was necessary to take deviating background oxidation of DNA sample preparation into account. The light grey area reflect the hmU-level fractions, which are generated by ROS dependent processes. The green area reflect the hmU-level fractions, which are generated by ROS independent processes. The assignment is based on the assumption, that hmU-levels in somatic tissue are exclusively ROS created lesions (derived from the cluster analysis in **Fig. 2b**). The dark grey area reflect the mean value ±SD of hmU/8-oxo-G ratios of the tissue data. The difference between the height of the hmU/8-oxo-G ratios of mESCs and the mean of the tissue data give the hmU-fraction which is formed by ROS-independent processes. In WT01, J1 and R1 cells about 67%, 83% and 74%, respectively, of the global hmU-levels are estimated to be created by ROS independent processes. Bars reflect biological mean values ±SD. The differences between mESCs (WT01, J1) and murine tissues are significant ($P = 9.3\text{x}10^{-5}$–$5.0\text{x}10^{-3}$; unpaired two-tailed t-test) except for mESCs (WT01, J1) and liver (P = 0.081, 0.15).

Supplementary Figure 4. Schematic representation of isotope tracing experiments with [^{13}C,^{15}N$_2$]-T (left; blue) and [^{13}C,D$_3$]-methionine (right; red) and exchange rates of derived genomic isotopologues. Small negligible deviations in the exchange rates are due to differential noise sources. LOD = limit of detection. wt = wild type mESCs; Tdg cm = Tdg / mESCs complemented with catalytic inactive Tdg (see Supplementary Fig. 6).

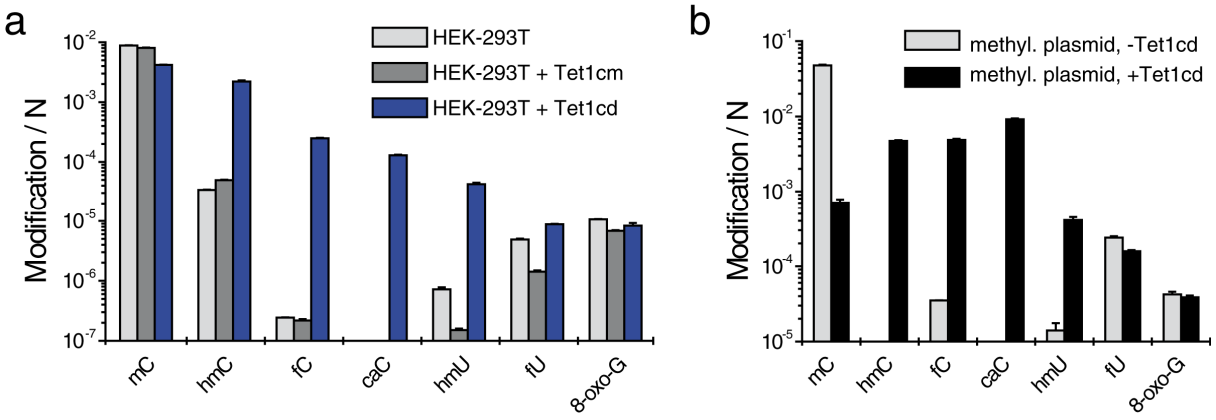
Nature Chemical Biology: doi:10.1038/nchembio.1532



Supplementary Figure 5. (**a**) Effect of Smug1 depletion on modification levels in mESCs (R1). Effect of SMUG1 (**b**) and TDG (**c**) depletion on modification levels in HEK-293T cells overexpressing Tet1cd. Shown is the percent change in modification content per nucleoside in cells co-transfected with esiRNAs targeting TDG or SMUG1 relative to co-transfection with control esiRNA. Depicted are technical mean values ± SD.

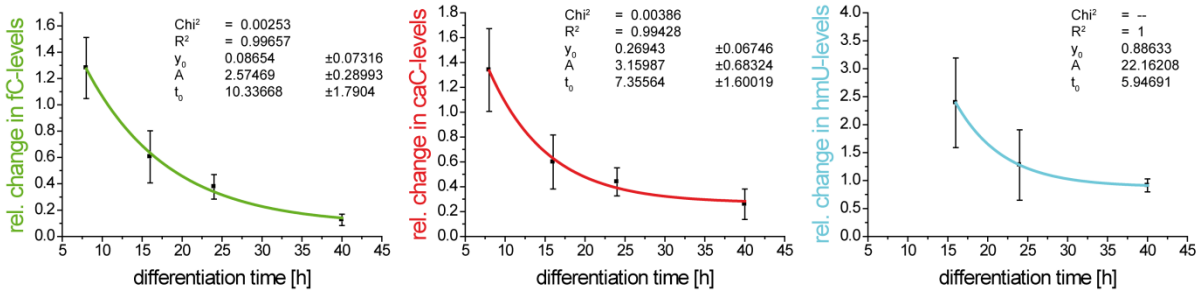


Supplementary Figure 6. Isotope tracing experiments with *Tdg*$^{+/-}$, *Tdg*$^{-/-}$ mESCs as well as *Tdg*$^{-/-}$ mESCs complemented with a catalytic mutant of Tdg (Tdg cm) grown in the presence of [^{13}C,D$_3$]-methionine (200 µM). The catalytic mutant of Tdg is not completely inactive (fC levels are between Tdg +/- and Tdg -/- cells, left). Only in case of *Tdg*-/- cells complemented with a catalytic inactive Tdg (Tdg cm) labeled hmU was detected, which originated from the deamination of labeled hmC. ~7% [^{13}C,D$_2$]-hmU over total hmU was observed. This corresponds to ~0.06% deamination of hmC to hmU under these conditions (2.5×10^{-4} total hmC / N; 2.2×10^{-6} total hmU / N). Labeled fU was not observed. Depicted fC-levels represent technical mean values ±SD.

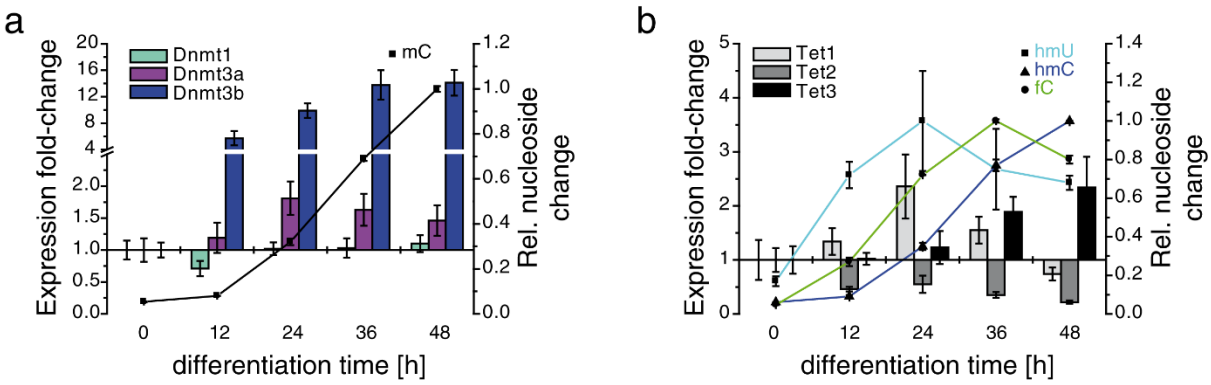


Supplementary Figure 7. Tet1 and Tet2 generate hmU in HEK-293T and in vitro (a) Effect of Tet1 overexpression on modified pyrimidines in HEK-293T cells. Modification levels in cells overexpressing wt and catalytic mutant versions of Tet1 catalytic domain (Tet1cd, blue bars and Tet1cm, gray bars, respectively), or a control construct (white bars). Depicted are mean values ±SD of technical triplicates on a logarithmic scale. **(b)** Pyrimidine modification levels in methylated plasmid DNA after treatment in vitro with Tet1cd. Depicted are mean values ±SD of technical duplicates. Note the logarithmic scale.

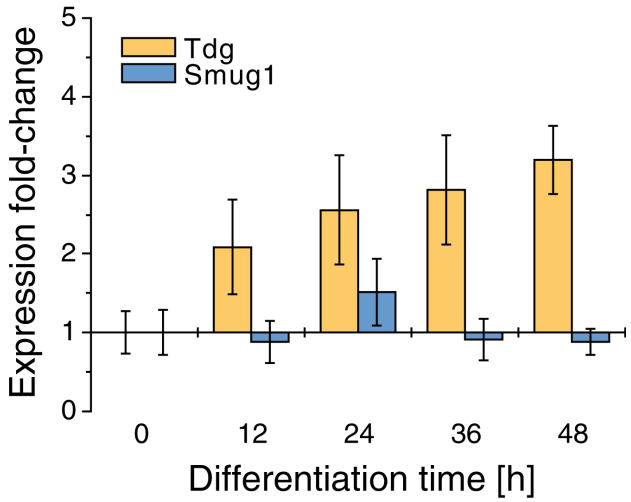
Nature Chemical Biology: doi:10.1038/nchembio.1532



Supplementary Figure 8. Exponential models for fitting the decay curve of fC, caC and hmU in combined data sets from differentiation of R1 and C57Bl6/129-derived mESCs (6 biological independent experiments). In a simplified approach a single exponential decay model ($y = y_0 + A*\exp(-x/t_0)$) was plotted using ORIGIN®. The parameters y_0 (offset), t_0 (time constant) and A (amplitude) of each decay function were iteratively optimized until the minimum of the Chi^2 value of the fitting was reached. Half-life times ($t_{1/2} = t_0*\ln 2$) for fC, caC and hmU were 7.2±1.2, 5.1±1.1 and 4.1 h, respectively.

