

Dissertation zur Erlangung des Doktorgrades
der Fakultät für Chemie und Pharmazie
der Ludwig-Maximilians-Universität München

**Untersuchung des Metabolismus nicht-kanonischer
DNA-Basen in embryonalen Stammzellen**

Angie Gabriele Kirchner

aus

Ulm, Deutschland

2020

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 selbstständig und ohne unerlaubte Hilfe erarbeitet.

München, den 09.01.2020

.....
Angie G. Kirchner

Dissertation eingereicht am: 09.01.2020

1. Gutachter: Prof. Dr. Thomas Carell

2. Gutachterin: Dr. Stefanie Kellner

Mündliche Prüfung am: 31.01.2020

„I am just a child who has never grown up.

I still keep asking these ‘how’ and ‘why’ questions.

Occasionally, I find an answer.”

Stephen Hawking

Teile dieser Arbeit wurden in folgenden Journalen oder auf Fachkonferenzen publiziert:

Publikationen

F. Spada, S. Schiffers, A. Kirchner, Y. Zhang, O. Kosmatchev, E. Korytiakova, R. Rahimoff, C. Ebert, T. Carell, *Oxidative and non-oxidative active turnover of genomic methylcytosine in distinct pluripotent states*, *bioRxiv* **2019**, 846584, 10.1101/846584.

K. Iwan*, R. Rahimoff*, A. Kirchner*, F. Spada*, A. S. Schröder, O. Kosmatchev, S. Ferizaj, J. Steinbacher, E. Parsa, M. Müller, T. Carell, *5-Formylcytosine to cytosine conversion by C-C bond cleavage in vivo*, *Nat. Chem. Biol.* **2018**, *14*, 72-78, 10.1038/nchembio.2531.

R. Rahimoff*, O. Kosmatchev*, A. Kirchner*, T. Pfaffeneder, F. Spada, V. Brantl, M. Müller, T. Carell, *5-Formyl- and 5-carboxydeoxycytidines do not cause accumulation of harmful repair intermediates in stem cells*, *J. Am. Chem. Soc.* **2017**, *139*, 10359-10364, 10.1021/jacs.7b04131.

M. Su, A. Kirchner, S. Stazzoni, M. Müller, M. Wagner, A. Schröder, T. Carell, *5-Formylcytosine could be a semipermanent base in specific genome sites*, *Angew. Chem. Int. Ed. Engl.* **2016**, *55*, 11797-11800, 10.1002/anie.201605994.

* Beitrag der Autoren zu gleichen Teilen.

Konferenzbeiträge

The 43rd FEBS Congress – Biochemistry Forever, The Czech Society for Biochemistry and Molecular Biology, Prague, Czech Republic (2018) – Posterpräsentation: *“Studies towards a C-C bond cleavage of 5-formyl-2'-deoxycytidine in mammalian cells”*

SFB 1309 Meeting – Chemical Biology of Epigenetic Modifications, Ludwig-Maximilians-University, Munich, Germany (2017) – Posterpräsentation: *“Investigation of enzymes involved in the C-C bond cleavage of 5-formyldeoxycytidine”*

Frontiers in Epigenetic Chemical Biology, The Royal Society, London, United Kingdom (2017) – Posterpräsentation: *“A novel reagent for the detection and quantification of base excision repair intermediates”*

Danksagung

An erster Stelle möchte ich meinem Doktorvater *Prof. Dr. Thomas Carell* für die herausfordernde Themenstellung und die exzellenten Arbeitsbedingungen danken. Ebenso danke ich dir für deine stete Unterstützung und das entgegengebrachte Vertrauen in meine Arbeit.

Dr. Stefanie Kellner danke ich ganz herzlich für die Übernahme des Koreferats.

Allen weiteren Mitgliedern der Prüfungskommission *Prof. Dr. Stylianos Michalakis, Prof. Dr. Lena Daumann, Dr. Dietmar Martin* und *Prof. Dr. Konstantin Karaghiosoff* danke ich vielmals für ihre Mitwirkung an der mündlichen Prüfung.

Frau *Slava Gärtner* danke ich für ihre Hilfe in allen organisatorischen Angelegenheiten.

Dr. Fabio Spada möchte ich für seine fachliche, wie auch persönliche Unterstützung und guten Ratschläge in allen Phasen meiner Promotion herzlichst danken. Deine Begeisterung für die Stammzellbiologie und die vielen lehrreichen Unterhaltungen haben mich zum Gelingen dieser Arbeit motiviert.

Dr. Markus Müller gebührt großer Dank für seine bedingungslose Hilfsbereitschaft in allen Belangen. Ebenso danke ich dir für die Ermutigung Ergebnisse kritisch zu hinterfragen und die angeregten wissenschaftlichen Diskussionen. Dank dir habe ich mich nicht nur persönlich weiterentwickelt, sondern auch meine graphischen Fähigkeiten verbessert.

Ein besonderer Dank geht auch an alle Korrekturleser *Dr. Sarah Schiffers, Eva Korytiakovà, Dr. Katharina Iwan, Franziska Traube, Dr. Fabio Spada* und *Dr. Markus Müller*.

Großer Dank gebührt meinen beiden Lieblingsmitstreitern *Dr. René Rahimoff* und *Dr. Katharina Iwan*. Unsere gute Zusammenarbeit hat den großen Erfolg unserer Themen erst möglich gemacht. Ich werde eure Effizienz, die angeregten Diskussionen und eure unerschütterliche Begeisterung für die Wissenschaft vermissen.

Auch dem „*BabySafe-Lab*“, *Johanna Bretzler* mit *Anna* sowie *Franziska Traube* mit *Malin* (alias *Erwin*), möchte ich für die gute Stimmung, die lustigen Momente und die gute Arbeitsatmosphäre danken. Ich kann mich glücklich schätzen so tolle Laborkolleginnen gehabt zu haben. *Franziska Traube* leider hatten wir nie die Gelegenheit zusammen die Weltherrschaft zu erobern, aber es wäre großartig geworden - und wer weiß, was die Zukunft bringt!

Meinen QQQ-Ladies *Dr. Jessica Steinbacher, Dr. Olesya Kosmatchev, Dr. Sarah Schiffers* und *Dr. Katharina Iwan* danke ich für die Prozessierung und Auswertung unzähliger DNA-Proben.

Herzlichst möchte ich mich auch bei *Kerstin Kurz, Kristof Hufnagel* und *Luis de la Osa de la Rosa* für die große Hilfe bei der täglichen Laborarbeit bedanken. Ohne euren Einsatz bei der DNA-Isolation, in der Zellkultur oder dem Probenverdau wäre vieles nicht so reibungslos verlaufen.

Ebenfalls möchte ich mich bei der „Bärenhöhle“ *Dr. Sidney Becker, Dr. Mirko Wagner, Dr. Katharina Iwan, Dr. Michael Stadlmeier, Dr. René Rahimoff* und *Dr. Nada Raddaoui* für die freundliche Aufnahme ins Team bedanken. Es war eine schöne Zeit mit euch und ein toller Ausklang meiner Promotion.

An dieser Stelle möchte ich mich bei meinen Praktikanten und Masteranden *Shqiponja Ferizaj, Peyton Randolph, David Schmidl* und *Ewelina Kamińska* für ihre Unterstützung und wissenschaftlichen Diskussionen bedanken. Es hat mir unglaublich viel Spaß gemacht, euch zu betreuen und euren wissenschaftlichen und persönlichen Fortschritt mitzuerleben. Ich wünsche euch viel Erfolg und alles Gute für eure Zukunft! *Ewelina Kamińska* es freut mich sehr so eine motivierte und würdige Nachfolgerin zu haben - ich bin mir sicher du wirst alle Herausforderungen meistern!

Auch allen weiteren Mitgliedern der Arbeitsgruppe möchte ich für die tolle Arbeitsatmosphäre und den guten Zusammenhalt danken.

Viele Kollegen aus der Arbeitsgruppe konnte ich in dieser Zeit als Freunde gewinnen. Ich bin sehr froh euch hier kennengelernt zu haben und blicke auf unzählige lustige Abende mit euch beim Kochen, beim „Topmodeln“, im Biergarten oder bei Mottopartys zurück. Auch meinen Freunden außerhalb der Uni möchte ich für ihre Unterstützung, Geduld und ihr Verständnis an langen Arbeitstagen danken. *Céline Kreiser* und *Carina Brosche* herzlichen Dank für die Korrektur meiner Arbeit – auch wenn ihr fachfremd seid, habt ihr tolle Arbeit geleistet!

Zum Schluss gilt mein herzlichster Dank meinen Eltern, die mich schon mein ganzes Leben begleiten und unterstützen. Ihr habt nie die Geduld und den Glauben an mich verloren und mir stets den nötigen Rückhalt geboten. Meinem Freund, *Dr. Dominik Leitz*, danke ich für seine Unterstützung, Liebe und guten Ratschläge. Ich hätte diese Zeit ohne dich nicht so gut gemeistert und bin dankbar, dich an meiner Seite zu haben.

Inhaltsverzeichnis

Zusammenfassung	I
Summary.....	V
1 Einleitung	1
1.1 Zelluläre Differenzierung und Epigenetik	1
1.2 Stammzellen als Modell der frühen Entwicklung.....	4
1.2.1 Gewinnung der pluripotenten Stammzellen.....	4
1.2.2 Naïve und <i>primed</i> Pluripotenz.....	6
1.3 Epigenetische Regulation	9
1.3.1 Chromatinstruktur und Histonmodifikationen	9
1.3.2 Nicht-kodierende RNA	11
1.4 DNA-Methylierung	12
1.4.1 DNA-Methyltransferasen	13
1.4.2 Dynamik in der zellulären Entwicklung.....	15
1.5 Tet-vermittelte Bildung der oxidierten mdC-Derivate	16
1.5.1 5-Hydroxymethyl-2'-desoxycytidin	17
1.5.2 5-Formyl-2'-desoxycytidin und 5-Carboxyl-2'-desoxycytidin.....	18
1.6 DNA-Demethylierung	20
1.6.1 Passive DNA-Demethylierung	20
1.6.2 Dynamik nicht-kanonischer DNA-Bausteine in der Embryonalentwicklung	21
1.6.3 Aktive DNA-Demethylierung.....	22
2 Zielsetzung	29
3 Ergebnisse und Publikationen.....	31
3.1 Studien zur DNA-Demethylierung über eine C-C Bindungsbruchreaktion	32
3.1.1 Umwandlung von 5-Formyl-desoxycytidin zu Desoxycytidin über einen C-C Bindungsbruch in lebenden Zellen.....	33
3.1.2 Untersuchung der aktiven Demethylierung in pluripotenten Zellen mittels isotopenmarkierter Cytidin-Derivate	44

3.1.3	Untersuchung einer enzymvermittelten DNA-Demethylierung über einen C-C Bindungsbruch	65
3.2	Studien zur Basenexzisionsreparatur-vermittelten aktiven Demethylierung	77
3.2.1	5-Formyl- und 5-Carboxydesoxycytidine verursachen keine schädlichen Reparatur-intermediate in murinen embryonalen Stammzellen	78
3.2.2	5-Formylcytosin ist vermutlich eine semipermanente Base an definierten Genompositionen	86
3.2.3	Involvierung der BER an der Entfernung modifizierter DNA-Basen	91
3.3	Studien einer oxidationsunabhängigen DNA-Demethylierung	102
3.3.1	Oxidationsunabhängige Demethylierung durch die DNA-Reparatur	103
4	Material und Methoden der unveröffentlichten Arbeiten	121
4.1	Material und Methoden zu Abschnitt 3.1.2	121
4.2	Material und Methoden zu Abschnitt 3.1.3	124
4.3	Material und Methoden zu Abschnitt 3.2.3	127
4.4	Material und Methoden zu Abschnitt 3.3.1	132
5	Abkürzungsverzeichnis	142
6	Literaturverzeichnis	146
7	Anhang: Zusatzmaterialien der veröffentlichten Arbeiten	163
7.1	Zusatzmaterialien zu Abschnitt 3.1.1	163
7.2	Zusatzmaterialien zu Abschnitt 3.2.1	197
7.3	Zusatzmaterialien zu Abschnitt 3.2.2	245

Zusammenfassung

Alle Zellen eines multizellulären Organismus besitzen eine identische Desoxyribonukleinsäure (DNA)-Sequenz, unterscheiden sich aber dennoch in ihrer Funktion und ihren Eigenschaften. Zu diesem Zweck werden bestimmte Gene im Zuge der zellulären Entwicklung zu definierten Zeitpunkten an- oder abgeschaltet. Zur Inaktivierung von Genen kann die DNA durch die DNA-Methyltransferasen (Dnmts) kovalent modifiziert werden, wobei 5-Methyl-2'-desoxycytidin (mdC) gebildet wird. Das Muster der Methylierung ist für zahlreiche zelluläre Prozesse von besonderer Bedeutung und bestimmt das Schicksal einer Zelle. Während die Entstehung von mdC bereits gut untersucht wurde, ist dessen Entfernung noch nicht vollständig verstanden. Mit der Entdeckung der Oxidationsprodukte der *Ten-eleven translocation* (Tet)-Enzyme: 5-Hydroxymethyl-2'-desoxycytidin (hmdC), 5-Formyl-2'-desoxycytidin (fdC) und 5-Carboxyl-2'-desoxycytidin (cadC) wurde der genetische Code um weitere nicht-kanonische DNA-Modifikationen erweitert. Diese DNA-Bausteine wurden vor kurzem als Intermediate eines aktiven DNA-Demethylierungsweges vorgeschlagen. Zugleich wird eine unabhängige Bedeutung für die DNA-Reparatur und die Kontrolle der Genexpression vermutet.

Die vorgelegte Dissertation beschäftigte sich mit der Untersuchung verschiedener Mechanismen der DNA-Demethylierung in der frühen Embryonalentwicklung, insbesondere beim Übergang vom naiven Präimplantations- zum *primed* Postimplantationsepiblast. Hierzu wurden drei potentielle Demethylierungswege genauer betrachtet und sind in Abbildung 1 dargestellt.

Zunächst wurde der Weg über eine direkte Demethylierung durch metabolische Markierung von mdC mit Hilfe der Isotopenverdünnungstechnik massenspektrometrisch verfolgt. Dabei wurde in einem *Pulse-Chase* Experiment eine rasche Abnahme an markierten mdCs im Verlauf des *Primings* beobachtet und gleichzeitig ein Anstieg des markierten Desaminierungsprodukts 2'-Desoxythymidin (dT) detektiert. Die entsprechenden höher oxidierten dC-Derivate blieben dabei länger in der genomischen DNA bestehen, was auf eine unterschiedliche Kinetik zur Entfernung dieser Spezies hindeutet. Durch *Doppelknockout* zweier für den Nukleotidmetabolismus wichtiger Desaminasen, der *Cytidine Deaminase* und der *Desoxycytidylate Deaminase* wurde erfahren, dass das markierte mdC vermutlich durch eine Desaminierungsreaktion im löslichen Nukleosid/-tid Pool als markiertes dT in die genomische DNA gelangt. Zudem wurden Hinweise auf eine Beteiligung der DNA-Reparatur-maschinerie, wie der *Mismatch*-Reparatur (MMR) oder der Basenexzisionsreparatur (BER), erhalten.

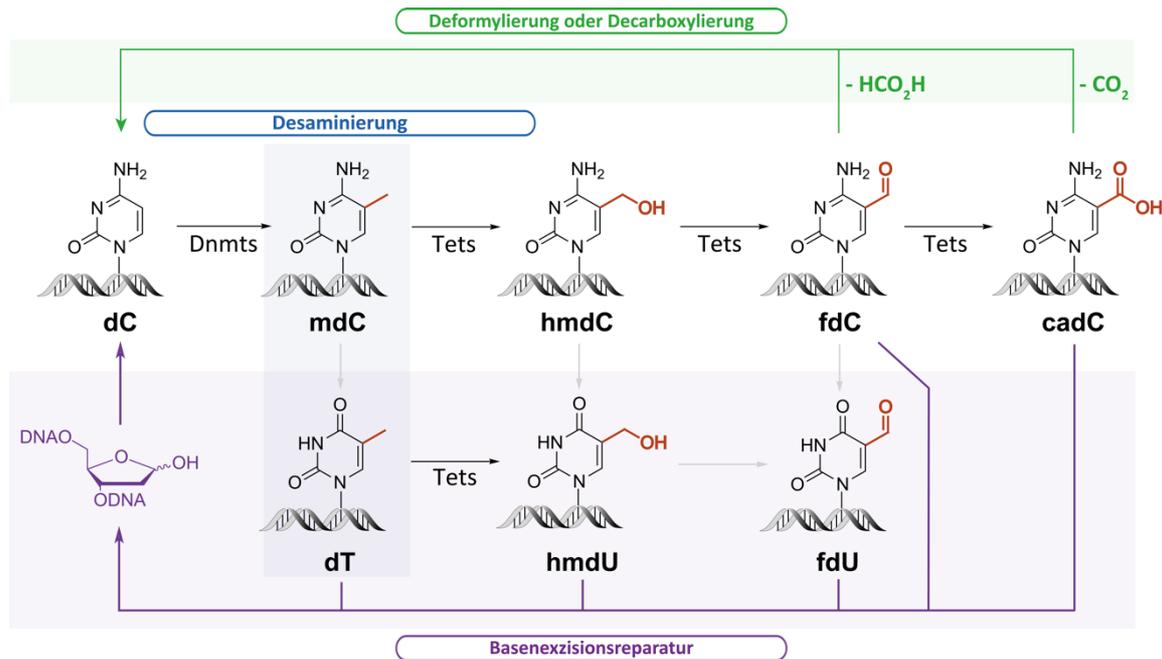


Abbildung 1: Untersuchte Wege der aktiven Demethylierung und Rückgewinnung von dC. Oxidations-unabhängige Demethylierung durch Desaminierung und DNA-Reparatur (blau). Der zweite und der dritte mutmaßliche Weg beinhalten die Exzision von fdC bzw. cadC im Verlauf der BER (violett) oder die direkte Demodifikation der oxidierten mdC-Derivate über einen C-C-Bindungsbruch (grün).

Der Schwerpunkt dieser Dissertation lag auf der Untersuchung der Entfernung der nicht-kanonischen DNA-Bausteine fdC und cadC im Zuge der aktiven Demethylierung. Kürzlich konnte gezeigt werden, dass die Thymin-DNA Glykosylase (Tdg) die modifizierten Basen erkennt und die glykosidische Bindung spaltet. Zurück bleibt eine abasische Stelle, welche im Rahmen der BER weiter prozessiert wird, um unmodifiziertes dC zurückzugewinnen. Als Teilprojekt dieser Arbeit wurde gezeigt, dass Tdg nur die Hälfte aller fdCs an einer bestimmten genomischen Stelle entfernt, was darauf hindeutet, dass dieser epigenetische Baustein an bestimmten Genompositionen durchaus stabil etabliert werden kann.

Die Hinweise verdichten sich zunehmend, dass die aktive Demethylierung über die Basenexzisionsreparatur stattfindet. Es ist jedoch schwer nachzuvollziehen, aus welchem DNA-Baustein eine abasische Stelle entsteht. Daher wurde eine UHPLC-MS/MS-basierte Methode zur Analyse dieser kurzlebigen BER-Intermediate entwickelt. Dies ermöglichte erstmals eine simultane Quantifizierung von abasischen Stellen und β -Eliminierungsprodukten. Eine Akkumulation der von dC stammenden, abasischen Stellen konnte nicht beobachtet werden. Es wurde vermutet, dass diese Intermediate zu schnell prozessiert werden, wodurch eine Exposition der abasischen Stelle nicht zugelassen wird. In einer kürzlich veröffentlichten Studie konnte gezeigt werden, dass die bifunktionalen Neil-Glykosylasen die Prozessierung der abasischen Stellen durch Tdg beschleunigen. Daher wäre es möglich, dass diese in Abwesenheit der Neil-Glykosylasen akkumulieren.

Mit der entwickelten Analysemethode wurde jedoch keine Anhäufung von abasischen Stellen oder β -Eliminierungsprodukten festgestellt. Es wurde spekuliert, dass Tdg aufgrund seiner hohen Affinität zu abasischen Stellen deren weitere Prozessierung beeinflusst. Daher wurden zur weiteren Untersuchung der Tdg-Beteiligung an der aktiven Demethylierung mit Hilfe des CRISPR-Cas9-Systems ein Tdg-*Knockout* und eine katalytisch inaktive Tdg-Mutante in embryonalen Stammzellen der Maus (mESCs) hergestellt.

In einem weiteren Teil dieser Arbeit wurde mit Hilfe von Isotopenverfolgungsexperimenten und der Fütterung alternativer DNA-Nukleoside eine direkte Demodifikation von fdC zu dC in mESCs und somatischen Zellen nachgewiesen. Hierzu wurden dem Kulturmedium fluorierte oder mit stabilen Isotopen markierte fdC-Derivate zugesetzt (Abbildung 2A) und die Bildung des Demodifizierungsproduktes mittels UHPLC-MS/MS verfolgt. Die Demodifikation fand in allen untersuchten Zelltypen statt und zwar unabhängig von Dnmt oder Tet Enzymen. Darüber hinaus wurde festgestellt, dass das Demodifizierungsprodukt [$^{13}\text{C}_5$][$^{15}\text{N}_2$]-2'-Desoxycytidin ($^{13}\text{C}_5^{15}\text{N}_2$ -dC) bzw. 2'-(R)-Fluorodesoxycytidin (F-dC) aktiv methyliert wird, was zur Annahme führte, dass die beobachtete C-C Bindungsbruchreaktion an der Genregulation beteiligt sein könnte.

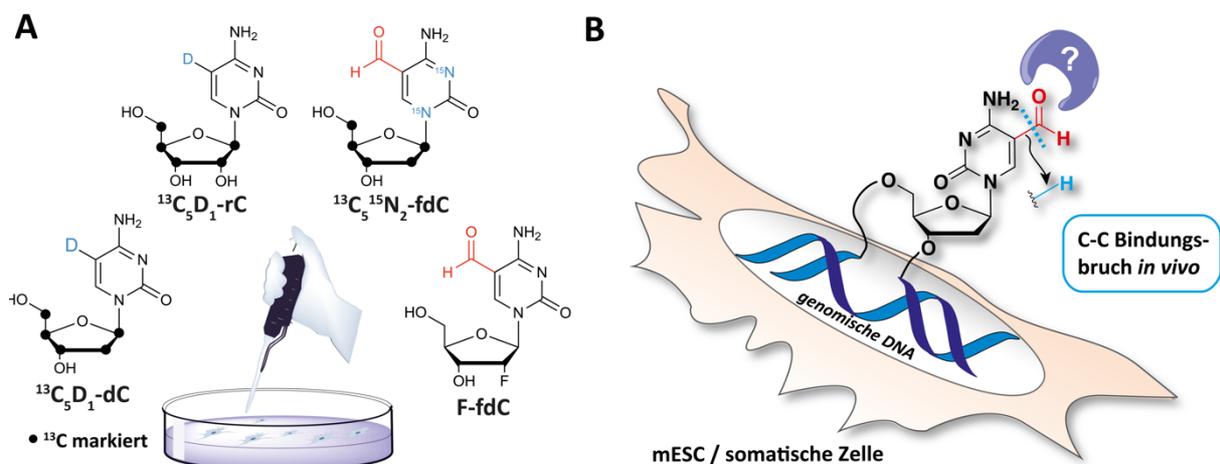


Abbildung 2: Demodifikation über einen C-C Bindungsbruch. **A)** Verwendete Sondenmoleküle: [$^{13}\text{C}_5$][$^{15}\text{N}_2$]-5-Formyl-2'-desoxycytidin ($^{13}\text{C}_5^{15}\text{N}_2$ -fdC), 2'-(R)-Fluoro-5-formyl-desoxycytidin (F-fdC), [$^{13}\text{C}_5$][D₁]-2'-Desoxycytidin ($^{13}\text{C}_5\text{D}_1$ -dC), [$^{13}\text{C}_5$][D₁]-Cytidin ($^{13}\text{C}_5\text{D}_1$ -rC). **B)** Schematische Darstellung des C-C Bindungsbruchs in pluripotenten oder verschiedenen somatischen Zellen.

Darüber hinaus konnten weitere Belege für die Existenz der beobachteten direkten C-C-Bindungsspaltung aus Isotopenverfolgungsstudien in mESCs und induzierten pluripotenten Stammzellen (iPSCs) mit verschiedenen deuterierten Sondenmolekülen gewonnen werden (siehe Abbildung 2A).

Ein starker Hinweis auf die C-C Bindungsspaltung wurde aus einem *Pulse-Chase* Experiment in *Neurogenin*-induzierbaren iPSCs nach Induktion der Neurogenese gewonnen. Hierbei wurde ein deutlicher Anstieg des Demodifizierungsproduktes $^{13}\text{C}_5\text{-dC}$ 72 h nach der Induktion beobachtet. Da diese Zellen ab dem 48 h Zeitpunkt postmitotisch sind, konnte von einer aktiven Akkumulation des Spaltungsproduktes ausgegangen werden.

Nach dem erfolgreichen Nachweis des C-C Bindungsbruchs als Mechanismus der aktiven Demethylierung wurde im Folgenden nach Faktoren gesucht, die diesen Prozess begünstigen könnten (siehe Abbildung 2B). So konnten gezeigt werden, dass sowohl Dnmt Enzyme selbst, als auch NAD-abhängige Enzyme, wie z.B. die Sirtuine, einen klaren Einfluss auf das F-dC zu F-fdC Verhältnis hatten. Die Identifizierung des verursachenden Faktors oder Enzyms erfordert jedoch weitere Untersuchungen. Unter Berücksichtigung der Ergebnisse aller bisherigen Experimente kann davon ausgegangen werden, dass weitere Prozesse an der beschriebenen Demodifikation beteiligt sind.

Summary

All cells of a multicellular organism, despite having an identical deoxyribonucleic acid (DNA) sequence, still differ in their function and properties. For this purpose, certain genes are switched on or off at defined timepoints in the course of cellular development. DNA can be covalently modified by DNA methyltransferases (Dnmts) to form 5-methyl-2'-deoxycytidine (mdC) to inactivate genes. The pattern of methylation is of particular importance for numerous cellular processes and determines the fate of a cell. While the creation of mdC has already been well investigated, its removal is not yet fully understood. With the discovery of the oxidation products of the Ten eleven translocation (Tet) enzymes: 5-hydroxymethyl-2'-deoxycytidine (hmdC), 5 formyl-2'-deoxycytidine (fdC) and 5-carboxyl-2'-deoxycytidine (cadC), the genetic code was recently extended by further non-canonical DNA modifications. Recently, these cytosine modifications have been proposed as intermediates of an active DNA demethylation pathway and could also exert independent roles in DNA repair and control of gene expression.

The Ph.D. thesis focuses on the investigation of different mechanisms of DNA demethylation during early embryonic development, particularly during the transition from the naïve preimplantation to the primed post-implantation epiblast. Three potential demethylation pathways were considered that are depicted in Figure 1.

First, the route of a direct demethylation of mdC was monitored by metabolic labelling and analyzed using isotope-dilution mass spectrometry. A pulse-chase experiment revealed during priming, a rapid decrease of the isotopically labelled mdC and a simultaneous increase of the labelled deamination product 2'-deoxythymidin (dT). The higher oxidized dC derivatives remained longer in the genomic DNA, suggesting different erasure kinetics. Double knockout of two deaminases important for the nucleoside metabolism, cytidine deaminase and deoxycytidylate deaminase, showed that the labelled mdC probably enters the genomic DNA as labelled dT through a deamination reaction in the soluble nucleoside/tide pool. In addition, evidence for the involvement of the DNA repair machinery, such as mismatch repair (MMR) and base excision repair (BER), was obtained.

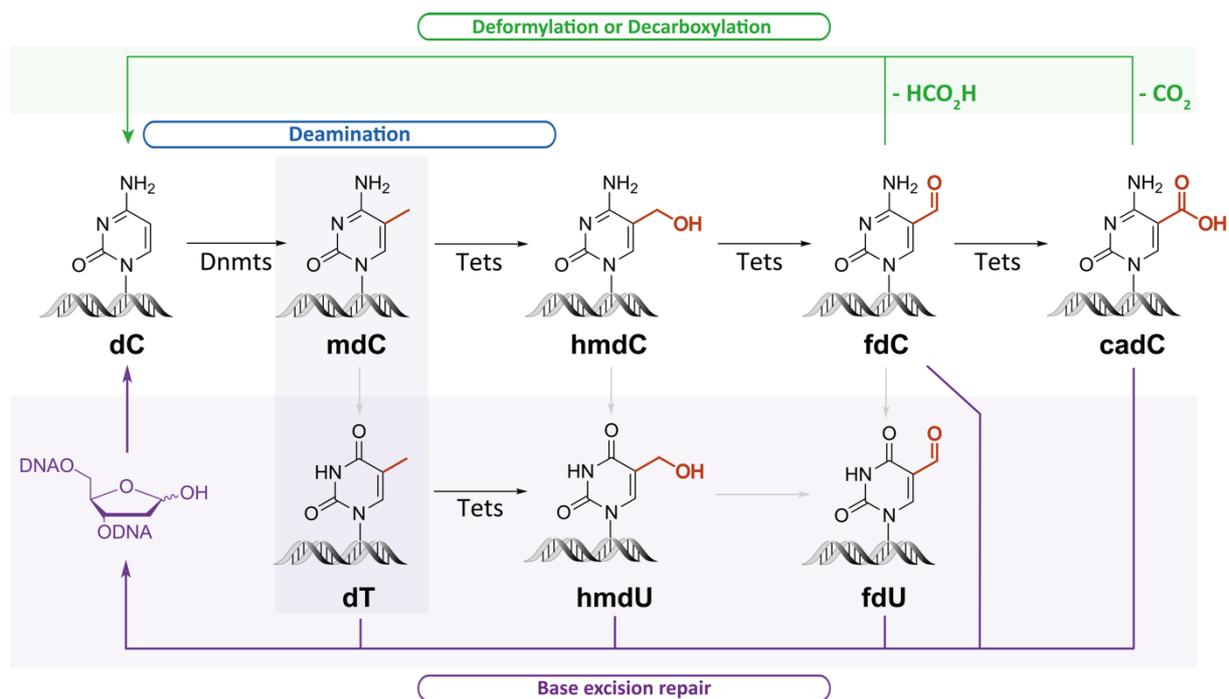


Figure 1: Investigated active demethylation pathways. Oxidation-independent demethylation of mdC via deamination and DNA repair (blue). The second and the third putative pathways involve excision of fdC and cadC, respectively, in the course of BER (violet) or direct demodification of the oxidized mdC derivatives via C-C bond cleavage (green).

The main focus of this work was the investigation of the removal of the non-canonical DNA building blocks fdC and cadC during active demethylation. Recently, it has been shown that the thymine DNA glycosylase (Tdg) recognizes modified bases and cleaves the glycosidic bond. What remains is an abasic site, which is processed further as part of the BER process to reestablish unmodified dC. We could show that Tdg removes only half of all the fdCs at a given genomic site suggesting that the epigenetic mark is permanent at specific positions.

Accumulated evidence suggests that active DNA demethylation processes through BER. However, it is difficult to determine from which DNA building block an abasic site is formed. Therefore, a UHPLC-MS/MS-based method for the analysis of the short-lived BER intermediates was developed. This allowed a simultaneous quantification of abasic sites and β -elimination products for the first time. Based on isotope tracing data it was concluded that dC does not contribute to the formation of abasic sites in the genome. It was therefore concluded that these intermediates are located in a closed complex with only limited exposure of the abasic site. A study published recently has revealed that the bifunctional *Nei endonuclease VIII-like* (Neil) glycosylases accelerate the processing of abasic sites by Tdg. It would therefore be possible for these to accumulate in the absence of the Neil glycosylases. However, no accumulation of abasic sites or β -elimination products was detected with the developed analytical method.

It has been suspected that Tdg affects further processing of abasic sites due to its high affinity to abasic sites. Thus, a Tdg knockout and a catalytically inactive mutant were generated in mouse embryonic stem cells (mESCs) to further investigate Tdg contribution to active demethylation using the CRISPR-Cas9 system.

In another study, isotope tracking experiments and feeding of alternative DNA nucleosides were used to demonstrate a direct demodification from fdC to dC in mESCs and somatic cells. Therefore, fluorinated or stable isotope-labelled fdC derivatives were added to the culture medium (Figure 2A) and the generation of the demodification product was traced by UHPLC-MS/MS. The demodification was observed in all investigated cell types and was independent of the presence of Dnmt or Tet enzymes. In addition, it was found that the demodification product [$^{13}\text{C}_5$][$^{15}\text{N}_2$]-2'-deoxycytidine ($^{13}\text{C}_5^{15}\text{N}_2$ -dC) and 2'-(R)-Fluoro-deoxycytidine (F-dC), respectively, is actively methylated, which led to the assumption that the observed demodification reaction could be involved in gene regulation.

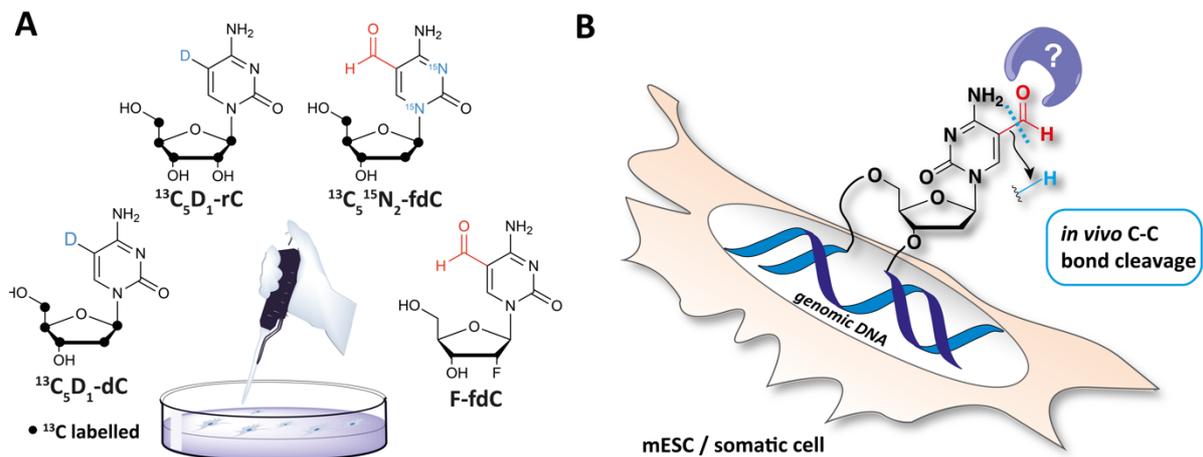


Figure 2: Demodification via C-C bond cleavage. **A)** Probe molecules used: [$^{13}\text{C}_5$][$^{15}\text{N}_2$]-5-Formyl-2'-deoxycytidine ($^{13}\text{C}_5^{15}\text{N}_2$ -fdC), 2'-(R)-Fluoro-5-formyldeoxycytidine (F-fdC), [$^{13}\text{C}_5$][D₁]-2'-Deoxycytidine ($^{13}\text{C}_5\text{D}_1$ -dC), [$^{13}\text{C}_5$][D₁]-Cytidine ($^{13}\text{C}_5\text{D}_1$ -rC). **B)** Schematic representation of the C-C bond cleavage reaction in pluripotent stem cells or various somatic cells.

In addition, further evidence for the existence of the observed direct C-C bond cleavage could be obtained from isotope tracing studies in mESCs and, in particular, induced pluripotent stem cells (iPSCs) using different deuterated probe molecules (Figure 2A). The strongest support was gained in a *Pulse-Chase* experiment post induction of neurogenesis in *Neurogenin*-inducible iPSCs. Thereby, a clear increase of the demodification product $^{13}\text{C}_5$ -dC was observed 72h post induction. As these cells are post-mitotic from the 48h timepoint on, an active accumulation of the C-C bond cleavage product could be assumed.

Following the successful demonstration of a C-C bond cleavage as active demethylation mechanism, the next step was to investigate factors that could favour this process (Figure 2B). As a result, it was shown that Dnmt enzymes and NAD-dependent enzymes, such as Sirtuins, had a clear effect on the F-dC to F-fdC ratio. However, the identification of the causative factor or enzyme requires further investigation. Considering the results of all previous experiments, it can be assumed that further processes are involved in order to generate the demodification described above.

1 Einleitung

Die Entwicklung eines vielzelligen Organismus beginnt mit der befruchteten Eizelle, der sogenannten Zygote. Durch eine Reihe von Zellteilungen, Differenzierungs- und Reprogrammierungsprozessen kann sich daraus ein vollständiges Lebewesen mit hunderten spezialisierten Zelltypen unterschiedlichster physiologischer Funktionen entwickeln.^[1] Alle Zellen enthalten das gleiche Erbgut, unterscheiden sich aber dennoch in Funktion und Eigenschaften. Wieso sind beispielsweise eineiige Zwillinge unterschiedlich? Die Tochterzellen einer einzigen Zygote unterziehen sich bei der Embryonalentwicklung unabhängigen mitotischen Teilungen. Der daraus entstehende Organismus ist genetisch identisch, kann aber bezüglich des Phänotyps und der Anfälligkeit gegenüber Krankheiten signifikante Unterschiede aufweisen.^[2] Eine mögliche Erklärung bieten epigenetische Veränderungen, die im Laufe des Alterungsprozesses der Individuen wirksam werden.^[3] In welchem Maße die Epigenetik für das Schicksal einer Zelle, ihrer Identität und den Phänotyp dieser Lebensform verantwortlich ist, muss jedoch noch vollständig geklärt werden.

1.1 Zelluläre Differenzierung und Epigenetik

Alle Zellen mehrzelliger Organismen besitzen eine identische Desoxyribonukleinsäure (DNA)-Sequenz. Dennoch sind diese somatischen Zellen in der Lage, sich auf unterschiedliche Zellfunktionen zu spezialisieren.^[4] Dies geschieht durch verschiedene Arten der Regulation von Genexpression, die in der Summe zu einem spezifischen Erscheinungsbild führen. Die Beobachtungen von *Conrad Waddington* lieferten bereits in der Mitte des letzten Jahrhunderts die Grundlage für den Begriff der Epigenetik.^[5] Dieser beschrieb erstmalig den Prozess der Epigenese. Bei diesem führte ein Genotyp im Zuge der Entwicklung zu einem bestimmten Phänotyp.^[6] Die Epigenetik beschreibt eine der DNA-Sequenz übergeordnete (altgr. *epi*) Informationsebene, wodurch die genetische Information in verschiedenen Zelltypen unterschiedlich genutzt wird. Erst einige Jahre später erkannten *Robin Holliday* und *Arthur Riggs*, dass Änderungen der Genexpressionsmuster vererbbar sind.^[7, 8] Nach heutigen Erkenntnissen beschreibt die Epigenetik molekulare Prozesse, die die Genexpression unabhängig von der eigentlichen DNA-Sequenz verändern oder regulieren.^[9] Äußere Einflüsse können DNA- und Histon-Modifikationen auslösen und Proteine rekrutieren, die das Zellschicksal dynamisch und vererbbar beeinflussen.^[10] In dem Modell nach *Conrad Waddington* wird dieser komplexe Prozess der Spezialisierung (Differenzierung) einer Zelle mit einer Kugel verglichen, die einen Abhang mit verzweigten Tälern hinunterrollt (siehe Abbildung 3).

Die Täler symbolisieren dabei die Möglichkeiten der Zelle für bestimmte Entwicklungsrichtungen. Ist ein Pfad einmal eingeschlagen, kann die Kugel nicht ohne weiteres den Hang hinaufrollen oder Anhöhen überqueren. Jedes Tal stellt ein anderes Aktivitätsmuster an Genen dar, das durch ein komplexes Netzwerk der epigenetischen Regulation etabliert wird. Das Gegenstück der Differenzierung bildet die Dedifferenzierung bzw. Reprogrammierung. Bei diesem Prozess können zelluläre Identitäten umgekehrt werden. Dies versetzt die Zelle in einen anderen Differenzierungs- oder den nicht-spezialisierten Ursprungszustand zurück. In *Waddingtons* Modell wurde dieser Weg nicht berücksichtigt, da die Entwicklung nicht spontan in Richtung undifferenzierter Zustände oder des Ursprungszustands „zurückrollen“ kann. Die Möglichkeit der Reprogrammierung ist aber in der modernen Forschung nicht mehr wegzudenken. Die Differenzierung und die Veränderung molekularer Muster des Epigenoms (epigenetische Reprogrammierung) beeinflussen entscheidend den Lebenszyklus eines Organismus schon während der frühen Entwicklungsphase bis hin zum Erwachsenenalter.

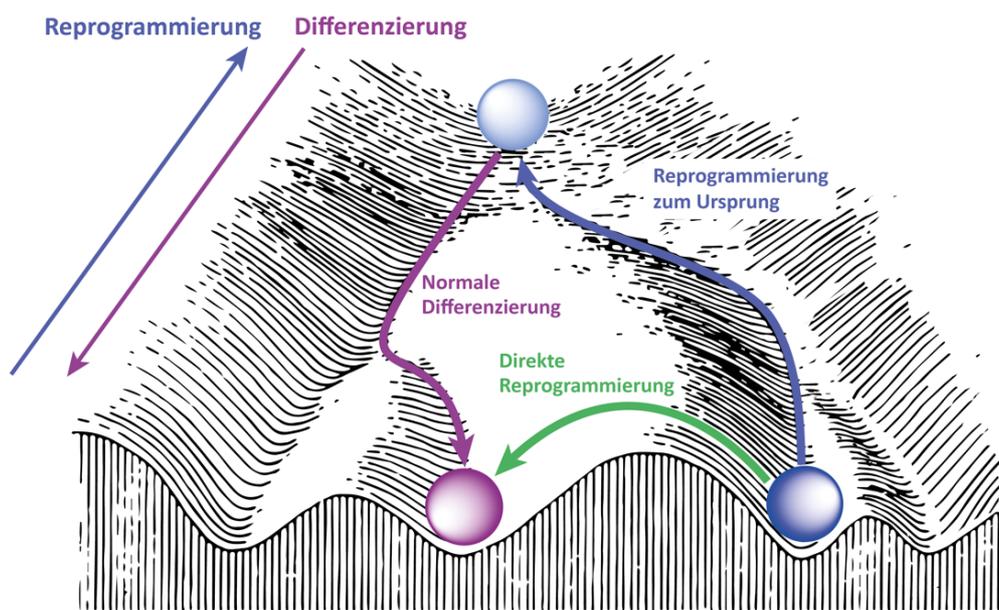


Abbildung 3: Modell der zellulären Differenzierung und Reprogrammierung. Dargestellt ist eine Abwandlung des *Waddington* Modells, indem das Zellschicksal der nicht spezialisierten Vorläuferzelle mit einer Kugel (hellblau) verglichen wird, die einen Abhang mit mehreren Tälern hinunterrollt. Diese Täler stellen jeweils vollständig differenzierte Zelltypen dar. Durch Reprogrammierung epigenetischer Markierungen können differenzierte Zellen den Abhang hinaufrollen und in ein anderes Tal wieder hinunterrollen. Mittels direkter Reprogrammierung bzw. Transdifferenzierung (grün) kann eine differenzierte Zelle (blau) direkt in eine andere differenzierte Zelle (violett) umgewandelt werden.^[11] (Abbildung nach *Srivastava et al.*)

Die Möglichkeit der künstlichen Reprogrammierung wurde von *John Gurdon* zum ersten Mal genutzt. Ihm gelang es aus einer bereits differenzierten somatischen Zelle einen vollständigen lebensfähigen Organismus zu entwickeln. Der Zellkern einer Epithelzelle aus einer Kaulquappe wurde dabei in eine entkernte Eizelle transplantiert. Faktoren im Zytosol der Oozyte setzten das Epigenom der somatischen Zelle in den Embryonalzustand zurück, wodurch sich wieder ein normaler Frosch entwickeln konnte.^[12] Dieser Vorgang wurde als somatischer Kerntransfer (engl. *Somatic Cell Nuclear Transfer*; SCNT) oder auch „Klonen“ bezeichnet. Ein weiteres bekanntes Beispiel der Reprogrammierung mittels SCNT ist das Klonschaf Dolly.^[13] Aus diesen Arbeiten ging die Hypothese hervor, dass auch *in vitro* die Möglichkeit besteht, den Differenzierungszustand von Zellen durch gezielte Gabe bestimmter Faktoren zu verändern und diese in einen stammzellähnlichen Zustand zu reprogrammieren. Dies gelang im Jahr 2006 *Shinya Yamanaka* unter Zuhilfenahme der vier Transkriptionsfaktoren Oct4, c-Myc, Klf4 und Sox2. In diesem Experiment wurden erstmals somatische Zellen in der Kulturschale in den Zustand embryonaler Stammzellen (engl. *Embryonic Stem Cells*; ESCs) zurückversetzt.^[14] Diese Zellen werden als induzierte pluripotente Stammzellen (engl. *Induced Pluripotent Stem Cells*; iPSCs) bezeichnet. Diese Zellen sind wiederum in der Lage in verschiedene Zelltypen zu differenzieren und bieten einen vielversprechenden Ansatz für die personalisierte und regenerative Medizin. Weiterhin wurde untersucht, ob ein Differenzierungszustand direkt in einen anderen überführt werden kann (dem Modell von *Waddington* nach, ist dies die Überquerung von Anhöhen; siehe Abbildung 3). Dieser Vorgang wird als direkte Reprogrammierung oder Transdifferenzierung bezeichnet.^[15-17]

Inwieweit transdifferenzierte Zellen für therapeutische Zwecke geeignet sind, ist Gegenstand aktueller Forschung. Epigenetische Anomalien und Fehlregulationen werden häufig bei neurodegenerativen Krankheiten, Krebs und Autoimmunerkrankungen beobachtet.^[4] Die Zellsysteme der iPSCs und ESCs werden derzeit intensiv als Krankheitsmodelle, zum Test neuer Arzneimittel, für toxikologische Tests und zur Zelltherapie genutzt. In Deutschland ist die Forschung an humanen embryonalen Stammzellen und Embryonen durch das Embryonenschutzgesetz und das Stammzellgesetz streng reguliert.^[18] Aus diesem Grund wird häufig auf den Organismus Maus als Modell zurückgegriffen. Durch die Generierung menschlicher iPSCs werden die ethischen Probleme beim Einsatz menschlicher Embryonen zur Generierung pluripotenter Modelle und differenzierter Derivate überwunden. Dies ermöglicht, frühe menschliche Entwicklungsübergänge und deren epigenetische Regulation zu untersuchen und größere Mengen an menschlichem differenziertem Gewebe ohne invasive Biopsieverfahren für Forschungszwecke zu erzeugen.

1.2 Stammzellen als Modell der frühen Entwicklung

Der Begriff Stammzelle wurde erstmals durch den deutschen Biologen *Ernst Haeckel* in der wissenschaftlichen Literatur eingeführt. Er beschrieb darin den einzelligen Vorfahren, aus dem sich alle vielzelligen Organismen entwickeln können.^[19] Diesen Zellen wurden seither die besonderen Eigenschaften zugeschrieben, sich nahezu unbegrenzt selbst erneuern und zu spezialisierten Zellen ausdifferenzieren zu können.^[20] Diese Vorläufer, die Stammzellen, gelten seither als einzigartiges Modell zur genetischen Manipulation von zellulären Prozessen und bieten unvorhergesehene Möglichkeiten für therapeutische Ansätze und die Simulierung von Krankheiten.

Stammzellen werden in zwei Arten unterteilt: Embryonale und somatische bzw. adulte Stammzellen. Embryonale Stammzellen haben die intrinsische und flexible zelluläre Eigenschaft, alle Zelltypen eines ausgewachsenen Organismus bilden zu können.^[21] Sie haben jedoch nicht die Fähigkeit eigenständig einen gesamten Organismus (inklusive der extraembryonalen Gewebe) zu bilden. Dies wird als totipotent bezeichnet und obliegt nur der Zygote und den frühen Blastomeren.^[22] Die bemerkenswerte Fähigkeit der ESCs, die drei Keimblätter Ektoderm, Entoderm und Mesoderm zu generieren, sowie Zellen der Keimbahn zu erzeugen, wird als Pluripotenz bezeichnet.^[23] Pluripotente Zellen sind lediglich in den frühen Embryonen zu finden. Sie existieren nur in einem sehr kurzen Zeitfenster. Zellen mit Stammzeleigenschaften, aber einem eingeschränkten Selbsterneuerungspotential werden als Vorläuferzellen (engl. *progenitor cells*) bezeichnet.^[18] Adulte Stammzellen werden aus den regenerativen Geweben ausgewachsener Organismen gewonnen. Sie haben die Fähigkeiten der Selbsterneuerung und der Differenzierung. Sie haben jedoch die Limitierung nur Zelltypen des Gewebes, dem sie entstammen, ausbilden zu können und werden daher als multipotent bezeichnet.^[18]

1.2.1 Gewinnung der pluripotenten Stammzellen

Die Studien an pluripotenten Stammzellen begannen mit der Entdeckung einer Mauslinie, die nach subkutaner Injektion spontan ein Karzinom ausbilden konnte, das Zellen aller Keimblätter, sowie undifferenzierte proliferative Zellen enthielt.^[24] Diese mehrfach ausdifferenzierten Teratokarzinome enthalten primitive, wachsende Zellen, die daraufhin embryonale Karzinomzellen (engl. *Embryonic Carcinoma Cells*; ECCs) genannt wurden.^[25] Die bereits für die ECCs optimierten Kulturbedingungen erlaubten in der Folge auch die Inkulturnahme pluripotenter embryonaler Zellen aus Präimplantationsembryonen.^[26, 27] Präimplantationsembryonen entwickeln sich aus der Zygote und diese entsteht durch Befruchtung der Oozyte. Im Gegensatz zu den ECCs sind diese ESCs genetisch normal. Sie haben die Fähigkeit, zur Generierung von Mäusen mit dem Erbmateriale unterschiedlicher Individuen (murine Chimäre) ohne, dass sie die Entwicklung von Tumoren einleiten.^[28]

Im Laufe der Embryonalentwicklung durchläuft die Zygote mehrere aufeinanderfolgende Zellteilungen, wie dem 2, 4, 8 und 16 Zell-Stadium (siehe Abbildung 4). Am Tag 3,5 der Embryonalentwicklung der Maus entsteht die Blastozyste. Diese beinhaltet die innere Zellmasse (engl. *inner cell mass*; ICM) und wird vom Trophoblasten umschlossen, der die extraembryonalen Gewebe wie die Plazenta ausbildet.^[29] In diesem Stadium der Entwicklung konnten erstmals pluripotente ESCs durch Entnahme von Zellen der ICM in Kultur gebracht werden.^[26, 27] Murine ESCs wurden zunächst auf einer Schicht von mitotisch inaktivierten Fibroblasten, den sog. *Feeder*-Zellen, in Medium mit fötalem Kälberserum kultiviert.^[30] Diese Kulturbedingungen waren zunächst nur auf Stammzellen eines Mäusestamms anwendbar.^[26, 27] Als jedoch wenig später entdeckt wurde, dass die *Feeder*-Zellen durch Zugabe des Zytokins *Leukemia Inhibitory Factor* (LIF) ersetzt werden konnten, wurden auch andere Mäusestämme für die Generierung von Stammzellen zugänglich.^[31-33] Die Inhibition zweier Signalkaskaden (siehe Abschnitt 1.2.2) ermöglichte schließlich die Gewinnung homogener Zellkulturen und durchbrach die Artenbarriere.^[34, 35] So konnten einige Zeit später auch pluripotente Stammzellen aus der Nagetierspezies Ratte gewonnen werden.^[36, 37]

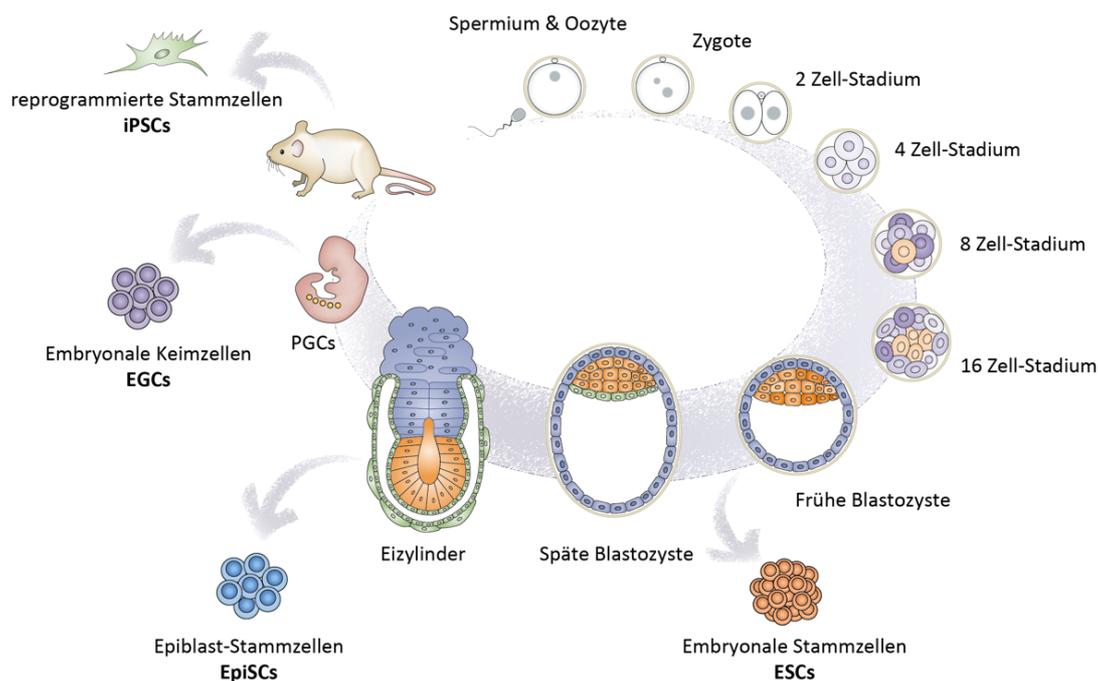


Abbildung 4: Möglichkeiten der Gewinnung pluripotenter Stammzellen. Abgebildet sind die unterschiedlichen Stadien während der Embryonalentwicklung sowie die Stammzellen die aus den unterschiedlichen Stadien in Kultur gebracht werden können. Die Entwicklung einer Maus beginnt mit der Befruchtung der Eizelle. Diese entwickelt sich weiter zur Zygote, dem 2, 4, 8 und 16 Zell-Stadium. Embryonale Stammzellen können der inneren Zellmasse des Blastozysten entnommen werden, Epiblast-Stammzellen dem Eizylinder, embryonale Keimzellen aus den primordialen Keimzellen des Fötus und reprogrammierte Stammzellen aus somatischen Zellen des ausgewachsenen Tieres.^[38, 39] ESCs: *Embryonic Stem Cells*, EpiSCs: *Epiblast Stem Cells*, EGCs: *Embryonic Germ Cells*, PGCs: *Primordial Germ Cells*, iPSCs: *Induced Pluripotent Stem Cells*.

Neben ESCs gibt es auch murine Epiblast-Stammzellen (engl. *Epiblast Stem Cells*; EpiSCs). Diese werden aus dem Postimplantationsembryo (Postimplantationsepiblast) entnommen und unterscheiden sich in ihrem Transkriptom sowie der Kulturbedingung stark von den ESCs.^[40-42] Humane embryonale Stammzellen, die ebenfalls mittlerweile in Kultur gehalten werden können, ähneln vorwiegend den EpiSCs.^[43, 44] Auch aus den Urkeimzellen der Embryonen, den primordialen Keimzellen (engl. *Primordial Germ Cells*; PGCs), können pluripotente Zellen gewonnen werden, die man als embryonale Keimzellen (engl. *Embryonic Germ Cells*; EGCs) bezeichnet.^[45, 46] Durch Entnahme muriner Gewebezellen und anschließender Reprogrammierung konnten murine iPSCs in Kultur gebracht werden.^[14] Kurze Zeit später gelang auch die Gewinnung humaner iPSCs aus menschlichen Fibroblasten.^[47, 48]

1.2.2 Naïve und *primed* Pluripotenz

Murine ESCs werden aus der ICM von Embryonen im Präimplantationsstadium gewonnen. Pluripotente Zellen dieses Stadiums (Embryonalentwicklungstage 3,5-4,5) zeichnen sich durch besondere Eigenschaften aus und werden als naïve bezeichnet.^[49] In Kultur bilden die naïven ESCs kompakte, kuppelförmige Kolonien (siehe Abbildung 5).

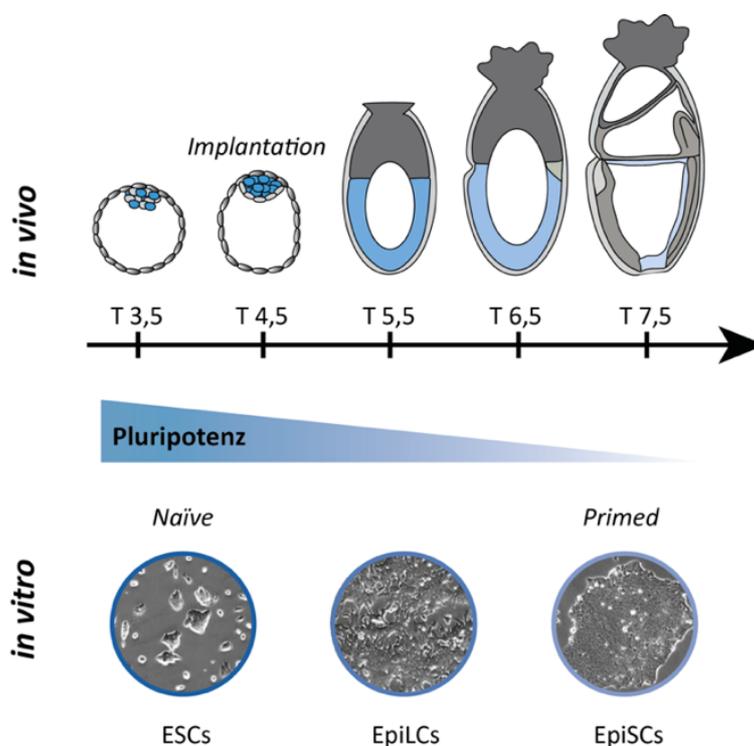


Abbildung 5: Abnahme der Pluripotenz im Verlauf der embryonalen Entwicklung und zugehörige *in vitro* Modelle. Die Abbildung zeigt den Verlust der Pluripotenz (blau) während der Embryonalentwicklung des murinen Embryos von Tag (T) 3,5-7,5 sowie mikroskopische Aufnahmen der entsprechenden *in vitro* Modelle. Die naïve Pluripotenz entspricht dem Präimplantations- und die *primed* Pluripotenz dem Postimplantationsembryo.^[23] (Abbildung nach *Morgani et al.*) ESCs (*Embryonic Stem cells*), EpiLCs (*Epiblast-Like Cells*), EpiSCs (*Epiblast Stem Cells*).

Näive ESCs exprimieren hohe Level der Pluripotenzfaktoren Oct4, Nanog und Sox2.^[50] Weiterhin zeichnen sie sich durch die Fähigkeiten aus, nach Blastozysteninjektion die ICM zu kolonisieren und somit zu somatischen Geweben und Keimzellen in Chimären beizutragen, sowie in weiblichen ESCs zwei aktive X-Chromosomen zu besitzen.^[50] Einige Tage nach der Implantation verlassen die Epiblast-Zellen allmählich das pluripotente Stadium und spezialisieren sich für den Eintritt in die Keimblätter-Entwicklung. Pluripotente Zellen mit den Merkmalen der Postimplantationsepiblasten (Embryonalentwicklungstage 6,5-8,5) werden als *primed* bezeichnet.^[50] Diese zur Differenzierung neigenden pluripotenten Zellen unterscheiden sich morphologisch wie auch in einigen anderen Eigenschaften von den naiven ESCs (Abbildung 5).^[51] Sie tragen nur wenig zur Chimärenbildung bei, leisten keinen Beitrag zur Bildung von Keimzellen und haben geringe Level des Pluripotenzfaktors Nanog. Weibliche Embryonen besitzen nur noch ein aktives X-Chromosom.^[44, 50] Die Pluripotenz geht am achten Entwicklungstag aufgrund der zunehmenden Spezialisierung für den Eintritt in die Keimblätter verloren.^[23]

Obwohl die Pluripotenz im Laufe der Embryonalentwicklung nur für einen kurzen Zeitraum besteht, können pluripotente Zellen verschiedener Entwicklungsstadien in Kultur eingefroren werden. Der Zusatz exogener Faktoren ermöglicht zudem eine unbegrenzte Kultivierung der Zelllinien in einem künstlich induzierten Selbsterneuerungszustand.^[51, 52] *In vitro* wird die Balance zwischen Pluripotenz und Differenzierung unter Standardbedingungen mit fötalem Kälberserum und dem Zytokin LIF (Serum/LIF) aufrechterhalten.^[23] Unter diesen Bedingungen entstehen heterogene Kulturen, bestehend aus Subpopulationen mit partieller Differenzierung, wobei die Selbsterneuerung der Zellen meist überwiegt und eine vollständige Differenzierung verhindert.^[23] Die zentrale Komponente des Serums ist das *Bone Morphogenic Protein 4* (BMP4), welches Signalkaskaden anregt, die eine Differenzierung der Zellen verhindern. LIF unterstützt die Selbsterneuerung der ESCs, indem es den Transkriptionsfaktor STAT3 (*Signal Transducer and Activator of Transcription 3*) aktiviert und dadurch pluripotenzfördernde Gene reguliert (siehe Abbildung 6).^[53] Homogenere Kulturen werden durch Zusatz von Inhibitoren erzielt.^[34] Die Zugabe des MEK-Inhibitors PD0325901 (PD) und des GSK3-Inhibitors CHIR99021 (CHIR; zusammen als 2i bezeichnet) reguliert Signalkaskaden von Wachstumsfaktoren, die für den Erhalt der Pluripotenz notwendig sind.^[34, 54] PD blockiert den FGF-Signalweg und verhindert die Differenzierung, während CHIR Einfluss auf den WNT-Signalweg nimmt und über Stabilisierung von β -Catenin die Expression pluripotenz-assoziiierter Gene fördert (siehe Abbildung 6). Zur Kultur der Stammzellen können sowohl Serum/LIF als auch 2i/LIF mit und ohne Serum verwendet werden.^[55] Unter 2i-Bedingungen wird LIF nicht zwingend benötigt. Das Zytokin wirkt sich aber positiv auf das Wachstum der Stammzellen aus.^[56-58] Je nach Kulturbedingung ähneln die Zellen einem anderen Abschnitt der embryonalen Entwicklung.

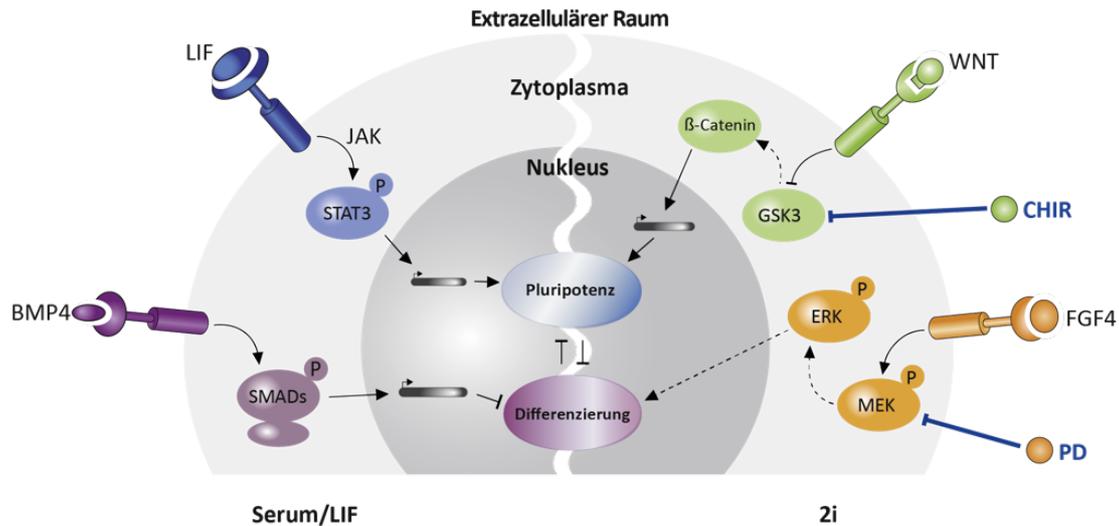


Abbildung 6: Signalwege zur Stabilisierung der Pluripotenz in Stammzellen. Vereinfachtes Schema verschiedener aktiver Signalkaskaden zur Kultur von ESCs. BMP4 aus Serum aktiviert SMADs und verhindert über Anregung bestimmter Gene die Differenzierung. LIF nimmt Einfluss auf die JAK-vermittelte Phosphorylierung von STAT3 und stimuliert das Pluripotenz-Netzwerk. CHIR99021 (CHIR) ahmt WNT-Signalisierung durch Inhibition von GSK3 nach und stabilisiert so β -Catenin. Eine Differenzierung wird ferner durch den MEK-Inhibitor PD0325901 (PD) verhindert, der die FGF-Signalkaskade inhibiert. Pluripotenz und Differenzierung sind in Balance.^[55] (Abbildung nach Hackett et al.) BMP4 (Bone Morphogenetic Protein 4), ERK (Extracellulär signal-regulierte Kinase), FGF4 (Fibroblast Growth Factor 4), GSK3 (Glycogen Synthase Kinase 3), JAK (Janus Kinase), LIF (Leukemia Inhibitory Factor), MEK (Mitogen-aktiviert Protein Kinase Kinase), STAT3 (Signal Transducer and Activator of Transcription 3), 2i (two inhibitors).

Unter 2i/LIF kultivierte naive ESCs entsprechen dem Präimplantationsepiblast. Eine weitere Möglichkeit ESCs zu kultivieren stellt die *alternative* 2i (a2i)-Bedingung dar. Das Transkriptom der mESCs unter diesen Kulturbedingungen ähnelt dem naiven Zustand. Der Methylierungsgrad ist jedoch weiter fortgeschritten.^[59] Für die a2i/LIF-Bedingung werden CHIR und ein Inhibitor des SRC-Signalweges, CGP77675, verwendet.^[60] Der SRC Inhibitor verhindert die Differenzierung und erhält die Selbsterneuerung sowie epigenetische Muster.^[59, 61] Werden diese Inhibitoren weggelassen, so entwickeln sich *primed* ESCs ähnlich des Postimplantationsepiblasten. So entsprechen Serum/LIF Kulturen den Epiblast-Zellen der Entwicklungsstadien 4,5-5,5.^[49, 59, 62] Eine weitere Möglichkeit zur Kultur von *primed* ESCs ist die CHIR/IWR1 Kulturbedingung. Diese wurde zunächst zur Kultur von EpiSCs verwendet und beinhaltet CHIR und den niedermolekularen Inhibitor IWR1.^[63, 64] IWR1 stabilisiert β -Catenin im Zytosol und wirkt sich positiv auf die Selbsterneuerung aus.^[63] Diese Bedingung stellt eine neue Methode zur Kultur von *primed* EpiSCs dar und kann auch für ESCs verwendet werden.^[65] Die bekannteste Kulturbedingung der EpiSCs ist die Zugabe der Zytokine FGF2 und Activin A in Abwesenheit von Serum.^[40, 41] Weiterhin ist es möglich EpiSCs aus ESCs zu gewinnen und ESCs aus EpiSCs zu reprogrammieren. Letzteres gelingt jedoch sehr ineffizient.^[66]

Ferner konnte festgestellt werden, dass ESCs unter kurzzeitiger (48 h) Zugabe von FGF2 und Activin A Epiblast-ähnliche Zellen (*Epiblast-Like Cells*; EpiLCs) ausbilden.^[23] Dieser Zustand liegt zwischen naïve und *primed* (siehe Abbildung 5), da das naïve, transkriptionelle Programm runterreguliert ist, aber die ersten Marker der Keimblätter bereits vorliegen.^[21, 67]

Die Entwicklung vom Präimplantations- bis zum Postimplantationsepiblast unterliegt zahlreichen streng kontrollierten epigenetischen Veränderungen.^[44] Dennoch können naïve und *primed* ESCs durch Änderung der jeweiligen Kulturbedingung ineinander überführt und so die epigenetische Umwandlung während der embryonalen Entwicklung *in vitro* nachgeahmt werden.

1.3 Epigenetische Regulation

Die genomische DNA der Maus enthält etwa 2,5 Milliarden Basenpaare.^[68] Um diese riesige Menge an genetischer Information im Zellkern zu speichern, liegt das genetische Material in Form von Chromatin vor. Die kleinste regulatorische Basiseinheit des Chromatins bildet das Nukleosom, bestehend aus DNA und Histon-Proteinen. Der DNA-Doppelstrang ist dabei mit einer Sequenzlänge von 146 bp um ein Histon-Oktamer gewunden, das aus jeweils zwei Heterodimeren der Histon-Monomere H2A, H2B, H3 und H4, besteht.^[69] Durch Zusammenlagerung der Nukleosomen bzw. des Chromatins zu einer höhergeordneten Struktur wird schließlich bei Bedarf das Chromosom ausgebildet. Sowohl die DNA als auch die Histon-Proteine der Nukleosomen sind mit besonderen Modifikationen versehen. Das Modifikationsmuster ist dabei streng regulierten Änderungen unterworfen und beeinflusst Histon/Histon- und DNA/Histon-Interaktionen sowie die Wechselwirkung mit benachbarten Nukleosomen.^[70] Dies hat einen direkten Einfluss auf die Packungsdichte der DNA und steuert damit die Expression spezifischer Genabschnitte. DNA- und Histonmodifikationen sind daher für die zelltypspezifische Regulation der Genexpression von großer Bedeutung.^[10]

1.3.1 Chromatinstruktur und Histonmodifikationen

Höhergeordnete Chromatinstrukturen und *Remodelling*-Komplexe sorgen für ein dynamisches Gleichgewicht zwischen einer gelockerten Form des Chromatins, dem Euchromatin, und einer kondensierten Form, dem sogenannten Heterochromatin. Dabei assoziiert man Euchromatin mit einer Aktivierung der Genexpression und Heterochromatin mit einem transkriptionell inaktiven Zustand. Die Genregulation kann weiterhin durch dreidimensionale Strukturen wie intra- oder interchromosomale Faltungen oder Schleifen beeinflusst werden.^[10] Dicht interagierende Chromatin-Bereiche, wie die *Topologically Associated Domains* (TADs), sind beispielsweise für die Organisation des Chromatins und die Weitergabe der epigenetischen Information von Bedeutung.^[71]

Die Bindung von Proteinen oder die Änderung der Chromatinstruktur durch Histon-Modifikationen beeinflusst die Genexpression zusätzlich. Zu diesem Zweck werden C- und N-terminale Fortsätze der Histone posttranslational reversibel und kovalent modifiziert. Beispiele dieser Proteinmodifikationen sind unter anderem die Acetylierung, Methylierung, Phosphorylierung, Ubiquitinylierung oder Sumoylierung der Histon-Fortsätze.^[70] Histone weisen einen hohen Anteil basischer Aminosäuren wie Lysin (K) und Arginin auf und sind unter physiologischen Bedingungen positiv geladen. Auf diese Weise wird eine elektrostatische Wechselwirkung mit dem negativ geladenen Phosphat-Rückgrat der DNA gewährleistet. Posttranslationale Modifikationen können die Wechselwirkung zwischen der DNA und den Histonen entscheidend beeinflussen. Die Einbringung der Modifikationen erfolgt durch spezifische, Histon-modifizierende Proteine, wobei zwischen *Reader*-, *Writer*- und *Eraser*-Proteinen unterschieden wird.^[10] Der Einfluss von Histonmodifikationen auf die Genregulation soll hier anhand eines *Writer/Reader/Eraser* Beispiels exemplarisch gezeigt werden: *Writer* sind beispielsweise Histonacetyltransferasen (HATs). HATs können Acetylgruppen auf Lysine im Histon-Fortsatz übertragen. Durch die Acetylgruppe wird die positive Nettoladung des Histons neutralisiert und die Bindung der DNA zum Nukleosom geschwächt, wodurch mehr Zugriff auf die DNA ermöglicht wird.^[72] Acetylierte Seitenketten werden wiederum von *Reader*-Proteinen über deren spezifische Erkennungsdomäne, die Bromodomäne, gebunden. Als Gegenspieler gelten *Eraser*-Proteine, wie Histondeacetylasen (HDACs), welche die Acetylgruppe wieder entfernen.^[73] Hierdurch kommt es zu einer transkriptionell inaktiven und kondensierten Form des Chromatins (siehe Abbildung 7).

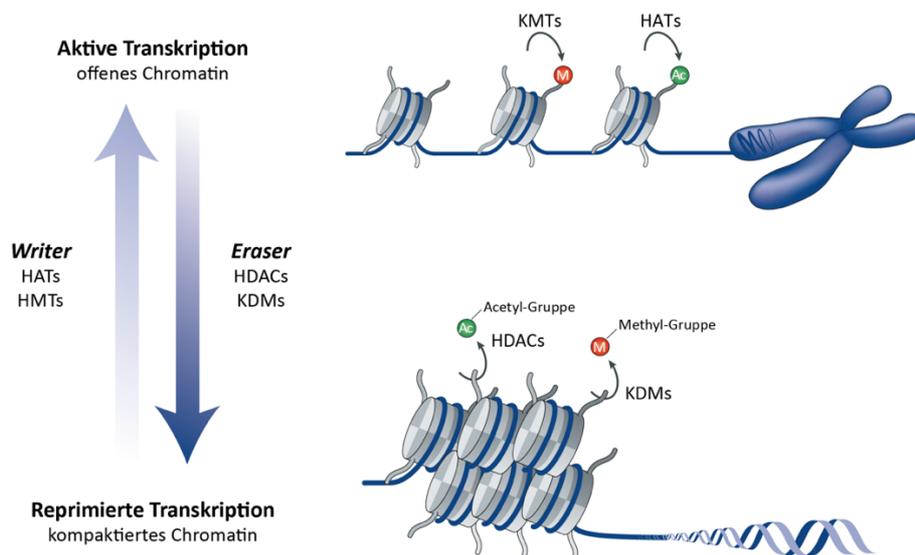


Abbildung 7: *Writer* und *Eraser* epigenetischer Markierungen. Abhängig vom Status der Transkription existiert Chromatin im offenen (Euchromatin) oder kompaktierten (Heterochromatin) Zustand. Acetylierung und Methylierung der Histon-Fortsätze durch Histonacetyltransferasen (HATs) bzw. Lysin-Methyltransferasen (KMTs) führen zu offenem Chromatin, was die Transkription erleichtert. Histon-Deacetylasen (HDACs) und Lysin-Demethylasen (KDMs) entfernen die Acetyl- bzw. Methyl-Gruppe und inaktivieren die Genexpression.^[10, 74]

Eine weitere Histon-Modifikation stellt die Methylierung dar. Diese wird von Lysin-Methyltransferasen eingeführt (KMTs) und durch Lysin-Demethylasen (KDMs) wieder entfernt. Die Methylierung kann sowohl aktivierend als auch inaktivierend wirken und wird ebenfalls von *Reader*-Proteinen erkannt, die wiederum weitere Proteinkomplexe rekrutieren.^[75] Zur Erkennung methylierter Histon-Seitenketten besitzen die *Reader*-Proteine Chromo- oder Tudordomänen.^[76] Auf diese Weise wird über Histon-Modifikationen die Interaktion von Proteinen mit der genomischen DNA reguliert und Transkriptionsfaktoren rekrutiert. Das Modifikationsmuster gibt Hinweise auf die Packungsdichte der DNA. So sind bestimmte Histon-Modifikationen Merkmale von Hetero- bzw. Euchromatin. Beispielsweise wird die Trimethylierung des Histon 3 am K27 (H3K27me3) oder die Histon 3 K9 Dimethylierung (H3K9me2) in stillgelegten Genbereichen mit hoher DNA-Kondensation beobachtet.^[77, 78] Dagegen wird die Histon 3 K4 Trimethylierung (H3K4me3) mit aktiven genomischen Bereichen assoziiert. Interessanterweise besitzen pluripotente Zellen große Mengen der transkriptionell aktivierenden Histonmodifikationen H3K4me3. Die lose gebundenen Histone ermöglichen einen erleichterten Zugang für Regulationsfaktoren und erlauben eine dynamische Änderung des Aktivitätszustandes.^[79] Dies ermöglicht den ESCs in multiple Zelllinien zu differenzieren. Das Chromatin differenzierter Zellen ist dagegen verdichtet und weist hohe Level an H3K27me3 auf.^[80]

1.3.2 Nicht-kodierende RNA

Eine weitere Art epigenetischer Regulatoren sind *long noncoding RNAs* (lncRNA). Diese sind 200 bp lange, nichtkodierende RNA-Stränge, also nicht zu Proteinen translatierte RNA-Stränge. Es wird zwischen vier Arten unterschieden: *Signal*, *Decoy*, *Guide* und *Scaffold*. Die *Signal* und *Decoy* lncRNAs sind an der An- und Abschaltung der Transkription beteiligt. *Guides* rekrutieren Chromatin modifizierende Proteine zur Regulation der Genexpression und *Scaffolds* bilden einen Ribonukleoproteinkomplex, welcher überwiegend Chromatin- bzw. Histon-Markierungen abändert.^[81] Die lncRNAs sind beispielsweise an der Allel-spezifischen genomischen Prägung beteiligt. Veränderungen in diesem System werden mit der Bildung von Tumoren in Verbindung gebracht.^[82]

1.4 DNA-Methylierung

Neben den kanonischen DNA-Bausteinen 2'-Desoxyadenosin (dA), 2'-Desoxyguanosin (dG), 2'-Desoxythymidin (dT) und 2'-Desoxycytidin (dC) enthält das Genom zahlreicher Organismen methylierte DNA-Bausteine wie 5-Methyl-2'-desoxycytidin (mdC), 4-Methyl-2'-desoxycytidin (4mdC) und 6-Methyl-2'-desoxyadenosin (6mdA).^[83] Die Modifikation 6mdA kommt in höheren Eukaryoten in geringen Mengen vor, während sie in einzelligen Eukaryoten und Prokaryoten weit verbreitet ist.^[84] 4mdC wurde bisher nur im genetischen Material von Bakterien beschrieben.^[83, 85] Die häufigste Methylierung in Eukaryoten ist mit etwa 1 % aller Nukleobasen mdC. mdC wird seit der Entdeckung oftmals als die fünfte Base des genetischen Codes bezeichnet.^[86] Während die DNA-Modifikation im Genom von Pilzen bis hin zu den Wirbeltieren weit verbreitet ist, kann sie in bestimmten Hefen, Nematoden oder Fliegen nicht nachgewiesen werden.^[87] In der DNA-Doppelhelix paaren dC und dG unter Ausbildung von drei Wasserstoffbrückenbindungen, wobei die Methylierung der C5-Position nicht an dieser Interaktion beteiligt ist (siehe Abbildung 8A). Die Methylierung an der C5-Position des Cytosins wurde von der Natur nicht zufällig gewählt. Diese Position ist einzigartig, da eine Methylierung und potentiell weitere Modifikationen mit nur minimalem Einfluss auf die DNA-Struktur und die Basenpaarung toleriert werden.^[88] Die Methylierung an der C5-Position ist nicht nur für dC, sondern auch für dT bzw. 5-Methyl-2'-desoxyuridin bekannt und verhilft der Zelle dT vom spontanen dC Desaminierungsprodukt 2'-Desoxyuridin (dU) zu unterscheiden.^[89]

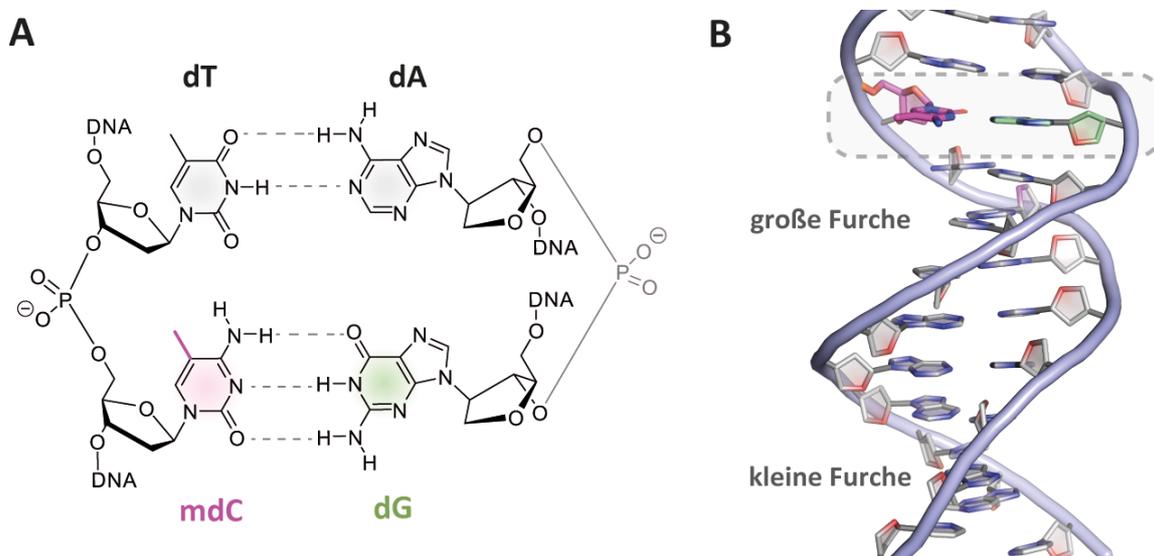


Abbildung 8: DNA-Struktur und räumliche Orientierung von mdC. **A)** Basenpaarungen in doppelsträngiger DNA. Drei bzw. zwei Wasserstoffbrückenbindungen werden zwischen mdC und dG bzw. dT und dA ausgebildet (gestrichelte Linien). **B)** Die Methylgruppe der DNA-Modifikation mdC ragt in die große Furche der Doppelhelix und beeinflusst dort die Wechselwirkung zwischen DNA und Protein. Die DNA-Struktur wurde mit *PyMOL* nach PDB:4GJU erstellt.^[90]

Aufgrund der exponierten Position der Methylgruppe in der großen Furche der DNA-Doppelhelix (siehe Abbildung 8B), wurde vermutet, dass dies Interaktionen mit Proteinen erleichtert. Tatsächlich werden interagierende Proteine durch mdC angezogen oder abgestoßen. Die DNA-Modifikation wird vorwiegend mit transkriptionell stillgelegtem Chromatin assoziiert. Durch Rekrutierung von Transkriptionsfaktoren kann sich DNA-Methylierung in besonderen Fällen auch aktivierend auswirken.^[91] Die DNA-Methylierung hat eine essentielle Bedeutung für die Kontrolle der Genexpression und damit die Säugetierentwicklung und gehört zu den wichtigsten epigenetischen Markern. In Säugetieren wird die DNA-Methylierung hauptsächlich im Kontext symmetrischer Cytosin-Phosphat-Guanin (CpG)-Dinukleotide auf beiden Strängen gefunden. Der CG-Gehalt im Säugetiergenom liegt bei etwa 42 %, wodurch CpG-Dinukleotide mit einer Häufigkeit von etwa 4,4 % erwartet werden würden. Tatsächlich treten die CpG-Dinukleotide nur mit etwa 20 % der erwarteten Häufigkeit auf.^[86] Eine Methylierung von CpH-Dinukleotiden (H= T, C oder A) in Säugetieren wurde bisher nur in Oozyten, ESCs und ausgereiften Neuronen festgestellt.^[92-94] Der Grund hierfür ist noch nicht vollständig geklärt, könnte aber zum Ausgleich des mdC-Mangels in CpG-armen Genombereichen dienen.^[95] Das Genom eines Menschen besitzt etwa 28 Millionen CpG-Stellen, von denen 60-80 % methyliert sind.^[96] Etwa 10 % der CpG-Dinukleotide liegen in Regionen mit höherer CpG-Dichte und werden als CpG-Inseln (CGIs) bezeichnet. Diese umfassen eine Länge von etwa 1 kb und liegen größtenteils hypomethyliert vor.^[97] CGIs sind häufig in Promotorregionen zu finden und sind an der Regulation der Transkription beteiligt.^[97] Die DNA-Methylierung dient in Säugetieren dem Erhalt zahlreicher zellulärer Funktionen und der Stabilität des Genoms und dient zur Rekrutierung von *Repressor*-Proteinen, die Bindestellen von Transkriptionsfaktoren blockieren und zur Stilllegung der Genexpression führen.^[96] So ist mdC an Prozessen, wie der X-Chromosomeninaktivierung, der Methylierung von regulativen Elementen und perizentromerischen Wiederholungen, der genomischen Prägung und der Stilllegung von Transposons beteiligt.^[87, 97-99] Ebenso wurde festgestellt, dass eine wachsende Zahl von menschlichen Krankheiten mit einer fehlerhaften DNA-Methylierung in Verbindung gebracht werden kann.^[100] Zusammen mit Histon- und Chromatin-Modifikationen bildet die DNA-Methylierung ein streng reguliertes Netzwerk zum Erhalt des Epigenoms.^[101]

1.4.1 DNA-Methyltransferasen

Die Methylierung wird durch die DNA-Methyltransferasen (Dnmts) eingebracht. Das humane Genom enthält fünf Dnmts mit großer Sequenzhomologie: hDNMT1, hDNMT2, hDNMT3A, hDNMT3B und hDNMT3L. Nur vier der DNA-Methyltransferasen Dnmt1, 2, 3a und 3b sind katalytisch aktiv.^[102] Dnmt1 ist im Allgemeinen für den Erhalt (engl. *maintenance*) der Methylierungsmuster nach der DNA-Replikation verantwortlich.

Das Enzym erkennt zusammen mit *Ubiquitin-like plant homeodomain and ring finger domain1* (Uhrf1) hemimethylierte CpG-Stellen und führt eine Methylierung im Gegenstrang ein. Die Vererbbarkeit des Methylierungsmusters gewährt die Weitergabe erworbener Eigenschaften an Nachkommen und fungiert damit als zentrales Mittel zu Übertragung der epigenetischen Information.^[38] Neue DNA-Methylierungsmuster werden durch die *de novo*-Methyltransferasen Dnmt3a und Dnmt3b eingebracht.^[103, 104] Deren Aktivität wird zudem durch das katalytisch inaktive Dnmt3l moduliert.^[105] Ein weiteres Mitglied der Familie, Dnmt2, wurde fälschlicherweise den DNA-Methyltransferasen zugeordnet. Es ist aber an der Methylierung von RNA beteiligt.^[106] Die Wichtigkeit der DNA-Methylierung wird bei der Betrachtung der von *Knockout*-Mäusen besonders deutlich. Deren Embryonen weisen deutliche Entwicklungsdefizite auf.^[104] Die Deletion von Dnmt1 führt zu einem genomweiten Verlust an DNA-Methylierung und zur Letalität am Embryonalentwicklungstag 9,5.^[87] Die Deletion von Dnmt3b ist letal, während Dnmt3a *Knockout*-Mäuse lebensfähig sind.^[104] Die Dnmt3l *Knockout*-Mäuse sind lebensfähig, besitzen jedoch keine *de novo*-Methylierung in der Keimbahn.^[107] In Nagetieren wurde kürzlich ein weitere *de novo*-Methyltransferase, Dnmt3c, entdeckt. Diese ist vermutlich an der Methylierung von Promotoren von Retrotransposons in der männlichen Keimbahn verantwortlich.^[108]

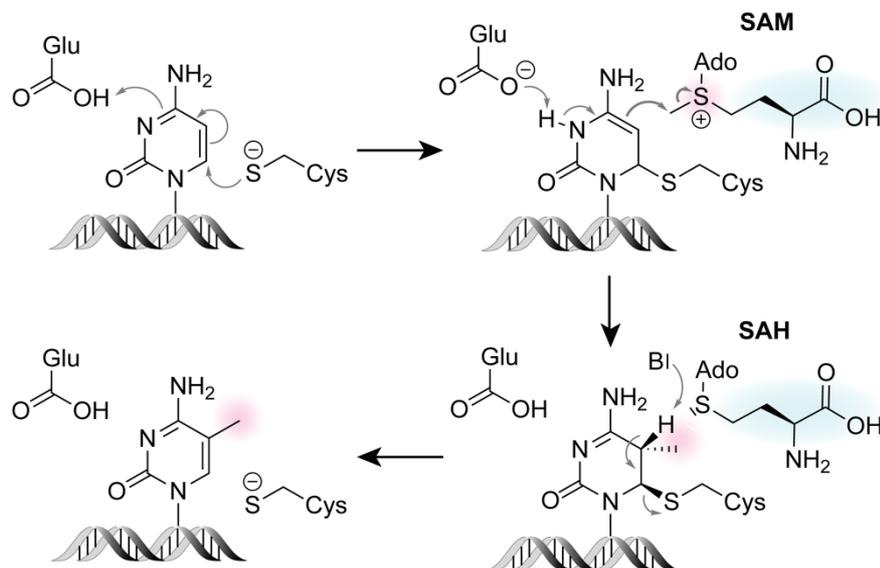


Abbildung 9: Katalytischer Mechanismus der DNA-Methyltransferasen. Ado: Adenosyl; B: Base; Cys: Cystein; Glu: Glutamin; SAM: S-Adenosylmethionin; SAH: S-Adosylhomocystein.

Alle DnmTs nutzen einen ähnlichen katalytischen Mechanismus und den Kofaktor S-Adenosylmethionin (SAM) als Methylgruppenspendor. Durch Herausdrehen der DNA-Base aus der DNA-Doppelhelix wird diese in die katalytische Tasche des Enzyms eingebracht.^[109] Die Methylierung wird durch einen nukleophilen Angriff an der C6-Position der Base durch eine konservierte Cystein-Seitenkette initiiert. Ein Glutamin-Rest protoniert die N3-Position.

Nach Bildung des Enamin-Intermediats erfolgt ein Angriff am Methylgruppendonor SAM. Nach Übertragung der Methylgruppe folgt eine Eliminierung des C5-Protons und Rearomatisierung der Nukleobase.^[109, 110] Der beschriebene Mechanismus zur Bildung von mdC durch die Dnmts ist in Abbildung 9 dargestellt.

1.4.2 Dynamik in der zellulären Entwicklung

In der frühen Entwicklung des Embryos unterliegt das DNA-Methylierungsmuster des gesamten Genoms dynamischen Änderungen. Nach der Befruchtung wird die DNA-Methylierung langsam zurückgesetzt und nach der Implantation wieder eingeführt (siehe Abschnitt 1.6.2). Während der Bildung der Urkeimzellen wird die Methylierung ebenfalls entfernt.^[111] Die mESCs gehören zu den am intensivsten genutzten Modellsystemen zur Untersuchung der epigenetischen Regulation. Die entscheidende Funktion der DNA-Methylierung für die Entwicklung der ESCs ist jedoch nicht vollständig verstanden. Eindeutig ist, dass die DNA-Methylierung nicht an der Aufrechterhaltung der Pluripotenz beteiligt ist.^[96] ESCs, in denen alle drei katalytisch aktiven DNA-Methyltransferasen ausgeschaltet sind, können unbegrenzt in Kultur gehalten werden und erfüllen alle Merkmale von Stammzellen. Diese Dnmt-defizienten Zellen zeigen jedoch eine beeinträchtigte Differenzierungsfähigkeit.^[112] Marker der Keimblätter können nicht hochreguliert und Pluripotenzfaktoren nicht abgeschaltet werden.^[113] ESCs, denen nur eine der drei DNA-Methyltransferasen fehlt, verlieren nach einiger Zeit die gesamte Methylierung, wenn auch unterschiedlich schnell.^[114] Einbringung von Dnmt1 in einen Dnmt1-Knockout stellt die DNA-Methylierung mit Ausnahme der *Imprinting*-Regionen wieder her.^[115]

Kürzlich wurde der Einfluss verschiedener ESC-Kulturbedingungen, Serum/LIF und 2i/LIF (siehe Abschnitt 1.2.2), auf die DNA-Methylierung untersucht. Dabei wurde festgestellt, dass es beim Übergang von Serum/LIF nach 2i/LIF-Bedingungen zu einer globalen DNA-Hypomethylierung kommt, die vergleichbar mit dem Methylierungsgrad der ICM ist.^[78, 116, 117] Der Verlust der DNA-Methylierung wird mit einer Abnahme von Dnmt3a, 3b und 3l in Verbindung gebracht.^[118] Es wird vermutet, dass Prdm14, ein Transkriptionsfaktor, unter 2i/LIF die Expression der DNA-Methyltransferasen herunterreguliert.^[118, 119] Interessanterweise reguliert Prdm14 wiederum einen Kofaktor, der zur Einbringung der Histon-Modifikation H3K9me2 verantwortlich ist.^[119, 120] Der Dnmt1 Kofaktor Uhrf1 bindet über seine Tudor-Domäne an H3K9me2.^[121] Die Histon-Modifikation liegt unter 2i/LIF-Bedingungen aber reduziert vor und könnte so die Rekrutierung von Uhrf1 bzw. Dnmt1 negativ beeinflussen.^[77, 78, 122] Auf welche Weise die DNA-Methylierung beim Übergang von Serum/LIF nach 2i/LIF tatsächlich entfernt wird, bedarf weiterer Untersuchungen. Die Erkenntnisse zeigen jedoch deutlich, dass die betroffenen epigenetischen Prozesse eng miteinander vernetzt sind.

1.5 Tet-vermittelte Bildung der oxidierten mdC-Derivate

Zur Entfernung des Methylierungsmusters steht die Säugetierzelle vor einer Herausforderung. Die chemische Stabilität der C-C Bindung zwischen der Methylgruppe und der C5-Position des Cytidins macht eine direkte enzymatische Spaltung sehr problematisch.^[123] Weiterhin ist in Säugetieren kein Enzym bekannt, welches mdC direkt entfernen kann.^[124] Aufgrund dessen wurde mdC lange Zeit als eine irreversible DNA-Modifikation angesehen. Zunächst wurde angenommen, dass die Modifikation durch Verdünnung im Zuge der DNA-Replikation aus dem Genom entfernt wird. Im Jahre 2000 wurde jedoch beobachtet, dass ein globaler Verlust an mdC in Zygoten von Mäusen in replikationsunabhängiger Weise stattfindet.^[125, 126] Obwohl das humane Protein TET1 bereits als anomaler Fusionspartner der H3K4-Methyltransferase MLL (*Mixed Lineage Leukemia*) in der akuten myeloische Leukämie (AML) beschrieben war, blieb die eigentliche Funktion des Proteins bis 2009 unentdeckt.^[127] Heute weiß man, dass die *Ten-eleven translocation* (Tet) Dioxygenasen zur Familie der Eisen und α -Ketoglutarat (α KG)-abhängigen Dioxygenasen gehören. Sie katalysieren die iterative Oxidation von mdC zu 5-Hydroxymethyl-2'-desoxycytidin (hmdC),^[128] 5-Formyl-2'-desoxycytidin (fdC)^[129, 130] und 5-Carboxyl-2'-desoxycytidin (cadC)^[130, 131] (siehe Abbildung 10).

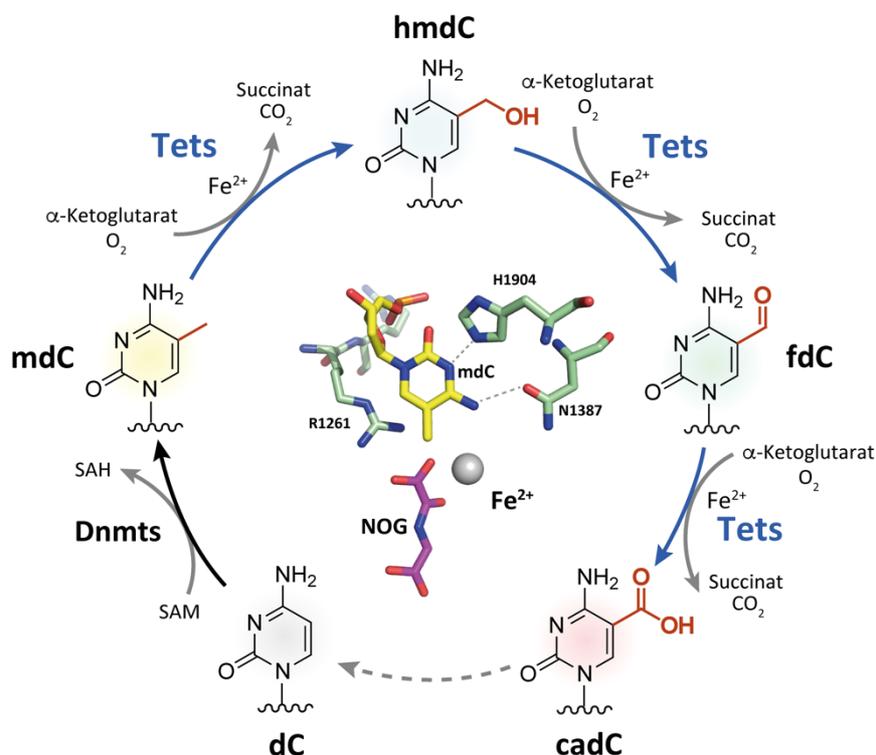


Abbildung 10: Entstehung der höher oxidierten mdC-Derivate durch Tet-Katalyse. Methylierung von dC durch die Dnmts und iterative Oxidation durch die Tet-Dioxygenasen. Es wird angenommen, dass dieser Weg schließlich zu unmodifiziertem dC zurückführt. Mittig: Aktives Zentrum des hTET2 mit den katalytisch wichtigen Resten, des Bindungspartners von mdC und dem Zentralmetall. Anstelle des natürlichen Cosubstrats α -Ketoglutarat wurde mit *N*-Oxalylglycin (NOG) co-kristallisiert. Die Abbildung wurde mit *PyMOL* nach PDB: 4NM6 erstellt.^[132]

Dabei zeigte sich, dass die gebildeten höher oxidierten mdC-Derivate eine wichtige Funktion als Intermediate der DNA-Demethylierung ausüben und an der epigenetischen Regulation der Genexpression beteiligt sind.^[133, 134] Im Säugetiergenom existieren drei Tet-Dioxygenasen Tet1, 2 sowie 3 und deren Isoformen.^[135] Diese bis zu 235 kDa großen Proteine besitzen eine C-terminale katalytische Domäne, die sich aus einer doppelsträngigen β -Helix (DSBH)-Domäne und einer Cysteinreichen Region zusammensetzt.^[128, 136] Die DSBH-Domäne beinhaltet dabei die Bindestelle für Fe^{2+} und αKG . Im aktiven Zentrum können die Tet-Enzyme einen Fe(IV)oxo -Komplex ausbilden, der aus einer Decarboxylierung von αKG zu Succinat resultiert. Durch Übertragung eines Sauerstoffatoms des Oxo-Komplexes auf die Methylgruppe von mdC entsteht hmdC, wobei das Fe(IV) zum Fe(II) reduziert wird.^[137] Tet1 und Tet3 besitzen weiterhin eine CXXC-Domäne am N-Terminus. Diese Domäne dient vielen Chromatin-assoziierten Proteinen als DNA-Bindungsmotiv unmethylierter CpG-Positionen.^[138] Das *Tet2*-Gen durchlief im Laufe der Evolution eine chromosomale Inversion, wobei die CXXC-Domäne verloren ging. Aus dieser Domäne entwickelte sich schließlich ein eigenständiges Protein, genannt IDAX (*Inhibition of the dcl and axin complex*).^[139] Es wird vermutet, dass IDAX mit Tet2 oder einer CXXC-losen Tet3 Isoform interagiert und so die Bindung verschiedener Tet-Isoformen mit der DNA reguliert.^[140] Kürzlich konnte die Kristallstruktur des humanen Tet2 im Komplex mit einer mdC enthaltenden DNA gelöst werden.^[132] Ähnlich der DNA-Methyltransferasen wird die Base aus der DNA herausgedreht (*Base Flipping*), um sie in der katalytischen Tasche des Enzyms zu positionieren. Zur katalytischen Umsetzung wird die Methylgruppe schließlich dem Cosubstrat αKG und dem Zentralmetall zugewandt. Bindung und Erkennung von mdC erfolgen über zwei Wasserstoffbrücken mit einer Histidin (H) und Arginin (R) Seitenkette in der katalytischen Tasche (siehe Abbildung 10).^[132, 141] Ferner wurde beobachtet, dass das Enzym keinen direkten Kontakt mit der Methylgruppe eingeht, wodurch genügend Platz für die weiteren Oxidationsprodukte fdC und cadC im aktiven Zentrum sichergestellt wird.^[132] Interessant ist diesem Zusammenhang auch, dass die Tet-Enzyme dT zu 5-Hydroxymethyl-2'-desoxyuridin (hmdU) zu oxidieren vermögen.^[142] Eine biologische Bedeutung des enzymatisch generierten hmdUs ist bisher noch nicht bekannt. Kürzlich entwickelte Methoden zur Sequenzierung von hmdU könnten einen großen Beitrag zur Aufklärung der Funktion und der Verteilung von hmdU im Genom liefern.^[143, 144]

1.5.1 5-Hydroxymethyl-2'-desoxycytidin

Die epigenetische Modifikation hmdC wurde erstmals 1952 als Bestandteil des Genoms geradzahlicher T-Bakteriophagen beschrieben.^[145] In Säugetieren wurde hmdC einige Jahre später ebenfalls entdeckt.^[146] Die Mengen waren allerdings fragwürdig hoch.^[146] Da diese Werte in einer weiterführenden Studie nicht bestätigt werden konnten,^[147] wurde hmdC zunächst nur für einen oxidativen Schaden von mdC gehalten.^[148]

Rao und Mitarbeiter konnten jedoch zeigen, dass hmdC durch enzymatische Oxidation aus mdC gebildet wird.^[128] Noch im selben Jahr konnten *Heintz und Kriaucionis* die Existenz der Modifikation in der genomischen DNA von Säugetieren bestätigen und die Level an hmdC genauer bestimmen. Die Studien ergaben einen hmdC-Gehalt von 0,1 % aller Cytidine in Stammzellen und einen Wert von 0,6 % pro dC in Purkinje-Neuronen.^[149] Unter Zuhilfenahme der *Liquid-Chromatography Tandem Mass Spectrometry* (LC-MS/MS) wurden die Level an hmdC im Gehirn und weiteren Organen der Maus bestimmt.^[150, 151] Die untersuchten Gewebe zeigten deutliche Unterschiede in ihren hmdC-Werten. Das humane Gehirn wies dabei mit 1,3 % die größten hmdC-Werte auf.^[152] Interessanterweise werden in Tumoren vorwiegend geringere Mengen an hmdC gefunden, während höhere Level an mdC vorliegen.^[153] Vermutlich verursacht der Sauerstoffmangel (Hypoxie) in Tumoren eine Reduktion der Tet-Aktivität, wobei weniger mdC zu hmdC oxidiert wird.^[154] Seit kurzem wird die Verteilung von hmdC auch als diagnostischer Biomarker für bestimmte Krebsarten genutzt.^[155] Während somatische Zellen konstante Werte an mdC von etwa 4-5 % aller Cytidine besitzen, wurde festgestellt, dass die Level an hmdC gewebespezifisch und altersabhängig variieren.^[151] Weiterhin konnte gezeigt werden, dass die Menge an hmdC nicht von der Expression des jeweiligen Gens abhängig ist. So weist das gleiche Gen, trotz unterschiedlich starker Expression, in verschiedenen Geweben unterschiedliche Mengen an hmdC auf.^[152, 156] Studien zeigten ebenfalls eine inverse Korrelation zwischen dem hmdC-Gehalt und dem Grad des zellulären Wachstums.^[156] Dies legt die Vermutung nahe, dass hmdC nur in postmitotischen und ausdifferenzierten Zellen akkumuliert. Dagegen konnten *Bachman et al.* zeigen, dass hmdC nach der Replikation mit Verzögerung neu gebildet wird, was auf einen Erhaltungsmechanismus von hmdC hindeutet.^[157] Untersuchungen zur Verteilung von hmdC konnten zeigen, dass hmdC fast ausschließlich in CpG-Dinukleotiden auftritt.^[158] Hierbei handelt es sich meist um eine asymmetrische Hydroxymethylierung der CpG-Dinukleotide. Eine Hydroxymethylierung an beiden Strängen tritt nur in etwa 20 % der hmCpG-Stellen auf.^[158, 159] Die Modifikation wird in Promotoren mit mittlerer Expression und durchschnittlicher CpG-Dichte gefunden.^[136, 160] Während mdC typischerweise mit Promotoren stillgelegter Gene assoziiert wird, markiert hmdC meist aktiv transkribierte Gene.^[161] Inwieweit hmdC eine eigenständige epigenetische Funktion ausübt oder vorwiegend als Intermediat der aktiven Demethylierung fungiert, bedarf weiterer Studien.

1.5.2 5-Formyl-2'-desoxycytidin und 5-Carboxyl-2'-desoxycytidin

Die nicht-kanonischen DNA-Bausteine fdC und cadC entstehen durch Tet-vermittelte Oxidation ausgehend von mdC und könnten daher als Intermediate einer enzymvermittelten Demethylierung fungieren.^[129, 130] *Pfaffeneder et al.* konnten erstmals fdC mit einer Häufigkeit von 0,02 % pro dG im Genom von Stammzellen massenspektrometrisch nachweisen.^[129] Kurze Zeit später wurde auch cadC als Tet-Oxidationsprodukt im Genom detektiert.^[130]

Nur wenig ist bisher über cadC bekannt, welches mit etwa 0,002 % pro dG 10-mal seltener als fdC vorkommt.^[85, 162] *He et al.* berichteten noch im selben Jahr, dass der *Knockout* der Thymin-DNA-Glykosylase (Tdg) zu einem Anstieg der fdC und cadC-Menge führt.^[131] Eine weitere Studie zeigte, dass diese Glykosylase fdC und cadC erkennt und ausschneiden kann.^[163] Die beiden DNA-Modifikationen kommen wesentlich seltener vor als hmdC. Der Grund hierfür ist nicht ganz klar. Als Ursachen werden eine schnelle Entfernung durch Tdg und eine ineffiziente Umsetzung von hmdC zu fdC und cadC diskutiert.^[38] Im Einklang mit einer effizienten Entfernung durch Tdg konnten in *Tdg-Knockout* mESCs 5-6-mal höhere Werte an fdC bzw. 8-9-mal mehr cadC detektiert werden.^[164] Neben mESCs wurden fdC und cadC auch in Neuronen und verschiedenen Geweben detektiert.^[152, 165] Die beiden DNA-Modifikationen wurden ebenfalls in regulativen Elementen des Genoms gefunden. Darunter fallen *active Enhancer*, die der Initiation der Transkription durch Rekrutierung von Protein-Komplexen an Promotoren dienen und *poised Enhancer*, die zwar übergangsweise stillgelegt, aber bereit zur Reaktivierung sind.^[166] Unter Zuhilfenahme neuester Sequenzieretechniken konnte so eine Anreicherung von fdC und cadC an diesen Genompositionen in mESCs und in verschiedenen embryonalen Geweben der Maus gezeigt werden.^[164, 167] Neben der Funktion als Intermediate einer aktiven Entfernung von mdC, könnten fdC und cadC auch als eigenständige epigenetische Signale fungieren. Diese Annahme ist im Einklang mit Interaktionsstudien von Proteinen aus Stammzellen, neuronalen Vorläuferzellen und dem Gehirn. Es konnten unterschiedliche *Reader*-Proteine für die verschiedenen DNA-Modifikationen identifiziert werden, die eine Beteiligung der modifizierten DNA-Bausteine an der DNA-Reparatur, der Regulation der Transkription und dem Chromatin-*Remodelling* vermuten lassen.^[168, 169] Die eigenständige biologische Funktion wurde durch Isotopenverfolgungsexperimente von *Bachmann et al.* und in Studien von *Su et al.* ebenfalls bekräftigt.^[170] Sie zeigen, dass fdC, wie auch hmdC, eine teilweise stabile DNA-Modifikation ist.^[171] Weiterhin konnten Studien eine Destabilisierung des DNA-Duplexes durch fdC und cadC beobachten.^[166] *Raiber et al.* beschrieben, dass fdC Einfluss auf die DNA-Konformation nimmt.^[172] Allerdings berichtete im Gegensatz dazu eine weitere Gruppe, dass fdC nicht zu einer Änderung der Konformation führt.^[88] Die Unterschiede dieser Studien könnten auf der Dynamik beruhen, die durch die modifizierten DNA-Basen ausgelöst wird. So konnte gezeigt werden, dass fdC die Flexibilität der DNA erhöht und zu einer verbesserten mechanischen Stabilität der Nukleosomen führt. Die Flexibilität der DNA wird bereits durch ein fdC erhöht.^[173] Diese erhöhte Flexibilität sorgt für eine bessere Zugänglichkeit DNA-modifizierender Proteine. *In vitro* Studien konnten weiterhin Schiff'sche-Base-Konjugate zwischen fdC und Lysin-Seitenketten von Histon-Proteinen nachweisen.^[174, 175] *Raiber et al.* konnten kürzlich zeigen, dass fdC mit dem Histon H3 in mESCs eine kovalente Bindung eingeht und so die Organisation der Nukleosomen steuert.^[176] Es konnte beobachtet werden, dass die fdC Histon H3 Interaktion Einfluss auf die Elongationsrate der RNA-Polymerase II (RNAP II) nimmt.^[176]

Wang *et al.* postulierten zuvor, dass fdC und cadC durch die RNAP II erkannt werden und übergangsweise deren Aktivität verlangsamen.^[177] Die RNAP II erkennt bereits kleine Unterschiede an der C5-Position von dC. Während mdC und hmdC keine Änderungen zeigen, beeinträchtigen fdC und cadC die Elongationsrate der RNAP II.^[178] Dies könnte ein Zeitfenster zur Rekrutierung von Proteinen zur Regulation der Transkription schaffen und einen Schlüsselschritt zur Kontrolle der Genexpression durch modifizierte DNA-Bausteine darstellen.

1.6 DNA-Demethylierung

In den vergangenen Jahren konnte eindeutig bewiesen werden, dass die DNA-Demethylierung während der Präimplantationsphase sowohl über replikationsabhängige, als auch aktive replikationsunabhängige enzymatische Prozesse erfolgt. Allerdings konnte der relative Beitrag der beiden Demethylierungswege und die dabei ablaufenden Mechanismen noch nicht vollständig verstanden werden. Darüber hinaus ist nicht klar, auf welche Weise die Tet-Enzyme zur aktiven Demethylierung beitragen und ob der Beitrag der Dioxygenasen in den verschiedenen Entwicklungsphasen variiert. Ebenso könnten Tet-unabhängige aktive Mechanismen eine Rolle spielen.

1.6.1 Passive DNA-Demethylierung

Das Methylierungsmuster wird während der DNA-Replikation durch die *maintenance* Methyltransferase Dnmt1 und dem Hilfsprotein Uhrf1 aufrechterhalten.^[179, 180] Kommt es zu einer Beeinträchtigung der Erhaltungsmethylierung, wie beispielsweise beim Übergang von *primed* nach naive ESCs, erfolgt ein Verlust der Methylierung im Zuge der DNA-Replikation.^[122] Über mehrere Replikationszyklen hinweg, kommt es auf diese Weise zu einer fortschreitenden, breiten und unspezifischen Demethylierung. Demzufolge führt die Abwesenheit der Erhaltungsmethylierung oder eine Störung der Aktivität der Methyltransferasen zu einem replikationsabhängigen Verlust von mdC. Dieser Prozess wird als passive DNA-Demethylierung bezeichnet. Auch die Tet-Enzyme könnten an einer passiven Demethylierung beteiligt sein. Über die Tet-Aktivität werden die höher oxidierten dC-Spezies hmdC, fdC und cadC, ausgehend von mdC generiert (siehe Abschnitt 1.5). Sie könnten die Methylierung stören, was eine zielgerichtete passive Demethylierung ermöglichen würde. Es konnte in der Tat gezeigt werden, dass Dnmt1 CpG-Dinukleotide mit oxidierten dC-Derivaten anders prozessiert. Dnmt1 bevorzugt hemimethylierte über hmdC-, fdC- oder cadC-modifizierte CpG-Stellen.^[181-183] Dagegen haben Dnmt3a und Dnmt3b eine vergleichbare Aktivität für mdC und hmdC enthaltende CpG-Positionen,^[181, 182] wobei sich fdC sogar positiv auf die Aktivität von Dnmt3a auswirkt.^[183]

Aufgrund der veränderten Dnmt-Aktivität an hmdC, fdC oder cadC enthaltenden Positionen, könnte auf diese Weise eine Demethylierung im Zuge der DNA-Replikation trotz aktiver Erhaltungsmethylierung stattfinden.

1.6.2 Dynamik nicht-kanonischer DNA-Bausteine in der Embryonalentwicklung

Kurz nach der Befruchtung werden die murine, wie auch die humane Zygote einer epigenetischen Reprogrammierung unterzogen, die unter anderem eine globale Demethylierung des paternalen und maternalen Genoms beinhaltet.^[38] Bereits im Jahr 2000 wurde beobachtet, dass das mdC-Signal des väterlichen Genoms schneller abnimmt, als durch passive Verdünnung im Zuge der DNA-Replikation erklärbar.^[125, 126] Derzeit wird angenommen, dass dies durch eine Kombination aus replikationsabhängiger Verdünnung sowie Tet3-vermittelter Oxidation (aktiv) geschieht, wohingegen die Demethylierung des maternalen Genoms vorwiegend in einer replikationsabhängigen Weise erfolgt.^[38] Abbildung 11A zeigt schematisch die Dynamik der DNA-Methylierung und der Tet-Oxidationsprodukte in den frühen Entwicklungsstadien.

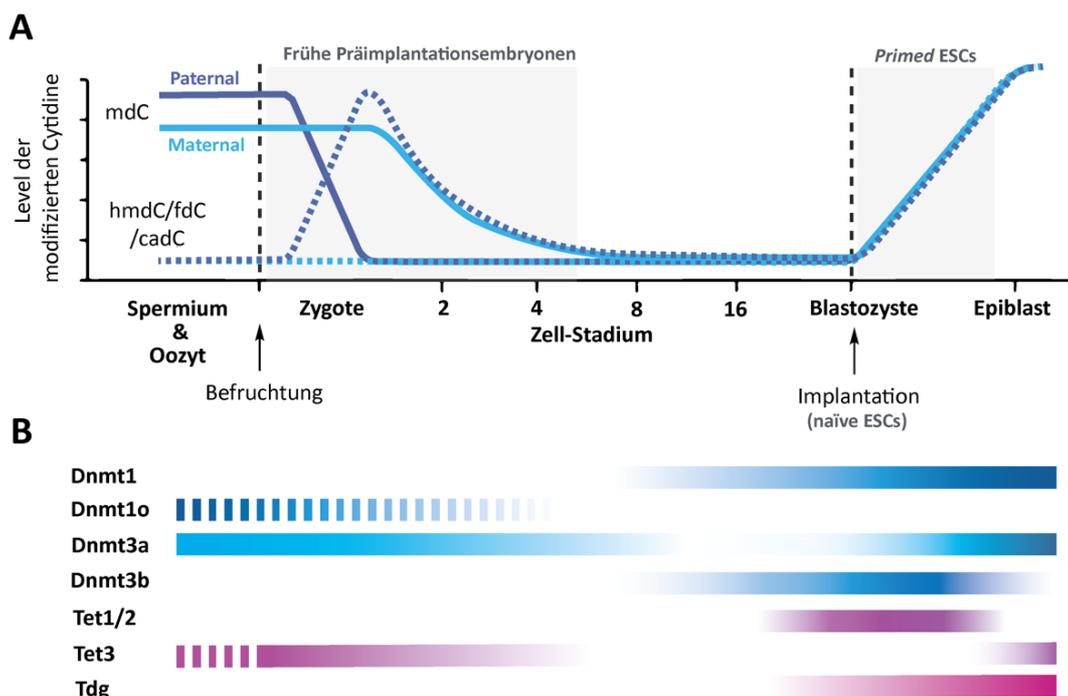


Abbildung 11: Dynamik der DNA-Methylierung vor der Implantation. A) In Präimplantationsembryonen werden das maternale (hellblaue Linie) und paternale Genom (dunkelblaue Linie) über passive und aktive Prozesse demethyliert. Der Verlauf der mdC-Oxidationsprodukte ist ebenfalls gezeigt (gestrichelte Linien). Im Stadium der Blastozyste wird die Methylierung zurückgewonnen. Dieser Prozess kann anhand der ESCs *in vitro* untersucht werden. **B)** Schematische Darstellung der Expressionslevel der beteiligten Enzyme. Die Abbildung wurde nach Wu *et. al* und Kohli *et. al* erstellt.^[184, 185]

Der schnelle Verlust der Methylierung im paternalen Genom wird von einer Generierung der DNA-Modifikationen hmdC, fdC und cadC begleitet,^[186-188] die anschließend replikationsabhängig aus dem Genom entfernt werden.^[187, 189, 190] Weiterhin werden im Stadium der Zygote sehr geringe Expressionslevel an *Tdg* mRNA detektiert (Abbildung 11B),^[184] weswegen eine Demodifizierung durch die Glykosylase unwahrscheinlich erscheint. Im Einklang damit ist auch die Detektion von fdC und cadC bis hin zum 4. Zell-Stadium.^[187] Interessanterweise sterben *Tdg* *Knockout*-Mäuse am Embryonalentwicklungstag 11,5. Die katalytische Aktivität von *Tdg* scheint daher erst zu einem späteren Zeitpunkt der Entwicklung essentiell zu sein.^[191] Bekannt ist, dass die Bildung der mdC-Oxidationsprodukte durch Tet3 katalysiert wird.^[186, 188, 192] Im maternalen Genom sind ebenfalls detektierbare Mengen der oxidierten mdC-Derivate vorzufinden. Allerdings sind die Level wesentlich geringer,^[186-188] werden aber ebenfalls durch eine Tet3-vermittelte Oxidation generiert.^[193, 194] Der Unterschied resultiert vermutlich aus einer Interaktion von Stella mit Tet3, was die enzymatische Aktivität der Dioxygenase inhibiert.^[195] Die globalen Level an mdC erreichen im Stadium der Blastozyste den niedrigsten Wert. Im korrespondierenden *in vitro* Modell, den naiven ESCs, liegt mit 1 % aller dC eine Hypomethylierung des Genoms vor.^[78] Erst nach Implantation wird das Methylierungsmuster wieder eingeführt, wobei die entsprechenden *primed* Kulturen Methylierungswerte von 4-5 % aller dC aufweisen.^[78, 116] *Amouroux et al.* konnten kürzlich zeigen, dass die Entfernung von Tet3 eine Akkumulation von hmdC verhindert, aber nicht den frühen Verlust des paternalen mdC abwendet.^[196] Ein alternativer Mechanismus der DNA-Demethylierung wurde daher angenommen. Im weiteren Verlauf dieser Studien wurden Hinweise erbracht, dass die Tet3-vermittelte Oxidation zur Sicherstellung der DNA-Demethylierung dienen könnte und auf diese Weise neu eingebrachtes mdC lokusspezifisch durch Oxidation zu hmdC entfernt.^[196] Eine positionsabhängige Oxidation könnte die DNA-Methylierung zum einen entfernen und zum anderen an definierten Positionen, wie *imprinted* Genen, erhalten.^[197-199] Weitere Studien sind notwendig um die ortsspezifischen Änderungen und essentiellen Faktoren dieses komplexen Prozesses hinreichend genau zu erfassen.

1.6.3 Aktive DNA-Demethylierung

Eine aktive Demethylierung wurde, wie im Abschnitt 1.6.2 beschrieben, bereits in der Zygote beobachtet.^[125, 188, 192] Ebenso wurde in postmitotischen Zellen des Gehirns und bei Änderungen des Zellschicksals eine lokusspezifische aktive Demethylierung gezeigt.^[200, 201] Auf welchem Weg die Demethylierung in diesen Systemen erfolgt, ist jedoch nicht vollständig geklärt. In Pflanzen können die DNA-Glykosylasen der *Repressor Of Silencing 1* (ROS1)/DEMETER (DME) Familie mdC direkt aus dem Genom ausschneiden und so kanonisches dC über die Basenexzisionsreparatur (BER) zurückgewinnen.^[202] In Säugetieren wurden derartige Enzyme nicht identifiziert, weswegen die Methylierung auf einem anderen Weg aus der genomischen DNA entfernt werden muss.

Seit der Entdeckung der Tet-Dioxygenasen rückten die höher oxidierten dC-Derivate, im Besonderen fdC und cadC, als Intermediate einer aktiven Demethylierung immer mehr in den Fokus. Durch die Entdeckung der Glykosylaseaktivität von Tdg gegenüber fdC und cadC, wurde vermutet, dass die BER an der aktiven Demethylierung ebenfalls beteiligt ist.^[163, 203] In diesem Kapitel werden daher drei mögliche Reaktionswege einer aktiven Demethylierung näher beleuchtet. Zwei davon beziehen die BER mit ein, der andere Weg basiert auf einer direkten C-C Bindungsspaltung. Diese potentiellen aktiven Demethylierungswege sind in Abbildung 12 dargestellt.

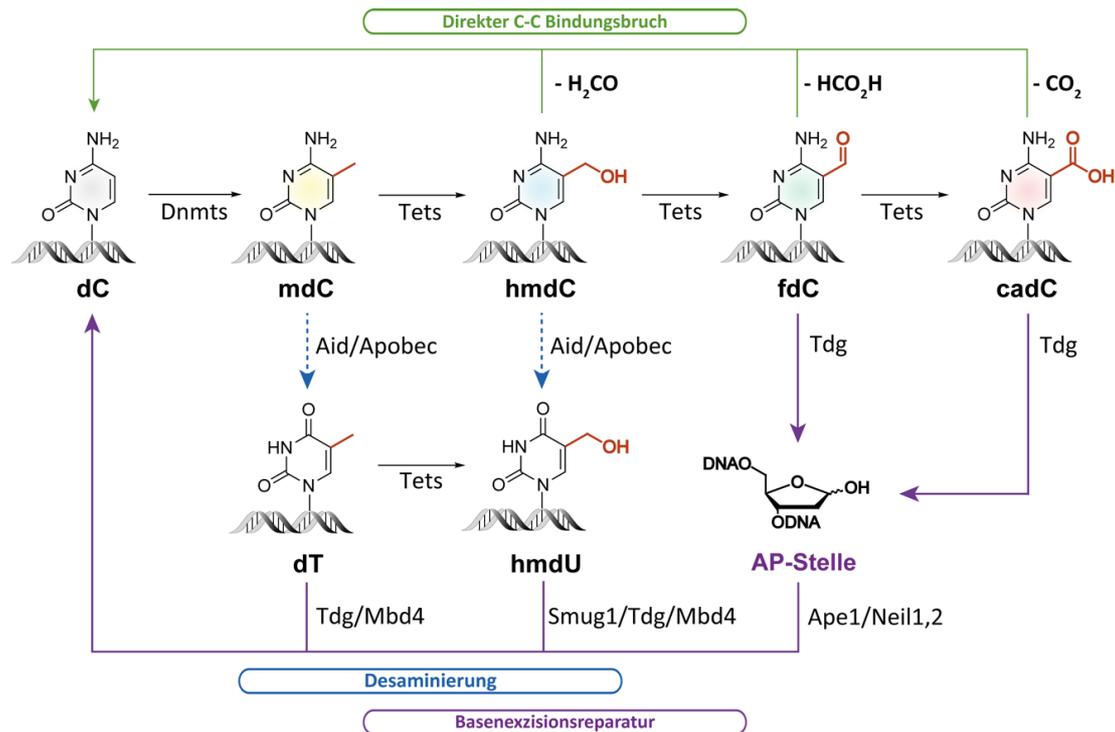


Abbildung 12: Wege der aktiven Demethylierung. Die nicht-kanonischen Bausteine mdC und hmdC könnten durch Desaminierung zu dT bzw. 5-Hydroxymethyl-2'-desoxyuridin (hmdU; blau) entfernt werden. Weitere Möglichkeiten sind die Demodifikation von fdC bzw. cadC über die BER (violett) und die direkte Demodifikation über einen C-C Bindungsbruch (grün). Der C-C Bindungsbruch würde zur Freisetzung von Formaldehyd (hmdC), Ameisensäure (fdC) oder Kohlenstoffdioxid (CO₂) führen. Alternativ wäre auch eine Übertragung auf unterschiedliche Akzeptormoleküle denkbar.

1.6.3.1 Aktive Demethylierung durch Tdg-vermittelte BER

Es wird vermutet, dass die BER neben ihrer Hauptaufgabe, der Reparatur der DNA, auch an der aktiven Demethylierung beteiligt ist. Dabei wird die modifizierte Base mittels schadenspezifischer Reparaturglykosylasen unter Ausbildung einer abasischen Stelle ausgeschnitten. Diese Enzyme spalten die N-glykosidische Bindung zwischen der Desoxyribose und der modifizierten bzw. fehlgepaarten Base. Unter den Glykosylasen der BER wird zwischen mono- und bifunktionalen Glykosylasen unterschieden.

Während monofunktionale Glykosylasen eine apurine oder apyrimidine (AP)-Stelle generieren, die weiter durch die Endonuklease *Apurinic/Apyrimidinic-Endonuclease 1* (Ape1) prozessiert wird, besitzen die bifunktionalen Glykosylasen eine Lyase-Aktivität. Diese führt zu einer β - oder β , δ -Eliminierung der abasischen Stelle. Die monofunktionale DNA-Glykosylase Tdg ist nach ihrer Fähigkeit benannt dT:dG Fehlpaarungen im Rahmen der BER zu prozessieren. Anschließende Exzision der fehlgepaarten Base durch die Glykosylase würde durch Ausbildung einer abasischer Stelle zu gefährlichen Einzelstrangbrüchen und im symmetrischen CpG-Kontext sogar zu einem Doppelstrangbruch führen. Zahlreiche Studien lassen die Vermutung zu, dass methyliertes Cytidin im Zuge der aktiven Demethylierung schrittweise durch die Tet-Enzyme oxidiert wird. Die höher oxidierten Derivate fdC bzw. cadC werden dabei durch Tdg aus dem Genom entfernt und unmodifiziertes Cytidin könnte so zurückgebildet werden.^[136, 163] Tdg gehört zur Familie der Uracil-DNA-Glykosylasen, die fehlgepaarte oder oxidierte Basen aus der doppelsträngigen DNA herausdrehen (*Base flipping*) und die glykosidische Bindung spalten können. Die steigenden Werte von fdC und cadC im *Knockout* der Glykosylase deuten darauf hin,^[136] dass die BER an der aktiven Demethylierung beteiligt sein könnte.^[204] Die nachweisliche Entfernung von fdC und cadC durch Tdg ist der bislang populärste Weg der aktiven Demethylierung. Die schnellere Exzision von fdC und cadC, verglichen mit der Entfernung von dT,^[163] lässt spekulieren ob die Prozessierung dieser Fehlpaarungen nicht die eigentliche Aufgabe von Tdg ist. Im Gegensatz zu anderen Glykosylasen interagiert Tdg mit Transkriptionsfaktoren,^[205] den Dnmts,^[206] Tet1 und Histon-modifizierenden Enzymen.^[207-209] Zudem arbeitet Tdg mit Ape1 zur Prozessierung der abasischen Stelle im Zuge der BER zusammen. Weiterhin gibt es Hinweise, dass die bisher als eigenständige bifunktionale DNA-Glykosylasen beschriebenen Enzyme *Nei endonuclease VIII-like* (Neil) 1 und 2 im Zusammenspiel mit Tdg die Prozessierung der abasischen Stelle beschleunigen könnten.^[210] Neil3 ist überwiegend als monofunktionale Glykosylase bekannt und führt zu einer β -Eliminierung.^[211, 212] Neil1 und Neil2 sind bifunktionale Glykosylasen, die oxidativ geschädigte Pyrimidine sowie chemisch veränderte Purine aus der DNA ausschneiden.^[213] Leonhardt und Mitarbeiter folgerten aus ihren Studien, dass Neil1 und Neil2 redundant mit Tdg arbeiten könnten.^[214] Schomacher *et al.* bestritten dies und schlugen vor, dass Tdg die bifunktionalen Glykosylasen anstelle von Ape1 rekrutiert, um die Desoxyribose schneller zu prozessieren.^[210] *In vitro* Studien von Slyvka *et al.* konnten zeigen, dass NEIL1 die Aktivität von Tdg stimuliert, was die Beobachtungen von Schomacher *et al.* unterstützt.^[215]

1.6.3.2 Aktive Demethylierung durch Desaminierung

Ein weiterer potentieller Weg der Demethylierung läuft über eine enzymatische Desaminierung ab. Als potentielle Desaminasen wurden die Enzyme der *Activation Induced cytidine Deaminase (Aid)*/*Apolipoprotein B mRNA-editing enzyme, catalytic polypeptide-like (ApoBec)*-Familie vorgeschlagen.^[184, 211] Durch eine Desaminierung von dC, mdC und hmdC, entstehen 2'-Desoxyuridin (dU), dT und hmdU fehlgepaart mit dG. Die so entstandene Fehlpaarung könnte anschließend durch DNA-Glykosylasen im Zuge der DNA-Reparatur behoben und unmodifiziertes dC zurückgewonnen werden. Resultierende dT:dG Fehlpaarungen könnten durch die Glykosylasen Tdg oder *Methyl-CpG-binding protein 4 (Mbd4)* repariert werden.^[213] Die Desaminase Aid arbeitet jedoch vorwiegend auf einzelsträngiger und nicht auf doppelsträngiger DNA.^[216-218] Weiterhin desaminiert Aid bevorzugt dC zu dU und im geringeren Maße mdC zu dT.^[219] Allerdings konnte durch *Knockout* von Aid in B-Zellen ein Effekt auf die mdC-Werte erzielt werden. Durch Aid-vermittelte Desaminierung von dC zu dU werden auf diese Weise die *Class Switch Recombination* und die somatische Hypermuation ausgelöst.^[220, 221] Bezugnehmend auf diese Beobachtungen wurde ein indirekter Mechanismus postuliert, der eine Desaminierung mit großräumiger DNA-Reparatur verbindet, wobei eine punktartige Desaminierung auch zur Entfernung benachbarter mdC führen könnte.^[222, 223] Dies würde zu einer umfassenderen DNA-Demethylierung führen.^[166] Weitere Studien konnten *in vitro* keine Desaminierung von hmdC und fdC zeigen.^[216, 219, 224] Ebenso ist eine Beteiligung von Aid an der DNA-Demethylierung umstritten.^[225-227] Eine Demethylierung im Gehirn von Mäusen wurde mit einer BER-gekoppelten Desaminierung in Verbindung gebracht, da ein Komplex bestehend aus Aid, Tdg und Gadd45a entdeckt wurde. Die Aktivität von Aid konnte jedoch nicht direkt nachgewiesen werden.^[191] Rai *et al* konnten zeigen, dass Aid in Zebrafischen durch Desaminierung eine aktive Demethylierung des Genoms einleitet.^[228] Diese Beobachtungen wurden jedoch bald darauf in Frage gestellt.^[229]

Eine Desaminierung von hmdC zu 5-Hydroxymethyl-2'-desoxyuridin (hmdU) ist generell denkbar. hmdU:dG Fehlpaarungen können durch die Glykosylasen Tdg, Mbd4 und *Single-strand-selective monofunctional Uracil-DNA glycosylase 1 (Smug1)* prozessiert werden.^[213] Neben der Rolle als oxidativer Schaden könnte hmdU auch eine regulative Funktion innehaben. In einer Studie wurden Aid/ApoBec in HEK-Zellen überexprimiert, wodurch in kotransfizierten hmdC enthaltenden Reportersträngen hmdU generiert wurde.^[200] Daraufhin wurde vermutet, dass hmdU als Intermediat der aktiven Demethylierung fungiert. Obwohl eine Studie Hinweise zugunsten einer Aktivität der Desaminasen gegenüber hmdC sammeln konnte,^[200] wurde in einer weiteren Studie nur eine geringe Aktivität dieser Enzyme festgestellt.^[219] Mit Hilfe einer Isotopenverfolgungsstudie konnte massenspektrometrisch die Generierung von hmdU durch eine Tet-vermittelte Oxidation von dT nachgewiesen werden. Ein geringer Beitrag einer Desaminierung von hmdC wurde nur in Abwesenheit von Tdg beobachtet.^[142]

Auf diese Weise generiertes hmdU ist vermutlich nur für eine kurze Zeit in der Zelle vorhanden. So lässt sich zusammenfassend sagen, dass diese Ergebnisse die Idee der aktiven Demethylierung durch Desaminierung von dC zu dU, mdC zu dT bzw. hmdC zu hmdU eher in Frage stellen. Aufgrund der widersprüchlichen Daten ist eine eindeutige Aussage über eine Beteiligung der Desaminierung, und damit einhergehender Reparatur, an der aktiven Demethylierung nicht zu treffen.

1.6.3.3 Aktive Demethylierung über einen direkten C-C Bindungsbruch

In der Zygote findet die aktive Demethylierung aufgrund der niedrigen Expressionslevel vermutlich unabhängig von Tdg statt (siehe Abschnitt 1.6.2).^[193] Dies lässt vermuten, dass es auch einen weiteren Weg zur Entfernung von mdC geben könnte. In der Theorie sind C-C Bindungsbruchreaktionen der C5-Position von hmdC, fdC und cadC mit steigender Wahrscheinlichkeit chemisch möglich.^[230] Dies würde den Reparaturweg der BER umgehen und unmodifiziertes dC ohne die Entstehung mutagener abasischer Stellen oder gefährlicher Strangbrüche wiederherstellen. Entsprechend könnten die beteiligten Enzyme die katalytische Aktivität einer Dehydroxylase, Deformylase oder Decarboxylase aufweisen. Das Vorkommen oxidativ demethylender, deformulierender und decarboxylierender Enzyme in der Natur ist bereits bekannt. Bereits gut untersucht wurde die oxidative Demethylierung der Lysin-spezifischen Demethylasen und der Jumonji-Enzyme.^[231] Ähnlich der Tet-Enzyme katalysieren die Jumonji-Dioxygenasen die Demethylierung der Lysin-Seitenketten von Histonfortsätzen in Abhängigkeit von Fe^{II} und α -Ketoglutarat.^[232] In Prokaryoten und Organellen fungieren Deformylasen als wichtiger Bestandteil der Proteinbiosynthese. Zur Initiation der Translation wird in diesen Organismen eine N-terminale Formyl-Gruppe an der Aminosäure Methionin genutzt und cotranslational mit Hilfe einer Peptid-Deformylase wieder entfernt.^[233]

In allen hier genannten Fällen ist die Formyl-Gruppe jedoch häufig an ein Heteroatom gebunden, was die Entfernung der Modifikation ermöglicht. Im Falle von fdC bzw. cadC liegt eine Bindung an ein aromatisches Csp²-Zentrum vor, was eine Defunktionalisierung bedeutend erschwert. Eine solche Aktivität ist nur als Redoxprozess denkbar. So wurde bereits für die Lanosterin-Demethylase und die Aldehyd-deformylierende Oxygenase eine Deformylierung an einem Csp³-Zentrum beobachtet.^[234-236] Die Deformylierung der Lanosterin-Demethylase ist ein wesentlicher Bestandteil der Cholesterin-Biosynthese.^[234] Dieses Enzym nutzt das Fe^{II}-Peroxy-Anion des Kofaktors Häm zum nukleophilen Angriff an der Formyl-Gruppe des Lanosterins, woraufhin Formiat freigesetzt wird.^[237] Ein besonderes Beispiel einer Decarboxylase ist die Isoorotat-Decarboxylase (IDCase) aus *Neurospora crassa*, die eine Demodifikation an der C5-Position eines Pyrimidins katalysiert.^[238] Im Zuge des *Thymine Salvage Pathways* wird Thymin zu Isoorotat umgewandelt. Durch schrittweise Oxidation der Basen wird Thymin zunächst über 5-Hydroxymethyluracil (hmU) und 5-Formyluracil (fU) zu 5-Carboxyluracil (caU) bzw. Isoorotat oxidiert.

Dies wird durch eine α -Ketoglutarat abhängige Fe^{II}-Oxygenase, *Thymine-7-Hydroxylase* (T7H), katalysiert, die ähnlich der Tet-Enzyme arbeitet.^[239] Mit Hilfe der IDCase wird die Base caU schließlich zu Uracil (U) decarboxyliert. Diese mechanistisch völlig unklare Reaktion wird vermutlich durch das Zn^{II} des aktiven Zentrums und einem Wassermolekül das als Nukleophil fungiert, katalysiert. Abbildung 13 zeigt den Katalysezyklus der IDCase und das aktive Zentrum co-kristallisiert mit dem Inhibitor 5-Nitro-U.^[238] Eine geringe Decarboxylaseaktivität der IDCase gegenüber des Nucleosids cadC konnte bereits beobachtet werden.^[240]

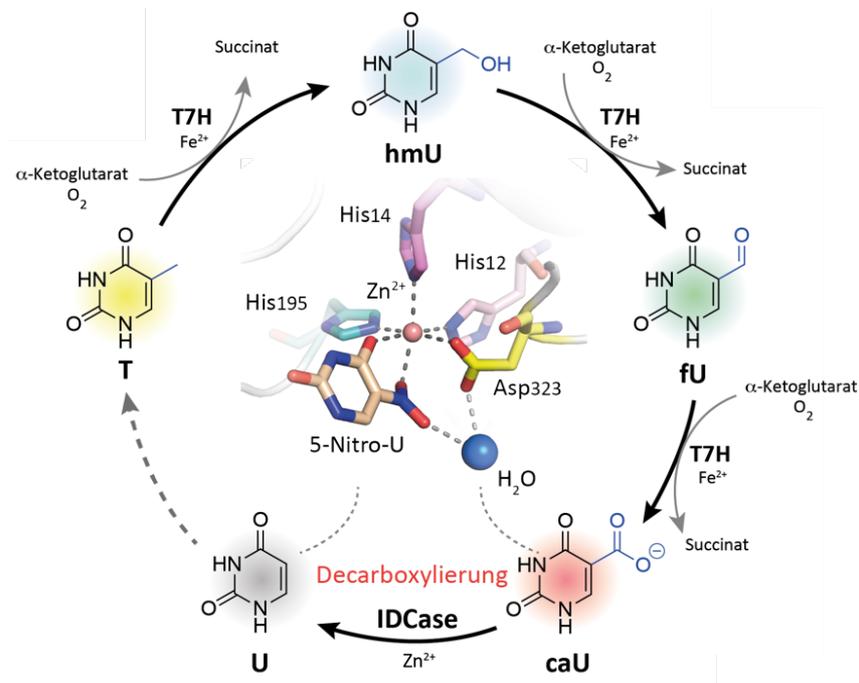


Abbildung 13: Decarboxylierung eines Pyrimidins im Zuge des *Thymine Salvage Pathway*. Das Schema zeigt die schrittweise Oxidation von T zu caU katalysiert durch die *Thymine-7-Hydroxylase* (T7H). Im nächsten Schritt decarboxyliert die Isoorotat-Decarboxylase (IDCase) Isoorotat (caU) zu U, welches zur Nucleotid-Biosynthese der Pyrimidine genutzt wird. Mittig befindet sich die Kristallstruktur des aktiven Zentrums der IDCase mit dem Inhibitor 5-Nitro-U. Die PyMOL-Struktur (PDB: 4HK6) wurde nach *Carell et al.* erstellt.^[85, 240]

Ein vergleichbares eukaryotisches Enzym konnte bisher noch nicht identifiziert werden. Die nächsten verwandten humanen Proteine sind die α -Amino- β -carboxymuconate- ϵ -semialdehyde decarboxylase (ACMSD) und die Cytosin-Desaminasen.^[85, 241, 242] Die direkte Entfernung der Carboxylgruppe des cadC durch eine putative Decarboxylase wäre der eingängigste und mechanistisch einfachste Weg. Bei der Suche nach dem Enzym mit Deformylase oder Decarboxylaseaktivität stieß man paradoxerweise auf die Dnmts. *In vitro* Studien mit bakteriellen C5-MTasen, sowie den entsprechenden murinen und humanen Enzymen, gelang die Detektion einer Decarboxylierung von cadC.^[243] Eine weitere *in vitro* Studie konnte eine geringe Decarboxylaseaktivität für DNMT3A und DNMT3B zusammen mit DNMT3L in Abwesenheit physiologischer Mengen des Kofaktors SAM beobachten.^[243, 244]

Chen et al. berichteten in diesem Zusammenhang ebenfalls von einer reversiblen Umwandlung von hmdC in kanonisches dC in Abwesenheit von SAM.^[245, 246] Allerdings scheint dieser Weg aufgrund der hohen intrazellulären SAM-Konzentration eher unwahrscheinlich. Zudem konnte *in vivo* bisher kein Hinweis auf eine Beteiligung der Dnmts an der aktiven Demethylierung erbracht werden. *Schiesser et al.* fanden jedoch Hinweise, dass die Decarboxylierung über einen Dnmt-ähnlichen Mechanismus (siehe Abschnitt 1.4.1) erfolgen könnte. In diesen Studien imitiert ein Thiol die Cystein-Seitenkette der hypothetischen Decarboxylase und führt einen nukleophilen Angriff an der C6-Position durch, woraufhin die Desaturierung der C5-C6 Bindung die Eliminierung von CO₂ erleichtert.^[230] Darüber hinaus konnte bei der Inkubation eines cadC-haltigen Oligonukleotids in Stammzelllysat eine Decarboxylaseaktivität beobachtet werden.^[247] Eine kürzlich veröffentlichte Isotopenverfolgungsstudie von *Iwan et al.* unterstützt die Idee eines direkten C-C Bindungsbruchs *in vivo*. Hierzu wurden isotope- und Fluor-markiertes fdC in Stammzellen sowie verschiedenen somatischen Zellen eingebracht und die Entstehung des Deformylierungsprodukts massenspektrometrisch verfolgt. Die erhaltenen Daten deuten eher auf eine Deformylierungs- als eine Decarboxylierungsreaktion hin.^[65] Die Identität der putativen Deformylase blieb jedoch ungeklärt.

2 Zielsetzung

Seit ihrer Entdeckung im frühen 20. Jahrhundert, etablierte sich die nicht-kanonische DNA-Base 5-Methylcytosin zur am besten charakterisierten epigenetischen DNA-Modifikation. Diese und die oxidierten Derivate 5-Hydroxymethylcytosin, 5-Formylcytosin und 5-Carboxylcytosin wird eine fundamentale Rolle bei der Steuerung zahlreicher biologischer Prozesse, wie der Genregulation und der epigenetischen Reprogrammierung zugeschrieben. Während der Prozess der DNA-Methylierung bereits gut untersucht wurde, ist die Entfernung von mdC noch nicht vollständig verstanden. Jüngste Entdeckungen ziehen zwei potentielle Wege der DNA-Demethylierung in Betracht: Der erste Weg führt über den Weg der Basenexzisionsreparatur und der Zweite über einen direkten C-C Bindungsbruch.

Diese Arbeit befasst sich vorrangig mit der Untersuchung der aktiven Demethylierung über eine C-C Bindungsbruchreaktion in Form einer Deformylierung bzw. einer Decarboxylierung (**Kapitel 3.1**). Als Sondenmoleküle wurden isotopenmarkiertes 5-Formyl-2'-desoxycytidin (fdC) und ein 2'-fluoriertes fdC-Derivat synthetisch hergestellt. Durch Einbringung der Markierung in verschiedene somatische und pluripotente Zelllinien sollte zunächst die Existenz der C-C Bindungsbruchreaktion in diesen Organismen nachgewiesen und der Metabolismus des inkorporierten, modifizierten DNA-Bausteins massenspektrometrisch verfolgt werden. Nach dem Nachweis des C-C Bindungsbruchs sollten der Mechanismus aufgeklärt und die beteiligten Enzyme identifiziert werden. Hierfür sollten verschiedene *Knockout*-Zelllinien und Inhibitor-Strategien genutzt werden.

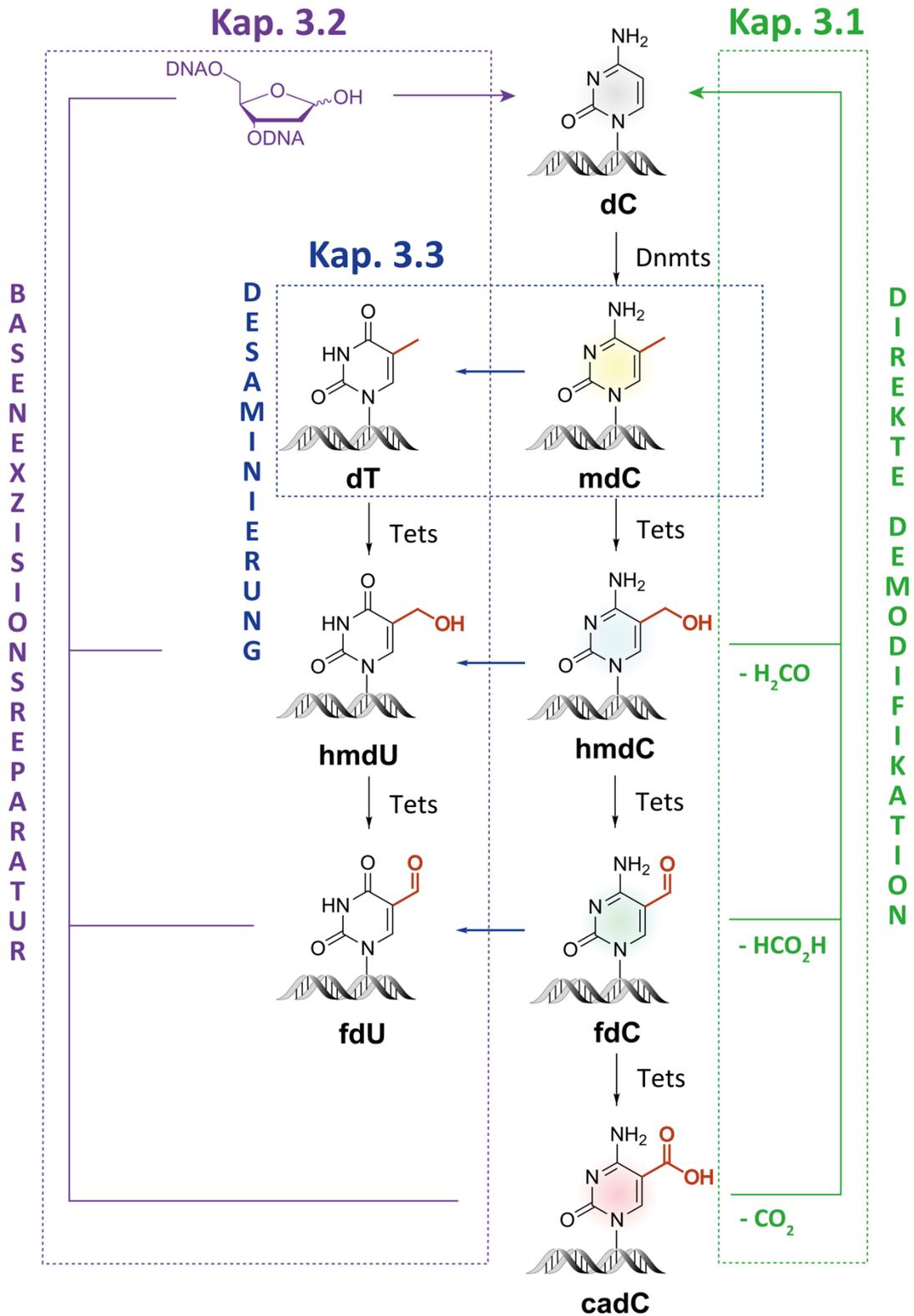
Als zweites Thema sollte die Entfernung der oxidierten Cytidin-Spezies fdC und cadC über die Basenexzisionsreparatur näher beleuchtet werden (**Kapitel 3.2**). Im Arbeitskreis wurde hierfür ein Derivatisierungsreagenz zur massenspektrometrischen Quantifizierung der BER-Intermediate entwickelt. Diese sensitive Methode soll zur genauen Bestimmung von abasischen Stellen und β -Eliminierungsprodukten *in vitro* und *in vivo* optimiert werden. Mittels stabiler Isotope sollten die entstandenen BER-Intermediate nach Einbringung markierter DNA-Bausteine ins Genom pluripotenter Zellen quantifiziert werden. Weiterhin sollte eine Quantifizierung der Intermediate in *Knockout*-Linien Aufschluss auf die Beteiligung der BER an der aktiven Demethylierung geben.

Ein weiteres Ziel dieser Arbeit ist die Erforschung eines alternativen DNA-Demethylierungsweges über die direkte Entfernung von mdC im Zuge der DNA-Reparatur (**Kapitel 3.3**). Mittels Zugabe von isotopenmarkiertem $^{13}\text{CD}_3$ -L-Methionin zum Kulturmedium der Zellen, sollte eine $^{13}\text{CD}_3$ -Markierung an der Methylgruppe von mdC eingeführt werden, was eine massenspektrometrische Verfolgung des Katabolismus ermöglichen sollte. Grundlage dieser Messung ist eine bereits im Arbeitskreis Carell etablierte Isotopenverfolgungsmethode.

Die Generierung von $^{13}\text{CD}_3\text{-mdU}$ aus $^{13}\text{CD}_3\text{-mdC}$ sollte in pluripotenten Zellen während des Übergangs vom Präimplantations- zum Postimplantationsepiblasten näher betrachtet werden. Durch die Generierung verschiedener Zelllinien unter Verwendung der CRISPR-Cas9 Genomeditierungstechnik sollten mögliche DNA-Reparaturwege zur Entfernung von mdC und Generierung von genomischem $^{13}\text{CD}_3\text{-mdU}$ betrachtet werden.

3 Ergebnisse und Publikationen

Übersicht der Kapitel



3.1 Studien zur DNA-Demethylierung über eine C-C Bindungsbruchreaktion

Zahlreiche Mechanismen wurden bereits für die DNA-Demethylierung von mdC und seinen Oxidationsprodukten postuliert. Dieses Kapitel befasst sich mit dem Nachweis eines alternativen DNA-Demethylierungsweges über eine C-C Bindungsbruchreaktion. Dieser Weg könnte über eine direkte Demodifikation in Form einer Deformylierung von fdC oder einer Decarboxylierung von cadC erfolgen, was zum Verlust der Aldehyd- bzw. Carboxy-Funktion und so zurück zu dC führen würde (Abbildung 14).

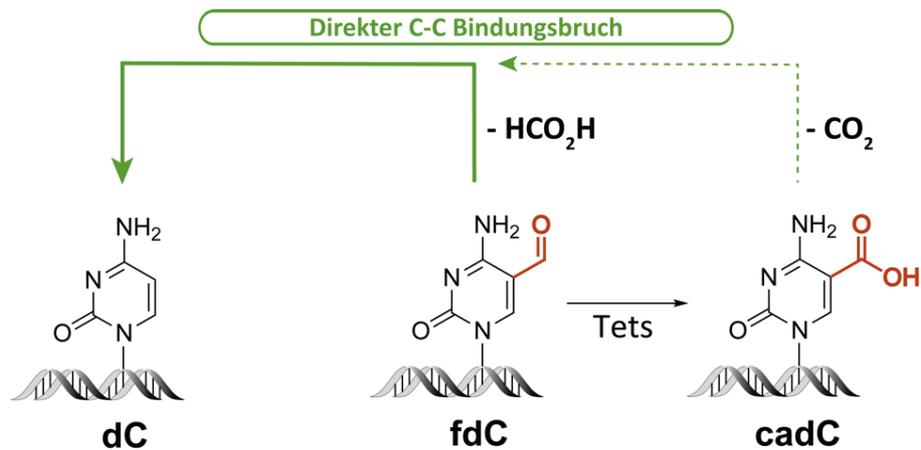


Abbildung 14: DNA-Demethylierung über einen C-C Bindungsbruch. Das Schema zeigt die Rückgewinnung von kanonischem dC über eine direkte Deformylierung von fdC bzw. Decarboxylierung von cadC.

3.1.1 Umwandlung von 5-Formyldesoxycytidin zu Desoxycytidin über einen C-C Bindungsbruch in lebenden Zellen

„5-Formylcytosine to cytosine conversion by C-C bond cleavage in vivo“ Katharina Iwan,[#] René Rahimoff,[#] Angie Kirchner,[#] Fabio Spada,[#] Arne S. Schröder, Olesea Kosmatchev, Shqiponja Ferizaj, Jessica Steinbacher, Edris Parsa, Markus Müller & Thomas Carell

#: Diese Autoren haben zu gleichen Teilen beigetragen.

Prolog

Schröder *et al.* konnten in einer vorangehenden Veröffentlichung zeigen, dass 2′-(R)-fluorierte Nucleoside von Tdg nicht als Substrat erkannt, jedoch von Polymerasen in die DNA eingebaut werden können.^[248] Demnach bieten fluorierte Nucleoside einen vielversprechenden Ansatz die Entfernung von fdC bzw. cadC in Abwesenheit der Basenexzisionsreparatur zu betrachten. Im nachfolgenden Manuskript wird 2′-(R)-Fluoro-5-formyl-2′-desoxycytidin (F-fdC), sowie isotonenmarkiertes [¹³C₅][¹⁵N₂]-5-Formyl-2′-desoxycytidin ([¹³C₅][¹⁵N₂]-fdC), in das Genom von Säugetierzellen eingebracht und der Metabolismus dieser fdC-Derivate mittels UHPLC-MS/MS verfolgt (Abbildung 15). Hierdurch konnte quantitativ gezeigt werden, dass nach Einbringung der fdC-Nucleoside eine direkte Umwandlung in unmodifiziertes 2′-(R)-Fluorodesoxycytidin (F-dC) bzw. [¹³C₅][¹⁵N₂]-dC stattfindet. Diese C-C Bindungsbruchreaktion könnte einen alternativen Demethylierungsweg darstellen der ohne die Bildung von DNA-Schäden abläuft.

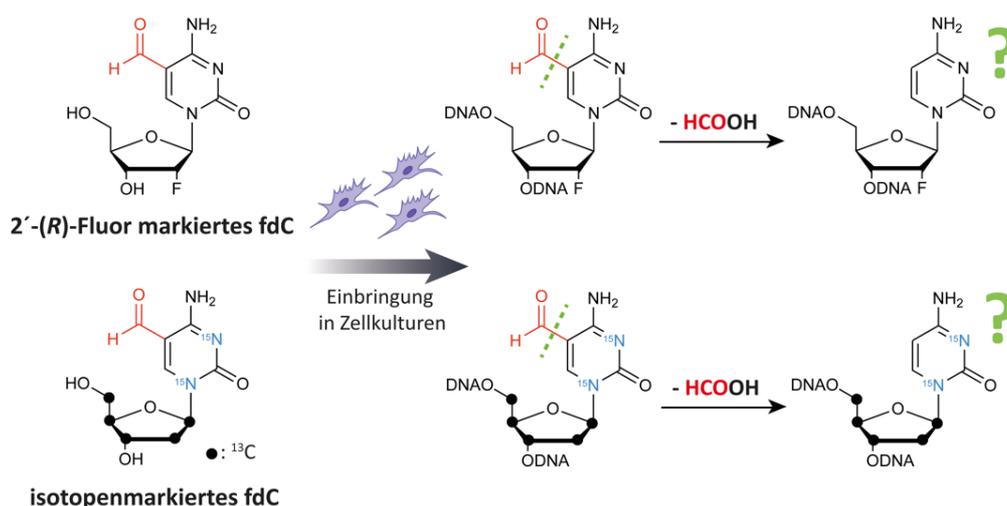


Abbildung 15: Studien zum C-C Bindungsbruch mit Hilfe der fluorierter und isotonenmarkierter Sondenmoleküle. Nach Einbringung in lebende Zellen, Isolation der genomischen DNA und enzymatischem Verdau, wird eine Demodifizierung von fdC zu dC mittels UHPLC-MS/MS beobachtet.

Autorenbeitrag

Alle verwendeten Zelllinien wurden von mir kultiviert. Die Isotopenmarkierungsexperimente und die Fütterung der fluoridierten Nucleoside wurden von mir entworfen und durchgeführt. Erste Markierungsexperimente mit F-fdC bzw. F-dC und die Studie zur Vitamin C Behandlung, wurden von *Fabio Spada* vorgenommen. Weiterhin wurde die genomische DNA aller Proben von mir isoliert. Die Extraktion des löslichen Nucleosid- bzw. Nucleotidpools wurde von mir in Zusammenarbeit mit *Shqiponja Ferizaj* vorgenommen. Die erhaltenen Daten wurden von den Erstautoren, gemeinsam mit *Thomas Carell*, interpretiert. Weiterhin war ich an der Abfassung des Manuskriptes beteiligt. Eine detaillierte Auflistung der Autorenbeiträge ist im nachfolgenden Manuskript zu finden.

Lizenz

Kopie der Publikation mit Erlaubnis des Verlags; Copyright 2017 Springer Nature.

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 Schröder¹, Olesea Kosmatchev¹, Shqiponja Ferizaj¹, Jessica Steinbacher¹, Edris Parsa¹, Markus Müller¹  & 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 Dnmt3l³. 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^{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 brain¹². 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 stable¹⁵, or at least semi-stable¹⁶, 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 Tdg²⁴. 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 nucleophile²⁶. The chemistry is therefore quite similar to the known reaction mechanisms employed by the Dnmt proteins³. Although chemically plausible, it is unclear whether such direct C–C bond cleavage reactions occur within the genome²⁸. 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 [¹³C₅][¹⁵N₂]fdC (1; Fig. 2a), in which all five C-atoms of the ribose ring were exchanged with ¹³C and the two in-ring nitrogen atoms replaced with ¹⁵N (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

¹Center for Integrated Protein Science Munich CiPSM at the Department of Chemistry, Ludwig-Maximilians-Universität München, Munich, Germany.

²These authors contributed equally to this work. *e-mail: Thomas.carell@lmu.de

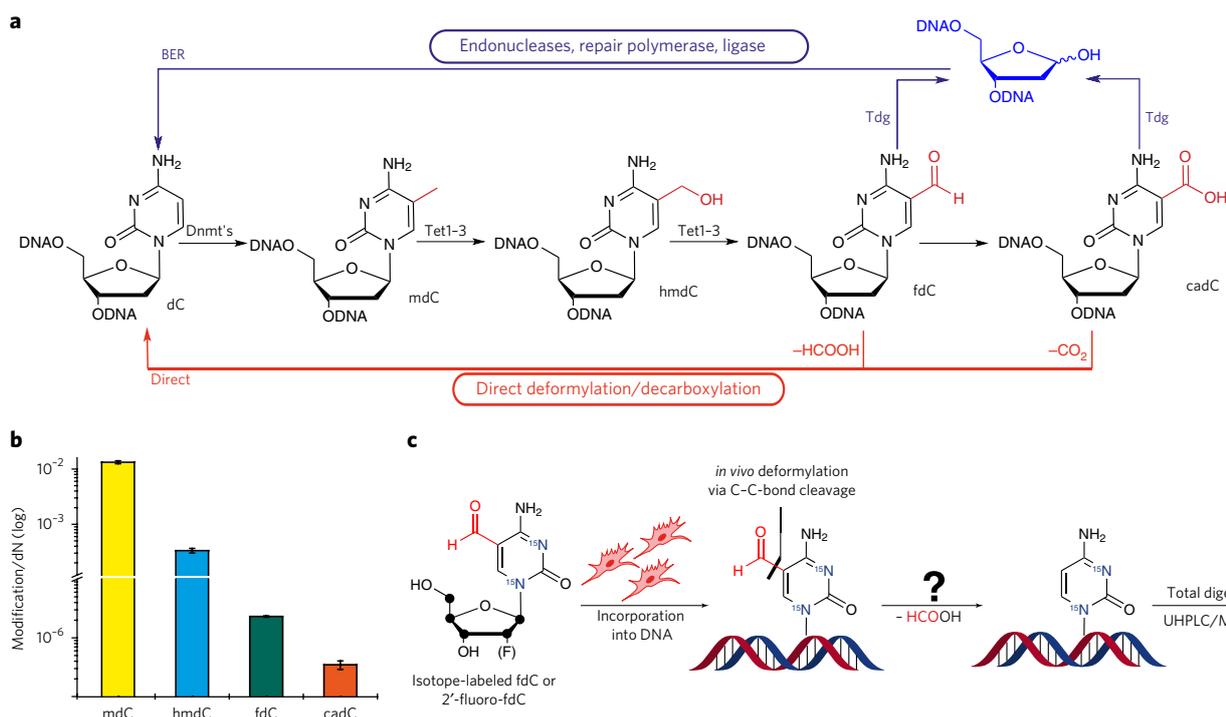


Figure 1 | Isotope tracing studies. (a) Suggested pathways of active demethylation. Thymine-DNA glycosylase (Tdg)-based cleavage of the glycosidic bond of 5-formyl-dC (fdC) and 5-carboxy-dC (cadC) results in an abasic site, which initiates a BER process that leads to the replacement of fdC and cadC by canonical dC (blue). Deformylation of fdC ($-\text{HCOOH}$) and decarboxylation ($-\text{CO}_2$) 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 [$^{13}\text{C}_5$][$^{15}\text{N}_2$]dC (**2**) and its deamination to [$^{13}\text{C}_5$][$^{15}\text{N}_2$]dU (**3**), followed by methylation of **3** to [$^{13}\text{C}_5$][$^{15}\text{N}_2$]dT (**4**). Alternatively, compound **1** can deaminate to [$^{13}\text{C}_5$][$^{15}\text{N}_2$]fdU (**5**) and, finally, the deformylated compound **2** can be methylated to [$^{13}\text{C}_5$][$^{15}\text{N}_2$]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 **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 [$^{13}\text{C}_5$][$^{15}\text{N}_2$]fdC is indeed demodified (Fig. 2c). To unequivocally prove the identity of **2**, an even heavier isotopically modified dC isotopologue, [$^{13}\text{C}_9$][$^{15}\text{N}_3$]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 $3\text{--}5 \times 10^{-7}$ of [$^{13}\text{C}_5$][$^{15}\text{N}_2$]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 studies²⁶, **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, [$^{13}\text{C}_9$][$^{15}\text{N}_3$]dC, and indeed detected the expected presence of the corresponding isotopically labeled deamination products [$^{13}\text{C}_9$][$^{15}\text{N}_3$]dT and [$^{13}\text{C}_9$][$^{15}\text{N}_3$]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

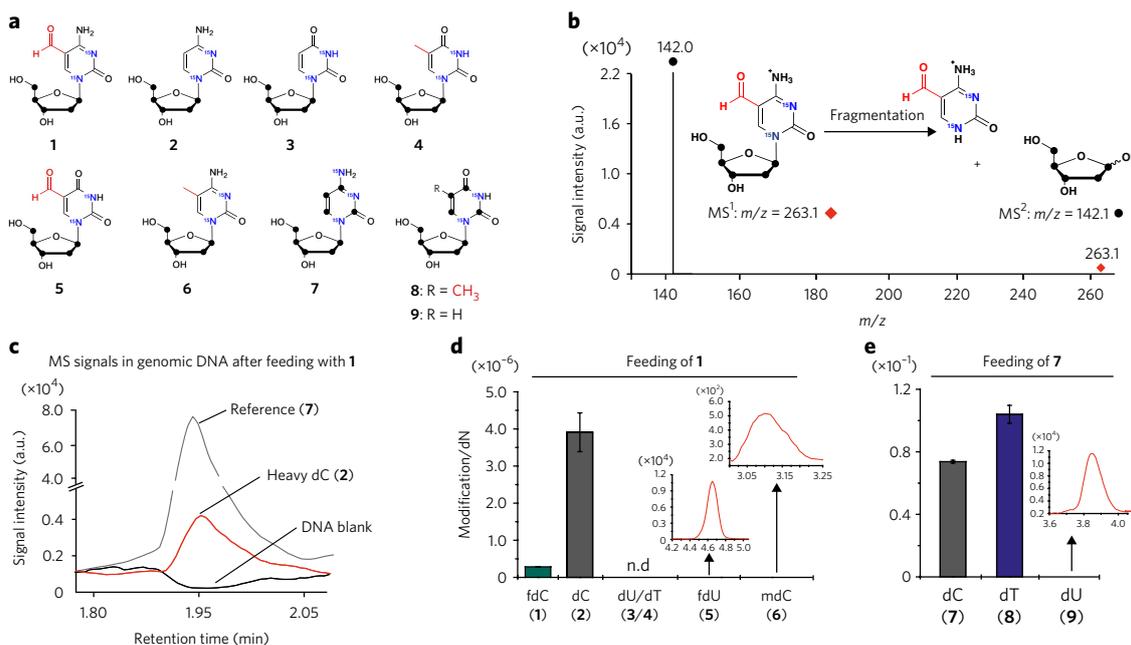


Figure 2 | Conversion of isotopically labeled fdC into dC in mESCs. (a) Overview of the compounds that may be detected after feeding of **1** or **7** to mESCs. (b) Feeding of **1** to mESCs results in incorporation of the isotopologue into the genomic DNA, as proven by its fingerprint MS transition. 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 methyltransferases³¹ 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 observed³².

To show that the fluorinated compounds are valid probe molecules, we first added **10** to the mESC culture at 0.5 μ M, 1.0 μ M or 2.5 μ M 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'-[D₃]F-mdC (**18**), (*R*)-2'-[¹⁵N₂]F-dC (**19**) and (*R*)-2'-[¹⁵N₂]F-fdC (**20**) and used them as internal standards for quantification (**Fig. 3a**). Upon feeding mESCs with 1 μ M **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 μ M 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

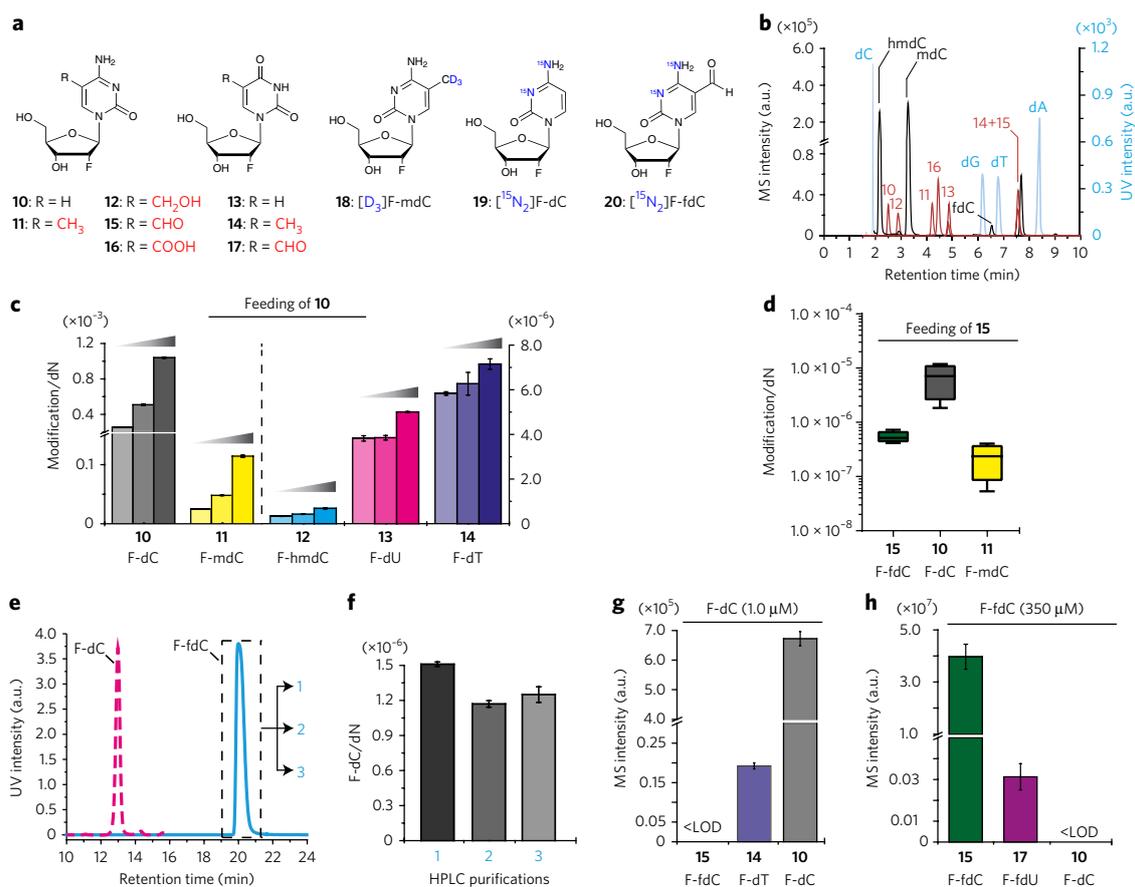


Figure 3 | F-dC is converted into F-dC within the genome. (a) Chemical structures of 2'-fluorinated dC and dU derivatives that were investigated and internal standards used. (b) Resulting UV and MS traces of the nucleosides under investigation. Light blue, canonical bases; black, natural dC derivatives; 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'-[D₃]F-mdC, (R)-2'-[¹⁵N₂]F-dC and (R)-2'-[¹⁵N₂]F-fdC (Fig. 3a), we found that remethylation of **10** results in levels of 2.8% ($\pm 0.3\%$) 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

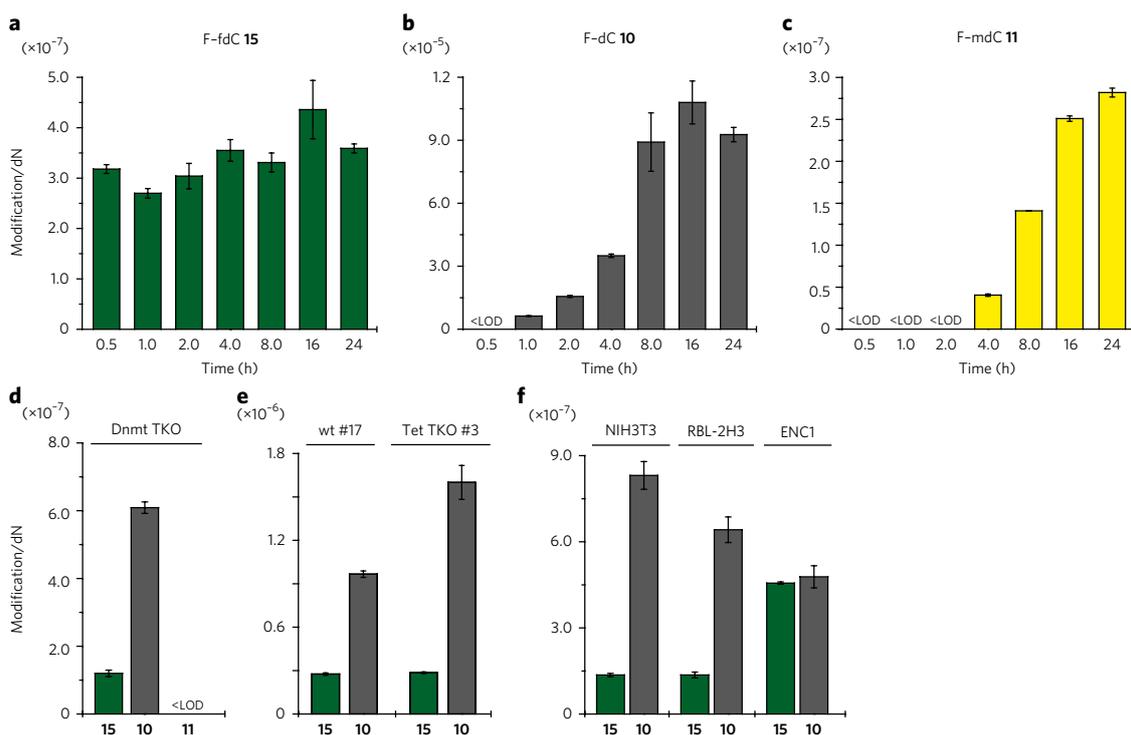


Figure 4 | Demodification of 2'-fluorinated fdC is a rapid process, does not require Dnmt or Tet enzymes and occurs also in somatic cell types.