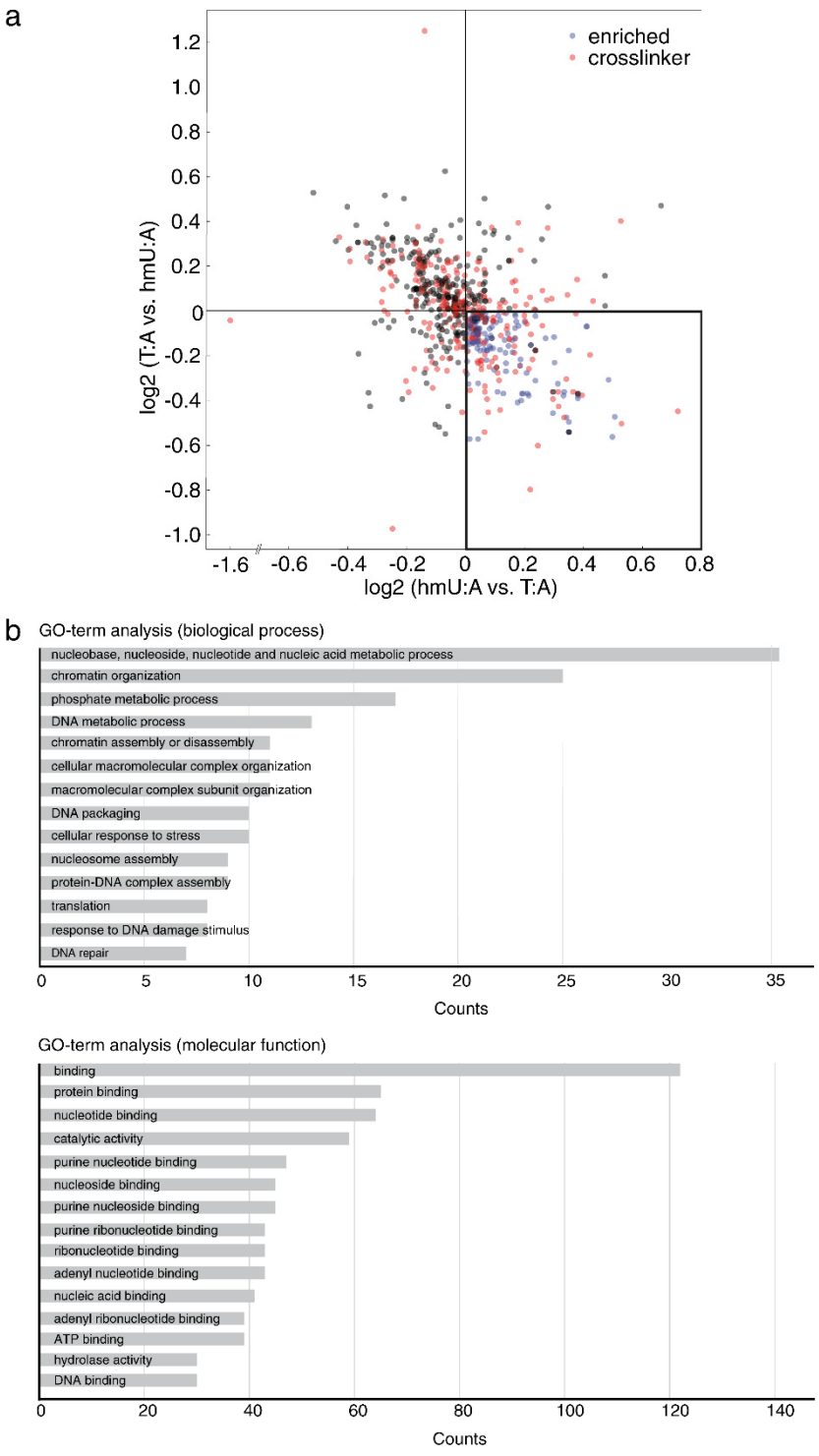


Supplementary Figure 9. Normalized transcript levels of Dnmts (c), Tet1–3 (d) and normalized modification levels of mC (c), hmC, fC and hmU (d) during differentiation of naïve mESCs in the presence of FGF-2 and ActA.



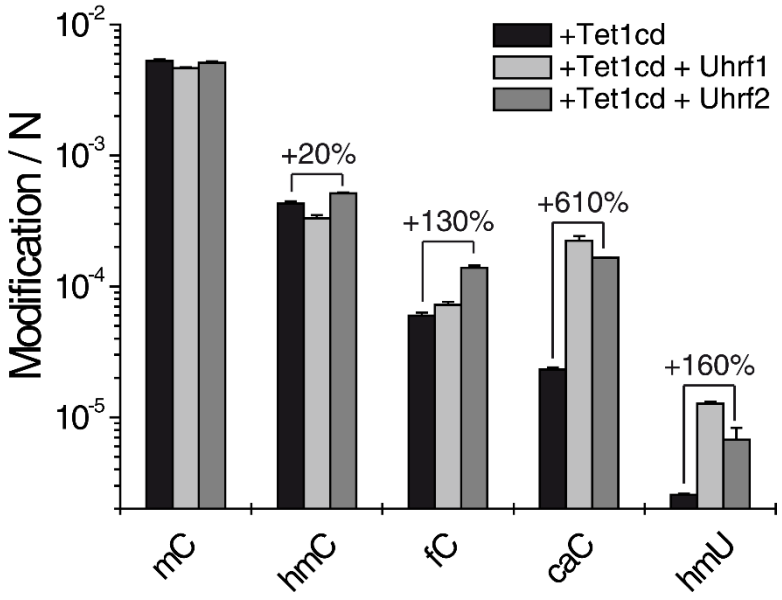
Supplementary Figure 10. Expression level analysis of Tdg and Smug1 during differentiation of mESCs in the presence of FGF-2 and ActA. Expression levels were quantified with respect to the housekeeping gene Gapdh and normalized to time point 0 h. Depicted are technical mean values ± SD.

Nature Chemical Biology: doi:10.1038/nchembio.1532



Supplementary Figure 11. (a) Scatterplot of proteins enriched with the hmU:A containing oligomer. Ratios of a forward and a reverse experiment are plotted. Specific readers in the forward and reverse experiment are marked in blue. Direct-specific readers are identified by the presence of the DNA-protein cross linker and marked in red. Gray dots are considered unspecific binders. See **Fig. 6** for detailed view. **(b)** Gene Ontology Analysis performed with DAVID Bioinformatics Resources 6.7[4]

Nature Chemical Biology: doi:10.1038/nchembio.1532



Supplementary Figure 12. Effect on modified pyrimidines in HEK-293T cells upon Tet1cd and Uhrf1 (light gray bars) or Uhrf2 (gray bars) co-overexpression. Depicted are mean values ±SD of technical triplicates on a logarithmic scale.

Nature Chemical Biology: doi:10.1038/nchembio.1532

Supplementary Table 1. Isotope tracing experiments by supplementing the growth medium of mES cells (LIF), differentiating mESCs (R1, without growth factors) and HEK-293T cells with either [^{13}C,$^{15}N_2$]-T (50 or 100 µM) or [^{13}C,D_3]-methionine (0.2 mM). Small deviations in the exchange yields are due to differential noise sources and are negligible. LOD = Limit of detection. In case of [^{13}C,D_1]-fU no difference was observed compared to the natural control.

cell type	growth medium	[^{13}C,$^{15}N_2$]-T / T [%]	[^{13}C,$^{15}N_2$]-hmU / hmU [%]	[^{13}C,$^{15}N_2$]-fU / fU [%]
mESC (2i)	100 µM natural T	0.1	< LOD	< LOD
mESC (2i)	100 µM [^{13}C,$^{15}N_2$]-T	76.0	78.2	74.6
HEK + Tet1cd (72h)	50 µM [^{13}C,$^{15}N_2$]-T	73.8	74.2	71.0

cell type	growth medium	[^{13}C,D_3]-mC / mC [%]	[^{13}C,D_2]-hmC / hmC [%]	[^{13}C,D_2]-hmU / hmU [%]	[^{13}C,D_1]-fU / fU [%]
mESC (LIF)	natural methionine	0.1	< LOD	< LOD	3.3
mESC (LIF; 0 h)	[^{13}C,D_3]-methionine	88.9	87.6	< LOD	3.0
diff. mESC (12 h)	[^{13}C,D_3]-methionine	89.3	88.4	< LOD	3.1
diff. mESC (24 h)	[^{13}C,D_3]-methionine	90.1	89.3	< LOD	3.2
diff. mESC (48 h)	[^{13}C,D_3]-methionine	90.4	90.5	< LOD	3.5
mESC *Tdg*$^{+/-}$	[^{13}C,D_3]-methionine	88.0	87.6	< LOD	3.7
mESC *Tdg*$^{-/-}$	[^{13}C,D_3]-methionine	87.4	87.2	< LOD	< LOD
mESC *Tdg*$^{-/-}$ + Tdg cm	[^{13}C,D_3]-methionine	86.9	86.7	7.4	2.2
HEK + Tet1cd (72h)	[^{13}C,D_3]-methionine	87.4	83.4	< LOD	< LOD

Nature Chemical Biology: doi:10.1038/nchembio.1532

Supplementary Table 2. Assessment of ROS dependent hmU and fU product ratio of T-oxidation in HEK-293T wild type cells where TET activity is lowest (related to Fig. 3c). Modified nucleosides / N are given as mean values plus SD of three independent technical replicates. When T is oxidized by ROS about 9.8% hmU and 90.2% fU is generated.

n	mC / N		hmC / N		fC / N		caC / N
	techn. mean	SD	techn. mean	SD	techn. mean	SD	techn. mean
1	6.14E-03	1.97E-04	2.92E-05	9.90E-08	2.56E-07	2.59E-09	n.d.
2	6.21E-03	1.32E-05	2.95E-05	2.21E-09	3.10E-07	2.06E-08	n.d.
3	5.76E-03	4.36E-05	3.32E-05	6.55E-07	2.86E-07	2.95E-09	n.d.
4	9.01E-03	2.87E-04	5.23E-05	5.01E-07	3.50E-07	5.00E-09	n.d.
5	8.80E-03	9.62E-05	3.39E-05	5.44E-07	2.43E-07	2.41E-09	n.d.
6	8.55E-03	7.69E-05	3.70E-05	2.17E-07	2.14E-07	1.30E-08	n.d.
biol. mean	7.41E-03		3.59E-05		2.76E-07		
biol. SD	1.52E-03		8.56E-06		4.90E-08		

n	hmU / N		fU / N		hmU/ (hmU+fU)	fU/ (hmU+fU)	8-oxo-G / N	
	techn. mean	SD	techn. mean	SD	[%]	[%]	techn. mean	SD
1	2.66E-07	5.54E-08	5.15E-06	1.59E-07	4.9	95.1	8.49E-06	1.30E-07
2	1.21E-06	1.62E-07	8.02E-06	6.66E-07	13.1	86.9	1.00E-05	1.98E-07
3	3.65E-08	7.35E-09	9.76E-07	8.75E-09	3.6	96.4	3.02E-06	7.02E-08
4	6.31E-07	9.10E-10	3.51E-06	3.28E-08	15.3	84.7	7.43E-06	1.34E-07
5	7.21E-07	5.93E-08	4.89E-06	1.59E-07	12.8	87.2	1.06E-05	1.14E-07
6	3.58E-07	4.95E-08	3.56E-06	9.64E-09	9.1	90.9	7.22E-06	1.15E-08
biol. mean	5.37E-07		4.35E-06		9.8	90.2	7.80E-06	
biol. SD	4.13E-07		2.33E-06		4.8	4.8	2.70E-06	

Nature Chemical Biology: doi:10.1038/nchembio.1532

Supplementary Note 1: oligonucleotide sequences for protein pull-down assays

Supplementary Table 3. DNA oligonucleotides used in protein pull-down studies.

ODN	Sequence (5´→ 3´)	Modifications
1	Biotin-GCA-TCC-GGT-CAY-CGT-TCC-TTC-GGA	Y = 5-octadienyl-U
2	Biotin-GCA-TCC-GGT-CAY-CAT-TCC-TTC-GGA	Y = 5-octadienyl-U
3	TCC-GAA-GGA-AXG-ATG-ACC-GGA-TGC	X= T
4		X= hmU
5		X= C
6		X= hmC
7	Biotin-GCT-CAC-GCT-AGY-CGA-CTC-CGT-GCA	Y = 5-octadienyl-U
8	TGC-ACG-GAG-TXG-ACT-AGC-GTG-AGC	X = T
9		Y = hmU

Hybridization scheme:

Pull-down 1: hmU:A vs. T:A : ODN4/2 vs. ODN3/2
Pull-down 2: hmU:G vs. C:G = ODN4/1 vs. ODN5/1
Pull-down 3: hmC:G vs. C:G = ODN6/1 vs. ODN5/1
Pull-down 4 (scrambled sequence): hmU:A vs. T:A = ODN9/7 vs. ODN8/7