(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 μ M 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 demethylated 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 proposed²⁵. 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 BER³⁶. 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

Tdg²⁴. 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 14 α -demethylase (CYP51A1), which oxidizes the C14 α 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 peroxy 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 peroxy 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 orotate⁴¹ and iso-orotate⁴² decarboxylase. Indeed, it was suggested that the iso-orotate decarboxylase could be a blueprint for a putatively existing cadC decarboxylase⁴³. 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

- Smith, Z.D. & Meissner, A. DNA methylation: roles in mammalian development. *Nat. Rev. Genet.* **14**, 204–220 (2013).
- Schübeler, D. Function and information content of DNA methylation. *Nature* **517**, 321–326 (2015).
- Jeltsch, A. & Jurkowska, R.Z. Allosteric control of mammalian DNA methyltransferases – a new regulatory paradigm *Nucleic Acids Res.* **44**, 8556–8575 (2016).
- Tahiliani, M. *et al.* Conversion of 5-methylcytosine to -hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930–935 (2009).
- Kriaucionis, S. & Heintz, N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **324**, 929–930 (2009).
- Pfaffeneder, T. *et al.* The discovery of 5-formylcytosine in embryonic stem cell DNA. *Angew. Chem. Int. Ed. Engl.* **50**, 7008–7012 (2011).
- Ito, S. *et al.* Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333**, 1300–1303 (2011).
- He, Y.F. *et al.* Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333**, 1303–1307 (2011).
- Globisch, D. *et al.* Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS One* **5**, e15367 (2010).
- Münzel, M., Globisch, D. & Carell, T. 5-Hydroxymethylcytosine, the sixth base of the genome. *Angew. Chem. Int. Ed. Engl.* **50**, 6460–6468 (2011).
- Pfaffeneder, T. *et al.* Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA. *Nat. Chem. Biol.* **10**, 574–581 (2014).
- Wagner, M. *et al.* Age-dependent levels of 5-methyl-, 5-hydroxymethyl-, and 5-formylcytosine in human and mouse brain tissues. *Angew. Chem. Int. Ed. Engl.* **54**, 12511–12514 (2015).
- Branco, M.R., Ficz, G. & Reik, W. Uncovering the role of 5-hydroxymethylcytosine in the epigenome. *Nat. Rev. Genet.* **13**, 7–13 (2011).
- Wu, H. & Zhang, Y. Mechanisms and functions of Tet protein-mediated 5-methylcytosine oxidation. *Genes Dev.* **25**, 2436–2452 (2011).
- Bachman, M. *et al.* 5-Formylcytosine can be a stable DNA modification in mammals. *Nat. Chem. Biol.* **11**, 555–557 (2015).
- Su, M. *et al.* 5-Formylcytosine could be a semipermanent base in specific genome sites. *Angew. Chem. Int. Ed. Engl.* **55**, 11797–11800 (2016).
- Raiber, E.A. *et al.* 5-Formylcytosine alters the structure of the DNA double helix. *Nat. Struct. Mol. Biol.* **22**, 44–49 (2015).
- Song, C.X. *et al.* Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. *Cell* **153**, 678–691 (2013).
- 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).
- 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).
- 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).
- 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).
- 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).
- 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).
- Wu, S.C. & Zhang, Y. Active DNA demethylation: many roads lead to Rome. *Nat. Rev. Mol. Cell Biol.* **11**, 607–620 (2010).
- Schiesser, S. *et al.* Deamination, oxidation, and C–C bond cleavage reactivity of 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxylcytosine. *J. Am. Chem. Soc.* **135**, 14593–14599 (2013).
- Liutkevičiūtė, Z. *et al.* Direct decarboxylation of 5-carboxylcytosine by DNA C5-methyltransferases. *J. Am. Chem. Soc.* **136**, 5884–5887 (2014).
- Schiesser, S. *et al.* Mechanism and stem-cell activity of 5-carboxylcytosine decarboxylation determined by isotope tracing. *Angew. Chem. Int. Ed. Engl.* **51**, 6516–6520 (2012).
- 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).
- 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).
- 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).
- 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).
- 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 14 α -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 isoortate decarboxylase: identification of the thymidine salvage pathway in lower eukaryotes. *Anal. Biochem.* **266**, 85–92 (1999).
43. Xu, S. *et al.* Crystal structures of isoortate 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 literature³². 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 $\mu\text{g}/\text{mL}$ streptomycin, 1 \times 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 reported⁴⁴. Under these conditions primed cells remained pluripotent for at least seven days as determined by epifluorescence with an Oct4-YFP knock-in cell line⁴⁵. 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/SvJae 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 $\mu\text{g}/\text{mL}$ 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 described⁵¹. 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 ; [¹³C₅][¹⁵N₂] fdC (**1**), 50 μM ; [¹³C₉][¹⁵N₃]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 sheared 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 H₂O were digested as follows: An aqueous solution (7.5 μL) of 480 μM ZnSO₄ 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'-[¹⁵N₂]F-dC (**19**), (R)-2'-[¹⁵N₂]F-hmdC (**37**) and (R)-2'-[¹⁵N₂]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 run⁴¹. The source-dependent parameters were as follows: gas temperature 80 °C, gas flow 15 L/min (N₂), nebulizer 30 psi, sheath gas heater 275 °C, sheath gas flow 11 L/min (N₂), 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 \times 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 Fluoro-nucleosides used for the quantification.

To obtain the internal calibration curves for exact quantification, we analyzed each standard, namely (R)-2'-[¹⁵N₂]F-fdC (**20**; $n = 205$ fmol), (R)-2'-[¹⁵N₂]F-dC (**19**; $n = 793$ fmol) and (R)-2'-[D₃]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 μ M in mESC culture medium at 37 °C and 5% CO₂ 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 H₂O (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 H₂O, the cartridges were dried *in vacuo*. The nucleosides were eluted with MeOH/MeCN (1:1), evaporated to dryness via speedvac and resuspended in H₂O.

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% CO₂ 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 H₂O 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 μ M F-fdC (**10**), 50 μ M [¹³C₃][¹⁵N₂]fdC (**1**) or 350 μ M F-fdC (**15**). Cells were washed twice with PBS (Sigma-Aldrich), harvested by trypsinization and pelleted by centrifugation for 3 min at 300g. 500 μ L ice-cold 50% (v/v) acetonitrile was added dropwise to the pellet and vortexed⁵⁵. 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 \times 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).

3.1.2 Untersuchung der aktiven Demethylierung in pluripotenten Zellen mittels isotopenmarkierter Cytidin-Derivate

3.1.2.1 Prolog

In Abschnitt 3.1.1 konnte durch isotopenmarkierte und fluorierte Sondenmoleküle eine C-C Bindungsbruchreaktion in lebenden Zellen nachgewiesen werden. Der Nachweis eines alternativen DNA-Demethylierungsweges über eine Deformylierungs- bzw. Decarboxylierungsreaktion wird in diesem Kapitel unter Zuhilfenahme deuterierter Sondenmoleküle fortgeführt. In der Gruppe wurden hierzu bereits Vorarbeiten geleistet, die in den Dissertationen von *Dr. Jessica Steinbacher* und *Dr. Arne Schröder* beschrieben sind.^[249, 250] Das nachfolgende Kapitel beinhaltet eine Weiterführung der bisherigen Markierungsstudien mit neuen Sondenmolekülen (Abbildung 16) und verbesserten UHPLC-MS/MS-Methoden. Dabei wird der Metabolismus der markierten dC-Derivate unter diversen Kulturbedingungen sowie unterschiedlichen pluripotenten Zelllinien betrachtet.

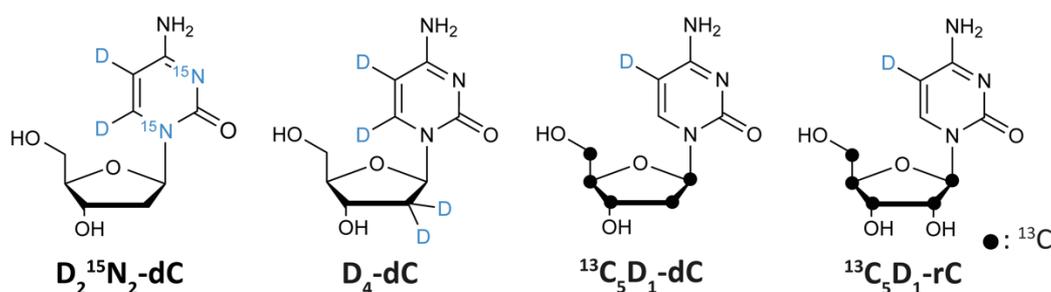


Abbildung 16: Chemische Strukturen der vier Sondenmoleküle.

Zur massenspektrometrischen Verfolgung des C-C Bindungsbruchs wurde zunächst ein 2'-Desoxycytidin-Derivat von *Dr. Arne Schröder* synthetisiert, welches zwei Deuterium- und zwei ¹⁵N-Markierungen in der Base aufweist. Dieses Molekül (D₂¹⁵N₂-dC) ist damit um vier Masseneinheiten schwerer als das natürlich vorkommende Homolog (dC). Aufgrund der kommerziell erhältlichen Ausgangsmaterialien war eine Isotopenreinheit von 95 % erreichbar. Nach Zugabe des Sondenmoleküls zum Kulturmedium wird dieses in die genomische DNA der Zellen eingebaut und kann durch die DNA-Methyltransferasen methyliert und die Tet-Dioxygenasen weiter oxidiert werden. Dabei geht die Deuterium-Markierung an der C5-Position verloren, wodurch das Molekül nur noch einen Massenunterschied von drei Einheiten aufweist. Nach einer Deformylierung oder Decarboxylierung entsteht so ein um drei Masseinheiten schwereres Cytidin. Dagegen geht durch Entfernung des markierten fdC oder cadC durch Tdg im Zuge der Basenexzisionsreparatur die Isotopenmarkierung am Cytidin gänzlich verloren oder es kommt zum Einbau der markierten Ausgangsverbindung. Abbildung 17 illustriert den eben beschriebenen Zusammenhang.

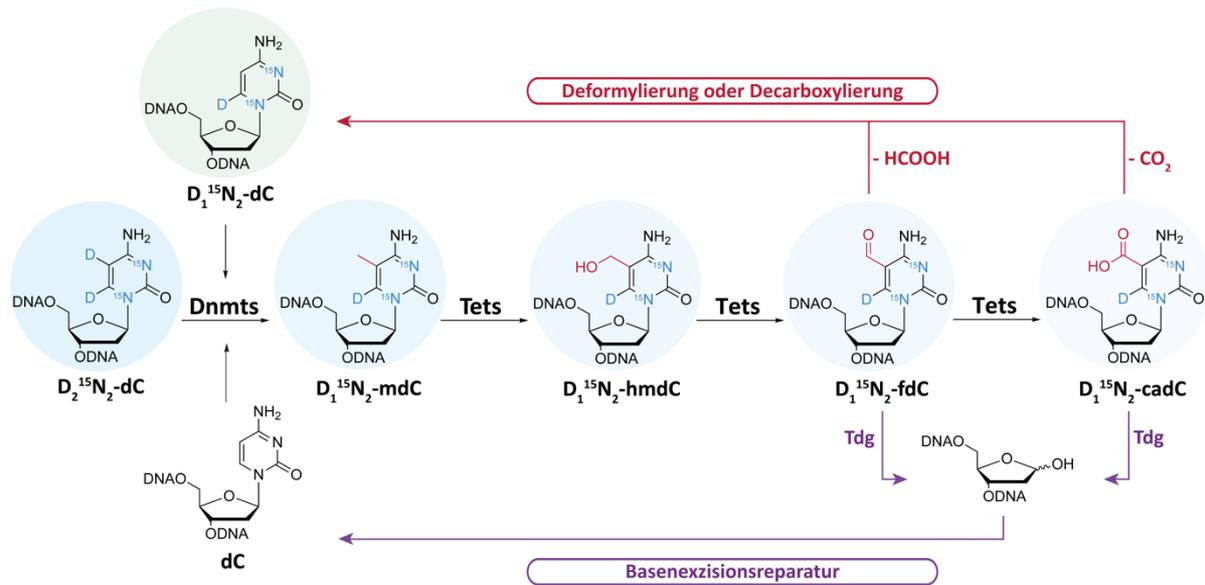


Abbildung 17: Unterscheidung der möglichen Wege einer aktiven DNA-Demethylierung durch Verfolgung der Isotopenmarkierung. Das Schema zeigt die unterschiedlichen isotopenmarkierten 2'-Desoxycytidin-Derivate nach Einbau von $D_2^{15}N_2$ -dC in die genomische DNA und die beiden putativen Demethylierungswege.

Die Akkumulation der $D_1^{15}N_2$ -dC Verbindung würde auf einen C-C Bindungsbruch schließen lassen, wobei jedoch die Verunreinigung der Ausgangsverbindung bei dieser Betrachtung nicht außer Acht gelassen werden darf. Da die Verunreinigung die gesuchte Spezies $D_1^{15}N_2$ -dC selbst ist, könnte es sich auch um ein falsch positives Ergebnis handeln, was überprüft werden muss.

In einem ersten Experiment wurde die Verbindung von *Dr. Fabio Spada* an humane induzierte pluripotente Stammzellen gefüttert. Erste Isotopenverfolgungsexperimente, durchgeführt von *Dr. Jessica Steinbacher*, zeigten eine Inkorporation der $D_2^{15}N_2$ -dC Verbindung von etwa 7 %. Diese war jedoch zu gering, um alle höher oxidierten Spezies verlässlich quantifizieren zu können. Die $D_1^{15}N_2$ -dC Spezies konnte ebenfalls detektiert werden. Es blieb jedoch unklar, ob es sich um die Verunreinigung der Ausgangsverbindung oder um eine aktive Entstehung über eine C-C Bindungsbruchreaktion handelte.

3.1.2.2 Ergebnisse

In einem ersten Schritt wurde eine Nachdeuterierung des Sondenmoleküls $D_2^{15}N_2$ -dC durch *Dr. René Rahimoff* vorgenommen. Es konnte eine Isotopenreinheit von 97 % erreicht werden. Um höhere Mengen an oxidierten Cytidin-Spezies und damit Substraten für einen möglichen C-C Bindungsbruch zu erzeugen, wurden *Epiblast-like* Zellen (EpiLCs) für weitere Experimente gewählt. *Pfaffeneder et al.* konnten bereits zeigen, dass diese Zellen nach 48 h hohe Mengen an fdC und cadC aufweisen, weshalb ein hoher Umsatz der oxidierten dC Verbindungen zu erwarten wäre.^[142] EpiLCs werden unter serumfreien Bedingungen kultiviert und das *Priming* durch Zugabe von Wachstumsfaktoren initiiert. Auf diese Weise können EpiLCs aus ESCs gebildet werden. Diese Zellen ähneln bezüglich des Transkriptoms dem Postimplantationsepiblasten.^[251] Auch unter serumfreien Bedingungen werden die naive mESCs unter Zugabe der beiden Inhibitoren PD0325901 und CHIR99021 sowie LIF kultiviert. Werden diese drei Komponenten entfernt und die Zytokine *Fibroblast Growth Factor-2* (FGF2) und Activin-A zugegeben, können EpiLCs erhalten werden.^[252] Diese *Priming*-Methode erzielt eine homogenere Kultur und erzeugt einen deutlichen Anstieg der oxidierten dC-Spezies. Abbildung 18A zeigt die unterschiedliche Menge an fdC und cadC beim Vergleich der Werte vor (0 h) und nach (48 h) Zugabe der Zytokine. Die fdC-Werte steigen um das Zehnfache an, während es bei cadC zu einer Verdopplung der Werte kommt. Die EpiLCs stellen somit ein vielversprechendes Modell zur Studie einer potentiellen Deformylierung bzw. Decarboxylierung dar.

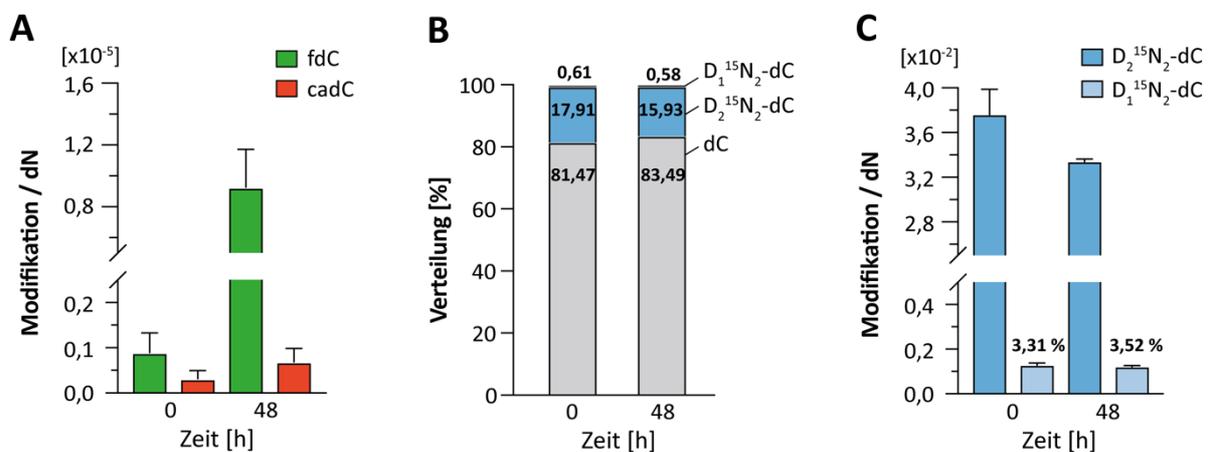


Abbildung 18: Untersuchung des C-C Bindungsbruchs in Epiblast-like Zellen. A) Absolute Quantifizierung der fdC und cadC Level in EpiLCs zu unterschiedlichen Zeitpunkten. Abgebildet sind die normierten Quantifizierungsergebnisse zweier biologischen Replikate, jeweils gemessen als technisches Triplikat. **B)** Relative Darstellung der markierten Cytidin-Spezies. **C)** Absolute Quantifizierung der $D_2^{15}N_2$ -dC und $D_1^{15}N_2$ -dC Nucleoside zum Zeitpunkt 0 h und 48 h. Berechnung der prozentualen Menge an $D_1^{15}N_2$ -dC bezogen auf die Summe aller markierter Verbindungen. Eine biologische Probe wurde als technisches Triplikat vermessen.

Im Zuge dessen wurden Stammzellen der Linie J1 für zwei Passagen unter naiven Bedingungen und Zusatz von $100 \mu\text{M D}_2^{15}\text{N}_2\text{-dC}$ serumfrei kultiviert. Diese Vorfütterung sollte sicherstellen, dass sich eine ausreichend große Menge des markierten Nukleosids im Genom der Zellen befindet. Innerhalb 48 h wurden aus den mESCs durch Zugabe der genannten Faktoren und $100 \mu\text{M D}_2^{15}\text{N}_2\text{-dC}$ EpiLCs generiert. Nach Isolation der DNA wurde diese enzymatisch verdaut und massenspektrometrisch untersucht. Die Messung und Auswertung der Daten wurde von *Dr. Jessica Steinbacher* unternommen.^[249]

In Abbildung 18B ist die Verteilung der isotopenmarkierten Cytidin-Derivate in Bezug auf dC dargestellt. Die relativen Mengen an dC, $\text{D}_2^{15}\text{N}_2\text{-dC}$ und $\text{D}_1^{15}\text{N}_2\text{-dC}$ sind als prozentualer Anteil aller dC Spezies dargestellt. Die Daten zeigen eine Inkorporation der $\text{D}_2^{15}\text{N}_2\text{-dC}$ Spezies von etwa 16-18 %. Auch die Level an $\text{D}_1^{15}\text{N}_2\text{-dC}$ sind mit etwa 0,6 % gut detektierbar. $\text{D}_1^{15}\text{N}_2\text{-dC}$ könnte sowohl durch die Verunreinigung im Zuge der Synthese, als auch durch eine Deformylierung bzw. Decarboxylierung ins Genom der Zellen gelangen. Interessanterweise nimmt der Anteil an $\text{D}_1^{15}\text{N}_2\text{-dC}$ während des *Primings* im Vergleich zum $\text{D}_2^{15}\text{N}_2\text{-dC}$ weniger stark ab. In Abbildung 18C sind die Werte der $\text{D}_2^{15}\text{N}_2\text{-dC}$ bzw. $\text{D}_1^{15}\text{N}_2\text{-dC}$ Verbindung auf alle Nukleoside bezogen, dargestellt. Nach 48 h ist trotz kontinuierlicher Kultivierung in isotopenhaltigem Medium ein geringer Abfall der D_2 -Modifikation zu erkennen. Bei Berechnung des prozentualen Anteils der $\text{D}_1^{15}\text{N}_2\text{-dC}$ Spezies bezogen auf die Summe der beiden deuterierten Verbindungen, ist nach 48 h ein Anstieg von 0,21 % zu beobachten. Dieser Anstieg könnte auf eine mögliche C-C Bindungsbruchreaktion hindeuten. Um diese Beobachtung zu bekräftigen, wurde ein weiteres Sondenmolekül entwickelt und untersucht.

Untersuchung der aktiven Demethylierung mit dem Sondenmolekül $\text{D}_4\text{-dC}$

Das neue Molekül sollte einen Massenunterschied von mindestens vier Einheiten zum natürlichen dC aufweisen und mindestens eine Isotopenmarkierung in der Desoxyribose aufweisen (siehe Abbildung 19). Der Verlust des Deuteriums an der C5-Position der Base sollte erneut Hinweise auf eine aktive Demethylierung über einen C-C Bindungsbruch geben. Die Isotopenreinheit des neu synthetisierten Nukleosids $\text{D}_4\text{-dC}$ betrug 98,34 % und ist damit um 1,34 % isotopener als die Vorläuferverbindung $\text{D}_2^{15}\text{N}_2\text{-dC}$. Analog des Metabolismus aus Abbildung 17 wird das markierte Nukleosid in die genomische DNA eingebracht, durch die Dnmts methyliert und mittels der Tet-Dioxygenasen oxidiert, wodurch die jeweilige oxidierte $\text{D}_3\text{-dC}$ -Spezies entsteht.

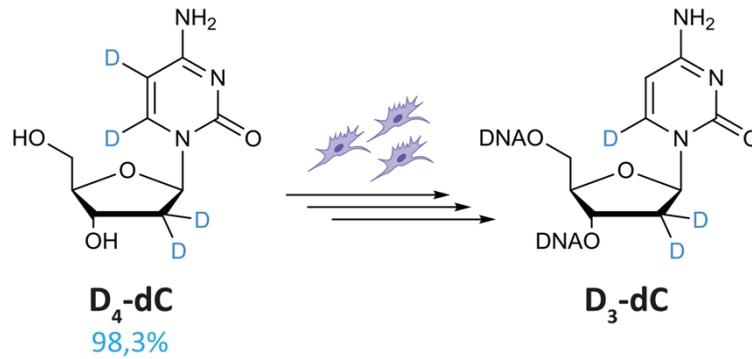


Abbildung 19: Verfolgung der aktiven Demethylierung mit Hilfe des isotopenmarkierten Nukleosids $\text{D}_4\text{-dC}$. Abgebildet ist die chemische Struktur des Moleküls und die, nach C-C Bindungsbruch, erwartete Verbindung $\text{D}_3\text{-dC}$. Die Reinheit des deuterierten Ausgangsmoleküls ist ebenfalls angegeben (hellblau).

Um einen direkten C-C Bindungsbruch ohne den Einfluss der BER auf die Entfernung von fdC und cadC betrachten zu können, wurde ein Experiment analog der $\text{D}_2^{15}\text{N}_2\text{-dC}$ Fütterung in *Tdg*-defiziente mESCs unternommen. Die verwendeten *Tdg*^{+/-} bzw. *Tdg*^{-/-} mESCs wurden durch Einbau einer Antibiotika-Resistenzkassette mittels homologer Rekombination aus der Parentallinie erzeugt.^[253] Durch Einbau dieser Kassette wurden Exon 6 und 7 des für *Tdg* kodierenden Gens gleichzeitig entfernt. Die *Tdg*^{+/-} Zellen enthalten das aktive Gen auf nur einem Allel. Durch kontinuierliche Steigerung der Konzentration des Antibiotikums, konnte aus der heterozygoten Zelllinie die *Tdg*^{-/-} Linie erhalten werden.^[253, 254] Die beiden Zelllinien wurden zunächst unter naiven 2i/LIF-Bedingungen für drei Tage kultiviert und dabei 200 μM $\text{D}_4\text{-dC}$ zugesetzt. Nach Ablauf der 72 h wurde ein Teil der Zellen geerntet und der andere Teil unter *Priming*-Bedingungen, sowie 200 μM $\text{D}_4\text{-dC}$ Zugabe, weiter kultiviert. Hierbei wurde die C/R-Bedingung zur Kultur der mESCs verwendet.^[255] Nach 72 h des *Primings* wurden die Zellen geerntet, die genomische DNA isoliert und diese massenspektrometrisch analysiert.

Die UHPLC-MS/MS-Ergebnisse der natürlichen und isotopenmarkierten dC-Derivate sind in Abbildung 20 dargestellt. Die Daten zeigen einen zweifachen Anstieg der natürlichen mdC-Werte in beiden Zelllinien unter *Priming*-Bedingungen. Weiterhin konnten in den *Tdg*^{+/-} Zellen etwa 39 % isotopenmarkiertes mdC pro natürlichem mdC in der DNA festgestellt werden. Im Falle der *Tdg*^{-/-} Zelllinie konnten 26 % mehr $\text{D}_3\text{-mdC}$ pro mdC detektiert werden. Es scheint als ob es sich entweder um einen Isotopeneffekt handelt oder die isotopenmarkierte Verbindung in der *Tdg*^{+/-} Zelllinie bevorzugt ausgeschnitten wird. Dieser Trend ist auch unter C/R-Bedingungen erkennbar. Bei der Betrachtung von hmdC fällt eine Abnahme der natürlichen Werte bei einem Vergleich von naiven und *primed* Kulturbedingungen auf, wobei erneut mehr $\text{D}_3\text{-hmdC}$ im Knockout zu detektieren ist. Ebenso finden sich unter naiven Bedingungen im *Tdg*^{-/-} erwartungsgemäß viermal höhere fdC Werte verglichen mit dem *Tdg*^{+/-}. Dieser Trend ist auch unter *Priming*-Bedingungen zu beobachten, wobei die Absolutwerte etwas geringer ausfallen. Das markierte fdC verhält sich entsprechend.

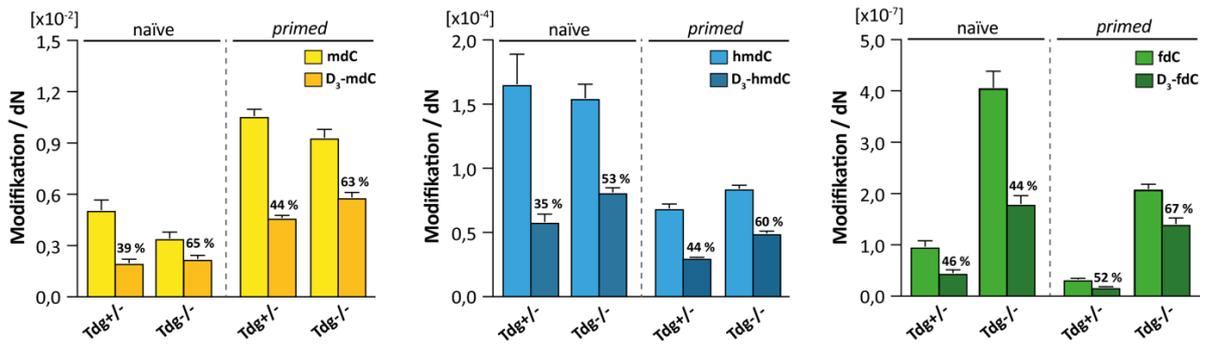


Abbildung 20: Absolute Quantifizierung der natürlichen und isotypenmarkierten Cytidin-Derivate nach D₄-dC Fütterung. Die Graphen zeigen die quantifizierten Werte aus *Tdg*^{+/-} sowie *Tdg*^{-/-} mESCs unter verschiedenen Kulturbedingungen. Die angegebenen Prozentzahlen beziehen sich auf die prozentuale Menge an isotypenmarkiertem D₃-mDC, D₃-hmdC bzw. D₃-fdC pro natürlichem mDC, hmdC bzw. fdC. Die Daten wurden aus einer biologischen Probe, gemessen als technisches Triplikate, erhoben.

Die generell prozentual etwas höhere Menge an isotypenmarkierter Spezies unter *Priming*-Bedingungen, könnte aus der Vorfütterung unter naiven Bedingungen resultieren. Die D₃-fdC, wie auch die natürlichen fdC-Werte befinden sich unter *Priming*-Bedingungen unter dem *Lower Limit of Quantification* (LLOQ), was die Messgenauigkeit beeinträchtigt. Die geringe Menge an analysierter genomischer DNA reicht hier nicht aus um fdC verlässlich zu quantifizieren. Größere Verdauansätze und eine Anpassung der Analyseverfahren könnten hier zu einem genaueren Ergebnis führen. Zusammenfassend konnte die D₄-dC Verbindung erfolgreich eingebaut und der Metabolismus der oxidierten dC Verbindungen verfolgt werden. Als nächstes sollte nun die Entstehung des C-C Bindungsbruchprodukts D₃-dC untersucht werden. Die ermittelte Markierungseffizienz und die jeweiligen prozentualen Anteile der markierten dC-Nukleoside bezogen auf alle deuterierten dC Spezies sind in Tabelle 1 dargestellt.

Tabelle 1: Verteilung der deuterierten Spezies in *Tdg*^{+/-} sowie *Tdg*^{-/-} unter verschiedenen Kulturbedingungen. Dargestellt ist die prozentuale Markierungsrate an D₄-dC bezogen auf alle dC-Nukleoside. Der prozentuale Anteil der jeweiligen deuterierten Verbindung ist bezogen auf die Summe (Σ) aller deuterierter dC-Spezies dargestellt. Die Daten wurden aus einer biologischen Probe, gemessen als technisches Triplikate erhoben.

		D ₄ -dC/dC [%]	D ₄ -dC/Σ [%]	D ₃ -dC/Σ [%]	D ₂ -dC/Σ [%]
naive	<i>Tdg</i> ^{+/-}	19,7	96,7	1,86	1,44
	<i>Tdg</i> ^{-/-}	28,5	97,2	1,84	0,95
primed	<i>Tdg</i> ^{+/-}	21,6	96,6	1,88	1,56
	<i>Tdg</i> ^{-/-}	25,7	97,1	1,89	1,02

Bei Betrachtung der D₄-dC-Spezies konnte eine Inkorporation von 19-28 % bezogen auf alle dC festgestellt werden. Weiterhin ist grundsätzlich eine höhere Inkorporation im *Knockout* zu beobachten. Beim Vergleich der prozentualen Menge an D₃-dC im *Tdg*^{+/-} und *Tdg*^{-/-} ist erneut eine geringe Anreicherung unter *Priming*-Bedingungen zu erkennen. Die ebenfalls detektierte D₂-dC-Spezies entsteht vermutlich aus der Verunreinigung oder durch einen spontanen D/H Austausch nach Angriff eines Nucleophils an der C6-Position, was zum Verlust des Deuteriums führen kann. Wird die C5-Position nun methyliert oder weiter oxidiert, entsteht eine D₂-Spezies, welche zudem schwerer vom natürlichen Isotopenhintergrund zu unterscheiden ist. Bei Betrachtung der Werte konnte festgestellt werden, dass die D₂-dC-Spezies beim Vergleich der beiden Zelllinien im *Knockout* etwas abnimmt.

Unter der Annahme, dass der C-C Bindungsbruch im Zuge der aktiven Demethylierung stattfindet, sollte eine Zelllinie als Negativkontrolle betrachtet werden, die kein mdC enthält und somit auch die Tet-Oxidationsprodukte nicht herstellen kann. Hierzu wurde eine *Dnmt Triple Knockout* mESC-Linie untersucht, die die DNA-Methyltransferasen Dnmt 1, 3a und 3b nicht enthält und mit dem korrespondierenden Wildtyp verglichen. Im Experiment wurden die mESCs unter naiven Bedingungen für 72 h mit der D₄-dC Spezies (100 µM) im Medium kultiviert. Ein Teil der Zellen wurde geerntet und die übrigen Zellen für weitere 72 h unter *Priming*-Bedingungen und Zugabe der isotope markierten Verbindung kultiviert. Im Anschluss wurde die DNA isoliert und das Genom auf die Menge an D₄-dC und D₃-dC analysiert. Zur Vollständigkeit wurde auch die Menge an D₂-dC bestimmt. In Tabelle 2 ist die Inkorporationsrate an D₄-dC bezogen auf alle dC-Nucleoside und die Quantifizierungsergebnisse als relative Menge aller deuterierter Spezies gezeigt.

Tabelle 2: Prozentualer Anteil der deuterierten Spezies in Dnmt TKO sowie Wildtyp mESCs unter verschiedenen Kulturbedingungen. Dargestellt ist die prozentuale Markierungsrate an D₄-dC bezogen auf alle dC-Nucleoside. Der prozentuale Anteil der jeweilige deuterierten Verbindung ist bezogen auf die Summe (Σ) aller deuterierter dC-Spezies dargestellt. Die Daten wurden aus zwei biologischen Replikaten, gemessen als technisches Triplikat erhoben.

		D ₄ -dC/dC [%]	D ₄ -dC/Σ [%]	D ₃ -dC/Σ [%]	D ₂ -dC/Σ [%]
naïve	Wildtyp	25,4	96,3	1,97	1,71
	Dnmt TKO	29,4	96,8	2,11	1,12
primed	Wildtyp	24,9	96,3	1,97	1,71
	Dnmt TKO	27,1	96,5	2,03	1,46

Die Inkorporationsrate an D₄-dC beträgt etwa 25-30 % was trotz der geringeren Fütterungskonzentration von 100 µM Fütterung bemerkenswert ist. Weiterhin ist die inkorporierte Menge an isotope markiertem Nucleosid abhängig von der jeweiligen Zelllinie. Die Inkorporation an D₄-dC ist im Wildtyp um 2 % geringer als im Dnmt TKO. Dies ist sowohl im naiven als auch im *primed* Zustand der Fall. Bei Betrachtung der D₃-dC Werte ist ein kleiner Anstieg in den jeweiligen Dnmt TKO Daten zu erkennen. Die Bildung von D₃-dC scheint daher unabhängig des aktiven Demethylierungsweges zu sein. Bei Betrachtung der D₂-dC Werte ist ein gegenläufiger Trend zu beobachten. Hier nimmt der Anteil an D₂-dC im TKO ab, was durch einen D/H-Austausch in Anwesenheit der Dnmts erklärt werden könnte. Geringe Unterschiede in den D₂-dC Werten können, aufgrund des natürlich vorkommenden Isotopenhintergrunds an ¹³C₂-dC, nicht verlässlich quantifiziert werden, was zu einer Verfälschung der Werte führen könnte. Weiterhin könnte der Vergleich von Dnmt TKO und Wildtyp zur Verfälschung der Ergebnisse aufgrund eines potentiellen D/H-Austausches an der C6-Position führen, weshalb eine weitere Zelllinie untersucht wurde.

Die Tet TKO Zelllinie enthält keine der drei Tet-Dioxygenasen Tet1, 2 und 3. Die Dnmts sind in beiden Zelllinien vorhanden und könnten in Wildtyp und TKO gleichermaßen zu einem D/H-Austausch beitragen. Die TKO-Linie dient als Negativkontrolle, da in diesen Zellen kein hmdC, fdC oder cadC entstehen kann. Ohne fdC bzw. cadC, sollte keine Deformylierung oder Decarboxylierung stattfinden können. Diese Linie wurde aus einer Wildtyp mESC-Linie mit Hilfe des CRISPR-Cas9 Systems erzeugt.^[256] Die Zellen wurden analog des Dnmt TKO Experiments kultiviert, markiert und geerntet. Tabelle 3 zeigt ebenso analog der Dnmt TKO Daten die prozentuale Verteilung der detektierten Spezies.

Tabelle 3: Prozentualer Anteil der deuterierten Spezies in Tet TKO sowie Wildtyp mESCs unter verschiedenen Kulturbedingungen. Dargestellt ist die prozentuale Markierungsrate an D₄-dC bezogen auf alle dC-Nucleoside. Der prozentuale Anteil der jeweilige deuterierten Verbindung ist bezogen auf die Summe (Σ) aller deuterierter dC-Spezies dargestellt. Die Daten wurden aus zwei biologischen Replikaten, gemessen als technisches Triplikat erhoben.

		D ₄ -dC/dC [%]	D ₄ -dC/Σ [%]	D ₃ -dC/Σ [%]	D ₂ -dC/Σ [%]
naïve	Wildtyp	26,3	96,5	2,02	1,50
	Tet TKO	28,3	96,7	2,02	1,30
<i>primed</i>	Wildtyp	25,4	96,4	2,01	1,63
	Tet TKO	29,7	96,8	2,09	1,08

Die Wildtyp und Tet TKO Zellen inkorporierten die deuterierte Verbindung mit 26-30 % aller dC Nucleoside. Zudem weisen die Tet TKO Zellen beim Vergleich mit dem Wildtyp eine höhere Inkorporationsrate auf.

Die Werte der D₃-dC Spezies sind nur im Tet TKO unter *Priming*-Bedingungen etwas höher. Die D₂-dC Verbindung zeigt, wie bereits im Dnmt TKO beobachtet, einen gegenläufigen Trend. Eine ausführlichere Analyse dieser Ergebnisse ist in der Dissertation von *Dr. Katharina Iwan* beschrieben.^[257]

Die hier gezeigten Daten stützen die in Abschnitt 3.1.1 gezeigten Beobachtungen, dass der C-C Bindungsbruch ohne Beteiligung der Tet und Dnmt Enzyme abläuft. Weiterhin ist unter *Priming*-Bedingungen meist eine Anreicherung des Demodifikationsprodukts D₃-dC zu beobachten. Allerdings lässt sich nicht ausschließen, dass die D₃-dC Verbindung auch über andere Wege wie einen D/H-Austausch oder durch bevorzugte Inkorporation der Verunreinigung entsteht. Weiterhin könnte die Existenz der D₂-dC Spezies, die ebenfalls aus einem C-C Bindungsbruch entstehen könnte, zu einer Verfälschung des Ergebnisses führen. Die beiden möglichen D₃-dC Spezies (Deuterium an der C5 oder C6-Position) konnten mit dieser Messmethode nicht unterschieden werden, was eine weitere Ungewissheit schafft. Aufgrund der zahlreichen entstandenen Spezies und der teils anspruchsvollen massenspektrometrischen Detektion, wurde ein neues Sondenmolekül entworfen.

Untersuchung der aktiven Demethylierung mit dem Sondenmolekül ¹³C₅D₁-dC

Die Molekülstruktur des verbesserten Sondenmoleküls ist in Abbildung 21 dargestellt. Alle C-Atome der Desoxyribose weisen eine ¹³C-Isotopenmarkierung auf und ein Deuterium befindet sich in 99,2 % Reinheit an der C5-Position der Base. Das Molekül ist demnach um sechs Masseinheiten schwerer als das natürliche Homolog und kann auf diese Weise gut vom natürlichen Isotopenhintergrund unterschieden werden. Das neue Sondenmolekül konnte mit einer verbesserten Reinheit hergestellt werden. Die Verunreinigung des potentiellen C-C Bindungsbruchprodukts ¹³C₅-dC beträgt nur noch 0,8 %.

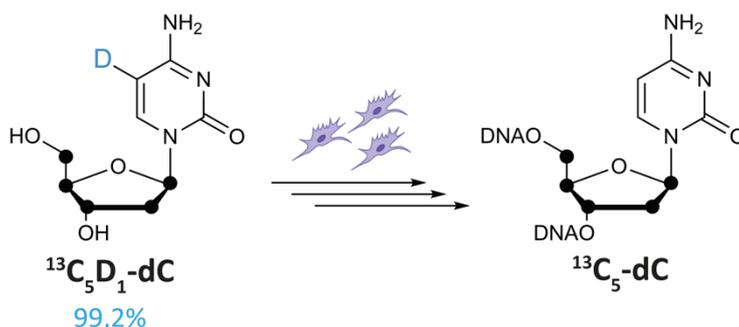


Abbildung 21: Verfolgung der aktiven Demethylierung mit Hilfe des isotopenmarkierten Nukleosids ¹³C₅D₁-dC. Abgebildet ist die chemische Molekülstruktur und das nach C-C Bindungsbruch erwartete Nukleosid ¹³C₅-dC. Die ausgefüllten Kreise zeigen die ¹³C-Markierung in der Desoxyribose. Die Reinheit des Ausgangsmoleküls ist ebenfalls angegeben (hellblau).

In bisherigen Experimenten wurde die Dynamik der markierten dC-Verbindungen außer Acht gelassen, weswegen im nächsten Schritt ein *Pulse-Chase* Experiment durchgeführt werden sollte. Hierzu sollte die Verdünnung der isotonenmarkierten dC-Verbindungen in Dnmt TKO mESCs und dem korrespondierenden Wildtyp im Zuge der DNA-Neusynthese verfolgt werden. Um die DNA-Synthese verfolgen zu können wurde dem Kulturmedium zusätzlich zu $100 \mu\text{M } ^{13}\text{C}_5\text{D}_1\text{-dC}$ auch $5 \mu\text{M } ^{13}\text{C}_{10}^{15}\text{N}_2\text{-dT}$ (dT^{+12}) beigemischt. Da Thymidin dafür bekannt ist, im geringeren Maße chemisch geschädigt zu werden und gute Markierungsraten von etwa 85 % zu ergeben, wurde dT^{+12} zur Verfolgung der DNA-Synthese gewählt.^[258] Zunächst wurden die Zellen für 36 h unter *Priming*-Bedingungen kultiviert, anschließend das Kulturmedium erneuert und die Isotopenmarkierung für 12 h zugegeben (*Pulse*). Nach dieser Zeit wurde nukleosidfreies Medium zugegeben (*Chase*) und die Zellen zu verschiedenen Zeitpunkten geerntet. Dies sollte den zeitlichen Rückgang der Isotopenmarkierung und Unterschiede zum Vorschein bringen. Der Ablauf des Experiments ist in Abbildung 22 schematisch dargestellt.

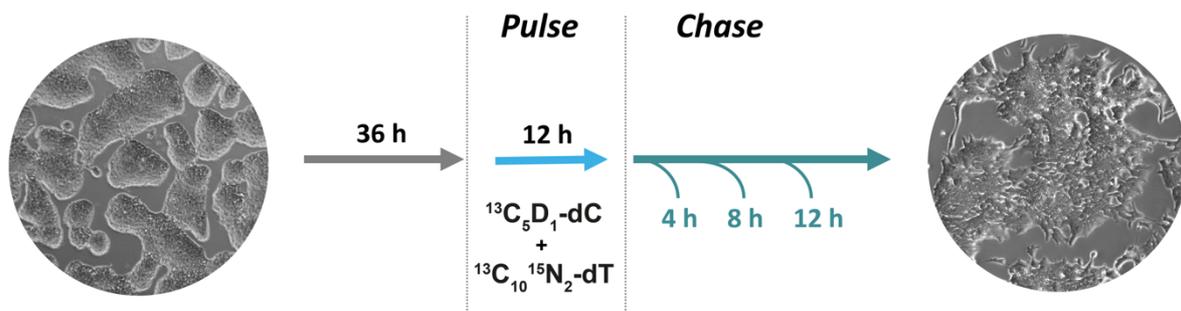


Abbildung 22: Ablauf des *Pulse-Chase* Experiments mit Dnmt TKO mESCs. Die Stammzellen werden zunächst für 36 h unter *Priming*-Bedingungen kultiviert. Nach Einbringung der isotonenmarkierten Nukleoside, erfolgt ein Wechsel des Kulturmediums. Anschließend werden die Zellen zu definierten Zeitpunkten geerntet. Die Mikroskopiebilder zeigen die mESCs unter naïve (links) und *primed* (rechts) Kulturbedingungen.

Die Verläufe der isotonenmarkierten Nukleoside sind in der nachfolgenden Abbildung 23 gezeigt. Die Menge an $^{13}\text{C}_5\text{D}_1\text{-dC}$ ist in beiden Zelltypen zu Beginn etwa gleich (siehe Abbildung 23A). Die Werte entsprechen einer Einbaurate von etwa 20 %. Auch die Abnahme findet in beiden Fällen etwa gleich schnell statt. Abbildung 23B zeigt den Abfall der $^{13}\text{C}_5\text{-dC}$ Spezies. Interessant ist hierbei, dass im Wildtyp, verglichen mit dem Dnmt TKO, nach 12 h etwa 35 % mehr $^{13}\text{C}_5\text{-dC}$ zu detektieren ist. Dies entspricht einer beachtlichen Menge von 1,3 % bezogen auf $^{13}\text{C}_5\text{D}_1\text{-dC}$ und ist ein weiterer Hinweis auf eine aktive Entstehung der $^{13}\text{C}_5\text{-dC}$ Verbindung. Nach der Hypothese ist der Wildtyp befähigt $^{13}\text{C}_5\text{D}_1\text{-dC}$ im Zuge der aktiven Demethylierung über eine Deformylierung oder Decarboxylierung in $^{13}\text{C}_5\text{-dC}$ umzuwandeln.

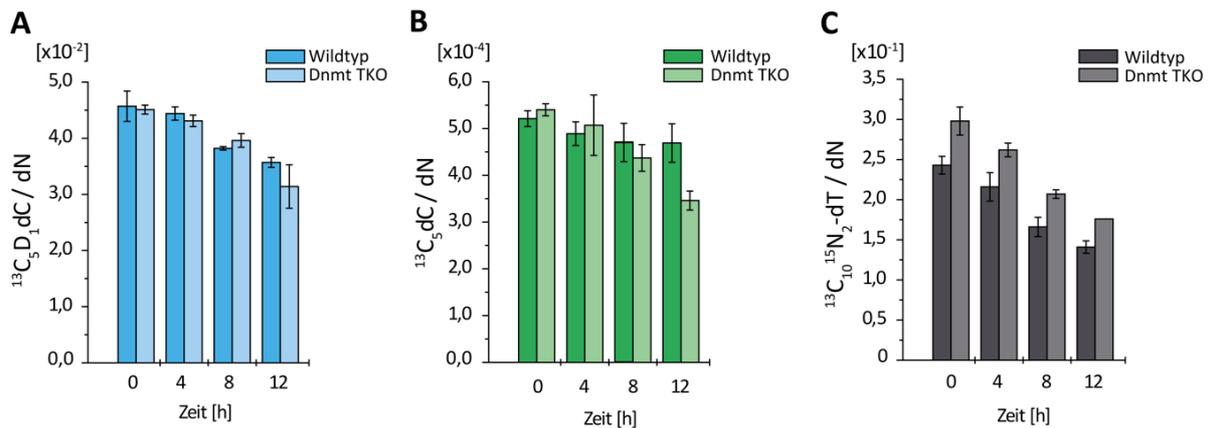


Abbildung 23: Ergebnisse des Pulse-Chase Experiments nach Fütterung der $^{13}\text{C}_5\text{D}_1\text{-dC}$ Spezies an pluripotente Zellen ohne die DNA-Methyltransferasen. **A)** Absolute Quantifizierung der $^{13}\text{C}_5\text{D}_1\text{-dC}$ Verbindung nach 0, 4, 8, 12 h im Dnmt TKO und dem korrespondierenden Wildtyp bezogen auf alle Nucleoside. **B)** Absolute Quantifizierung der $^{13}\text{C}_5\text{-dC}$ Verbindung im TKO und dem korrespondierenden Wildtyp. **C)** Quantifizierungsergebnisse der $^{13}\text{C}_{10}\text{N}_2\text{-dT}$ Verbindung in beiden Stammzelllinien. Die Daten wurden aus einer biologischen Probe, gemessen als technisches Triplikate ermittelt.

Im Dnmt TKO sollte dieser Weg nicht möglich sein, da der erste Schritt, die Methylierung durch Abwesenheit der Dnmt-Enzyme, nicht stattfinden kann. So ist die recht große Menge an $^{13}\text{C}_5\text{-dC}$ im Dnmt TKO verwunderlich, da die Spezies in diesen Zellen nicht generiert werden sollte. Es handelt sich somit entweder um die Inkorporation der Verunreinigung oder einen alternativen Weg der Demethylierung. Weiterhin liegen zum Zeitpunkt 0 h 87 % aller dT im Genom des Wildtyps markiert vor. Zudem ist der Dnmt TKO nahezu vollständig markiert (siehe Abbildung 23C). Im nächsten Schritt sollten die Werte der deuterierten Verbindungen auf dT^{+12} normiert werden, um die Verdünnung durch die DNA-Synthese mit einzubeziehen. Die folgende Abbildung 24 stellt die Normierung der eben beschriebenen Quantifizierungsergebnisse auf dT^{+12} dar. Die Menge an $^{13}\text{C}_5\text{D}_1\text{-dC}$ steigt bezogen auf dT^{+12} in beiden Zelltypen an. Es wäre anzunehmen, dass die Verbindung mindestens im gleichen Maße abnimmt wie das markierte dT und daher wäre weder ein Anstieg noch eine Abnahme zu erwarten. Vermutlich befindet sich nach 12 h noch eine zu große Menge an $^{13}\text{C}_5\text{D}_1\text{-dC}$ im löslichen Pool, das verzögert ins Genom eingebaut wird. Zugabe von unmarkiertem dC könnte diesen Effekt abschwächen. Interessant ist der Trend der normierten $^{13}\text{C}_5\text{-dC}$ Werte. Im Dnmt TKO bleiben die Werte etwa konstant, während im Wildtyp eine deutliche Zunahme der $^{13}\text{C}_5\text{-dC}$ Verbindung zu erkennen ist. Im Laufe der 12 h wächst der Anteil an $^{13}\text{C}_5\text{-dC}$ im Wildtyp bezogen auf den 0 h Zeitpunkt um etwa 55 % an. Im Dnmt TKO ist nur ein kleiner Anstieg von etwa 16 % zu verzeichnen. Die Zunahme an $^{13}\text{C}_5\text{-dC}$ im Wildtyp spricht für eine aktive Demethylierung über einen C-C Bindungsbruch. Weiterhin zeigen die Daten die Notwendigkeit einer Normierung auf die DNA-Synthese, da die beiden Zelltypen die markierten Spezies in unterschiedlichen Mengen einbauen.

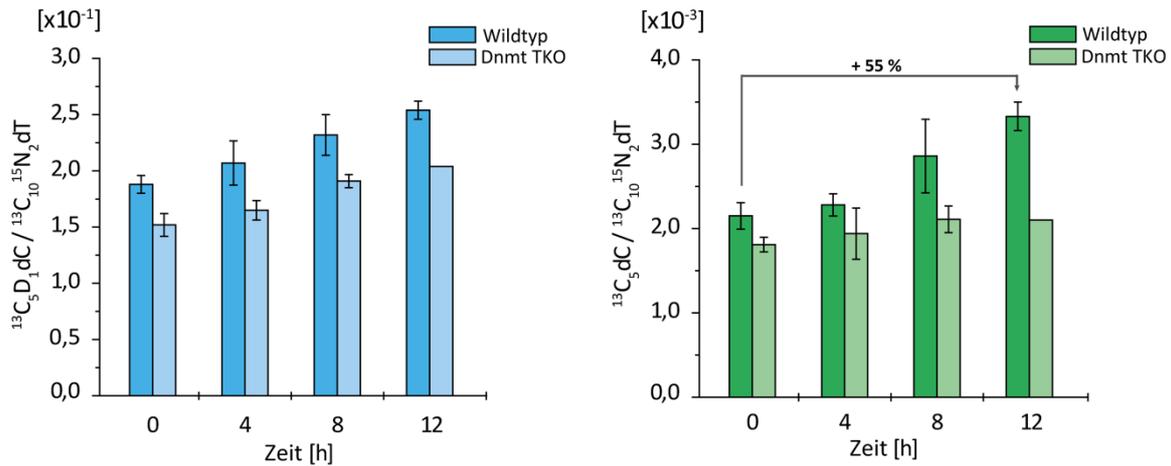


Abbildung 24: Normierung der Quantifizierungsergebnisse des der *Pulse-Chase* Experiments auf $^{13}\text{C}_{10}\text{-}^{15}\text{N}_2\text{-dT}$ in Abwesenheit der Dnmt-Enzyme. Absolute Quantifizierung der $^{13}\text{C}_5\text{D}_1\text{-dC}$ und $^{13}\text{C}_5\text{-dC}$ Mengen pro isotopenmarkiertem dT in Wildtyp und Dnmt TKO mESCs. Die angegebenen Zeitpunkte entsprechend den Erntezeitpunkten der Zellen im Laufe des *Chase*. Die Daten wurden aus einer biologischen Probe, gemessen als technisches Triplikat erhoben.

Das entsprechende Experiment wurde mit der Tet TKO mESC Linie und dem korrespondierenden Wildtyp ebenfalls durchgeführt. Diese Zellen können die markierte Spezies zwar methylieren aber nicht weiter oxidieren. Dieses Experiment soll nochmals Aufschluss über die Beteiligung eines D/H-Austausches geben und die Abhängigkeit der beobachteten Reaktion von den Tet-Dioxygenasen betrachten. Wird die Entstehung der $^{13}\text{C}_5\text{-dC}$ Verbindung durch die Dnmt-Enzyme katalysiert, so sollte dies in beiden Zellen nun im gleichen Maße stattfinden. Entsteht die Verbindung durch einen C-C Bindungsbruch, so sollte das Ergebnis dem des Dnmt TKOs ähneln. Abbildung 25 zeigt die erhaltenen Werte der isotopenmarkierten Verbindungen normiert auf das schwere dT⁺¹².

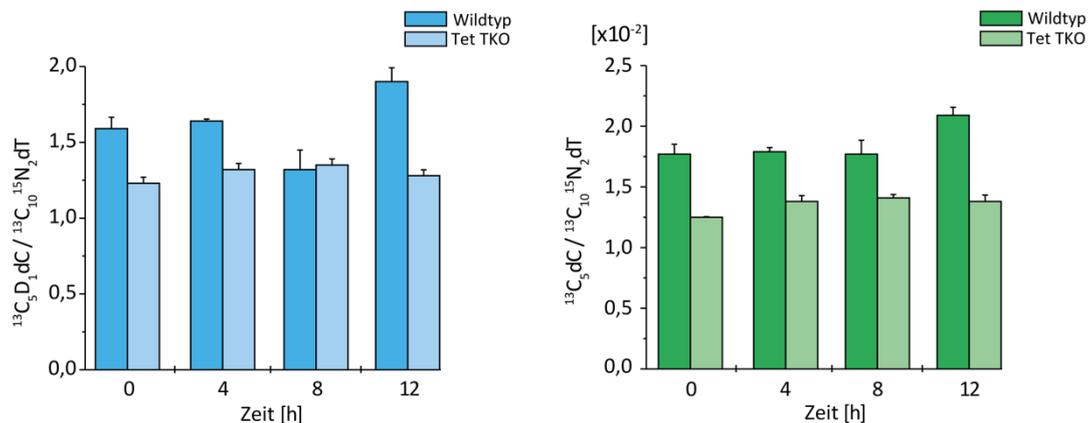


Abbildung 25: Normierung der Quantifizierungsergebnisse des *Pulse-Chase* Experiments auf $^{13}\text{C}_{10}\text{-}^{15}\text{N}_2\text{-dT}$ in Abwesenheit der Tet-Enzyme. Absolute Quantifizierung der $^{13}\text{C}_5\text{D}_1\text{-dC}$ Mengen pro isotopenmarkiertem dT⁺¹² in Wildtyp und Tet TKO mESCs. Die angegebene Zeitdauer, entspricht den Erntezeitpunkten der Zellen im Laufe des *Chase*. Die Daten wurden aus einer biologischen Probe, gemessen als technisches Triplikat erhoben.

Im Wildtyp wird mehr $^{13}\text{C}_5\text{D}_1\text{-dC}$ eingebaut und ist auch nach 12 h noch in einer größeren Menge im Genom vorhanden. Demnach ist ein Anstieg an $^{13}\text{C}_5\text{-dC}$ im Wildtyp zu diesem Zeitpunkt nicht verwunderlich. Die Werte beider markierter Verbindungen bleiben im Tet TKO etwa gleich. Die Generierung des $^{13}\text{C}_5\text{-dC}$ sollte im Tet-TKO nicht über den angenommenen Weg der aktiven Demethylierung ablaufen und müsste als Hintergrund abgezogen werden. Demnach würde die Zunahme an $^{13}\text{C}_5\text{-dC}$ im Wildtyp für einen C-C Bindungsbruchreaktion im Zuge der aktiven Demethylierung sprechen.

Untersuchung der aktiven Demethylierung mit den Sondenmolekülen $^{13}\text{C}_5\text{D}_1\text{-dC}$ und $^{13}\text{C}_5\text{D}_1\text{-rC}$

Die bisher genutzten Sondenmoleküle weisen keine vollständige Isotopenreinheit auf, weswegen nicht klar ist, ob der gesehene Effekt tatsächlich aus einer C-C Bindungsbruchreaktion oder einer Verunreinigung kommt. In Abschnitt 3.2.3.2 wurde bereits festgestellt, dass RNA-Nukleoside zu einer höheren Inkorporationsrate führen. Der RNA-Baustein wird über den *RNA Salvage Pathway* in die genomische DNA eingebaut. Da die Ribonukleotidreduktase auf Diphosphat-Ebene arbeitet, wird der Weg über die löslichen Desaminasen größtenteils umgangen. Im nächsten Schritt erfolgt direkt die Phosphorylierung zum Triphosphat, wodurch vermutlich mehr markiertes Nukleosid eingebaut wird. Als neues Sondenmolekül wurde daher das RNA-Derivat der $^{13}\text{C}_5\text{D}_1\text{-dC}$ Verbindung von *Dr. René Rahimoff* synthetisiert.^[259] Weiterhin wurde eine neue UHPLC-MS/MS Detektionsmethode von *Dr. Katharina Iwan* entwickelt.^[257] Die chemische Struktur und die potentiell aus einem C-C Bindungsbruch entstehende Verbindung sind in Abbildung 26 dargestellt. Schließlich wurde $^{13}\text{C}_5\text{D}_1\text{-rC}$ mit einer Isotopenreinheit von 99,7 % synthetisiert und war demnach nur zu 0,3 % an $^{13}\text{C}_5\text{-rC}$ verunreinigt.

Die *Pulse-Chase* Experimente in Dnmt TKO und Tet TKO mESCs wurden mit dieser Verbindung wiederholt. Die Inkorporation konnte durch Fütterung der Ribo-Verbindung auf etwa 43 % $^{13}\text{C}_5\text{D}_1\text{-dC}$ erhöht werden, was einer doppelt so hohen Menge, verglichen mit der Desoxy-Verbindung entspricht. Die Ergebnisse dieses Experiments deuten ebenfalls auf die Existenz eines C-C Bindungsbruchs hin. Die Daten und eine genauere Diskussion dieser, sind jedoch in der Dissertation von *Dr. Katharina Iwan* aufgeführt.^[140]

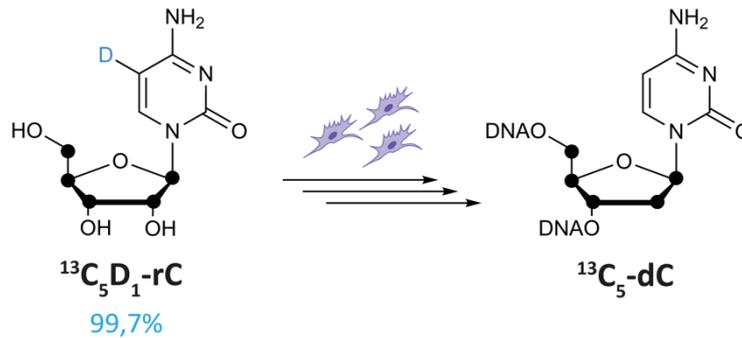


Abbildung 26: Verfolgung der aktiven Demethylierung mit Hilfe des isotonenmarkierten Nukleosids $^{13}\text{C}_5\text{D}_1\text{-rC}$. Abgebildet ist die chemische Struktur des Sondenmoleküls und die durch C-C Bindungsbruch entstandene Verbindung $^{13}\text{C}_5\text{-dC}$. Die ausgefüllten Kreise zeigen die ^{13}C -Markierung. Die Reinheit des Ausgangsmoleküls ist ebenfalls angegeben (hellblau).

In einem weiteren Experiment wurden die bereits beschriebenen $Tdg^{+/-}$ und $Tdg^{-/-}$ mESCs mit der isotonenmarkierten Spezies gefüttert, um erneut die C-C Bindungsbruchreaktion in Abwesenheit des Tdg/BER-Demethylierungsweges betrachten zu können. Hierzu wurden die Zellen für drei Tage unter naiven Bedingungen kultiviert. Ein Teil der Zellen wurde geerntet und der andere nochmals für drei Tage mit der Ribo- bzw. Desoxy-Verbindung unter *Priming*-Bedingungen kultiviert. Für die Fütterung wurde stets eine Konzentration von $200\ \mu\text{M}$ $^{13}\text{C}_5\text{D}_1\text{-dC}$ bzw. $50\ \mu\text{M}$ $^{13}\text{C}_5\text{D}_1\text{-rC}$ zum jeweiligen Kulturmedium zugegeben. Nach Isolation der genomischen DNA wurde diese massenspektrometrisch analysiert. Die Ergebnisse sind in Tabelle 4 dargestellt.

Tabelle 4: Verteilung der markierten Spezies in $Tdg^{+/-}$ sowie $Tdg^{-/-}$ unter naiven und *primed* Kulturbedingungen. Dargestellt ist Markierungsrate an $^{13}\text{C}_5\text{D}_1\text{-dC}$ bezogen auf alle dC-Nukleoside. Der prozentuale Anteil der jeweiligen isotonenmarkierten Verbindung ist bezogen auf die Summe (Σ) beider markierter dC-Spezies dargestellt. Die Daten wurden aus einer biologischen Probe, gemessen als technisches Triplikat erhoben.

			$^{13}\text{C}_5\text{D}_1\text{-dC/dC}$ [%]	$^{13}\text{C}_5\text{D}_1\text{-dC}/\Sigma$ [%]	$^{13}\text{C}_5\text{-dC}/\Sigma$ [%]
naive	$^{13}\text{C}_5\text{D}_1\text{-dC}$	$Tdg^{+/-}$	29,8	99,70	0,30
		$Tdg^{-/-}$	37,1	99,68	0,32
	$^{13}\text{C}_5\text{D}_1\text{-rC}$	$Tdg^{+/-}$	45,1	99,68	0,32
		$Tdg^{-/-}$	50,4	99,68	0,32
primed	$^{13}\text{C}_5\text{D}_1\text{-dC}$	$Tdg^{+/-}$	31,5	99,68	0,32
		$Tdg^{-/-}$	36,0	99,72	0,28
	$^{13}\text{C}_5\text{D}_1\text{-rC}$	$Tdg^{+/-}$	46,0	99,65	0,35
		$Tdg^{-/-}$	46,8	99,66	0,34

Die Markierungseffizienz nach Fütterung der $^{13}\text{C}_5\text{D}_1\text{-dC}$ Spezies beträgt unter naive Kulturbedingungen etwa 30-37 %, wohingegen nach $^{13}\text{C}_5\text{D}_1\text{-rC}$ Fütterung Werte von 46-50 % erreicht werden konnten. Ein ähnliches Ergebnis lieferte die Markierungsrate unter *Priming*-Bedingungen. Demnach ist die Art der Kulturbedingung nicht ausschlaggebend für die Markierungseffizienz. Wird die $^{13}\text{C}_5\text{-dC}$ Spezies bezogen auf die Summe der markierten dC-Nukleoside betrachtet, so ist keine auffällige Änderung des Wertes unter den analysierten Bedingungen zu beobachten. Auch die Steigerung der Markierungseffizienz nimmt keinen Einfluss auf die $^{13}\text{C}_5\text{-dC}$ Werte.

Nach diesen Ergebnissen wird klar, dass es trotz Abwesenheit des Tdg/BER-Prozesses zu keinem deutlichen Anstieg des C-C Bindungsbruchproduktes kommt. Ebenso hat die Kulturbedingung, bei Betrachtung dieser mESC-Linie, keinen Einfluss auf die $^{13}\text{C}_5\text{-dC}$ Werte. Zusammenfassend scheinen auf diese Weise kultivierte mESCs kein geeignetes Modellsystem zum eindeutigen Nachweis eines C-C Bindungsbruchs zu sein, was auf die hohe Teilungsrate und die kontinuierliche Inkorporation der markierten Verbindung zurückzuführen sein könnte.

Untersuchung der aktiven Demethylierung in postmitotischen Zellen

Die bisherigen *Pulse-Chase* Experimente lassen die Frage offen inwieweit Rückstände der deuterierten Spezies im löslichen Nukleosid/-tid Pool die erhaltenen Daten beeinflussen. Eine kontinuierliche Neubildung sowie Re-Metabolisierung der $^{13}\text{C}_5\text{-dC}$ Verbindung könnte im Gleichgewichtszustand ein signifikantes Ergebnis verdecken. Dazu kommt, dass sich der zu beobachtende Prozess komplementär zur passiven Demethylierung verhält und die genutzten Sondenmoleküle nur in sich aktiv teilende Zellsysteme eingebracht werden können. Daher wurde für die nachfolgende Studie ein Modellsystem gewählt, wo der Übergang von proliferierenden zu postmitotischen Zellen aktiv herbeigeführt werden kann. So könnte die Markierungs- von der Beobachtungsphase separiert werden und möglicherweise eine vermehrte Aktivität des aktiven Prozesses nachgewiesen werden.

Als Modellsystem wurden die humanen, induzierten pluripotenten Stammzellen (iPSCs) gewählt. An diese Zellen wurden sowohl die dC- als auch die rC-Verbindung gefüttert und eine neuronale Differenzierung zum Erhalt postmitotischer Neuronen ausgelöst. Die Neurogenese wurde in diesen iPSCs durch eine bicistronische, Doxycyclin-induzierbare Neurogenin-1 und 2 Expressionskassette induziert. Die resultierenden Zellen werden als *small molecule-inducible neurogenin iPS-line* (iNGNs) bezeichnet und bilden bereits nach vier Tagen eine bipolare Morphologie mit langen Neuriten aus. Neuronen mit anregbaren Aktionspotentialen werden nach zwei Wochen in Kultur beobachtet.^[260] Kurz nach Induktion der Differenzierung nimmt die Wachstumsrate der Zellen stark ab. Dies wurde für postmitotische Neuronen bereits beobachtet und scheint damit ein geeignetes Modellsystem für das geplante Experiment zu sein.^[261]

In Zusammenarbeit mit *Dr. Jessica Steinbacher* wurden bereits die Level der natürlichen oxidierten dC-Spezies massenspektrometrisch quantifiziert. Im Laufe der Differenzierung konnte eine Zunahme an hmdC beobachtet werden. Die fdC und cadC-Level blieben während der Differenzierung konstant und betragen etwa $1,0 \times 10^{-6}$ bzw. $2,5 \times 10^{-7} / \text{dN}$.^[249] In den ersten 24 h nach Induktion konnte ein geringer Abfall der Werte für beide Spezies beobachtet werden, was vermutlich auf die noch stattfindende DNA-Synthese zurückzuführen ist. Erst nach etwa 48 h findet in diesen Zellen keine DNA-Neusynthese mehr statt.^[249] Dieses Zellsystem scheint sich gut zur Beobachtung eines C-C Bindungsbruchs zu eignen, da ein Neueinbau der markierten Spezies nach 48 h nicht mehr möglich ist und die oxidierten dC-Verbindungen gut detektierbar sind. Der Ablauf des Fütterungsexperiments wird in Abbildung 27 gezeigt.

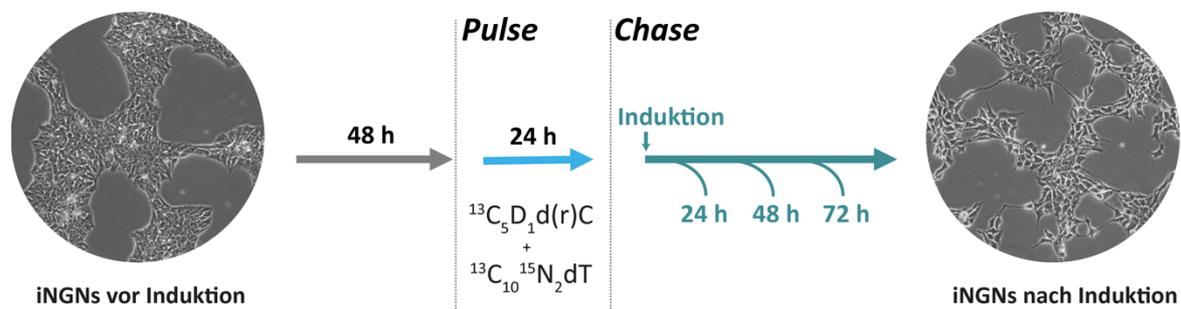


Abbildung 27: Experimenteller Ablauf zum Nachweis der C-C-Bindungsbruchreaktion in iNGNs nach neuronaler Differenzierung. Die Mikroskopiebilder zeigen die Zellen vor und nach Induktion. Die Zellen wurden für 24 h markiert und die Neurogenese nach Wechsel des Kulturmediums begonnen.

Die Zellen wurden zunächst ausgesät und für 48 h kultiviert. Anschließend wurde dem Kulturmedium jeweils $140 \mu\text{M}$ $^{13}\text{C}_5\text{D}_1\text{-dC}$ bzw. $50 \mu\text{M}$ $^{13}\text{C}_5\text{D}_1\text{-rC}$ sowie jeweils $0,5 \mu\text{M}$ dT^{+12} zugesetzt. Nach 24 h wurde das Kulturmedium gegen nukleosidfreies Medium ausgetauscht und die Induktion durch Zugabe von Doxycyclin gestartet. Die Zellen wurden vor Induktion und zu den Zeitpunkten 24, 48 und 72 h nach Induktion geerntet. Nach Isolation der genomischen DNA wurde diese massenspektrometrisch untersucht. In Abbildung 28 ist die prozentuale Verteilung aller dC-Spezies nach Markierung durch die beiden deuterierten Nucleoside zu den unterschiedlichen Erntezeitpunkten illustriert. Die Daten zeigen einen deutlichen Unterschied in der Markierungseffizienz der beiden gefütterten Nucleoside.

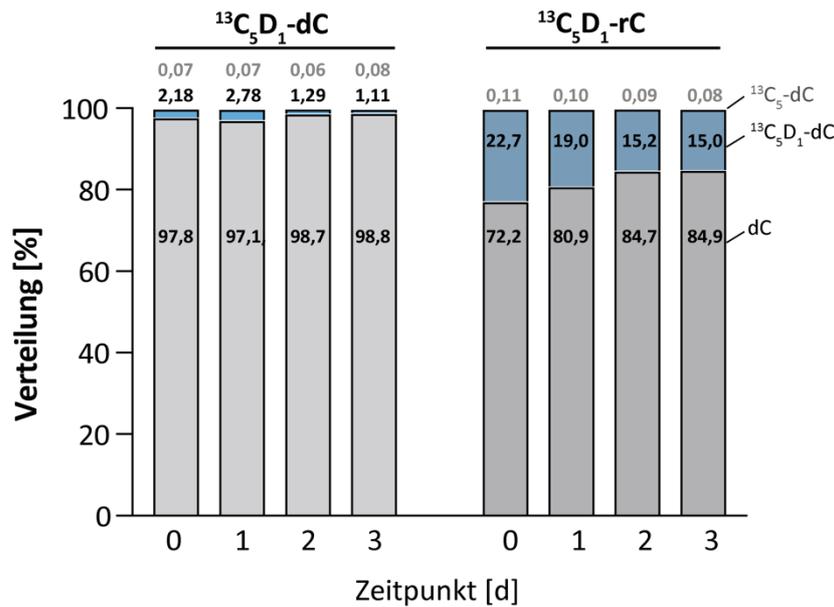


Abbildung 28: Prozentuale Verteilung der gemessenen isotopenmarkierten Spezies nach Induktion der iNGNs.

Prozentuale Darstellung aller gemessener Cytidin-Spezies nach Fütterung der beiden Sondenmoleküle. Die Zeitpunkte beziehen sich auf den Beginn der neuronalen Differenzierung. Die erhaltenen Daten beziehen sich auf eine biologische Probe gemessen als technisches Triplikat.

Trotz geringerer Fütterungskonzentration wird das deuterierte rC-Nukleosid etwa zehnmal effizienter aufgenommen. Zum Zeitpunkt 0 h ist eine Markierungsrate von 2,2 % für die $^{13}\text{C}_5\text{D}_1\text{-dC}$ Verbindung und etwa 23 % für die rC-Verbindung zu beobachten. Die zeitliche Abnahme nach Induktion im Zuge des *Chase* ist ebenfalls gut zu sehen. Der Verlust an Markierung ist erwartungsgemäß in den ersten beiden Tagen am deutlichsten. Dies bestätigt, dass in den Zellen nach 48 h keine DNA-Synthese mehr stattfindet und eine passive Verdünnung der dC-Spezies ausbleibt. Die resultierende $^{13}\text{C}_5\text{-dC}$ Verbindung nimmt nach Fütterung von $^{13}\text{C}_5\text{D}_1\text{-dC}$ im Laufe der Differenzierung (72 h) um 0,01% zu, während sie im Fall der rC-Verbindung minimal abnimmt.

Als nächstes wurden die erhaltenen Quantifizierungsdaten auf alle Nukleoside bezogen. In Abbildung 29 sind die normierten absoluten Quantifizierungsergebnisse der zu betrachtenden Verbindungen resultierend aus den beiden Fütterungen dargestellt. Die Quantifizierungsergebnisse der $^{13}\text{C}_5\text{D}_1\text{-dC}$ Level zeigen eine kontinuierliche Abnahme in beiden Fällen (siehe Abbildung 29A). Ab dem 48 h Zeitpunkt scheinen die Werte jedoch stabil zu bleiben. Auch der große Unterschied in der Markierungsrate kommt zum Vorschein. Die absoluten Werte der beiden Testmoleküle zum 0 h Zeitpunkt unterscheiden sich etwa um das Zehnfache. Für die absolute Quantifizierung der $^{13}\text{C}_5\text{-dC}$ Verbindung ergeben sich unterschiedliche Trends (siehe Abbildung 29B). Während die Werte nach rC-Fütterung kontinuierlich abnehmen, ist nach dC-Fütterung nach drei Tagen Induktion eine Zunahme zu sehen. Die $^{13}\text{C}_5\text{-mdC}$ Werte verhalten sich ähnlich der $^{13}\text{C}_5\text{D}_1\text{-dC}$ Level und nehmen kontinuierlich ab (siehe Abbildung 29C).

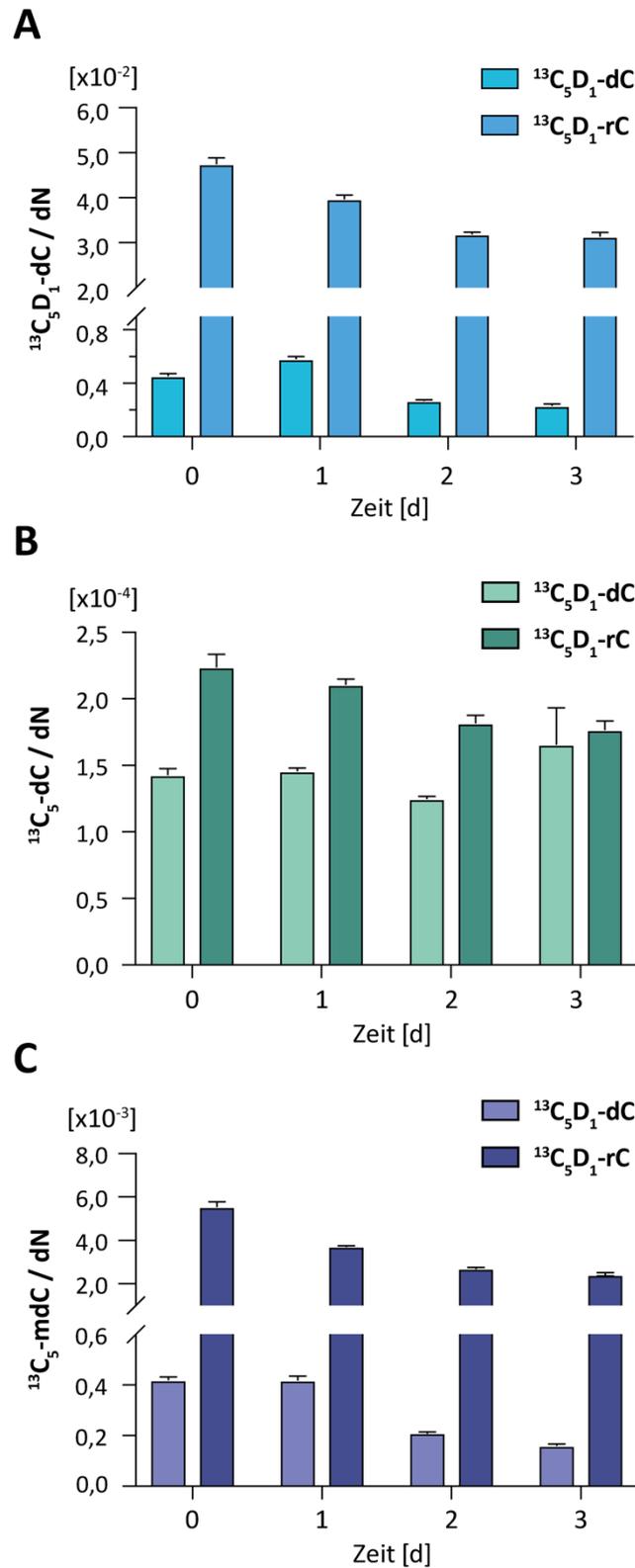


Abbildung 29: Absolute Quantifizierung der isotonenmarkierten dC-Derivate aus induzierten iNGNs. Die Daten zeigen die Messergebnisse nach Fütterung von $^{13}\text{C}_5\text{D}_1\text{-dC}$ bzw. $^{13}\text{C}_5\text{D}_1\text{-rC}$ an iNGNs zu unterschiedlichen Zeitpunkten nach neuronaler Differenzierung. Die Daten wurden aus einer biologischen Probe, gemessen als technisches Triplikat erhoben.

Zur besseren Vergleichbarkeit wurden die absoluten Quantifizierungsergebnisse der dC-Derivate auf die Menge an isopenmarkiertem dT^{+12} normiert, welches ebenfalls gefüttert wurde. Dies sollte Hinweise auf eine übrige, stattfindende DNA-Synthese und daraus resultierende passive Verdünnung der jeweiligen Spezies geben. Die Normierung der $^{13}C_5D_1$ -dC Werte ist in Abbildung 30A dargestellt und Abbildung 30B zeigt die Ergebnisse nach $^{13}C_5D_1$ -rC Markierung. Rechts in der Abbildung ist jeweils die Menge an markiertem dT^{+12} , auf alle Nucleoside bezogen, gezeigt. Mit Ausnahme des 24 h Zeitpunkts der dC-Fütterung, nehmen die Werte erwartungsgemäß kontinuierlich ab.

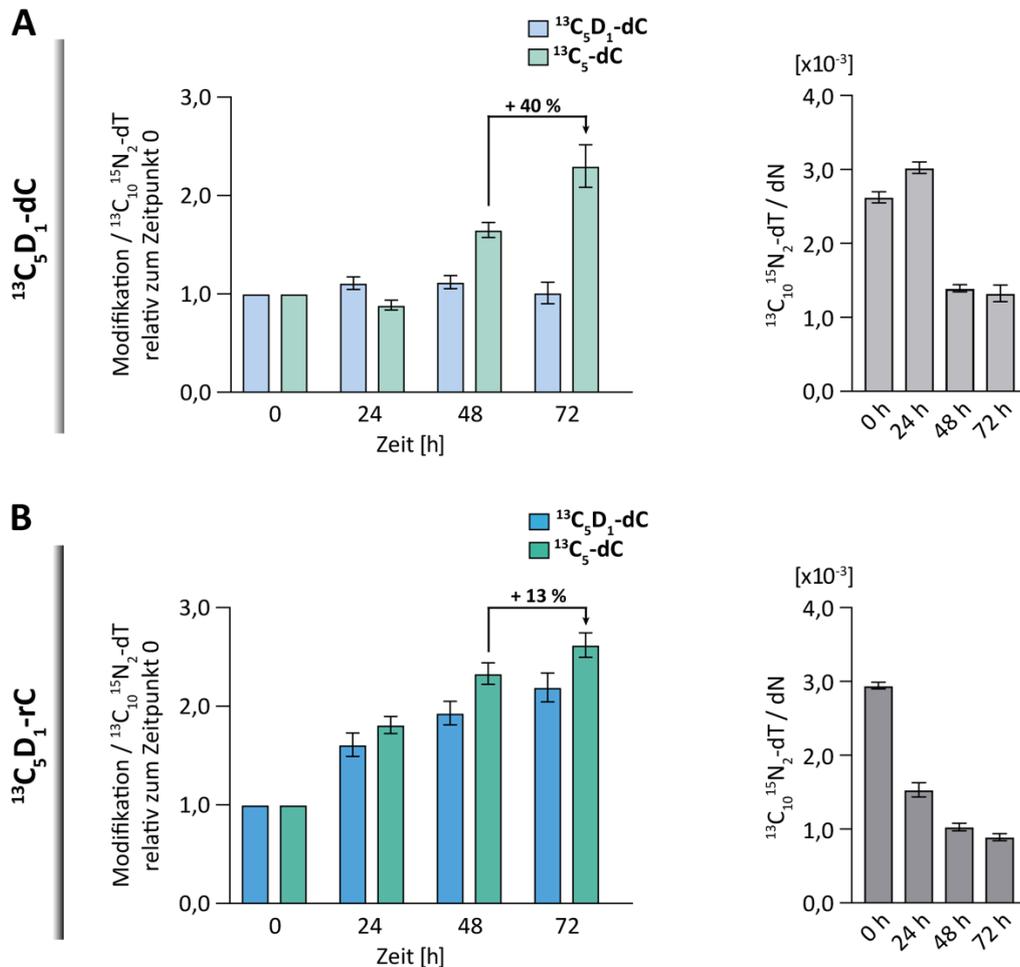


Abbildung 30: Verfolgung einer C-C Bindungsbruchreaktion nach neuronaler Differenzierung der iNGNs.

A) Normierte Quantifizierungsergebnisse nach Fütterung von $^{13}C_5D_1$ -dC an iNGNs zu unterschiedlichen Zeitpunkten nach Induktion. **B)** Normierte Quantifizierungsergebnisse nach Fütterung von $^{13}C_5D_1$ -rC an iNGNs zu unterschiedlichen Zeitpunkten nach Induktion (eine biologische Probe wurde als technisches Triplikat vermessen).

Nach Normierung bleiben die $^{13}C_5D_1$ -dC/ dT^{+12} Verhältnisse nahezu konstant. Die $^{13}C_5D_1$ -dC-Verbindung nimmt durch passive Verdünnung im gleichen Maße ab wie das markierte dT^{+12} . Nach der Arbeitshypothese entsteht die $^{13}C_5$ -dC Verbindung durch eine C-C-Bindungsbruchreaktion der metabolisierten $^{13}C_5D_1$ -dC Spezies.

Die Normierung zeigt einen deutlichen Anstieg der $^{13}\text{C}_5\text{-dC}$ Verbindung nach 48 bzw. 72 h in der mit dC-gefütterten Kultur. Der Wert steigt um etwa 40 % in den letzten 24 h des Experiments. An diesem Zeitpunkt findet nahezu keine DNA-Synthese mehr statt, wie die dT^{+12} -Werte zeigen. Diese Beobachtung spricht für eine C-C Bindungsbruchreaktion in diesen postmitotischen neuronalen Vorläufern. Bei Betrachtung des Experiments nach $^{13}\text{C}_5\text{D}_1\text{-rC}$ Markierung ist eine Korrelation der $^{13}\text{C}_5\text{D}_1\text{-dC}$ und $^{13}\text{C}_5\text{-dC}$ Verbindungen nach Normierung auf das isotopenmarkierte dT^{+12} zu beobachten. Beide markierten Spezies nehmen im Zuge der Differenzierung zu. Es scheint, als ob die Verbindung trotz des Rückgangs an DNA-Synthese stetig in die genomische DNA eingebaut wird. Vermutlich ist diese Beobachtung auf den Verbleib an markiertem Material im löslichen Nukleosid/-tid-Pool zurückzuführen. Ebenso könnte die Verbindung $^{13}\text{C}_5\text{D}_1\text{-rC}$ über den RNA-Metabolismus kontinuierlich freigesetzt werden und anschließend über die Ribonukleotidreduktase zu $^{13}\text{C}_5\text{D}_1\text{-dC}$ umgewandelt werden oder im Zuge der DNA-Reparatur ins Genom eingebaut werden, was den Anstieg in Abwesenheit der DNA-Synthese erklären könnte. Ebenso könnte es sich um Spuren mitochondrialer DNA handeln. Dies verhindert, dass die Vorteile dieses Experiments überhaupt zum Tragen kommen. Zusammengefasst zeigt im Besonderen die $^{13}\text{C}_5\text{D}_1\text{-dC}$ Fütterung der postmitotischen, neuronalen Vorläuferzellen Hinweise auf eine aktive Entstehung und Akkumulation des hypothetischen C-C Bindungsbruchprodukts $^{13}\text{C}_5\text{-dC}$ und sollte daher wiederholt werden.

3.1.2.3 Zusammenfassung und Ausblick

Zusammengefasst wurden durch Markierung pluripotenter Stammzelllinien mit deuterierten Sondenmolekülen weitere Hinweise auf die Beteiligung einer direkten Demodifikation an der DNA-Demethylierung gewonnen. Es stellte sich heraus, dass sich die Sondenmoleküle $^{13}\text{C}_5\text{D}_1\text{-dC}$ und $^{13}\text{C}_5\text{D}_1\text{-rC}$ am besten zum Nachweis des C-C Bindungsbruchs eignen. Durch diese Verbindungen konnten erstmals konkrete Hinweise auf eine direkte Demodifikation in Form einer Deformylierung oder Decarboxylierung erbracht werden. *Pulse-Chase* Experimente in *Dnmt* pro- und defizienten *primed* mESCs zeigten einen deutlichen Anstieg des Demodifikationsproduktes. In Tet TKO und dem korrespondierenden Wildtyp ist dies erkennbar, jedoch weniger stark ausgeprägt. Um die aktive Generierung des Demodifikationsproduktes final zu bestätigen, sollten die Proben allerdings an einem hochauflösenden Massenspektrometer vermessen werden, da der Massenunterschied zwischen Ausgangsmaterial und Demodifikationsprodukt nur eine Masseneinheit beträgt. Weiterhin konnten konkrete Hinweise auf die Existenz des C-C Bindungsbruchs nach Markierung von iNGNs erhalten werden. Im Verlauf einer neuronalen Differenzierung wurde hier ein Anstieg des Demodifikationsproduktes beobachtet, was trotz der Limitierungen des Massenspektrometers und der 0,8 %igen Verunreinigung an $^{13}\text{C}_5\text{-dC}$, Hinweise auf einen C-C Bindungsbruch in diesen Zellen liefert.

Im weiteren Verlauf sollten weitere biologische Replikate der iNGN-Fütterung mit $^{13}\text{C}_5\text{D}_1\text{-dC}$ bzw. rC durchgeführt werden. Interessant wäre hierbei auch eine Reduktion der rC -Konzentration auf etwa $25\ \mu\text{M}$ oder weniger, um eine Akkumulation der Nukleoside im löslichen Nukleosid/-tid Pool zu verringern. Eventuell könnten so Ergebnisse ähnlich der $^{13}\text{C}_5\text{D}_1\text{-dC}$ Markierung erzielt werden. Weiterhin sollten das Verdauprotokoll und die Messmethode zur Detektion der markierten hmdC bzw. fdC Verbindungen optimiert werden. Zur Betrachtung eines BER-unabhängigen Deformylierungs- bzw. Decarboxylierungsweges könnte die $\text{Tdg}^{-/-}$ mESC-Linie in EpiLCs differenziert werden. Dies würde eine homogenere Zellkultur mit höheren Leveln der natürlichen oxidierten dC -Spezies ergeben und könnte ein vielversprechendes Modell zum Nachweis des C-C Bindungsbruchs in Abwesenheit des Tdg /BER-Demethylierungsweges darstellen. Weiterhin könnte die ENC1-Zelllinie ebenfalls zu Neuronen differenziert werden (siehe Abschnitt 3.1.1). Dies würde ein weiteres Kulturmodell zum Nachweis einer Akkumulation von $^{13}\text{C}_5\text{-dC}$ in postmitotischen Zellen liefern.

3.1.2.4 Projektbeiträge

Die erste Synthese des D_4dCs wurde von *Alexander Schön* während seiner Masterarbeit unter der Anleitung von *Dr. René Rahimoff* durchgeführt. Alle weiteren Metaboliten zur Verfütterung ($\text{D}_2^{15}\text{N}_2\text{-dC}$, $^{13}\text{C}_5\text{D}_1\text{-dC}$ und $^{13}\text{C}_5\text{D}_1\text{-rC}$) wurden von *Dr. René Rahimoff* synthetisiert. Die Zellkulturversuche der beschriebenen Experimente wurden von mir entwickelt und durchgeführt. *Ewelina Kamińska* führte unter meiner Anleitung die Differenzierung der humanen reprogrammierten Zellen durch. Auch die Isolation aller DNA-Proben wurde von mir vorgenommen. Die Entwicklung der Quantifizierungsmethoden, der enzymatische Verdau, die Messung und Auswertung der Daten wurden von *Dr. Katharina Iwan* und *Dr. Jessica Steinbacher* vorgenommen. Die Daten wurden gemeinsam von *Dr. Katharina Iwan*, *Dr. René Rahimoff* und mir interpretiert.

3.1.3 Untersuchung einer enzymvermittelten DNA-Demethylierung über einen C-C Bindungsbruch

3.1.3.1 Prolog

Die Veröffentlichung aus Abschnitt 3.1.1 und die Ergebnisse aus Abschnitt 3.1.2 geben Hinweise auf eine direkte C-C Bindungsspaltung. Neben der Beteiligung der BER an der aktiven Demethylierung rückt daher die viel diskutierte Route über einen C-C Bindungsbruch ebenfalls in den Fokus. Zu diesem Zeitpunkt ist jedoch unklar, inwieweit die beschriebene Bindungsspaltung an der aktiven Demethylierung beteiligt ist und ob diese über eine Deformylierung oder eine Decarboxylierung stattfindet. Eindeutig ist, dass es sich hierbei um einen Dnmt- und Tet-unabhängigen Weg zur Entfernung von fdC bzw. cadC aus dem Säugetiergenom handelt,^[65] der ohne die Generierung von schädlichen abasischen Stellen abläuft.^[262] Dieses Kapitel befasst sich mit der Fragestellung, welches Enzym oder welche Faktoren eine derartige Reaktion bewerkstelligen könnten (Abbildung 31). Im Zuge dieser Studien soll untersucht werden, ob es sich bei der beobachteten Reaktion aus Abschnitt 3.1.1 um eine enzymatisch katalysierte Reaktion oder um einen Übertrag der Formyl-Gruppe auf eine bestimmte Entität handelt.

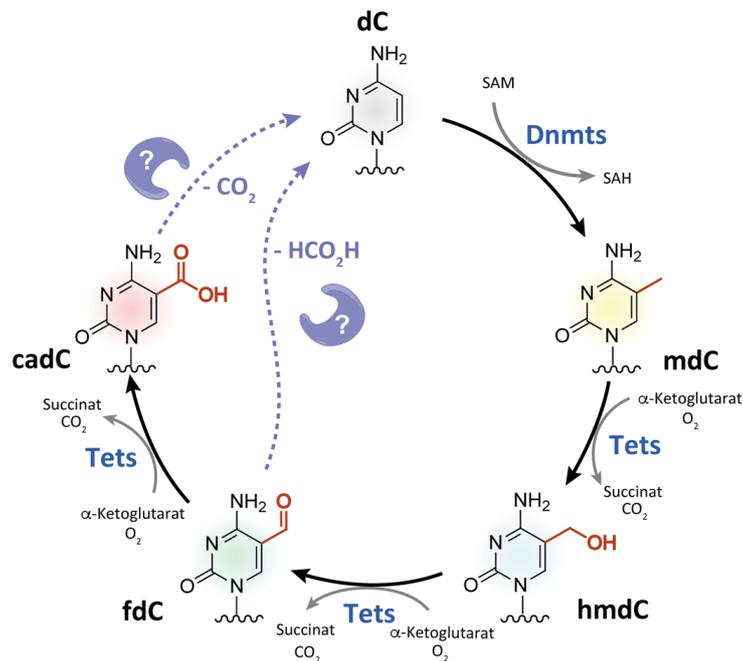


Abbildung 31: Methylierung von dC und oxidativer Demethylierungsweg über einen C-C Bindungsbruch.

Zunächst wird dC mit Hilfe der DNA-Methyltransferasen methyliert. Das entstandene mdC wird nachfolgend durch die Tet-Dioxygenasen schrittweise oxidiert. Die Rückgewinnung von dC könnte über eine direkte Deformylierung von fdC bzw. Decarboxylierung von cadC ablaufen. Die benötigten Enzyme oder Faktoren, die diese Reaktion bewerkstelligen könnten, sind bisher jedoch nicht bekannt.

3.1.3.2 Ergebnisse und Diskussion

Zunächst sollte der Fragestellung nachgegangen werden, ob der C-C Bindungsbruch während oder direkt nach der Inkorporation der Nucleoside in die genomische DNA stattfindet und Hinweise auf beteiligte Enzyme oder Faktoren geben. Da die Einbringung der modifizierten Nucleoside in die genomische DNA im Zuge der DNA-Synthese erfolgt, sollten die Zellen zunächst vor Eintritt in die S-Phase festgehalten werden. Als Modellsystem wurden *Chinese Hamster Ovary* (CHO)-K1 Zellen gewählt, da diese als leicht zu synchronisierendes Zellsystem gelten und bereits eine C-C Bindungsbruchreaktion nach Fütterung mit F-dC zeigten.^[65]

Isotopenverfolgungsexperimente mit synchronisierten Säugetierzellen

Nach Synchronisation der Zellen an der G1/S-Grenze wurden die Zellen nach Wiederaufnahme des Zellzyklus mit [¹³C₅][¹⁵N₂]-fdC gefüttert und die zeitliche Entstehung des Deformylierungsprodukts [¹³C₅][¹⁵N₂]-dC massenspektrometrisch verfolgt. Die Synchronisation der CHO-K1 erfolgte durch einen zweifachen Thymidinblock.^[263] Die hohe Konzentration an Thymidin inhibiert die Ribonucleotidreduktase und so die Synthese der dNTPs. Dadurch wird die DNA-Synthese verlangsamt und die Zellen arretieren am Übertritt zur S-Phase.^[264] Zur vollständigen Synchronisation wurde der Block zweimal durchgeführt. Abbildung 32A zeigt schematisch den Ablauf des Experiments. Die erfolgreiche Synchronisation wurde mit Hilfe eines Durchflusszytometers nachgewiesen. Als Referenz dienten nicht-synchronisierte Zellen. Durch den interkalierenden Fluoreszenzfarbstoff Propidiumiodid wurde der DNA-Gehalt und anhand dessen die Verteilung der Zellzyklus-Phasen bestimmt. Dies ergab für nicht-synchronisierte Zellen eine Verteilung von 41 % G1, 51 % S und 7 % G2/M. Nach Synchronisation der Zellen befanden sich 93 % in der G1, 7 % in der S und keine Zellen in der G2/M-Phase. Anschließend wurde das Kulturmedium erneuert und gleichzeitig [¹³C₅][¹⁵N₂]-fdC zugegeben. Nach Eintritt in die S-Phase wurde der Einbau des isotope markierten Nucleosids im Laufe der S-Phase verfolgt. Zum Vergleich wurden Zellen in der frühen und späten S-Phase geerntet. Nach Isolation der genomischen DNA wurde die Deformylierungsaktivität massenspektroskopisch untersucht. Die Quantifizierungsergebnisse sind in Abbildung 32B dargestellt. Die Daten zeigen einen Anstieg an [¹³C₅][¹⁵N₂]-fdC in synchronisierten Zellen verglichen mit der Kontrolle. Nach 7h, in der späten S-Phase, ist ein noch deutlicherer Anstieg der eingebauten Menge an markiertem Nucleosid zu erkennen.

Weiterhin konnte, sowohl in synchronisierten, als auch nicht-synchronisierten Zellen das Resultat des C-C Bindungsbruchs [¹³C₅][¹⁵N₂]-dC detektiert werden. Da auch in nicht-synchronisierten Zellen 51 % der Zellen in der S-Phase sind, ist ein Einbau des [¹³C₅][¹⁵N₂]-fdC und damit das Auftreten von [¹³C₅][¹⁵N₂]-dC zu erwarten. In nicht-synchronisierten und besonders in synchronisierten Zellen nimmt die Menge an isotope markiertem dC mit der Zeit deutlich zu.

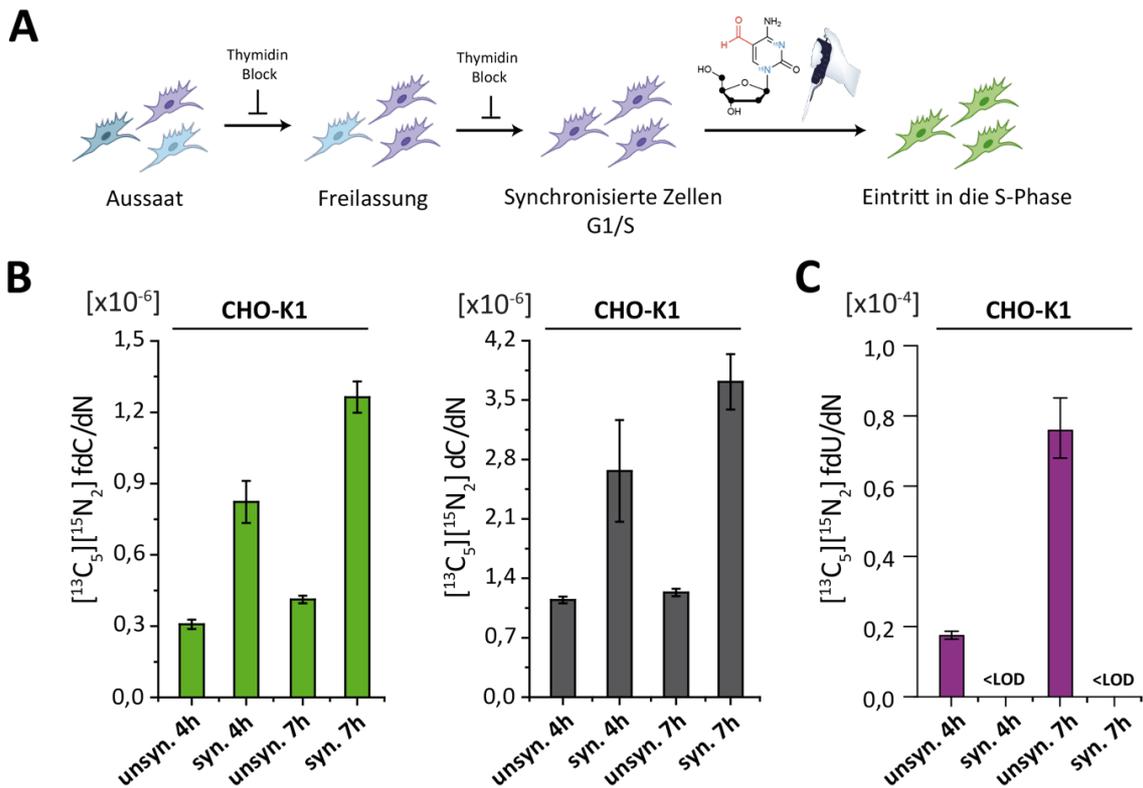


Abbildung 32: Synchronisation von CHO-K1 zur Ermittlung der Zellzyklusabhängigkeit des C-C Bindungsbruchs.

A) Schematischer Ablauf des Thymidinblocks mit anschließender Einbringung des markierten Nucleosids. **B)** Normierte Quantifizierungsergebnisse der synchronisierten (syn.) und unsynchronisierten (unsyn.) CHO-K1 Zellen 4 bzw. 7 h nach Eintritt in die S-Phase. **C)** Unabhängiges Experiment zur Quantifizierung der isotonenmarkierten 5-Formyldeoxyuridin-Spezies nach Synchronisation der CHO-K1 Zellen. Die Daten wurden jeweils aus einer biologischen Probe, gemessen als technisches Triplikat, erhoben. LOD: *Limit of Detection*

Weiterhin wurde beobachtet, dass [¹³C₅][¹⁵N₂]-5-Formyldeoxyuridin ([¹³C₅][¹⁵N₂]-fdU) in nicht-synchronisierten Zellen detektierbar ist und in synchronisierten Zellen unter dem *Limit of Detection* (LOD) liegt (Abbildung 32C). Diese Uridin-Verbindung kann spontan oder durch enzymatische Desaminierung im löslichen Nucleosid/-tid Pool entstehen. Allerdings deutet die beachtliche Menge von $7,66 \times 10^{-5}$ /dN an isotonenmarkiertem fdU eher auf eine enzymatische Entstehung hin.

Die C-C Bindungsbruchreaktion geschieht demnach während oder direkt nach der Inkorporation in die genomische DNA. Die gute Korrelation der Daten legt die Vermutung nahe, dass die Deformylierung bereits auf Nucleosid/-tid Ebene oder direkt während der DNA-Synthese stattfindet. Die Befunde deuten weiterhin darauf hin, dass synchronisierte Zellen das isotonenmarkierte Nucleosid zu schnell einbauen, um eine Desaminierung durch die Desaminasen des löslichen Pools zu ermöglichen, was gegen einen C-C Bindungsbruch im löslichen Nucleosid- bzw. Nucleotidpool sprechen würde. Eine alternative Erklärung wäre eine Überfrachtung des löslichen Pools an dT aufgrund des Thymidin-Blocks, wodurch dT-Derivate, wie fdU, zunächst nicht eingebaut werden.

Eine Analyse des löslichen Pools der synchronisierten Zellen könnte an dieser Stelle Aufschluss geben. Zusammenfassend konnte aus dieser Untersuchung gefolgert werden, dass der C-C Bindungsbruch direkt nach dem Einbau in die genomische DNA erfolgen muss.

Beteiligung DNA-bindender Enzyme an der C-C Bindungsbruchreaktion

Unter der Annahme, dass die Deformylierung auf DNA-Ebene stattfindet, wurden im weiteren Verlauf DNA-interagierende Enzyme als potenzielle Deformylasen näher betrachtet. Hierzu wurde in der Veröffentlichung aus Abschnitt 3.1.1 bereits eine Tet *Triple Knockout* (TKO) Zelllinie untersucht. Diese embryonale Stammzelllinie enthält kein Tet1, 2, sowie 3 und kann daher die höher oxidierten Cytidin-Spezies hmdC, fdC und cadC nicht bilden. Die Quantifizierungsdaten zeigten, dass die C-C Bindungsbruchreaktion nicht durch die Tet-Enzyme katalysiert wird.

Studien einer möglichen Deformylierung und Decarboxylierung zeigten bereits, dass ein nukleophiler Angriff an der C6-Position des Cytidins einen C-C Bindungsbruch erleichtern könnte.^[230] Eine derartige Reaktion wird von den Dnmt-Enzymen katalysiert.^[265] Eine geringe Dehydroxylase- und Decarboxylaseaktivität der Dnmts in Abwesenheit von SAM wurde bereits beobachtet.^{[243], [244, 246]} Aus diesem Grund wurde untersucht, ob die Dnmt-Enzyme die beobachtete C-C Bindungsbruchreaktion katalysieren könnten. In Abschnitt 3.1.1 wurde eine Dnmt TKO Zelllinie, welche die Enzyme Dnmt1, 3a und 3b nicht enthält analysiert. Die Ergebnisse der massenspektrometrischen Analyse zeigen keinen Einfluss auf die quantifizierten Mengen an F-dC. Im Zuge dieser Untersuchungen wurden die *Knockouts* der einzelnen Methyltransferasen ebenfalls analysiert. Zur Betrachtung der Deformylierungsreaktion wurden die Zellen für 72 h unter C/R-Bedingungen kultiviert. In den letzten 24 h wurde das Medium ausgetauscht und F-fdC zugesetzt.

Nach Isolation der genomischen DNA und absoluter Quantifizierung der fluorierten Derivate mittels UHPLC-MS/MS wurden zwei Normierungen zur besseren Vergleichbarkeit der einzelnen *Knockouts* durchgeführt. Zunächst wurden die jeweiligen F-dC- auf die F-fdC-Werte bezogen und das erhaltene Verhältnis auf den Wildtyp normiert. Abbildung 33 zeigt das Ergebnis dieser Normierung und die dazugehörigen natürlichen mdC-Werte.

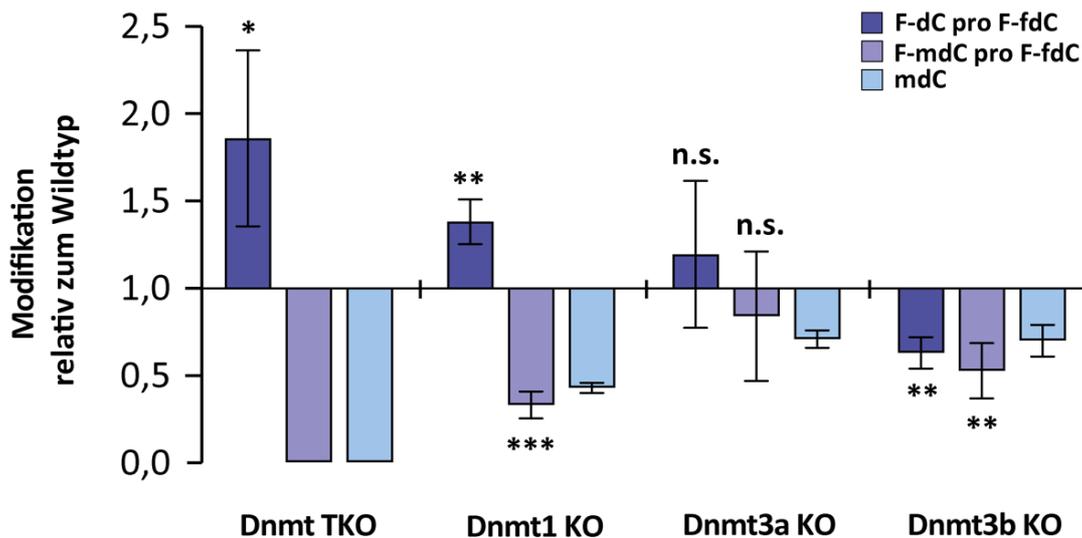


Abbildung 33: Verhältnis von 2'-fluorierten Nucleosiden und natürlichem m5C in verschiedenen Dnmt-Knockouts relativ zum Wildtyp. Dargestellt sind das F-dC (blau) bzw. F-mdC (hellviolett) zu F-fdC Verhältnis der Dnmt KOs und die quantifizierte Menge an natürlichem m5C (hellblau) jeweils bezogen auf den Wildtyp. Abgebildet sind die normierten Quantifizierungsergebnisse dreier biologischer Replikate, gemessen als technisches Triplikat.

Beim Vergleich der errechneten Verhältnisse von F-dC zu F-fdC bzw. F-mdC zu F-fdC wurden deutliche Unterschiede zwischen den Zelllinien festgestellt. Die Inkorporation an F-fdC war in den untersuchten Zelllinien etwa gleich und schien daher nicht zu den beobachteten Variationen beizutragen. Eine Erhöhung des F-dC zu F-fdC Verhältnisses, wie im Dnmt TKO und im Dnmt1 KO beobachtet, würde demnach eine vermehrte C-C Bindungsbruchreaktion bedeuten. Die *maintenance* Methyltransferase Dnmt1 erhält den Großteil des Methylierungsmusters, da diese hemimethylierte CpG-Stellen nach der Replikation bindet und remethyliert.^[96] Auch die gemessenen natürlichen m5C-Werte zeigen den Beitrag von Dnmt1 an der Methylierung. Während im Dnmt TKO kein m5C detektierbar ist, erfolgt im Dnmt1 KO eine Reduktion des natürlichen m5C um etwa 50%. Ebenso nimmt die Menge an remethyliertem F-dC (F-mdC) um einen ähnlichen Prozentsatz ab. Das Verhältnis von F-dC zu F-fdC verhält sich genau entgegengesetzt. Der Verlust an *maintenance* Methylierung führt also zu einer größeren Menge an F-dC, was auf eine fehlende Remethylierung zurückzuführen sein könnte. Allerdings sind die absoluten F-mdC Werte um zwei Zehnerpotenzen geringer als die F-dC Werte. Es scheint daher unwahrscheinlich, dass der Beitrag der fehlenden Remethylierung maßgebend für die erhaltene Menge an F-dC ist.

Die Betrachtung der *de novo* Methyltransferasen Dnmt3a und 3b zeigte sich hingegen ein anderes Bild. Die *de novo* Methyltransferasen binden unmethylierte DNA im CpG-Kontext und führen dort eine Methylierung ein.^[96] Sowohl im Dnmt3a KO, als auch im Dnmt3b KO kommt es zu einer Abnahme an natürlichem m5C um etwa 20%.

Während der Verlust von Dnmt3a zu keiner signifikanten Änderung der F-dC bzw. F-mdC zu F-fdC Verhältnisse führte, war gegenteiliges der Fall für Dnmt3b. Hier konnte erstmals eine signifikante Abnahme beobachtet werden. Im Dnmt3b KO nimmt das F-dC zu F-fdC Verhältnis etwa um 40 % signifikant ab. Erstaunlich ist, dass der Dnmt TKO einen gegenläufigen Trend trotz des Verlustes an Dnmt3b zeigt. Da der Verlust an natürlichem mdC im Dnmt3b um einen ähnlichen Prozentsatz erfolgt wie die Menge an F-dC bzw. F-mdC zu F-fdC, wäre ein Zusammenhang denkbar. Allerdings sind hier auch die Unterschiede der absoluten Werte zu beachten. Die Abnahme des F-dC zu F-fdC Verhältnisses im Dnmt3b legt jedoch die Vermutung nahe, dass dieses Protein am beobachteten C-C Bindungsbruch beteiligt sein könnte, aber selbst nicht das gesuchte demodifizierende Enzym ist. Ansonsten sollte kein F-dC detektierbar sein. Um auszuschließen, dass es sich bei dem beobachteten Rückgang um einen Effekt einzig dieser KO-Linie handelt, wurde in einem weiteren Experiment eine unabhängige Dnmt3b KO Linie untersucht. Diese Studie führte zum selben Ergebnis (Daten nicht gezeigt) und bekräftigt somit die Beteiligung von Dnmt3b.

Um den Mechanismus der C-C Bindungsbruchreaktion zu verstehen, wurde auch die Zunahme an F-dC genauer betrachtet. Da der Verlust an Dnmt1 zu einer Erhöhung des F-dC-Wertes führt und den größten Anteil an Methylierung ins Genom einbringt, erschien eine genauere Betrachtung literaturbekannter Interaktionspartner von Dnmt1 interessant. Eine Überlegung führte zur Interaktion der *maintenance* DNA-Methyltransferase mit dem DNA-Reparaturenzym *MutL homolog 1* (Mlh1) und der bereits nachgewiesenen Beteiligung von Dnmt1 an der *Mismatch*-Reparatur (MMR).^[266, 267] Aufgrund der Zusammenarbeit von Dnmt1 und Mlh1 ist ein ähnlicher Effekt auf das F-dC zu F-fdC Verhältnis denkbar. Bisherige Studien konnten zeigen, dass die beiden Proteine an dieselben Genompositionen rekrutiert werden.^[268] Daher wäre es denkbar, dass die Bindungsbruchreaktion an eben diesen Positionen stattfindet. Zur Untersuchung dessen wurde eine Mlh1 KO mESC-Linie massenspektrometrisch analysiert. Die Generierung dieser mESC-Linie wird in Abschnitt 3.3 beschrieben. Erste Quantifizierungsergebnisse zeigten höhere Werte der oxidierten dC-Spezies im Mlh1 KO (siehe Abbildung 53). Die Werte an hmdC, fdC und cadC waren etwa zweimal höher als im entsprechenden Wildtyp. Um die Beteiligung von Mlh1 an der Entstehung von F-dC zu untersuchen, wurden die Zellen für 72 h unter C/R-Bedingungen kultiviert und in den letzten 24 h der Kultur mit F-fdC gefüttert. Nach Isolation der genomischen DNA wurde der Metabolismus des Nukleosids mittels UHPLC-MS/MS verfolgt. Die Ergebnisse der absoluten Quantifizierung sind in Abbildung 34 aufgezeigt.

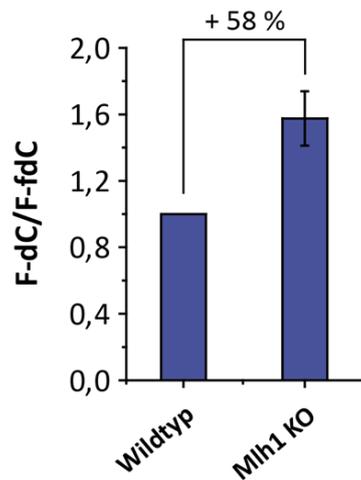


Abbildung 34: Beteiligung des Mismatch Repair Proteins Mlh1 an der C-C Bindungsbruchreaktion. Verhältnis von demodifiziertem zu 2'-fluoriniertem fdC im Mlh1 KO und dem korrespondierenden Wildtyp. Abgebildet sind die normierten Quantifizierungsergebnisse zweier biologischer Replikate, gemessen als technisches Triplikat.

Beim Vergleich des Mlh1 KOs mit dem korrespondierenden Wildtyp konnte eine Zunahme des F-dC zu F-fdC Verhältnisses um etwa 60 % beobachtet werden. Interessanterweise ist diese Zunahme vergleichbar mit der Zunahme im Dnmt1 KO. Dies spricht für die Hypothese, dass es sich um dieselben Genompositionen handeln könnte, die bereits in Abwesenheit von Dnmt1 zu einem Anstieg des F-dC zu F-fdC Verhältnisses führten.

Identifikation beteiligter Enzyme durch Inhibitorstudien

Im weiteren Verlauf wurden die *Poly-ADP-Ribose Polymerases* (PARPs) auf ihre Beteiligung an der Deformylierung untersucht. Die PARPs stellen eine große Familie multifunktionaler Enzyme dar. In Mäusen existieren 16 Mitglieder dieser Proteinfamilie, wobei PARP1 das am häufigsten vorkommende Mitglied darstellt.^[269] Diese Enzymklasse katalysiert den Transfer einer ADP-Ribose von Nicotinamiddinukleotid (NAD⁺) auf ein Zielprotein und ist wichtig für die zelluläre Homöostase sowie die DNA-Reparatur.^[270, 271] Durch Verwendung der bekannten PARP-Inhibitoren Olaparib und Veliparib sollte deren Einfluss auf die hier untersuchte C-C Bindungsbruchreaktion betrachtet werden.^[272] F-fdC wurde hierfür 72 h mit dem jeweiligen Inhibitor an die mESCs gefüttert. Die Konzentration von Olaparib (*Selleckchem*) im Kulturmedium betrug 20 nM. Für Veliparib (*Selleckchem*) wurde eine Konzentration von 1 µM verwendet. Nach Isolation der genomischen DNA und massenspektrometrischer Untersuchung konnte keine Änderung des F-dC zu F-fdC Verhältnisses detektiert werden. Daher wurde gefolgert, dass die PARP-Enzyme nicht an der Deformylierung beteiligt sind.

Demethylierungen, Deformylierungen und Decarboxylierungen sind in der Natur bereits bekannt (siehe Abschnitt 1.6.3). Ebenso existieren Enzyme die eine Deacetylierung katalysieren können. Entsprechend wäre eine Deformylierung über einen ähnlichen Weg denkbar. Die Enzymklasse der Histon-Deacetylasen (HDACs) katalysiert die Entfernung posttranslational eingeführter Acetyl-Gruppen der Lysin-Seitenketten von Histonen. Die Acetylierung führt zu einer gelockerten Chromatinstruktur und so zu einer besseren Erreichbarkeit der DNA für Transkriptionsfaktoren.^[273] Diese Proteinfamilie besitzt in Säugetieren 11 Mitglieder, die in vier Gruppen eingeteilt werden können. Sie unterscheiden sich in ihrer Struktur, dem Expressionsmuster, der enzymatischen Funktion und zellulären Lokalisation.^[273] Die Klassen I, II und IV nutzen einen Mechanismus in Abhängigkeit von Zink, während die Klasse III, die sogenannten Sirtuine, NAD⁺ abhängig sind.^[73] Zur Untersuchung der HDACs wurden zwei bekannte Inhibitoren Trichostatin A (10 nM; *Sigma*) und Valporinsäure (0,5 mM; *Selleckchem*) verwendet.^[274, 275] Für das Experiment wurden die mESCs für drei Tage unter C/R-Bedingungen kultiviert und dabei F-dC und der jeweiligen HDAC-Inhibitor zugesetzt. Anschließend wurde die genomische DNA isoliert und der Metabolismus des fluorierten Nucleosids mittels UHPLC-MS/MS analysiert. Die Inhibition durch Valporinsäure und Trichostatin A hatte keinen Einfluss auf die Deformylierung. Die HDACs der Klassen I, II und IV sind daher vermutlich nicht an der Deformylierung beteiligt.

Im nächsten Schritt wurde mit Hilfe des Breitband-Inhibitors Nicotinamid (NAM) die Aufgabe der Klasse III HDACs näher beleuchtet.^[276] Hierzu wurde der Inhibitor NAM (10 mM; *Sigma*) für 72 h zum Kulturmedium zugegeben und die Zellen unter den bereits genannten *Priming*-Bedingungen kultiviert (Abbildung 35A). Nach Isolation der genomischen DNA und massenspektrometrischer Analyse war eine signifikante Abnahme des F-dC zu F-fdC Verhältnisses erkennbar. Zur besseren Vergleichbarkeit der einzelnen biologischen Replikate wurden die F-dC auf die F-fdC Werte bezogen und die erhaltenen Verhältnisse auf den Wildtyp normiert. Die Behandlung mit dem Breitband-Inhibitor lieferte eine signifikante Reduktion des F-dC zu F-fdC Verhältnisses um etwa 24 % (siehe Abbildung 35B). Zur genaueren Untersuchung des NAM-Effekts wurde der Inhibitor für eine unterschiedliche Zeitdauer zugegeben und gleichzeitig das entstandene NAD⁺ mit Hilfe eines NAD⁺/NADH Detektionsassays (NAD⁺/NADH-Glo™ Assay, *Promega*) beobachtet. Dies sollte Rückschlüsse auf die Inhibition der Klasse III HDACs durch Zunahme an NAD⁺, dem Edukt der katalysierten Reaktion, geben. Nach 24 h NAM im Medium ist bereits ein Anstieg an NAD⁺ im Vergleich zur Kontrolle erkennbar. Dasselbe Ergebnis lieferte die NAM Behandlung für 72 h. Der Unterschied zwischen den NAD⁺-Werten nach 24 bzw. 72 h war sehr gering (Abbildung 35C). Demnach müsste dies auch für das F-dC zu F-fdC Verhältnis der Fall sein.

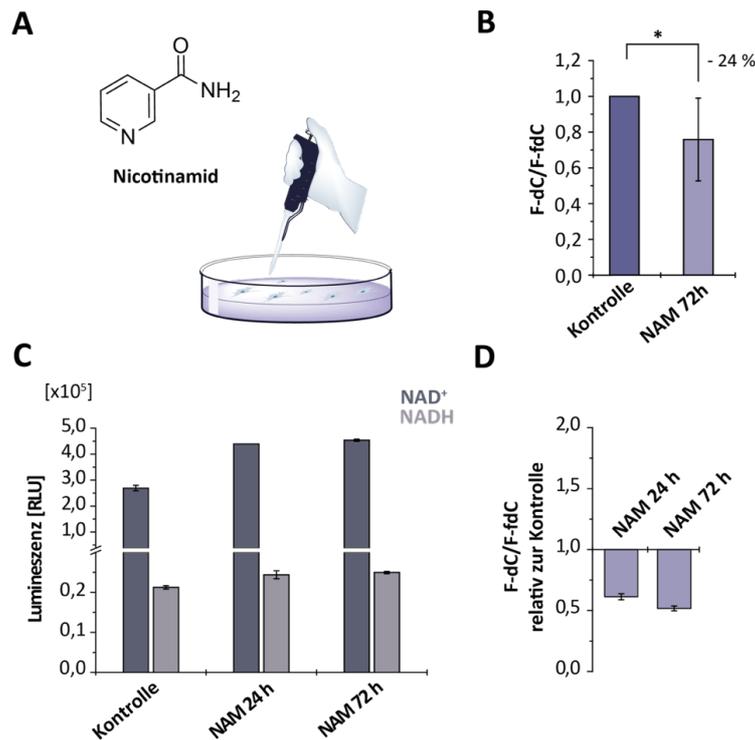


Abbildung 35: Untersuchung der Deformylierungsreaktion nach Behandlung der mESCs mit NAM.

A) Nicotinamid wird dem Kulturmedium zugesetzt. **B)** Verhältnis von demodifiziertem zu 2'-fluoriniertem fdC nach NAM Zugabe. Abgebildet sind die normierten Quantifizierungsergebnisse von vier biologischen Replikaten, gemessen als technische Triplikate. **C)** Ergebnisse des NAD⁺/NADH Detektionsassays nach NAM Behandlung. Dargestellt sind die Ergebnisse einer biologischen Probe, gemessen als technisches Triplikate. **D)** F-dC zu F-fdC nach NAM Zugabe für eine unterschiedliche Zeitdauer. Normierte Quantifizierungsergebnisse einer biologischen Probe, gemessen als technisches Triplikate.

Zur Untersuchung dessen wurde die DNA der behandelten mESCs isoliert und massenspektrometrisch analysiert. Tatsächlich ist der Unterschied zwischen 24 h und 72 h NAM Behandlung ebenfalls gering (Abbildung 35D). In beiden Fällen kam es zu einer Reduktion des F-dC zu F-fdC Verhältnisses um etwa 40 %. Die Zeitdauer der NAM Behandlung scheint daher nicht ausschlaggebend für den beobachteten Effekt zu sein.

Um einen direkten Nachweis der Beteiligung der Klasse III HDACs, der sogenannten Sirtuine, an der Deformylierungsreaktion zu erhalten, wurden spezifischere Inhibitoren verwendet. Hierzu wurden die beiden Inhibitoren Selisistat (50 μ M, Sigma, EX-527) und SirReal2 (10 μ M, Sigma) untersucht. Selisistat ist ein bekannter Sirtuin1-Inhibitor und SirReal2 ein Inhibitor von Sirtuin2.^[277, 278] Erneut wurde F-fdC an die Stammzellen zusammen mit dem jeweiligen Inhibitor verfüttert und das Verhältnis der fluorinierten Nukleoside massenspektrometrisch verfolgt. Die Daten zeigten eine große Varianz innerhalb der biologischen Replikate und führten daher zu keinem eindeutigen Ergebnis. Eine Kontrolle des Inhibitionsgrades wäre hier von Vorteil, da die Variabilität der erhaltenen Daten auf eine Varianz in der Inhibition zurückzuführen sein könnte.

Beteiligung der Klasse III Histon-Deacetylasen an der C-C Bindungsbruchreaktion

In Säugetieren existieren sieben Sirtuine, die nach ihrer biologischen Funktion in Klassen eingeteilt werden können. Einige Sirtuine katalysieren die Deacetylierung von Peptiden oder Proteinen und nutzen hierzu NAD^+ als Substrat, welches dadurch in NAM umgewandelt wird. Es entsteht das deacetylierte Produkt und die 2'-O-Acetyl-ADP-Ribose (siehe Abbildung 36A).^[279] Es wäre denkbar, dass diese Enzyme neben einer Deacetylierung auch eine Deformylierung nach einem ähnlichen Mechanismus katalysieren könnten. Beispielsweise könnte durch Deacetylierung von Lysin-Seitenketten entstandene freie Amine als Nukleophil an der Carbonylgruppe des Aldehyds angreifen und so die Formyl-Gruppe auf das Lysin übertragen. Als einzige Sirtuine mit Deacetylaseaktivität befinden sich Sirtuin1 und Sirtuin6 im Nukleus.^[280] Um eindeutige Hinweise auf eine Beteiligung der Sirtuine an der Deformylierungsreaktion zu bekommen wurden im Folgenden mESC KO-Linien der Sirtuine1 und 6 untersucht.

Erneut wurden die Zellen mit F-fdC unter *Priming*-Bedingungen kultiviert und anschließend die genomische DNA isoliert. Nach UHPLC-MS/MS Analyse wurden die erhaltenen F-dC zu F-fdC Verhältnisse auf den Wildtyp normiert. Es konnte eine signifikante Reduktion der Deformylierung um etwa 60 % im Sirtuin1 KO relativ zum Wildtyp beobachtet werden (Abbildung 36B). Dieses Ergebnis gibt einen deutlichen Hinweis auf die Beteiligung von Sirtuin1 an der beobachteten C-C Bindungsbruchreaktion. Allerdings zeigt dieses Ergebnis auch, dass Sirtuin1 entweder eine sekundäre Rolle spielt oder nicht das einzige Enzym mit Deformylaseaktivität ist. Im Falle des Sirtuin6 KOs kam es zu einer Reduktion der Aktivität um etwa 15 %. Dieser Rückgang des F-dC zu F-fdC Verhältnisses war jedoch nicht signifikant (Abbildung 36B). Um sicherzustellen, dass es sich hierbei um eine signifikante Reduktion der Deformylierungsaktivität handelt, müssten weitere biologische Replikate untersucht werden.

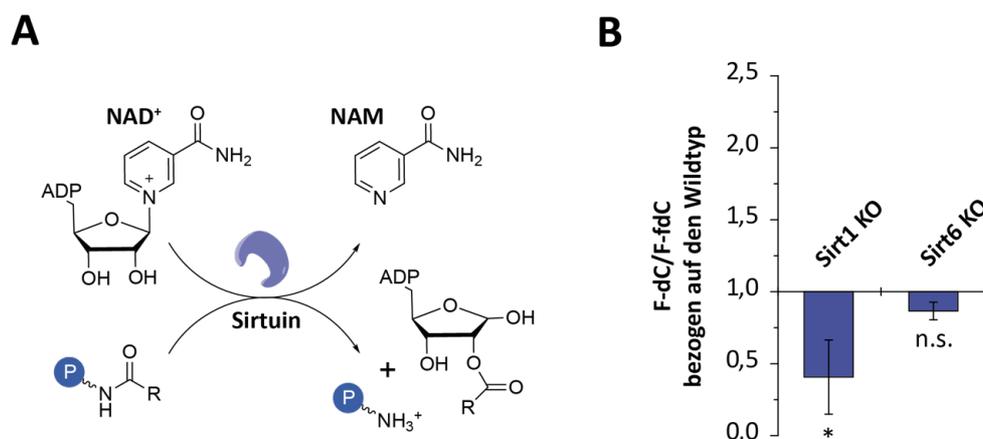


Abbildung 36: Katalytische Funktion der Sirtuine und der C-C Bindungsbruchstudie in Sirtuin-Knockout mESCs.

A) Enzymatische Reaktion der Sirtuine.^[279] **B)** Verhältnis von F-dC zu F-fdC in den untersuchten Sirtuin (Sirt)-Knockouts relativ zum Wildtyp. Abgebildet sind die normierten Quantifizierungsergebnisse von vier biologischen Replikaten (Sirt1) bzw. einer biologischen Probe (Sirt6), gemessen als technisches Triplikat.

Ferner wäre denkbar, dass die Abwesenheit von Sirtuin1 bzw. 6 zu einer Hochregulation anderer Sirtuine führt, die ebenfalls diese Reaktion katalysieren. Als weiteren Hinweis auf die Beteiligung von Sirtuin1 sollte die direkte Deformylaseaktivität von Sirtuin1 *in vitro* nachgewiesen werden. Hierzu wurde ein rekombinantes Sirtuin1 mit einem fdC enthaltenden Oligonukleotid inkubiert. Zur Bestimmung der Sirtuin-Aktivität wurde der von *Seidel et al.* entwickelte Assay genutzt.^[281] Bei diesem wird ein acetyliertes Substrat durch das aktive Sirtuin deacetyliert, woraufhin das Substrat nach tryptischem Verdau ein Fluorophor freisetzt. Das Fluoreszenzsignal korreliert somit mit der Aktivität des Enzyms.^[281] Die Ergebnisse des Assays zur Überprüfung der Aktivität des Sirtuins sind in Abbildung 37A gezeigt. Dargestellt ist das Fluoreszenzsignal resultierend aus dem vollständig deacetylierten Kontrollsubstrat und zeigt das maximal erreichbare Fluoreszenzsignal. Die weiteren beiden Werte zeigen das acetylierte Peptid mit und ohne Enzym. Ein Fluoreszenzsignal kann nur nach Abspaltung der Acetyl-Gruppe erhalten werden. Der Assay zeigt, dass mit 5 μM humanen Sirtuin1 (hSIRT1) das Substrat (50 μM) vollständig umgesetzt werden kann.

Im durchgeführten Experiment wurde ein $^{15}\text{N}_2$ -fdC enthaltendes, doppelsträngiges Oligonukleotid (30 mer) mit dem aktiven hSIRT1 inkubiert. Die Aktivität des Enzyms wurde stets mit Hilfe des eben beschriebenen Assays überprüft. Der Doppelstrang enthielt im CpG-Kontext entweder fdC gegenüber dC oder dC gegenüber mdC (siehe Abbildung 37B).

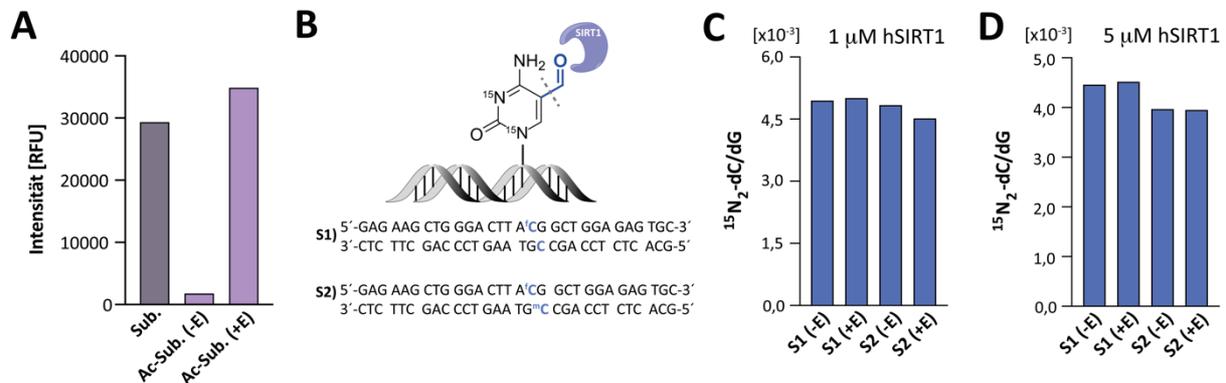


Abbildung 37: Aktivitätsassay und *in vitro* Experiment. A) Aktivitätsbestimmung des Sirtuins. Sub.: Substrat; Ac-Sub.: acetyliertes Substrat; E: Enzym. **B)** Schematische Darstellung des Assays. **C)** Quantifizierungsergebnisse nach Inkubation mit 1 μM Sirtuin inkubiert für 60 min. **D)** Siehe C mit 5 μM Sirtuin und 60 min Inkubationszeit.

Ein erfolgreicher Nachweis der Deformylaseaktivität des Sirtuins würde zur vermehrten Detektion von $^{15}\text{N}_2$ -dC im Doppelstrang führen. Abbildung 37C zeigt die massenspektrometrischen Ergebnisse des *in vitro* Assays mit 1 μM hSIRT1 und einer Inkubation von 60 min. Ein $^{15}\text{N}_2$ -dC Signal konnte detektiert werden, allerdings waren die Werte mit und ohne Enzym in etwa gleich. Das detektierte Signal ist vermutlich auf den natürlichen Isotopenhintergrund zurückzuführen. Das Experiment wurde mit einer verlängerten Inkubationszeit (24 h) wiederholt, führte jedoch zum selben Ergebnis.

Das gleiche Experiment wurde ebenfalls mit einer höheren Enzymkonzentration (5 μ M hSIRT1) durchgeführt. Abbildung 37D zeigt das Ergebnis, wobei auch hier kein Unterschied an $^{15}\text{N}_2$ -dC mit und ohne Enzym detektiert wurde. Die Ergebnisse dieses Experiments zeigen, dass das hSIRT1 zwar aktiv ist, aber das isotopenmarkierte fdC im Oligonukleotid nicht zu deformylieren vermag.

Es bleibt weiterhin ungeklärt ob weitere Hilfsproteine benötigt werden oder die Klasse III HDACs selbst die deformylierende Entität sind. In der Literatur gibt es bereits Hinweise auf ein Zusammenspiel von Tet- und Dnmt-Enzymen mit den Sirtuinen.^[279, 282, 283] Ferner wurde eine Dnmt3b-vermittelte Rekrutierung von Sirtuin1 an kondensiertes Chromatin publiziert.^[284] Da im Dnmt3b KO eine Reduktion des F-dC zu F-fdC Verhältnisses beobachtet werden konnte, wäre eine Beteiligung von Dnmt3b an der Demodifikation durchaus denkbar. Zur genauen Aufklärung dieses Zusammenspiels bedarf es weiterer Untersuchungen.

3.1.3.3 Projektbeiträge

Die verwendeten Nukleoside wurden von *Dr. René Rahimoff* und *Eva Korytiaková* synthetisiert. *Dr. Fabio Spada* führte die Behandlung muriner embryonaler Stammzellen mit HDAC oder PARP-Inhibitoren durch. Die Synchronisation der CHO-K1 wurde von *Shqiponja Ferizaj* im Rahmen ihrer Masterarbeit unter meiner Anleitung durchgeführt. Alle weiteren Zellkulturversuche und biochemischen Experimente wurden von mir durchgeführt. Die Isolation der genomischen DNA aller Proben wurde von mir in Zusammenarbeit mit *Luis de la Osa de la Rosa* vorgenommen. *Ewelina Kamińska* führte unter meiner Anleitung den *in vitro* Aktivitätsassay durch. Der enzymatische Verdau wurde von *Dr. Katharina Iwan* mit Unterstützung von *Kerstin Kurz* unternommen. Die UHPLC-MS/MS Messung und Datenauswertung wurde ebenfalls von *Dr. Katharina Iwan* durchgeführt. Die Interpretation der Daten erfolgte von *Dr. Katharina Iwan* und mir gemeinsam. *Ewelina Kamińska*, *Eva Korytiaková*, *Alexander Schön* und *Dr. Sarah Schiffers* führen das Projekt fort.

3.2 Studien zur Basenexzisionsreparatur-vermittelten aktiven Demethylierung

Allgemein akzeptiert ist die Entfernung nichtkanonischer DNA-Basen im Zuge der Basenexzisionsreparatur. Wie lange nichtkanonische Basen und ihre Reparaturintermediate im Genom tatsächlich bestehen bleiben und ob die modifizierten Nucleobasen eine eigenständige Funktion inne haben ist bislang noch nicht vollständig geklärt. Dieses Kapitel befasst sich mit einer Quantifizierung der BER-Reparaturintermediate sowie der Dynamik der höher oxidierten dC-Spezies fdC und cadC. Weiterhin wird die Beteiligung mono- bzw. bifunktionaler Glykosylasen an der Entfernung oxidierter dC sowie dT-Derivate massenspektrometrisch untersucht. Abbildung 38 zeigt die betrachteten nicht-kanonischen Nucleobasen und eine schematische Darstellung der Demodifikation über die BER.

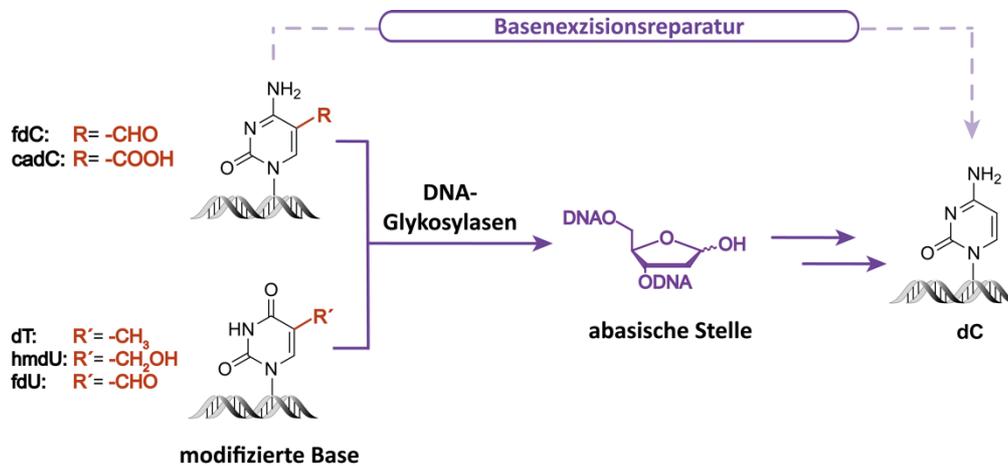


Abbildung 38: DNA-Demethylierung über die Basenexzisionsreparatur. Das Schema zeigt die Entfernung modifizierter Nucleobasen und Rückgewinnung von kanonischem dC. Nach Spaltung der glykosidischen Bindung durch die DNA-Glykosylasen entsteht eine abasische Stelle die im Rahmen der Basenexzisionsreparatur weiter prozessiert wird.

3.2.1 5-Formyl- und 5-Carboxydesoxycytidine verursachen keine schädlichen Reparaturintermediate in murinen embryonalen Stammzellen

„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

#: Diese Autoren haben zu gleichen Teilen beigetragen.

Prolog

In welcher Anzahl abasische Stellen tatsächlich im Genom von Säugetierzellen vorkommen, ist nicht genau geklärt. Die Verknüpfung mit massenspektrometrischen Analysemethoden ist daher ideal, um AP-Stellen, β - oder β , δ -Eliminierungsintermediate exakt zu quantifizieren, was mit den bisherigen Methoden nicht erreicht werden konnte.^[285-287] Im nachfolgenden Manuskript wird die Entwicklung eines neuartigen Derivatisierungsreagenzes beschrieben, dass in Kombination mit UHPLC-MS/MS die genaue Quantifizierung der AP-Stellen und parallel dazu die Bestimmung der β -Eliminierungsintermediate erlaubt (Abbildung 39). Hierzu wurden Zellkulturen mit schweren Nukleosiden markiert, um so die Entstehung der AP-Stelle aus dem ursprünglichen Nukleosid nachverfolgen zu können. Weiterhin wurde eine Verknüpfung der BER mit der aktiven Demethylierung untersucht. Die Markierungsexperimente mit schwerem Cytidin zeigten nur eine geringe Anzahl an β -Eliminierungsintermediaten und auch nur wenige AP-Stellen, was eine Beteiligung des BER Prozesses an der aktiven Demethylierung in Frage stellt.

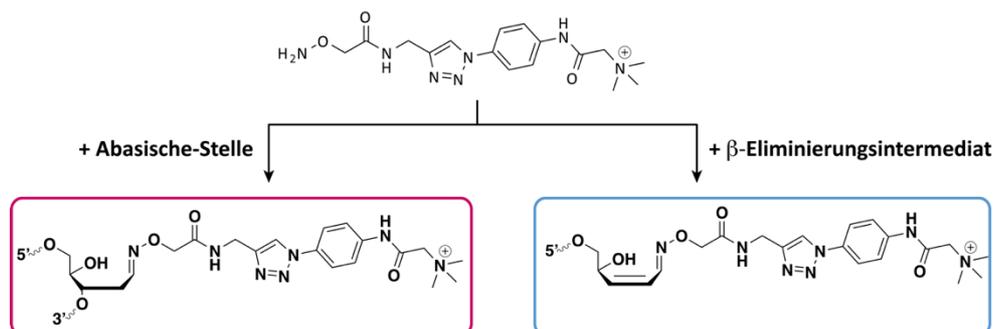


Abbildung 39: Chemische Struktur des Reagenzes zur Derivatisierung und Quantifizierung von abasischen Stellen und β -Eliminierungsintermediaten.

Autorenbeitrag

Alle verwendeten Zelllinien wurden von mir kultiviert. Die Isotopenmarkierungsexperimente, Inhibitor-, sowie *Knockdown*-Studien wurden von mir entworfen und durchgeführt. Die *Knockdown*-Effizienz wurde von mir mittels RT-qPCR nachgewiesen. Der *in vitro* Glykosylase-Assay wurde von mir durchgeführt. Die Isolation der genomischen DNA wurde von mir durchgeführt. Die Derivatisierung der genomischen DNA wurde in enger Zusammenarbeit mit *Dr. René Rahimoff* vorgenommen. Weiterhin wurden die erhaltenen Ergebnisse und Daten von mir, gemeinsam mit den anderen Autoren diskutiert und interpretiert. Zudem war ich an der Abfassung des Manuskripts beteiligt.

Lizenz

Kopie der Publikation mit Erlaubnis des Verlags; Copyright 2017 American Chemical Society.

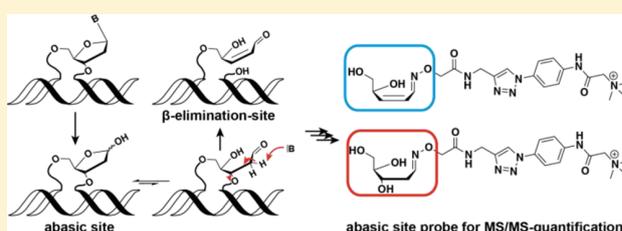
5-Formyl- and 5-Carboxydeoxycytidines Do Not Cause Accumulation of Harmful Repair Intermediates in Stem Cells

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

S 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 (β E) 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 β E-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 β E-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) I (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 (β E) reaction (Figure 1A, step c), which gives transient β E-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 β E-site intermediates are generated in the genome in response to the presence of fdC and cadC.

Here, we report the development of reagent 1¹⁵ (Figure 1B) that in combination with highly sensitive UHPLC-triple quadrupole mass spectrometry (MS)¹⁶ enables quantification of both AP-sites and β E intermediates with unprecedented sensitivity (~ 100 intermediates per genome). The method was used to quantify the steady-state levels of AP- and β E-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 β E-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

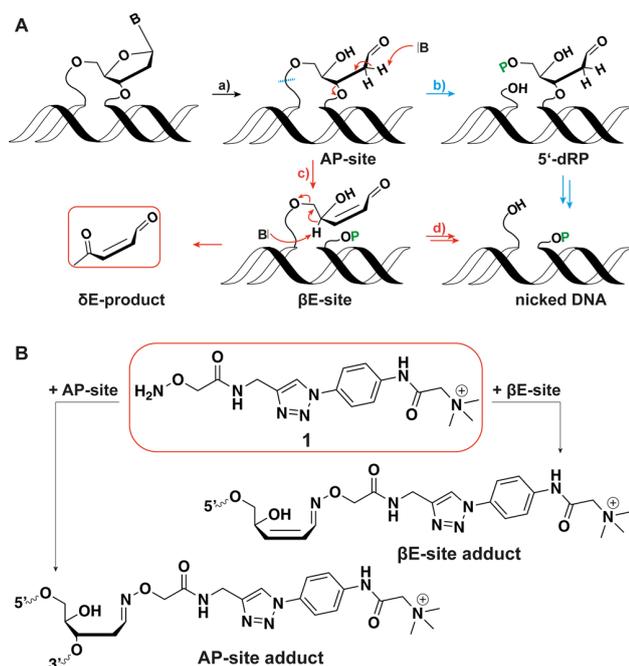


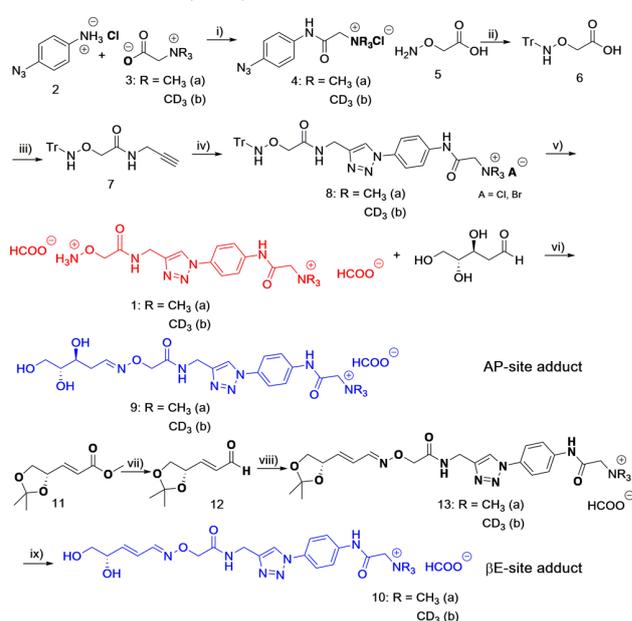
Figure 1. Mechanism of the base excision repair process. (A) Chemistry of BER with formation of AP- and β E-sites. Blue pathway: Monofunctional glycosylases act in concert with Ape1 and Pol β . Red pathway: Bifunctional glycosylases. (B) Depiction of reagent **1** and of the reaction products that are formed when **1** reacts with AP- and β E-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 β E-sites that are embedded in the negatively charged DNA duplex. The coupling to mass spectrometry is essential to distinguish AP- and β E-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 azidoamide **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 β E-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)⁴⁴



^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-SM e_2 , H $_2$ O/DCM (1:1), rt, 16 h, 77%; (v) 6 M HCl, DCM/H $_2$ O (1:1), rt, 1 h, quant.; (vi) H $_2$ O, 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 $_2$ O (1:1), rt, 16 h, 68%, (2) pTSA-H $_2$ O, H $_2$ O, 25 °C, o/n, HPLC (2 \times), 14%.

reacted both isotopologues **1a/b** with 2'-deoxyribose, which provided the AP-site reaction products **9a/b**. To prepare the β E 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 $m/z = 478.2 \rightarrow 450.2$ (quantifier) and $m/z = 478.2 \rightarrow 192.1$ (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 $m/z = 487.3 \rightarrow 459.3$ and $m/z = 487.3 \rightarrow 201.2$. Similar data were obtained for the β E-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 β E-adducts **9** and **10** have advantageous ionization (permanent positive charge) and fragmentation (triazole that leads to a neutral loss of N_2)

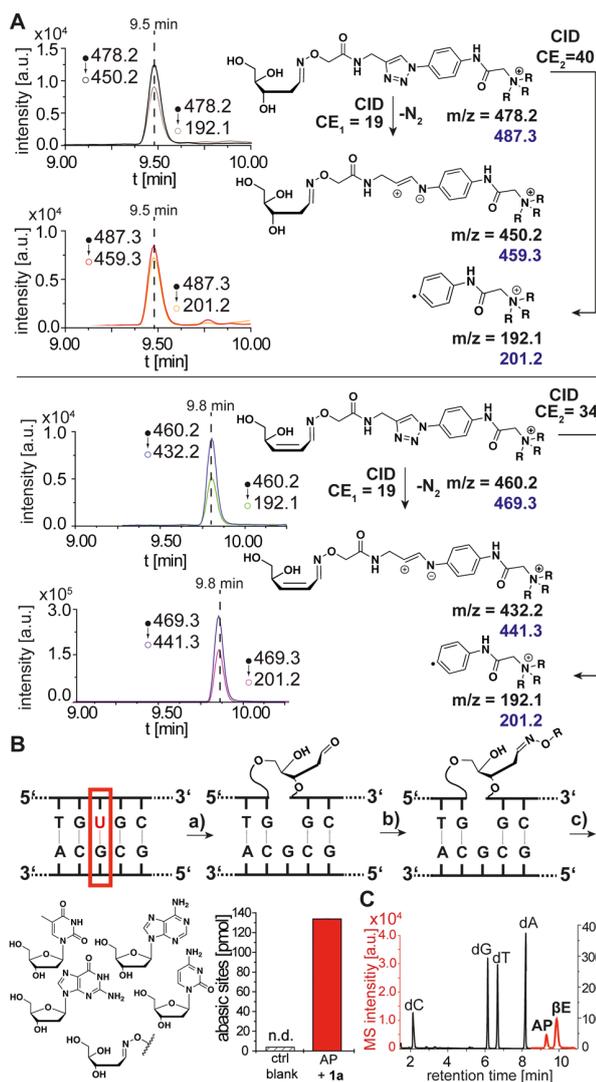


Figure 2. (A) Fragmentation pattern of 9a/b and 10a/b, which give defined daughter ions after loss of N₂ or an aryl radical. (B) Reaction of reagent 1 with a defined AP-site created inside a DNA duplex. Quantification of the reaction product 9a after the treatment of the oligodeoxynucleotide with 1a. (C) UHPLC-chromatogram of the canonical bases plus the AP (9a)- and β E-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 glyco-

lyase 1 (OGG1) and observed the formation of both an AP- and a β E-site (Supporting Information, section 3.7).

Quantification of Abasic- and β E-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 β E-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 β E-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} β E-

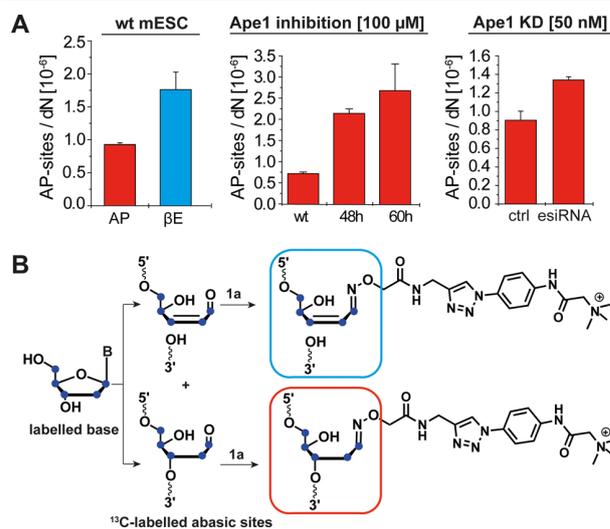


Figure 3. Quantification of BER intermediates and isotope tracing studies. (A) Quantitative data of global levels of AP- and β E-sites from mESCs. Inhibition and knockdown of Ape1 yields an increased amount of AP-sites. (B) Feeding of mESC cultures with labeled nucleosides results in the formation of ribose-labeled AP- and β E-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 β E-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 β E-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 \sim 9200 β E-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-

and β E-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 β E-sites (Figure S9). Increasing levels of AP- and β E-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 Ape1²⁸ 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 CRT0044867³⁰ 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 β E-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 β E-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 ¹²C atoms were exchanged against ¹³C. Correspondingly, the AP- and β E-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

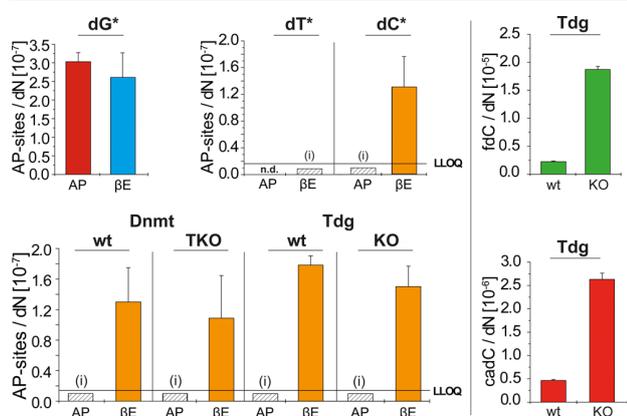


Figure 4. Quantification of labeled AP- and β E-intermediates of the isotope tracing studies and quantification of the fdC and cadC content in Tdg wt and KO cells. n.d., not detectable. (i) Levels were over the LOD, but under LLOQ.

¹³C₁₀-dG an incorporation of 93%. ¹³C₉-dC was incorporated to 40%, and the ¹³C₉-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 β E-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 β E-sites. We measured 2.7×10^{-7} sites per dN, which corresponds to a steady-state level of around ~1500 AP- and β E-sites derived from dG (Figure 4). The high level of β E-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 β E-sites are below 10^{-8} per dN, which amounts to less than ~100 β E-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). β E-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 β E-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 β E-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 β E-site changes, arguing that at the steady state the overwhelming majority of β E-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 β E-sites that are negligible. If BER is however a major process, the fact that we do not see changing AP- and β E-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 β E-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

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 β E-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.; Hagemeyer, 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.; Michalakakis, 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.; Michalakakis, 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.

- (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.

3.2.2 5-Formylcytosin ist vermutlich eine semipermanente Base an definierten Genompositionen

„5-Formylcytosine could be a semipermanent base in specific genome sites“ Meng Su, [Angie Kirchner](#), Samuele Stazzioni, Markus Müller, Mirko Wagner, Arne S. Schröder & Thomas Carell

Prolog

Als Oxidationsprodukt der Tet-Enzyme gilt die epigenetische Base fdC als Intermediat eines aktiven Demethylierungsweges. Die genaue Funktion ist jedoch nicht bekannt. Um mehr über die Rolle von fdC an definierten Genompositionen zu verstehen, wurden verschiedene Sequenziermethoden entwickelt. Dies liefert zwar Informationen über die Verteilung der Base zu einem bestimmten Zeitpunkt, vernachlässigt aber die Dynamik der epigenetischen Informationsschicht. Um einen tieferen Einblick in die Funktion der Base an definierten Genompositionen zu erlangen, ist es daher wünschenswert fdC an einer bestimmten Stelle über einen gewissen Zeitraum hinweg zu verfolgen. Das nachfolgende Manuskript beschreibt hierzu die Entwicklung einer sequenzspezifischen chemischen Derivatisierungsmethode, die in Kombination mit der *droplet digital (dd) PCR*, eine derartige Verfolgung ermöglicht. Erste Ergebnisse zeigen, dass die Glykosylase Tdg nur die Hälfte der fdCs an bestimmten Genompositionen in pluripotenten Zellen entfernt. Dies lässt vermuten, dass es sich bei der epigenetischen Base fC um eine DNA-Modifikation mit semipermanenten Eigenschaften handelt.

Autorenbeitrag

Die verwendeten, murinen Stammzelllinien wurden von mir kultiviert und entsprechend der durchgeführten Experimente behandelt. Die Isolation der genomischen DNA wurde von mir in Zusammenarbeit mit *Dr. Meng Su* durchgeführt. Ferner war ich an der Dateninterpretation und der Verfassung des Manuskripts beteiligt.

Lizenz

Kopie der Publikation mit Erlaubnis des Verlags; Copyright 2016 *John Wiley and Sons*.



5-Formylcytosine Could Be a Semipermanent Base in Specific Genome Sites

Meng Su, Angie Kirchner, Samuele Stazzoni, Markus Müller, Mirko Wagner, Arne Schröder, and Thomas Carell*

Abstract: 5-Formyl-2'-deoxycytosine (fdC) is a recently discovered epigenetic base in the genome of stem cells, with yet unknown functions. Sequencing data show that the base is enriched in CpG islands of promoters and hence likely involved in the regulation of transcription during cellular differentiation. fdC is known to be recognized and excised by the enzyme thymine-DNA-glycosylase (Tdg). As such, fdC is believed to function as an intermediate during active demethylation. In order to understand the function of the new epigenetic base fdC, it is important to analyze its formation and removal at defined genomic sites. Here, we report a new method that combines sequence-specific chemical derivatization of fdC with droplet digital PCR that enables such analysis. We show initial data, indicating that the repair protein Tdg removes only 50 % of the fdCs at a given genomic site, arguing that fdC is a semipermanent base.

DNA contains besides the sequence information a second, epigenetic information level, which encodes how actively the controlled gene is transcribed.^[1] Today, next to the four canonical bases, four additional epigenetic bases are known.^[2] These are 5-methylcytosine (5mC), 5-hydroxymethylcytosine (5hmC),^[3] 5-formylcytosine (5fC),^[4] and 5-carboxycytosine (5caC)^[5] (Figure 1a). Over the last years, sensitive mass spectrometry based methods have helped to reveal the global levels of these epigenetic bases in stem cells^[4,6] and tissues including the brain.^[7] In order to learn about the levels and the distribution of the epigenetic bases at specific sites in the genome, different sequencing methods have been developed^[8] in which selective chemical derivatization of the bases is performed,^[9] sometimes in combination with bisulfite sequencing.^[9c,10] Although these methods provide information about the distribution of the bases at a given time point, it is a hallmark of epigenetic information that it changes dynamically. To gain deeper insight into the dynamics of the epigenetic information layer at a single position in the genome, it is therefore essential to develop methods that can be used to follow the changes of, for example, fdC at a specific location in the genome over time.^[11] A perfect

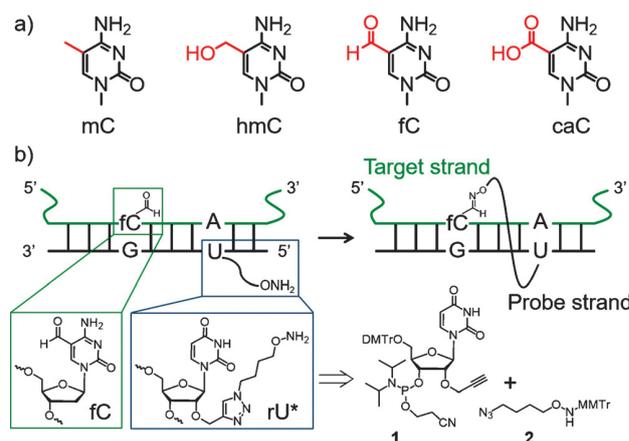


Figure 1. a) Structures of cytosine epigenetic modifications; b) schematic representation of the fdC detection strategy, and building blocks used including the click-chemistry-based assembly of the rU* probe molecule.

method will ultimately allow parallel monitoring of fdC dynamics at different genomic sites.

The central question addressed in this manuscript is: Are the measured global data of the past averages from different processes at different positions in the genome, or do they reflect what is happening at an individual site in the genome? To answer this question, we developed a sequence-specific chemical derivatization method that we then used in combination with droplet digital PCR to monitor the epigenetic base fdC at different loci directly in the genome of stem cells.

For the sequence-specific localization of fdC in the genome, we utilize a small probe oligonucleotide (Figure 1b, Table S1), which contains a hydroxylamine tether that is able to form a covalent linkage with fdC so that the probe strand is subsequently tightly bound to the target.^[4] We examined systematically different linker lengths, linker attachment points, and distances. Best results were obtained when we incorporated the 2'-O-propargyl uridine using its phosphoramidite **1** into the probe oligonucleotide and attached the azido-C4-hydroxylamine **2** using the Cu^I-catalyzed version of the Huisgen reaction (click reaction).^[12] We protected the hydroxylamine unit for the click reaction with a monomethoxytrityl group (MMTr), which was cleaved afterwards with acetic acid at 25 °C. This brief exposure of the probe oligonucleotide to acidic conditions did not cause significant depurination. After solid-phase synthesis, click modification of the oligonucleotide, and a final purification step (Figure S1), we obtained oligonucleotides with different sequen-

[*] M. Sc. M. Su, M. Sc. A. Kirchner, M. Sc. S. Stazzoni, Dr. M. Müller, Dr. M. Wagner, M. Sc. A. Schröder, Prof. Dr. T. Carell
Center for Integrated Protein Science at the
Department of Chemistry, Ludwig-Maximilians-Universität München
Butenandtstrasse 5–13, 81377 München (Germany)
E-mail: thomas.carell@lmu.de
Homepage: <http://www.carellgroup.de>

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.201605994>.

ces and lengths containing an rU-hydroxylamine base (rU*) at different positions for reaction with the fdC base on the target strand. For the following experiments, we prepared 13-mer long oligonucleotides.

To investigate at which position the linker in the probe strand would react best with fdC in the target strand, we varied the position of rU* relative to fdC and explored different reaction conditions. Excellent results were finally obtained when probe strand **P1**, containing rU* exactly four base pairs in the 5'-direction relative to fdC, was hybridized to the fdC target strand **T1** in the presence of catalytic amounts of 4-methoxyaniline (Figure 2a). With this catalyst, the crosslinking reaction is complete after 24 h with yields exceeding 95%. Without the catalyst, only about 50% yield could be obtained (Figure 2b).

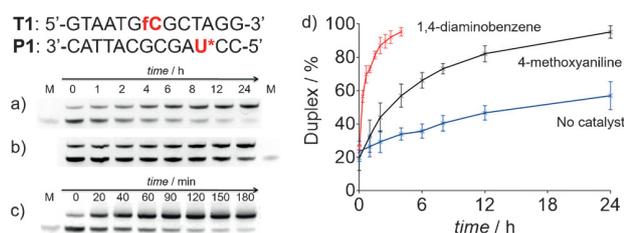


Figure 2. Denaturing PAGE gel showing the duplex formation between **T1** and **P1** at 25°C: a) with the catalyst 4-methoxyaniline; b) without a catalyst; c) with the catalyst 1,4-diaminobenzene. d) Quantification of the DNA duplex formation during the reaction. Black: catalyst 4-methoxyaniline, blue: no catalyst, red: catalyst 1,4-diaminobenzene. Error bars represent the standard error of the mean calculated from three replicates. Conditions: 2 μ M oligonucleotides, 100 mM NaCl, 10 mM NaOAc buffer pH 6.0, 10 mM 4-methoxyaniline. M = single-strand marker. The time point 0 is after re-annealing.

In order to increase the rate of the reaction, we tested other catalysts. We observed the best results when we used 1,4-diaminobenzene as a catalyst, in which case the crosslinking reaction between **T1** and **P1** is completed already after 3 h (Figure 2c). Duplex formation (**T1:P1**) was analyzed using denaturing PAGE and quantified by fluorescence (Figure 2d).

When fdC is located one base pair further away from rU* without changes in the probe strand, we observe slower reaction (Figure S2). These results show that rU* placed four or five bases away from fdC in the 5'-direction to fdC allows the tether to reach the formyl group of fdC via the major groove of the duplex (Figure S3).

MALDI-TOF data confirmed that the crosslinks form as expected (Figure S4). For the crosslinked duplex **T1:P1**, we obtained the correct molecular weight with $m/z_{\text{found}} = 8081.9$ ($m/z_{\text{calc}} = 8084.7$). As expected, the oxime formation reaction between **T1** and **P1** leads to a higher melting temperature of the hybridized and crosslinked duplex (Figure S5). Typically, we observed that the un-crosslinked 13-mer duplex melts at around 44°C. The duplex after crosslink formation shows a melting temperature of above 80°C.

Because pyrimidine bases are able to react with nucleophiles also at the C6 position in a Michael-type reaction, which is the basis for bisulfite sequencing, we next tested

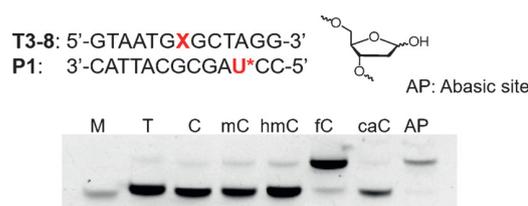


Figure 3. Denaturing PAGE gel showing duplex formation of **T3-8** and **P1** at 25°C after 24 h.

whether the reaction of rU* is possible with other pyrimidines (Figure 3). To our delight, hybridization of the rU*-containing probe strand with target strands containing dT, dC, mdC, hmdC, and cadC (**T3-8**) gave no reaction. Reaction is, however, observed with abasic sites. This is important because fdC and cadC are substrates for base excision repair and hence could in principle be precursors for abasic sites.^[13] In this sense, rU* always reports the presence of fdC and also potentially of fdC- and cadC-derived abasic sites.

We finally turned our procedure for sequence-specific fdC detection into a method for detecting single fdC bases at a defined position in whole genomes. To this end, we coupled the chemistry to droplet digital PCR^[14]-based amplification and readout.

In the first step genomic target DNA (**Tg**) was isolated from mouse embryonic stem cells (mESCs) at different time points during priming from naive cells. We also isolated genomic DNA from mESCs with a knockout of the Tdg repair enzyme (Tdg^{-/-}) to block excision and repair of fdC and cadC. We finally also isolated genomic DNA from mESCs lacking any of the three methyltransferases (Dnmt1, 3a, and 3b). These stem cells lack mdC and are hence unable to produce the oxidized xdC ($x = \text{hm}, \text{f}, \text{and ca}$) epigenetic bases. This genomic DNA served in our studies consequently as a negative control. For analysis, we selected two different fdC sites that were reported to have high fdC contents.^[10c] We focused initially on the 30020539th site of chromosome 16 *Mus musculus* (MM9) located on the exon 3 of 632428C04Rik. It was found to contain 23% of fdC based on redBS-sequencing. The second site we studied was the 8846677th site of chromosome 15, which is located in noncoding DNA. This site was reported to contain 32% of fdC.