Supplementary Note 2: LC-UV-ESI-MS/MS analysis of DNA

Supplementary Table 4. Compound-dependent LC-MS/MS-parameters used for the analysis of genomic DNA. CE: collision energy; CAV: collision cell accelerator voltage; EMV: electron multiplier voltage. The nucleosides were analyzed in the positive ($[M+H]^+$ species) as well as in the negative ($[M-H]^-$ species) 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 1.5–4.0 min								
$[^{15}N_2]$-caC	274.08	Wide	158.03	Wide	170	5	5	Positive
caC	272.09	Wide	156.04	Wide	170	5	5	Positive
$[^{15}N_2,D_2]$-hmC	262.12	enhanced	146.07	enhanced	40	27	1	Positive
hmC	258.11	enhanced	142.06	enhanced	40	27	1	Positive
$[D_3]$-mC	245.13	enhanced	129.09	enhanced	30	60	1	Positive
mC	242.11	enhanced	126.07	enhanced	30	60	1	Positive
C	228.1	enhanced	112.05	enhanced	1	1	0	Positive
time segment 4.0–6.0 min								
$[D_2]$-hmU	259.09	Wide	216.08	Wide	48	7	5	Negative
$[D_2]$-hmU	259.09	Wide	126.05	Wide	48	7	5	Negative
hmU	257.08	Wide	214.07	Wide	48	7	5	Negative
hmU	257.08	Wide	124.04	Wide	48	7	5	Negative
$[^{15}N_2]$-fU	257.06	Wide	213.05	Wide	48	6	5	Negative
fU	255.06	Wide	212.06	Wide	48	6	5	Negative
time segment 6.0–9.0 min								
$[^{15}N_5]$-8-oxo-G	289.08	Wide	173.04	Wide	120	9	7	Positive
8-oxo-G	284.1	Wide	168.05	Wide	120	9	7	Positive
$[^{15}N_2]$-fC	258.09	Wide	142.04	Wide	120	5	5	Positive
fC	256.09	Wide	140.05	Wide	120	5	5	Positive

Nature Chemical Biology: doi:10.1038/nchembio.1532

Supplementary Table 5. Compound-dependent LC-MS/MS-parameters used for the analysis of genomic DNA obtained from cells which were grown in medium supplemented with labeled thymidine ([^{13}C,^{15}N$_2$]-T). CE: collision energy; CAV: collision cell accelerator voltage; EMV: electron multiplier voltage. The nucleosides were analyzed in the positive ([M+H]$^+$ species) as well as in the negative ([M-H]$^-$ species) 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 1.5–4.0 min								
[^{13}C,^{15}N$_2$]-caC	275.09	wide	159.04	wide	65	5	5	Positive
caC	272.09	wide	156.04	wide	65	5	5	Positive
[^{13}C,^{15}N$_2$]-hmC	261.11	enhanced	145.06	enhanced	40	27	1	Positive
hmC	258.11	enhanced	142.06	enhanced	40	27	1	Positive
[^{13}C,^{15}N$_2$]-mC	245.13	enhanced	129.09	enhanced	30	60	1	Positive
mC	242.11	enhanced	126.07	enhanced	30	60	1	Positive
[^{13}C,^{15}N$_2$]-C	231.1	enhanced	115.05	enhanced	40	1	3	Positive
C	228.1	enhanced	112.1	enhanced	40	1	3	Positive
time segment 4.0–6.0 min								
[^{13}C,^{15}N$_2$]-hmU	260.08	wide	215.07	wide	50	7	5	Negative
hmU	257.08	wide	214.07	wide	50	7	5	Negative
[^{13}C,^{15}N$_2$]-fU	258.06	wide	213.05	wide	50	6	5	Negative
fU	255.06	wide	212.06	wide	50	6	5	Negative
time segment 6.0–9.0 min								
[^{15}N$_5$]-8-oxo-G	289.08	wide	173.04	wide	80	9	7	Positive
8-oxo-G	284.1	wide	168.05	wide	80	9	7	Positive
[^{13}C,^{15}N$_2$]-fC	259.09	wide	143.04	wide	80	5	5	Positive
fC	256.09	wide	140.05	wide	80	5	5	Positive
[^{13}C,^{15}N$_2$]-T	246.1	enhanced	130.05	enhanced	30	40	3	Positive
T	243.1	enhanced	127.05	enhanced	30	40	3	Positive

Supplementary Table 6. Compound-dependent LC-MS/MS-parameters used for the analysis of genomic DNA obtained from cells which were grown in medium supplemented with labeled (*methyl*-^{13}C,D$_3$)-methionine.CE: collision energy; CAV: collision cell accelerator voltage; EMV: electron multiplier voltage. The nucleosides were analyzed in the positive ([M+H]$^+$ species) as well as in the negative ([M-H]$^-$ species) 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 1.5–4.0 min								
[^{13}C]-caC	273.09	wide	157.04	wide	65	5	5	Positive
caC	272.09	wide	156.04	wide	65	5	5	Positive
[^{13}C,D$_2$]-hmC	261.12	enhanced	145.08	enhanced	40	27	1	Positive
hmC	258.11	enhanced	142.06	enhanced	40	27	1	Positive
[^{13}C,D$_3$]-mC	246.14	enhanced	130.09	enhanced	30	60	1	Positive
mC	242.11	enhanced	126.07	enhanced	30	60	1	Positive
C-dN	228.1	enhanced	112.1	enhanced	40	1	3	Positive
time segment 4.0–6.0 min								
[^{13}C,D$_2$]-hmU	260.09	wide	217.09	wide	60	7	5	Negative
hmU	257.08	wide	214.07	wide	60	7	5	Negative
[^{13}C,D]-fU	257.07	wide	214.07	wide	60	6	5	Negative
fU	255.06	wide	212.06	wide	60	6	5	Negative
time segment 6.0–9.0 min								
[^{15}N$_5$]-8-oxo-G	289.08	wide	173.04	wide	80	9	7	Positive
8-oxo-G	284.1	wide	168.05	wide	80	9	7	Positive
[^{13}C,D]-fC	258.1	wide	142.06	wide	80	5	5	Positive
fC	256.09	wide	140.05	wide	80	5	5	Positive

Nature Chemical Biology: doi:10.1038/nchembio.1532

Spiking amounts of labeled internal standards for quantitative LC-MS/MS analysis

The quantification of nucleosides of genomic DNA isolated from mESC or mouse tissue was carried out with the following amounts of internal standards: 51.03 pmol [D_3]-mC, 7.655 pmol [$^{15}N_2$,D_2]-hmC, 45.6 fmol [$^{15}N_2$]-fC, 43.0 fmol [$^{15}N_2$]-caC, 108.9 fmol [$^{15}N_5$]-8-oxo-G; 160.1 fmol [D_2]-hmU and 180.0 fmol [$^{15}N_2$]-fU. The quantification of nucleosides of genomic DNA isolated from HEK293 cells overexpressing Tet was carried out with the following amounts of internal standards: 34.02 pmol [D_3]-mC, 5.103 pmol [$^{15}N_2$,D_2]-hmC, 303.8 fmol [$^{15}N_2$]-fC, 215.1 fmol [$^{15}N_2$]-caC, 108.9 fmol [$^{15}N_5$]-8-oxo-G; 160.1 fmol [D_2]-hmU and 180.0 fmol [$^{15}N_2$]-fU. Genomic DNA samples isolated from cells grown in media supplemented with either [^{13}C,$^{15}N_2$]-T or [^{13}C,D_3]-methionine were not spiked with internal standards except [$^{15}N_5$]-8-oxo-G.

Validation of the LC-UV-MS/MS quantification method:

Method validation, in particular linearity, precision, and accuracy (i.e. determined from matrix samples spiked with isotopically labeled internal standards) of the established method were investigated. Validation for the established LC-UV-ESI-MS/MS quantification method was based on three different series (i.e., calibration functions and quality control samples) accomplished on different days. Each calibration standard (5-8 standard concentrations) was analyzed five times. Each validation experiment was complemented by matrix blanks (analyzed in triplicates) to ensure selectivity and specificity of the method. Linear regression was applied by Origin® 6.0 (*Microcal*™) to obtain calibration curves. Therefore, the ratio of the area under the curve (A/A*) of the unlabeled nucleoside to the internal standard (*) was plotted against the ratio of the amount of substance (n/n*) of the unlabeled nucleoside to the internal standard (*) (see **Supplementary Fig. 13**). Calibration functions were calculated without weighting. Additionally, acceptable accuracy (80–120%) as well as precision (<20% RSD) was required. Accuracy was proven by computing the amount of substance n from the obtained A/A* ratios of the calibration standards using the respective calibration function. Here, accuracy was defined as the ratio of the used amount of substance to the calculated amount of substance in percent and had to be between 80–120% for each standard concentration. Precision was defined as follows: technical replicates of A/A* ratios for each calibration standard had to have relative standard deviations (RSD) smaller than 20%. The lower limit of quantification (LLOQ) was defined as the lowest concentration fulfilling the requirements of accuracy and precision and achieving a response of at least three times the response compared with the blank response. A compilation of absolute and relative LLOQs is shown in **Supplementary Table 7**.

Quality control samples to evaluate intra-batch precision (see below) were investigated using a biological sample spiked with internal standards. Long-term stability of aqueous solutions of the labeled and unlabeled nucleosides at a storage temperature of −20 °C was investigated over two months including several freeze and thaw cycles by analyzing the MS/MS-responses with each batch.

Nature Chemical Biology: doi:10.1038/nchembio.1532

Supplementary Table 7. Compilation of absolute lower limits of quantification [fmol] (LLOQ; see Supplementary Fig. 13) and relative LLOQs [per N] depending on the amount of DNA, which is digested. The relative LLOQs were computed by generating ratios of the absolute LLOQ [pmol] to the total amount of nucleosides (N; [pmol]) in the respective amount of DNA [µg]. The total amount of nucleosides was obtained by using the average molar mass of 308.91 g mol^{-1} for the monomeric DNA entity by taking the GC-content (21% C or G) in mouse into account.

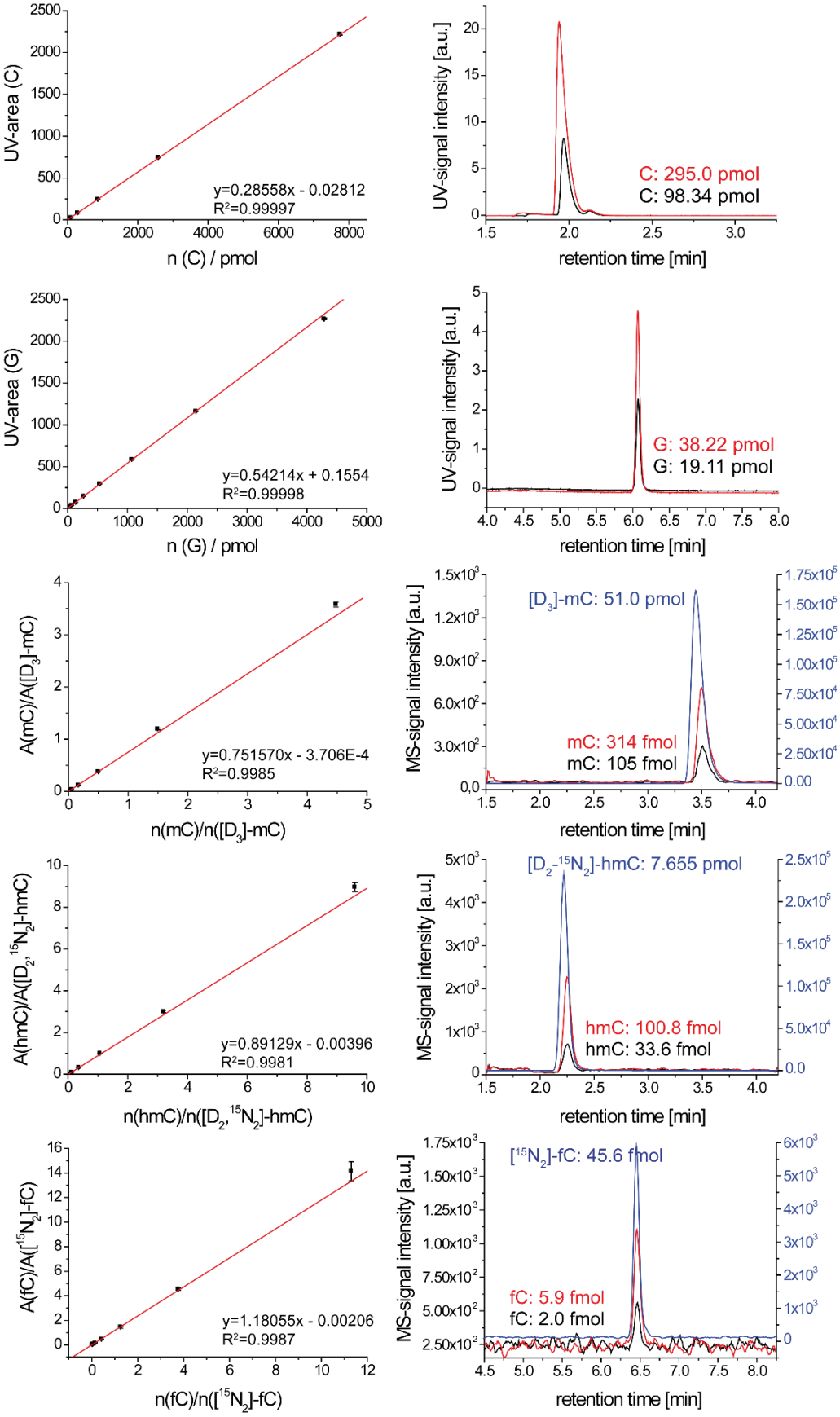
	absolute LLOQ [fmol]	**relative LLOQ [per N]**	**relative LLOQ [per N]**	**relative LLOQ [per N]**
DNA amount		**5 µg**	**10 µg**	**25 µg**
mC	104.5	6.5E-06	3.2E-06	1.3E-06
hmC	100.8	6.2E-06	3.1E-06	1.2E-06
fC	2.1	1.3E-07	6.5E-08	2.6E-08
caC	2.0	1.2E-07	6.2E-08	2.5E-08
8-oxo-G	2.0	1.2E-07	6.2E-08	2.5E-08
dU	14.1	8.7E-07	4.4E-07	1.7E-07
hmU	6.4	4.0E-07	2.0E-07	7.9E-08
fU	5.7	3.5E-07	1.8E-07	7.0E-08