For the first site, we reacted a 25-mer probe (**P2**, SI) containing the rU* base with **Tg** using 1,4-diaminobenzene as the catalyst. In the absence of fdC, a covalent bond between **P2** and **Tg** cannot form. To remove the excess of probe, we loaded the **Tg:P2** complex onto an NEB Monarch DNA cleanup column and rinsed the column with wash buffer to elute oligonucleotides shorter than the 50-mer, which is the unbound **P2**. After this washing, we eluted the **Tg:P2** complex with TE buffer. UV/Vis analysis of the eluted material showed a typical gDNA spectrum. We next added a 70-mer 5'-phosphorylated reporter strand (**R1**, SI) which hybridizes with an 18-nt stretch directly adjacent to the probe strand and ligated both probe and the reporter at 60°C upon addition of Ampligase to form **R1-P2** as depicted schematically in Figure 4a.

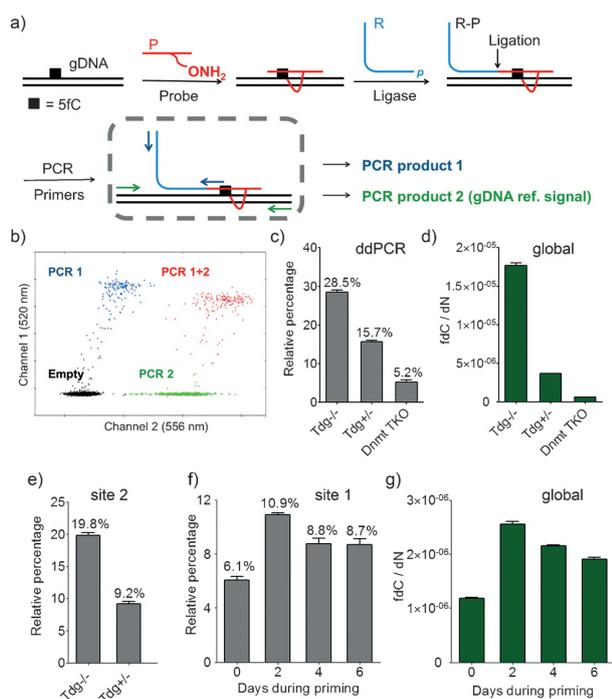


Figure 4. The fdC sequencing method: a) schematic representation of the method, black line: gDNA; red segments: fdC probe; blue segments: reporter strands; arrows: PCR primer pairs. b) Typical 2D plot of droplet fluorescence. c) Cluster ratios for position 1 in $Tdg^{-/-}$, $Tdg^{+/-}$, and Dnmt TKO mES cells. d) Global fdC quantification in $Tdg^{-/-}$, $Tdg^{+/-}$, and Dnmt TKO mES cells using our LC-MS method. e) Cluster ratios for position 2 in $Tdg^{-/-}$ and $Tdg^{+/-}$ mES cells. f) Cluster ratios for position 1 in wild-type mES cells at different days after priming. g) Global quantification data for the wild-type mES cells measured by LC-MS.

We next added two sets of primers to the assay (blue and green arrows, Figure 4a) to amplify the ligation product relative to the target duplex. Importantly, the blue primers recognize only the hybrid **R1–P2** probe generated in the ligation step, while the green primers indicate the presence of gDNA. The amplification was monitored with two different TaqMan probes which showed fluorescence at 520 and 556 nm. This relative detection is needed to normalize on the amount of input gDNA. Because conventional real-time PCR is known to become inaccurate when copy number differences less than the 10-fold need to be resolved, we used droplet digital PCR. In this method, small droplets are generated with one droplet containing a maximum of one of the fully assembled analysis constructs shown in Figure 4a. The PCR reaction takes place in the droplets, producing a specific signal. Subsequent fluorescence-based counting of each individual droplet yields numbers from which one can accurately calculate the amount of fdC, even if the fdC values are very low. A representative plot of the data is given in Figure 4b. Empty drops give no PCR signal (black dots in Figure 4b). Drops containing only **Tg** give only the PCR signal from the green primers (green dots in Figure 4b). Blue dots are obtained due to the dissociation of the ligated product **R1–P2** from **Tg** in the ligation process which is

performed at 60°C for 10 h. The red signals are finally generated from droplets that contain both PCR products. For the calculation, please see the Supporting Information.

Using the method, we first studied mES cells lacking the Tdg enzyme ($Tdg^{-/-}$). A rather high level of 28.5% fdC was measured at the first locus (Figure 4c) in agreement with the results from redBS-seq.^[10c] When we performed the study, however, with mESCs having an active Tdg repair enzyme ($Tdg^{+/-}$) we measured that the fdC level drops at this particular position to 15.7% (Figure 4c). This is very important because it shows that Tdg removes only half of the fdCs at a given site and also unusual since repair glycosylases are known to find basically all possible substrates. The result underpins the high dynamics of fdC formation and repair at a given site. When we studied the fdC content at this location in mESCs lacking any methyltransferase (Dnmt TKO) the fdC level drops as expected to a little more than 5%, showing that the reported levels of fdC in the $Tdg^{+/-}$ cells are real and not an artifact. In order to elucidate whether single-site fdC levels (Figure 4c) follow global genomic fdC levels, we quantified the total levels of fdC in these cells (Figure 4d). These global data are in good agreement with the data obtained from single-site fdC quantifications. Thus, our new data make a scenario where fdC is fully removed at one site and shielded from repair at another place unlikely. Instead, fdC is even at a given position only partially removed in a cell population. Alternatively, it may be that Tdg removes fdC differently on the two chromosomes, which however needs further investigation.

In order to verify the data, we repeated the Tdg study at a second genomic site (8846677th nucleoside of chromosome 15). For this site, we designed a new probe strand **P3** and a new reporter strand **R2** and performed again ddPCR with two sets of primers (Figure S5). When we compare the data obtained from $Tdg^{-/-}$ cells with the data from $Tdg^{+/-}$ cells, again only a 50% reduction of the fdC level is shown at this position, in full agreement with the data above obtained from the first position (Figure 4e).

We finally performed a kinetic study in which we monitored the fdC development at the first position during priming of stem cells (Figure 4f). We see that the fdC levels rise at the given position with a strong increase in the early phase of priming, followed by a small decline phase and finally stable values (Figure 4f) again in agreement with the global data that we measured using our reported method (Figure 4g).

The fact that our method provides the same trends as seen in the global data at a single genomic site makes us confident that our method is robust and reliable in reporting what happens at an individual site. Because single-site and global data correlate, we have now first evidence that the reported global trends are reflecting what happens at each individual fdC site, rather than evening out largely different dynamics at separate sites. Another interesting result of this study is that the repair enzyme Tdg removes only half of the fdC bases at a given genomic site in an mESC population, which argues that fdC is a semipermanent base at a given position in the genome.

Experimental Section

Probe crosslinking: gDNA solution (1.2 µg), fdC probe (1 µM, 2 µL), NaH₂PO₄-Na₂HPO₄ buffer (200 mM, pH 6.0, 2 µL), NaCl aq. (1.5 M, 2 µL), and doubly distilled (dd) H₂O were mixed to a final volume of 18 µL. The mixture was heated to 95 °C for 3 min and then cooled down rapidly to 25 °C. 1,4-Diaminobenzene aq. (10 mM, 2 µL) was added and the reaction vial was shaken for 6 h at 25 °C. The mixture was neutralized with Na₂HPO₄ aq. (200 mM, 40 µL) before purification with the NEB Monarch PCR DNA Cleanup Kit.

Ligation: The above-described gDNA solution (300 ng), reporter strand (20 nM, 1 µL), Ampligase reaction buffer (10×, 2 µL), Ampligase from Epicentre (10 U), and ddH₂O were mixed to a final volume of 20 µL. The mixture was heated to 95 °C for 3 min, and then 94 °C for 1 min, 60 °C for 1 h, and back to 94 °C for 10 cycles. Then, the reaction mixture was diluted with Tris-HCl buffer (200 mM, pH 7.6, 50 µL) before purification using the NEB Monarch PCR DNA Cleanup Kit.

Droplet digital PCR (ddPCR) was conducted on a Bio-Rad QX100 ddPCR System. For one reaction, gDNA (6 ng), four primers (18 µM each, 1 µL), two TaqMan probes (5 µM each, 1 µL), digital PCR Supermix for Probes (no dUTP, 2×, 10 µL), and ddH₂O were mixed to a final volume of 20 µL. PCR cycle: heating at 95 °C for 10 min, 94 °C for 30 s, and 64 °C for 1 min for 35 cycles, then 98 °C for 10 min and cooling to 12 °C, with a temperature ramp of 2 °Cs⁻¹. For a detailed description please see the Supporting Information.

Acknowledgments

We thank K. Hufnagel for preparing the phosphoramidites of the epigenetic bases. This project received funding from Deutsche Forschungsgemeinschaft SFB1032 and the Excellence Cluster CiPS^M (EXC114). Further support was obtained from the European Union's Horizon 2020 research and innovation program under grant agreement no. 642023 (ITN clickgene).

Keywords: 5-formylcytosine · click chemistry · droplet digital PCR · epigenetic bases · genomic DNA

How to cite: *Angew. Chem. Int. Ed.* **2016**, *55*, 11797–11800
Angew. Chem. **2016**, *128*, 11974–11978

- [1] P. A. Jones, *Nat. Rev. Genet.* **2012**, *13*, 484–492.
[2] T. Carell, C. Brandmayr, A. Hienzsch, M. Müller, D. Pearson, V. Reiter, I. Thoma, P. Thumbs, M. Wagner, *Angew. Chem. Int. Ed.* **2012**, *51*, 7110–7131; *Angew. Chem.* **2012**, *124*, 7220–7242.
[3] a) S. Kriaucionis, N. Heintz, *Science* **2009**, *324*, 929–930; b) 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.
[4] T. Pfaffeneder, B. Hackner, M. Truss, M. Münzel, M. Müller, C. Deiml, C. Hagemeyer, T. Carell, *Angew. Chem. Int. Ed.* **2011**, *50*, 7008–7012; *Angew. Chem.* **2011**, *123*, 7146–7150.
[5] a) S. Ito, L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He, Y. Zhang, *Science* **2011**, *333*, 1300–1303; b) 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.
[6] S. Schiesser, B. Hackner, T. Pfaffeneder, M. Müller, C. Hagemeyer, M. Truss, T. Carell, *Angew. Chem. Int. Ed.* **2012**, *51*, 6516–6520; *Angew. Chem.* **2012**, *124*, 6622–6626.
[7] a) M. Münzel, D. Globisch, T. Brückl, M. Wagner, V. Welzmler, S. Michalakis, M. Müller, M. Biel, T. Carell, *Angew. Chem. Int. Ed.* **2010**, *49*, 5375–5377; *Angew. Chem.* **2010**, *122*, 5503–5505; b) D. Globisch, M. Münzel, M. Müller, S. Michalakis, M. Wagner, S. Koch, T. Brückl, M. Biel, T. Carell, *PLoS ONE* **2010**, *5*, e15367; c) M. Wagner, J. Steinbacher, T. F. J. Kraus, S. Michalakis, B. Hackner, T. Pfaffeneder, A. Perera, M. Müller, A. Giese, H. A. Kretzschmar, T. Carell, *Angew. Chem. Int. Ed.* **2015**, *54*, 12511–12514; *Angew. Chem.* **2015**, *127*, 12691–12695.
[8] a) N. Plongthongkum, D. H. Diep, K. Zhang, *Nat. Rev. Genet.* **2014**, *15*, 647–661; b) M. J. Booth, E.-A. Raiber, S. Balasubramanian, *Chem. Rev.* **2015**, *115*, 2240–2254.
[9] a) W. A. Pastor, U. J. Pape, Y. Huang, H. R. Henderson, R. Lister, M. Ko, E. M. McLoughlin, Y. Brudno, S. Mahapatra, P. Kapranov, M. Tahiliani, G. Q. Daley, X. S. Liu, J. R. Ecker, P. M. Milos, S. Agarwal, A. Rao, *Nature* **2011**, *473*, 394–397; b) E.-A. Raiber, D. Beraldi, G. Ficiz, H. Burgess, M. Branco, P. Murat, D. Oxley, M. Booth, W. Reik, S. Balasubramanian, *Genome Biol.* **2012**, *13*, R69; c) C.-X. Song, K. E. Szulwach, Q. Dai, Y. Fu, S.-Q. Mao, L. Lin, C. Street, Y. Li, M. Poidevin, H. Wu, J. Gao, P. Liu, L. Li, G.-L. Xu, P. Jin, C. He, *Cell* **2013**, *153*, 678–691; d) B. Xia, D. Han, X. Lu, Z. Sun, A. Zhou, Q. Yin, H. Zeng, M. Liu, X. Jiang, W. Xie, C. He, C. Yi, *Nat. Methods* **2015**, *12*, 1047–1050.
[10] a) M. Yu, G. C. Hon, K. E. Szulwach, C.-X. Song, L. Zhang, A. Kim, X. Li, Q. Dai, Y. Shen, B. Park, J.-H. Min, P. Jin, B. Ren, C. He, *Cell* **2012**, *149*, 1368–1380; b) M. J. Booth, M. R. Branco, G. Ficiz, D. Oxley, F. Krueger, W. Reik, S. Balasubramanian, *Science* **2012**, *336*, 934–937; c) M. J. Booth, G. Marsico, M. Bachman, D. Beraldi, S. Balasubramanian, *Nat. Chem.* **2014**, *6*, 435–440; d) X. Lu, C.-X. Song, K. Szulwach, Z. Wang, P. Weidenbacher, P. Jin, C. He, *J. Am. Chem. Soc.* **2013**, *135*, 9315–9317.
[11] a) A. Nomura, K. Sugizaki, H. Yanagisawa, A. Okamoto, *Chem. Commun.* **2011**, *47*, 8277–8279; b) J. Duprey, G. A. Bullen, Z.-Y. Zhao, D. M. Bassani, A. F. A. Peacock, J. Wilkie, J. H. R. Tucker, *ACS Chem. Biol.* **2016**, *11*, 717–721.
[12] a) P. M. E. Gramlich, S. Warncke, J. Gierlich, T. Carell, *Angew. Chem. Int. Ed.* **2008**, *47*, 3442–3444; *Angew. Chem.* **2008**, *120*, 3491–3493; b) J. Willibald, J. Harder, K. Sparrer, K.-K. Conzelmann, T. Carell, *J. Am. Chem. Soc.* **2012**, *134*, 12330–12333.
[13] A. Maiti, A. C. Drohat, *J. Biol. Chem.* **2011**, *286*, 35334–35338.
[14] B. J. Hindson, K. D. Ness, D. A. Masquelier, P. Belgrader, N. J. Heredia, A. J. Makarewicz, I. J. Bright, M. Y. Lucero, A. L. Hiddessen, T. C. Legler, T. K. Kitano, M. R. Hodel, J. F. Petersen, P. W. Wyatt, E. R. Steenblock, P. H. Shah, L. J. Bousse, C. B. Troup, J. C. Mellen, D. K. Wittmann, N. G. Erndt, T. H. Cauley, R. T. Koehler, A. P. So, S. Dube, K. A. Rose, L. Montesclaros, S. Wang, D. P. Stumbo, S. P. Hodges, S. Romine, F. P. Milanovich, H. E. White, J. F. Regan, G. A. Karlin-Neumann, C. M. Hindson, S. Saxonov, B. W. Colston, *Anal. Chem.* **2011**, *83*, 8604–8610.
[15] M. Wendeler, L. Grinberg, X. Wang, P. E. Dawson, M. Baca, *Bioconjugate Chem.* **2014**, *25*, 93–101.

Received: June 20, 2016

Published online: August 25, 2016

3.2.3 Involvierung der BER an der Entfernung modifizierter DNA-Basen

3.2.3.1 Prolog

In der genomischen DNA existieren neben den oxidierten Cytidin-Derivaten auch die Thymidin-Derivate 5-Hydroxymethyldeoxyuridin (hmdU) und 5-Formyldeoxyuridin (fdU).^[85] Diese werden sowohl spontan durch *Reactive Oxygen Species* (ROS), als auch enzymatisch gebildet und wurden bereits in der genomischen DNA zahlreicher Organismen entdeckt.^[162, 288] Sowohl hmdU als auch fdU führen zu genomischer Instabilität und potentiell mutagenen DNA-Schäden. In diesem Kapitel wird zunächst der Beitrag der Glykosylase Smug1 an der Entfernung von genomischem hmdU und fdU in pluripotenten Zellen untersucht und die daraus resultierenden Reparaturintermediate unter Verwendung der in Abschnitt 3.2.1 beschriebenen Methode quantifiziert. Abschließend wird die Beteiligung von Tdg und den Neil-Enzymen an der aktiven Demethylierung mit verschiedenen *Knockout*-mESC Linien eingehender betrachtet.

3.2.3.2 Ergebnisse

Quantifizierung von hmdU, fdU und abasischer Stellen in Smug1^{-/-} mESCs

Zur Bestimmung der Anzahl an oxidierten dT-Derivaten im Genom muriner Stammzellen wurde eine mESC-Linie in der die DNA-Glykosylase Smug1 ausgeschaltet wurde massenspektrometrisch untersucht. Diese *Smug1^{-/-}* mESC-Linie wurde im Rahmen meiner Masterarbeit mit Hilfe des CRISPR-Cas9 Systems generiert. Die Verifizierung der *Knockout*-Zelllinie erfolgte im Zuge dieser Dissertation und wird in Abschnitt 3.2.3.5 beschrieben. Zur massenspektrometrischen Bestimmung wurde die Zelllinie für fünf Tage unter Serum/LIF *Priming*-Bedingungen kultiviert. Nach Isolation der genomischen DNA wurden die hmdU-Level bestimmt und die Zahl der AP-Stellen unter Zuhilfenahme des entwickelten Derivatisierungsreagenzes (Abschnitt 3.2.1) quantifiziert. Die Ergebnisse sind in Abbildung 40A dargestellt. Es konnte gezeigt werden, dass die hmdU-Werte im *Smug1^{-/-}* verglichen mit dem Wildtyp etwa dreimal höher sind. Im Wildtyp befinden sich demnach ~15000 hmdU pro Genom (LLOQ: $4,0 \times 10^{-7}$ /dN; entspricht ~2000 hmdU pro Genom),^[142] wohingegen im *Knockout* ~45000 vorliegen. Dies lässt vermuten, dass weniger abasische Stellen in Abwesenheit der Glykosylase generiert werden, da hmdU hauptsächlich von Smug1 ausgeschnitten wird.^[289] Nach Derivatisierung der genomischen DNA, wie bereits in Abschnitt 3.2.1 beschrieben, wurde die DNA mittels UHPLC-MS/MS untersucht. Ein geringer Unterschied konnte detektiert werden (siehe Abbildung 40A). Ob es sich um einen tatsächlichen Unterschied handelt, müsste mit weiteren biologischen Replikaten überprüft werden. Die Fähigkeit der monofunktionalen Glykosylasen Tdg und Mbd4 hmdU aus einzel- und doppelsträngiger DNA auszuschneiden ist bereits bekannt.^[166, 213]

In Abwesenheit von Smug1 könnten diese Glykosylasen diese Aufgabe übernehmen, was die Differenz an AP-Stellen relativieren könnte. Der Anstieg des β -Eliminierungsproduktes, könnte durch Zusammenarbeit der monofunktionalen Glykosylasen mit den bifunktionalen Neil-Glykosylasen erklärt werden.^[210] Kemmerich *et al.* konnten bereits eine größere Toleranz in *Smug1*^{-/-} Fibroblasten gegenüber externer hmdU Zugabe beobachten.^[289] Diese Toleranz des generierten Smug1 KO sollte nun ebenfalls dahingehend untersucht werden. Hierzu wurden dem Kulturmedium unterschiedliche Mengen an hmdU in den letzten 24 h des fünftägigen *Primings* zugesetzt. Nach DNA-Isolation erfolgte eine Analyse der hmdU-Level. Die Ergebnisse sind in Abbildung 40B illustriert.

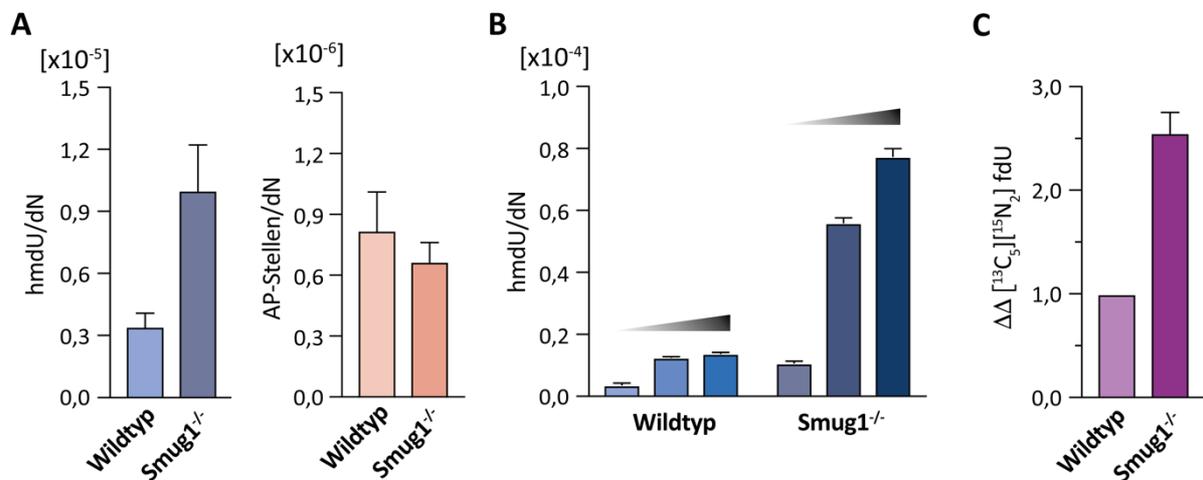


Abbildung 40: Level der oxidierten Uracil-Derivate und abasische Stellen in Smug1 defizienten mESCs. **A)** Quantifizierungsdaten der Modifikation hmdU und der globalen abasischen Stellen im *Smug1*^{-/-} und dem korrespondierenden Wildtyp. Abgebildet sind die normierten Quantifizierungsergebnisse dreier biologischer Replikate, gemessen als technisches Triplikat. **B)** Genomische hmdU Werte nach externer Zugabe von hmdU (Farbgradient= 0 μ M, 1 μ M und 2 μ M). Dargestellt sind die normierten Quantifizierungsergebnisse einer biologischen Probe (N=1), gemessen als technisches Triplikat. **C)** Quantifizierte Daten nach Zugabe von 50 μ M ¹³C₅¹⁵N₂-fdC. Die Flächenintegrale wurden zunächst in Relation gesetzt Δ_{1a} : $A(^{13}\text{C}_5^{15}\text{N}_2\text{-fdU})/A(\text{dG})$, sowie Δ_{1b} : $A(^{13}\text{C}_5^{15}\text{N}_2\text{-fdC})/A(\text{dG})$ und anschließend auf die inkorporierte Nukleosid-Menge Δ_2 : Δ_{1a}/Δ_{1b} bezogen. N=3, gemessen als technisches Triplikat.

Mit steigender hmdU Konzentration sind, wie zu erwarten, höhere Werte an hmdU im Genom der mESCs zu beobachten. Im *Smug1*^{-/-} sind jedoch deutlich höhere hmdU-Werte im Genom detektierbar. Im Smug1 KO konnten $\sim 4,3 \times 10^5$ hmdU pro Genom eingebaut werden, während der Wildtyp bei einer Konzentration von 2 μ M $\sim 5,4 \times 10^4$ inkorporiert. Vermutlich führt die Aufnahme von hmdU zu einer verstärkten Basenexzisionsreparatur und so zur Entstehung schädlicher abasischer Stellen, die zu Strangbrüchen führen können. Die Menge von $\sim 3,0 \times 10^4$ hmdU im Genom, scheint der Grenzwert für diese Wildtyp mESC-Linie zu sein.

Höhere Konzentrationen an hmdU führten zu deutlich sichtbarem Zelltod. Die resultierende Zahl abasischer Stellen dieses Experiments wurde ebenfalls untersucht, führte aber zu keinem eindeutigen Ergebnis aufgrund der hohen Datenvarianz. Die Prozessierung der abasischen Stellen, geschieht vermutlich zu schnell und durch eine massenspektrometrische Quantifizierung kann nur eine Momentaufnahme der Anzahl abasischer Stellen beobachtet werden. Die Generierung vergleichbarer biologischer Replikate erwies sich hier als sehr schwierig und war abhängig von mehreren Faktoren, wie Zelldichte und Dauer der hmdU-Fütterung. Mehr biologische Replikate, kürzere Fütterungszeiten und eine Normierung auf hmdU-Inkorporation könnten hier zu besseren Ergebnissen führen.

Quantifizierung der Reparaturintermediate nach Einbringung von $^{13}\text{C}_5^{15}\text{N}_2\text{-fdC}$ in *Smug1*^{-/-} mESCs

Im Verlauf dieser Arbeit (siehe Abbildung 32) wurde eine Desaminierung von gefüttertem, isotopenmarkiertem fdC zu fdU beobachtet. In einem unabhängigen Experiment sollte daher die Entstehung von fdU in mESCs untersucht werden. Hierzu wurde isotopenmarkiertes $^{13}\text{C}_5^{15}\text{N}_2\text{-fdC}$ in die genomische DNA der *Smug1*^{-/-} mESC-Linie, sowie des korrespondierenden Wildtyps, eingebracht und die daraus entstandene Menge an genomischem, markierten fdU massenspektrometrisch bestimmt. Da die Menge an markiertem fdU über dem *Upper Limit of Quantification* (ULOQ) lag, konnte das Level nicht absolut bestimmt werden. Zum Vergleich der erhaltenen Daten wurden die MS-Flächenintegrale auf die jeweilige UV-dG Fläche, sowie die Menge an inkorporiertem fdC bezogen und in Relation zum Wildtyp gesetzt (Abbildung 40C). Im *Knockout* konnte eine 2,5-fache Zunahme an isotopenmarkiertem fdU beobachtet werden. Dies bestätigt die Glykosylaseaktivität von *Smug1*, neben hmdU, auch fdU aus dem Genom zu entfernen.^[290] Auch die Erhöhung der globalen fdU-Level im *Knockout* entspricht etwa der Zunahme an hmdU, was die Vermutung nahe legt, dass hmdU die Verläuferverbindung von fdU nach Oxidation durch ROS darstellt.

Die Level der BER-Intermediate wurden nach Fütterung mit isotopenmarkiertem fdC ebenfalls bestimmt. Die markierten β -Eliminierungsprodukte lagen unter dem Detektionslimit, wohingegen die markierten AP-Stellen bestimmbar waren. Im *Smug1*^{-/-} konnte ein Anstieg der markierten AP-Stellen ($3,5 \times 10^{-7}/\text{dN}$) im Vergleich zum Wildtyp ($1,8 \times 10^{-7}/\text{dN}$) beobachtet werden. Dies stimmt mit der Hypothese überein, dass in Abwesenheit von *Smug1* andere Glykosylasen der BER-Familie fdU ausschneiden und es so zu einer Zunahme an AP-Stellen kommt. Welchen prozentualen Anteil die jeweilige mono- oder bifunktionale Glykosylase tatsächlich an der Entfernung von hmdU bzw. fdU hat, bedarf jedoch weiterer Untersuchungen.

Quantifizierung abasischer Stellen und β -Eliminierungsprodukte in Neil-Knockout mESCs

Im weiteren Verlauf wurden bifunktionale Glykosylasen untersucht, die bereits mit der aktiven Demethylierung in Verbindung gebracht wurden. Dabei wurden zunächst *Knockout* mESC-Linien der Neil-Glykosylasen betrachtet. Die Neil1, Neil2 sowie Neil1 und 2 *Doppelknockout* mESCs wurden in der Arbeitsgruppe von *Christof Niehrs* unter Verwendung des CRISPR-Cas9 Systems generiert und zur Verfügung gestellt.^[291] Zur Analyse wurden die Zellen, ausgehend einer 2i/LIF-Kultur, für fünf Tage unter Serum/LIF-Bedingungen *geprimed*. Nach anschließender Isolation der genomischen DNA wurden die Zellen mittels UHPLC-MS/MS untersucht. Die Quantifizierungsergebnisse der natürlichen höher oxidierten dC-Spezies der einzelnen *Knockout*-Linien sind in Abbildung 41A dargestellt. In allen *Knockout*-Linien sind geringere mdC-Werte erkennbar. Die geringeren Level an hmdC und fdC in den Neil-*Knockouts* ist wahrscheinlich durch die geringe Ausgangsmenge der Vorläuferverbindung mdC begründet. Die cadC Werte konnten aufgrund der geringen Abundanz nicht bestimmt werden. Die 8oxodG-Level befinden sich im Normalbereich und zeigen keine *Knockout*-spezifische Änderung.

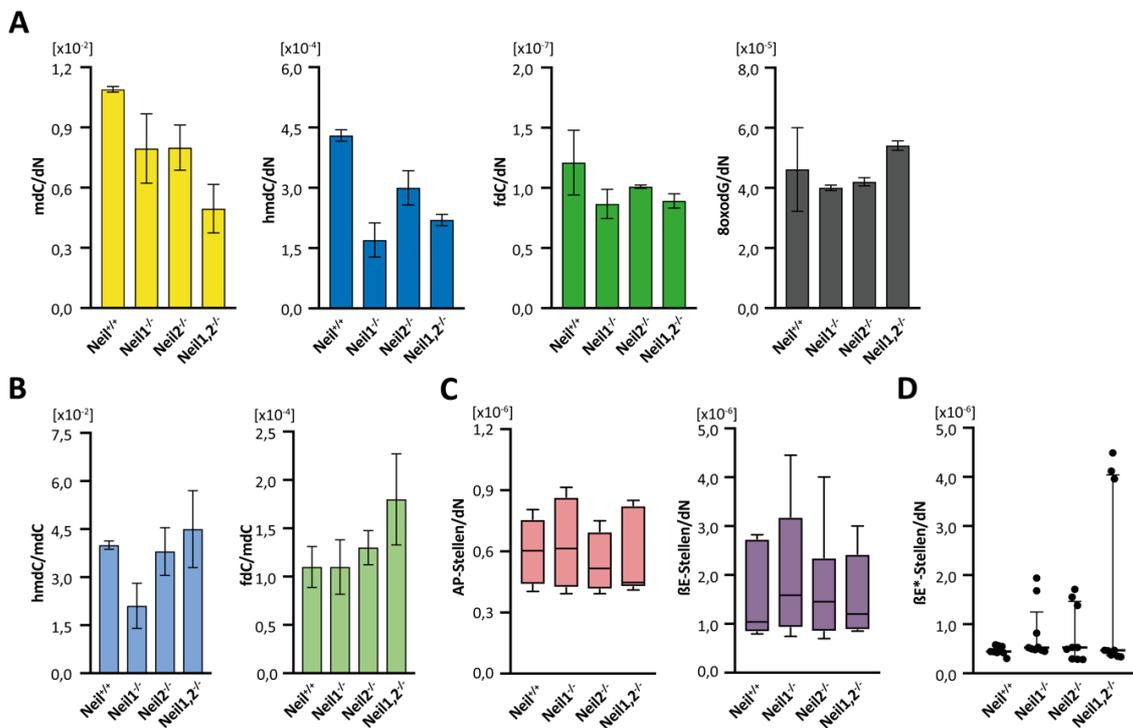


Abbildung 41: Beteiligung der Neil-Enzyme an der aktiven Demethylierung. **A)** Quantifizierungsergebnisse der natürlichen höher oxidierten Cytidin-Spezies. Abgebildet sind die normierten Quantifizierungsergebnisse zweier biologischer Replikate, gemessen als technisches Triplikat. **B)** Normierung der natürlichen höher oxidierten Cytidin-Derivate auf 5-Methyldeoxycytidin. **C)** Level der globalen abasischen Stellen und β -Eliminierungsprodukte in den mESCs. Abgebildet sind die normierten Quantifizierungsergebnisse aus sechs biologischen Replikaten, gemessen als technisches Triplikat. Dargestellt ist der Fehlerbereich mit Median. **D)** Isotopenmarkierte β -Eliminierungsprodukte nach Fütterung mit $^{13}\text{C}_9^{15}\text{N}_3\text{-rC}$. Abgebildet sind die normierten Quantifizierungsergebnisse dreier biologischer Replikate, gemessen als technisches Triplikat. Dargestellt ist der Interquartilabstand mit Median.

Zur besseren Vergleichbarkeit der Zelllinien wurden die natürlichen hmdC- bzw. fdC-Level auf mdC normiert (siehe Abbildung 41B). Im Vergleich zum Wildtyp sind die normierten hmdC-Werte im *Neil1*^{-/-} geringer, in dem *Neil2*^{-/-} und im *Doppelknockout* höher. Im Gegensatz hierzu sind die normierten fdC-Werte in den *Einzelknockouts* verglichen mit dem Wildtyp etwa gleich. Im *Doppelknockout* ist aufgrund der Normierung ein deutlicher Anstieg zu beobachten. Um dieses Ergebnis interpretieren zu können, müssten weitere biologische Daten erhoben werden. Als nächstes wurden weitere Untersuchungen mit Hilfe des in Abschnitt 3.2.1 entwickelten Reagenzes zur Quantifizierung der BER-Intermediate unternommen. Nach Derivatisierung mit dem Detektionsreagenz wurde die genomische DNA massenspektrometrisch untersucht.

Die Ergebnisse sind in Abbildung 41C dargestellt. Die Neil KOs zeigen keine signifikanten Unterschiede bezüglich der Menge an abasischen Stellen. In allen vier mESC-Linien sind etwa 3200 abasische Stellen pro Genom zu beobachten. Bei Betrachtung der globalen β -Eliminierungsintermediate sind, vor allem in den *Knockouts*, deutliche Abweichungen innerhalb der biologischen Replikate zu erkennen, was zu großen Fehlerbalken führt. Auch hier ist keine *Knockout*-spezifische Änderung zu beobachten. Abschließend sollten von dC stammende abasischen Stellen und β -Eliminierungsprodukte durch Einbau einer isotope markierten [¹³C₉¹⁵N₃]-Cytidin-Spezies quantifiziert werden. Da diese Verbindung am Zucker um fünf Masseneinheiten schwerer ist, als das natürliche Homolog, sollten so Erkenntnisse über die BER-Intermediate aus dC-Derivaten erhalten werden. Die Inkorporation der isotope markierten Desoxy-Verbindung betrug in diesen Zelllinien etwa 15 % aller dC. Diese Markierungseffizienz erschien sehr gering, weshalb auf die entsprechende Ribo-Verbindung (¹³C₉¹⁵N₃-rC) für weitere Experimente zurückgegriffen wurde. Die Ribo-Verbindung wird in der Zelle zunächst zum Diphosphat phosphoryliert, dann als Teil des RNA-*Salvage-Pathway* durch die Ribonukleotidreduktase zur Desoxy-Verbindung reduziert und nach Phosphorylierung zum Triphosphat im Zuge der DNA-Replikation ins Genom eingebaut.^[292] Da die löslichen Desaminasen, Cda und Dctd, keine Ribo-Verbindung umsetzen, wird so der Verlust an Markierung durch Desaminierung zum dU-Derivat deutlich reduziert.^[293, 294] So konnte durch Zugabe der Ribo-Verbindung eine Markierungseffizienz von etwa 40 % aller dC erzielt werden. Nach Fütterung der Ribo-Verbindung für fünf Tage unter *Priming*-Bedingungen, wurde die genomische DNA isoliert, derivatisiert und massenspektrometrisch untersucht. Die markierten abasischen Stellen konnten nicht quantifiziert werden, lagen jedoch über dem *Limit of Detection* (LOD). Die quantifizierten, markierten β -Eliminierungsprodukte sind in Abbildung 41D gezeigt. Trotz Normierung der β -Eliminierungsprodukte auf 100 % Markierungseffizienz, war auch hier eine große Varianz zwischen den biologischen Replikaten erkennbar.

In einem biologischen Replikat kam es in den *Knockouts* zu einer Zunahme an β -Eliminierungsprodukten. Zusammengefasst geben die Daten keinen Hinweis auf eine Beteiligung der Neil-Glykosylasen an der aktiven Demethylierung. Die erhaltenen Messwerte bewegen sich in einem Bereich, indem eine geringe Datenvarianz zu relativ großen Abweichungen führen kann. Ein Beitrag an der aktiven Demethylierung kann auf diese Weise nicht bestätigt werden. Falls die Neil-Enzyme als bifunktionale Glykosylasen tatsächlich die von Tdg hinterlassene AP-Stelle weiter prozessieren, wäre die Anzahl an markierten β -Eliminierungsprodukten im Vergleich zum Wildtyp geringer. Ferner könnte Ape1 in Abwesenheit der Neil-Glykosylasen diese Aufgabe übernehmen, was durch ein *Knockdown*-Experiment untersucht werden könnte.

Quantifizierung oxidierter dC-Derivate in *Tdg*^{-/-} mESCs und der katalytischen Mutante N151A

Um die Rolle von Tdg an der aktiven Demethylierung eingehender zu untersuchen, sollte eine *Tdg*^{-/-} mESC Linie und eine entsprechende katalytische Mutante hergestellt werden. Die katalytische Mutante bindet dabei fdC bzw. cadC, vermag diese aber nicht mehr aus der genomischen DNA zu entfernen.^[191, 295] Die Generierung der Zelllinien wird in Abschnitt 3.2.3.5 beschrieben. Als nächstes wurden die natürlichen höher oxidierten Cytidin-Modifikationen in den generierten Zelllinien massenspektrometrisch analysiert. Hierzu wurden die Zellen vier Tage unter Serum/LIF-Bedingungen kultiviert. Nach Isolierung der genomischen DNA wurde diese enzymatisch verdaut und die erhaltenen Nucleoside mittels UHPLC-MS/MS analysiert. Die Ergebnisse der absoluten Quantifizierung sind in Abbildung 42 dargestellt.

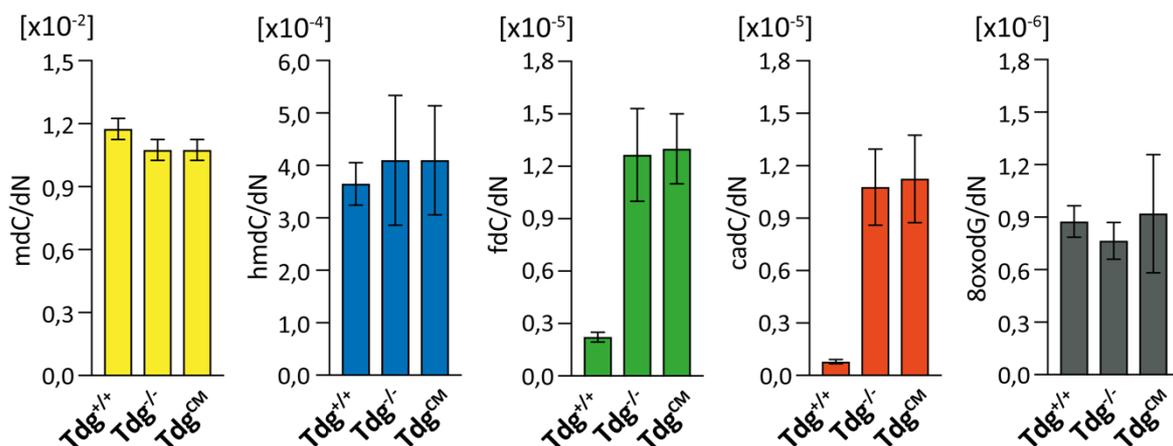


Abbildung 42: Level der natürlichen höher oxidierten Cytidin-Derivate in *Tdg*-*Knockout*, *Tdg*-katalytischer Mutante und dem Wildtyp. Abgebildet sind die normierten Quantifizierungsergebnisse aus vier biologischen Replikaten, gemessen als technische Triplikate.

Die Quantifizierungsdaten zeigen vergleichbare Mengen an mdC und hmdC in allen drei Zelllinien. Allerdings sind die mdC-Werte im *Tdg*^{-/-} und der katalytischen Mutante etwas niedriger, was auf eine Beeinflussung der Remethylierung in diesen Zellen hindeutet.

Die *Tdg*^{-/-} mESCs und die katalytische Mutante unterscheiden sich deutlich vom Wildtyp anhand der fdC- und cadC-Werte. Im Falle des *Knockouts* und der Mutante kommt es zu einem etwa sechsfachen Anstieg der fdC Werte und für cadC ist der Anstieg etwa 13-fach. Dieser Phänotyp bestätigt zusätzlich die erfolgreiche Generierung der *Tdg*^{-/-} mESCs und einer katalytischen Mutante die, aufgrund eines Aminosäureaustauschs (N151A), nicht mehr vermag Tdg auszuschneiden. Der bedeutend höhere Anstieg der cadC-Werte könnte auf eine Anhäufung der Modifikation hindeuten, da diese sich am Ende des postulierten DNA-Demethylierungsweges befindet. Interessant wäre hier, auch die Menge der BER-Intermediate nach Markierung der DNA mit isopenmarkiertem Cytidin zu überprüfen. Unter Zuhilfenahme der katalytischen Mutante könnte zudem die Protein-Protein Interaktion mit den Neil-Glykosylasen untersucht werden.

3.2.3.3 Ausblick

In diesem Kapitel wurde die Generierung einer *Smug1*^{-/-} mESC-Linie beschrieben. Diese könnte für weitere metabolische Untersuchungen genutzt werden. Isotopenmarkiertes hmdU könnte verwendet werden, um die von hmdU stammenden abasischen Stellen oder β -Eliminierungsprodukte eindeutiger quantifizieren zu können. Weiterhin könnten Letalitätsstudien unternommen werden. Die *Smug1*^{-/-} Zelllinie weist zwar eine höhere Toleranz gegenüber hmdU auf, zeigt aber ab einer bestimmten Inkubationsdauer oder Konzentration eine deutliche Letalität. Interessant ist hierbei, dass *Smug1* nicht die einzige Glykosylase ist, die hmdU auszuschneiden vermag. In dessen Abwesenheit könnten Glykosylasen, wie *Mbd4* oder *Tdg* diese Aufgabe übernehmen. Mit Hilfe der *Smug1*^{-/-} Zelllinie könnte weiterhin eine Sequenzierung von hmdU vorgenommen werden. Derartige Sequenzierungen waren bisher nur in *Leishmania major* möglich, wo hmdU ein natürlicher Bestandteil des Genoms ist.^[143, 144] Dies könnte nun auch in embryonalen Stammzellen wiederholt werden und Hinweise auf eine Beteiligung von hmdU an der DNA-Demethylierung geben. Im Vergleich mit den Wildtyp mESCs könnte der *Smug1-Knockout* genutzt werden um hmdU Positionen die durch Tet-Oxidation, ausgehend von dT, entstanden sind zu identifizieren. So könnte durch Sequenzierung eine Lokalisation von hmdU regulativen Elementen der DNA nachgewiesen werden.^[191]

Die generierte *Tdg*^{-/-} bzw. *Tdg*^{CM} Zelllinie könnte genutzt werden, um die Entstehung der fC-Base im Rahmen der BER zu detektieren. Der lösliche Nukleosid/-tid Pool dieser Zellen könnte extrahiert werden und eine Derivatisierung mit Hilfe des in Abschnitt 3.2.1 beschriebenen Reagenzes zur Quantifizierung der BER-Intermediate vorgenommen werden. Die quantifizierte Menge an ausgeschnittenem fdC könnte Hinweise auf die Reparaturaktivität der Glykosylase im Zuge der Differenzierung der Stammzellen geben oder eine Beteiligung bifunktionaler Glykosylasen bestätigen.

3.2.3.4 Projektbeiträge

Die Zellkulturexperimente wurden von mir entwickelt und durchgeführt. Auch die Isolation aller DNA Proben wurde von mir vorgenommen. Die Derivatisierung wurde von mir in Zusammenarbeit mit *René Rahimoff* und *Sarah Schiffers* unternommen. Der enzymatische Verdau, die Messung und Auswertung der Daten des *Smug1^{-/-}*, sowie der *Neil Knockout*-Linien wurde von *Olesea Kosmatchev* und *Sarah Schiffers* vorgenommen. Der enzymatische Verdau der Tdg-Zelllinien wurde von *Kerstin Kurz* durchgeführt und gemessen und von mir ausgewertet. Alle Daten wurden von *Olesea Kosmatchev*, *René Rahimoff*, *Sarah Schiffers* und mir gemeinsam interpretiert. *Eva Korytiakovà* und *Sarah Schiffers* führen das Projekt fort.

3.2.3.5 Appendix: Generierung und Verifizierung der *Knockout*-Linien

Smug1 Knockout mES-Zelllinie

Die *Knockout* mES-Zelllinie der Glykosylase Smug1 wurde im Rahmen meiner Masterarbeit unter Verwendung des CRISPR-Cas9 Systems generiert. Mit Hilfe des Cas9-*guide* RNA (gRNA) Komplexes wurde ein Doppelstrangbruch im Exon1 des für Smug1 kodierenden DNA-Abschnittes eingeführt (siehe Abbildung 43A). Dadurch wird das zelleigene Reparatursystem aktiviert und zufällige Insertionen oder Deletionen von Nucleobasen an der Stelle des Strangbruchs eingeführt. Durch diese Mutationen kommt es zu einer Verschiebung des Leserasters und ein funktionales Protein kann nicht mehr erzeugt werden. Abbildung 43B zeigt eine schematische Darstellung des Smug1 Proteins mit katalytischen Domänen und der Position der Mutation. Das katalytische Zentrum des Proteins wird *downstream* der Mutation kodiert und sollte nach Verschiebung des Leserasters defekt sein. Durch *Limiting Dilution* wurde eine klonale Zelllinie gewonnen und der modifizierte DNA-Abschnitt sequenziert. Die Änderung der DNA-Sequenz und die daraus resultierende Aminosäureabfolge sind in Abbildung 43C dargestellt. Dies sollte ein verändertes Protein mit etwa 19 kDa ergeben. Zur vollständigen Verifizierung des *Knockouts* wurde ein Western Blot der *Knockout*- und der dazugehörigen Wildtyp-Linie durchgeführt. Dieser ist in Abbildung 43D dargestellt und liefert aufgrund der zahlreichen kreuzreaktiven Banden kein eindeutiges Ergebnis. Gemeinsam geben die Sequenzierergebnisse, die UHPLC-MS/MS Messungen (3.2.3.2) und der Western Blot Hinweise auf die Generierung des Smug1 *Knockouts*.

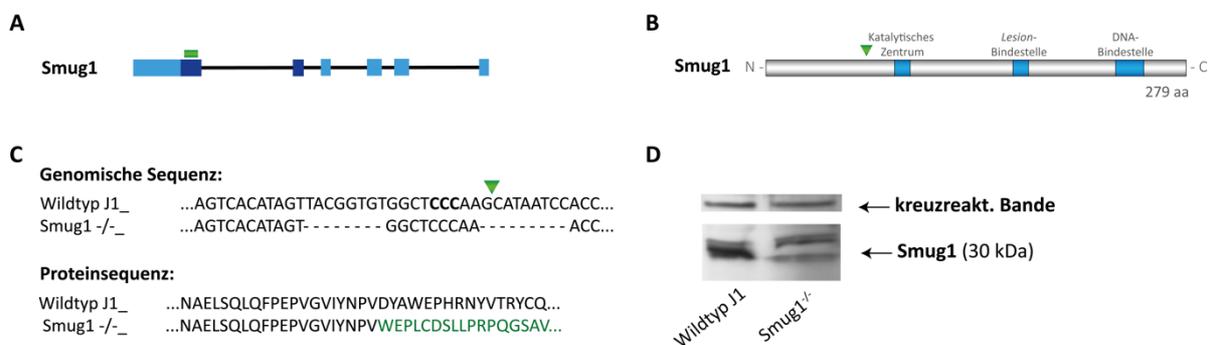


Abbildung 43: Verifizierung des Smug1-Knockouts. A) Schematische Darstellung des murinen Smug1 Genabschnitts mit kodierenden (dunkelblau) und nichtkodierenden (hellblau) Exons, sowie der Bindestelle der gRNA (grün). B) Schematischer Aufbau des Smug1 Proteins mit den wichtigsten Domänen. Das grüne Dreieck illustriert die Position der erzeugten Mutation. C) Sequenzierungsergebnisse des modifizierten Genabschnitts mit resultierender Proteinsequenz und der Alterierung (grün). D) Western Blot zum Nachweis des *Knockouts*.

Tdg Knockout und katalytische Mutante mES-Zelllinien

Zur weiteren Aufklärung der Beteiligung der BER an der aktiven Demethylierung, wurde eine *Knockout* mES-Zelllinie der Glykosylase Tdg unter Zuhilfenahme des CRISPR-Cas9 Systems hergestellt. Hierzu wurde endogen eine Reparaturvorlage eingeführt und mit Hilfe dieser das Gen ausgeschaltet. Eine ähnliche Strategie wurde zur Generierung der katalytisch inaktiven Mutante (*Tdg^{CM}*, N151A) genutzt. Die N151A Mutante wurde bereits in Embryonen von Mäusen, murinen Fibroblasten und *in vitro* in einem humanen Zellsystem (N140A) erzeugt.^[191, 295] Hier konnte gezeigt werden, dass eine Mutation von Asparagin zu Alanin zwar noch zu einer Bindung von fdC führt, aber dieses nicht mehr ausgeschnitten werden kann.^[295, 296] Zunächst wurden, unter Zuhilfenahme des *CRISPR-Design Tools* (F. Zhang, MIT), die gRNA-Sequenzen zur Rekrutierung der Endonuklease an die Zielsequenz ermittelt und in das Expressionsplasmid kloniert. Eine detaillierte Beschreibung des Arbeitsablaufs ist in meiner Masterarbeit aufgeführt.^[297] Zur Herstellung der *Tdg^{-/-}* mESC-Linie wurden zwei Plasmide mit den jeweiligen gRNAs, Cas9 und der einzelsträngigen Reparaturvorlage in Wildtyp mESCs transfiziert. Durch Zugabe eines NHEJ-Inhibitors wurde die Einbringung der Reparaturvorlage durch die HDR favorisiert. Abbildung 44A zeigt die Exon-Positionen, an denen Cas9 einen Strangbruch, geführt durch die gRNA, einführen sollte. Für den *Knockout* wurde Exon 2 gewählt, um auch das alternative Startkodon der natürlich vorkommenden kürzeren Isoform zu adressieren. Die Editierung wurde deshalb für eine Position nach dem zweiten möglichen ATG entworfen. Das katalytische Zentrum liegt im Exon 4, weswegen für die katalytische Mutante eine Reparaturvorlage für diesen Genabschnitt entwickelt wurde. Weiterhin sollte die Reparaturvorlage ein Signal für eine erfolgreiche Editierung in die genomische DNA einführen. Für die katalytische Mutante konnte eine *SmaI*-Restriktionsschnittstelle durch die Änderung der Basenabfolge entfernt werden. Im Fall der *Tdg^{-/-}* Zelllinie wurde, durch Einführung eines Stoppcodons, eine *BamHI*-Schnittstelle eingeführt. Weiterhin wurde die PAM-Sequenz mutiert, sodass Cas9 das Reparaturtemplat nicht erneut schneiden kann und multiple Insertionen der Reparaturvorlage vermieden werden. Nach Transfektion der Zellen mit den Plasmiden, sowie der Reparaturvorlage wurde eine *Limiting Dilution* zur Isolierung einer klonalen Zelllinie unternommen. Die gewonnenen Zelllinien wurden anschließend genotypisiert. Hierzu wurde der Ziellokus mittels PCR amplifiziert und mit dem jeweiligen Restriktionsenzym verdaut. Durch Gelelektrophorese wurde das Bandenmuster im Anschluss untersucht. Die Ergebnisse der positiven Klone und der Wildtyp-Kontrolle sind in Abbildung 44B dargestellt. Der Restriktionsverdau zur Identifizierung des eingebrachten Reparaturtemplats für die *Knockout*-Linie ergibt zwei Banden für eine Einbringung, da die Vorlage die Schnittstelle für das Restriktionsenzym enthält. Die unverdaute Kontrolle, sowie nicht-editierte Klone zeigen nur ein Fragment, das PCR-Produkt. Genau umgekehrt ist es für den Fall der katalytischen Mutante (*Tdg^{CM}*). Hier wurde die Restriktionsschnittstelle durch Einbringung der Reparaturvorlage entfernt. In der verdauten

katalytischen Mutante ist demnach keine Aktivität des Enzyms zu sehen. Im unbehandelten Wildtyp hingegen schon. Auf diese Weise wurden im Zuge des Screenings positive Klone ausgewählt. Zur eindeutigen Verifizierung wurden die Klone im nächsten Schritt sequenziert. Die Ergebnisse sind in Abbildung 44C zusammengefasst und bestätigen die Resultate des Restriktionsverbaus. Weiterhin ist die daraus resultierende Proteinsequenz mit den eingeführten Mutationen gezeigt.

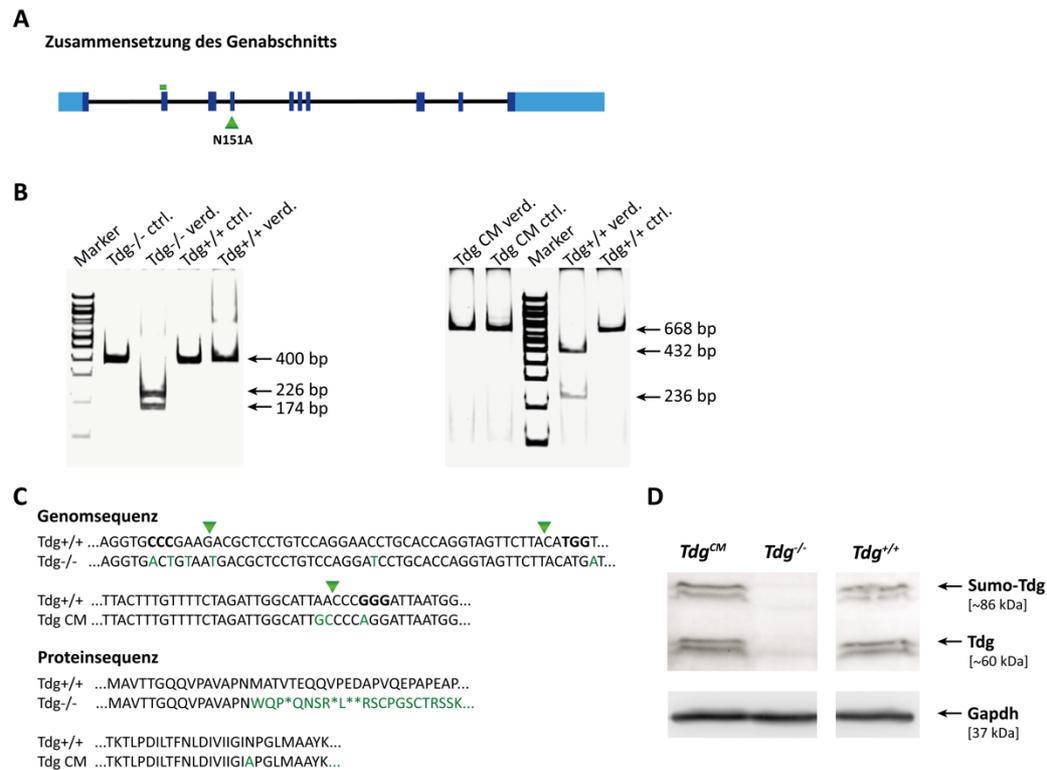


Abbildung 44: Strategie und Verifizierung der Tdg Knockout und katalytische Mutante mESC-Linie.

A) Schematische Darstellung des murinen Tdg Genabschnitts mit kodierenden (dunkelblau) und nichtkodierenden (hellblau) Exons, sowie der Bindestelle der *guide* RNAs (grün). **B)** Ergebnisse der Restriktionsverbaus zur Genotypisierung der Zelllinie. **C)** Sequenzierungsergebnisse des modifizierten Genabschnitts mit resultierender Proteinsequenz und der Alterierung (grün, Stoppcodon *). **D)** Western Blot zum Nachweis des *Knockouts*.^[298]

Im letzten Schritt wurde die Expression des Proteins in den generierten Zelllinien überprüft. Hierzu wurde ein Western Blot mit einem Tdg-spezifischen Antikörper durchgeführt (Abbildung 44D). Deutlich zu sehen ist, dass die *Tdg*^{-/-} Zelllinie kein Tdg Protein enthält. Dahingegen sind in Wildtyp und *Tdg*^{CM} mESCs die Protein-Banden in gleicher Intensität vorhanden. Die Doppelbande resultiert entweder aus der Nutzung des alternativen Startkodons der zweiten Tdg-Isoform oder ist ein verkürztes Degradationsprodukt des Proteins durch die Zellyse.^[205, 298] Die höhermolekulare Doppelbande bei etwa 86 kDa ist vermutlich auf eine „sumoylierte“ Form von Tdg zurückzuführen.^[299] Als Ladekontrolle wurde das Protein Gapdh detektiert, welches ebenfalls die gleiche Intensität in allen mESCs aufweist.

3.3 Studien einer oxidationsunabhängigen DNA-Demethylierung

Neben einer oxidationsabhängigen Demethylierung, katalysiert durch die Tet-Dioxygenasen, wurde eine enzymatische Desaminierung von mdC als alternativer DNA-Demethylierungsweg vorgeschlagen. In Folge einer Desaminierung könnte mdC direkt im Genom oder in Folge der DNA-Reparatur im Nukleosid/-tid Pool durch lösliche Desaminasen zu dT umgewandelt werden. Ob und in welchem Maße eine derartige Reaktion in lebenden Zellen stattfindet ist Gegenstand nachfolgender Studie.

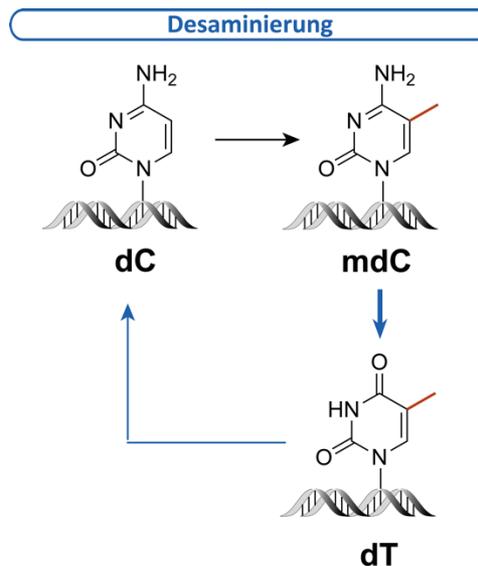


Abbildung 45: Oxidationsunabhängige DNA-Demethylierung durch Desaminierung von mdC zu dT.

3.3.1 Oxidationsunabhängige Demethylierung durch die DNA-Reparatur

3.3.1.1 Prolog

Ob tatsächlich eine Desaminierung von mdC zu dT stattfindet, sollte durch Isotopenverfolgungsstudien nachgewiesen werden. Bei dieser Untersuchung wird $^{13}\text{CD}_3$ -L-Methionin zum Kulturmedium hinzugegeben und der Metabolismus der Isotopenmarkierung verfolgt (siehe Abbildung 46). Katalysiert durch die DNA-Methyltransferasen überträgt der isotope markierte Kofaktor SAM (S-Adenosyl-L-Methionin) eine $^{13}\text{CD}_3$ -Methylgruppe auf die C5-Position von dC. Das entstandene $^{13}\text{CD}_3$ -mdC ($m^{+4}\text{dC}$) ist somit um vier Masseneinheiten schwerer als das natürliche Isotopolog. Die $^{13}\text{CD}_3$ -Methylgruppe wird im Zuge der aktiven Demethylierung durch die Tet-Dioxygenasen oxidiert, wobei die höher oxidierten Cytidin-Spezies ebenfalls eine Isotopenmarkierung tragen.

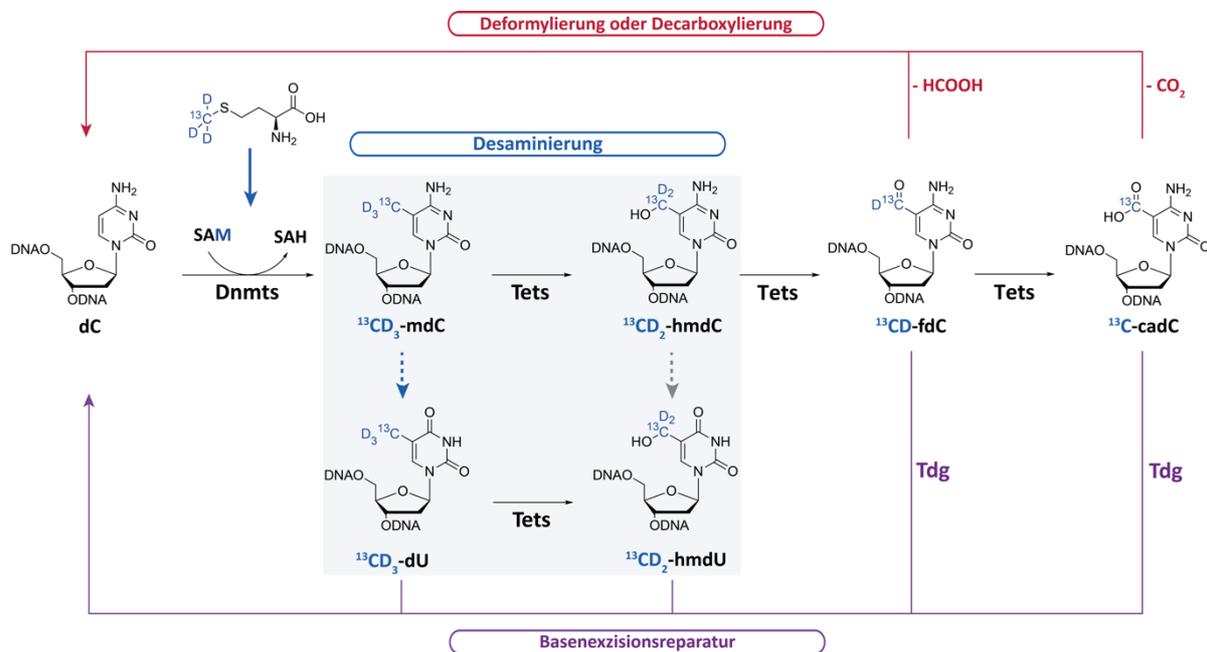


Abbildung 46: Untersuchung der Entfernung von mdC aus dem Genom durch metabolische Markierung mit $^{13}\text{CD}_3$ -L-Methionin.

Zur Detektion der isotope markierten Metaboliten wurde eine massenspektrometrische Methode zur absoluten Quantifizierung von Dr. Jessica Steinbacher und Dr. Olesya Kosmatchev entwickelt.^[249, 300] Hierdurch konnte die Existenz des Desaminierungsprodukts dT bzw. 5-Methyl-2'-desoxyuridin (mdU) im Säugetiergenom nachgewiesen werden.^[142, 249, 300] Weitere massenspektrometrische Untersuchungen zeigten nach metabolischer Markierung einer Dnmt TKO mESC-Linie und dem korrespondierenden Wildtyp, dass $^{13}\text{CD}_3$ -dT bzw. $^{13}\text{CD}_3$ -mdU ($m^{+4}\text{dU}$) ausschließlich in Wildtyp mESCs auftritt. Die erhaltenen Quantifizierungsergebnisse sind in Abbildung 47 gezeigt.

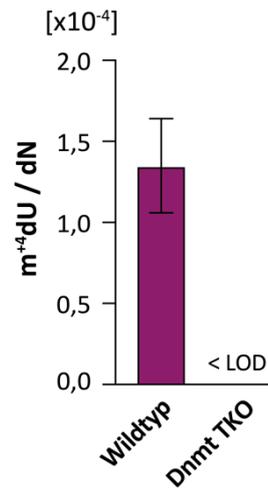


Abbildung 47: Absolute Quantifizierung von $m^{+4}dU$ in Dnmt TKO und Wildtyp mESCs nach metabolischer Markierung mit $^{13}CD_3$ -L-Methionin. Die Quantifizierungsergebnisse aus drei biologischen Replikaten, gemessen als technisches Triplikat sind dargestellt.^[249, 300]

Da die Methylgruppe des natürlichen dT biosynthetisch aus Serin im Zusammenspiel mit 5,10-Methylentetrahydrofolat und der Thymidylatsynthase übertragen wird, kann isotopenmarkiertes $m^{+4}dU$ nur aus $m^{+4}dC$ entstehen. Die Verbindung kann demnach nur gebildet werden, wenn die DNA-Methyltransferasen in der DNA $^{13}CD_3$ -m dC generiert haben. Weiterhin wurde die Bildung von $m^{+4}dU$ auch in Abwesenheit der Tet-Enzyme beobachtet und lässt auf einen oxidationsunabhängigen Vorgang schließen.^[249, 300] Die Existenz von genomischem $m^{+4}dU$ könnte erste Hinweise auf einen alternativen Demethylierungsweg geben und könnte auf zwei Arten entstehen: 1) Desaminierung von $m^{+4}dC$ auf DNA-Ebene; 2) Entfernung von $m^{+4}dC$ über die DNA-Reparatur und einer Desaminierung von löslichem $m^{+4}dC$ im Nukleosid/-tid Pool.

Das nachfolgende Kapitel beschäftigt sich eingehender mit diesem putativen Demethylierungsweg. Unter verschiedenen Kulturbedingungen wird die Entstehung und Entfernung von $m^{+4}dC$ in mESCs näher betrachtet. Weiterhin wird die Beteiligung verschiedener DNA-Reparaturwege an der aktiven Demethylierung und Bildung von $m^{+4}dU$ untersucht. Hierzu werden *Knockout*-Linien der Schlüsselenzyme verschiedener Reparaturwege in mESCs hergestellt.

3.3.1.2 Ergebnisse und Diskussion

In der frühen Embryonalentwicklung ändern sich die Level an genomischem mdC beim Übergang vom naïve Präimplantations- zum *primed* Postimplantationsepiplast. Zur Untersuchung dieser Zustände werden embryonale Stammzellen als *in vitro* Modell unter verschiedenen Kulturbedingungen betrachtet. Eine homogene, undifferenzierte Stammzellkultur kann durch Kultivierung der Zellen unter 2i/LIF- oder *alternative* 2i/LIF-Bedingungen (a2i) erhalten werden.^[34, 116] Im Gegensatz zu naïven 2i/LIF-Bedingungen modelliert a2i/LIF ein weiter fortgeschrittenes Stadium der embryonalen Entwicklung.^[301] Dies spiegelt sich auch bei der Betrachtung der natürlichen mdC-Level wider (siehe Abbildung 48A und B). Unter 2i/LIF werden niedrigere Mengen an mdC im Genom detektiert als unter a2i/LIF-Kulturbedingungen. Dieser Unterschied wird in höheren Passagen besonders deutlich.^[59, 78]

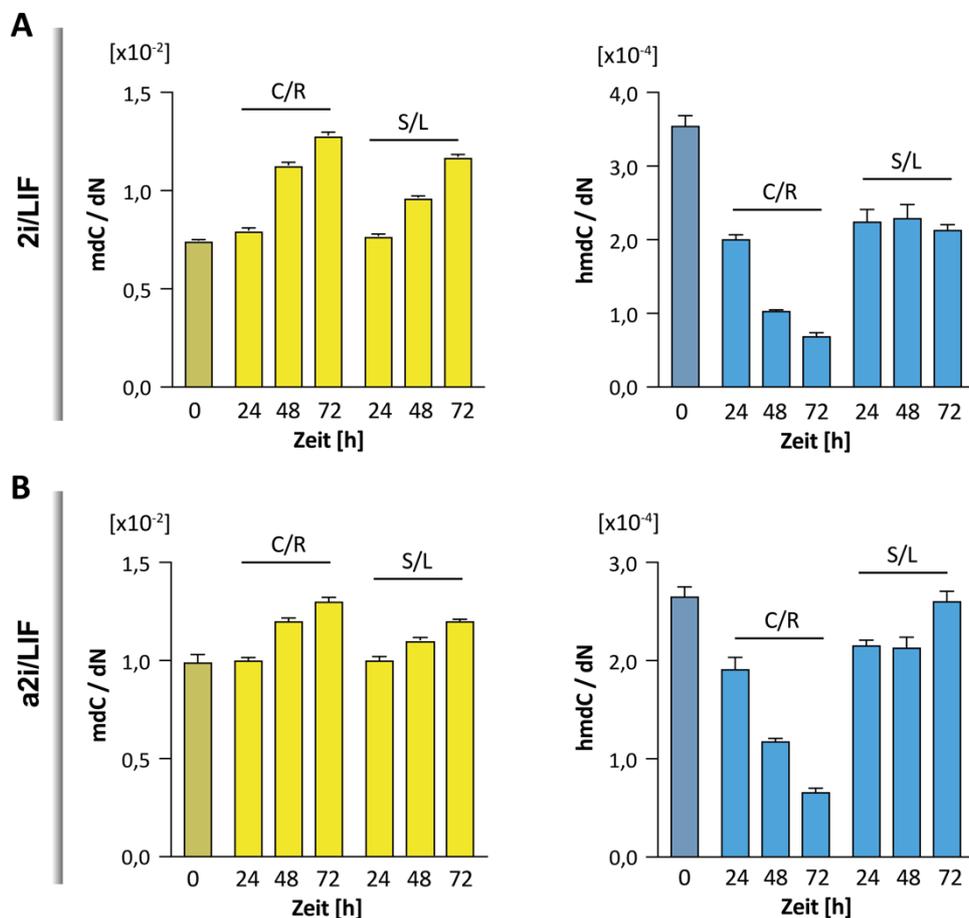


Abbildung 48: Natürliche Cytidin-Modifikationen unter verschiedenen Kulturbedingungen. A) Absolute Quantifizierung der natürlichen mdC und hmdC Level nach *Priming* aus einer 2i/LIF Ursprungskultur. **B)** Quantifizierungsergebnisse nach *Priming* aus einer a2i/LIF Ursprungskultur. Dargestellt sind die Ergebnisse einer biologischen Probe (N=1), gemessen als technisches Triplikat. 2i/L: CHIR/PD/LIF-; a2i/LIF: CHIR/CGP/LIF-; C/R: CHIR/IWR1-Kulturbedingung; S/L: Serum/LIF-Kulturbedingung

Stammzellen, die kontinuierlich unter Serum/LIF kultiviert werden, enthalten globale Methylierungswerte von etwa 4 % aller Cytidine.^[78] Die CpG-Dinukleotide sind durchschnittlich zu 70-80 % methyliert.^[302] Unter naiven 2i/LIF-Bedingungen wird üblicherweise ein sehr niedriger Methylierungsgrad von etwa 1 % aller Cytidine erreicht und ist vergleichbar mit Zellen der ICM.^[78] Im beschriebenen Experiment konnte jedoch unter 2i/LIF eine Methylierung etwa $0,75 \times 10^{-2}$ bezogen auf alle Nukleoside (dN) beobachtet werden, was einer prozentualen Menge von 3.6 % aller dC entspricht. Dieser Wert liegt über dem erwarteten Methylierungsgrad naiver Zellen und ist vermutlich darauf zurückzuführen, dass die Zellen nicht lange genug unter 2i/LIF kultiviert wurden. Unter a2i/LIF liegt ein Methylierungsgrad von etwa $1,0 \times 10^{-2}$ /dN (4,7 % aller dC) vor und spiegelt einen weiter fortgeschrittenen Punkt in der Embryonalentwicklung wider.^[59, 303] Für 5-Hydroxymethyl-2'-desoxycytidin (hmdC), dem ersten Tet-Oxidationsprodukt, werden Level von etwa 0,2 % aller dC beobachtet. Unter 2i/LIF-Bedingungen sind die hmdC-Werte jedoch etwas höher als die der a2i/LIF-Bedingung. Um die zeitliche Änderung von mdC und der Tet-Oxidationsprodukte während der Entwicklung zum Postimplantationsepiblast zu beobachten, wurden die Zellen weiter unter *Priming*-Bedingungen kultiviert. Hierzu wurden die mESCs aus der naiven 2i/LIF-Bedingung (Abbildung 48A) oder dem a2i/LIF-Stadium (Abbildung 48B) unter Serum/LIF (S/L) bzw. CHIR/IWR1 (C/R) für drei Tage kultiviert und massenspektrometrisch untersucht.^[63] Die C/R-Bedingungen können als Methode eines synchroneren *Primings* der ESCs im Vergleich zu S/L beschrieben werden.^[63] Im Gegensatz zu C/R fluktuieren die mESCs unter Serum/LIF-Bedingungen zwischen verschiedenen Entwicklungsstadien. Die Zellkultur ist eher heterogen.^[304] Die C/R-Bedingung führt aus beiden Kulturmethoden heraus zu einem stärkeren Anstieg der mdC-Level. Aus beiden Bedingungen heraus wird nach 72 h ein Wert von $1,3 \times 10^{-2}$ pro dN erreicht, was einer Methylierung von 6,1 % aller Cytidine entspricht. Demnach werden neben den CpG- auch CpH-Stellen (H=A, C oder T) methyliert. Die Methylierung der CpH-Positionen wurde in diesem Entwicklungsstadium bereits beschrieben.^[305, 306] Interessant ist hier auch die Bildung des Tet-Oxidationsprodukts hmdC. Unter C/R-Bedingungen nehmen die hmdC-Werte in beiden Fällen stark ab, während sie unter S/L-Bedingungen zunächst leicht abfallen, aber vorwiegend stabil bleiben. *Habibi et al.* beschrieben bereits einen Rückgang der hmdC-Level beim Vergleich von ESC und *primed* EpiSCs.^[116] Die Werte an fdC sowie cadC folgen demselben Trend, konnten aber aufgrund der niedrigen Werte nicht verlässlich quantifiziert werden. In weiteren Experimenten wurde beobachtet, dass die hmdC-Werte nach Passagieren der Zellen unter S/L-Bedingungen zunehmen.^[170] Dieser Trend wurde auch für die weiteren natürlichen dC-Modifikationen fdC und cadC beobachtet. Dieses Experiment zeigt, dass die verwendeten Kulturbedingungen unterschiedliche Auswirkungen auf die Level von mdC und den Tet-Oxidationsprodukten haben können. Die verschiedenen Kulturbedingungen scheinen ein gutes Modellsystem zu sein die Vorgänge der frühen Embryonalentwicklung vom Präimplantations- zum Postimplantationsepiblast zu rekapitulieren.

Mittels metabolischer Markierung durch [$^{13}\text{C}_3$]-L-Methionin wurde als nächstes der Umsatz von genomischem mdC in mESCs verfolgt. Das markierte Methionin wird dem Kulturmedium für fünf Tage unter naiven a2i-Bedingungen zugesetzt. Zur Untersuchung des *primed*-Zustandes wurden die Zellen zunächst für zwei Tage in a2i-Bedingungen vormarkiert und dann für drei Tage unter C/R-Bedingungen gefüttert. Diese Behandlung stellt eine maximale Markierung des mdC sicher. An Tag 4 wird zusätzlich $^{13}\text{C}_{10}^{15}\text{N}_2$ -dT (dT $^{+12}$) für 24 h als Proliferationsmarker zugegeben. Dies ergibt gewöhnlich eine Markierungseffizienz von etwa 85 % aller dT. Anschließend wird das Kulturmedium gegen markierungsfreies Medium ausgetauscht (*Chase*) und die Zellen im zeitlichen Abstand von 12 h geerntet. Der Ablauf des Experiments ist in Abbildung 49A illustriert.

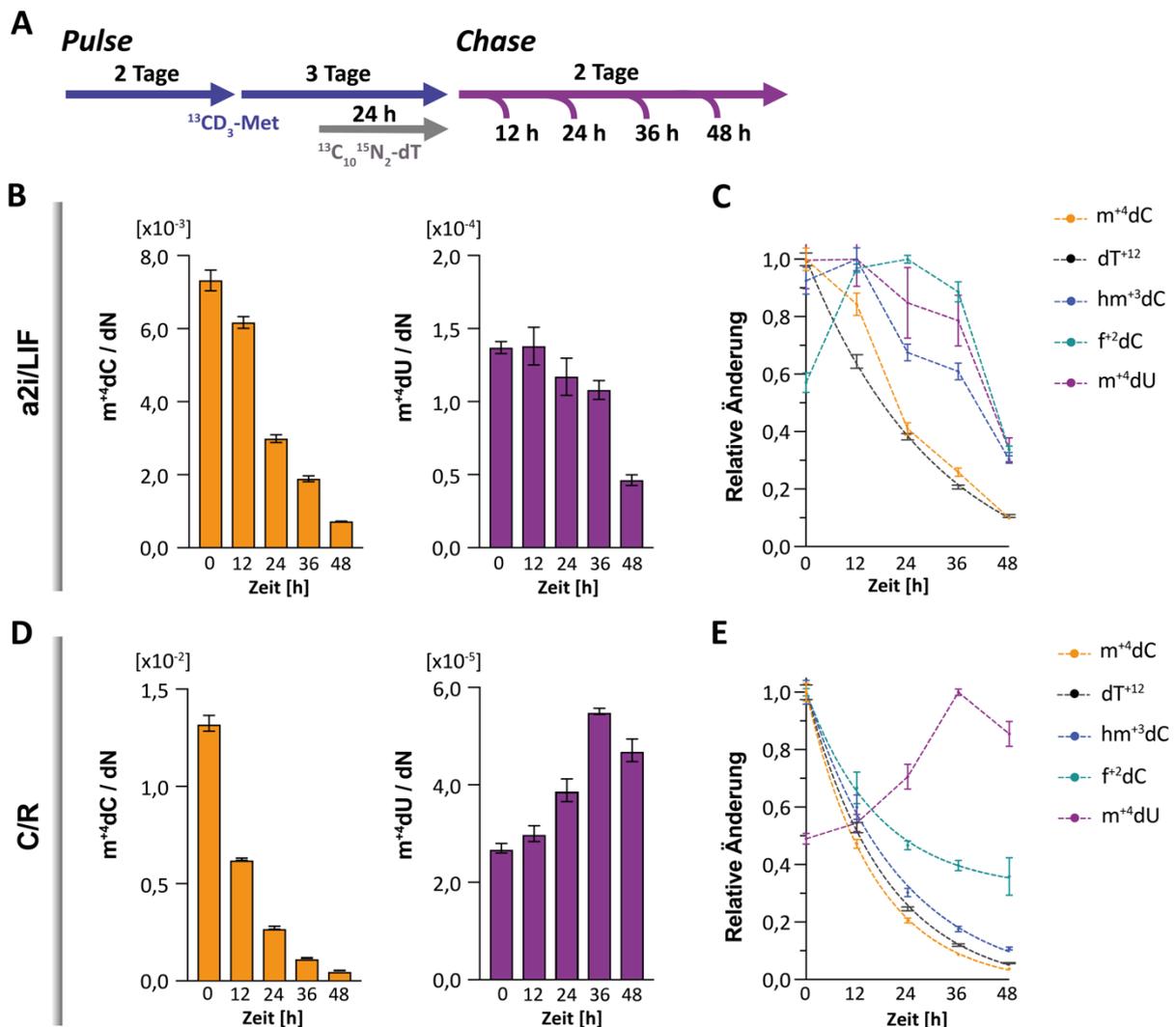


Abbildung 49: Pulse-Chase Experiment in Wildtyp mESCs. **A)** Schematische Darstellung des experimentellen Ablaufs. $^{13}\text{CD}_3$ -Met: $^{13}\text{CD}_3$ -L-Methionin **B)** Pulse-Chase Experiment unter a2i-Kulturbedingungen. **C)** Relative Änderung der markierten mdC, hmdC, fdC, mdU und dT-Spezies unter a2i-Bedingungen nach Chase. **D)** Pulse-Chase Experiment unter C/R-Kulturbedingungen. **E)** Relative Änderung der markierten mdC, hmdC, fdC, mdU und dT-Spezies unter C/R-Bedingungen nach dem Chase. Dargestellt sind die Ergebnisse einer biologischen Probe gemessen als technisches Triplikat.

Nach Isolierung der genomischen DNA und enzymatischem Verdau auf Nukleosidebene wurde die Menge an isotope markierten Nukleosiden massenspektrometrisch analysiert. In Abbildung 49B sind die Werte an $m^{+4}dC$ und dem Desaminierungsprodukt $m^{+4}dU$ bezogen auf alle Nukleoside unter a2i/LIF-Bedingungen dargestellt. Beide Werte nehmen im Laufe des *Chase*-Experiments ab, wobei $m^{+4}dC$ deutlich schneller abnimmt als $m^{+4}dU$. Aufgrund der zusätzlichen Markierung mit dT^{+12} kann im *Chase* auf die passive Verdünnung durch die DNA-Synthese geschlossen werden. Entsprechend wurde die relative Änderung der isotope markierten dC-Modifikationen, $m^{+4}dU$ und dT^{+12} in Abbildung 49C dargestellt. Die Produkte zeigen, dass mdC und die Tet-Oxidationsprodukte unter a2i/LIF-Bedingungen langsamer aus dem Genom entfernt werden als dT^{+12} durch passive Verdünnung. Unter C/R-Bedingungen (siehe Abbildung 49D) verläuft die Abnahme an $m^{+4}dC$ im Laufe des *Chase* rascher als unter a2i-Bedingungen. Im Gegensatz dazu nimmt das Desaminierungsprodukt $m^{+4}dU$ auf alle Nukleoside bezogen bis zum 36 h Zeitpunkt zu. Dieser Trend ist auch in Abbildung 49E zu erkennen. Hier wird die relative Änderung der markierten dC Spezies und dT^{+12} während des *Chase*-Zeitraums dargestellt. Die Abbildung zeigt, dass die $m^{+4}dC$ Werte schneller abnehmen als dT^{+12} durch passive Verdünnung. Dies lässt den Schluss zu, dass während des *Primings* $m^{+4}dC$ aktiv aus dem Genom entfernt wird. Die höher oxidierten dC Verbindungen hmdC und fdC nehmen langsamer ab als dT^{+12} und sind vermutlich entweder stabile Genommodifikationen oder werden hier durch Tet-Oxidation aktiv aus mdC gebildet.^[157, 307] Der Beitrag der Tet-Oxidation könnte durch eine Wiederholung des Experiments in einer *Tet-Triple Knockout* mESC-Linie näher beleuchtet werden. Interessanterweise nimmt das isotope markierte $m^{+4}dU$ unter *Priming*-Bedingungen im zeitlichen Verlauf erst nach 48 h ab. Da $m^{+4}dC$ als einzige dC-Modifikation schneller abnimmt als die Tet-Oxidationsprodukte, liegt die Annahme nahe, dass $m^{+4}dC$ über $m^{+4}dU$ während des *Primings* aktiv entfernt wird.

Der größte Umsatz von mdC zu mdU geschieht im löslichen Nukleosid/-tid Pool

In den nachfolgenden Experimenten wird untersucht, ob das detektierte $m^{+4}dU$ durch eine Desaminierung direkt im Genom entsteht oder nach Entfernung von $m^{+4}dC$ aus dem Genom im Zuge der DNA-Reparatur durch die Desaminasen des löslichen Nukleosid/-tid Pool erzeugt wird. Eine direkte Desaminierung würde zur Ausbildung von dT:dG Fehlpaarungen führen. Aufgrund ihrer Mutagenität werden diese Fehlpaarungen schnell prozessiert, was allerdings mit der Gefahr von Strangbrüchen einhergeht. Aufgrund der beachtlichen Level an genomischem $m^{+4}dU$ können Fehlpaarungen als alleinige Quelle ausgeschlossen werden. Konkurrenzexperimente mit unmarkiertem dT zeigten zunächst einen Rückgang von $m^{+4}dU$ im Genom.^[249, 300] Der größte Teil an inkorporiertem $m^{+4}dU$ entsteht vermutlich im löslichen Nukleosid/-tid Pool, wobei der Einbau durch dT-Kompetition unterdrückt werden kann.^[249, 300] Die Cytidin-Desaminase (Cda) und die Deoxycytidylat-Desaminase (Dctd) sind die beiden bekannten Desaminasen des löslichen Pools.

Im Zuge meiner Masterarbeit wurden *Knockout*-Stammzelllinien mit Hilfe des CRISPR-Cas9 Systems hergestellt, in welchen die beiden Desaminasen Cda und Dctd nicht vorhanden sind. Die Verifizierung erfolgte im Rahmen dieser Dissertation und ist in Abschnitt 3.3.1.5 beschrieben.