Intra-batch assay and quantification data processing:

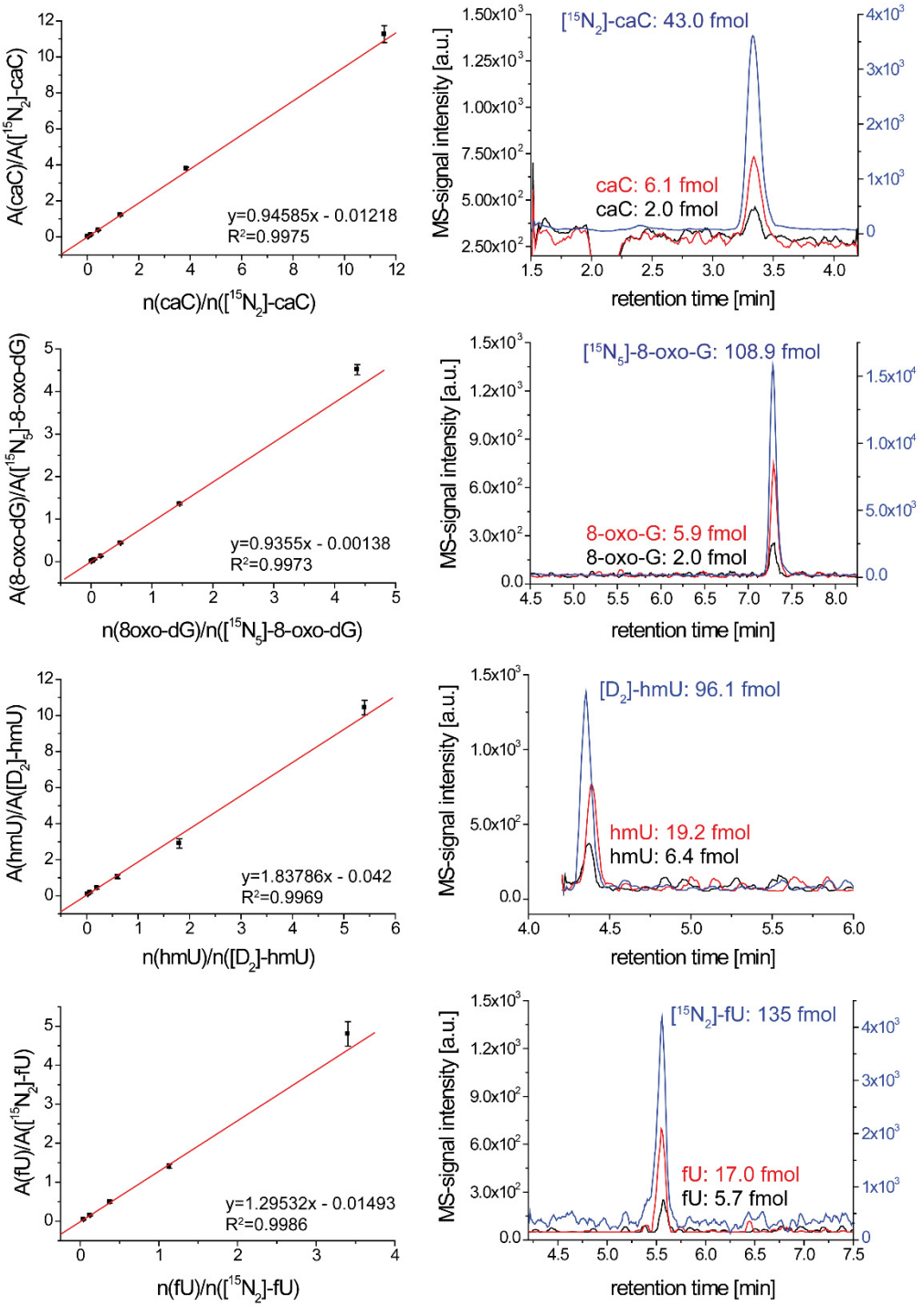
In order to evaluate intra-batch precision (see below) quality samples were investigated using a biological sample spiked with internal standards. The intra-batch-assay was performed for the LC-ESI-MS/MS analysis of the nucleosides G, C, mC, hmC, fC, caC, hmU, fU and 8-oxo-G. For this, a representative mESC DNA sample was analyzed. Technical replicates (n=5; each 4 µg DNA) were independently prepared using the below described digestion protocol. For data processing MassHunter Quantitative Analysis from *Agilent* was used. The area under the curve (A) was determined by LC-MS/MS for mC, hmC, fC, caC, hmU, fU, 8-oxo-G and for the corresponding labeled internal standards (A*); the area under the curve (A_{UV}) for G and C was determined by LC-UV. The amount of substance (n; pmol) of each nucleoside was computed by using the calibration curves (see **Supplementary Fig. 13**). The total sample volume was 40 µL, the injection volume after sample filtration was 29 µL. Therefore, the obtained values of G and C by LC-UV quantification were corrected by the factor given by the ratio of 40 µL/29 µL. Careful monitoring of the exact pipetting and injection volumes was therefore necessary. The obtained absolute amounts (pmol) of the DNA modifications (X= C, mC, fC, caC, hmU, fU, 8-oxo-G) were then related to the amount of G (pmol) giving ratios of X / G in %. The sum of X / G was defined as 100%. These values were then transferred in X / N values, considering that the G content is 21% in mouse.

The determined A/A* ratios of the DNA nucleosides to the labeled internal standards (see **Supplementary Table 8**) showed a high precision (RSD = 3.9-18%) for each nucleoside. The necessity in using labeled internal standards for quantification is shown by comparing these results with the relative standard deviation (RSD = 7.6-49.5%) of the uncorrected mass signal (A) of the respective DNA modification. Moreover, in order to gain precision between sample batches measured on different days (data not shown), it is even more important to use internal standards. No memory effect was observed during blank experiments performed after several measurements of a sample. The blank analyses were not contaminated by carry-over.

Nature Chemical Biology: doi:10.1038/nchembio.1532



Nature Chemical Biology: doi:10.1038/nchembio.1532



Supplementary Figure 13. UV and LC-MS/MS calibration curves and representative chromatograms of C (UV), G (UV), mC/[D_3]-mC, hmC/[$^{15}N_2$,D_2]-hmC, fC/[$^{15}N_2$]-fC, caC/[$^{15}N_2$]-caC, 8-oxo-G/[$^{15}N_5$]8-oxo-G, U/[$^{15}N_2$]-U, hmU/[D_2]-hmU, fU/[$^{15}N_2$]-fU. These were obtained by applying the compound-dependent parameters summarized in Supplementary Table 4. For hmU the MS/MS transition 257→214 was used. Depicted are the means of five technical replicates of one sample batch. Error bars reflect SD. Linearity was given across the following compound amounts in 29 µL injection volume: 10.63–7751 pmol C; 19.11–4892 pmol G; 104.5 fmol–228.6 pmol mC; 100.8 fmol–73.45 pmol hmC; 2.1–515.0 fmol fC; 2.0–496.6 fmol caC; 2.0–475.7 fmol 8-oxo-G; 6.4–519.3 fmol hmU; 5.7–459.6 fmol fU. The amounts of the labeled internal standards in 29 µL injection volume were as follows: 51.03 pmol [D_3]-mC; 7.655 pmol [$^{15}N_2$,D_2]-hmC; 45.6 fmol [$^{15}N_2$]-fC; 43.0 fmol [$^{15}N_2$]-caC; 108.9 fmol [$^{15}N_5$]-8-oxo-G; 96.1 fmol [D_2]-hmU; 135.0 fmol [$^{15}N_2$]-fU.

Nature Chemical Biology: doi:10.1038/nchembio.1532

Supplementary Table 8. Intra-batch-assay and quantification data processing.