Um die Menge an zytosolischer Desaminierung massenspektrometrisch zu bestimmen, wurden drei klonalen *Doppelknockout*-Linien und der entsprechende Wildtyp für fünf Tage mit 0,2 mM- $^{13}\text{CD}_3$ -L-Methionin und 100 μM $^{13}\text{C}_9^{15}\text{N}_3$ -dC (dC^{+12}) unter *Priming*-Bedingungen kultiviert. Zur Verfolgung der DNA-Synthese wurde 7,5 μM $^{13}\text{C}_{10}^{15}\text{N}_5$ -dG zugegeben. Die stattfindenden, metabolischen Wege der Isotopenmarkierungen sind in Abbildung 50 schematisch dargestellt. Zur graphischen Vereinfachung sind die Nucleoside mit der größtmöglichen Isotopenmarkierung gezeigt.

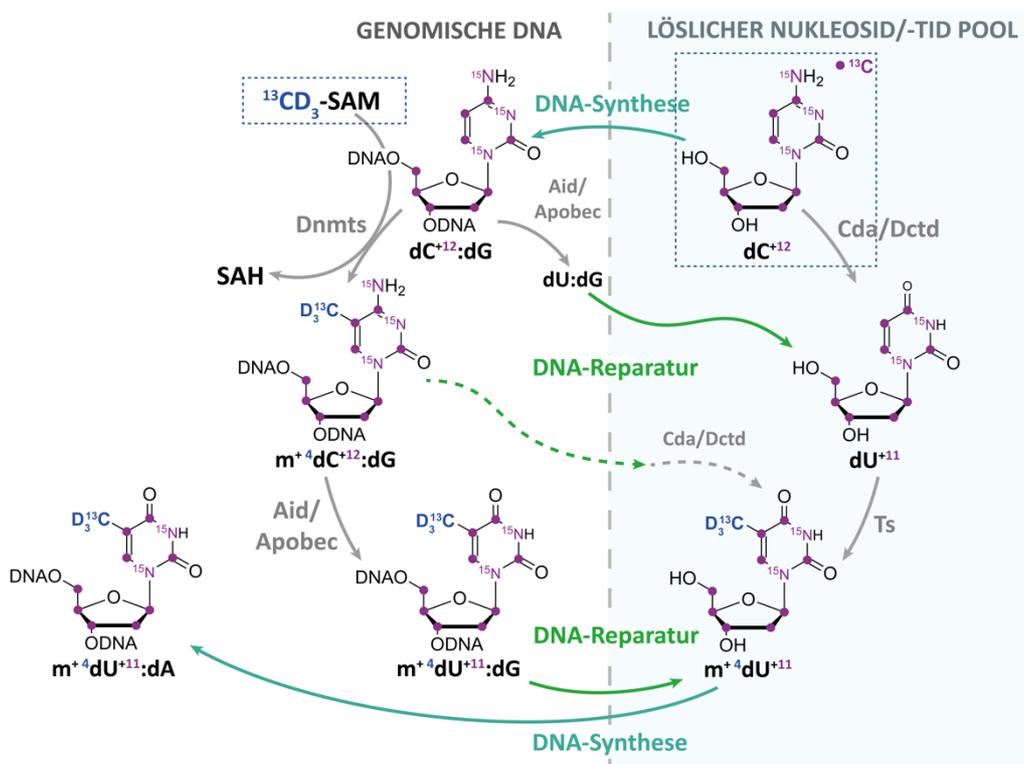


Abbildung 50: Mögliche Wege der Metabolisierung der eingebrachten Isotopenmarkierungen nach Zugabe von $^{13}\text{CD}_3$ -SAM und $^{13}\text{C}_9^{15}\text{N}_3$ -dC (dC^{+12}) zum Kulturmedium.

Nach Einbringung der dC^{+12} -Spezies in die Zelle befindet sich das Nucleosid zunächst im löslichen Nucleosid/-tid Pool. Dort kann das markierte Cytidin durch die Desaminasen Cda und Dctd zu dU^{+11} hydrolytisch desaminiert werden. Im nächsten Schritt findet eine Methylierung durch die Thymidylatsynthase (Ts) statt, welche mdU^{+11} ergibt. Die Methylgruppe liegt an dieser Stelle nicht markiert vor, da Ts 5,10-Methylentetrahydrofolat als Methyl-Quelle nutzt, wobei die übertragene Methylgruppe aus Serin stammt.

In der genomischen DNA wird ein Teil des eingebrachten, markierten dC^{+12} durch die Dnmt-Enzyme unter Verwendung des Kofaktors SAM zu mdC umgewandelt. In der genomischen DNA können daher drei isotoopenmarkierte dC-Verbindungen auftreten: mdC^{+12} , sowie die $^{13}CD_3$ -markierten $m^{+4}dC$ bzw. $m^{+4}dC^{+12}$ -Verbindungen. Die Existenz der mdC^{+12} -Spezies ist durch eine Restmenge an natürlichem Methionin im Kulturmedium zu begründen. Das natürliche Methionin wird vermutlich durch das Serum eingebracht. Die Existenz beider Spezies, $m^{+4}dC$ und $m^{+4}dC^{+12}$, resultiert aus der unvollständigen Markierung nach dC^{+12} -Fütterung und beträgt in Wildtyp mESCs etwa 20-30 % aller dC.

Durch Desaminierung über die *Activation-induced deaminase* (Aid) oder die *Apolipoprotein B mRNA-editing enzyme catalytic polypeptide* (ApoBec)-Desaminasen könnte sowohl markiertes dU als auch mdU fehlgepaart mit dG entstehen. Dieser Prozess wird in der Literatur vielseitig diskutiert, konnte aber nicht nachgewiesen werden.^[308] Diese dU-Verbindung kann durch die DNA-Reparatur entfernt werden und gelangt zurück in den löslichen Pool. Durch den Pyrimidin-*Salvage Pathway* können die markierten mdU Spezies gepaart mit dA wiederum im Zuge der DNA-Synthese zurück in die genomische DNA gelangen.

Die massenspektrometrischen Quantifizierungsergebnisse des beschriebenen Markierungsexperiments (siehe Abbildung 50) sind in Abbildung 51 dargestellt. Zunächst wurde die Verteilung der unterschiedlichen markierten mdC -Derivate in den verschiedenen Zelllinien betrachtet (Abbildung 51A). Neben dem natürlichen mdC konnten die isotoopenmarkierten Verbindungen $m^{+4}dC$, mdC^{+12} und $m^{+4}dC^{+12}$ detektiert werden. Im Wildtyp konnte eine Markierungseffizienz von 72 % $m^{+4}dC$ ($dC + ^{13}CD_3$ -SAM) aller mdC beobachtet werden. Dieses wird durch Übertragung der $^{13}CD_3$ -Methylgruppe von SAM mittels Dnmt-Katalyse aus unmarkiertem dC gebildet. Aus einer Markierungseffizienz von etwa 20 % an dC^{+12} , konnten 1,22 % der mdC^{+12} ($dC^{+12} + CH_3$ -SAM) und 22 % der $m^{+4}dC^{+12}$ ($dC^{+12} + ^{13}CD_3$ -SAM) Verbindung in der genomischen DNA quantifiziert werden. Das prozentual höhere Vorkommen der $m^{+4}dC$ bzw. $m^{+4}dC^{+12}$ Verbindungen ist aufgrund der sehr effizienten Markierung durch den Kofaktor SAM erklärbar. Die Menge an 1,22% mdC^{+12} entstammt durch Methylierung von gefüttertem dC^{+12} mit natürlichem CH_3 -SAM. Letzteres ist wohl eine Verunreinigung durch natürliches Methionin aus dem Serum des Kulturmediums. Bei Betrachtung der CD-DKO Zelllinien ist die Menge an $m^{+4}dC^{+12}$ auf 7-10 % aller mdC -Spezies reduziert. Eine Abnahme ist ebenfalls für die mdC^{+12} Verbindung zu verzeichnen. Dagegen nimmt die Menge an $m^{+4}dC$ um etwa 12 % zu. Vermutlich wird das markierte dC^{+12} von den CD-DKO mESCs weniger effizient in die DNA inkorporiert.

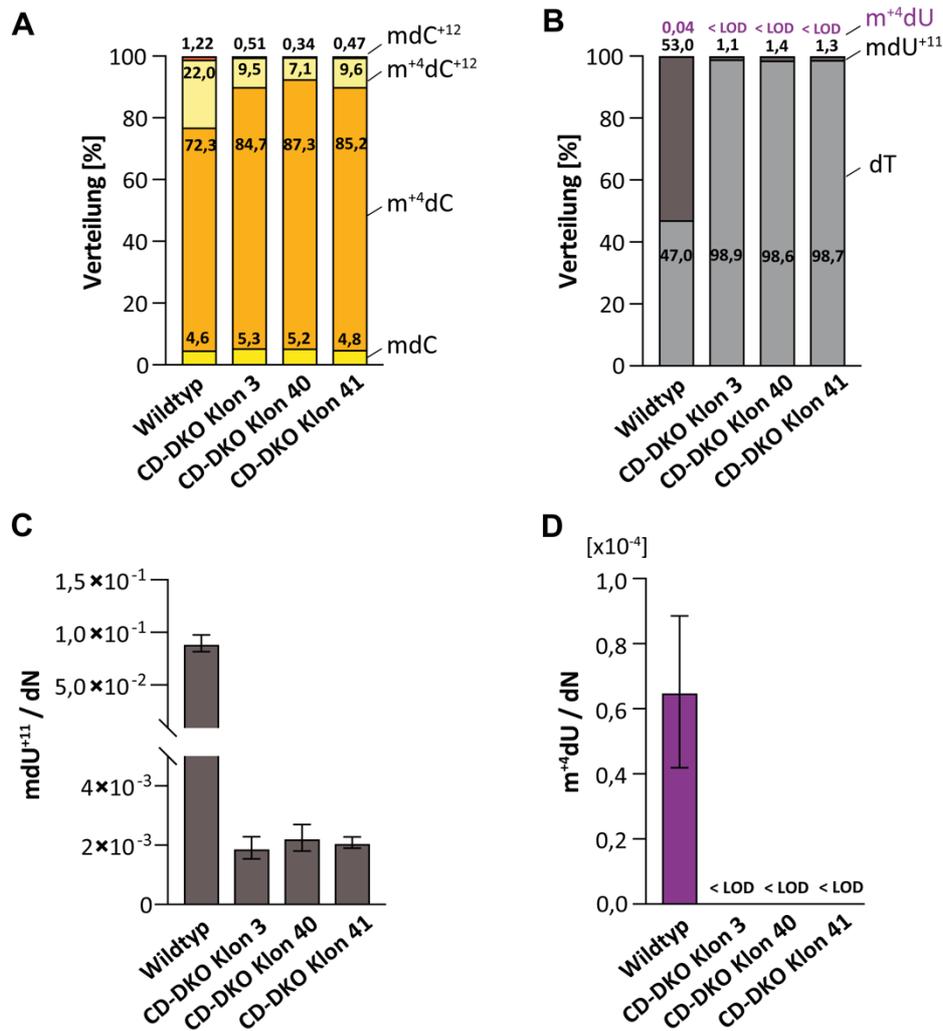


Abbildung 51: Quantifizierungsergebnisse nach Verfolgung der metabolischen Markierung im Doppelknockout der löslichen Desaminasen. A) Verteilung der isotonenmarkierten mdC-Spezies nach Einbringung der Isotonenmarkierung ins Genom. **B)** Verteilung der detektierten isotonenmarkierten mdU bzw. dT-Nukleoside normiert auf ¹³C₁₀¹⁵N₅-dG. **C)** Absolute Quantifizierung der mdU⁺¹¹ Verbindung. **D)** Absolute Quantifizierung der m⁺⁴dU Verbindung. Dargestellt sind die Ergebnisse einer biologischen Probe, gemessen als technisches Duplikat.

In den Wildtyp mESCs entsteht aus dC⁺¹² durch die Aktivität der Desaminasen des löslichen Pools die Verbindung dU⁺¹¹. Dies sollte in den CD-DKO mESCs nicht möglich sein. Im weiteren Verlauf wird diese Verbindung durch die Ts zu mdU⁺¹¹ methyliert. Da die Methylgruppe biosynthetisch ausgehend von Serin übertragen wird, liegt sie nicht markiert vor. Die Quantifizierungsergebnisse der möglichen mdU Isotopologe sind in Abbildung 51B als prozentualer Anteil aller dT-Verbindungen dargestellt. Im Wildtyp konnte ein Anteil von 53 % mdU⁺¹¹ (aus dU⁺¹¹ + CH₃-SAM) detektiert werden, während im CD-DKO der prozentuale Anteil bei 1,1-1,4 % lag. Die deutliche Abnahme der mdU⁺¹¹ Verbindung im DKO erklärt sich aus der Abwesenheit der Desaminasen Cda und Dctd. Die Menge an mdU⁺¹¹ im CD-DKO ist deutlich geringer als im Wildtyp (siehe Abbildung 51C), entspricht aber einer beachtlichen Menge von etwa 2,0 × 10⁻³ bezogen auf alle Nukleoside.

mdU⁺¹¹ könnte entweder aus einer Restaktivität der löslichen Nukleosid/-tid Pool Desaminasen (CD-DKO besitzt noch Aktivität der Desaminasen) oder durch eine Desaminierung von mdC⁺¹² auf genomischer Ebene entstehen. Neben dem Desaminierungsprodukt mdU⁺¹¹, aus dU⁺¹¹ + CH₃-SAM, konnte im Wildtyp zusätzlich m⁺⁴dU mit einem Anteil von 0,04 % detektiert werden. Dies entspricht im Wildtyp einer Menge von 0,6x10⁻⁴ m⁺⁴dU bezogen auf alle Nukleoside (Abbildung 51D). Dagegen liegt die Menge an m⁺⁴dU in den CD-DKO mESCs unter dem *Limit of Detection* (LOD). Diese Verbindung kann nur durch Desaminierung von m⁺⁴dC entstehen, welches durch Übertragung der Methyl-Gruppe von ¹³CD₃-SAM auf natürliches dC in die DNA gelangt. Als Entstehungswege kommen folgende Möglichkeiten in Frage (siehe auch Abbildung 50, grün): a) Desaminaseaktivität der Aid/Apobec-Familie auf genomischer DNA-Ebene und Entfernung der entstandenen mdU:dG Fehlpaarung durch die DNA-Reparatur sowie Re-Inkorporation als dT im Zuge der DNA-Synthese. b) Entfernung von mdC durch DNA-Reparatur und Desaminierung im löslichen Nukleosid/-tid Pool sowie Re-Inkorporation durch die DNA-Synthese als dT gegenüber dA. DNA-Reparaturprozesse, wie die *long-patch* Basenexzisionsreparatur oder die *Mismatch*-Reparatur rücken daher als mögliche Entstehungswege von m⁺⁴dU aus m⁺⁴dC in den Fokus und erscheinen wahrscheinlicher als eine genomische Desaminierung. Weiterhin gibt dieses Experiment keinen Hinweis auf die Entstehung von m⁺⁴dC aus m⁺⁴dU durch eine Desaminierung im Genom.

Beteiligung der DNA-Reparatur an der Umwandlung von mdC zu mdU

Zunächst sollte die Involvierung der Basenexzisionsreparatur an der oxidationsunabhängigen Entfernung von mdC betrachtet werden. Die DNA-Glykosylasen Tdg und Mbd4 entfernen fehlgepaartes dT (bzw. mdU) gegenüber dG und wurden bereits mit der Entfernung von genomischem mdC, durch ein Zusammenspiel aus Desaminierung und BER, in Verbindung gebracht.^{[191, 228], [309]} Eine bevorzugte Glykosylaseaktivität von Mbd4 für TpG-Stellen im symmetrischen mCpG-Kontext wurde ebenfalls beobachtet.^[310] Erste Experimente in *Tdg*^{-/-} bzw. *Mbd4*^{-/-} mESCs zeigten bereits einen Einfluss auf die genomischen m⁺⁴dU Level.^[249, 300] Eine orthogonale Aktivität der beiden Glykosylasen wurde daher vermutet. Um einen stärkeren Effekt auf die m⁺⁴dU Level zu erzielen, wurde unter Zuhilfenahme des CRISPR-Cas9 Systems eine *Doppelknockout*-mESC Linie generiert (beschrieben in Abschnitt 3.3.1.5). Zunächst wurden die oxidierten Cytidin-Spezies der einzelnen *Knockouts* und der erhaltenen *Doppelknockout*-Linie massenspektrometrisch analysiert. Das Quantifizierungsergebnis zeigt einen geringeren fdC bzw. cadC Wert im *Mbd4*^{-/-} verglichen mit der *Tdg*^{+/-} Linie. Dieser Unterschied könnte auf die verschiedenen genetischen Hintergründe der beiden Zelllinien zurückzuführen sein. Im *Doppelknockout* ist dagegen ein geringer Anstieg der fdC und cadC-Werte verglichen mit der *Tdg*^{-/-} Zelllinie (siehe Abbildung 52A) zu beobachten.

In vitro Experimente zeigten bereits eine geringe Glykosylaseaktivität von Mbd4 gegenüber fdC.^[163] Weitere biologische Proben wären notwendig, um eine konkretere Aussage über die Beteiligung von Mbd4 an der Entfernung dieser oxidierten dC-Spezies treffen zu können.

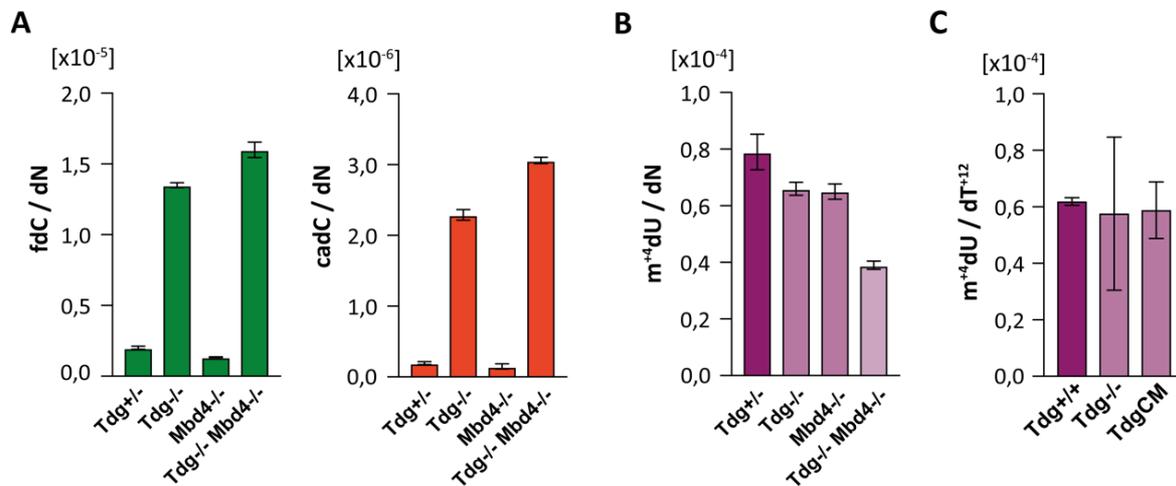


Abbildung 52: Einfluss der BER auf die mdC Reparatur. A) Quantifizierungsergebnisse der natürlichen höher oxidierten Cytidin-Spezies. **B)** Absolute Quantifizierungsergebnisse der m⁺⁴dU-Verbindungen im Doppelknockout im Vergleich zu den Einzelknockouts der BER-Glykosylasen. Dargestellt sind die Ergebnisse einer biologischen Probe gemessen als technisches Triplikat. **C)** Ergebnisse der absoluten Quantifizierung von m⁺⁴dU in der mittels CRISPR-Cas9 hergestellten *Tdg*^{-/-} Zelllinie und der katalytischen Mutante. Dargestellt sind die Ergebnisse eines biologischen Duplikats gemessen als technisches Triplikat nach Normierung auf dT⁺¹².

Als nächstes wurden die *Einzel-* und *Doppelknockout*-mESCs in Medium mit ¹³CD₃-L-Methionin kultiviert und die Akkumulation der m⁺⁴dU Spezies massenspektrometrisch untersucht. Beim Vergleich der *Tdg*^{+/-} mit den *Einzelknockouts* *Tdg*^{-/-} sowie *Mbd4*^{-/-} konnte eine Abnahme von m⁺⁴dU festgestellt werden (Abbildung 52B). Im *Doppelknockout* ist der m⁺⁴dU-Wert nochmals reduziert, was vermuten lässt, dass beide Glykosylasen an der Entfernung des markierten mdU beteiligt sind. Aus diesen Daten geht hervor, dass vermutlich beide Glykosylasen, Tdg und Mbd4, an der Entfernung des Desaminierungsproduktes beteiligt sind.

In einem weiteren Experiment wurden die Level an m⁺⁴dU in einer unabhängigen *Tdg*^{-/-}, einer katalytisch inaktiven Tdg-Mutante sowie dem korrespondierenden *Tdg*^{+/+} Wildtyp ebenfalls untersucht. Die Generierung dieser Zelllinie ist in Abschnitt 3.2.3.5 beschrieben. In unabhängigen Experimenten wurden geringe Unterschiede in der Geschwindigkeit der DNA-Synthese zwischen Wildtyp und *Knockout*-Linie festgestellt. Zur besseren Vergleichbarkeit der m⁺⁴dU-Level wurde dem Kulturmedium daher neben ¹³CD₃-L-Methionin markiertes ¹³C₁₀¹⁵N₂-dT (dT⁺¹²) beigefügt. Die erhaltenen m⁺⁴dU-Werte wurden schließlich auf das ebenso inkorporierte dT⁺¹² normiert und sind in Abbildung 52C dargestellt.

Nach Normierung konnte kein Unterschied an m^4dU zwischen Wildtyp und *Knockout* bzw. katalytischer Mutante festgestellt werden, was gegen eine Beteiligung von Tdg an der Generierung von m^4dU spricht. Um auszuschließen, dass der beobachtete Effekt der *Doppelknockout*-Linie aus Abbildung 52B aus einer unterschiedlichen DNA-Syntheserate resultiert, sollte dieses Experiment wiederholt und ebenfalls auf dT^{+12} normiert werden.

Beteiligung der *Mismatch*-Reparatur an der Umwandlung von mdC zu mdU

Zunächst sollte untersucht werden inwieweit das *Mismatch*-Reparaturprotein Mlh1 zur globalen Änderung der m^4dC -Level beiträgt. Hierfür wurde eine Mlh1 *Knockout*-Linie mit Hilfe des CRISPR-Cas9 Systems generiert (beschrieben in Abschnitt 3.3.1.5). Die generierte Zelllinie wurde als nächstes massenspektrometrisch untersucht. Die Quantifizierung der natürlichen dC-Modifikationen ergab einen Anstieg der hmdC, fdC und cadC Werte trotz eines geringen Abfalls von mdC im *Mlh1*^{-/-}. Die Quantifizierungsergebnisse sind in Abbildung 53A dargestellt.

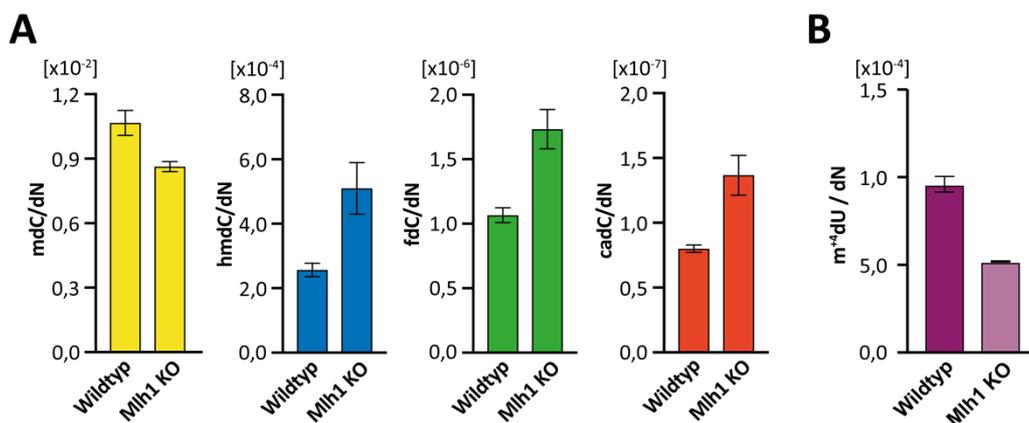


Abbildung 53: Einfluss der *Mismatch*-Reparatur auf die Reparatur von mdC. A) Quantifizierungsergebnisse der natürlichen höher oxidierten Cytidin-Spezies. **B)** Absolute Quantifizierungsergebnisse der m^4dU -Verbindung in Mlh1 *Knockout* und Wildtyp Zellen. Dargestellt sind die Ergebnisse einer biologischen Probe gemessen als technisches Triplikat.

Dies spricht für eine potentielle Beteiligung der *Mismatch*-Reparatur (MMR) an der Entfernung der oxidierten mdC-Derivate. Weiterhin wurde dieselbe Beobachtung bereits im Falle der *Aid*^{-/-} mESCs gemacht. In der Literatur ist eine Beteiligung von Aid an der MMR ebenfalls bekannt.^[311] Die Abnahme der mdC-Level könnte in beiden Fällen auch durch eine fehlgeschlagene Remethylierung erklärt werden. Nach dieser Theorie würde es zu einer Akkumulation der Intermediate hmdC, fdC und cadC kommen, die ohne Aid oder Mlh1 nicht weiter zu dC demodifiziert werden können. Ohne dC wäre kein Substrat für eine potentielle Remethylierung vorhanden.

Als nächstes wurden die Zellen mit isotoopenmarkiertem Methionin unter *Priming*-Bedingungen kultiviert. Nach Analyse der genomischen DNA konnte eine Reduktion der m⁴dU-Level beobachtet werden (siehe Abbildung 53B). Im Vergleich zum Wildtyp ist eine Abnahme um etwa 50 % im *Mlh1-Knockout* zu sehen. Da die MMR bis zu 2 kb umfassen kann, ist diese beachtliche Reduktion durchaus erklärbar. Dieser Befund passt ebenfalls zur Entfernung mehrerer markierter mdC-Stellen aus dem Genom. Nach Prozessierung durch die MMR gelangen die Nukleotide in den löslichen Pool, werden dort desaminiert und als mdU bzw. dT gepaart mit dA wieder inkorporiert. Im Falle des *Knockouts* ist dies nicht möglich und die Menge an markiertem mdU sinkt. Auch in *Knockout-Zellen* eines weiteren wichtigen MMR-Proteins, *MutS protein homolog 2 (Msh2)*, konnte ein ähnlicher Trend beobachtet werden.^[300] Die Daten passen zu der Annahme, dass ein enzymatischer Prozess DNA-Läsionen erzeugt und so ein Reparaturprozess aktiviert wird, welcher sich über mehrere benachbarte mdC-enhaltende Bereiche erstreckt. Ein ähnliches Modell wurde bereits für die *Immunoglobulin Class-Switch Recombination* beschrieben. Die direkte Desaminierung von dC durch Aid rekrutiert dabei DNA-Reparaturprozesse, wie die *long-patch* BER oder die nicht-kanonische MMR, zur Reparatur des Doppelstranges.^[312, 313]

3.3.1.3 Zusammenfassung

Die erhaltenen Ergebnisse deuten auf eine aktive Demethylierung von genomischem mdC in *primed* murinen Stammzellen hin. Das *Pulse-Chase* Experiment zeigt eine Abnahme von genomischem mdC, die nicht durch eine passive Verdünnung begründet werden kann. Gleichzeitig wird ein Anstieg des mdC Desaminierungsprodukts mdU während des *Primings* der mESCs beobachtet. Die höher oxidierten dC-Derivate bleiben dabei länger in der genomischen DNA bestehen. Eine direkte genomische Desaminierung konnte nicht beobachtet werden. Der *Doppelknockout* der Desaminasen *Cda* und *Dctd* bestätigt weiterhin, dass das markierte mdU größtenteils durch Desaminierung von mdC im löslichen Nukleosid/-tid Pool entsteht. Hinweise auf eine Beteiligung der DNA-Reparaturprozesse MMR und BER konnten ebenso erhalten werden. Die *Knockout*-mESCs der einzelnen Glykosylasen zeigen unter *primed* Kulturbedingungen einen Einfluss auf die Bildung Desaminierungsprodukt mdU. Zur Klärung der Fragestellung ob das markierte mdU aktiv gesteuert oder eher zufällig in das Genom der Stammzellen gelangt, bedarf es weiterer Untersuchungen.

Teile dieses Kapitels wurden bereits im *Preprint*-Journal BioRxiv (*Spada et al.*, 846584) veröffentlicht.

3.3.1.4 Zusammenfassung der Projektbeiträge

Die Generierung aller *Knockout* mESC-Linien wurde von mir geplant und durchgeführt. Die Verifizierung wurde ebenfalls von mir vorgenommen. Die Planung und Ausführung der Zellkulturexperimente wurden von *Dr. Fabio Spada* und mir durchgeführt. Die Isolation der DANN-Proben wurde von mir in Zusammenarbeit mit *Luis de la Osa de la Rosa* unternommen. Der enzymatische Verdau, die Messung und die Auswertung der Daten wurden von *Dr. Olesya Kosmatchev*, *Dr. Jessica Steinbacher* oder *Sarah Schiffers* vorgenommen. Alle Daten wurden von *Dr. Fabio Spada*, *Sarah Schiffers* und mir gemeinsam interpretiert. *Dr. Fabio Spada*, *Yingqian Zhang* und *Dr. Sarah Schiffers* führen das Projekt fort.

3.3.1.5 Appendix: Generierung und Verifizierung der *Knockout*-Linien

Cda und *Dctd* Doppelknockout mES-Zelllinie

Cda tritt in der Zelle als Homotetramer (Monomer 16 kDa) auf und katalysiert die hydrolytische Desaminierung von Cytidin bzw. Desoxycytidin-Nucleosiden zu den jeweiligen Uridin-Derivaten.^[294] *Dctd* hingegen ist 20 kDa groß und katalysiert die Reaktion von Cytidin- bzw. Desoxycytidinmonophosphat zum Uridinmonophosphat.^[314] Diese Enzyme sind Schlüsselenzyme des Pyrimidin-Nucleosid/-tid *Salvage Pathways* und arbeiten im löslichen Nucleosid/-tid Pool der Zelle. Es wurde aber auch eine Lokalisation im Kern beobachtet.^[315] Die *Cda* und *Dctd* Doppelknockout mES-Zelllinie wurde durch Kotransfektion zweier *guide* RNA (gRNA) und Cas9 bewerkstelligt. Die Endonuklease Cas9 erzeugt, geführt durch die gRNA, einen Doppelstrangbruch im ersten Exon im Falle von *Cda* und im vierten Exon im Falle von *Dctd* (siehe Abbildung 54A).

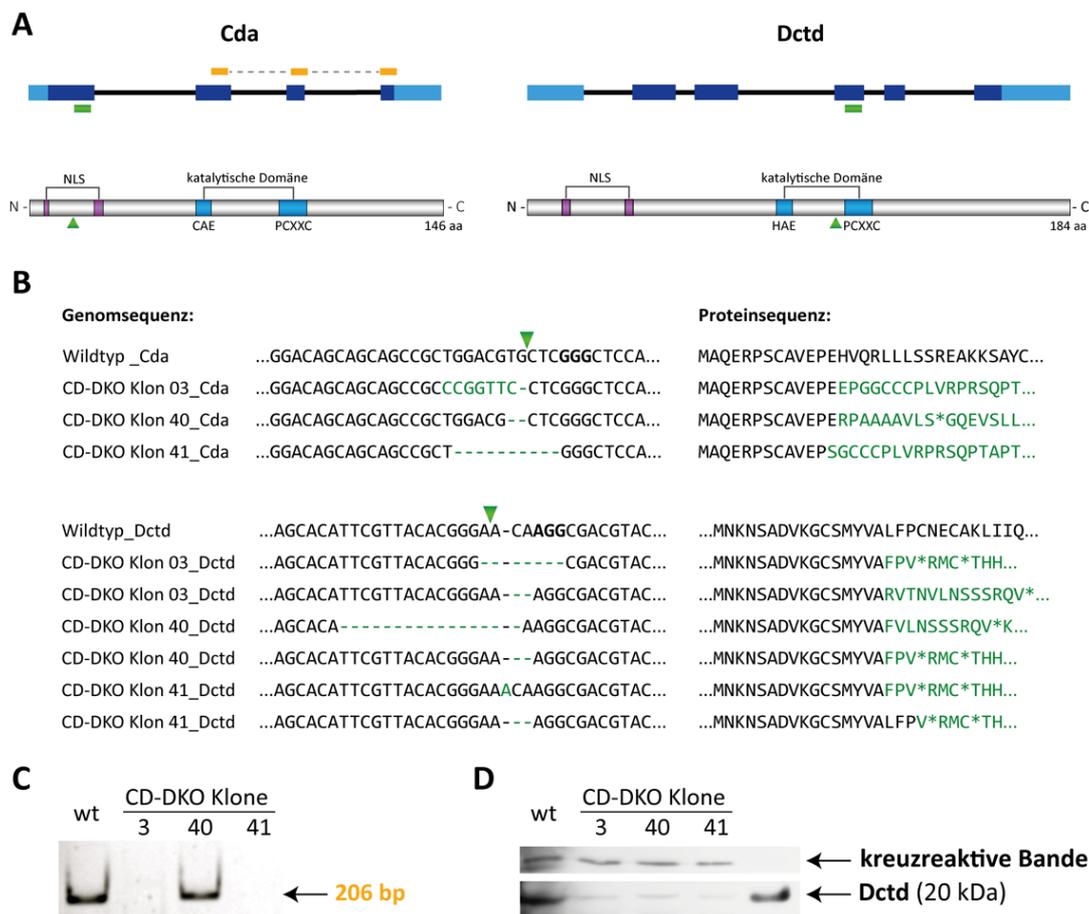


Abbildung 54: Strategie und Verifizierung der *Cda*-*Dctd* Doppelknockout Zelllinie. **A)** Schema des Genabschnitts mit kodierenden (blau) und nichtkodierenden Exons (hellblau), sowie der Bindestelle der gRNA (grün). Das Amplifikationsprodukt der RT-PCR ist in orange abgebildet. Darstellung des jeweiligen Proteins mit wichtigen Domänen und der CRISPR-Position (grün).^[294, 315] **B)** Sequenzierungsergebnisse und eingeführte Mutationen im CD-DKO (*Cda*, *Dctd*-Doppelknockout). Vorhergesagte Proteinsequenz resultierend aus den Sequenzierungsergebnissen. **C)** Reverse Transcriptase-PCR (RT-PCR) zur Untersuchung der Expression der *Cda* mRNA. **D)** Western Blot zur Verifizierung des *Dctd*-Knockouts.

Durch die Mutationen wird das Leseraster verschoben und es entsteht ein frühzeitiges Stoppcodon. Die CRISPR-*Targeting*-Stelle liegt in beiden Fällen 5' von der Region, die für das katalytische Zentrum kodiert und sollte zu keinem funktionalen Protein führen (siehe Abbildung 54B). Das zelleigene Reparatursystem führt Insertionen und Deletionen an der Stelle des Strangbruchs ein. Diese wurden durch Sequenzierung des Ziellokus verifiziert. Es konnten drei klonale Zelllinien identifiziert werden (siehe Abbildung 54C). Die Sequenzierung zeigt im Falle des Cda-Genabschnitts in allen drei klonalen Zelllinien eine homozygote Mutation (Abbildung 54C). Für Dctd konnten zwei unterschiedliche Mutationen auf den beiden Allelen pro Zellklon festgestellt werden (siehe Abbildung 54C). Bei Betrachtung der Proteinsequenz wird deutlich, dass alle Mutationen zu einem verfrühten Stoppcodon und so zum *Knockout* des Zielproteins führen sollten (Abbildung 54D). Zur Überprüfung des *Knockouts* des Zielproteins Cda wurde eine *Reverse Transkriptase*-PCR (RT-PCR) durchgeführt. Diese amplifiziert die mRNA des Zielgens. Im Falle eines verfrühten Stoppcodons wird die mRNA durch den *Nonsense Mediated Decay Pathway* abgebaut. Die RT-PCR zeigt die durch Amplifikation erwartete Bande im Wildtyp und im Klon 40 (siehe Abbildung 54E). Die Klone 3 und 41 scheinen keine mRNA für Cda und somit kein Protein aufzuweisen. Ob in Klon 40 tatsächlich das Protein vorhanden ist oder nicht bleibt fraglich. Der Western Blot zur Detektion des Dctd-Proteins zeigt eine schwache Bande in allen *Doppelknockout* Zelllinien (Abbildung 54F). Dies könnte auf eine reduzierte Expression des Dctd Proteins hinweisen oder eine kreuzreaktive Bande sein. Eine Restaktivität von Dctd im CD-DKO (Cda, Dctd-*Doppelknockout*) kann daher nicht ausgeschlossen werden.

Tdg und Mbd4 Doppelknockout mES-Zelllinie

Als Parentallinie wurde eine *Tdg*^{-/-} mESC-Linie genutzt und darin Mbd4 ausgeschaltet. Unter Zuhilfenahme einer Reparaturvorlage sollte gezielt ein Stoppcodon zur Ausschaltung des Proteins im ersten Exon eingeführt werden. Die verwendete Strategie nutzt die Cas9-Endonuklease zusammen mit zwei gRNA und die Vorlage. Die Reparaturvorlage enthält ein vorzeitiges Stoppcodon und führt gleichzeitig eine Restriktionsschnittstelle zur Genotypisierung ein. Nach Doppelstrangbruch durch Cas9 erfolgt die Einbringung des Templats durch *Homology Directed Repair* (HDR). Nach *Limiting Dilution* wurden die klonalen Zelllinien genotypisiert. Hierzu wurde der mutierte Abschnitt durch PCR amplifiziert und einem Restriktionsverdau unterzogen. Das Bandenmuster wurde mit Hilfe einer Polyacrylamid-gelelektrophorese (PAGE) untersucht. Die Ergebnisse sind in Abbildung 55A dargestellt. Die Kontrolle (*Mbd4*^{+/+}) enthält einen nicht mutierten Genabschnitt. Die PCR-Amplifikation ergab eine Bande bei 499 bp, wohingegen nach *PstI*-Verdau (verd.) zwei Banden (123 und 376 bp) zu sehen sind. Dies ist zu erwarten, da der Locus bereits eine *PstI*-Restriktionsschnittstelle enthält. Im positiven Klon hingegen sind drei Banden zu erkennen, da hier eine zusätzliche *PstI*-Restriktionsschnittstelle eingeführt wurde. Es kommt daher zu Banden von 68, 123 und 308 bp.

Die Bande bei 499 bp ist in diesem Zellklon nicht mehr zu sehen, was für einen vollständigen Verdau spricht. Eine schwache Bande bei 376 bp ist zu erkennen und könnte auf ein nicht-mutiertes Allel hindeuten.

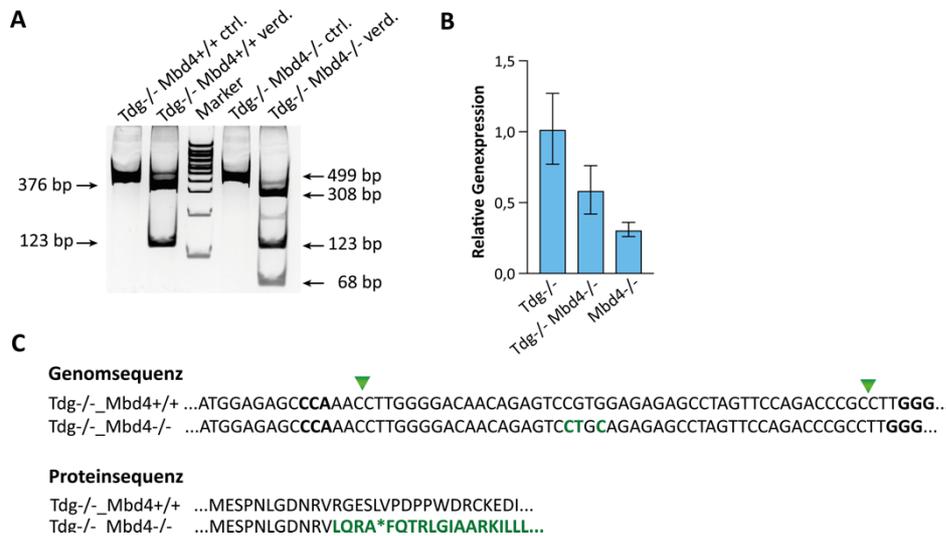


Abbildung 55: Generierung und Verifizierung des *Tdg*^{-/-} *Mbd4*^{-/-} Doppelknockouts. **A)** Natives Polyacrylamidgel nach Restriktionsverdau des Doppelknockouts und der korrespondierenden parentalen Zelllinie, sowohl vor (ctrl.) und nach (verd.) Verdau. **B)** RT-qPCR zur Überprüfung der Expression des *Mbd4*-Gens. **C)** Sequenzierungsergebnisse des modifizierten Genabschnitts mit resultierender Proteinsequenz und der Alterierung (grün).

Zur Verifizierung der *Knockout*-Linie wurde daher die Menge des mRNA-Transkripts mittels RT-qPCR untersucht. Die Bindestellen der Primer wurden zu diesem Zweck für die letzten Exons entworfen, da von hier aus auch der Abbau stattfinden sollte. Die Ergebnisse der RT-qPCR sind in Abbildung 55B aufgezeigt. Als *Housekeeping*-Gen wurde *Gapdh* verwendet. Neben der parentalen *Tdg*^{-/-} wurde auch eine bereits etablierte *Mbd4*^{-/-} mESC-Linie als Positivkontrolle untersucht. Diese wurde durch *Gene Targeting* in Stammzellen der Linie E14-Tg2a generiert.^[316] Eine Abnahme der Genexpressionslevel des *Mbd4*-Transkripts wird in *Mbd4*^{-/-} und den *Tdg*^{-/-} *Mbd4*^{-/-} mESCs beobachtet. Die Reduktion des *Mbd4*-Genexpressionslevels deutet darauf hin, dass mindestens ein Allel des putativen *Doppelknockouts* mutiert wurde. Ob es sich jedoch bei der generierten Zelllinie um einen tatsächlichen *Doppelknockout* handelt, erfordert weitere Evaluierung. Die Sequenzierungsergebnisse zeigen die Einführung der Mutation auf beiden Allelen. Durch Sanger-Sequenzierung konnte nur die Sequenz der eingeführten Reparaturvorlage beobachtet werden. Die daraus resultierende Proteinsequenz zeigt ein vorzeitiges Stoppcodon und eine Alterierung der Aminosäuresequenz (siehe Abbildung 55C). Zur vollständigen Verifizierung der *Doppelknockout*-Linie sollte jedoch ein Western Blot zur Untersuchung der Proteinexpression durchgeführt werden. Erste Versuche hierzu schlugen fehl. Die genutzten Antikörper ergaben neben kreuzreaktiven Banden auch eine Bande in der bereits etablierten *Mbd4*^{-/-} Zelllinie.

Knockout von *Mlh1* in mES-Zellen

Zunächst wurden zwei gRNA entworfen mit dem Ziel, das Exon1 des Gens zu mutieren. Durch die Cas9 Nickase (Cas9n) werden Einzelstrangbrüche erzeugt und das Reparaturtemplat durch HDR eingebaut. Da Einzelstrangbrüche im Zuge der BER repariert werden, sollten auf diese Weise *off-target* Mutationen minimiert werden. Das Templat soll eine Restriktionsschnittstelle sowie Punktmutationen in den Ziellokus einführen. Die Verschiebung des Leserasters führt zu vier Stoppcodons, wodurch kein funktionales Protein mehr entstehen kann. Nach Transfektion und *Limiting Dilution* wurde eine Genotypisierung der Zellklone unternommen. Mittels PCR wurde der Ziellokus amplifiziert und im Anschluss einem Restriktionsverdau unterzogen. Der Verdau wurde mit einem PAGE-Gel analysiert. Die Ergebnisse sind in Abbildung 56B dargestellt.

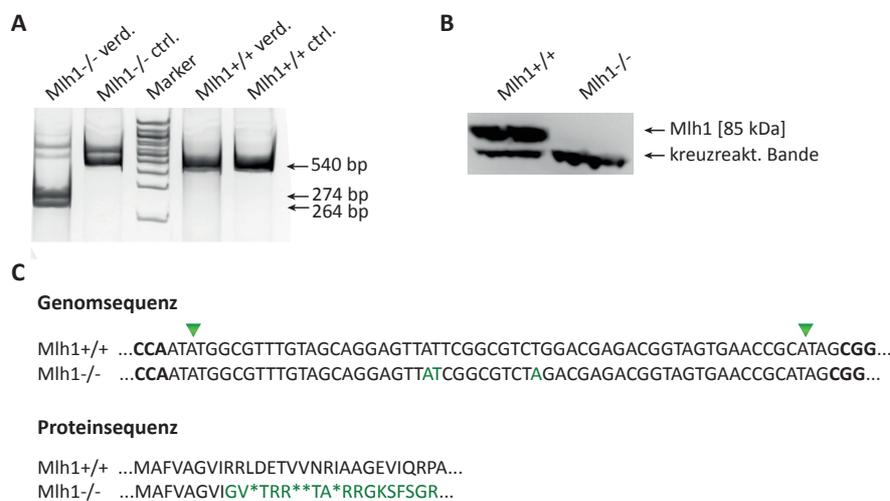


Abbildung 56: Generierung und Verifizierung der *Mlh1*^{-/-} mESCs. A) Natives Polyacrylamidgel nach Restriktionsverdau des *Knockouts* und der korrespondierenden parentalen Zelllinie, sowohl vor (ctrl.) als auch nach (verd.) Verdau. B) Western Blot zum Nachweis des *Knockouts*. C) Sequenzierungsergebnisse des modifizierten Genabschnitts mit resultierender Proteinsequenz und der Alterierung (grün).

Die Kontrolle zeigt eine PCR-Bande von 540 bp. Nach erfolgreicher Einbringung des Reparaturtemplats ist eine Restriktionsschnittstelle vorhanden, die nach Verdau zu Banden von 264 und 274 bp führt. Dies konnte für den *Knockout*-Klon beobachtet werden. Als Kontrolle wurde ein negativer Klon ebenfalls amplifiziert und einem Restriktionsverdau unterzogen. Dieser zeigt keine Fragmente. Auch die Entstehung des Proteins wurde anhand eines Western Blots untersucht und zeigt im *Knockout* kein Protein (Abbildung 56B). In Abbildung 56C sind die Sequenzierungsergebnisse des Ziellokus dargestellt. Die eingeführten Punktmutationen sind markiert und die Genomsequenz wurde in die entstehende Proteinsequenz translatiert. Die vier Stoppcodons sind ebenfalls dargestellt und zeigen, dass kein funktionales Protein mehr entstehen kann. Die *Mlh1*^{-/-} mESC-Linie konnte erfolgreich hergestellt werden.

4 Material und Methoden der unveröffentlichten Arbeiten

4.1 Material und Methoden zu Abschnitt 3.1.2

4.1.1 Zelllinien

Die verwendeten J1-mESCs (*murine Embryonic Stem Cells*) entstammen der Zelllinie 129/Sv.^[317] Diese Zellen stammen von *Kerry Tucker* (Universität Heidelberg). Die Dnmt-TKO mESCs wurden von *Masaki Okano* (*RIKEN Center for Developmental Biology*) bereitgestellt.^[112] Sowohl der Wildtyp E14Tg2a als auch die Tet-TKO mESCs wurden von *Yi Zhang* (*Harvard Stem Cell Institute*) bezogen.^[256, 318] Die iNGN-Zellen wurden von *Stylianos Michalakis* (LMU) zur Verfügung gestellt.^[260] Die Zelllinien *Tdg*^{+/-} und *Tdg*^{-/-} wurden von *Primo Schär* (Universität Basel) bereitgestellt.

4.1.2 Isotopenmarkierungsexperimente

D₂¹⁵N₂-dC:

Zur Generierung von *Epiblast-like* Zellen wurden die murinen Stammzellen der Wildtyp Linie J1 für zwei Passagen in serumfreiem 2i/LIF Medium kultiviert. Dieses Medium beinhaltet DMEM *High Glucose* (*Sigma-Aldrich*), Hams F-12 (*Sigma-Aldrich*), Neurobasal Medium (*Life Technologies*), 2 mM L-Glutamin, 1x Penicillin-Streptomycin (100x, *Sigma-Aldrich*), 0.1 mM β -Mercaptoethanol, 0,5x NS21 (50x, *Life Technologies*) und 0,5x N2 (100x, hergestellt von *Fabio Spada*). Zur Kultur der Zellen unter naiven Bedingungen wurden 1000 U/mL LIF (*ORF Genetics*), 3.0 μ M GSK3 Inhibitor CHIR99021 und 1.0 μ M Mek Inhibitor PD0325901 (2i; *Selleckchem*) dem Medium zugesetzt. Für das Experiment wurden 3,0 x 10⁴ Zellen/cm² in einer mit Geltrex (1:1000 in 1:1 DMEM/Hams F-12 mit 1x Penicillin-Streptomycin; *Thermo Fisher Scientific*) beschichteten Kulturschale ausgesät. Die Zellen wurden für zwei Tage in serumfreiem 2i/LIF Medium, sowie mit 100 μ M D₂¹⁵N₂-dC (hergestellt von *Dr. René Rahimoff*) kultiviert.

Anschließend wurde das Medium gewechselt und die Differenzierung zu EpiLCs durch Zugabe von serumfreiem Medium mit den Wachstumsfaktoren 12 ng/mL FGF-2, 20 ng/mL Activin A und 1 % *Knockout Serum Replacement* (KSR), sowie unter Zugabe von 100 μ M D₂¹⁵N₂-dC gestartet.^[251, 252] Das Basismedium der mESCs beinhaltet DMEM *High Glucose* (*Sigma-Aldrich*), 2 mM L-Glutamin, 1x Penicillin-Streptomycin (100x, *Sigma-Aldrich*), 0.1 mM β -Mercaptoethanol, 1x *Non-essential Amino Acid Solution* (100x, *Sigma-Aldrich*).

Die Stammzellen wurde auf mit Gelatine (*Sigma-Aldrich*) beschichteten Platten durch Zugabe von 1000 U/mL LIF (*ORF Genetics*), 3 μM CHIR99021 und 1 μM PD0325901 unter naiven 2i/LIF Bedingungen kultiviert. Die *Tdg*^{+/-} und *Tdg*^{-/-} Zellen wurden zunächst unter 2i/LIF Bedingungen und 100 μM D₂¹⁵N₂-dC für drei Tage kultiviert.

Ein Teil der Zellen wurde zur Analyse geerntet (siehe Abschnitt 4.1.3). Der übrige Teil der Zellen wurde unter CHIR/IWR1-Bedingungen (siehe Abschnitt 3.1.1) und 100 μM D₂¹⁵N₂-dC für drei Tage weiter *geprimed*.

D₄-dC:

Für die Markierungsexperimente wurden die Zellen drei Tage unter 2i/LIF Bedingungen zusammen mit 200 μM D₄-dC (*Tdg*^{+/-} und *Tdg*^{-/-}) bzw. 100 μM D₄-dC (Dnmt-TKO, Tet-TKO, sowie der korrespondierende Wildtyp) kultiviert. Zum dreitägigen *Priming* wurden die Zellen in Basismedium zusammen mit 3 μM CHIR99021 und dem Wnt-Inhibitor IWR1-endo mit 2,5 μM (C/R-Bedingung) kultiviert. Die markierten Nukleoside wurden in derselben Konzentration wie unter 2i/LIF Bedingungen zugegeben.

¹³C₅D₁-dC und ¹³C₅D₁-rC:

Die *Tdg*^{+/-} und *Tdg*^{-/-} Zellen wurden zunächst unter 2i/LIF Bedingungen und 200 μM ¹³C₅D₁-dC bzw. 50 μM ¹³C₅D₁-dC für drei Tage kultiviert. Ein Teil der Zellen wurde geerntet (siehe Abschnitt 4.1.3). Der übrige Teil der Zellen wurde unter C/R-Bedingungen mit 200 μM ¹³C₅D₁-dC bzw. 50 μM ¹³C₅D₁-dC ebenfalls für drei Tage kultiviert.

Im *Pulse-Chase* Experiment wurden die Dnmt-TKO und Wildtyp Zellen für 36 h unter C/R Bedingungen kultiviert. Anschließend wurde das Medium gewechselt und 100 μM ¹³C₅D₁-dC, sowie 5 μM ¹³C₁₀¹⁵N₂-dT für 12 h zugegeben. Nach Ablauf der Zeit wurden die Zellen zweimal mit PBS-Puffer gewaschen und frisches C/R-Medium ohne die isotope-markierten Nukleoside zugegeben. Die unmarkierten Kontrollen, sowie die Chase Kulturen wurden nach 0, 4, 8 und 12 h geerntet. Auf diese selbe Weise wurde das Experiment mit den Tet-TKO Zellen durchgeführt.

Im *Pulse-Chase* Experiment wurden die Dnmt-TKO und Wildtyp Zellen für 24 h unter C/R-Bedingungen kultiviert. Anschließend wurde das Medium gewechselt und 200 μM ¹³C₅D₁-rC, sowie 5 μM ¹³C₁₀¹⁵N₂-dT für 24 h zugegeben. Nach Ablauf der Zeit wurden die Zellen zweimal mit PBS-Puffer gewaschen und frisches C/R Medium ohne die isotope-markierten Nukleoside zugegeben. Die unmarkierten Kontrollen, sowie die *Chase*-Kulturen wurden nach 0, 4, 8 und 12 h geerntet. Auf dieselbe Weise wurde das Experiment mit den Tet TKO Zellen durchgeführt.

Die INGN-Zellen wurden auf Geltrex beschichteten Kulturschalen (1:1000) in E9-Medium kultiviert. Das E7-Basismedium beinhaltet DMEM *high glucose* (Sigma), Hams F-12 (Sigma), 2 mM L-Alanyl-L-Glutamin (200 mM, Sigma), 64 mg/L L-Ascorbinsäure-2-phosphat (Sigma), 1x Penicillin-Streptomycin (100x, Sigma), 0,1 mM Nicotinamide, 77,6 nM Natriumselenit, 11,18 g/L NaCl. Vor Nutzung des Mediums wurden 10 µg/mL humanes Holo-Transferrin und 20 µg/mL Insulin zugegeben.

Das E7-Medium wurde durch einen 0,2 µM Filter steril gemacht. Um das E9-Medium zu vervollständigen wurden die Wachstumsfaktoren hFGF-2 100 ng/mL und hTGF-β1 2 ng/mL zugegeben. Im *Pulse-Chase* Experiment wurden die Zellen unter diesen Bedingungen für 24 h 140 µM ¹³C₅D₁-dC und 0,5 µM ¹³C₁₀¹⁵N₂-dT bzw. 50 µM ¹³C₅D₁-rC und 0,5 µM ¹³C₁₀¹⁵N₂-dT kultiviert. Anschließend wurde das Medium gewechselt und nukleosidfreies E7-Basismedium mit Zugabe von Doxycyclin 0,5 µg/mL verwendet. Das Doxycyclin induziert die Bildung von Neurogenin 1 sowie 2, der Expressionskassette und sorgt für die Differenzierung der Zellen in die neuronale Linie. Zeitpunkte wurden vor Induktion (0 h) und jeweils nach 24 h, 48 h und 72 h nach Induktion geerntet.

4.1.3 Isolation der genomischen DNA

Die genomische DNA wurde, wie in Abschnitt 3.1.1 beschrieben isoliert.

4.1.4 Massenspektrometrische Analyse

Die massenspektrometrische Analyse erfolgte, wie in den Dissertationen von *Dr. Jessica Steinbacher* und *Dr. Katharina Iwan* beschrieben.^[249, 257]

4.2 Material und Methoden zu Abschnitt 3.1.3

4.2.1 Verwendete Zelllinien

Die verwendeten R1-mESCs (embryonalen Stammzellen der Maus, *murine Embryonic Stem Cells*) als auch die *Sirt1*^{-/-} mESCs wurden von *Michael W. McBurney* (University of Ottawa) bereitgestellt. Die *Dnmt1*^{-/-} Zelllinie, sowie die *Dnmt3a*^{-/-} und *Dnmt3b*^{-/-} mESCs wurden von *En Li* (*Novartis Institutes for Biomedical Research, Shanghai*) zur Verfügung gestellt. Die verwendeten J1-mESCs (embryonalen Stammzellen der Maus, *murine Embryonic Stem Cells*) entstammen der Zelllinie 129/Sv/Jae. Die verwendeten *Kindlin3*^{+/+} Zellen wurden von *Reinhard Fässler* zur Verfügung gestellt.^[319] Die IB10-Zelllinie wurde von *Hein te Riele* (*Netherlands Cancer Institute*) bezogen und ist ein Subklon der E14 Zelllinie entstammt dem genetischen Hintergrund 129/Ola.^[320] Die *Mlh1*^{-/-} Stammzelllinie wurde von mir hergestellt. Die Generierung mittels CRISPR-Cas9 wird in Abschnitt 3.3 beschrieben. Die CHO-K1 Zelllinie wurde von ATCC (ATCC®CCL-61™) bezogen.

4.2.2 Kulturbedingungen der Säugetierzellen

Die Kultur und Fütterung der mESCs erfolgte, wie in Abschnitt 3.1.1 und 3.3 beschrieben. Die CHO-K1 Zellen wurden in DMEM/F12 Medium kultiviert. Dieses beinhaltet für 100 mL des Mediums 44 mL HAMs-F12 (N4888, Sigma), 44 mL DMEM (D6546, Sigma), 10 mL FBS (*Heat Inactivated*, Invitrogen), 1 mL L-Alanyl-L-Glutamin (finale Konzentration 2 mM, Sigma), 1 mL Penicillin-Streptomycin-Lösung (1x, Sigma). Die adhärenenten Zellen wurden in einem Kulturgefäß bei 37 °C und 5 % CO₂ kultiviert.

4.2.3 Synchronisation der CHO-K1

CHO-K1 Zellen wurden 24 h in DMEM/F12 bis zu einer Konfluenz von 30 % kultiviert. Anschließend wurde dem Kulturmedium 2 mM Thymidin für 16 h zugesetzt. Das Medium wurde erneuert und die Zellen für 8 h weiter kultiviert. Der Thymidin-Block wurde erneut für 16 h durchgeführt. Die CHO-K1 wurden anschließend aus der Blockade befreit und 50 µM ¹³C₅¹⁵N₂-5-fdC zugesetzt. Es wurden Kulturen nach 0, 4 und 8 h geerntet.

4.2.4 Durchflusszytometrische Analyse

Die Zellen wurden mit PBS gewaschen und anschließend mit TrypLEexpress (Sigma) aus der Kulturschale gelöst. Erneut wurde zweimal mit PBS gewaschen. Anschließend wurden die Zellen 3 min bei 300 xg zentrifugiert. Der Überstand wurde verworfen und das Zellpellet in 200 µL HSF-Lösung (0,1 % (v/v) Triton-X-100; 0,1 % (w/v) Natriumcitrat; 50 µg/mL Propidiumiodid in PBS) resuspendiert. Die Zellsuspension wurde im Dunkeln bei 4 °C über Nacht inkubiert. Der DNA-Gehalt wurde mittels Durchflusszytometrie (BD Biosciences, FACSCanto II) vermessen. Die Daten wurden mit Hilfe der Software FlowJo 887 analysiert.

4.2.5 Oligonukleotide

Die synthetischen Oligonukleotide (Tabelle 5) wurden von *Ella Biotech* oder *Merck* bezogen. Das genutzte isotopenmarkierte $^{15}\text{N}_2$ -fdC (^fC) Phosphoramidit wurde von *Alexander Schön* synthetisiert.

Tabelle 5: Genutzte Oligonukleotidsequenzen.

Bezeichnung	Sequenz	Verwendung
<i>Top strand</i> _fdC	5'-GAGAAGCTGGGACTTA f CGGCTGGAGAGTGC-3'	Sirtuin-Assay
<i>Bottom strand</i> _mdC	5'-GCACTCTCCAGC m CGTAAGTCCCAGCTTCTC-3'	Sirtuin-Assay
<i>Bottom strand</i> _dC	5'-GCACTCTCCAGCCGTAAGTCCCAGCTTCTC-3'	Sirtuin-Assay

4.2.6 Isolation der genomischen DNA

Die genomische DNA der Zellen wurde, wie im Abschnitt 3.1.1 beschrieben, isoliert.

4.2.7 NAD⁺/NADH Detektionsassay

Zur Analyse der NAD⁺-Level wurden $1,0 \times 10^5$ der mESCs des Experiments entnommen. Diese wurden zweimal mit PBS gewaschen und in 50 µL PBS resuspendiert. Anschließend wurde nach der Herstelleranleitung des NAD⁺/NADH-GloTM Assays der Firma *Promega* verfahren. Die Luciferase Aktivität wurde an einem *Tecan GENios Pro Microplate Reader* in einer weißen 96-Well Platte visualisiert.

4.2.8 Sirtuin-Aktivitätassay

Der Assay zur Bestimmung der Aktivität des hSIRT1 (bereitgestellt von Prof. Dr. Dirk Schwarzer, Universität Tübingen) wurde nach dem Protokoll von Seidel *et al.* durchgeführt.^[281] Als Testpeptid wurde das acetylierte Substrat aus Seidel *et al.* in einer Konzentration von 50 μM mit 1 oder 5 μM des hSIRT1 für 1 h bei 37 °C inkubiert. Das, nach Abspaltung der Acetylgruppe und Trypsin-Verdau, erhaltene Fluoreszenzsignal wurde an einem Tecan GENios Pro Microplate Reader in einer schwarzen 96-Well Platte visualisiert. Zur Untersuchung der Aktivität des Sirtuins gegenüber eines fdC-enthaltenden doppelsträngigen Oligonukleotids (Tabelle 5) wurden 1500 pmol des jeweiligen Doppelstranges für 1 h oder über Nacht bei 37 °C mit 1 bzw. 5 μM hSIRT1 umgesetzt. Zur nachfolgenden massenspektrometrischen Analyse wurde das Oligonukleotid unter Verwendung des *Oligo Clean & Concentrator Kits* (Zymo Research) nach Herstellerangaben vorbereitet. Die Elution erfolgte in 30 μL *dd* H₂O. Die anschließende massenspektrometrische Analyse wurde von Dr. Katharina Iwan durchgeführt (siehe Abschnitt 4.2.9).

4.2.9 Massenspektrometrische Analyse

Die massenspektrometrische Analyse erfolgte, wie in der Dissertation Dr. Katharina Iwan beschrieben oder nach Abschnitt 3.1.1.^[65, 249]

4.3 Material und Methoden zu Abschnitt 3.2.3

4.3.1 Medien und Puffer

Siehe hierzu Abschnitt 4.4.1 des nachfolgenden Kapitels.

4.3.2 Zellbiologische Experimente

Die verwendete murine embryonale Stammzelllinie J1 entstammt der Linie 129/Sv/Jae.^[317] Die *Smug1*^{-/-} Zelllinie wurde auf Basis der J1 mESC-Linie im Rahmen meiner Masterarbeit generiert.^[297] Weiterhin wurden die *Neil*^{+/+}, *Neil1*^{-/-}, *Neil2*^{-/-} und die *Neil1,2-Doppelknockout* mESC-Linien von *Christof Niehrs* zur Verfügung gestellt. Die verwendeten *Kindlin3*^{+/+} Zellen wurden von *Reinhard Fässler* zur Verfügung gestellt.^[319] Auf Basis der *Kindlin3*^{+/+} mESCs wurden die *Tdg*^{-/-} und katalytische Mutante (CM) Stammzelllinie von mir hergestellt. Alle verwendeten mESCs wurden, wie im Abschnitt 3.2.1 oder 3.1.1 beschrieben, kultiviert. Andernfalls wird auf eine abweichende Kulturbedingung hingewiesen.

Für das Fütterungsexperiment wurden die Wildtyp und *Smug1*^{-/-} mESCs fünf Tage unter *Priming*-Bedingungen, bestehend aus Basalmedium und 1000 U/mL LIF, kultiviert. In den letzten 24 h des Experiments wurde das Medium ausgetauscht und die entsprechende Menge an hmdU (0, 1 und 2 µM) zugesetzt. Die Bestimmung der markierten fdU-Level wurde durch zweitägige Kultivierung der Zellen unter C/R-Bedingungen und gleichzeitiger Zugabe von 60 µM ¹³C₅¹⁵N₂-fdC erzielt. Zur Quantifizierung der BER-Intermediate wurden die *Neil-Knockout* mESCs unter C/R-Bedingungen, wie in Abschnitt 3.1.1 beschrieben, für drei Tage unter Zugabe von 100 µM ¹³C₉¹⁵N₃-rC kultiviert. Zur Bestimmung der natürlichen Cytidin-Modifikationen wurden die Zellen für fünf Tage unter Serum/LIF-Bedingungen kultiviert. Die Isolation der genomischen DNA wurde in Abschnitt 3.2.1 beschrieben.

Zur Bestimmung der globalen Level der höher oxidierten Cytidin-Spezies wurden die *Kindlin3*^{+/+} und die erzeugten *Tdg*^{-/-} und *Tdg*^{CM} mESCs für vier Tage unter Serum/LIF-Bedingungen kultiviert. Die genomische DNA wurde nach dem Protokoll aus Abschnitt 3.1.1 isoliert.

4.3.3 Massenspektrometrische Analyse

Die massenspektrometrische Analyse erfolgte analog der Vorschrift aus *Pfaffeneder et. al* bzw. *Schiesser et. al* oder wie in Abschnitt 3.2.1 beschrieben.^[142, 230]

4.3.4 Oligonukleotide

Die synthetischen Oligonukleotide (Tabelle 6) wurden von *Merck* oder *IDT* bezogen.

Tabelle 6: Genutzte Oligonukleotidsequenzen.

Bezeichnung	Sequenz	Verwendung
mTdg_KO gRNA FF	5'-CACCGTGCACCAGGTAGTTCTTACA-3'	guide RNA
mTdg_KO gRNA FR	5'-AAACTGTAAGAACTACCTGGTGCAC-3'	guide RNA
mTdg_KO gRNA RF	5'-CACCGTCCTGGACAGGAGCGTCTTC-3'	guide RNA
mTdg_KO gRNA RR	5'-AAACGAAGACGCTCCTGTCCAGGAC-3'	guide RNA
mTdg_CM gRNA F	5'-CACCGTTCTAGATTGGCATTAAACC-3'	guide RNA
mTdg_CM gRNA R	5'-AAACGGGTTAATGCCAATCTAGAAC-3'	guide RNA
Cas9(n) FWD	5'-ATGGACTATCATATGCTTACCGTAAC-3'	Sequenzierung
mTdg_KO FWD	5'-TTGTTCTTTCAGCTATTCTCTGGA-3'	PCR
mTdg_KO REV	5'-CCCACCCCTGTTGTTAC-3'	PCR
mTdg_CM FWD	5'-ATTAGCCTTACTTGGCAGGGG-3'	PCR
mTdg_CM REV	5'-ACTCCGCTATTACCAACGCT-3'	PCR
Tdg KO	5'-CCAGCAGTAGCTCCTAACATGGCAACCGT GACAGAACAGCAGGTGACTGTAATGACG CTCCTGTCCAGGATCCTGCACCAGGTAGTT CTTACATGATCTCCCTTGAATGTCTCCGCCT CCAGCAGAGCAATCAGTGAACAC-3'	Reparaturvorlage
Tdg CM	5'-GTCTGAGTTTTTAAAAACAAAAGTGTCA ATTACTTTGTTTTCTAGATTGGCATTGCCC CAGGATTAATGGCTGCTTACAAAGGACAT CACTACCCTGGGCCT-3'	Reparaturvorlage

4.3.5 Ligation des Expressionsvektors

Die Ligation des Expressionsvektors erfolgte wie in Abschnitt 4.4.5 des vorherigen Kapitels beschrieben.

4.3.6 Transfektion

Die mESCs wurden wie im Abschnitt 3.3 beschrieben unter a2i/LIF-Bedingungen kultiviert. Zur Transfektion wurden 6×10^5 Zellen in einer 6 Well Platte mit *Lipofectamine2000* (*Invitrogen*) transfiziert. Hierzu wurden 100 μ L Opti-MEM (Sigma) mit 8 μ L Lipofectamin gemischt. Weiterhin wurden im Falle der *Tdg*^{-/-} Generierung 100 μ L Opti-MEM mit 670 ng Cas9-Plasmid, 670 ng gRNA-F_Cas9n, 670 ng gRNA-R_Cas9n, 100 ng GFP-Plasmid und 5 μ L Reparaturtemplat (10 μ M) gemischt und zum Lipofectamin-Mix gegeben.

Für die katalytische Mutante wurde nach gleichem Prinzip verfahren und 2000 ng gRNA-Cas9, 100 ng GFP-Plasmid und 5 µL Reparaturtemplat (10 µM) verwendet. Die Gesamtmischung wurde 5 min bei Raumtemperatur inkubiert. Der Mix wurde in der Kulturschale vorgelegt und die Zellmischung hinzugegeben. Das Kulturmedium enthielt außerdem 1 µM des Inhibitors Scr7 (*Sigma*). Nach 24 h Inkubation wurde das Medium erneuert und für 48 h Kulturmedium unter Zusatz von 0,5 µg/mL Puromycin und Scr7 zugegeben. Nach der Selektion wurde eine *Limiting Dilution* zum Erhalt klonaler Zellpopulationen durchgeführt. Nach etwa 1,5 Wochen wurden Replikaklone kultiviert und auf Einbringung des Reparaturtemplats getestet.

4.3.7 Genotypisierung und Sequenzierung

Die Replikaklone wurden mittels *QuickExtraction* Puffer (*Epicentre*) nach Herstellerangaben lysiert. Anschließend wurde eine PCR zur Amplifikation des Ziellokus durchgeführt. Das PCR-Programm und Pipettierschema sind in Tabelle 7 aufgezeigt.

Tabelle 7: Pipettierschema und Programm der PCR zur Genotypisierung der Zellklone.

	Ausgangskonzentration	Menge [µL]	Programm
Zellysat	200 ng	x	95 °C 5 min
HotStar Taq MM (<i>Qiagen</i>)	-	10	95 °C 30 s
FWD Primer	10 µM	0,5	60 °C 20 s
REV Primer	10 µM	0,5	72 °C 10 s
dd H ₂ O		ad 20	72 °C 10 min
			4 °C ∞

← 34x

Nach erfolgreicher Amplifikation wurde das PCR-Amplifikat zur Genotypisierung mit dem jeweiligen Restriktionsenzym, wie in Tabelle 8 beschrieben, verdaut.

Tabelle 8: Pipettierschema und Programm des Restriktionsverdau zur Detektion der mutierten Klone.

	Ausgangskonzentration	Menge [µL]
Amplifiziertes Produkt	-	10
Cutsmart Puffer	10x	4
Enzym	10 µM	1
dd H ₂ O		ad 40

Zur Überprüfung der Einbringung des *Tdg*^{-/-} Reparaturtemplates wurde *BamHI* (*NEB*) verwendet und für das *Tdg*^{CM} Reparaturtemplat *SmaI* (*NEB*). Im Falle von *SmaI* wurde für 1,5 h bei 25 °C inkubiert und im Falle von *BamHI* wurde für 1,5h bei 37 °C inkubiert.

Die Proben wurden anschließend auf ein natives 10 %-iges TBE-PAGE (0,75 mm, 10 Well) geladen. Das Pipettierschema befindet sich in Tabelle 15. Hierzu wurden 25 µL des Reaktionsansatzes mit 5x Ladepuffer geladen und die Elektrophorese für 1 h bei 90 V durchgeführt. Zur Abschätzung der Bandengröße wurden 4 µL des 100 bp Marker (NEB) verwendet. Als Laufpuffer wurde 0,5x TBE Puffer verwendet. Zur Visualisierung wurde *SybrGreenI* (1:10000) in 0,5x TBE genutzt und das Gel darin für 15 min inkubiert. Die Banden wurden an der Geldokumentationskammer Las 3000 (*Fujifilm*) sichtbar gemacht. Alle Sequenzierungen wurden von *Eurofins Genomics* durchgeführt.

4.3.8 Western Blot

Zur Probenvorbereitung wurden $2,5 \times 10^5$ mESCs für zwei Tage unter Serum/LIF-Bedingungen (12 Well-Platte) kultiviert. Die Zellen wurden trypsinisiert und zweimal mit PBS gewaschen. Anschließend wurden die Zellen in 90 µL RIPA Puffer (10 mM Tris pH 7,4; 0,5 mM EDTA; 1% Triton-X; 1 % Deoxycholate; 150 mM NaCl; 2,5 mM MgCl₂; 1 mM DTT; 1x Protease Inhibitor Tablette (*Roche*)) lysiert. Die Lysate wurden mit Ladepuffer (50 mM Tris pH 6,8; 2 % (w/v) SDS; 10 % (v/v) Glycerin; 0,2 % (w/v) Bromphenolblau; 20% (v/v) β-Mercaptoethanol) kurz erhitzt. Die Proben (35 µL) wurden auf ein 10 %-iges Laemmli-Gel (1,5 mm; 10 well) geladen. Als Protein-Standard wurden 3 µL MagicMark XP Standard (*ThermoFisher*; LC5603) und 4 µL *Blue Prestained Protein Standard, Broad Range* (11-250 kDa) (*NEB*, P7706) genutzt. Die Gelelektrophorese wurde bei 120 V für 30 min und 200 V für 60 min in 1x SDS-Laufpuffer (25 mM Tris pH 8,9; 192 mM Glycin; 0,1 % SDS) durchgeführt. Zum Transfer wurde eine PVDF Membran (*GE Healthcare*) und vorgekühlter Towbin-Puffer (25 mM Tris; 192 mM Glycin; 20 % (v/v) Methanol; 0,05 % (w/v) SDS) benutzt. Die Membran wurde 30 s in Methanol aktiviert.

Das Whatman Filterpapier (*Sigma-Aldrich*) und das Laemmli-Gel (siehe Tabelle 19) wurden nach der Elektrophorese für 15 min in Towbin-Puffer äquilibriert. Der Transfer (*wet transfer*) wurde bei 4°C und 23 V über Nacht durchgeführt. Nach dem Transfer wurde die PVDF Membran für 1 h bei Raumtemperatur mittels 5 % (w/v) Milchpulver in TBS-T (20 mM Tris pH 7,4; 150 mM NaCl; 0,1 % (v/v) Tween-20) blockiert. Der primäre Antikörper (*Santa Cruz Biotechnology*, Anti-Tdg (D-11), sc-376652 (*mouse anti human*) bzw. Anti-Smug1 (*rabbit anti human*; bereitgestellt von *Geir Slupphaug*) wurde 1:200 bzw. 1:500 in 5 % (w/v) Milchpulver in TBS-T verdünnt.^[321, 322] Die Blockier-Lösung wurde abgenommen und der verdünnte primäre Antikörper bei 4 °C über Nacht inkubiert. Nach dieser Inkubationszeit wurde die Antikörper-Lösung verworfen und die Membran dreimal für jeweils 5 min mit TBS-T gewaschen. Der HRP-konjugierte sekundäre Antikörper (*Sigma-Aldrich*, A667 (*goat anti rabbit*) oder der sekundäre Antikörper (*Sigma-Aldrich*, A3682 (*goat anti mouse*)) wurden 1:10000 in 5 % (w/v) Milchpulver in TBS-T verdünnt und für eine Stunde bei Raumtemperatur inkubiert.

Im Anschluss wurde die Membran zweimal für jeweils 5 min mit TBS-T gewaschen und zweimal für jeweils 5 min mit TBS. Das *SuperSignal West Pico* Chemilumineszenz Substrat (*Thermo Scientific*; 34077) wurde zur Detektion verwendet.

4.4 Material und Methoden zu Abschnitt 3.3.1

4.4.1 Medien und Puffer

LB-Medium: Trypton 10 g, Hefeextrakt 5 g, Natriumchlorid 10 g; ad 1 L dH_2O ; Autoklavieren

LB-Agar: Zu einem Liter LB-Medium wird 15 g Agar gegeben und im Anschluss autoklaviert.

10x TBE Puffer: 0.89 M Tris-Base; 0.89 M Borsäure; 20 mM EDTA; pH 8,3

1x TBS Puffer: 20 mM Tris; 150 mM NaCl; pH 7,6; ad 1 L ddH_2O

0,5 M Tris-HCl: 60,6 g Tris; pH 6,8; ad 1 L ddH_2O

1,5 M Tris-HCl: 181,71 g Tris; pH 8,8; ad 1 L ddH_2O

Gel-Puffer: 3 M Tris pH 8,45; 0,3 % SDS

Kathoden-Puffer: 0,1 M Tris; 0,1 M Tricin; 0,1 % SDS; pH 8,25 (pH nicht einstellen)

Anoden-Puffer: 0,2 M Tris pH 8,9

5x Nicht-Reduzierender Ladepuffer: 250 mM Tris pH 6,8; 10 % (w/v) SDS; 50 % (v/v) Glycerin; 0,1 % (w/v) Bromphenolblau; 4x reduzierenden Ladepuffer: Zugabe von 20 % (v/v) β -Mercaptoethanol

1x RIPA Puffer: 10 mM Tris pH 7,4; 0,5 mM EDTA; 1% Triton-X; 1 % Deoxycholat; 150 mM NaCl; 2,5 mM $MgCl_2$; 1 mM DTT; 1x Protease Inhibitor Tablette (*Roche*)

1x Cell Lysis Puffer: 10 mM Tris pH 7,4; 1 mM EDTA; 1 % Triton-X; 100 mM NaCl; 1x Protease Inhibitor Tablette (*Roche*)

1x SDS-Laufpuffer: 25 mM Tris pH 8,9; 192 mM Glycin; 0,1 % SDS

1x Towbin-Puffer: 25 mM Tris; 192 mM Glycin

4.4.2 Zelllinien

Die verwendete murine embryonale Stammzelllinie J1 entstammt der Linie 129/Sv/Jae.^[317] Die Dnmt-TKO mESCs wurden von *Masaki Okano (RIKEN Center for Developmental Biology)* bereitgestellt.^[112] Die *Cda^{-/-} Dctd^{-/-}* Zelllinie wurde auf Basis der mESC-Linie J1 im Rahmen meiner Masterarbeit generiert.^[297] Die *Mbd4^{-/-}* mESCs wurden von *Adrian Bird (Wellcome Centre for Cell Biology - Edinburgh)* zur Verfügung gestellt.^[316] Weiterhin wurde die *Mlh1^{-/-}* mESC Linie auf Grundlage der IB10-mESCs (Subklon der E14-mESCs) von mir im Rahmen der Doktorarbeit generiert.^[320] Die IB10-Linie wurde von *Hein te Riele (Netherlands Cancer Institute)* zur Verfügung gestellt. Die *Tdg^{+/-}* und *Tdg^{-/-}* Zelllinie wurde von *Primo Schär (Universität Basel)* bezogen.^[253, 254] Diese *Tdg^{-/-}* Zelllinie wurde als Parentallinie zur Herstellung der *Tdg^{-/-} Mbd4^{-/-}* mESC-Linie genutzt.

Die verwendeten *Kindlin3^{+/+}* Zellen wurden von *Reinhard Fässler* zur Verfügung gestellt.^[319] Die Herstellung der verwendeten *Tdg^{-/-}* und *Tdg^{CM}* mESCs wird in Abschnitt 3.2.3 beschrieben. Alle verwendeten mESCs wurden, wie in den Abschnitten 3.2.1 und 3.1.1 beschrieben, entweder unter naiven 2i/LIF oder a2i/LIF Bedingungen kultiviert. Für die 2i-Bedingung wurde das beschriebene Basalmedium mit 1000 U/mL LIF (*ORF Genetics*), 3 μ M CHIR99021 und 1 μ M PD0325901 (beide von *Axon Medchem*) versetzt.^[34] Für die a2i-Bedingung wurde das Medium wie für die 2i-Bedingung hergestellt aber mit dem SRC-Inhibitor CGP77675 (1,5 μ M *Axon Medchem*) anstatt PD0325901 versetzt.^[59] Zur Kultivierung der Zellen unter Serum/LIF *Priming*-Bedingungen wurden die beiden Inhibitoren CHIR99021 und PD0325901 weggelassen. Das CHIR/IWR1 (C/R) *Priming* verwendet Basalmedium mit 3 μ M CHIR99021 und 2,5 μ M des *Tankyrase*-Inhibitors IWR1-endo.^[63]

4.4.3 Zellbiologische Experimente

Zur Untersuchung der natürlichen mdC und hmdC Modifikationen in *Kindlin3^{+/+}* mESCs wurden die Zellen für zwei Passagen unter naiven 2i/LIF oder a2i/LIF Bedingungen kultiviert. Anschließend wurden die Zellen unter C/R oder Serum/LIF *Priming*-Bedingungen kultiviert und nach 24, 48 oder 72 h geerntet. Zur Markierung der Zellen mit ¹³CD₃-L-Methionin wurden die Zellen wie in *Pfaffeneder et al.* beschrieben in Methionin-freiem Medium unter Zusatz von natürlichem oder isotopenmarkiertem Methionin kultiviert.^[142] Für das *Pulse-Chase* Experiment in *Kindlin3^{+/+}* mESCs wurden die Zellen für fünf Tage unter a2i/LIF oder für zwei Tage unter a2i/LIF und anschließend drei Tage unter C/R-Bedingungen in Medium mit markiertem Methionin kultiviert. Am Tag vier in a2i/LIF-Bedingungen wurde zusätzlich 5 μ M ¹³C₁₀¹⁵N₂-dT zur Verfolgung der DNA-Synthese zugesetzt. Nach Passagieren der Zellen am fünften Tag wurden alle zugesetzten markierten Nukleoside durch Wechsel des Mediums entfernt und die Zellen zu den jeweiligen Zeitpunkten geerntet.

Für die Verfolgung der markierten Verbindungen in Wildtyp und *Cda^{-/-} Dctd^{-/-}* mESCs wurden die Zellen für fünf Tage unter C/R-Bedingungen mit 0,2 mM ¹³CD₃-L-Methionin (*Sigma-Aldrich*) im Kulturmedium und Zugabe von 100 μ M ¹³C₉¹⁵N₃-dC sowie 7,5 μ M ¹³C₁₀¹⁵N₅-dG (beide bezogen von *Silantes*) kultiviert.

Zur Untersuchung der Änderung der mdC-Level in den BER und MMR *Knockout* mESCs wurden die Zellen für fünf Tage in C/R *Priming*-Medium mit markiertem ¹³CD₃-L-Methionin kultiviert. Die Kulturen wurden jeweils nach drei Tagen passagiert und für weitere 48 h wachsen gelassen.

4.4.4 Oligonukleotide

Die synthetischen Oligonukleotide (Tabelle 9) wurden von *Merck* oder *IDT* bezogen.

Tabelle 9: Verwendete Oligonukleotidsequenzen.

Bezeichnung	Sequenz	Verwendung
mMlh1_KO gRNA FF	5'-CACCGACGGTAGTGAACCGCATAG-3'	guide RNA
mMlh1_KO gRNA FR	5'-AAACCTATGCGGTTCACTACCGTC-3'	guide RNA
mMlh1_KO gRNA RF	5'-CACCGTCCTGCTACAAACGCCATAT-3'	guide RNA
mMlh1_KO gRNA RR	5'-AAACATATGGCGTTTGTAGCAGGAC-3'	guide RNA
mTdg_Mbd4_DKO gRNA FF	5'-CACCGCCTAGTTCAGACCCGCCTT-3'	guide RNA
mTdg_Mbd4_DKO gRNA FR	5'-AAACAAGGCGGGTCTGGAAGTGGC-3'	guide RNA
mTdg_Mbd4_DKO gRNA RF	5'-CACCGACTCTGTTGTCCCAAGGTT-3'	guide RNA
mTdg_Mbd4_DKO gRNA RR	5'-AAACAACCTTGGGGACAACAGAGTC-3'	guide RNA
Cas9(n) FWD	5'-ATGGACTATCATATGCTTACCGTAAC-3'	Sequenzierung
mMlh1_KO FWD	5'-GAAGAGCGGACCGTGAACCTT-3'	PCR
mMlh1_KO REV	5'-CAAGTCGTGAGGACCTTCGG-3'	PCR
mMd4_KO FWD	5'-GGCATCTGTAGCTGAAGCGG-3'	PCR
mMd4_KO REV	5'-GTGCATCATTTTAGCCCGGA-3'	PCR
mMbd4_F	5'-TATGGCAACGACTCCTACCG-3'	qPCR
mMbd4_R	5'-CACCACAGGGCCCTTAGAAA-3'	qPCR
mCda_F	5'-AAGGCCATCTCCGAAGGGTA-3'	RT-PCR
mCda_R	5'-TCTTCAGGTCCAAACGAGGC-3'	RT-PCR
mGapdh_F	5'-CATGGCCTCCGTGTTCTA-3'	qPCR ^[323]
mGapdh_R	5'-CTTCACCACCTTCTTGATGTCATC-3'	qPCR ^[323]
mMlh1_KO	5'-GAACTTGAGCGTGAGGAGCTCGAGTGATT GGCTGACTGGGAACTCGGGCGCCAATATGGCG TTTGTAGCAGGAGTTATCGGCGTCTAGACGAGA CGGTAGTGAACCGCATAGCGGCGGGGAAGTC ATTCAGCGCCGCCAATGCTATCAAAGAGATG ATAGAA-3'	Reparaturvorlage
mMbd4_KO	5'-CTGGAGCTGCAGCTGCGCCCCAGTGGACCAA GCAGCAGGGATGGAGAGCCCAAACCTTGGGGAC AACAGAGTCCTGCAGAGAGCCTAGTTCAGACCC GCCTTGGGATCGCTGGCAAGTTATTAACCTCTCT CGGGCCTCAGAGGAGTCGCTCTT-3'	Reparaturvorlage

4.4.5 Ligation des Expressionsvektors

Zur Klonierung der gRNA in den jeweiligen pSpCas9(n)-2A-Puro Expressionsvektor (*Addgene* Plasmide pSpCas9-2A-Puro PX 459 (*Addgene* Plasmid 48139) und pSpCas9n-2A-Puro PX 462 (*Addgene* Plasmid 48141)) wurden die synthetisierten Oligonukleotide zunächst am 5'-Ende phosphoryliert. Mit Hilfe des Enzyms T4-Polynukleotidkinase (*NEB*), welche die γ -Phosphatgruppe von ATP an das 5'-Ende der DNA überträgt, lassen sich fehlende Phosphatgruppen anhängen.

Gleichzeitig wird ein Temperaturgradient gefahren, um ein *Annealing* der Einzelstränge zu gewährleisten. Tabelle 10 zeigt das verwendete Pipettierschema und Programm der Phosphorylierungs- und Hybridisierungsreaktion.

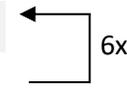
Tabelle 10: Pipettierschema und Programm der Phosphorylierungs- und *Annealing*-Reaktion der gRNA Oligonukleotide.

	Ausgangskonzentration	Menge [μL]	Programm
gRNA (<i>forward</i>)	100 μM	1	37 °C 10 min
gRNA (<i>reverse</i>)	100 μM	1	95 °C 5 min
T4 Ligase Puffer	10x	1	ramp down 5 °C/min
T4 PNK	10.000 U/mL	1	25 °C ∞
dd H₂O	-	6	

Die doppelsträngigen, phosphorylierten Oligonukleotide wurden 1:200 mit ddH₂O verdünnt (verd.) und über die *Bpil*-Restriktionsstelle in den Cas9-Vektor kloniert. Der Vektor wurde zunächst an der Restriktionsschnittstelle geöffnet und das doppelsträngige Oligonukleotid aufgrund ungleicher Enden directional mittels Ligase in den Vektor kloniert. Pipettierschema und Programm der Reaktion sind in Tabelle 11 abgebildet.

Tabelle 11: Pipettierschema und Programm des Restriktionsverdau mit *Bpil* (*Thermo Scientific*) und Ligation der gRNA in den Vektor.

	Ausgangskonzentration	Menge [μL]	Programm
pSpCas9(n)-2A-Puro	100 ng/ μL	1	37 °C 5 min
verd. Oligo-Duplex		2	21 °C 5 min
Tango Puffer	10x	2	
DTT	10 mM	1	
ATP	10 mM	1	
Bpil	5000 U/mL	1	
T4 Ligase	400.000 U/mL	1	
ddH₂O		11	



Anschließend wurde der ligierte Vektor mit der *PlasmidSafe Exonuklease* (Epicentre) verdaut. In Tabelle 12 sind das verwendete Pipettierschema und Programm gezeigt.

Tabelle 12: Pipettierschema und Programm der Behandlung des Ligationsansatzes mit Exonuklease.

	Ausgangskonzentration	Menge [μ l]	Programm
Ligationsreaktion		11	37 °C 10 min
<i>PlasmidSafe</i> Puffer	10x	1,5	95 °C 5 min
<i>PlasmidSafe</i> Exonuklease		1	
ATP	10 mM	1,5	

Der ligierte pSpCas9-Vektor wurde anschließend in chemisch kompetente 10-beta Zellen (*NEB*) nach Herstellerangaben eingebracht. Die Bakterien wurden auf *Lysogeny broth* (LB)-Agar Festmedium ausplattiert und über Nacht bei 37 °C inkubiert. Weiterhin enthielten die Platten zur Selektion positiver Kolonien 100 μ g/mL des Antibiotikums Carbenicillin. Zur Vervielfältigung des Expressionsvektors wurden Flüssigkulturen der gewachsenen Bakterienstämme in LB-Medium mit 100 μ g/mL Carbenicillin bei 260 rpm und 37 °C über Nacht kultiviert. Die Plasmid-DNA wurde unter Zuhilfenahme des *GeneJET Miniprep Kit* (*Thermo Scientific*) nach Herstellerangaben isoliert.

4.4.6 Transfektion

Die mESCs wurden wie im Abschnitt 3.2.13.3 beschrieben unter 2i/LIF-Bedingungen kultiviert. Zur Transfektion wurden 6×10^5 Zellen in einer 6 well Platte mit *Lipofectamine2000* (*Invitrogen*) transfiziert. Hierzu wurden 100 μ L Opti-MEM (*Sigma-Aldrich*) mit 8 μ L Lipofectamin gemischt. Weiterhin wurden im Falle der *Tdg^{-/-} Mbd4^{-/-}* Generierung 100 μ L Opti-MEM mit 670 ng Cas9-Plasmid, 670 ng gRNA-F_Cas9n, 670 ng gRNA-R_Cas9n, 100 ng GFP-Plasmid und 3 μ L Reparaturtemplat (10 μ M) gemischt und zum Lipofectamin-Mix gegeben. Bei der Generierung des *Mlh1^{-/-}* wurde analog verfahren, mit der Abwandlung, dass das Cas9-Plasmid nicht kotransfiziert wurde. Die Gesamtmischung wurde 5 min bei Raumtemperatur inkubiert. Der Mix wurde in der Kulturschale vorgelegt und die Zellmischung hinzugegeben. Das Kulturmedium enthielt außerdem 1 μ M des Inhibitors Scr7 (*Sigma*). Nach 24 h Inkubation wurde das Medium erneuert und für 48 h Kulturmedium unter Zusatz von 1,0 μ g/mL Puromycin und Scr7 zugegeben. Nach der Selektion wurde eine Grenzverdünnungsklonierung zum Erhalt klonaler Zellpopulationen durchgeführt. Nach etwa 1,5 Wochen wurden Replikaklone kultiviert und auf Einbringung des Reparaturtemplats getestet.

4.4.7 Genotypisierung und Sequenzierung

Die Replikklone wurden mittels *QuickExtraction* Puffer (*Epicentre*) nach Herstellerangaben lysiert. Anschließend wurde eine PCR zur Amplifikation des Ziellokus durchgeführt. Das PCR-Programm und Pipettierschema sind in Tabelle 13 aufgezeigt.

Tabelle 13: Pipettierschema und Programm der PCR zur Genotypisierung der Zellklone.

	Ausgangskonzentration	Menge [μ L]	Programm
Zelllysat	400 ng	x	95 °C 5 min
Taq Polymerase (NEB)	5000 U/mL	0,1	95 °C 30 s
FWD Primer	10 μ M	0,5	60 °C 30 s
REV Primer	10 μ M	0,5	68 °C 40 s
DMSO	100 %	1,0	68 °C 10 min
Polymerase-Puffer (NEB)	10x	2,5	4 °C ∞
ddH₂O		ad 25	

← 40x

Nach erfolgter Amplifikation wurde das PCR-Produkt zur Genotypisierung mit dem jeweiligen Restriktionsenzym, wie in Tabelle 14 beschrieben, verdaut.

Tabelle 14: Pipettierschema und Programm des Restriktionsverdau zur Detektion der mutierten Klone.

	Ausgangskonzentration	Menge [μ L]
Amplifiziertes Produkt	-	25
Cutsmart Puffer	10x	3,0
Enzym	10 μ M	0,5
ddH₂O		ad 30

Zur Überprüfung der Einbringung des *Mlh1*^{-/-} Reparaturtemplates wurde *XbaI* (*NEB*) verwendet und für das *Tdg*^{-/-} *Mbd4*^{-/-} Reparaturtemplat *PstI* (*NEB*). Beide Reaktionsansätze wurden für 1,5 h bei 37 °C inkubiert. Die Proben wurden anschließend auf ein natives 10 %-iges TBE-PAGE-Gel (0,75 mm, 10 Well) geladen. Das Pipettierschema befindet sich in Tabelle 15. Hierzu wurden 15 μ L des Reaktionsansatzes mit 5x Ladepuffer geladen und die Elektrophorese für 1 h bei 120 V durchgeführt. Zur Abschätzung der Bandengröße wurden 4 μ L des 100 bp Marker (*NEB*) verwendet. Als Laufpuffer wurde 0,5x TBE Puffer verwendet. Zur Visualisierung wurde SybrGreenI (1:10000) in 0,5x TBE genutzt und das Gel darin für 15 min inkubiert. Die Banden wurden an der Geldokumentationskammer Las 3000 (*Fujifilm*) sichtbar gemacht.

Tabelle 15: Pipettierschema des verwendeten nativen TBE-Gels.

	Menge
30 % Acrylamid/Bisacryl.	6,67 mL
10 % TBE-Puffer	2 mL
TEMED	20 µL
10 % APS	200 µL
ddH₂O	ad 20 mL

Alle Sequenzierungen wurden von *Eurofins Genomics* durchgeführt.

4.4.8 Reverse Transkriptase PCR

Nach Isolierung der Gesamt-RNA wurde die mRNA mit Hilfe des *iScript™ cDNA Synthesis Kits* (BioRad) in cDNA überschrieben. Dieses Kit enthält hierzu oligo(dT) und zufällige Hexamer-Primer, sowie eine Reverse Transkriptase. Das Pipettierschema und das genutzte *Thermocycler*-Programm sind in Tabelle 16 dargestellt.

Tabelle 16: Pipettierschema und Programm der cDNA Synthese.

Ansatz	Menge [µL]	Programm
Total RNA (500 ng)		25 °C 5 min
<i>iScript</i> Reverse Transkriptase	1,0	46 °C 20 min
5x <i>iScript Reaction Mix</i>	4,0	95 °C 1 min
ddH ₂ O	ad 20	4 °C ∞

Die erhaltene cDNA (25 ng/µL) wurde im Anschluss zur RT-PCR Reaktion genutzt. Das verwendete Primer-Paar ist in Tabelle 9 dargestellt. Das Pipettierschema und Programm der RT-PCR sind in Tabelle 17 beschrieben.

Tabelle 17: Pipettierschema und Programm der RT-PCR Reaktion.

Ansatz	Menge [μL]	PCR Programm
cDNA	1,5	95 °C 5 min
Taq Polymerase (NEB)	0,2	95 °C 30 s
dNTPs (10 mM)	1,0	60 °C 30 s
DMSO	2,5	68 °C 40 s
Forward Primer (10 μM)	1,0	68 °C 5 min
Reverse Primer (10 μM)	1,0	4 °C ∞
10x ThermoPol Puffer	5	
ddH ₂ O	ad 50	

← 34x

Im Anschluss wurden die Proben (jeweils 25 μL) mit Ladepuffer auf ein 8%-iges Polyacrylamidgel geladen und die Elektrophorese bei 100 V für 40 min durchgeführt. Anschließend wurde das Gel mit Sybr-Green I (1:10000; *Invitrogen*) gefärbt und detektiert.

4.4.9 Western Blot

Zur Proben-Vorbereitung wurden 4×10^6 mES-Zellen in 100 μL *Cell Lysis Puffer* lysiert. Die Lysate wurden mit reduzierendem Ladepuffer (Abschnitt 4.4.1) kurz erhitzt. Die Proben (20 μL) und gereinigtes hDctd (1: 12500; von *Gladys Maley* zur Verfügung gestellt) wurden auf ein 10%-iges Tricin-Gel (1 mm; 10 Well; Tabelle 18) geladen. Zur Verifizierung des *Mlh1-Knockouts* wurden die Zellen in 90 μL RIPA-Puffer lysiert. Die *Mlh1-Knockout* Proben (35 μL) wurden auf ein 12 %-iges Laemmli-Gel (1,5 mm; 10 Well; Tabelle 19) geladen. Als Protein-Standard wurden 3 μL MagicMark XP Standard (*ThermoFisher* LC5603) und 4 μL *Blue Prestained Protein Standard, Broad Range* (11-250 kDa) (*NEB*, P7706) genutzt. Die Gelelektrophorese wurde bei 60 V für 20 min und 110 V für 50 min in Anoden-Puffer und Kathoden-Puffer durchgeführt. Die Gelelektrophorese der *Mlh1*-Proben wurde bei 20 mA für 3 h in 1x SDS-Laufpuffer durchgeführt. Zum Transfer wurde eine PVDF Membran (*GE Healthcare*) und vorgekühlter Towbin-Puffer mit 20% (v/v) Methanol und 0,1% (w/v) SDS benutzt. Im Falle der *Mlh1*-Proben wurde 20 % Methanol und 0,05 % SDS (w/v) verwendet. Die Membran wurde 30 s in Methanol aktiviert. Das Whatman-Filterpapier (*Sigma-Aldrich*) und das Tricin-Gel wurden nach der Elektrophorese für 15 min in Towbin-Puffer äquilibriert. Der Transfer (*wet transfer*) wurde bei 4 °C und 25 V über Nacht durchgeführt. Nach dem Transfer wurde die PVDF-Membran für 1 h bei Raumtemperatur mittels 5% (w/v) Milchpulver in TBS-T (1xTBS mit 1 % (v/v) Tween-20) blockiert.

Der primäre Antikörper Anti-hDctd (*rabbit anti human*; bereitgestellt von Gladys Maley) wurde 1:300 in 5 mL 5 % (w/v) Milchpulver in TBS-T verdünnt.^[324, 325] Zur Detektion des Mlh1-Proteins wurde der primäre Antikörper Anti-Mlh1 (*rabbit anti human*; C-20; sc-582; Santa Cruz Biotechnology) 1:400 in 5 mL 5 % (w/v) Milchpulver in TBS-T verdünnt. Die Blockier-Lösung wurde abgenommen und der verdünnte primäre Antikörper bei 4°C über Nacht inkubiert. Nach dieser Inkubationszeit wurde die Antikörper-Lösung verworfen und die Membran dreimal für jeweils 5 min mit TBS-T gewaschen. Der HRP-konjugierte sekundäre Antikörper (*Sigma-Aldrich*, A667 (*goat anti rabbit*)) wurde 1:10000 in 5 mL 5 % (w/v) Milchpulver in TBS-T verdünnt und für eine Stunde bei Raumtemperatur inkubiert. Im Anschluss wurde die Membran zweimal für jeweils 5 min mit TBS-T gewaschen und zweimal für jeweils 5 min mit TBS. Das *SuperSignal West Pico* Chemilumineszenz Substrat (*Thermo Scientific* 34077) wurde zur Detektion verwendet.

Tabelle 18: Pipettierschema des verwendeten Tricin-Gels.

	Sammelgel (4 %)	Trenngel (10 %)
30 % Acrylamid/Bisacryl.	4,0 mL	10 mL
Gelpuffer	7,5 mL	10 mL
87 % Glycerin	-	3,2 mL
ddH₂O	ad 30 mL	ad 30 mL

Tabelle 19: Pipettierschema des verwendeten Laemmli-Gels.

	Sammelgel (4 %)	Trenngel (10 %)
30 % Acrylamid/Bisacryl.	1,98 mL	5,00mL
0,5 M Tris-HCl pH 6,8	3,78 mL	-
1,5 M Tris-HCl pH 8,8	-	3,75 mL
10 % SDS	150 µL	150 µL
TEMED	15 µL	15 µL
10 % APS	150 µL	150 µL
ddH₂O	ad 15 mL	ad 15 mL

4.4.10 Isolation der genomischen DNA

Die genomische DNA der Zellen wurde, wie im Abschnitt 3.1.1 beschrieben, isoliert.

4.4.11 Massenspektrometrische Analyse

Die massenspektrometrische Analyse erfolgte wie in den Dissertationen von *Dr. Jessica Steinbacher* und *Dr. Olesya Kosmatchev* beschrieben oder wurde von *Sarah Schiffers* nach ihrer optimierten Methode durchgeführt.^[249, 300]

5 Abkürzungsverzeichnis

-/-	<i>Knockout</i> beider Allele
°C	Grad Celsius
+/-	Knockout eines Allels
2i	<i>Two inhibitors</i>
4mdC	4-Methyl-2'-desoxycytidin
6mdA	6-Methyl-2'-desoxyadenosin
8oxodG	8-Oxo-2'-desoxyguanosin
ACMSD	<i>α-Amino-β-carboxymuconate-ε-semialdehyde decarboxylase</i>
Aid/AID	<i>Activation-induced deaminase</i>
AML	<i>Akute Myeloische Leukämie</i>
AP-Stelle	<i>Apurinic/Apyrimidinic-Stelle</i>
Ape1	<i>Apurinic/Apyrimidinic-Endonuclease 1</i>
Apobec	<i>Apolipoprotein B mRNA-editing enzyme catalytic polypeptide</i>
APS	Ammoniumpersulfat
ATP	Adenosintriphosphat
BER	Basenexzisionsreparatur
BMP4	<i>Bone Morphogenic Protein 4</i>
bp	Basenpaar
C/R	CHIR/IWR1
cadC	5-Carboxy-2'-desoxycytidin
Cas9n	Cas9 <i>Nickase</i>
caU	5-Carboxyluracil
Cda	Cytosin-Desaminase
cDNA	<i>complementary DNA</i>
CHIR	CHIR99021
CHO-K1	<i>Chinese Hamster Ovary K1</i>
CpG	dC/dG-Dinukleotid
CRISPR	<i>Clustered Regularly Interspaced Palindromic Repeats System</i>
dA	2'-Desoxyadenosin
dC	2'-Desoxycytidin
dCDP	2'-Desoxycytidin-5'-diphosphat
dCMP	2'-Desoxycytidin-5'-monophosphat
Dctd	Deoxycytidylat-Desaminase
ddPCR	<i>droplet digital PCR</i>
dG	2'-Desoxyguanosin
dN	2'-Desoxynukleosid
DNA	Desoxyribonukleinsäure (<i>Deoxyribonucleic Acid</i>)
Dnmt	DNA-Methyltransferase
dNTP	2'-Desoxynukleosidtriphosphat
dT	2'-Desoxythymidin
DTT	Dithiothreitol
dTTP	2'-Desoxycytidin-5'-triphosphat

dU	2'-Desoxyuridin
dUMP	2'-Desoxyuridin-5'-monophosphat
<i>E. coli</i>	<i>Escherichia Coli</i>
ECC	<i>Embryonic Carcinoma Cell</i>
EDTA	Ethylendiamintetraessigsäure
EGC	<i>Embryonic Germ Cell</i>
EpILC	<i>Epiblast-Like Cell</i>
EpISC	<i>Epiblast Stem Cell</i>
ERK	<i>Extracellular signal-Regulated Kinase</i>
ESC	<i>Embryonic Stem Cell</i>
F-dC	2'-(R)-Fluorodesoxycytidin
F-dU	2'-(R)-Fluorodesoxyuridin
F-fdC	2'-(R)-Fluoro-5-formyldesoxycytidin
F-mdC	2'-(R)-Fluoro-5-methyldesoxycytidin
FBS	Fötale Kälberserum
fC	5-Formylcytosin
fdC	5-Formyl-2'-desoxycytidin
fdU	5-Formyl-2'-desoxyuridin
FGF	<i>Fibroblast Growth Factor</i>
fU	5-Formyluracil
g	Erdbeschleunigung
gRNA	<i>guide RNA</i>
H	Histidin
HAT	Histonacetyltransferase
HDAC	Histondeacetylase
HDR	<i>Homology Directed Repair</i>
HEK	<i>Human Embryonic Kidney</i>
hmdC	5-Hydroxymethyl-2'-desoxycytidin
hmdU	5-Hydroxymethyl-2'-desoxyuridin
HPLC	<i>High Performance Liquid Chromatography</i>
ICM	<i>Inner Cell Mass</i>
ICR	<i>Imprinting Control Region</i>
IDAX	<i>Inhibition of the Dcl and AXin complex</i>
IDCase	Isoorotat-Decarboxylase
iPSCs	<i>Induced Pluripotent Stem Cells</i>
JAK	<i>Janus Kinase</i>
JBP	J-bindendes Proteine
K	Lysin
Kap.	Kapitel
kDa	Kilodalton
KDM	Lysin-Demethylase
KMT	Lysin-Methyltransferase
KO	<i>Knockout</i>
LIF	<i>Leukemia Inhibitory Factor</i>
LLOQ	<i>Lower Limit of Quantification</i>
LOD	<i>Limit of Detection</i>

Mbd4	<i>Methyl-CpG-binding domain protein 4</i>
mdC	5-Methyl-2'-desoxycytidin
mdU	5-Methyl-2'-desoxyuridin
MeCN	Acetonitril
MEK	<i>Mitogen-activated Protein Kinase Kinase</i>
MeOH	Methanol
mESC	<i>murine Embryonic Stem Cells</i>
mL	Milliliter
Mlh1	<i>MutL homolog 1</i>
MLL	<i>Mixed Lineage Leukemia</i>
MMR	<i>Mismatch-Reparatur</i>
mRNA	<i>messenger RNA</i>
MS	Massenspektrometrie
MS/MS	Tandemmassenspektrometrie
N	Anzahl
NAD ⁺	Nicotinamidadenindinukleotid
NADH	Nicotinamidadenindinukleotid-Hydrogen
NAM	Nicotinamid
NEB	<i>New England Biolabs</i>
Neil	<i>Nei endonuclease VIII-like</i>
NHEJ	<i>Non-Homologous End Joining</i>
Oct4	Octamer-bindender Transkriptionsfaktor 4
PAGE	Polyacrylamidgelelektrophorese
PARP	<i>Poly-ADP-Ribose Polymerase</i>
PBS	<i>Phosphate-Buffered Saline</i>
PCNA	<i>Proliferating-Cell-Nuclear-Antigen</i>
PCR	<i>Polymerase-Kettenreaktion</i>
PD	PD0325901
PGC	<i>Primordial Germ Cell</i>
PNK	Polynukleotidkinase
R	<i>Arginin</i>
RFU	<i>Relative Fluorescence Units</i>
RNA	Ribonukleinsäure
RNAP II	RNA-Polymerase II
ROS	<i>Reactive Oxygen Species</i>
rRNA	ribosomale RNA
RT-PCR	<i>Reverse Transcriptase-PCR</i>
RT-qPCR	<i>Real-Time quantitative PCR</i>
S/L	Serum/LIF
SAM	S-Adenosyl-L-Methionin
SDS	Sodiumdodecylsulfat
Sirt	Sirtuin
Smug1	<i>Single-strand selective monofunctional uracil DNA glycosylase 1</i>
STAT3	<i>Signal Transducer and Activator of Transcription 3</i>
T	Tag
T7H	<i>Thymine-7-Hydroxylase</i>

TAE	TRIS-Acetat-EDTA
Tdg	Thymin-DNA-Glykosylase
TEMED	<i>N,N,N',N'</i> -Tetramethylethyldiamin
Tet	<i>Ten-eleven translocation</i> Enzym
TKO	<i>Triple Knockout</i>
Ts	Thymidylatsynthase
TSA	Trichostatin A
U	Uracil
UHPLC	<i>Ultrahigh Performance Liquid Chromatography</i>
Uhrf1	<i>Ubiquitin-like plant homeodomain and ring finger domain1</i>
ULOQ	<i>Upper Limit of Quantification</i>
v/v	<i>volume per volume</i>
w/v	<i>weight per volume</i>
α KG	α -Ketoglutarat
β -EI	β -Eliminierungsintermediat
μ L	Mikroliter

6 Literaturverzeichnis

- [1] P. W. Hill, R. Amouroux, P. Hajkova, *DNA demethylation, Tet proteins and 5-hydroxymethylcytosine in epigenetic reprogramming: an emerging complex story*, *Genomics* **2014**, *104*, 324-333.
- [2] A. H. Wong, Gottesman, II, A. Petronis, *Phenotypic differences in genetically identical organisms: the epigenetic perspective*, *Hum. Mol. Genet.* **2005**, *14*, R11-18.
- [3] M. F. Fraga, E. Ballestar, M. F. Paz, S. Ropero, F. Setien, M. L. Ballestar, D. Heine-Suner, J. C. Cigudosa, M. Urioste, J. Benitez, M. Boix-Chornet, A. Sanchez-Aguilera, C. Ling, E. Carlsson, P. Poulsen, A. Vaag, Z. Stephan, T. D. Spector, Y. Z. Wu, C. Plass, *et al.*, *Epigenetic differences arise during the lifetime of monozygotic twins*, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 10604-10609.
- [4] L. Holtzman und C. A. Gersbach, *Editing the Epigenome: Reshaping the Genomic Landscape*, *Annu. Rev. Genomics Hum. Genet.* **2018**, *19*, 43-71.
- [5] C. H. Waddington, *The epigenotype. 1942*, *Int. J. Epidemiol.* **2012**, *41*, 10-13.
- [6] C. H. Waddington, *The strategy of the genes: a discussion of some aspects of theoretical biology*, George Allen & Unwin, London, **1957**.
- [7] P. A. Jeggo und R. Holliday, *Azacytidine-induced reactivation of a DNA repair gene in Chinese hamster ovary cells*, *Mol. Cell* **1986**, *6*, 2944-2949.
- [8] V. E. A. Russo, R. A. Martienssen, A. D. Riggs, *Epigenetic mechanisms of gene regulation*, Cold Spring Harbor Laboratory Press, Plainview, New York, **1996**.
- [9] A. Bird, *Perceptions of epigenetics*, *Nature* **2007**, *447*, 396-398.
- [10] Z. Chen, S. Li, S. Subramaniam, J. Y.-J. Shyy, S. Chien, *Epigenetic Regulation: A New Frontier for Biomedical Engineers*, *Annu. Rev. Biomed. Eng.* **2017**, *19*, 195-219.
- [11] D. Srivastava und N. DeWitt, *In Vivo Cellular Reprogramming: The Next Generation*, *Cell* **2016**, *166*, 1386-1396.
- [12] J. B. Gurdon, *The developmental capacity of nuclei taken from intestinal epithelium cells of feeding tadpoles*, *J. Embryol. Exp. Morphol.* **1962**, *10*, 622-640.
- [13] I. Wilmut, A. E. Schnieke, J. McWhir, A. J. Kind, K. H. Campbell, *Viable offspring derived from fetal and adult mammalian cells*, *Nature* **1997**, *385*, 810-813.
- [14] K. Takahashi und S. Yamanaka, *Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors*, *Cell* **2006**, *126*, 663-676.
- [15] R. L. Davis, H. Weintraub, A. B. Lassar, *Expression of a single transfected cDNA converts fibroblasts to myoblasts*, *Cell* **1987**, *51*, 987-1000.
- [16] S. J. Tapscott, R. L. Davis, M. J. Thayer, P. F. Cheng, H. Weintraub, A. B. Lassar, *MyoD1: a nuclear phosphoprotein requiring a Myc homology region to convert fibroblasts to myoblasts*, *Science* **1988**, *242*, 405-411.
- [17] H. Weintraub, S. J. Tapscott, R. L. Davis, M. J. Thayer, M. A. Adam, A. B. Lassar, A. D. Miller, *Activation of muscle-specific genes in pigment, nerve, fat, liver, and fibroblast cell lines by forced expression of MyoD*, *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 5434-5438.
- [18] S. Kühl und M. Kühl, *Stammzellbiologie*, Ulmer, Stuttgart, **2012**.
- [19] E. Häckel, *Natürliche Schöpfungsgeschichte von Dr. Ernst Häckel, Professor in Jena. Berlin bei Georg Reimer. 1868 S. 568*, *Arch. Pharm.* **1869**, *189*, 282-283.
- [20] M. Ramalho-Santos und H. Willenbring, *On the Origin of the Term Stem Cell*, *Cell Stem Cell* **2007**, *1*, 35-38.
- [21] A. Smith, *Formative pluripotency: the executive phase in a developmental continuum*, *Development* **2017**, *144*, 365-373.
- [22] R. Jaenisch und R. Young, *Stem Cells, the Molecular Circuitry of Pluripotency and Nuclear Reprogramming*, *Cell* **2008**, *132*, 567-582.
- [23] S. Morgani, J. Nichols, A.-K. Hadjantonakis, *The many faces of Pluripotency: in vitro adaptations of a continuum of in vivo states*, *BMC Dev. Biol.* **2017**, *17*, 7.

- [24] L. C. Stevens und C. C. Little, *Spontaneous Testicular Teratomas in an Inbred Strain of Mice*, *Proc. Natl. Acad. Sci. USA* **1954**, *40*, 1080-1087.
- [25] G. R. Martin, *Teratocarcinomas and mammalian embryogenesis*, *Science* **1980**, *209*, 768-776.
- [26] M. J. Evans und M. H. Kaufman, *Establishment in culture of pluripotential cells from mouse embryos*, *Nature* **1981**, *292*, 154-156.
- [27] G. R. Martin, *Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells*, *Proc. Natl. Acad. Sci. USA* **1981**, *78*, 7634-7638.
- [28] A. Bradley, M. Evans, M. H. Kaufman, E. Robertson, *Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines*, *Nature* **1984**, *309*, 255-256.
- [29] H. Niwa, *How is pluripotency determined and maintained?*, *Development* **2007**, *134*, 635-646.
- [30] G. R. Martin und M. J. Evans, *Differentiation of clonal lines of teratocarcinoma cells: formation of embryoid bodies in vitro*, *Proc. Natl. Acad. Sci. USA* **1975**, *72*, 1441-1445.
- [31] A. G. Smith, J. K. Heath, D. D. Donaldson, G. G. Wong, J. Moreau, M. Stahl, D. Rogers, *Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides*, *Nature* **1988**, *336*, 688-690.
- [32] R. L. Williams, D. J. Hilton, S. Pease, T. A. Willson, C. L. Stewart, D. P. Gearing, E. F. Wagner, D. Metcalf, N. A. Nicola, N. M. Gough, *Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells*, *Nature* **1988**, *336*, 684-687.
- [33] A. G. Smith und M. L. Hooper, *Buffalo rat liver cells produce a diffusible activity which inhibits the differentiation of murine embryonal carcinoma and embryonic stem cells*, *Dev. Biol.* **1987**, *121*, 1-9.
- [34] Q.-L. Ying, J. Wray, J. Nichols, L. Batlle-Morera, B. Doble, J. Woodgett, P. Cohen, A. Smith, *The ground state of embryonic stem cell self-renewal*, *Nature* **2008**, *453*, 519.
- [35] J. Nichols, K. Jones, J. M. Phillips, S. A. Newland, M. Roode, W. Mansfield, A. Smith, A. Cooke, *Validated germline-competent embryonic stem cell lines from nonobese diabetic mice*, *Nat. Med.* **2009**, *15*, 814-818.
- [36] M. Buehr, S. Meek, K. Blair, J. Yang, J. Ure, J. Silva, R. McLay, J. Hall, Q.-L. Ying, A. Smith, *Capture of Authentic Embryonic Stem Cells from Rat Blastocysts*, *Cell* **2008**, *135*, 1287-1298.
- [37] P. Li, C. Tong, R. Mehrian-Shai, L. Jia, N. Wu, Y. Yan, R. E. Maxson, E. N. Schulze, H. Song, C.-L. Hsieh, M. F. Pera, Q.-L. Ying, *Germline Competent Embryonic Stem Cells Derived from Rat Blastocysts*, *Cell* **2008**, *135*, 1299-1310.
- [38] X. Wu und Y. Zhang, *TET-mediated active DNA demethylation: mechanism, function and beyond*, *Nat. Rev. Genet.* **2017**, *18*, 517-534.
- [39] M. Hemberger, W. Dean, W. Reik, *Epigenetic dynamics of stem cells and cell lineage commitment: digging Waddington's canal*, *Nat. Rev. Mol. Cell Biol.* **2009**, *10*, 526-537.
- [40] I. G. M. Brons, L. E. Smithers, M. W. B. Trotter, P. Rugg-Gunn, B. Sun, S. M. Chuva de Sousa Lopes, S. K. Howlett, A. Clarkson, L. Ahrlund-Richter, R. A. Pedersen, L. Vallier, *Derivation of pluripotent epiblast stem cells from mammalian embryos*, *Nature* **2007**, *448*, 191-195.
- [41] P. J. Tesar, J. G. Chenoweth, F. A. Brook, T. J. Davies, E. P. Evans, D. L. Mack, R. L. Gardner, R. D. G. McKay, *New cell lines from mouse epiblast share defining features with human embryonic stem cells*, *Nature* **2007**, *448*, 196-199.
- [42] Q.-L. Ying und A. Smith, *The Art of Capturing Pluripotency: Creating the Right Culture*, *Stem Cell Rep.* **2017**, *8*, 1457-1464.
- [43] J. A. Thomson, J. Itskovitz-Eldor, S. S. Shapiro, M. A. Waknitz, J. J. Swiergiel, V. S. Marshall, J. M. Jones, *Embryonic Stem Cell Lines Derived from Human Blastocysts*, *Science* **1998**, *282*, 1145-1147.
- [44] J. H. Hanna, K. Saha, R. Jaenisch, *Pluripotency and Cellular Reprogramming: Facts, Hypotheses, Unresolved Issues*, *Cell* **2010**, *143*, 508-525.
- [45] Y. Matsui, K. Zsebo, B. L. M. Hogan, *Derivation of pluripotential embryonic stem cells from murine primordial germ cells in culture*, *Cell* **1992**, *70*, 841-847.
- [46] H. G. Leitch, K. Blair, W. Mansfield, H. Ayetey, P. Humphreys, J. Nichols, M. A. Surani, A. Smith, *Embryonic germ cells from mice and rats exhibit properties consistent with a generic pluripotent ground state*, *Development* **2010**, *137*, 2279-2287.

- [47] J. Yu, M. A. Vodyanik, K. Smuga-Otto, J. Antosiewicz-Bourget, J. L. Frane, S. Tian, J. Nie, G. A. Jonsdottir, V. Ruotti, R. Stewart, I. I. Slukvin, J. A. Thomson, *Induced Pluripotent Stem Cell Lines Derived from Human Somatic Cells*, *Science* **2007**, *318*, 1917-1920.
- [48] K. Takahashi, K. Tanabe, M. Ohnuki, M. Narita, T. Ichisaka, K. Tomoda, S. Yamanaka, *Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors*, *Cell* **2007**, *131*, 861-872.
- [49] T. Boroviak, R. Loos, P. Bertone, A. Smith, J. Nichols, *The ability of inner-cell-mass cells to self-renew as embryonic stem cells is acquired following epiblast specification*, *Nat. Cell Biol.* **2014**, *16*, 513-528.
- [50] J. Nichols und A. Smith, *Naive and Primed Pluripotent States*, *Cell Stem Cell* **2009**, *4*, 487-492.
- [51] L. Weinberger, M. Ayyash, N. Novershtern, J. H. Hanna, *Dynamic stem cell states: naive to primed pluripotency in rodents and humans*, *Nat. Rev. Mol. Cell Biol.* **2016**, *17*, 155-169.
- [52] J. Nichols und A. Smith, *Pluripotency in the Embryo and in Culture*, *Cold Spring Harbor Perspect. Biol.* **2012**, *4*, a008128.
- [53] H. Niwa, T. Burdon, I. Chambers, A. Smith, *Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3*, *Genes Dev.* **1998**, *12*, 2048-2060.
- [54] J. Bain, L. Plater, M. Elliott, N. Shpiro, C. J. Hastie, H. McLauchlan, I. Klevernic, J. Simon C. Arthur, Dario R. Alessi, P. Cohen, *The selectivity of protein kinase inhibitors: a further update*, *Biochem. J.* **2007**, *408*, 297-315.
- [55] Jamie A. Hackett und M. A. Surani, *Regulatory Principles of Pluripotency: From the Ground State Up*, *Cell Stem Cell* **2014**, *15*, 416-430.
- [56] J. Wray, T. Kalkan, S. Gomez-Lopez, D. Eckardt, A. Cook, R. Kemler, A. Smith, *Inhibition of glycogen synthase kinase-3 alleviates Tcf3 repression of the pluripotency network and increases embryonic stem cell resistance to differentiation*, *Nat. Cell Biol.* **2011**, *13*, 838-845.
- [57] B. J. Merrill, *Wnt Pathway Regulation of Embryonic Stem Cell Self-Renewal*, *Cold Spring Harbor Perspect. Biol.* **2012**, *4*, a007971.
- [58] D. ten Berge, D. Kurek, T. Blauwkamp, W. Koole, A. Maas, E. Eroglu, R. K. Siu, R. Nusse, *Embryonic stem cells require Wnt proteins to prevent differentiation to epiblast stem cells*, *Nat. Cell Biol.* **2011**, *13*, 1070-1075.
- [59] M. Yagi, S. Kishigami, A. Tanaka, K. Semi, E. Mizutani, S. Wakayama, T. Wakayama, T. Yamamoto, Y. Yamada, *Derivation of ground-state female ES cells maintaining gamete-derived DNA methylation*, *Nature* **2017**, *548*, 224-227.
- [60] T. Shimizu, J. Ueda, J. C. Ho, K. Iwasaki, L. Poellinger, I. Harada, Y. Sawada, *Dual Inhibition of Src and GSK3 Maintains Mouse Embryonic Stem Cells, Whose Differentiation Is Mechanically Regulated by Src Signaling*, *Stem Cells* **2012**, *30*, 1394-1404.
- [61] M. A. Meyn, S. J. Schreiner, T. P. Dumitrescu, G. J. Nau, T. E. Smithgall, *Src Family Kinase Activity Is Required for Murine Embryonic Stem Cell Growth and Differentiation*, *Mol. Pharmacol.* **2005**, *68*, 1320-1330.
- [62] J. Martin Gonzalez, Sophie M. Morgani, Robert A. Bone, K. Bonderup, S. Abelchian, C. Brakebusch, Joshua M. Brickman, *Embryonic Stem Cell Culture Conditions Support Distinct States Associated with Different Developmental Stages and Potency*, *Stem Cell Rep.* **2016**, *7*, 177-191.
- [63] H. Kim, J. Wu, S. Ye, C.-I. Tai, X. Zhou, H. Yan, P. Li, M. Pera, Q.-L. Ying, *Modulation of β -catenin function maintains mouse epiblast stem cell and human embryonic stem cell self-renewal*, *Nat. Commun.* **2013**, *4*, 2403.
- [64] J. Wu und Juan C. Izpisua Belmonte, *Dynamic Pluripotent Stem Cell States and Their Applications*, *Cell Stem Cell* **2015**, *17*, 509-525.
- [65] K. Iwan, R. Rahimoff, A. Kirchner, F. Spada, A. S. Schröder, O. Kosmatchev, S. Ferizaj, J. Steinbacher, E. Parsa, M. Müller, T. Carell, *5-Formylcytosine to cytosine conversion by C-C bond cleavage in vivo*, *Nat. Chem. Biol.* **2018**, *14*, 72-78.
- [66] M. Tosolini und A. Jouneau, *From Naive to Primed Pluripotency: In Vitro Conversion of Mouse Embryonic Stem Cells in Epiblast Stem Cells*, *Methods Mol. Biol.* **2016**, *1341*, 209-216.

- [67] T. Kalkan und A. Smith, *Mapping the route from naive pluripotency to lineage specification*, *Philos. Trans. R. Soc. Lond., Ser. B: Biol. Sci.* **2014**, 369, 20130540.
- [68] A. T. Chinwalla, L. L. Cook, K. D. Delehaunty, G. A. Fewell, L. A. Fulton, R. S. Fulton, T. A. Graves, L. W. Hillier, E. R. Mardis, J. D. McPherson, T. L. Miner, W. E. Nash, J. O. Nelson, M. N. Nhan, K. H. Pepin, C. S. Pohl, T. C. Ponce, B. Schultz, J. Thompson, E. Trevaskis, *et al.*, *Initial sequencing and comparative analysis of the mouse genome*, *Nature* **2002**, 420, 520-562.
- [69] K. Luger, A. W. Mäder, R. K. Richmond, D. F. Sargent, T. J. Richmond, *Crystal structure of the nucleosome core particle at 2.8 Å resolution*, *Nature* **1997**, 389, 251-260.
- [70] T. Kouzarides, *Chromatin Modifications and Their Function*, *Cell* **2007**, 128, 693-705.
- [71] V. Narendra, P. P. Rocha, D. An, R. Raviram, J. A. Skok, E. O. Mazzone, D. Reinberg, *CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation*, *Science* **2015**, 347, 1017-1021.
- [72] J. L. Workman und R. E. Kingston, *Alteration of Nucleosome Structure as a Mechanism of Transcriptional Regulation*, *Annu. Rev. Biochem.* **1998**, 67, 545-579.
- [73] E. Seto und M. Yoshida, *Erasers of Histone Acetylation: The Histone Deacetylase Enzymes*, *Cold Spring Harbor Perspect. Biol.* **2014**, 6, a018713.
- [74] S. C. Joosten, K. M. Smits, M. J. Aarts, V. Melotte, A. Koch, V. C. Tjan-Heijnen, M. van Engeland, *Epigenetics in renal cell cancer: mechanisms and clinical applications*, *Nat. Rev. Urol.* **2018**, 15, 430-451.
- [75] D. Pasini, K. H. Hansen, J. Christensen, K. Agger, P. A. C. Cloos, K. Helin, *Coordinated regulation of transcriptional repression by the RBP2 H3K4 demethylase and Polycomb-Repressive Complex 2*, *Genes Dev.* **2008**, 22, 1345-1355.
- [76] M. J. Bottomley, *Structures of protein domains that create or recognize histone modifications*, *EMBO Rep.* **2004**, 5, 464-469.
- [77] H. Marks, T. Kalkan, R. Menafra, S. Denissov, K. Jones, H. Hofemeister, J. Nichols, A. Kranz, A. Francis Stewart, A. Smith, Hendrik G. Stunnenberg, *The Transcriptional and Epigenomic Foundations of Ground State Pluripotency*, *Cell* **2012**, 149, 590-604.
- [78] H. G. Leitch, K. R. McEwen, A. Turp, V. Encheva, T. Carroll, N. Grabole, W. Mansfield, B. Nashun, J. G. Knezovich, A. Smith, M. A. Surani, P. Hajkova, *Naive pluripotency is associated with global DNA hypomethylation*, *Nat. Struct. Mol. Biol.* **2013**, 20, 311-316.
- [79] E. Meshorer und T. Misteli, *Chromatin in pluripotent embryonic stem cells and differentiation*, *Nat. Rev. Mol. Cell Biol.* **2006**, 7, 540-546.
- [80] B. E. Bernstein, T. S. Mikkelsen, X. Xie, M. Kamal, D. J. Huebert, J. Cuff, B. Fry, A. Meissner, M. Wernig, K. Plath, R. Jaenisch, A. Wagschal, R. Feil, S. L. Schreiber, E. S. Lander, *A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells*, *Cell* **2006**, 125, 315-326.
- [81] Kevin C. Wang und Howard Y. Chang, *Molecular Mechanisms of Long Noncoding RNAs*, *Mol. Cell* **2011**, 43, 904-914.
- [82] M. J. C. Steenman, S. Rainier, C. J. Dobry, P. Grundy, I. L. Horon, A. P. Feinberg, *Loss of imprinting of IGF2 is linked to reduced expression and abnormal methylation of H19 in Wilms' tumour*, *Nat. Genet.* **1994**, 7, 433-439.
- [83] M. A. Sánchez-Romero, I. Cota, J. Casadesús, *DNA methylation in bacteria: from the methyl group to the methylome*, *Curr. Opin. Microbiol.* **2015**, 25, 9-16.
- [84] M. J. Koziol, C. R. Bradshaw, G. E. Allen, A. S. Costa, C. Frezza, *Identification of Methylated Deoxyadenosines in Genomic DNA by dA6m DNA Immunoprecipitation*, *Bio-Protoc.* **2016**, 6.
- [85] T. Carell, M. Q. Kurz, M. Müller, M. Rossa, F. Spada, *Non-canonical Bases in the Genome: The Regulatory Information Layer in DNA*, *Angew. Chem. Int. Ed* **2018**, 57, 4296-4312.
- [86] E. S. Lander, L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczyk, R. LeVine, P. McEwan, K. McKernan, *et al.*, *Initial sequencing and analysis of the human genome*, *Nature* **2001**, 409, 860-921.
- [87] D. M. Messerschmidt, B. B. Knowles, D. Solter, *DNA methylation dynamics during epigenetic reprogramming in the germline and preimplantation embryos*, *Genes Dev.* **2014**, 28, 812-828.

- [88] J. S. Hardwick, D. Ptchelkine, A. H. El-Sagheer, I. Tear, D. Singleton, S. E. V. Phillips, A. N. Lane, T. Brown, *5-Formylcytosine does not change the global structure of DNA*, *Nat. Struct. Mol. Biol.* **2017**, *24*, 544-552.
- [89] R. Savva, K. McAuley-Hecht, T. Brown, L. Pearl, *The structural basis of specific base-excision repair by uracil–DNA glycosylase*, *Nature* **1995**, *373*, 487-493.
- [90] D. Renciuik, O. Blacque, M. Vorlickova, B. Spingler, *Crystal structures of B-DNA dodecamer containing the epigenetic modifications 5-hydroxymethylcytosine or 5-methylcytosine*, *Nucleic Acids Res.* **2013**, *41*, 9891-9900.
- [91] C. G. Spruijt und M. Vermeulen, *DNA methylation: old dog, new tricks?*, *Nat. Struct. Mol. Biol.* **2014**, *21*, 949-954.
- [92] 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, *Human DNA methylomes at base resolution show widespread epigenomic differences*, *Nature* **2009**, *462*, 315-322.
- [93] 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. Haghghi, T. J. Sejnowski, *et al.*, *Global epigenomic reconfiguration during mammalian brain development*, *Science* **2013**, *341*, 1237905.
- [94] J. U. Guo, D. K. Ma, H. Mo, M. P. Ball, M. H. Jang, M. A. Bonaguidi, J. A. Balazer, H. L. Eaves, B. Xie, E. Ford, K. Zhang, G. L. Ming, Y. Gao, H. Song, *Neuronal activity modifies the DNA methylation landscape in the adult brain*, *Nat. Neurosci.* **2011**, *14*, 1345-1351.
- [95] J.-C. Shen, W. M. Rideout, III, P. A. Jones, *The rate of hydrolytic deamination of 5-methylcytosine in double-stranded DNA*, *Nucleic Acids Res.* **1994**, *22*, 972-976.
- [96] Z. D. Smith und A. Meissner, *DNA methylation: roles in mammalian development*, *Nat. Rev. Genet.* **2013**, *14*, 204-220.
- [97] A. M. Deaton und A. Bird, *CpG islands and the regulation of transcription*, *Genes Dev.* **2011**, *25*, 1010-1022.
- [98] G. R. Wyatt, *Occurrence of 5-methylcytosine in nucleic acids*, *Nature* **1950**, *166*, 237-238.
- [99] P. A. Jones, *Functions of DNA methylation: islands, start sites, gene bodies and beyond*, *Nat. Rev. Genet.* **2012**, *13*, 484-492.
- [100] K. D. Robertson, *DNA methylation and human disease*, *Nat. Rev. Genet.* **2005**, *6*, 597-610.
- [101] J. Du, L. M. Johnson, S. E. Jacobsen, D. J. Patel, *DNA methylation pathways and their crosstalk with histone methylation*, *Nat. Rev. Mol. Cell Biol.* **2015**, *16*, 519-532.
- [102] F. Lyko, *The DNA methyltransferase family: a versatile toolkit for epigenetic regulation*, *Nat. Rev. Genet.* **2017**, *19*, 81-92.
- [103] M. Okano, S. Xie, E. Li, *Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases*, *Nat. Genet.* **1998**, *19*, 219-220.
- [104] M. Okano, D. W. Bell, D. A. Haber, E. Li, *DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development*, *Cell* **1999**, *99*, 247-257.
- [105] M. G. Goll und T. H. Bestor, *Eukaryotic cytosine methyltransferases*, *Annu. Rev. Biochem.* **2005**, *74*, 481-514.
- [106] M. G. Goll, F. Kirpekar, K. A. Maggert, J. A. Yoder, C.-L. Hsieh, X. Zhang, K. G. Golic, S. E. Jacobsen, T. H. Bestor, *Methylation of tRNA^{Asp} by the DNA methyltransferase homolog Dnmt2*, *Science* **2006**, *311*, 395-398.
- [107] D. Bourc'his und T. H. Bestor, *Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L*, *Nature* **2004**, *431*, 96-99.
- [108] J. Barau, A. Teissandier, N. Zamudio, S. Roy, V. Nalesso, Y. Hérault, F. Guillou, D. Bourc'his, *The DNA methyltransferase DNMT3C protects male germ cells from transposon activity*, *Science* **2016**, *354*, 909-912.
- [109] S. Klimasauskas, S. Kumar, R. J. Roberts, X. Cheng, *HhaI methyltransferase flips its target base out of the DNA helix*, *Cell* **1994**, *76*, 357-369.
- [110] A. Jeltsch, W. Nellen, F. Lyko, *Two substrates are better than one: dual specificities for Dnmt2 methyltransferases*, *Trends Biochem. Sci.* **2006**, *31*, 306-308.

- [111] Y. Atlasi und H. G. Stunnenberg, *The interplay of epigenetic marks during stem cell differentiation and development*, *Nat. Rev. Genet.* **2017**, *18*, 643-658.
- [112] A. Tsumura, T. Hayakawa, Y. Kumaki, S. i. Takebayashi, M. Sakaue, C. Matsuoka, K. Shimotohno, F. Ishikawa, E. Li, H. R. Ueda, J. i. Nakayama, M. Okano, *Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b*, *Genes Cells* **2006**, *11*, 805-814.
- [113] M. Jackson, A. Krassowska, N. Gilbert, T. Chevassut, L. Forrester, J. Ansell, B. Ramsahoye, *Severe Global DNA Hypomethylation Blocks Differentiation and Induces Histone Hyperacetylation in Embryonic Stem Cells*, *Mol. Cell. Biol.* **2004**, *24*, 8862-8871.
- [114] T. Chen, Y. Ueda, J. E. Dodge, Z. Wang, E. Li, *Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b*, *Mol. Cell. Biol.* **2003**, *23*, 5594-5605.
- [115] K. L. Tucker, C. Beard, J. Dausmann, L. Jackson-Grusby, P. W. Laird, H. Lei, E. Li, R. Jaenisch, *Germ-line passage is required for establishment of methylation and expression patterns of imprinted but not of nonimprinted genes*, *Genes Dev.* **1996**, *10*, 1008-1020.
- [116] E. Habibi, Arie B. Brinkman, J. Arand, Leonie I. Kroeze, Hindrik H. D. Kerstens, F. Matarese, K. Lepikhov, M. Gut, I. Brun-Heath, Nina C. Hubner, R. Benedetti, L. Altucci, Joop H. Jansen, J. Walter, Ivo G. Gut, H. Marks, Hendrik G. Stunnenberg, *Whole-Genome Bisulfite Sequencing of Two Distinct Interconvertible DNA Methylomes of Mouse Embryonic Stem Cells*, *Cell Stem Cell* **2013**, *13*, 360-369.
- [117] G. Ficiz, Timothy A. Hore, F. Santos, Heather J. Lee, W. Dean, J. Arand, F. Krueger, D. Oxley, Y.-L. Paul, J. Walter, Simon J. Cook, S. Andrews, Miguel R. Branco, W. Reik, *FGF Signaling Inhibition in ESCs Drives Rapid Genome-wide Demethylation to the Epigenetic Ground State of Pluripotency*, *Cell Stem Cell* **2013**, *13*, 351-359.
- [118] C. Bock und A. Wutz, *DNA methylation: a matter of culture*, *Nat. Struct. Mol. Biol.* **2013**, *20*, 249-251.
- [119] Z. Ma, T. Swigut, A. Valouev, A. Rada-Iglesias, J. Wysocka, *Sequence-specific regulator Prdm14 safeguards mouse ESCs from entering extraembryonic endoderm fates*, *Nat. Struct. Mol. Biol.* **2010**, *18*, 120-127.
- [120] H. G. Leitch, M. A. Surani, P. Hajkova, *DNA (De)Methylation: The Passive Route to Naïvety*, *Trends Genet.* **2016**, *32*, 592-595.
- [121] P. Karagianni, L. Amazit, J. Qin, J. Wong, *ICBP90, a Novel Methyl K9 H3 Binding Protein Linking Protein Ubiquitination with Heterochromatin Formation*, *Mol. Cell. Biol.* **2008**, *28*, 705-717.
- [122] F. von Meyenn, M. Iurlaro, E. Habibi, Ning Q. Liu, A. Salehzadeh-Yazdi, F. Santos, E. Petrini, I. Milagre, M. Yu, Z. Xie, Leonie I. Kroeze, Tatyana B. Nesterova, Joop H. Jansen, H. Xie, C. He, W. Reik, Hendrik G. Stunnenberg, *Impairment of DNA Methylation Maintenance Is the Main Cause of Global Demethylation in Naive Embryonic Stem Cells*, *Mol. Cell* **2016**, *62*, 848-861.
- [123] E. Kriukienė, Z. Liutkevičiūtė, S. Klimašauskas, *5-Hydroxymethylcytosine – the elusive epigenetic mark in mammalian DNA*, *Chem. Soc. Rev.* **2012**, *41*, 6916-6930.
- [124] N. Bhutani, David M. Burns, Helen M. Blau, *DNA Demethylation Dynamics*, *Cell* **2011**, *146*, 866-872.
- [125] W. Mayer, A. Niveleau, J. Walter, R. Fundele, T. Haaf, *Demethylation of the zygotic paternal genome*, *Nature* **2000**, *403*, 501-502.
- [126] J. Oswald, S. Engemann, N. Lane, W. Mayer, A. Olek, R. Fundele, W. Dean, W. Reik, J. Walter, *Active demethylation of the paternal genome in the mouse zygote*, *Curr. Biol.* **2000**, *10*, 475-478.
- [127] R. Ono, T. Taki, T. Taketani, M. Taniwaki, H. Kobayashi, Y. Hayashi, *LCX, leukemia-associated protein with a CXXC domain, is fused to MLL in acute myeloid leukemia with trilineage dysplasia having t(10;11)(q22;q23)*. *Cancer Res.* **2002**, *62*, 4075-4080.
- [128] 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, *Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1*, *Science* **2009**, *324*, 930-935.

- [129] T. Pfaffeneder, B. Hackner, M. Truss, M. Münzel, M. Müller, C. A. Deiml, C. Hagemeyer, T. Carell, *The discovery of 5-formylcytosine in embryonic stem cell DNA*, *Angew. Chem. Int. Ed* **2011**, *50*, 7008-7012.
- [130] S. Ito, L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He, Y. Zhang, *Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine*, *Science* **2011**, *333*, 1300-1303.
- [131] 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, *Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA*, *Science* **2011**, *333*, 1303-1307.
- [132] L. Hu, Z. Li, J. Cheng, Q. Rao, W. Gong, M. Liu, Y. G. Shi, J. Zhu, P. Wang, Y. Xu, *Crystal structure of TET2-DNA complex: insight into TET-mediated 5mC oxidation*, *Cell* **2013**, *155*, 1545-1555.
- [133] X. Lu, B. S. Zhao, C. He, *TET Family Proteins: Oxidation Activity, Interacting Molecules, and Functions in Diseases*, *Chem. Rev.* **2015**, *115*, 2225-2239.
- [134] H. Zhao und T. Chen, *Tet family of 5-methylcytosine dioxygenases in mammalian development*, *J. Hum. Genet.* **2013**, *58*, 421-427.
- [135] M. Ravichandran, R. Z. Jurkowska, T. P. Jurkowski, *Target specificity of mammalian DNA methylation and demethylation machinery*, *Org. Biomol. Chem.* **2018**, *16*, 1419-1435.
- [136] W. A. Pastor, U. J. Pape, Y. Huang, H. R. Henderson, R. Lister, M. Ko, E. M. McLoughlin, Y. Brudno, S. Mahapatra, P. Kapranov, M. Tahiliani, G. Q. Daley, X. S. Liu, J. R. Ecker, P. M. Milos, S. Agarwal, A. Rao, *Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells*, *Nature* **2011**, *473*, 394-397.
- [137] C. Loenarz und C. J. Schofield, *Oxygenase Catalyzed 5-Methylcytosine Hydroxylation*, *Chem. Biol.* **2009**, *16*, 580-583.
- [138] L. M. Iyer, M. Tahiliani, A. Rao, L. Aravind, *Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids*, *Cell Cycle* **2009**, *8*, 1698-1710.
- [139] N. Liu, M. Wang, W. Deng, C. S. Schmidt, W. Qin, H. Leonhardt, F. Spada, *Intrinsic and extrinsic connections of Tet3 dioxygenase with CXXC zinc finger modules*, *PLoS One* **2013**, *8*, e62755.
- [140] M. Ko, J. An, H. S. Bandukwala, L. Chavez, T. Aijo, W. A. Pastor, M. F. Segal, H. Li, K. P. Koh, H. Lahdesmaki, P. G. Hogan, L. Aravind, A. Rao, *Modulation of TET2 expression and 5-methylcytosine oxidation by the CXXC domain protein IDAX*, *Nature* **2013**, *497*, 122-126.
- [141] H. Hashimoto, J. E. Pais, X. Zhang, L. Saleh, Z. Q. Fu, N. Dai, I. R. Correa, Jr., Y. Zheng, X. Cheng, *Structure of a Naegleria Tet-like dioxygenase in complex with 5-methylcytosine DNA*, *Nature* **2014**, *506*, 391-395.
- [142] T. Pfaffeneder, F. Spada, M. Wagner, C. Brandmayr, S. K. Laube, D. Eisen, M. Truss, J. Steinbacher, B. Hackner, O. Kotljarova, D. Schuermann, S. Michalakis, O. Kosmatchev, S. Schiesser, B. Steigenberger, N. Raddaoui, G. Kashiwazaki, U. Müller, C. G. Spruijt, M. Vermeulen, *et al.*, *Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA*, *Nat. Chem. Biol.* **2014**, *10*, 574-581.
- [143] F. Kawasaki, D. Beraldi, R. E. Hardisty, G. R. McInroy, P. van Delft, S. Balasubramanian, *Genome-wide mapping of 5-hydroxymethyluracil in the eukaryote parasite Leishmania*, *Genome Biol.* **2017**, *18*, 23.
- [144] F. Kawasaki, S. Martínez Cuesta, D. Beraldi, A. Mahtey, R. E. Hardisty, M. Carrington, S. Balasubramanian, *Sequencing 5-Hydroxymethyluracil at Single-Base Resolution*, *Angew. Chem. Int. Ed* **2018**, *57*, 9694-9696.
- [145] G. R. Wyatt und S. S. Cohen, *A new pyrimidine base from bacteriophage nucleic acids*, *Nature* **1952**, *170*, 1072-1073.
- [146] N. W. Penn, R. Suwalski, C. O'riley, K. Bojanowski, R. Yura, *Presence of 5-Hydroxymethylcytosine in Animal Deoxyribonucleic Acid*, *Biochem. J.* **1972**, *126*, 781-790.
- [147] R. M. Kothari und V. Shankar, *5-Methylcytosine content in the vertebrate deoxyribonucleic acids: species specificity*, *J. Mol. Evol.* **1976**, *7*, 325-329.

- [148] J. R. Wagner und J. Cadet, *Oxidation Reactions of Cytosine DNA Components by Hydroxyl Radical and One-Electron Oxidants in Aerated Aqueous Solutions*, *Acc. Chem. Res.* **2010**, *43*, 564-571.
- [149] S. Kriaucionis und N. Heintz, *The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain*, *Science* **2009**, *324*, 929-930.
- [150] M. Münzel, D. Globisch, T. Brückl, M. Wagner, V. Welzmler, S. Michalakis, M. Müller, M. Biel, T. Carell, *Quantification of the Sixth DNA Base Hydroxymethylcytosine in the Brain*, *Angew. Chem. Int. Ed* **2010**, *49*, 5375-5377.
- [151] D. Globisch, M. Münzel, M. Müller, S. Michalakis, M. Wagner, S. Koch, T. Brückl, M. Biel, T. Carell, *Tissue Distribution of 5-Hydroxymethylcytosine and Search for Active Demethylation Intermediates*, *PLoS One* **2010**, *5*, e15367.
- [152] M. Wagner, J. Steinbacher, T. F. J. Kraus, S. Michalakis, B. Hackner, T. Pfaffeneder, A. Perera, M. Müller, A. Giese, H. A. Kretschmar, T. Carell, *Age-dependent levels of 5-methyl-, 5-hydroxymethyl-, and 5-formylcytosine in human and mouse brain tissues*, *Angew. Chem. Int. Ed* **2015**, *54*, 12511-12514.
- [153] E.-A. Raiber, D. Beraldi, S. Martínez Cuesta, G. R. McInroy, Z. Kingsbury, J. Becq, T. James, M. Lopes, K. Allinson, S. Field, S. Humphray, T. Santarius, C. Watts, D. Bentley, S. Balasubramanian, *Base resolution maps reveal the importance of 5-hydroxymethylcytosine in a human glioblastoma*, *npj Genomic Med.* **2017**, *2*, 6.
- [154] B. Thienpont, J. Steinbacher, H. Zhao, F. D'Anna, A. Kuchnio, A. Ploumakis, B. Ghesquière, L. Van Dyck, B. Boeckx, L. Schoonjans, E. Hermans, F. Amant, V. N. Kristensen, K. P. Koh, M. Mazzone, M. L. Coleman, T. Carell, P. Carmeliet, D. Lambrechts, *Tumour hypoxia causes DNA hypermethylation by reducing TET activity*, *Nature* **2016**, *537*, 63-68.
- [155] W. Li, X. Zhang, X. Lu, L. You, Y. Song, Z. Luo, J. Zhang, J. Nie, W. Zheng, D. Xu, Y. Wang, Y. Dong, S. Yu, J. Hong, J. Shi, H. Hao, F. Luo, L. Hua, P. Wang, X. Qian, *et al.*, *5-Hydroxymethylcytosine signatures in circulating cell-free DNA as diagnostic biomarkers for human cancers*, *Cell Res.* **2017**, *27*, 1243-1257.
- [156] D. Gackowski, E. Zarakowska, M. Starczak, M. Modrzejewska, R. Olinski, *Tissue-Specific Differences in DNA Modifications (5-Hydroxymethylcytosine, 5-Formylcytosine, 5-Carboxylcytosine and 5-Hydroxymethyluracil) and Their Interrelationships*, *PLoS One* **2015**, *10*, e0144859.
- [157] M. Bachman, S. Uribe-Lewis, X. Yang, M. Williams, A. Murrell, S. Balasubramanian, *5-Hydroxymethylcytosine is a predominantly stable DNA modification*, *Nat. Chem.* **2014**, *6*, 1049-1055.
- [158] M. Yu, G. C. Hon, K. E. Szulwach, C. X. Song, L. Zhang, A. Kim, X. Li, Q. Dai, Y. Shen, B. Park, J. H. Min, P. Jin, B. Ren, C. He, *Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome*, *Cell* **2012**, *149*, 1368-1380.
- [159] H. Wu, X. Wu, L. Shen, Y. Zhang, *Single-base resolution analysis of active DNA demethylation using methylase-assisted bisulfite sequencing*, *Nat. Biotechnol.* **2014**, *32*, 1231-1240.
- [160] K. E. Szulwach, X. Li, Y. Li, C. X. Song, J. W. Han, S. Kim, S. Namburi, K. Hermetz, J. J. Kim, M. K. Rudd, Y. S. Yoon, B. Ren, C. He, P. Jin, *Integrating 5-hydroxymethylcytosine into the epigenomic landscape of human embryonic stem cells*, *PLoS Genet.* **2011**, *7*, e1002154.
- [161] C. Luo, P. Hajkova, J. R. Ecker, *Dynamic DNA methylation: In the right place at the right time*, *Science* **2018**, *361*, 1336-1340.
- [162] A. Hofer, Z. J. Liu, S. Balasubramanian, *Detection, Structure and Function of Modified DNA Bases*, *J. Am. Chem. Soc.* **2019**, *141*, 6420-6429.
- [163] A. Maiti und A. C. Drohat, *Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites*, *J. Biol. Chem.* **2011**, *286*, 35334-35338.
- [164] L. Shen, H. Wu, D. Diep, S. Yamaguchi, A. C. D'Alessio, H. L. Fung, K. Zhang, Y. Zhang, *Genome-wide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics*, *Cell* **2013**, *153*, 692-706.

- [165] D. Condliffe, A. Wong, C. Troakes, P. Proitsi, Y. Patel, L. Chouliaras, C. Fernandes, J. Cooper, S. Lovestone, L. Schalkwyk, J. Mill, K. Lunnon, *Cross-region reduction in 5-hydroxymethylcytosine in Alzheimer's disease brain*, *Neurobiol. Aging* **2014**, *35*, 1850-1854.
- [166] E.-A. Raiber, R. Hardisty, P. van Delft, S. Balasubramanian, *Mapping and elucidating the function of modified bases in DNA*, *Nat. Rev. Chem.* **2017**, *1*, 0069.
- [167] M. Iurlaro, G. R. McInroy, H. E. Burgess, W. Dean, E. A. Raiber, M. Bachman, D. Beraldi, S. Balasubramanian, W. Reik, *In vivo genome-wide profiling reveals a tissue-specific role for 5-formylcytosine*, *Genome Biol.* **2016**, *17*, 141.
- [168] M. Iurlaro, G. Ficz, D. Oxley, E. A. Raiber, M. Bachman, M. J. Booth, S. Andrews, S. Balasubramanian, W. Reik, *A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation*, *Genome Biol.* **2013**, *14*, R119.
- [169] C. G. Spruijt, F. Gnerlich, A. H. Smits, T. Pfaffeneder, P. W. Jansen, C. Bauer, M. Munzel, M. Wagner, M. Müller, F. Khan, H. C. Eberl, A. Mensinga, A. B. Brinkman, K. Lephikov, U. Müller, J. Walter, R. Boelens, H. van Ingen, H. Leonhardt, T. Carell, *et al.*, *Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives*, *Cell* **2013**, *152*, 1146-1159.
- [170] M. Su, A. Kirchner, S. Stazzoni, M. Müller, M. Wagner, A. Schröder, T. Carell, *5-Formylcytosine Could Be a Semipermanent Base in Specific Genome Sites*, *Angew. Chem. Int. Ed* **2016**, *55*, 11797-11800.
- [171] M. Bachman, S. Uribe-Lewis, X. P. Yang, H. E. Burgess, M. Iurlaro, W. Reik, A. Murrell, S. Balasubramanian, *5-Formylcytosine can be a stable DNA modification in mammals*, *Nat. Chem. Biol.* **2015**, *11*, 555-557.
- [172] E.-A. Raiber, P. Murat, D. Y. Chirgadze, D. Beraldi, B. F. Luisi, S. Balasubramanian, *5-Formylcytosine alters the structure of the DNA double helix*, *Nat. Struct. Mol. Biol.* **2014**, *22*, 44.
- [173] T. T. M. Ngo, J. Yoo, Q. Dai, Q. Zhang, C. He, A. Aksimentiev, T. Ha, *Effects of cytosine modifications on DNA flexibility and nucleosome mechanical stability*, *Nat. Commun.* **2016**, *7*, 10813.
- [174] S. Ji, H. Shao, Q. Han, C. L. Seiler, N. Y. Tretyakova, *Reversible DNA-Protein Cross-Linking at Epigenetic DNA Marks*, *Angew. Chem. Int. Ed* **2017**, *56*, 14130-14134.
- [175] F. Li, Y. Zhang, J. Bai, M. M. Greenberg, Z. Xi, C. Zhou, *5-Formylcytosine Yields DNA-Protein Cross-Links in Nucleosome Core Particles*, *J. Am. Chem. Soc.* **2017**, *139*, 10617-10620.
- [176] E.-A. Raiber, G. Portella, S. Martínez Cuesta, R. Hardisty, P. Murat, Z. Li, M. Iurlaro, W. Dean, J. Spindel, D. Beraldi, Z. Liu, M. A. Dawson, W. Reik, S. Balasubramanian, *5-Formylcytosine organizes nucleosomes and forms Schiff base interactions with histones in mouse embryonic stem cells*, *Nat. Chem.* **2018**, *10*, 1258-1266.
- [177] L. Wang, Y. Zhou, L. Xu, R. Xiao, X. Lu, L. Chen, J. Chong, H. Li, C. He, X.-D. Fu, D. Wang, *Molecular basis for 5-carboxycytosine recognition by RNA polymerase II elongation complex*, *Nature* **2015**, *523*, 621-625.
- [178] M. W. Kellinger, C.-X. Song, J. Chong, X.-Y. Lu, C. He, D. Wang, *5-formylcytosine and 5-carboxylcytosine reduce the rate and substrate specificity of RNA polymerase II transcription*, *Nat. Struct. Mol. Biol.* **2012**, *19*, 831-833.
- [179] M. Bostick, J. K. Kim, P. O. Esteve, A. Clark, S. Pradhan, S. E. Jacobsen, *UHRF1 plays a role in maintaining DNA methylation in mammalian cells*, *Science* **2007**, *317*, 1760-1764.
- [180] J. Sharif, M. Muto, S. Takebayashi, I. Suetake, A. Iwamatsu, T. A. Endo, J. Shinga, Y. Mizutani-Koseki, T. Toyoda, K. Okamura, S. Tajima, K. Mitsuya, M. Okano, H. Koseki, *The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA*, *Nature* **2007**, *450*, 908-912.
- [181] H. Hashimoto, Y. Liu, A. K. Upadhyay, Y. Chang, S. B. Howerton, P. M. Vertino, X. Zhang, X. Cheng, *Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation*, *Nucleic Acids Res.* **2012**, *40*, 4841-4849.
- [182] J. Otani, H. Kimura, J. Sharif, T. A. Endo, Y. Mishima, T. Kawakami, H. Koseki, M. Shirakawa, I. Suetake, S. Tajima, *Cell cycle-dependent turnover of 5-hydroxymethyl cytosine in mouse embryonic stem cells*, *PLoS one* **2013**, *8*, e82961.

- [183] D. Ji, K. Lin, J. Song, Y. Wang, *Effects of Tet-induced oxidation products of 5-methylcytosine on Dnmt1- and DNMT3a-mediated cytosine methylation*, *Mol. BioSyst.* **2014**, *10*, 1749-1752.
- [184] H. Wu und Y. Zhang, *Reversing DNA Methylation: Mechanisms, Genomics, and Biological Functions*, *Cell* **2014**, *156*, 45-68.
- [185] R. M. Kohli und Y. Zhang, *TET enzymes, TDG and the dynamics of DNA demethylation*, *Nature* **2013**, *502*, 472-479.
- [186] M. Wossidlo, T. Nakamura, K. Lepikhov, C. J. Marques, V. Zakhartchenko, M. Boiani, J. Arand, T. Nakano, W. Reik, J. Walter, *5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming*, *Nat. Commun.* **2011**, *2*, 241.
- [187] A. Inoue, L. Shen, Q. Dai, C. He, Y. Zhang, *Generation and replication-dependent dilution of 5fC and 5caC during mouse preimplantation development*, *Cell Res.* **2011**, *21*, 1670-1676.
- [188] K. Iqbal, S. G. Jin, G. P. Pfeifer, P. E. Szabo, *Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine*, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 3642-3647.
- [189] N. Rougier, D. Bourc'his, D. M. Gomes, A. Niveleau, M. Plachot, A. Paldi, E. Viegas-Pequignot, *Chromosome methylation patterns during mammalian preimplantation development*, *Genes Dev.* **1998**, *12*, 2108-2113.
- [190] A. Inoue und Y. Zhang, *Replication-Dependent Loss of 5-Hydroxymethylcytosine in Mouse Preimplantation Embryos*, *Science* **2011**, *334*, 194.
- [191] 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, *et al.*, *Thymine DNA Glycosylase Is Essential for Active DNA Demethylation by Linked Deamination-Base Excision Repair*, *Cell* **2011**, *146*, 67-79.
- [192] 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, *The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes*, *Nature* **2011**, *477*, 606-610.
- [193] 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, *Active and passive demethylation of male and female pronuclear DNA in the mammalian zygote*, *Cell Stem Cell* **2014**, *15*, 447-459.
- [194] L. Shen, A. Inoue, J. He, Y. Liu, F. Lu, Y. Zhang, *Tet3 and DNA Replication Mediate Demethylation of Both the Maternal and Paternal Genomes in Mouse Zygotes*, *Cell Stem Cell* **2014**, *15*, 459-471.
- [195] C. Bian und X. Yu, *PGC7 suppresses TET3 for protecting DNA methylation*, *Nucleic Acids Res.* **2013**, *42*, 2893-2905.
- [196] R. Amouroux, B. Nashun, K. Shirane, S. Nakagawa, P. W. S. Hill, Z. D'Souza, M. Nakayama, M. Matsuda, A. Turp, E. Ndjetehe, V. Encheva, N. R. Kudo, H. Koseki, H. Sasaki, P. Hajkova, *De novo DNA methylation drives 5hmC accumulation in mouse zygotes*, *Nat. Cell Biol.* **2016**, *18*, 225-233.
- [197] L. Wang, J. Zhang, J. Duan, X. Gao, W. Zhu, X. Lu, L. Yang, J. Zhang, G. Li, W. Ci, W. Li, Q. Zhou, N. Aluru, F. Tang, C. He, X. Huang, J. Liu, *Programming and Inheritance of Parental DNA Methylation in Mammals*, *Cell* **2014**, *157*, 979-991.
- [198] J. R. Peat, W. Dean, S. J. Clark, F. Krueger, S. A. Smallwood, G. Ficiz, J. K. Kim, J. C. Marioni, T. A. Hore, W. Reik, *Genome-wide bisulfite sequencing in zygotes identifies demethylation targets and maps the contribution of TET3 oxidation*, *Cell Rep.* **2014**, *9*, 1990-2000.
- [199] J. Arand, M. Wossidlo, K. Lepikhov, J. R. Peat, W. Reik, J. Walter, *Selective impairment of methylation maintenance is the major cause of DNA methylation reprogramming in the early embryo*, *Epigenet. Chromatin* **2015**, *8*, 1.
- [200] Junjie U. Guo, Y. Su, C. Zhong, G.-I. Ming, H. Song, *Hydroxylation of 5-Methylcytosine by TET1 Promotes Active DNA Demethylation in the Adult Brain*, *Cell* **2011**, *145*, 423-434.
- [201] W. A. Pastor, L. Aravind, A. Rao, *TETonic shift: biological roles of TET proteins in DNA demethylation and transcription*, *Nat. Rev. Mol. Cell Biol.* **2013**, *14*, 341-356.

- [202] J.-K. Zhu, *Active DNA Demethylation Mediated by DNA Glycosylases*, *Annu. Rev. Genet.* **2009**, *43*, 143-166.
- [203] A. Maiti, A. Z. Michelson, C. J. Armwood, J. K. Lee, A. C. Drohat, *Divergent Mechanisms for Enzymatic Excision of 5-Formylcytosine and 5-Carboxylcytosine from DNA*, *J. Am. Chem. Soc.* **2013**, *135*, 15813-15822.
- [204] A. C. Drohat und C. T. Coey, *Role of Base Excision "Repair" Enzymes in Erasing Epigenetic Marks from DNA*, *Chem. Rev.* **2016**, *116*, 12711-12729.
- [205] S. Um, M. Harbers, A. Benecke, B. t. Pierrat, R. Losson, P. Chambon, *Retinoic Acid Receptors Interact Physically and Functionally with the T:G Mismatch-specific Thymine-DNA Glycosylase*, *J. Biol. Chem.* **1998**, *273*, 20728-20736.
- [206] Y.-Q. Li, P.-Z. Zhou, X.-D. Zheng, C. P. Walsh, G.-L. Xu, *Association of Dnmt3a and thymine DNA glycosylase links DNA methylation with base-excision repair*, *Nucleic Acids Res.* **2006**, *35*, 390-400.
- [207] A. R. Weber, C. Krawczyk, A. B. Robertson, A. Kusnierczyk, C. B. Vagbo, D. Schuermann, A. Klungland, P. Schär, *Biochemical reconstitution of TET1-TDG-BER-dependent active DNA demethylation reveals a highly coordinated mechanism*, *Nat. Commun.* **2016**, *7*, 10806.
- [208] M. Tini, A. Benecke, S.-J. Um, J. Torchia, R. M. Evans, P. Chambon, *Association of CBP/p300 Acetylase and Thymine DNA Glycosylase Links DNA Repair and Transcription*, *Mol. Cell* **2002**, *9*, 265-277.
- [209] R. A. Henry, P. Mancuso, Y.-M. Kuo, R. Tricarico, M. Tini, P. A. Cole, A. Bellacosa, A. J. Andrews, *Interaction with the DNA Repair Protein Thymine DNA Glycosylase Regulates Histone Acetylation by p300*, *Biochemistry* **2016**, *55*, 6766-6775.
- [210] L. Schomacher, D. Han, M. U. Musheev, K. Arab, S. Kienhofer, A. von Seggern, C. Niehrs, *Neil DNA glycosylases promote substrate turnover by Tdg during DNA demethylation*, *Nat. Struct. Mol. Biol.* **2016**, *23*, 116-124.
- [211] M. Bochtler, A. Kolano, G.-L. Xu, *DNA demethylation pathways: Additional players and regulators*, *BioEssays* **2017**, *39*, 1-13 e201600178.
- [212] S. Z. Krokeide, J. K. Laerdahl, M. Salah, L. Luna, F. H. Cedervik, A. M. Fleming, C. J. Burrows, B. Dalhus, M. Bjørås, *Human NEIL3 is mainly a monofunctional DNA glycosylase removing spiroimidiohydantoin and guanidinohydantoin*, *DNA Repair* **2013**, *12*, 1159-1164.
- [213] H. E. Krokan und M. Bjørås, *Base excision repair*, *Cold Spring Harbor Perspect. Biol.* **2013**, *5*, a012583.
- [214] U. Müller, C. Bauer, M. Siegl, A. Rottach, H. Leonhardt, *TET-mediated oxidation of methylcytosine causes TDG or NEIL glycosylase dependent gene reactivation*, *Nucleic Acids Res.* **2014**, *42*, 8592-8604.
- [215] A. Slyvka, K. Mierzejewska, M. Bochtler, *Nei-like 1 (NEIL1) excises 5-carboxylcytosine directly and stimulates TDG-mediated 5-formyl and 5-carboxylcytosine excision*, *Sci. Rep.* **2017**, *7*, 9001.
- [216] G. Rangam, K. M. Schmitz, A. J. Cobb, S. K. Petersen-Mahrt, *AID enzymatic activity is inversely proportional to the size of cytosine C5 orbital cloud*, *PLoS One* **2012**, *7*, e43279.
- [217] R. Bransteitter, P. Pham, M. D. Scharff, M. F. Goodman, *Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase*, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 4102-4107.
- [218] M. Larijani, D. Frieder, T. M. Sonbuchner, R. Bransteitter, M. F. Goodman, E. E. Bouhassira, M. D. Scharff, A. Martin, *Methylation protects cytidines from AID-mediated deamination*, *Mol. Immunol.* **2005**, *42*, 599-604.
- [219] C. S. Nabel, H. Jia, Y. Ye, L. Shen, H. L. Goldschmidt, J. T. Stivers, Y. Zhang, R. M. Kohli, *AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation*, *Nat. Chem. Biol.* **2012**, *8*, 751-758.
- [220] P. M. Dominguez, M. Teater, N. Chambwe, M. Kormaksson, D. Redmond, J. Ishii, B. Vuong, J. Chaudhuri, A. Melnick, A. Vasanthakumar, L. A. Godley, F. N. Papavasiliou, O. Elemento, R. Shakhovich, *DNA Methylation Dynamics of Germinal Center B Cells Are Mediated by AID*, *Cell Rep.* **2015**, *12*, 2086-2098.

- [221] M. A. Hogenbirk, M. R. Heideman, A. Velds, P. C. van den Berk, R. M. Kerkhoven, B. van Steensel, H. Jacobs, *Differential programming of B cells in AID deficient mice*, *PLoS One* **2013**, *8*, e69815.
- [222] D. M. Franchini, C. F. Chan, H. Morgan, E. Incorvaia, G. Rangam, W. Dean, F. Santos, W. Reik, S. K. Petersen-Mahrt, *Processive DNA demethylation via DNA deaminase-induced lesion resolution*, *PLoS One* **2014**, *9*, e97754.
- [223] F. Santos, J. Peat, H. Burgess, C. Rada, W. Reik, W. Dean, *Active demethylation in mouse zygotes involves cytosine deamination and base excision repair*, *Epigenet. Chromatin* **2013**, *6*, 39.
- [224] E. K. Schutsky, C. S. Nabel, A. K. F. Davis, J. E. DeNizio, R. M. Kohli, *APOBEC3A efficiently deaminates methylated, but not TET-oxidized, cytosine bases in DNA*, *Nucleic Acids Res.* **2017**, *45*, 7655-7665.
- [225] N. Bhutani, J. J. Brady, M. Damian, A. Sacco, S. Y. Corbel, H. M. Blau, *Reprogramming towards pluripotency requires AID-dependent DNA demethylation*, *Nature* **2010**, *463*, 1042-1047.
- [226] K. M. Foshay, T. J. Looney, S. Chari, F. F. Mao, J. H. Lee, L. Zhang, C. J. Fernandes, S. W. Baker, K. L. Clift, J. Gaetz, C. G. Di, A. P. Xiang, B. T. Lahn, *Embryonic stem cells induce pluripotency in somatic cell fusion through biphasic reprogramming*, *Mol. Cell* **2012**, *46*, 159-170.
- [227] R. Kumar, L. DiMenna, N. Schrode, T. C. Liu, P. Franck, S. Munoz-Descalzo, A. K. Hadjantonakis, A. A. Zarrin, J. Chaudhuri, O. Elemento, T. Evans, *AID stabilizes stem-cell phenotype by removing epigenetic memory of pluripotency genes*, *Nature* **2013**, *500*, 89-92.
- [228] K. Rai, I. J. Huggins, S. R. James, A. R. Karpf, D. A. Jones, B. R. Cairns, *DNA Demethylation in Zebrafish Involves the Coupling of a Deaminase, a Glycosylase, and Gadd45*, *Cell* **2008**, *135*, 1201-1212.
- [229] N. Shimoda, K. Hirose, R. Kaneto, T. Izawa, H. Yokoi, N. Hashimoto, Y. Kikuchi, *No Evidence for AID/MBD4-Coupled DNA Demethylation in Zebrafish Embryos*, *PLoS One* **2014**, *9*, e114816.
- [230] S. Schiesser, T. Pfaffeneder, K. Sadeghian, B. Hackner, B. Steigenberger, A. S. Schröder, J. Steinbacher, G. Kashiwazaki, G. Hofner, K. T. Wanner, C. Ochsenfeld, T. Carell, *Deamination, oxidation, and C-C bond cleavage reactivity of 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxycytosine*, *J. Am. Chem. Soc.* **2013**, *135*, 14593-14599.
- [231] L. Shen, C.-X. Song, C. He, Y. Zhang, *Mechanism and Function of Oxidative Reversal of DNA and RNA Methylation*, *Annu. Rev. Biochem.* **2014**, *83*, 585-614.
- [232] S. M. Kooistra und K. Helin, *Molecular mechanisms and potential functions of histone demethylases*, *Nat. Rev. Mol. Cell Biol.* **2012**, *13*, 297-311.
- [233] D. Mazel, S. Pochet, P. Marlière, *Genetic characterization of polypeptide deformylase, a distinctive enzyme of eubacterial translation*, *EMBO J.* **1994**, *13*, 914-923.
- [234] T. Y. Hargrove, Z. Wawrzak, J. Liu, W. D. Nes, M. R. Waterman, G. I. Lepesheva, *Substrate Preferences and Catalytic Parameters Determined by Structural Characteristics of Sterol 14 α -Demethylase (CYP51) from *Leishmania infantum**, *J. Biol. Chem.* **2011**, *286*, 26838-26848.
- [235] K. G. Aukema, T. M. Makris, S. A. Stoian, J. E. Richman, E. Münck, J. D. Lipscomb, L. P. Wackett, *Cyanobacterial aldehyde deformylase oxygenation of aldehydes yields n-1 aldehydes and alcohols in addition to alkanes*, *ACS Catal.* **2013**, *3*, 2228-2238.
- [236] C. Jia, M. Li, J. Li, J. Zhang, H. Zhang, P. Cao, X. Pan, X. Lu, W. Chang, *Structural insights into the catalytic mechanism of aldehyde-deformylating oxygenases*, *Protein Cell* **2015**, *6*, 55-67.
- [237] G. I. Lepesheva, T. Y. Hargrove, Y. Kleshchenko, W. D. Nes, F. Villalta, M. R. Waterman, *CYP51: A major drug target in the cytochrome P450 superfamily*, *Lipids* **2008**, *43*, 1117-1125.
- [238] J. A. Smiley, M. Kundracik, D. A. Landfried, V. R. Barnes, A. A. Axhemi, *Genes of the thymidine salvage pathway: Thymine-7-hydroxylase from a *Rhodotorula glutinis* cDNA library and iso-orotate decarboxylase from *Neurospora crassa**, *Biochim. Biophys. Acta* **2005**, *1723*, 256-264.
- [239] W. Li, T. Zhang, J. Ding, *Molecular basis for the substrate specificity and catalytic mechanism of thymine-7-hydroxylase in fungi*, *Nucleic Acids Res.* **2015**, *43*, 10026-10038.
- [240] S. Xu, W. Li, J. Zhu, R. Wang, Z. Li, G.-L. Xu, J. Ding, *Crystal structures of iso-orotate decarboxylases reveal a novel catalytic mechanism of 5-carboxyl-uracil decarboxylation and shed light on the search for DNA decarboxylase*, *Cell Res.* **2013**, *23*, 1296-1309.