	A_{UV}(G)	n(G) [pmol]	A_{UV}(C)	n(C) [pmol]	C / G [%]	C / N
techn. replicate 1	956	2431	485	2341	96.3	2.02E-01
techn. replicate 2	980	2493	495	2391	95.9	2.01E-01
techn. replicate 3	1043	2654	531	2566	96.7	2.03E-01
techn. replicate 4	979	2492	498	2406	96.5	2.03E-01
techn. replicate 5	982	2498	498	2406	96.3	2.02E-01
techn. mean value	988	2513	501	2422	96.4	2.02E-01
SD	33	83	18	85	0.3	6.29E-04
RSD [%]	3.3	3.3	3.5	3.5	0.3	0.3
	A(mC)	A([D_3]-mC)	A(mC)/ A([D_3]-mC	n(mC) [pmol]	mC / G [%]	mC / N
techn. replicate 1	636358	510501	1.247	91.9	3.78	7.94E-03
techn. replicate 2	678284	512180	1.324	97.7	3.92	8.23E-03
techn. replicate 3	565889	447103	1.266	93.4	3.52	7.39E-03
techn. replicate 4	664238	546597	1.215	89.6	3.60	7.56E-03
techn. replicate 5	719777	570445	1.262	93.1	3.73	7.82E-03
techn. mean value	652909	517365	1.263	93.1	3.71	7.79E-03
SD	57206	46591	0.040	2.9	0.16	3.30E-04
RSD [%]	8.8	9.0	3.1	3.1	4.2	4.2
	A(hmC)	A([D_2,$^{15}N_2$]-hmC)	A(hmC)/ A([D_2,$^{15}N_2$]-hmC)	n(hmC) [pmol]	hmC / G [%]	hmC / N
techn. replicate 1	80754	84338	0.958	8.51	0.350	7.35E-04
techn. replicate 2	115869	108774	1.065	9.46	0.380	7.97E-04
techn. replicate 3	264594	269496	0.982	8.72	0.329	6.90E-04
techn. replicate 4	139093	133830	1.039	9.23	0.371	7.78E-04
techn. replicate 5	116163	106074	1.095	9.73	0.389	8.18E-04
techn. mean value	143295	140502	1.028	9.13	0.364	7.64E-04
SD	70941	74213	0.057	0.51	0.024	5.12E-05
RSD [%]	49.5	52.8	5.6	5.6	6.7	6.7
	A(fC)	A([$^{15}N_2$]-fC)	A(fC)/ A([$^{15}N_2$]-fC)	n(fC) [pmol]	fC / G [%]	fC / N
techn. replicate 1	163427	31747	5.148	0.199	8.18E-03	1.72E-05
techn. replicate 2	178366	32585	5.474	0.211	8.48E-03	1.78E-05
techn. replicate 3	196827	35504	5.544	0.214	8.07E-03	1.69E-05
techn. replicate 4	193392	33755	5.729	0.221	8.88E-03	1.86E-05
techn. replicate 5	193493	35959	5.381	0.208	8.32E-03	1.75E-05
techn. mean value	185101	33910	5.455	0.211	8.38E-03	1.76E-05
SD	14070	1816	0.214	0.008	3.17E-04	6.66E-07
RSD [%]	7.6	5.4	3.9	3.9	3.8	3.8
	A(caC)	A([$^{15}N_2$]-caC)	A(caC)/ A([$^{15}N_2$]-caC)	n(caC) [pmol]	caC / G [%]	caC / N
techn. replicate 1	1444	6666	0.217	0.0104	4.28E-04	8.99E-07
techn. replicate 2	1735	8205	0.211	0.0102	4.08E-04	8.57E-07
techn. replicate 3	2111	9709	0.217	0.0104	3.93E-04	8.26E-07
techn. replicate 4	1985	8301	0.239	0.0114	4.59E-04	9.63E-07
techn. replicate 5	1927	8581	0.225	0.0108	4.31E-04	9.05E-07
techn. mean value	1840	8292	0.222	0.0106	4.24E-04	8.90E-07
SD	260	1089	0.011	0.0005	2.48E-05	5.20E-08
RSD [%]	14.1	13.1	4.8	4.6	5.8	5.8
	A(hmU)	A([D_2]-hmU)	A(hmU)/ A([D_2]-hmU)	n(hmU) [pmol]	hmU / G [%]	hmU / N
techn. replicate 1	894	5799	0.154	0.0090	3.69E-04	7.74E-07
techn. replicate 2	1278	6368	0.201	0.0130	5.23E-04	1.10E-06
techn. replicate 3	1561	7679	0.203	0.0133	5.00E-04	1.05E-06

Nature Chemical Biology: doi:10.1038/nchembio.1532

techn. replicate 4	1695	8148	0.208	0.0137	5.49E-04	1.15E-06
techn. replicate 5	1236	9010	0.137	0.0075	3.00E-04	6.29E-07
techn. mean value	1333	7401	0.181	0.0113	4.48E-04	9.41E-07
SD	312	1309	0.033	0.0029	1.08E-04	2.27E-07
RSD [%]	23.4	17.7	18.1	25.3	24.2	24.2
	A(fU)	**A([$^{15}N_2$]-fU)**	**A(fU)/ A([$^{15}N_2$]-fU)**	**n(fU) [pmol]**	**fU / G [%]**	**fU / N**
techn. replicate 1	7527	19661	0.383	0.0553	2.27E-03	4.78E-06
techn. replicate 2	8672	22769	0.381	0.0550	2.21E-03	4.63E-06
techn. replicate 3	11884	23930	0.497	0.0711	2.68E-03	5.63E-06
techn. replicate 4	10143	27784	0.365	0.0528	2.12E-03	4.45E-06
techn. replicate 5	14349	30152	0.476	0.0682	2.73E-03	5.73E-06
techn. mean value	10515	24859	0.420	0.0605	2.40E-03	5.04E-06
SD	2695	4148	0.061	0.0085	2.82E-04	5.93E-07
RSD [%]	25.6	16.7	14.5	14.0	11.7	11.7
	A(8oxo-G)	**A([$^{15}N_5$]-8oxo-G)**	**A(8oxo-G)/ A([$^{15}N_5$]-8oxo-G)**	**n(8oxo-G) [pmol]**	**8oxo-G / G [%]**	**8oxoG / N**
techn. replicate 1	237655	245919	0.966	0.113	4.63E-03	9.73E-06
techn. replicate 2	256991	266182	0.965	0.113	4.51E-03	9.48E-06
techn. replicate 3	310924	294412	1.056	0.123	4.64E-03	9.74E-06
techn. replicate 4	282245	299479	0.942	0.110	4.41E-03	9.26E-06
techn. replicate 5	327930	305105	1.075	0.125	5.01E-03	1.05E-05
techn. mean value	283149	282219	1.001	0.117	4.64E-03	9.75E-06
SD	37187	25225	0.060	0.007	2.28E-04	4.79E-07
RSD [%]	13.1	8.9	6.0	6.0	4.9	4.9

Nature Chemical Biology: doi:10.1038/nchembio.1532

Supplementary Note 3: LC-MS/MS quantification results of genomic DNA

Supplementary Table 9. LC-MS/MS quantification results of mESC (WT01, J1 and R1) wild type, knock down (KD) and knock out cells (related to Fig. 1c,, 2a and 4a). The results of independent biological replicates are shown, the biological mean values / N and the biological standard deviation (SD). n.d. = not detected.

biol. replicate	C / N	mC / N	hmC / N	fC / N	caC / N	hmU / N	fU / N	8-oxo-G / N
mESC (WT01) 1	2.01E-01	8.08E-03	6.53E-04	9.35E-06	6.01E-07	6.94E-07	3.51E-06	6.75E-06
mESC (WT01) 2	2.02E-01	7.71E-03	6.41E-04	9.81E-06	5.90E-07	7.04E-07	2.52E-06	5.95E-06
mESC (WT01) 3	2.02E-01	7.78E-03	6.82E-04	9.26E-06	5.24E-07	4.81E-07	1.60E-06	4.02E-06
mESC (WT01) 4	2.01E-01	8.17E-03	6.92E-04	1.03E-05	6.44E-07	4.73E-07	2.17E-06	4.90E-06
mESC (WT01) 5	2.02E-01	6.72E-03	8.63E-04	1.36E-05	7.25E-07	3.83E-07	4.34E-06	5.76E-06
mESC (WT01) 6	2.03E-01	6.43E-03	7.70E-04	1.32E-05	3.81E-07	5.14E-07	1.30E-06	3.17E-06
mESC (WT01) 7	2.03E-01	6.10E-03	8.18E-04	1.08E-05	6.77E-07	4.78E-07	1.03E-06	3.47E-06
biol. mean value	2.02E-01	7.28E-03	7.31E-04	1.09E-05	5.92E-07	5.32E-07	2.35E-06	4.86E-06
SD	7.75E-04	8.48E-04	8.63E-05	1.81E-06	1.13E-07	1.21E-07	1.21E-06	1.36E-06

biol. replicate	C / N	mC / N	hmC / N	fC / N	caC / N	hmU / N	fU / N	8-oxo-G / N
Tet1 KD (WT01) 1	2.01E-01	8.99E-03	2.76E-04	3.31E-06	2.21E-07	1.42E-07	7.98E-07	3.21E-06
Tet1 KD (WT01) 2	2.02E-01	7.77E-03	1.39E-04	2.53E-06	2.20E-07	6.44E-08	2.00E-06	3.73E-06
Tet1 KD (WT01) 3	2.03E-01	6.37E-03	2.36E-04	6.51E-06	2.07E-07	3.07E-07	4.13E-06	7.08E-06
biol. mean value	2.02E-01	7.71E-03	2.17E-04	4.11E-06	2.16E-07	1.71E-07	2.31E-06	4.67E-06
SD	1.33E-03	1.31E-03	7.03E-05	2.11E-06	7.87E-09	1.24E-07	1.69E-06	2.10E-06

biol. replicate	C / N	mC / N	hmC / N	fC / N	caC / N	hmU / N	fU / N	8-oxo-G / N
Tet2 KD (WT01) 1	2.00E-01	9.41E-03	4.45E-04	6.12E-06	2.38E-07	9.51E-08	5.18E-07	2.82E-06
Tet2 KD (WT01) 2	2.02E-01	7.25E-03	3.38E-04	4.52E-06	9.02E-08	3.23E-07	6.01E-07	1.78E-06
Tet2 KD (WT01) 3	2.02E-01	7.36E-03	3.98E-04	6.25E-06	3.36E-07	1.21E-07	2.50E-06	5.21E-06
Tet2 KD (WT01) 4	2.02E-01	7.69E-03	3.90E-04	6.78E-06	3.37E-07	n.d.	1.16E-06	6.52E-06
biol. mean value	2.02E-01	7.93E-03	3.93E-04	5.92E-06	2.50E-07	1.80E-07	1.20E-06	4.08E-06
SD	1.04E-03	1.01E-03	4.36E-05	9.73E-07	1.16E-07	1.25E-07	9.16E-07	2.17E-06

biol. replicate	C / N	mC / N	hmC / N	fC / N	caC / N	hmU / N	fU / N	8-oxo-G / N
mESC (J1) 1	2.04E-01	5.53E-03	3.61E-04	2.18E-06	7.95E-07	1.48E-06	3.98E-06	6.68E-06
mESC (J1) 2	2.04E-01	5.63E-03	3.32E-04	2.17E-06	9.30E-07	1.41E-06	4.19E-06	6.64E-06
biol. mean value	2.04E-01	5.58E-03	3.47E-04	2.17E-06	8.62E-07	1.44E-06	4.08E-06	6.66E-06
SD	4.96E-05	6.97E-05	2.06E-05	8.94E-09	9.55E-08	4.91E-08	1.51E-07	2.93E-08

biol. replicate	C / N	mC / N	hmC / N	fC / N	caC / N	hmU / N	fU / N	8-oxo-G / N
DNMT1 -/- (J1) 1	2.07E-01	2.72E-03	1.80E-04	1.36E-06	3.10E-07	1.68E-06	4.56E-06	7.22E-06
DNMT1 -/- (J1) 2	2.07E-01	2.71E-03	1.77E-04	1.55E-06	2.94E-07	1.56E-06	3.90E-06	6.73E-06
biol. mean value	2.07E-01	2.72E-03	1.79E-04	1.46E-06	3.02E-07	1.62E-06	4.23E-06	6.97E-06
SD	5.80E-06	3.35E-06	2.28E-06	1.31E-07	1.11E-08	8.47E-08	4.73E-07	3.45E-07

biol. replicate	C / N	mC / N	hmC / N	fC / N	caC / N	hmU / N	fU / N	8-oxo-G / N
DNMT3ab -/- (J1) 1	2.10E-01	4.12E-04	3.01E-05	2.46E-07	n.d.	8.02E-07	6.25E-07	2.68E-06
DNMT3ab -/- (J1) 2	2.10E-01	3.76E-04	3.54E-05	5.67E-07	n.d.	2.60E-06	7.86E-07	5.73E-06
DNMT3ab -/- (J1) 3	2.10E-01	3.81E-04	3.03E-05	3.71E-07	n.d.	2.27E-06	3.94E-07	2.32E-06
DNMT3ab -/- (J1) 4	2.10E-01	2.67E-04	2.67E-05	4.52E-07	n.d.	1.71E-06	4.50E-07	1.79E-06
DNMT3ab -/- (J1) 5	2.10E-01	2.50E-04	1.07E-05	3.78E-07	n.d.	7.41E-07	4.51E-06	7.68E-06
DNMT3ab -/- (J1) 6	2.10E-01	2.32E-04	1.60E-05	2.64E-07	n.d.	5.28E-07	2.87E-06	5.50E-06
biol. mean value	2.10E-01	3.20E-04	2.48E-05	3.80E-07		1.44E-06	1.61E-06	4.28E-06
SD	9.60E-05	7.83E-05	9.51E-06	1.19E-07		8.76E-07	1.70E-06	2.35E-06

biol. replicate	C / N	mC / N	hmC / N	fC / N	caC / N	hmU / N	fU / N	8-oxo-G / N
mESC (R1) 1	2.01E-01	8.37E-03	2.00E-04	1.04E-06	n.d.	1.52E-06	2.56E-06	1.21E-05
mESC (R1) 2	2.03E-01	6.76E-03	3.30E-04	2.00E-06	3.03E-07	9.23E-07	1.20E-06	n.d.
mESC (R1) 3	2.01E-01	8.68E-03	2.70E-04	2.12E-06	n.d.	1.77E-06	4.67E-06	1.05E-05
biol. mean value	2.02E-01	7.94E-03	2.67E-04	1.72E-06	3.03E-07	1.40E-06	2.81E-06	8.42E-06
SD	9.92E-04	1.03E-03	6.52E-05	5.91E-07		4.35E-07	1.75E-06	5.00E-06

Nature Chemical Biology: doi:10.1038/nchembio.1532

Supplementary Table 10. LC-MS/MS quantification results of different murine organs of 3 months old wild type individuals (*n*) (related to Fig. 1d, 2 and Supplementary Fig. 2). Compiled are mean values / N obtained from three independent technical replicates and the standard deviation (SD).

n	organ	DNA isolation	C / N		mC / N		hmC / N		fC / N	
			techn. mean	SD	techn. mean	SD	techn. mean	SD	techn. mean	SD
1	cerebellum		2.01E-01	2.30E-04	8.06E-03	2.40E-04	6.59E-04	1.08E-05	3.45E-07	3.14E-08
1	cerebellum	BHT, + Desf.,	2.01E-01	2.67E-04	8.27E-03	2.59E-04	7.37E-04	1.76E-05	2.86E-07	1.30E-08
1	cerebellum	BHT, + Desf., +THU	2.01E-01	3.66E-04	8.52E-03	3.45E-04	7.29E-04	2.14E-05	2.97E-07	2.03E-08
2	cerebellum	BHT, + Desf., +THU	2.00E-01	3.92E-04	9.81E-03	2.80E-04	6.37E-04	4.00E-05	2.73E-07	1.16E-08
3	cerebellum	BHT, + Desf., +THU	2.00E-01	3.95E-04	8.95E-03	3.45E-04	6.90E-04	2.91E-05	2.61E-07	1.57E-08
4	cerebellum	BHT, + Desf., +THU	2.00E-01	4.55E-04	9.46E-03	3.60E-04	6.69E-04	2.45E-05	2.59E-07	3.89E-09
1	kidney		2.01E-01	1.72E-04	8.13E-03	1.59E-04	4.37E-04	2.79E-05	2.27E-07	1.10E-08
1	kidney	BHT, + Desf.	2.02E-01	3.38E-04	7.79E-03	3.29E-04	4.33E-04	8.84E-06	2.25E-07	2.75E-08
1	kidney	BHT, + Desf., +THU	2.02E-01	1.70E-04	7.71E-03	1.76E-04	4.27E-04	2.15E-05	2.11E-07	1.19E-08
2	kidney	BHT, + Desf., +THU	2.01E-01	4.44E-04	8.59E-03	3.75E-04	3.78E-04	1.35E-05	1.86E-07	1.43E-08
3	kidney	BHT, + Desf., +THU	2.01E-01	1.46E-04	8.20E-03	1.02E-04	3.73E-04	1.67E-05	1.88E-07	8.82E-09
4	kidney	BHT, + Desf., +THU	2.01E-01	3.17E-04	8.30E-03	2.45E-04	3.79E-04	2.06E-05	1.94E-07	2.45E-08
1	cortex	BHT, + Desf., +THU	2.00E-01	4.48E-04	9.06E-03	3.37E-04	1.12E-03	7.21E-05	4.09E-07	4.17E-08
2	cortex	BHT, + Desf., +THU	1.99E-01	2.05E-04	9.29E-03	1.48E-04	1.37E-03	2.05E-05	4.34E-07	1.14E-08
3	cortex	BHT, + Desf., +THU	2.00E-01	4.65E-04	9.23E-03	3.73E-04	1.14E-03	7.48E-06	4.24E-07	1.92E-08
1	hippocampus	BHT, + Desf., +THU	1.98E-01	1.88E-04	1.08E-02	2.11E-04	1.56E-03	6.83E-05	4.49E-07	2.82E-08
2	hippocampus	BHT, + Desf., +THU	1.98E-01	2.13E-04	1.08E-02	1.73E-04	1.55E-03	3.08E-05	7.08E-07	4.67E-08
3	hippocampus	BHT, + Desf., +THU	1.97E-01	1.78E-04	1.09E-02	1.44E-04	1.71E-03	2.66E-05	3.89E-07	2.66E-08
1	heart	BHT, + Desf., +THU	2.02E-01	3.25E-04	7.96E-03	2.87E-04	4.08E-04	3.32E-05	1.85E-07	1.66E-08
2	heart	BHT, + Desf., +THU	2.02E-01	1.68E-04	7.87E-03	1.31E-04	3.84E-04	2.30E-05	1.48E-07	2.64E-09
3	heart	BHT, + Desf., +THU	2.01E-01	4.25E-04	8.07E-03	3.42E-04	4.56E-04	4.39E-05	1.41E-07	1.35E-08
1	liver	BHT, + Desf., +THU	2.01E-01	2.32E-04	8.35E-03	1.92E-04	2.44E-04	3.24E-06	1.66E-07	2.23E-08
2	liver	BHT, + Desf., +THU	2.02E-01	3.16E-04	8.10E-03	2.61E-04	2.73E-04	8.57E-06	1.78E-07	2.26E-08
3	liver	BHT, + Desf., +THU	2.01E-01	3.04E-04	8.42E-03	2.39E-04	2.82E-04	3.44E-05	1.81E-07	2.91E-08

n	organ	DNA isolation	hmU / N		fU / N		8-oxo-G / N	
			techn. mean	SD	techn. mean	SD	techn. mean	SD
1	cerebellum		7.27E-08		2.03E-06	2.29E-07	5.09E-06	1.92E-07
1	cerebellum	BHT, + Desf.,	n.d.		6.89E-07	2.65E-08	2.96E-06	6.04E-08
1	cerebellum	BHT, + Desf., +THU	n.d.		8.05E-07	7.34E-08	3.30E-06	3.79E-08
2	cerebellum	BHT, + Desf., +THU	1.30E-07	7.94E-08	1.02E-06	4.83E-08	3.86E-06	2.70E-07
3	cerebellum	BHT, + Desf., +THU	6.01E-08	3.67E-08	9.33E-07	9.52E-08	3.66E-06	2.26E-07
4	cerebellum	BHT, + Desf., +THU	1.07E-07	7.00E-08	1.03E-06	8.63E-08	4.67E-06	5.69E-07
1	kidney		n.d.		7.58E-07	7.38E-08	3.66E-06	2.63E-07
1	kidney	BHT, + Desf.	n.d.		7.00E-07	5.30E-08	3.24E-06	3.73E-07
1	kidney	BHT, + Desf., +THU	n.d.		7.16E-07	1.73E-08	3.48E-06	2.63E-07
2	kidney	BHT, + Desf., +THU	1.08E-07	7.89E-08	1.08E-06	1.45E-07	3.73E-06	1.42E-07
3	kidney	BHT, + Desf., +THU	1.29E-07	9.06E-08	1.17E-06	1.67E-07	3.91E-06	5.99E-08
4	kidney	BHT, + Desf., +THU	n.d.		9.34E-07	6.36E-08	3.65E-06	1.57E-07
1	cortex	BHT, + Desf., +THU	1.46E-07	6.19E-09	1.28E-06	1.22E-07	2.51E-06	3.74E-08
2	cortex	BHT, + Desf., +THU	8.24E-08	2.81E-08	1.30E-06	7.35E-08	3.13E-06	1.16E-07
3	cortex	BHT, + Desf., +THU	2.05E-07	5.83E-08	1.43E-06	3.72E-08	3.58E-06	2.83E-07
1	hippocampus	BHT, + Desf., +THU	2.74E-07	6.06E-08	2.18E-06	1.29E-07	6.30E-06	2.17E-07
2	hippocampus	BHT, + Desf., +THU	6.00E-07	5.50E-08	3.20E-06	3.46E-07	1.31E-05	7.69E-07

Nature Chemical Biology: doi:10.1038/nchembio.1532

3	**hippocampus**	BHT, + Desf., +THU	4.68E-08	1.85E-08	9.85E-07	2.26E-07	5.03E-06	8.96E-08
1	**heart**	BHT, + Desf., +THU	4.14E-07	7.28E-08	1.97E-06	3.46E-07	7.85E-06	1.69E-06
2	**heart**	BHT, + Desf., +THU	9.81E-08	4.23E-09	1.15E-06	7.63E-08	2.48E-06	1.18E-07
3	**heart**	BHT, + Desf., +THU	1.04E-07	3.36E-08	1.55E-06	7.55E-08	3.14E-06	2.27E-07
1	**liver**	BHT, + Desf., +THU	n.d.		1.13E-06	9.97E-08	2.69E-06	3.93E-07
2	**liver**	BHT, + Desf., +THU	1.48E-07	8.67E-08	1.09E-06	1.96E-07	2.43E-06	9.41E-08
3	**liver**	BHT, + Desf., +THU	2.34E-08		1.05E-06	1.18E-07	2.39E-06	4.69E-08

Supplementary Table 11. LC-MS/MS quantification results of n = 6 independent mESC differentiation experiments without growth factors (related to Fig. 5a,b). Compiled are mean values / N obtained from three independent technical measurements and their standard deviation (SD). K = C57Bl6/129 derived mES cell line.

n	Sample	C / N		mC / N		hmC / N		fC / N		caC / N	
	mESC diff. time	techn. mean	SD	techn. mean	SD	techn. mean	SD	techn. mean	SD	techn. mean	SD
1	**K, t= 0h**	2.06E-01	1.47E-04	3.79E-03	1.52E-04	4.25E-04	8.84E-06	1.68E-05	9.20E-07	1.21E-06	5.80E-08
1	**K, t= 8h**	2.06E-01	9.23E-05	3.43E-03	9.96E-05	5.33E-04	1.27E-05	1.71E-05	7.46E-07	1.09E-06	2.97E-08
1	**K, t= 16h**	2.06E-01	1.07E-04	3.88E-03	1.15E-04	4.93E-04	9.98E-06	6.69E-06	2.33E-07	3.92E-07	4.81E-08
2	**R1, t= 0h**	2.08E-01	1.38E-04	1.87E-03	1.05E-04	3.89E-04	3.33E-05	2.17E-05	5.00E-07	1.52E-06	6.10E-08
2	**R1, t= 8h**	2.08E-01	2.55E-05	1.60E-03	1.73E-05	5.50E-04	1.14E-05	2.91E-05	1.12E-06	1.91E-06	3.30E-08
2	**R1, t= 16h**	2.08E-01	7.54E-05	1.74E-03	7.53E-05	5.87E-04	1.28E-05	1.89E-05	4.48E-07	1.04E-06	2.38E-08
3	**K, t= 0h**	2.05E-01	6.83E-05	4.24E-03	6.36E-05	3.84E-04	6.03E-06	9.95E-06	1.46E-07	1.74E-07	1.70E-08
3	**K, t= 8h**	2.05E-01	7.08E-05	4.08E-03	7.14E-05	4.85E-04	5.17E-06	1.47E-05	7.35E-07	2.60E-07	1.84E-08
3	**K, t= 24h**	2.05E-01	5.00E-05	4.65E-03	5.74E-05	4.48E-04	7.86E-06	4.00E-06	1.67E-07	8.74E-08	2.02E-08
4	**R1, t= 0h**	2.07E-01	6.37E-05	2.49E-03	5.86E-05	4.61E-04	1.43E-05	1.71E-05	4.28E-07	3.25E-07	4.80E-08
4	**R1, t= 8h**	2.07E-01	5.45E-05	2.27E-03	4.25E-05	5.53E-04	1.69E-05	2.63E-05	8.05E-07	5.71E-07	2.45E-08
4	**R1, t= 24h**	2.06E-01	5.12E-05	3.00E-03	5.28E-05	5.93E-04	7.13E-06	8.39E-06	2.73E-07	1.74E-07	2.04E-08
5	**K, t=0 h**	2.04E-01	1.02E-04	5.38E-03	1.04E-04	3.89E-04	4.27E-06	7.66E-06	2.13E-07	2.29E-07	1.73E-08
5	**K, t=8 h**	2.04E-01	1.74E-04	5.10E-03	1.72E-04	4.65E-04	4.64E-06	1.01E-05	5.36E-07	2.37E-07	8.38E-10
5	**K, t=16 h**	2.04E-01	2.04E-04	5.88E-03	2.09E-04	5.22E-04	9.13E-06	4.40E-06	2.13E-07	1.92E-07	1.34E-08
5	**K, t=24 h**	2.04E-01	1.84E-04	5.40E-03	1.79E-04	4.01E-04	6.47E-06	2.06E-06	1.87E-07	1.00E-07	1.16E-08
5	**K, t=40 h**	2.04E-01	1.71E-04	5.72E-03	1.71E-04	3.00E-04	1.20E-06	1.20E-06	5.32E-08	7.93E-08	
6	**R1, t=0 h**	2.06E-01	2.16E-05	3.56E-03	3.01E-05	5.85E-04	1.80E-05	2.08E-05	4.72E-07	5.49E-07	5.15E-08
6	**R1, t=8 h**	2.05E-01	9.98E-05	3.83E-03	8.48E-05	7.63E-04	1.51E-05	2.05E-05	5.32E-07	8.74E-07	9.94E-09
6	**R1, t=16 h**	2.06E-01	4.53E-05	3.44E-03	3.95E-05	6.89E-04	7.42E-06	1.20E-05	5.65E-07	3.03E-07	8.59E-09
6	**R1, t=24 h**	2.06E-01	1.19E-04	3.57E-03	8.65E-05	6.16E-04	3.52E-05	7.18E-06	2.03E-07	1.55E-07	2.87E-08
6	**R1, t=40 h**	2.05E-01	6.38E-06	4.95E-03	1.56E-05	4.42E-04	1.11E-05	1.97E-06	4.48E-08	9.46E-08	1.84E-08

n	Sample	hmU / N		fU / N		8-oxo-G / N	
	mESC diff. time	techn. mean	SD	techn. mean	SD	techn. mean	SD
1	**K, t= 0h**	3.75E-07	1.67E-09	3.89E-06	2.73E-07	6.66E-06	2.74E-07
1	**K, t= 8h**	1.08E-06	1.57E-07	4.83E-06	3.45E-07	7.79E-06	3.05E-07
1	**K, t= 16h**	1.01E-06	1.76E-07	5.11E-06	5.69E-07	8.07E-06	5.57E-07
2	**R1, t= 0h**	1.14E-06	3.65E-07	3.82E-06	4.10E-07	6.70E-06	1.72E-07
2	**R1, t= 8h**	2.97E-06	4.13E-07	5.23E-06	9.97E-08	9.08E-06	5.89E-07
2	**R1, t= 16h**	3.75E-06	6.55E-07	5.15E-06	5.72E-07	8.81E-06	5.37E-07
3	**K, t= 0h**	8.76E-07	2.02E-08	8.86E-06	9.44E-07	1.35E-05	9.84E-07
3	**K, t= 8h**	1.29E-06	1.76E-07	6.90E-06	9.81E-07	1.00E-05	6.53E-07
3	**K, t= 24h**	6.55E-07	5.89E-08	6.72E-06	7.12E-07	9.73E-06	5.55E-07

Nature Chemical Biology: doi:10.1038/nchembio.1532

4	**R1, t= 0h**	1.22E-06	1.58E-07	8.62E-06	7.83E-07	1.26E-05	8.90E-08
4	**R1, t= 8h**	1.91E-06	2.69E-07	6.02E-06	4.56E-07	8.05E-06	2.35E-07
4	**R1, t= 24h**	1.82E-06	3.67E-07	6.60E-06	6.25E-07	1.06E-05	6.27E-07
5	**K, t=0 h**	9.98E-07	2.95E-07	7.04E-06	1.35E-06	9.98E-06	9.08E-07
5	**K, t=8 h**	2.97E-06	2.62E-07	6.89E-06	5.67E-07	9.64E-06	5.85E-07
5	**K, t=16 h**	1.43E-06	3.46E-07	2.71E-06	1.92E-07	3.82E-06	3.04E-07
5	**K, t=24 h**	8.01E-07	2.16E-07	6.60E-06	5.93E-07	9.72E-06	1.33E-07
5	**K, t=40 h**	8.30E-07	3.44E-08	7.78E-06	1.31E-07	1.13E-05	3.07E-07
6	**R1, t=0 h**	1.85E-06	3.00E-07	8.62E-06	1.94E-07	1.12E-05	7.13E-07
6	**R1, t=8 h**	3.77E-06	6.22E-07	4.01E-06	7.44E-07	6.02E-06	1.41E-06
6	**R1, t=16 h**	3.94E-06	8.01E-08	5.93E-06	3.24E-07	7.61E-06	2.01E-07
6	**R1, t=24 h**	3.84E-06	3.90E-07	8.92E-06	1.03E-06	1.23E-05	2.83E-07
6	**R1, t=40 h**	1.84E-06	9.73E-08	8.44E-06	2.94E-07	1.26E-05	1.02E-06

Supplementary Table 12. Relative modification levels of combined data sets from differentiation (0–40 h) of R1 and C57Bl6/129-derived mESCs without growth factors (related to Fig. 5a and Supplementary Fig. 8). In order to obtain these, the absolute modification levels of t = 0 h time points compiled in Supplementary Table 11 were set as 1 and the modification levels of later time points respectively related to these. Summarized are the biological mean values at each differentiation time point and the standard deviation (SD).

	relative C / N		**relative mC / N**		**relative hmC / N**	
time	**biol. mean**	**SD**	**biol. mean**	**SD**	**biol. mean**	**SD**
t = 0 h	1.00000E+00		1.000E+00		1.000E+00	
t = 8 h	1.00024E+00	1.2480E-03	9.435E-01	7.579E-02	1.270E+00	8.082E-02
t = 16 h	9.98952E-01	1.4308E-03	1.004E+00	7.157E-02	1.296E+00	1.629E-01
t = 24 h	9.98556E-01	1.4742E-03	1.077E+00	9.513E-02	1.133E+00	1.176E-01
t = 40 h	9.96347E-01	3.4067E-03	1.228E+00	2.314E-01	7.634E-01	1.034E-02

	relative fC / N		**relative caC / N**	
time	**biol. mean**	**SD**	**biol. mean**	**SD**
t = 0 h	1.000E+00		1.000E+00	
t = 8 h	1.280E+00	2.318E-01	1.339E+00	3.329E-01
t = 16 h	6.051E-01	1.980E-01	5.991E-01	2.175E-01
t = 24 h	3.769E-01	9.373E-02	4.393E-01	1.122E-01
t = 40 h	1.255E-01	4.352E-02	2.589E-01	1.225E-01

	relative hmU / N		**relative fU / N**		**relative 8-oxo-G / N**	
time	**biol. mean**	**SD**	**biol. mean**	**SD**	**biol. mean**	**SD**
t = 0 h	1.000E+00		1.000E+00		1.000E+00	
t = 8 h	2.258E+00	6.597E-01	9.221E-01	3.417E-01	9.024E-01	3.178E-01
t = 16 h	2.390E+00	8.012E-01	9.337E-01	4.755E-01	8.978E-01	4.410E-01
t = 24 h	1.278E+00	6.283E-01	8.742E-01	1.357E-01	9.086E-01	1.623E-01
t = 40 h	9.129E-01	1.155E-01	1.043E+00	8.856E-02	1.129E+00	3.665E-04

Nature Chemical Biology: doi:10.1038/nchembio.1532

Supplementary Table 13. LC-MS/MS quantification results of mESC differentiation with the growth factors FGF-2 and ActA (related to Supplementary Fig. 9). Modified nucleosides / N are given as mean values plus SD of three independent technical replicates.

Sample	C / N		mC / N		hmC / N		fC / N	
EpiLC diff. time	techn. mean	SD	techn. mean	SD	techn. mean	SD	techn. mean	SD
t= 0h	2.096E-01	1.40E-05	3.92E-04	1.35E-05	4.57E-05	6.48E-07	2.88E-07	4.52E-09
t= 12h	2.093E-01	1.01E-05	5.80E-04	4.94E-06	6.85E-05	5.47E-06	1.65E-06	1.34E-07
t= 24h	2.074E-01	1.17E-04	2.29E-03	1.02E-04	2.60E-04	1.41E-05	4.45E-06	6.30E-09
t= 36h	2.044E-01	9.71E-05	4.96E-03	7.38E-05	5.81E-04	2.30E-05	6.16E-06	5.27E-08
t= 48h	2.020E-01	8.85E-05	7.23E-03	8.83E-05	7.53E-04	8.90E-07	4.92E-06	1.31E-07

Sample	hmU / N		fU / N		8-oxo-G / N	
EpiLC diff. time	techn. mean	SD	techn. mean	techn. mean	SD	techn. mean
t= 0h	4.76E-07	7.05E-08	1.87E-06	9.10E-08	4.00E-06	1.10E-07
t= 12h	2.01E-06	1.90E-07	3.91E-06	7.84E-07	9.80E-06	3.07E-07
t= 24h	2.78E-06	7.20E-07	2.42E-06	3.65E-07	5.19E-06	1.45E-07
t= 36h	2.10E-06	5.82E-07	2.20E-06	1.51E-07	5.42E-06	4.12E-07
t= 48h	1.90E-06	1.03E-07	3.74E-06	3.74E-07	6.69E-06	3.08E-07

Supplementary Table 14. HEK-293T wild type vs. HEK + Tet1cm vs. HEK + Tet1cd (related to Supplementary Fig. 7a). Modified nucleosides / N are given as mean values of three independent technical replicates.

	HEK-293T wild type		HEK + Tet1cm		HEK + Tet1cd	
Nucleosides	techn. mean	SD	techn. mean	SD	techn. mean	SD
C / N	2.01E-01	9.67E-05	2.02E-01	1.09E-04	2.03E-01	3.91E-04
mC / N	8.80E-03	9.62E-05	8.04E-03	1.09E-04	4.18E-03	3.93E-05
hmC / N	3.39E-05	5.44E-07	4.95E-05	1.04E-06	2.21E-03	9.42E-05
fC / N	2.43E-07	2.41E-09	2.18E-07	1.17E-08	2.48E-04	5.96E-06
caC / N	n.d.		n.d.		1.29E-04	3.35E-06
hmU / N	7.21E-07	5.93E-08	1.51E-07	9.07E-09	4.24E-05	2.51E-06
fU / N	4.89E-06	1.59E-07	1.42E-06	6.66E-08	8.79E-06	1.07E-07
8-oxo-G / N	1.06E-05	1.14E-07	6.83E-06	1.47E-07	8.31E-06	9.03E-07

Supplementary Table 15. DNA modification levels of Tet1 *in vitro* assay (related to Supplementary Fig. 7b). Plasmid DNA with full CpG methylation was treated with commercially available Tet1.

plasmid	mC / N	hmC / N	fC / N	caC / N	hmU / N	fU / N	8-oxo-G / N
untreated	4.8E-02	n.d.	n.d.	n.d.	n.d.	4.4E-06	8.6E-06
untreated	4.6E-02	n.d.	n.d.	n.d.	n.d.	4.0E-06	9.4E-06
+Tet1	7.6E-04	4.8E-03	5.0E-03	9.0E-03	3.8E-04	1.5E-04	4.0E-05
+Tet1	6.5E-04	4.7E-03	4.8E-03	9.3E-03	4.4E-04	1.6E-04	3.7E-05
-Tet1	4.8E-02	n.d.	3.5E-05	n.d.	1.6E-05	2.5E-04	4.0E-05
-Tet1	4.7E-02	n.d.	3.5E-05	n.d.	1.2E-05	2.3E-04	4.5E-05

Nature Chemical Biology: doi:10.1038/nchembio.1532

Supplementary Table 16. Effect of Smug1 depletion on modification levels in mESCs (R1) and effect of TDG and SMUG1 depletion on modification levels in HEK-293T cells overexpressing Tet1cd (related to Supplementary Fig. 5). Percent change values for modified nucleosides of cells treated with esiRNA (targeting Smug1/SMUG1 or TDG) with respect to unrelated control esiRNA. The absolute modification content of HEK-293T cells was normalized based on Tet1cd expression levels (determined by TECAN reading). The percent change is given as a mean value of three independent technical replicates.

	mESC Smug1 KD		HEK-293T + Tet1cd / TDG KD		HEK-293T + Tet1cd / SMUG1 KD	
Nucleosides	**Percent change**	**SD**	**Percent change**	**SD**	**Percent change**	**SD**
hmC	-0.15	4.07	-2.22	2.61	-3.43	2.24
fC	-1.26	3.29	38.15	1.25	0.77	2.05
caC	14.04	29.24	33.51	2.72	6.59	3.24
hmU	37.14	9.47	-10.41	4.74	46.46	5.65
fU	71.00	15.73	-4.36	2.34	22.66	1.91
8-oxo-G	17.27	14.26	-5.36	2.04	-6.97	1.77

Supplementary Table 17. HEK-293T wild type, HEK with Tet1cd-overexpression, Tet1cd/Uhrf1 co-overexpression or Tet1cd/Uhrf2 co-overexpression (related to Supplementary Fig. 12). Modified nucleosides / N are given as mean values of three independent technical replicates.

	HEK-293T wt		HEK + Tet1cd		HEK + Tet1cd + Uhrf1		HEK + Tet1cd + Uhrf2	
Nucleosides	**techn. mean**	**SD**	**techn. mean**	**SD**	**techn. mean**	**SD**	**techn. mean**	**SD**
C / N	2.04E-01	2.97E-05	2.04E-01	1.24E-04	2.05E-01	8.32E-05	2.04E-01	1.15E-04
mC / N	6.01E-03	2.74E-05	5.29E-03	1.22E-04	4.63E-03	8.93E-05	5.11E-03	1.15E-04
hmC / N	2.86E-05	1.57E-06	4.29E-04	1.85E-05	3.31E-04	1.93E-05	5.13E-04	9.85E-06
fC / N	4.51E-07	5.21E-08	5.97E-05	3.43E-06	7.21E-05	4.24E-06	1.39E-04	5.61E-06
caC / N	1.77E-07	2.67E-09	2.32E-05	8.50E-07	2.23E-04	1.84E-05	1.66E-04	1.21E-06
hmU / N	8.18E-07	8.89E-08	2.55E-06	6.26E-08	1.27E-05	4.30E-07	6.74E-06	1.59E-06
fU / N	7.60E-06	3.39E-07	5.40E-06	5.26E-07	5.64E-06	8.42E-07	1.22E-05	2.10E-06
8-oxo-G / N	1.27E-05	7.08E-07	9.23E-06	3.22E-07	6.93E-06	1.09E-06	1.62E-05	8.43E-07

Nature Chemical Biology: doi:10.1038/nchembio.1532

Supplementary Note 4: materials in cell culture

Supplementary Table 18. Overexpression plasmids and esiRNAs used in HEK-293T cell experiments.

Figure	Experiment	Sample	Plasmid DNA	esiRNA
S7a	HEK-293T +/- Tet1xx	HEK + Tet1cd	GFP-Tet1cd (7.5 µg)	x
		HEK + Tet1cm	mCh-Tet1cm (7.5 µg)	x
		Wild type	pCMV6-Cdk5Rap1-v2 (7.5 µg)	x
S5b,S5c	Tet1cd with TDG or SMUG1 KD	HEK + Tet1cd	GFP-Tet1cd (10 µg)	CDK5RAP1 esiRNA (5 µg)
		HEK + Tet1cd with TDG KD	GFP-Tet1cd (10 µg)	TDG esiRNA (5 µg)
		HEK + Tet1cd with SMUG1 KD	GFP-Tet1cd (10 µg)	SMUG1 esiRNA (5 µg)
S12	HEK-293T	Tet1cd	GFP-Tet1cd (6 µg)	x
	HEK-293T	Tet1cd + Uhrf1	GFP-Tet1cd, GFP-Uhrf1[1] (each 6 µg)	x
	HEK-293T	Tet1cd + Uhrf2	GFP-Tet1cd, GFP-Uhrf2[1] (each 6 µg)	x

Supplementary Table 19. Knockdown (KD) efficiencies by Tet relative to SCR shRNAs.

	FWD	REV	Reference
Tet1	GAGCCTGTTCCTCGATGTGG	CAAACCCACCTGAGGCTGTT	*Ito et. al.*[2]
Tet2	TGTTGTTGTCAGGGTGAGAATC	TCTTGCTTCTGGCAAACTTACA	*Ito et. al.*[2]
actin	AAGGCCAACCGTGAAAAGAT	GTGGTACGACCAGAGGCATAC	This work

Supplementary Table 20. Primers for qPCR analysis of Tet, Dnmt, Tdg and Smug1 of EpiLC differentiation and Smug1 knockdown samples.

	FWD	REV	Reference
Gapdh	CATGGCCTTCCGTGTTCCTA	CTTCACCACCTTCTTGATGTCATC	*Szwagierzcak et al.*[3]
Tet1	CCAGGAAGAGGCGACTACGTT	TTAGTGTTGTGTGAACCTGATTTATTGT	*Szwagierzcak et al.*[3]
Tet2	ACTTCTCTGCTCATTCCCACAGA	TTAGCTCCGACTTCTCGATTGTC	*Szwagierzcak et al.*[3]
Tet3	GAGCACGCCAGAGAAGATCAA	CAGGCTTTGCTGGGACAATC	*Szwagierzcak et al.*[3]
Dnmt1	CCTAGTTCCGTGGCTACGAGGAG	TCTCTCTCCTCTGCAGCCGACTC	This work
Dnmt3a	GCTTTCTTCTCAGCCTCCCT	CCATGCCAAGACTCACCTTC	This work
Dnmt3b	CTGGCACCCTCTTCTTCATT	ATCCATAGTGCCTTGGGACC	This work
Tdg	GTCTGTTCATGTCGGGGCTGAGTGAG	CTGCAGTTTCTGCACCAGGATGCGC	This work
Smug1	CACTGGGGCCTACCCATGA	CTCCCAAGCATAATCCACCG	This work

Nature Chemical Biology: doi:10.1038/nchembio.1532

Supplementary Note 5: correlation analysis results of modification levels

Supplementary Table 21. Correlation analysis of DNA modification levels comparing murine tissues from three months old individuals (cortex, hippocampus, cerebellum, heart, liver and kidney). Pearson coefficients (p) and significance values (s) are summarized. n=24 independent DNA samples (see Supplementary Table 10). Highlighted in gray are strong to very strong correlations ($|p| > 0.7$) with significance levels (s) lower than 0.001 (marked with *). Additionally, moderate correlations ($0.7 > |p| > 0.6$) with significance level lower than 0.001 are highlighted in light gray.

		C	mC	hmC	fC	8-oxo-G	hmU	fU
C	**p**	1.000	-0.986*	-0.935*	-0.832*	-0.534	-0.466	-0.479
	s		0.000	0.000	0.000	0.007	0.022	0.018
mC	**p**	-0.986*	1.000	0.863*	0.769*	0.534	0.464	0.460
	s	0.000		0.000	0.000	0.007	0.022	0.024
hmC	**p**	-0.935*	0.863*	1.000	0.887*	0.481	0.420	0.472
	s	0.000	0.000		0.000	0.017	0.041	0.020
fC	**p**	-0.832*	0.769*	0.887*	1.000	0.663*	0.586	0.649*
	s	0.000	0.000	0.000		0.000	0.003	0.001
8-oxo-G	**p**	-0.534	0.534	0.481	0.663*	1.000	0.837*	0.835*
	s	0.007	0.007	0.017	0.000		0.000	0.000
hmU	**p**	-0.466	0.464	0.420	0.586	0.837*	1.000	0.871*
	s	0.022	0.022	0.041	0.003	0.000		0.000
fU	**p**	-0.479	0.460	0.472	0.649*	0.835*	0.871*	1.000
	s	0.018	0.024	0.020	0.001	0.000	0.000	

Supplementary Table 22. Correlation analysis of DNA modification levels during early mESC differentiation (0-40 h). Pearson coefficients (p) and significance values (s) are summarized. n=22 independent DNA samples (see Supplementary Table 11). Highlighted in gray are strong to very strong correlations ($|p| > 0.7$) with significance levels (s) lower than 0.001 (marked with *). Additionally, weak correlations of hmC/hmU with mC/C are highlighted in pale pink.

		C	mC	hmC	fC	caC	8-oxo-G	hmU	fU
C	**p**	1	-0.997*	0.324	0.806*	0.699*	-0.032	0.299	-0.145
	s		0.000	0.141	0.000	0.000	0.889	0.176	0.518
mC	**p**	-0.997*	1	-0.399	-0.815*	-0.693*	0.057	-0.356	0.160
	s	0.000		0.066	0.000	0.000	0.801	0.104	0.476
hmC	**p**	0.324	-0.399	1	0.404	0.182	-0.312	0.783*	-0.227
	s	0.141	0.066		0.062	0.417	0.157	0.000	0.309
fC	**p**	0.806*	-0.815*	0.404	1	0.797*	-0.266	0.303	-0.299
	s	0.000	0.000	0.062		0.000	0.232	0.170	0.177
caC	**p**	0.699*	-0.693*	0.182	0.797*	1	-0.435	0.165	-0.567
	s	0.000	0.000	0.417	0.000		0.043	0.463	0.006
8-oxo-G	**p**	-0.032	0.057	-0.312	-0.266	-0.435	1	-0.080	0.959*
	s	0.889	0.801	0.157	0.232	0.043		0.723	0.000
hmU	**p**	0.299	-0.356	0.783*	0.303	0.165	-0.080	1	-0.022
	s	0.176	0.104	0.000	0.170	0.463	0.723		0.921
fU	**p**	-0.145	0.160	-0.227	-0.299	-0.567	0.959*	-0.022	1
	s	0.518	0.476	0.309	0.177	0.006	0.000	0.921	

Nature Chemical Biology: doi:10.1038/nchembio.1532

Supplementary References

1. Pichler, G. et al. Cooperative DNA and histone binding by Uhrf2 links the two major repressive epigenetic pathways. *J Cell Biochem* **112**, 2585-93 (2011).
2. Ito, S. et al. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333**, 1300-3 (2011).
3. Szwagierczak, A., Bultmann, S., Schmidt, C.S., Spada, F. & Leonhardt, H. Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic Acids Res* **38**, e181 (2010).
4. Huang, D.W., Sherman, B.T., Lempicki R.A. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. *Nature Protoc.***4**(1):44-57 (2009).

Nature Chemical Biology: doi:10.1038/nchembio.1532