- [241] L. Huo, F. Liu, H. Iwaki, T. Li, Y. Hasegawa, A. Liu, *Human α -amino- β -carboxymuconate- ϵ -semialdehyde decarboxylase (ACMSD): a structural and mechanistic unveiling*, *Proteins* **2015**, *83*, 178-187.
- [242] T. Li, H. Iwaki, R. Fu, Y. Hasegawa, H. Zhang, A. Liu, *α -Amino- β -carboxymuconic- ϵ -semialdehyde Decarboxylase (ACMSD) Is a New Member of the Amidohydrolase Superfamily*, *Biochemistry* **2006**, *45*, 6628-6634.
- [243] Z. Liutkeviciute, E. Kriukiene, J. Licyte, M. Rudyte, G. Urbanaviciute, S. Klimasauskas, *Direct decarboxylation of 5-carboxylcytosine by DNA C5-methyltransferases*, *J. Am. Chem. Soc.* **2014**, *136*, 5884-5887.
- [244] Z. Liutkeviciute, G. Lukinavicius, V. Masevicius, D. Daujotyte, S. Klimasauskas, *Cytosine-5-methyltransferases add aldehydes to DNA*, *Nat. Chem. Biol.* **2009**, *5*, 400-402.
- [245] C. C. Chen, K. Y. Wang, C. K. Shen, *The mammalian de novo DNA methyltransferases DNMT3A and DNMT3B are also DNA 5-hydroxymethylcytosine dehydroxymethylases*, *J. Biol. Chem.* **2012**, *287*, 33116-33121.
- [246] C. C. Chen, K. Y. Wang, C. K. Shen, *DNA 5-methylcytosine demethylation activities of the mammalian DNA methyltransferases*, *J. Biol. Chem.* **2013**, *288*, 9084-9091.
- [247] S. Schiesser, B. Hackner, T. Pfaffeneder, M. Müller, C. Hagemeyer, M. Truss, T. Carell, *Mechanism and stem-cell activity of 5-carboxycytosine decarboxylation determined by isotope tracing*, *Angew. Chem. Int. Ed* **2012**, *51*, 6516-6520.
- [248] A. S. Schröder, E. Parsa, K. Iwan, F. R. Traube, M. Wallner, S. Serdjukow, T. Carell, *2'-(R)-Fluorinated mC, hmC, fC and caC triphosphates are substrates for DNA polymerases and TET-enzymes*, *Chemical Communications* **2016**, *52*, 14361-14364.
- [249] J. A. Steinbacher, Dissertation, Ludwig-Maximilians-Universität München, **2017**.
- [250] A. S. Schröder, Dissertation, Ludwig-Maximilians-Universität München, **2016**.
- [251] K. Hayashi, H. Ohta, K. Kurimoto, S. Aramaki, M. Saitou, *Reconstitution of the Mouse Germ Cell Specification Pathway in Culture by Pluripotent Stem Cells*, *Cell* **2011**, *146*, 519-532.
- [252] K. Hayashi und M. Saitou, *Generation of eggs from mouse embryonic stem cells and induced pluripotent stem cells*, *Nat. Protoc.* **2013**, *8*, 1513-1524.
- [253] D. Cortázar, 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. Schär, *Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability*, *Nature* **2011**, *470*, 419-423.
- [254] C. Kunz, F. Focke, Y. Saito, D. Schuermann, T. Lettieri, J. Selfridge, P. Schär, *Base Excision by Thymine DNA Glycosylase Mediates DNA-Directed Cytotoxicity of 5-Fluorouracil*, *PLoS Biol.* **2009**, *7*, e1000091.
- [255] X. Zhou, J. P. Chadarevian, B. Ruiz, Q.-L. Ying, *Cytoplasmic and Nuclear TAZ Exert Distinct Functions in Regulating Primed Pluripotency*, *Stem Cell Rep.* **2017**, *9*, 732-741.
- [256] F. Lu, Y. Liu, L. Jiang, S. Yamaguchi, Y. Zhang, *Role of Tet proteins in enhancer activity and telomere elongation*, *Genes Dev.* **2014**, *28*, 2103-2119.
- [257] K. Iwan, Dissertation, Ludwig-Maximilians-Universität München, **2018**.
- [258] T. Lindahl, *Instability and decay of the primary structure of DNA*, *Nature* **1993**, *362*, 709-715.
- [259] R. J. Rahimoff, Dissertation, Ludwig-Maximilians-Universität München, **2018**.
- [260] V. Busskamp, N. E. Lewis, P. Guye, A. H. M. Ng, S. L. Shipman, S. M. Byrne, N. E. Sanjana, J. Murn, Y. Li, S. Li, M. Stadler, R. Weiss, G. M. Church, *Rapid neurogenesis through transcriptional activation in human stem cells*, *Mol. Syst. Biol.* **2014**, *10*, 760.
- [261] R. D. Bhardwaj, M. A. Curtis, K. L. Spalding, B. A. Buchholz, D. Fink, T. Björk-Eriksson, C. Nordborg, F. H. Gage, H. Druid, P. S. Eriksson, J. Frisé, *Neocortical neurogenesis in humans is restricted to development*, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 12564-12568.
- [262] R. Rahimoff, O. Kosmatchev, A. Kirchner, T. Pfaffeneder, F. Spada, V. Brantl, M. Müller, T. Carell, *5-Formyl- and 5-Carboxydeoxycytidines Do Not Cause Accumulation of Harmful Repair Intermediates in Stem Cells*, *J. Am. Chem. Soc.* **2017**, *139*, 10359-10364.
- [263] C. J. Bostock, D. M. Prescott, J. B. Kirkpatrick, *An evaluation of the double thymidine block for synchronizing mammalian cells at the G1-S border*, *Exp. Cell Res.* **1971**, *68*, 163-168.

- [264] G. Bjursell und P. Reichard, *Effects of Thymidine on Deoxyribonucleoside Triphosphate Pools and Deoxyribonucleic Acid Synthesis in Chinese Hamster Ovary Cells*, *J. Biol. Chem.* **1973**, 248, 3904-3909.
- [265] Q. Du, Z. Wang, V. L. Schramm, *Human DNMT1 transition state structure*, *Proc. Natl. Acad. Sci. USA* **2016**, 113, 2916-2921.
- [266] K.-Y. Wang und C. K. James Shen, *DNA methyltransferase Dnmt1 and mismatch repair*, *Oncogene* **2004**, 23, 7898-7902.
- [267] G. Guo, W. Wang, A. Bradley, *Mismatch repair genes identified using genetic screens in Blm-deficient embryonic stem cells*, *Nature* **2004**, 429, 891-895.
- [268] N. Ding, E. M. Bonham, B. E. Hannon, T. R. Amick, S. B. Baylin, H. M. O'Hagan, *Mismatch repair proteins recruit DNA methyltransferase 1 to sites of oxidative DNA damage*, *J Mol Cell Biol* **2016**, 8, 244-254.
- [269] M. O. Hottiger, P. O. Hassa, B. Lüscher, H. Schüler, F. Koch-Nolte, *Toward a unified nomenclature for mammalian ADP-ribosyltransferases*, *Trends Biochem. Sci.* **2010**, 35, 208-219.
- [270] K. A. Menear, C. Adcock, R. Boulter, X.-I. Cockcroft, L. Copsey, A. Cranston, K. J. Dillon, J. Drzewiecki, S. Garman, S. Gomez, H. Javaid, F. Kerrigan, C. Knights, A. Lau, V. M. Loh, I. T. W. Matthews, S. Moore, M. J. O'Connor, G. C. M. Smith, N. M. B. Martin, *4-[3-(4-Cyclopropanecarbonylpiperazine-1-carbonyl)-4-fluorobenzyl]-2H-phthalazin-1-one: A Novel Bioavailable Inhibitor of Poly(ADP-ribose) Polymerase-1*, *J. Med. Chem.* **2008**, 51, 6581-6591.
- [271] C. K. Donawho, Y. Luo, Y. Luo, T. D. Penning, J. L. Bauch, J. J. Bouska, V. D. Bontcheva-Diaz, B. F. Cox, T. L. DeWeese, L. E. Dillehay, D. C. Ferguson, N. S. Ghoreishi-Haack, D. R. Grimm, R. Guan, E. K. Han, R. R. Holley-Shanks, B. Hristov, K. B. Idler, K. Jarvis, E. F. Johnson, *et al.*, *ABT-888, an Orally Active Poly(ADP-Ribose) Polymerase Inhibitor that Potentiates DNA-Damaging Agents in Preclinical Tumor Models*, *Clin. Cancer. Res.* **2007**, 13, 2728-2737.
- [272] L. Henneman, M. H. van Miltenburg, E. M. Michalak, T. M. Braumuller, J. E. Jaspers, A. P. Drenth, R. de Korte-Grimmerink, E. Gogola, K. Szuhai, A. Schlicker, R. Bin Ali, C. Pritchard, I. J. Huijbers, A. Berns, S. Rottenberg, J. Jonkers, *Selective resistance to the PARP inhibitor olaparib in a mouse model for BRCA1-deficient metaplastic breast cancer*, *Proc. Natl. Acad. Sci. USA* **2015**, 112, 8409-8414.
- [273] M. Haberland, R. L. Montgomery, E. N. Olson, *The many roles of histone deacetylases in development and physiology: implications for disease and therapy*, *Nat. Rev. Genet.* **2009**, 10, 32-42.
- [274] D. M. Vigushin, S. Ali, P. E. Pace, N. Mirsaidi, K. Ito, I. Adcock, R. C. Coombes, *Trichostatin A is a histone deacetylase inhibitor with potent antitumor activity against breast cancer in vivo*, *Clin. Cancer. Res.* **2001**, 7, 971-976.
- [275] C. J. Phiel, F. Zhang, E. Y. Huang, M. G. Guenther, M. A. Lazar, P. S. Klein, *Histone Deacetylase Is a Direct Target of Valproic Acid, a Potent Anticonvulsant, Mood Stabilizer, and Teratogen*, *J. Biol. Chem.* **2001**, 276, 36734-36741.
- [276] J. L. Avalos, K. M. Bever, C. Wolberger, *Mechanism of Sirtuin Inhibition by Nicotinamide: Altering the NAD⁺ Cosubstrate Specificity of a Sir2 Enzyme*, *Mol. Cell* **2005**, 17, 855-868.
- [277] X. Zhao, D. Allison, B. Condon, F. Zhang, T. Gheyi, A. Zhang, S. Ashok, M. Russell, I. MacEwan, Y. Qian, J. A. Jamison, J. G. Luz, *The 2.5 Å Crystal Structure of the SIRT1 Catalytic Domain Bound to Nicotinamide Adenine Dinucleotide (NAD⁺) and an Indole (EX527 Analogue) Reveals a Novel Mechanism of Histone Deacetylase Inhibition*, *J. Med. Chem.* **2013**, 56, 963-969.
- [278] T. Rumpf, M. Schiedel, B. Karaman, C. Roessler, B. J. North, A. Lehotzky, J. Oláh, K. I. Ladwein, K. Schmidtkunz, M. Gajer, M. Pannek, C. Steegborn, D. A. Sinclair, S. Gerhardt, J. Ovádi, M. Schutkowski, W. Sippl, O. Einsle, M. Jung, *Selective Sirt2 inhibition by ligand-induced rearrangement of the active site*, *Nat. Commun.* **2015**, 6, 6263.
- [279] H. Jing und H. Lin, *Sirtuins in epigenetic regulation*, *Chem. Rev.* **2015**, 115, 2350-2375.
- [280] A. Bedalov, S. Chowdhury, J. A. Simon, in *Methods Enzymol.*, Vol. 574, Academic Pres, United States, **2016**, 183-211.

- [281] J. Seidel, C. Klockenbusch, D. Schwarzer, *Investigating Deformylase and Deacylase Activity of Mammalian and Bacterial Sirtuins*, *ChemBioChem* **2016**, *17*, 398-402.
- [282] J. Sun, X. He, Y. Zhu, Z. Ding, H. Dong, Y. Feng, J. Du, H. Wang, X. Wu, L. Zhang, X. Yu, A. Lin, T. McDonald, D. Zhao, H. Wu, W.-K. Hua, B. Zhang, L. Feng, K. Tohyama, R. Bhatia, *et al.*, *SIRT1 Activation Disrupts Maintenance of Myelodysplastic Syndrome Stem and Progenitor Cells by Restoring TET2 Function*, *Cell Stem Cell* **2018**, *23*, 355-369.
- [283] J. Heo, J. Lim, S. Lee, J. Jeong, H. Kang, Y. Kim, J. W. Kang, H. Y. Yu, E. M. Jeong, K. Kim, M. Kucia, S. J. Waigel, W. Zacharias, Y. Chen, I.-G. Kim, M. Z. Ratajczak, D.-M. Shin, *Sirt1 Regulates DNA Methylation and Differentiation Potential of Embryonic Stem Cells by Antagonizing Dnmt3l*, *Cell Rep.* **2017**, *18*, 1930-1945.
- [284] K. Kashiwagi, K. Nimura, K. Ura, Y. Kaneda, *DNA methyltransferase 3b preferentially associates with condensed chromatin*, *Nucleic Acids Res.* **2011**, *39*, 874-888.
- [285] H. Ide, K. Akamatsu, Y. Kimura, K. Michiue, K. Makino, A. Asaeda, Y. Takamori, K. Kubo, *Synthesis and damage specificity of a novel probe for the detection of abasic sites in DNA*, *Biochemistry* **1993**, *32*, 8276-8283.
- [286] K. Kubo, H. Ide, S. S. Wallace, Y. W. Kow, *A novel, sensitive, and specific assay for abasic sites, the most commonly produced DNA lesion*, *Biochemistry* **1992**, *31*, 3703-3708.
- [287] K. P. Roberts, J. A. Sobrino, J. Payton, L. B. Mason, R. J. Turesky, *Determination of Apurinic/Apyrimidinic Lesions in DNA with High-Performance Liquid Chromatography and Tandem Mass Spectrometry*, *Chem. Res. Toxicol.* **2006**, *19*, 300-309.
- [288] J. H. Gommers-Ampt und P. Borst, *Hypermodified bases in DNA*, *FASEB J.* **1995**, *9*, 1034-1042.
- [289] K. Kemmerich, F. A. Dingler, C. Rada, M. S. Neuberger, *Germline ablation of SMUG1 DNA glycosylase causes loss of 5-hydroxymethyluracil- and UNG-backup uracil-excision activities and increases cancer predisposition of Ung^{-/-}Msh2^{-/-} mice*, *Nucleic Acids Res.* **2012**, *40*, 6016-6025.
- [290] A. Masaoka, M. Matsubara, R. Hasegawa, T. Tanaka, S. Kurisu, H. Terato, Y. Ohyama, N. Karino, A. Matsuda, H. Ide, *Mammalian 5-formyluracil-DNA glycosylase. 2. Role of SMUG1 uracil-DNA glycosylase in repair of 5-formyluracil and other oxidized and deaminated base lesions*, *Biochemistry* **2003**, *42*, 5003-5012.
- [291] D. Han, L. Schomacher, K. M. Schüle, M. Mallick, M. U. Musheev, E. Karaulanov, L. Krebs, A. von Seggern, C. Niehrs, *NEIL1 and NEIL2 DNA glycosylases protect neural crest development against mitochondrial oxidative stress*, *eLife* **2019**, *8*, e49044.
- [292] J. W. Fairman, S. R. Wijerathna, M. F. Ahmad, H. Xu, R. Nakano, S. Jha, J. Prendergast, R. M. Welin, S. Flodin, A. Roos, P. Nordlund, Z. Li, T. Walz, C. G. Dealwis, *Structural basis for allosteric regulation of human ribonucleotide reductase by nucleotide-induced oligomerization*, *Nat. Struct. Mol. Biol.* **2011**, *18*, 316-322.
- [293] F. Maley und G. F. Maley, in *Prog. Nucleic Acid Res. Mol. Biol.*, Vol. 39 (Eds.: W. E. Cohn, K. Moldave), Academic Press, **1990**, pp. 49-80.
- [294] A.-H. Teh, M. Kimura, M. Yamamoto, N. Tanaka, I. Yamaguchi, T. Kumasaka, *The 1.48 Å Resolution Crystal Structure of the Homotetrameric Cytidine Deaminase from Mouse*, *Biochemistry* **2006**, *45*, 7825-7833.
- [295] U. Hardeland, M. Bentele, J. Jiricny, P. Schär, *Separating Substrate Recognition from Base Hydrolysis in Human Thymine DNA Glycosylase by Mutational Analysis*, *J. Biol. Chem.* **2000**, *275*, 33449-33456.
- [296] L. S. Pidugu, J. W. Flowers, C. T. Coey, E. Pozharski, M. M. Greenberg, A. C. Drohat, *Structural Basis for Excision of 5-Formylcytosine by Thymine DNA Glycosylase*, *Biochemistry* **2016**, *55*, 6205-6208.
- [297] A. Kirchner, Masterarbeit, Ludwig-Maximilians-Universität München, **2014**.
- [298] U. Hardeland, R. Steinacher, J. Jiricny, P. Schär, *Modification of the human thymine-DNA glycosylase by ubiquitin-like proteins facilitates enzymatic turnover*, *EMBO J.* **2002**, *21*, 1456-1464.

- [299] P. Neddermann, P. Gallinari, T. Lettieri, D. Schmid, O. Truong, J. J. Hsuan, K. Wiebauer, J. Jiricny, *Cloning and Expression of Human G/T Mismatch-specific Thymine-DNA Glycosylase*, *J. Biol. Chem.* **1996**, *271*, 12767-12774.
- [300] O. Kosmatchev, Dissertation, Ludwig-Maximilians-Universität München, **2018**.
- [301] J. Choi, A. J. Huebner, K. Clement, R. M. Walsh, A. Savol, K. Lin, H. Gu, B. Di Stefano, J. Brumbaugh, S.-Y. Kim, J. Sharif, C. M. Rose, A. Mohammad, J. Odajima, J. Charron, T. Shioda, A. Gnirke, S. Gygi, H. Koseki, R. I. Sadreyev, *et al.*, *Prolonged Mek1/2 suppression impairs the developmental potential of embryonic stem cells*, *Nature* **2017**, *548*, 219-223.
- [302] M. B. Stadler, R. Murr, L. Burger, R. Ivaneck, F. Lienert, A. Schöler, E. v. Nimwegen, C. Wirbelauer, E. J. Oakeley, D. Gaidatzis, V. K. Tiwari, D. Schübeler, *DNA-binding factors shape the mouse methylome at distal regulatory regions*, *Nature* **2011**, *480*, 490-495.
- [303] F. von Meyenn, Rebecca V. Berrens, S. Andrews, F. Santos, Amanda J. Collier, F. Krueger, R. Osorno, W. Dean, Peter J. Rugg-Gunn, W. Reik, *Comparative Principles of DNA Methylation Reprogramming during Human and Mouse In Vitro Primordial Germ Cell Specification*, *Dev. Cell* **2016**, *39*, 104-115.
- [304] Sophie M. Morgani, Maurice A. Canham, J. Nichols, Alexei A. Sharov, Rosa P. Migueles, Minoru S. H. Ko, Joshua M. Brickman, *Totipotent Embryonic Stem Cells Arise in Ground-State Culture Conditions*, *Cell Rep.* **2013**, *3*, 1945-1957.
- [305] K. Shirane, K. Kurimoto, Y. Yabuta, M. Yamaji, J. Satoh, S. Ito, A. Watanabe, K. Hayashi, M. Saitou, H. Sasaki, *Global Landscape and Regulatory Principles of DNA Methylation Reprogramming for Germ Cell Specification by Mouse Pluripotent Stem Cells*, *Dev. Cell* **2016**, *39*, 87-103.
- [306] Y. Zhang, Y. Xiang, Q. Yin, Z. Du, X. Peng, Q. Wang, M. Fidalgo, W. Xia, Y. Li, Z.-a. Zhao, W. Zhang, J. Ma, F. Xu, J. Wang, L. Li, W. Xie, *Dynamic epigenomic landscapes during early lineage specification in mouse embryos*, *Nat. Genet.* **2018**, *50*, 96-105.
- [307] M. Bachman, S. Uribe-Lewis, X. Yang, H. E. Burgess, M. Iurlaro, W. Reik, A. Murrell, S. Balasubramanian, *5-Formylcytosine can be a stable DNA modification in mammals*, *Nat. Chem. Biol.* **2015**, *11*, 555-557.
- [308] D.-M. Franchini, K.-M. Schmitz, S. K. Petersen-Mahrt, *5-Methylcytosine DNA Demethylation: More Than Losing a Methyl Group*, *Annu. Rev. Genet.* **2012**, *46*, 419-441.
- [309] B. Hendrich, U. Hardeland, H.-H. Ng, J. Jiricny, A. Bird, *The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites*, *Nature* **1999**, *401*, 301-304.
- [310] C. Bertoni, A. Rustagi, T. A. Rando, *Enhanced gene repair mediated by methyl-CpG-modified single-stranded oligonucleotides*, *Nucleic Acids Res.* **2009**, *37*, 7468-7482.
- [311] J. Chaudhuri und F. W. Alt, *Class-switch recombination: interplay of transcription, DNA deamination and DNA repair*, *Nat. Rev. Immunol.* **2004**, *4*, 541-552.
- [312] S. Schanz, D. Castor, F. Fischer, J. Jiricny, *Interference of mismatch and base excision repair during the processing of adjacent U/G mispairs may play a key role in somatic hypermutation*, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 5593-5598.
- [313] S. Bregenhorn, L. Kallenberger, M. Artola-Borán, J. Peña-Díaz, J. Jiricny, *Non-canonical uracil processing in DNA gives rise to double-strand breaks and deletions: relevance to class switch recombination*, *Nucleic Acids Res.* **2016**, *44*, 2691-2705.
- [314] A. Marx und A. Alian, *The First Crystal Structure of a dTTP-bound Deoxycytidylate Deaminase Validates and Details the Allosteric-Inhibitor Binding Site*, *J. Biol. Chem.* **2015**, *290*, 682-690.
- [315] A. Somasekaram, A. Jarmuz, A. How, J. Scott, N. Navaratnam, *Intracellular localization of human cytidine deaminase. Identification of a functional nuclear localization signal*, *J. Biol. Chem.* **1999**, *274*, 28405-28412.
- [316] C. B. Millar, J. Guy, O. J. Sansom, J. Selfridge, E. MacDougall, B. Hendrich, P. D. Keightley, S. M. Bishop, A. R. Clarke, A. Bird, *Enhanced CpG Mutability and Tumorigenesis in MBD4-Deficient Mice*, *Science* **2002**, *297*, 403-405.
- [317] E. Li, T. H. Bestor, R. Jaenisch, *Targeted mutation of the DNA methyltransferase gene results in embryonic lethality*, *Cell* **1992**, *69*, 915-926.

- [318] M. Hooper, K. Hardy, A. Handyside, S. Hunter, M. Monk, *HPRT-deficient (Lesch–Nyhan) mouse embryos derived from germline colonization by cultured cells*, *Nature* **1987**, 326, 292-295.
- [319] M. Moser, B. Nieswandt, S. Ussar, M. Pozgajova, R. Fässler, *Kindlin-3 is essential for integrin activation and platelet aggregation*, *Nat. Med.* **2008**, 14, 325-330.
- [320] E. Robanus-Maandag, M. Dekker, M. van der Valk, M.-L. Carrozza, J.-C. Jeanny, J.-H. Dannenberg, A. Berns, H. te Riele, *p107 is a suppressor of retinoblastoma development in pRb-deficient mice*, *Genes Dev.* **1998**, 12, 1599-1609.
- [321] B. Kavli, O. Sundheim, M. Akbari, M. Otterlei, H. Nilsen, F. Skorpen, P. A. Aas, L. Hagen, H. E. Krokan, G. Slupphaug, *hUNG2 is the major repair enzyme for removal of uracil from U:A matches, U:G mismatches, and U in single-stranded DNA, with hSMUG1 as a broad specificity backup*, *J. Biol. Chem.* **2002**, 277, 39926-39936.
- [322] H. S. Pettersen, O. Sundheim, K. M. Gilljam, G. Slupphaug, H. E. Krokan, B. Kavli, *Uracil–DNA glycosylases SMUG1 and UNG2 coordinate the initial steps of base excision repair by distinct mechanisms*, *Nucleic Acids Res.* **2007**, 35, 3879-3892.
- [323] A. Szwagierczak, C. S. Schmidt, S. Bultmann, F. Spada, H. Leonhardt, *Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA*, *Nucleic Acids Res.* **2010**, 38, e181.
- [324] K. X. Weiner, R. S. Weiner, F. Maley, G. F. Maley, *Primary structure of human deoxycytidylate deaminase and overexpression of its functional protein in Escherichia coli*, *J. Biol. Chem.* **1993**, 268, 12983-12989.
- [325] G. F. Maley, A. P. Lobo, F. Maley, *Properties of an affinity-column-purified human deoxycytidylate deaminase*, *Biochim. Biophys. Acta* **1993**, 1162, 161-170.

7 Anhang: Zusatzmaterialien der veröffentlichten Arbeiten

7.1 Zusatzmaterialien zu Abschnitt 3.1.1

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

* Beitrag der Autoren zu gleichen Teilen

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

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).

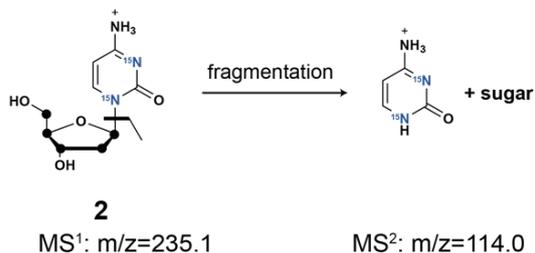
compound	Precursor ion (m/z)	MS1 Resolution	Product Ion (m/z)	MS2 Resolution	Dwell time [ms]	CE (V)	CAV (V)	Polarity
Time segment 1.5-3.3 min								
sh-cadC	279.10	Wide	158.03	Wide	50	5	5	Positive
$[^{15}N_2]$ -cadC	274.08	Wide	158.03	Wide	20	5	5	Positive
cadC	272.09	Wide	156.04	Wide	20	5	5	Positive
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

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).

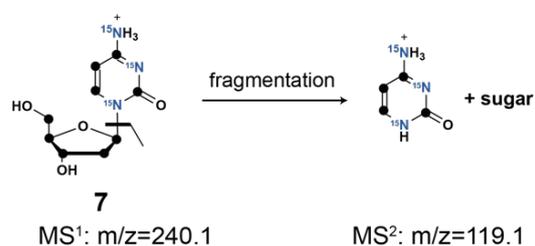
compound	Precursor ion (m/z)	MS1 Resolution	Product Ion (m/z)	MS2 Resolution	Dwell time [ms]	CE (V)	CAV (V)	Polarity
Time segment 1.5-3.3 min								
$[^{13}\text{C}_9, ^{15}\text{N}_3]\text{-dC}$	240.10	Unit	119.10	Unit	70	30	5	Positive
dC	228.1	Wide	112.10	Wide	70	1	3	Positive
Time segment 3.3-4.8 min								
dU	227.07	Wide	184.06	Wide	80	5	5	Negative
$[^{13}\text{C}_9, ^{15}\text{N}_2]\text{-dU}$	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
$[^{13}\text{C}_9, ^{15}\text{N}_2]\text{-dT}$	254.12	Unit	133.06	Unit	70	3	5	Positive
$[\text{D}_3]\text{-dT}$	246.10	Unit	130.05	Unit	30	3	5	Positive

Supplementary Figures

MS-signal for *in vivo* deformylated fdC

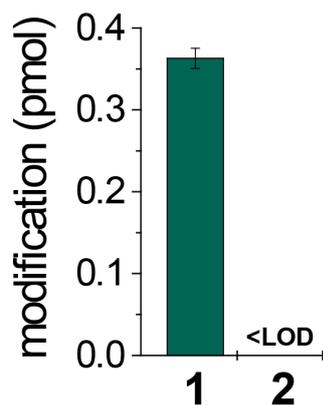


MS-signal for internal standard

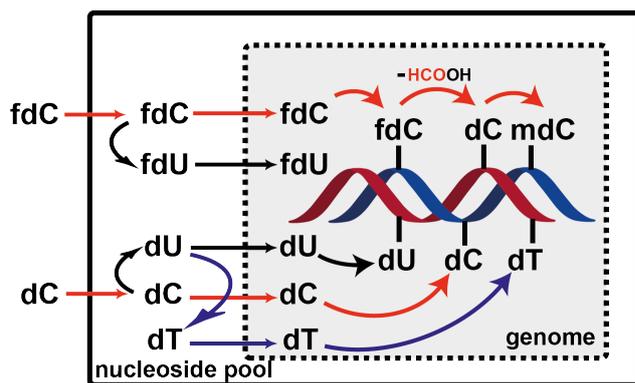


Supplementary Figure 1: Characteristic MS transition of [¹³C₅][¹⁵N₂]-dC and [¹³C₉][¹⁵N₃]-dC.

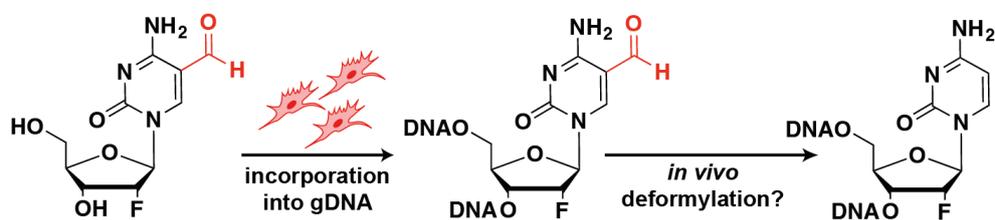
Fingerprint MS-fragmentation pathways and detected MS signals of **2** and **7** used to track isotopically labelled dC and its derivatives in mESC cultures.



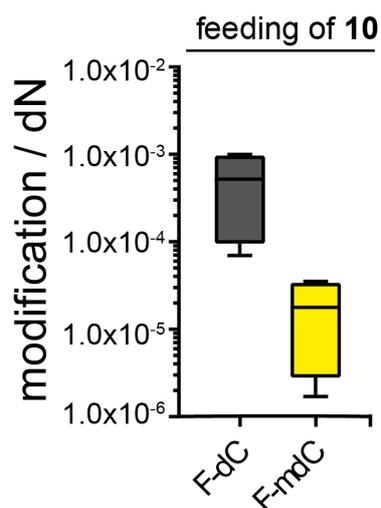
Supplementary Figure 2: Quantitative data of isotopically labelled pyrimidine derivatives in the soluble nucleoside pool of mESCs after feeding [¹³C₅][¹⁵N₂]-fdC shows its incorporation and no formation of [¹³C₅][¹⁵N₂]-dC. LC-MS/MS analysis of the soluble nucleoside pool from J1 mESCs cultured in the 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.



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

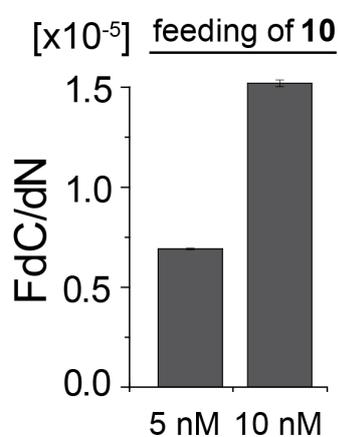


Supplementary Figure 4: Schematic overview of the feeding experiment with (*R*)-2'-F-fdC to mESCs. Feeding of **15** results in the incorporation of (*R*)-2'-F-fdC into genomic DNA and is converted *in vivo* to (*R*)-2'-F-dC by a C-C bond cleavage reaction.



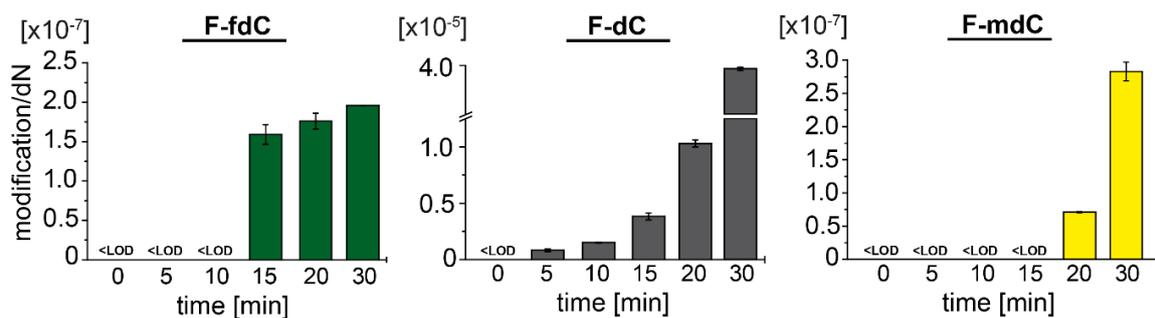
Supplementary Figure 5: Quantitative data of fluorinated pyrimidine derivatives after treatment with 10 show incorporation of 10 and 3% re-methylation.

LC-MS/MS analysis of genomic DNA from J1 mES cells cultured in the presence of $1.0 \mu\text{M}$ (*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.



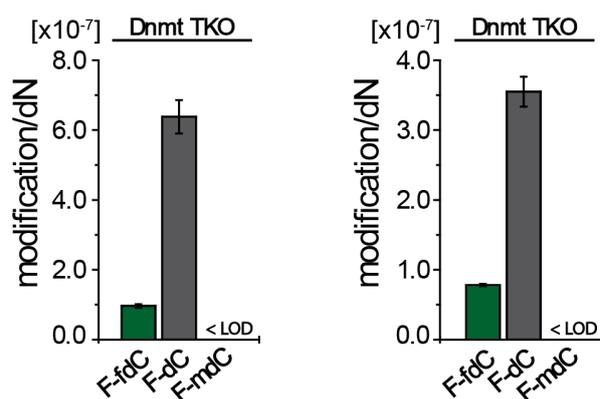
Supplementary Figure 6: Quantification data of incorporated (*R*)-2'-F-dC after feeding of 10 at two concentrations. Levels similar to those resulting from feeding of (*R*)-2'-F-fdC were observed.

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.



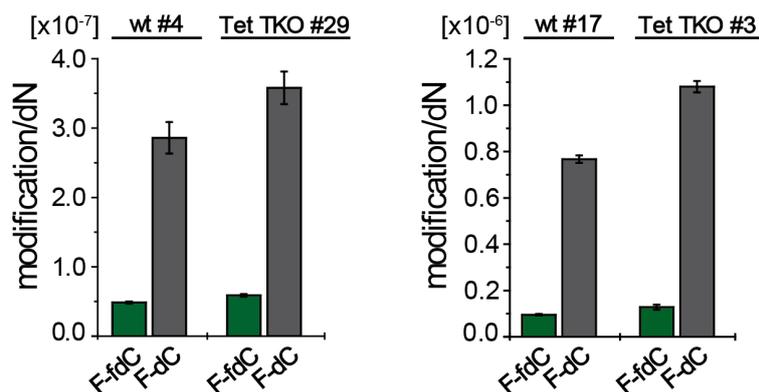
Supplementary Figure 7: Time dependent co-feeding with 15 and 10 show incorporation of 10 already after 5 min, 15 after 15 min and 11, resulting from both compounds, after 20 min.

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 μ M) and (*R*)-2'-F-dC (1.0 μ M) for 0, 5, 10, 15, 20 and 30 min. Mean values and s.d. of technical triplicates from single cultures are shown.



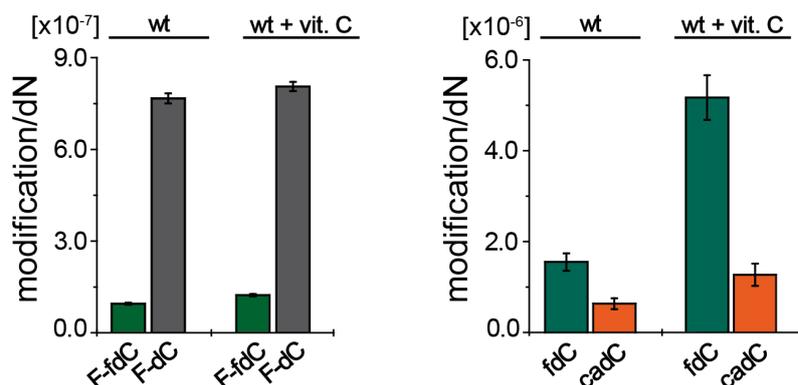
Supplementary Figure 8: Catalytically active DNA methyltransferases are required for methylation of (*R*)-2'-F-dC, but not for deformylation of (*R*)-2'-F-fdC (related to Fig. 4d).

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.



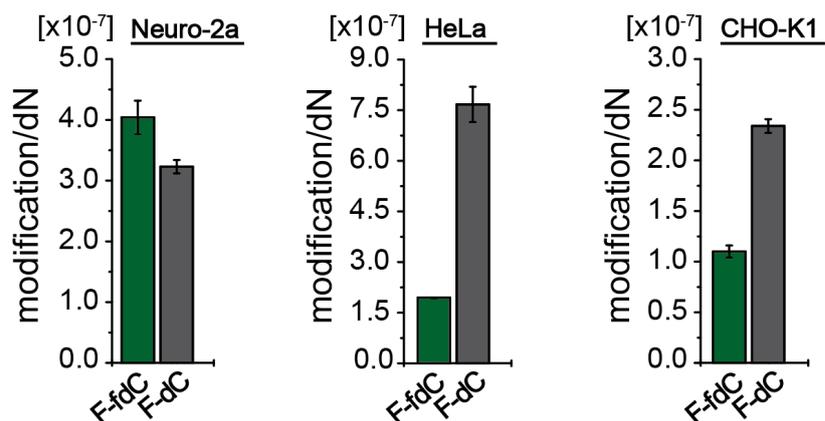
Supplementary Figure 9: Tet enzymes are not required for deformylation of (*R*)-2'-F-fdC (related to Fig. 4e).

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.



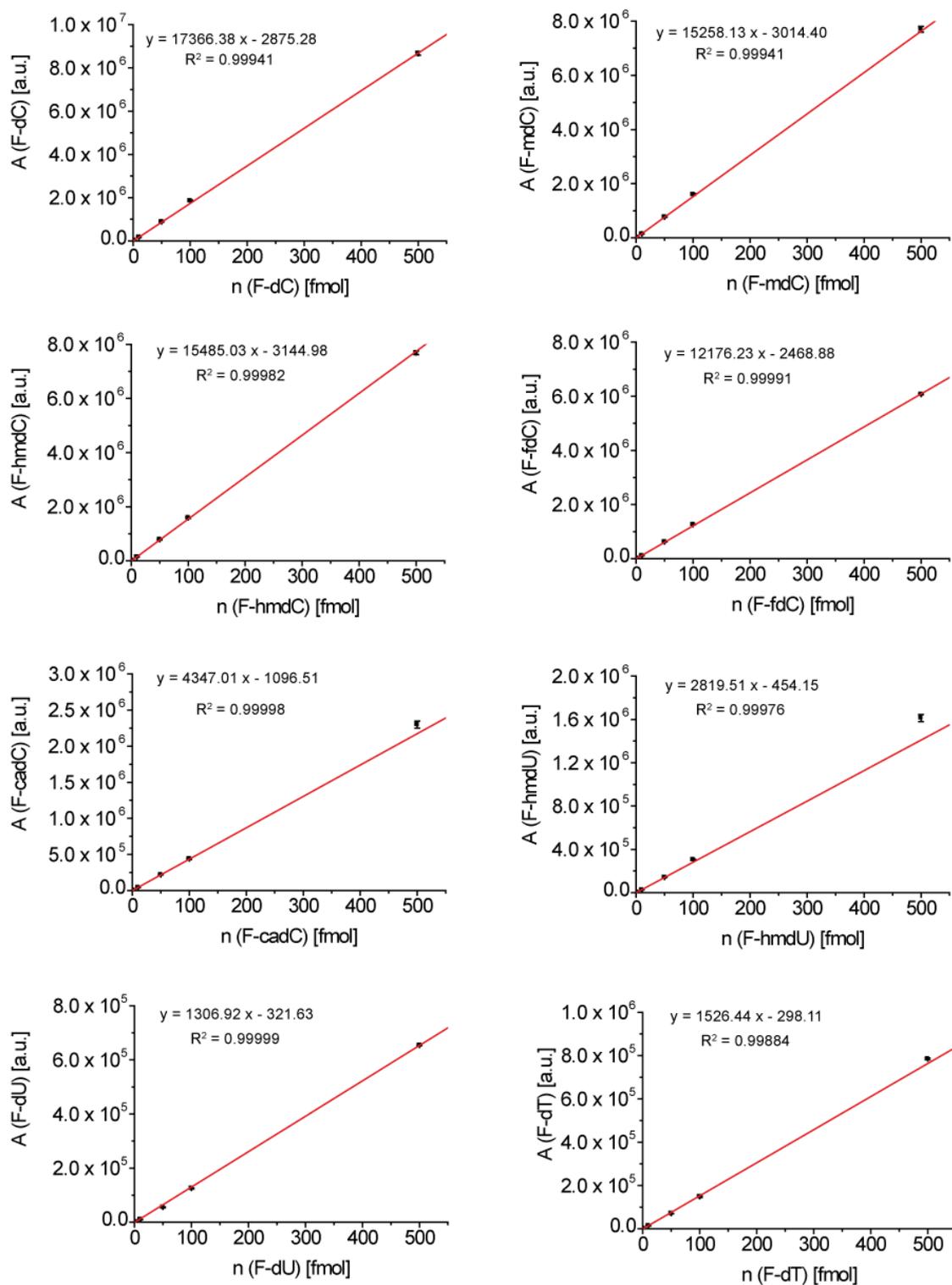
Supplementary Figure 10: Deformylation of (*R*)-2'-F-fdC is not affected by treatment with ascorbic acid-2-phosphate.

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.

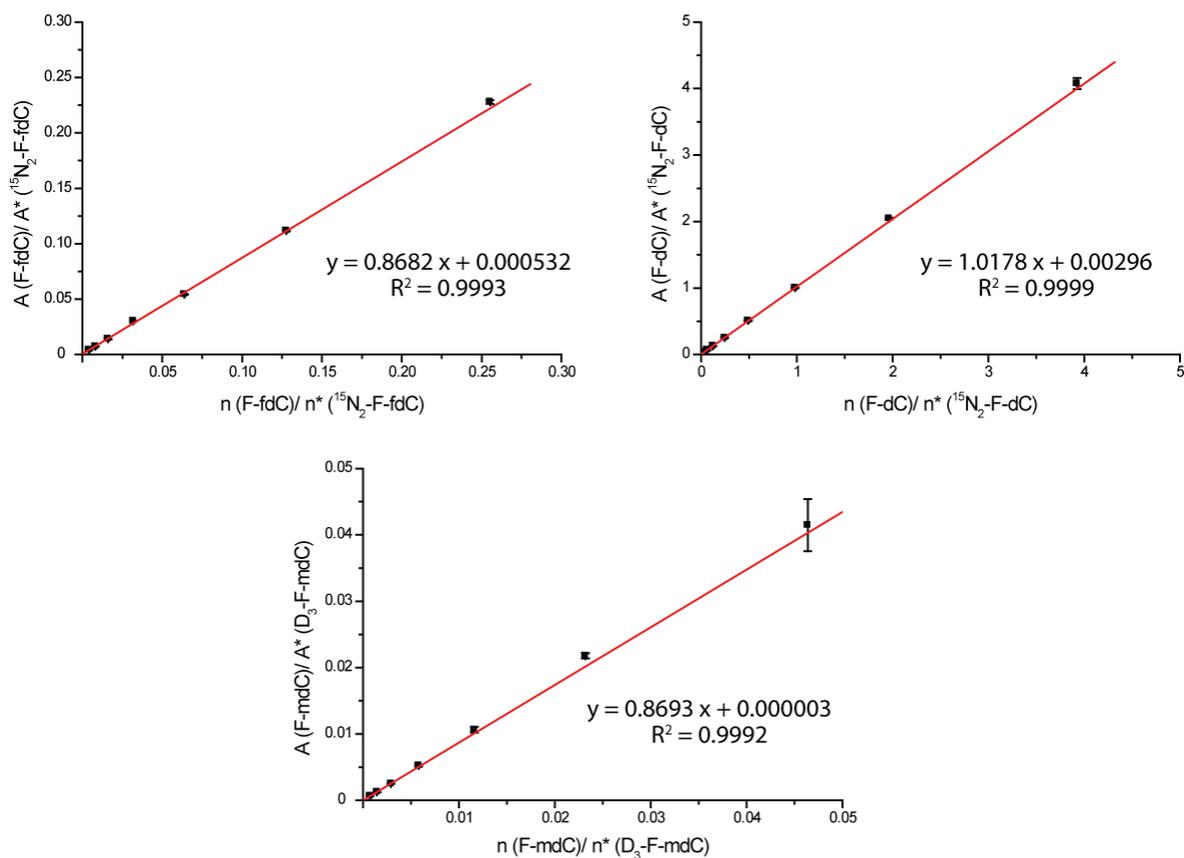


Supplementary Figure 11: Deformylation of (*R*)-2'-F-fdC to (*R*)-2'-F-dC occur also in non-pluripotent mouse and human cell lines (related to Fig. 4f).

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.



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



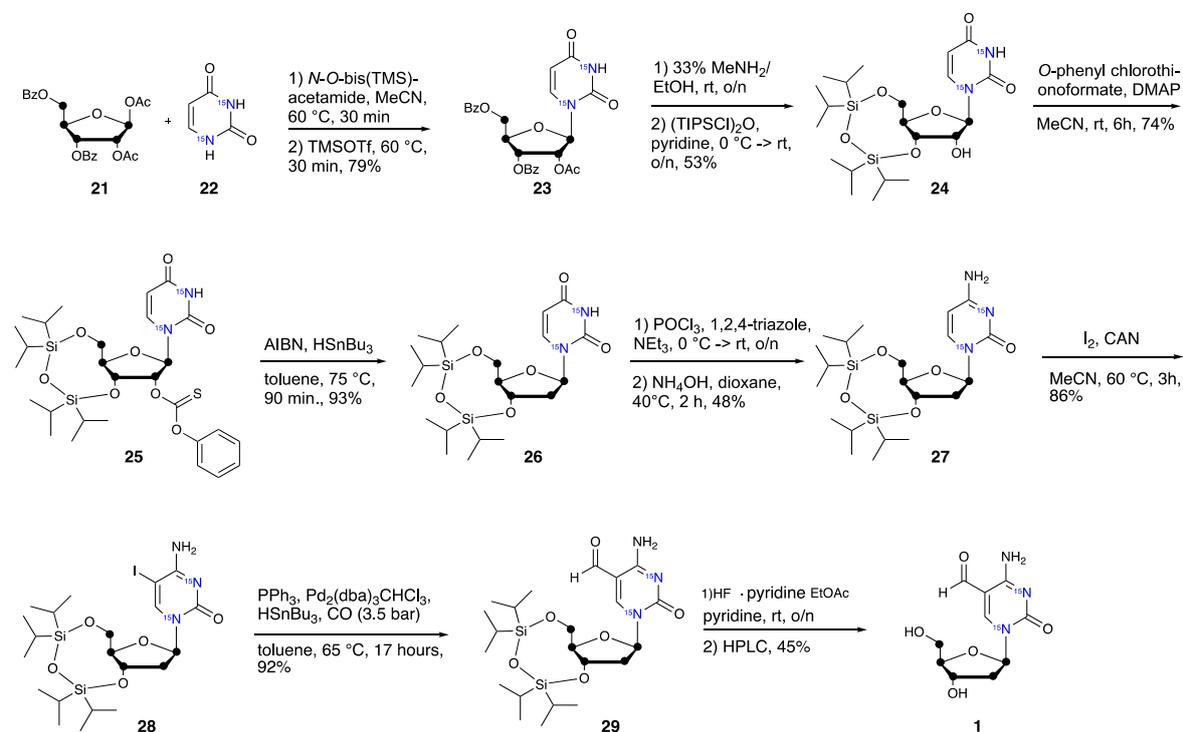
Supplementary Figure 13: Internal calibration curves for the exact quantification of (R)-2'-F-fdC, (R)-2'-F-dC and (R)-2'-F-mdC with the corresponding linear equation and coefficient of determination.

Synthetic procedures

General information

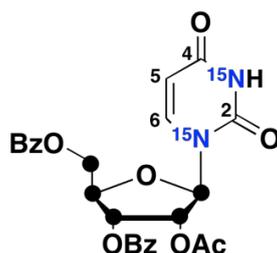
Isotopically labelled **7** was bought from *B.A.C.H. UG*, 2'-deoxy-2'-(*R*)-fluorocytidine **10** was purchased from *Carbosynth*. Compounds **8** and **9** resulted from the feeding experiment of **7**, compounds **11** - **14** resulted from the feeding of **10**. **15** and **30** were synthesized 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'-¹³C₅, ¹⁵N₂)cytidine (**1**)



Scheme 1: Synthesis of ¹³C₅-¹⁵N₂-labelled fdC **1**.

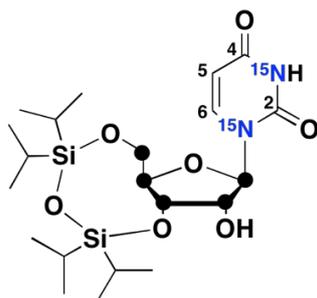
2'-O-acetyl-3',5'-di-O-benzoyl(1',2',3',4',5'-¹³C₅, N¹,N³-¹⁵N₂)-uridine (23)



The ¹³C ribose fragment **21** and ¹⁵N₂ urea **22** were synthesized according to the literature.^{3,4} Bis-benzoyl and bis-acetyl protected ¹³C ribose **21** (2.00 g, 4.47 mmol, 1.0 eq), ¹⁵N₂-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 NaHCO₃ (75 mL) and extracted with DCM (3 x 75 mL). The combined organic layers were dried over Na₂SO₄ 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.

¹H-NMR (400 MHz, CDCl₃, 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-H} = 3.2 Hz, 1H, 5-H), 6.41 (d, ¹J_{C-H} = 169.1 Hz, 1H, 1'-H), 5.96 (d, ¹J_{C-H} = 150.3 Hz, 1H, 3'-H), 5.74 (d, ¹J_{C-H} = 153.0 Hz, 2'-H), 5.57 5.74 (m, 1H, 6-H), 4.97 (d, ¹J_{C-H} = 154.6 Hz, 1H, 4'-H), 4.82 (d, ¹J_{13C-H} = 150.1 Hz, 1H, 5'-H), 2.06 (s, 3H, Ac-CH₃). **¹³C-NMR (101 MHz, CDCl₃, ppm):** δ = 169.7 (Ac-C=O), 166.0 (Bz-C=O), 165.3 (Bz-C=O), 162.4 (d, ¹J_{C-N} = 9.5 Hz, C-4), 150.1 (t, ¹J_{C-N} = 19.4 Hz, C-2), 139.2 (d, ¹J_{C-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-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-CH₃). **¹⁵N-NMR (41 MHz, CDCl₃, ppm):** δ = -224.4, -224.5. **HRMS (ESI⁺):** calc. for C₂₀¹³C₅H₂₆N¹⁵N₂O₉⁺ [M+NH₄]⁺: 519.1772, found: 519.1784. **HRMS (ESI⁻):** calc. for C₂₀¹³C₅H₂₁¹⁵N₂O₉ [M-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).

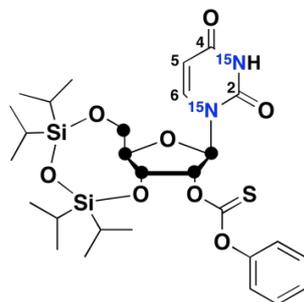
3',5'-(1,1,3,3,-tetraisopropyldisiloxane-1,3-diyl)-(1',2',3',4',5'-¹³C₅, N¹,N³-¹⁵N₂)-uridine (24)



First, **23** (1.75 g, 3.49 mmol, 1.0 eq) was deprotected by stirring the compound in a solution of MeNH₂ 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 NH₄Cl (50 mL), extracted twice with EtOAc (50 mL) and the organic layers dried over Na₂SO₄. After removal of the solvent *in vacuo*, pure product was isolated after column chromatography (30% EtOAc/Hex → 40% EtOAc/Hex), which yielded 920 mg (1.86 mmol, 53%) of **24** as an off white foam.

¹H-NMR (599 MHz, CDCl₃, ppm): δ = 9.23 (d, ²J_{N-H} = 90.9 Hz, 1H, N-H), 7.72 (td, ³J = 8.1 Hz, ²J_{N-H} = 2.2 Hz, 1H, 5-H), 5.88 (d, ¹J_{C-H} = 174.6 Hz, 1H, 1'-H), 5.69 (ddd, ³J = 7.8 Hz, ³J_{N-H} = 4.6 Hz, ³J_{N-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(CH₃)₂, Si-CH). **¹³C-NMR (151 MHz, CDCl₃, ppm):** δ = 163.3 (d, ¹J_{C-N} = 9.5 Hz, C-4), 150.0 (t, ¹J_{C-N} = 18.4 Hz, 2-C), 139.9 (d, ¹J_{C-N} = 12.7 Hz, C-5), 105.4 (d, ²J_{C-N} = 34.7 Hz, C-6), 90.9 (dd, ¹J_{C-C} = 42.6 Hz, ¹J_{C-N} = 11.0 Hz, 1'-C), 81.9 (t, J = 42.3 Hz, 2'-C), 75.1 (t, ¹J_{C-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(CH₃)), 13.4 - 12.8 (4 x Si-C(CH₃)). **¹⁵N-NMR (41 MHz, CDCl₃, ppm):** δ = -224.4, -224.5. **HRMS (ESI⁺):** calc. for C₁₆¹³C₅H₃₉¹⁵N₂O₇Si₂⁺ [M+H]⁺: 494.2399, found: 494.2412. **HRMS (ESI⁻):** calc. for C₁₆¹³C₅H₃₇¹⁵N₂O₇Si₂⁻ [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).

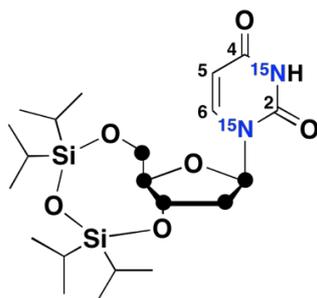
3',5'-(1,1,3,3,-tetraisopropylidisiloxane-1,3-diyl)-2'-O-phenoxythiocarbonyl-(1',2',3',4',5'-¹³C₅, N¹,N³-¹⁵N₂)-uridine (25)



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 μ L, 0.86 mmol, 1.5 eq) and 4-(diethylamino)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 NaHCO₃ 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 Na₂SO₄ 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.

¹H-NMR (599 MHz, CDCl₃, ppm): δ = 8.27 (d, ²*J*_{N-H} = 91.1 Hz, 1H, N-H), 7.71 (dt, ³*J* = 8.1 Hz, ²*J*_{N-H} = 2.2 Hz, 1H, 5-H), 7.42 (t, ³*J* = 8.0 Hz, 2H, H_{Ar}), 7.30 (t, ³*J* = 7.6 Hz, 1H, H_{Ar}), 7.12 (d, ³*J* = 8.3 Hz, 2H, H_{Ar}), 6.15 (td, ¹*J*_{C-H} = 164.2 Hz, ³*J* = 5.5 Hz, 1H, 2'-H), 6.07 (d, ¹*J*_{C-H} = 174.8 Hz, 1H, 1'-H), 5.72 (ddd, ³*J* = 7.7 Hz, ³*J*_{N-H} = 4.9 Hz, ³*J*_{N-H} = 2.5 Hz, 1H, 6-H), 4.55 (ddt, ¹*J*_{C-H} = 142.4 Hz, ³*J* = 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(CH₃)₂, Si-CH). **¹³C-NMR (151 MHz, CDCl₃, ppm):** δ = 193.7 ((O)₂C=S), 162.5 (d, ¹*J*_{C-N} = 9.5 Hz, C-4), 153.4 (C_{Ar}), 149.3 (t, ¹*J*_{C-N} = 18.5 Hz, 2-C), 139.4 (d, ¹*J*_{C-N} = 12.8 Hz, C-5), 129.6 (2 x C_{Ar}), 126.7 (C_{Ar}), 121.7 (2 x C_{Ar}), 102.3 (d, ²*J*_{C-N} = 7.5 Hz, C-6), 88.6 (dd, ¹*J*_{C-C} = 53.5 Hz, ¹*J*_{C-N} = 12.2 Hz, 1'-C), 83.7 (t, *J* = 41.9 Hz, 2'-C), 82.2 (t, ¹*J*_{C-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(CH₃)), 13.4 - 12.4 (4 x Si-C(CH₃)). **¹⁵N-NMR (41 MHz, CDCl₃, ppm):** δ = -223.5, -223.6. **HRMS (ESI⁺):** calc. for C₂₃¹³C₅H₄₃¹⁵N₂O₈SSi₂⁺ [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).

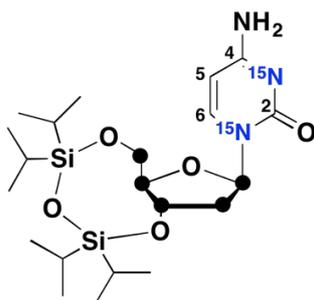
3',5'-(1,1,3,3,-tetraisopropylidisiloxane-1,3-diyl)-2'-deoxy-(1',2',3',4',5'-¹³C₅, N¹,N³-¹⁵N₂)-uridine (26)



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 μ L, 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 Na₂SO₄. Flash-chromatography of the crude (25% EtOAc/Hex --> 30% EtOAc/Hex) yielded **26** as a white foam (180 mg, 0.38 mmol, 93%).

¹H-NMR (599 MHz, CDCl₃, ppm): δ = 8.43 (d, ²J_{N-H} = 91.1 Hz, 1H, N-H), 7.71 (dt, ³J = 8.1 Hz, ²J_{N-H} = 2.2 Hz, 1H, 5-H), 6.19 (dd, ¹J_{C-H} = 175 Hz, ³J = 6.5 Hz, 1H, 1'-H), 5.70 - 5.67 (m, 1H, 6-H), 4.57 (d, ¹J_{C-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(CH₃)₂, Si-CH). **¹³C-NMR (151 MHz, CDCl₃, ppm):** δ = 162.9 (d, ¹J_{C-N} = 9.4 Hz, C-4), 149.8 (t, ¹J_{C-N} = 18.5 Hz, 2-C), 139.6 (d, ¹J_{C-N} = 12.9 Hz, C-5), 101.6 (d, ²J_{C-N} = 7.0 Hz, C-6), 85.1 (t, ¹J_{C-C} = 41.9 Hz, 4'-C), 84.3 (dd, ¹J_{C-C} = 36.4 Hz, ¹J_{C-N} = 10.6 Hz, 1'-C), 67.0 (t, ¹J_{C-C} = 39.2 Hz, 3'-C), 60.0 (d, J = 37.0 Hz, 5'-C), 39.9 (t, ¹J_{C-C} = 37 Hz, 2'-C), 17.5 - 16.7 (Si-C(CH₃)), 13.4 - 12.4 (4 x Si-C(CH₃)). **¹⁵N-NMR (41 MHz, CDCl₃, ppm):** δ = -224.1, -224.2. **HRMS (ESI⁺):** calc. for C₁₆¹³C₅H₃₉¹⁵N₂O₆Si₂⁺ [M+H]⁺: 478.2450, found: 478.2451. **HRMS (ESI⁻):** calc. for C₁₆¹³C₅H₃₇¹⁵N₂O₆Si₂⁻ [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).

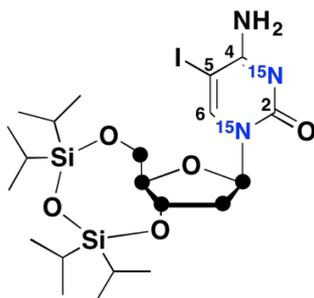
3',5'-(1,1,3,3,-tetraisopropylidisiloxane-1,3-diyl)-2'-deoxy-(1',2',3',4',5'-¹³C₅, N¹,N³-¹⁵N₂)-cytidine (27)



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, POCl₃ (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 Na₂SO₄ and the volatiles were removed *in vacuo*. This residue was then diluted with dioxane (15 mL), charged with NH₄OH (4 mL, 25%) and stirred at 40 °C for three hours. The mixture was subsequently neutralized with saturated aqueous NH₄Cl (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.

¹H-NMR (800 MHz, CDCl₃, ppm): δ = 7.90 (ddd, ³J = 7.3 Hz, ²J_{N-H} = 2.7 Hz, 1.1 Hz, 1H, 5-H), 6.17 (d, ¹J_{C-H} = 174.7 Hz, ³J = 7.7 Hz, 1H, 1'-H), 5.64 (dd, ³J = 7.4 Hz, ²J_{N-H} = 3.4 Hz, 6-H), 4.46 (d, ¹J_{C-H} = 137.8 Hz, 1H, 3'-H), 4.25 (d, ¹J_{C-H} = 149.2 Hz, 1H, 4'-H), 4.10 (d, ¹J_{C-H} = 139.7 Hz, 1H, 5'-H), 3.86 (d, ¹J_{C-H} = 143.6 Hz, 1H, 5'-H), 2.61 (d, ¹J_{C-H} = 134.2 Hz, 1H, 2'-H), 2.61 (d, ¹J_{C-H} = 134.2 Hz, 1H, 2'-H), 2.41 (d, ¹J_{C-H} = 143.2 Hz, 1H, 2'-H), 1.10 - 0.98 (m, 28H, Si(CH₃)₂, Si-CH). **¹³C-NMR (201 MHz, CDCl₃, ppm):** δ = 164.6 (d, ¹J_{C-N} = 6.2 Hz, C-4), 154.6 (t, ¹J_{C-N} = 12.2 Hz, 2-C), 140.3 (d, ¹J_{C-N} = 12.8 Hz, C-5), 92.0 (d, ²J_{C-N} = 7.0 Hz, C-6), 84.1 (d, ¹J_{C-C} = 40.4 Hz, 1'-C), 83.9 (t, ¹J_{C-C} = 42.8 Hz, 4'-C), 65.7 (t, ¹J_{C-C} = 36.9, 3'-C), 59.9 (d, J = 43.0 Hz, 5'-C), 38.9 (t, ¹J_{C-C} = 36.9 Hz, 2'-C), 16.5 - 15.8 (Si-C(CH₃)), 12.4 - 11.4 (4 x Si-C(CH₃)). **¹⁵N-NMR (41 MHz, CDCl₃, ppm):** δ = -217.7, -217.9. **HRMS (ESI⁺):** calc. for C₁₆¹³C₅H₄₀N¹⁵N₂O₅Si₂⁺ [M+H]⁺: 477.2609, found: 477.2610. **HRMS (ESI⁻):** calc. for C₁₆¹³C₅H₃₈N¹⁵N₂O₅Si₂⁻ [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).

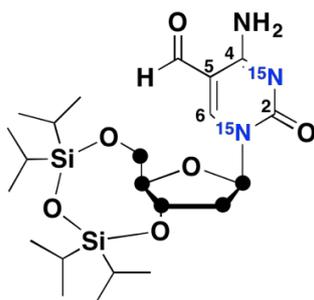
3',5'-(1,1,3,3,-tetraisopropylidisiloxane-1,3-diyl)-5-iodo-2'-deoxy-(1',2',3',4',5'-¹³C₅, N¹,N²-¹⁵N₂)-cytidine (28)



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 Na₂S₂O₃ solution (5 mL). After extraction with DCM (3 x 25 mL), the organic phases were dried over Na₂SO₄ 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%).

¹H-NMR (599 MHz, CDCl₃, ppm): δ = 8.07 (sd, ²J_{N-H} = 2.4 Hz, 1H, 5-H), 5.98 (dd, ¹J_{C-H} = 175.8 Hz, ³J = 6.8 Hz, 1H, 1'-H), 4.49 (d, ¹J_{C-H} = 126.2 Hz, 1H, 3'-H), 4.29 (d, ¹J_{C-H} = 146.0 Hz, 1H, 4'-H), 4.13 (d, ¹J_{C-H} = 137.9 Hz, 1H, 5'-H), 3.86 (d, ¹J_{C-H} = 136.7 Hz, 1H, 5'-H), 2.63 (d, ¹J_{C-H} = 133.4 Hz, 1H, 2'-H), 2.44 (d, ¹J_{C-H} = 136.0 Hz, 1H, 2'-H), 1.11 - 0.99 (m, 28H, Si(CH₃)₂, Si-CH). **¹³C-NMR (151 MHz, CDCl₃, ppm):** δ = 163.7 (d, ¹J_{C-N} = 5.2 Hz, C-4), 154.6 (t, ¹J_{C-N} = 10.5 Hz, 2-C), 146.5 (d, ¹J_{C-N} = 12.6 Hz, C-5), 85.7 - 84.9 (1'-C + 4'-C), 66.5 (t, ¹J_{C-C} = 39.2, 3'-C), 59.7 (d, J = 43.2 Hz, 5'-C), 55.3 (C-6), 39.9 (t, ¹J_{C-C} = 37.0 Hz, 2'-C), 17.8 - 16.9 (8 x Si-C(CH₃)), 13.5 - 12.3 (4 x Si-C(CH₃)). **¹⁵N-NMR (41 MHz, CDCl₃, ppm):** δ = -209.7, -209.9. **HRMS (ESI⁺):** calc. for C₁₆¹³C₅H₃₉I¹⁵N₂O₅Si₂⁺ [M+H]⁺: 603.1576, found: 603.1581. **HRMS (ESI⁻):** calc. for C₁₆¹³C₅H₃₇I¹⁵N₂O₅Si₂⁻ [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).

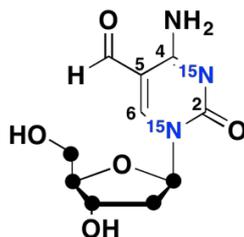
3',5'-(1,1,3,3,-tetraisopropylidisiloxane-1,3-diyl)-5-formyl-2'-deoxy-(1',2',3',4',5'-¹³C₅, N¹,N³-¹⁵N₂)-cytidine (29)



A high pressure autoclave glass vessel was charged with PPh₃ (13 mg, 0.05 mmol, 0.6), Pd₂(dba)₃CHCl₃ (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.

¹H-NMR (400 MHz, CDCl₃, ppm): δ = 9.50 (s, 1H, CHO), 8.53 (sd, ²J_{N-H} = 2.7 Hz, 1H, 5-H), 6.01 (dd, ¹J_{C-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-H} = 135.1 Hz, 1H, 2'-H), 2.44 (d, ¹J_{C-H} = 137.5 Hz, 1H, 2'-H), 1.10 - 0.96 (m, 28H, Si(CH₃)₂, Si-CH). **¹³C-NMR (151 MHz, CDCl₃, ppm):** δ = 187.5 (CHO), 162.9 (d, ¹J_{C-N} = 5.9 Hz, C-4), 153.3 (t, ¹J_{C-N} = 7.9 Hz, 2-C), 152.7 (d, ¹J_{C-N} = 14.3 Hz, C-5), 104.8 (C-6), 86.1 - 84.9 (1'-C + 4'-C), 65.9 (t, ¹J_{C-C} = 38.2, 3'-C), 59.4 (d, J = 42.9 Hz, 5'-C), 39.5 (t, ¹J_{C-C} = 36.9 Hz, 2'-C), 17.5 - 16.8 (8 x Si-C(CH₃)), 13.5 - 12.4 (4 x Si-C(CH₃)). **¹⁵N-NMR (41 MHz, CDCl₃, ppm):** δ = -206.6, -206.8. **HRMS (ESI⁺):** calc. for C₁₇¹³C₅H₄₀N¹⁵N₂O₆Si₂⁺ [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).

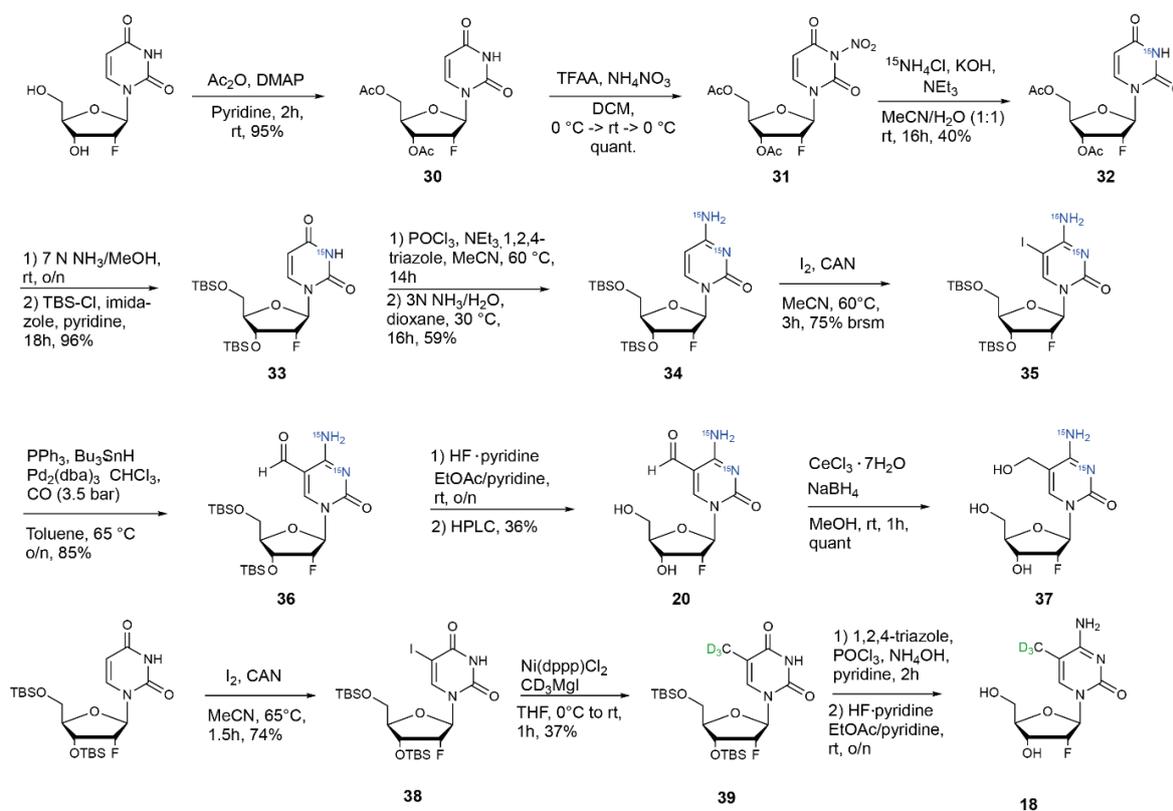
5-formyl-2'-deoxy-(1',2',3',4',5'-¹³C₅, N¹,N³-¹⁵N₂)-cytidine (1)



29 (30 mg, 0.06 mmol, 1.0 eq) was dissolved in EtOAc (1 mL) and pyridine (24 μ L) and HF•pyridine (16 μ L, 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 μ L) 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.

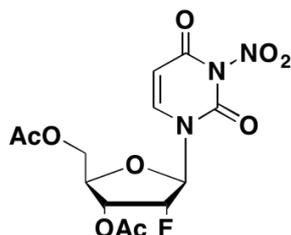
¹H-NMR (400 MHz, D₂O, ppm): δ = 9.50 (s, 1H, CHO), 8.84 (sd, ²J_{N-H} = 2.9 Hz, 1H, 5-H), 6.15 (dt, ¹J_{C-H} = 175.4 Hz, ³J = 5.8 Hz, 1H, 1'-H), 4.61 (d, ¹J_{C-H} = 151.9 Hz, 1H, 3'-H), 4.30 - 3.93 (m, 2H, 4'-H + 5'-H), 3.76 - 3.53 (m, 1H, 5'-H), 2.74 (d, ¹J_{C-H} = 136.0 Hz, 1H, 2'-H), 2.53 (d, ¹J_{C-H} = 134.5 Hz, 1H, 2'-H). **¹³C-NMR (151 MHz, D₂O, ppm):** δ = 190.5 (CHO), 161.7 (d, ¹J_{C-N} = 7.4 Hz, C-4), 154.7 (d, ¹J_{C-N} = 14.2 Hz, C-5), 154.0 (t, ¹J_{C-N} = 11.3 Hz, 2-C), 105.4 (C-6), 87.8 - 86.5 (1'-C + 4'-C), 69.3 (t, ¹J_{C-C} = 36.5, 3'-C), 60.5 (d, J = 41.5 Hz, 5'-C), 40.0 (t, ¹J_{C-C} = 36.0 Hz, 2'-C). **¹⁵N-NMR (41 MHz, D₂O, ppm):** δ = -209.9, -210.2. **HRMS (ESI⁺):** calc. for C₅¹³C₅H₁₃N¹⁵N₂NaO₅⁺ [M+Na]⁺: 285.0856, found: 285.0857.

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



Scheme 2: Synthesis of $^{15}\text{N}_2$ -labelled 2'-(*R*)-F-fdC **20**, 2'-(*R*)-F-hmdC **37** and $[\text{D}_3]$ -2'-(*R*)-F-mdC **18**.

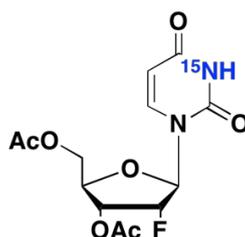
3',5'-di-*O*-acetyl-2'-deoxy-2'-(*R*)-fluoro-3-nitrouridine (**31**)



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.

$^1\text{H-NMR}$ (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, $^3J_{\text{H-F}}$ = 17.2 Hz, 3J = 1.4 Hz, 1H, 1'-H), 5.41 (ddd, $^2J_{\text{H-F}}$ = 51.6 Hz, 3J = 4.8 Hz, 3J = 1.4 Hz, 1H, 2'-H), 5.13 (ddd, $^3J_{\text{H-F}}$ = 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)-CH₃), 2.10 (s, 3H, C5'-O-(C=O)-CH₃). **$^{13}\text{C-NMR}$ (151 MHz, CDCl_3 , ppm):** δ = 170.2 (C5'-O-(C=O)-CH₃), 169.9 (C3'-O-(C=O)-CH₃), 154.9 (C-4), 145.1 (C-2), 140.1 (C-6), 102.0 (C-5), 92.0 (d, $^2J_{\text{C-F}}$ = 37.3 Hz, C-1'), 90.8 (d, $^1J_{\text{C-F}}$ = 193.1 Hz, C-2'), 79.3 (C-4'), 69.3 (d, $^2J_{\text{C-F}}$ = 16.0 Hz, C-3'), 61.8 (C-5'), 20.8 (C5'-O-(C=O)-CH₃), 20.5 (C3'-O-(C=O)-CH₃). **$^{19}\text{F-NMR}$ (377 MHz, CDCl_3 , ppm):** δ = -199.4 (dt, $^2J_{\text{F-H}}$ = 51.8 Hz, $^3J_{\text{F-H}}$ = 18.4 Hz). **HRMS (ESI⁺):** calc. for $\text{C}_{13}\text{H}_{18}\text{FN}_4\text{O}_9^+$ [M+NH₄]⁺: 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.

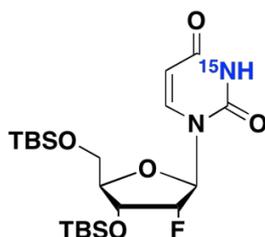
3',5'-di-O-acetyl-2'-deoxy-2'-(R)-fluoro-(N³-¹⁵N)-uridine (32)



First, ¹⁵NH₄Cl (400 mg, 7.34 mmol, 1.2 eq) was dissolved in H₂O (15 mL) and KOH (412 mg, 7.34 mmol, 1.2 eq) and NEt₃ (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. NH₄Cl (100 mL). The aqueous layers were extracted with DCM (3 x 100 mL), organic layers were collected and dried over Na₂SO₄. 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.

¹H-NMR (800 MHz, DMSO-d₆, ppm): δ = 7.71 (d, ³J = 8.1 Hz, 1H, 6-H), 5.87 (dd, ³J_{H-F} = 22.4 Hz, ³J = 2.0 Hz, 1H, 1'-H), 5.69 - 5.68 (m, 1H, 5-H), 5.33 (ddd, ²J_{H-F} = 52.5 Hz, ³J = 5.3 Hz, ³J = 2.0 Hz, 1H, 2'-H), 5.26 (ddd, ³J_{H-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)-CH₃), 2.04 (s, 3H, C5'-O-(C=O)-CH₃). **¹³C-NMR (202 MHz, DMSO-d₆, ppm):** δ = 170.2 (C5'-O-(C=O)-CH₃), 169.5 (C3'-O-(C=O)-CH₃), 163.2 (sd, ¹J_{C-N} = 8.7 Hz, C-4), 150.1 (sd, ¹J_{C-N} = 17.6 Hz, C-2), 142.9 (C-6), 102.1 (C-5), 90.8 (d, ¹J_{C-F} = 186.3 Hz, C-2'), 90.5 (d, ²J_{C-F} = 37.0 Hz, C-1'), 77.9 (C-4'), 69.8 (d, ²J_{C-F} = 14.3 Hz, C-3'), 62.7 (C-5'), 20.5 (C5'-O-(C=O)-CH₃), 20.3 (C3'-O-(C=O)-CH₃). **¹⁹F-NMR (377 MHz, DMSO-d₆, ppm):** δ = -198.0 (ddd, ²J_{H-F} = 52.5 Hz, ³J_{H-F} = 22.3, ³J_{H-F} = 17.2). **¹⁵N-NMR (41 MHz, DMSO-d₆, ppm):** δ = -222.0. **HRMS (ESI⁺):** calc. for C₁₃H₁₉FN₂¹⁵NO₇⁺ [M+NH₄]⁺: 349.1172, found: 349.1174. **HRMS (ESI⁻):** calc. for C₁₃H₁₄FN¹⁵NO₇⁻ [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.

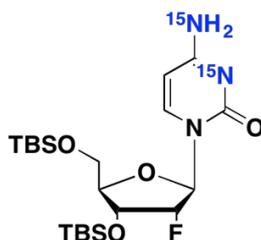
3',5'-bis-O-(tert-butyl(dimethyl)silyl)-2'-deoxy-2'-(R)-fluoro-(N³-¹⁵N)-uridine (33)



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. NaHCO₃ (100 mL) and extracted with DCM (3 x 75 mL) and the combined organic layers were dried over Na₂SO₄. 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.

¹H-NMR (599 MHz, CDCl₃, ppm): δ = 7.92 (d, ³*J* = 8.2 Hz, 1H, 6-H), 6.06 (dd, ³*J*_{H-F} = 15.3 Hz, ³*J* = 2.0 Hz, 1H, 1'-H), 5.69 (dd, ³*J* = 8.2 Hz, ³*J*_{H-N} = 2.6 Hz, 1H, 5-H), 4.77 (ddd, ²*J*_{H-F} = 52.3 Hz, ³*J* = 4.3 Hz, ³*J* = 2.0 Hz, 1H, 2'-H), 4.29 (ddd, ³*J*_{H-F} = 18.9 Hz, ³*J* = 7.2 Hz, ³*J* = 4.2 Hz, 1H, 3'-H), 4.08 (dd, ³*J* = 7.2 Hz, ³*J* = 1.6 Hz, 1H, 4'-H), 4.05 (dd, ²*J* = 11.7 Hz, ³*J* = 2.1 Hz, 1H, 5'-H), 3.78 (dd, ²*J* = 11.7 Hz, ³*J* = 1.8 Hz, 1H, 5'-H), 0.93 - 0.90 (m, 18H, Si(CH₃)₂(^tBu)), 0.12 - 0.10 (m, 12H, Si(CH₃)₂(^tBu)). **¹³C-NMR (151 MHz, CDCl₃, ppm):** δ = 163.1 (d, ¹*J*_{C-N} = 8.7 Hz, 4-C), 150.0 (sd, ¹*J*_{C-N} = 17.8 Hz, 2-C), 139.9 (C-6), 102.5 (C-5), 93.2 (d, ¹*J*_{C-F} = 193.3 Hz, C-2'), 87.9 (d, ²*J*_{C-F} = 33.8 Hz, C-1'), 83.9 (C-4'), 68.7 (d, ²*J*_{C-F} = 16.2 Hz, C-3'), 60.9 (C-5'), 26.1 (Si(C(CH₃)₃)), 25.7 (Si(C(CH₃)₃)), 18.5 (Si(C(CH₃)₃)), 18.2 (Si(C(CH₃)₃)), -4.5 (Si(CH₃)₂^tBu), -4.9 (Si(CH₃)₂^tBu), -5.3 (Si(CH₃)₂^tBu), -5.4 (Si(CH₃)₂^tBu). **¹⁹F-NMR (377 MHz, CDCl₃, ppm):** δ = -202.5 (ddd, ²*J*_{H-F} = 52.4 Hz, ³*J*_{H-F} = 19.0, ³*J*_{H-F} = 15.3). **¹⁵N-NMR (41 MHz, CDCl₃, ppm):** δ = -224.4. **HRMS (ESI⁺):** calc. for C₂₁H₄₀FN¹⁵NO₅Si₂⁺ [M]⁺: 476.2425, found: 476.2429. **HRMS (ESI⁻):** calc. for C₂₁H₃₈FN¹⁵NO₅Si₂⁻ [M-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*⁴-¹⁵N₂)-cytidine (34)

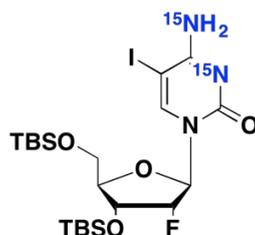


In a round bottom flask, 1,2,4-triazole (1.04 g, 15.1 mmol, 9.9 eq) was dissolved in MeCN (30 mL), cooled to 0 °C and stirred for ten minutes. To the resulting turbid solution, NEt₃ (2.00 mL, 14.4 mmol, 9.4 eq) was added dropwise and the mixture was stirred at 0 °C for another 20 minutes. Subsequently, 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 NaHCO₃ (100 mL) and the resulting aqueous layer was extracted with DCM (3 x 100 mL). Combined organic layers were dried over Na₂SO₄ 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 ¹⁵NH₃ (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 NH₄Cl (50 mL), extracted with DCM (3 x 50 mL) and the combined organic layers were backwashed with brine (100 mL) and subsequently dried over Na₂SO₄. 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.

¹H-NMR (800 MHz, CDCl₃, ppm): δ = 8.07 (d, ³J = 7.5 Hz, 1H, 6-H), 6.03 (dd, ³J_{H-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-F} = 52.2 Hz, ³J = 3.9 Hz, 1H, 2'-H), 4.23 (ddd, ³J_{H-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(CH₃)₂(^tBu)), 0.89 (s, 9 H, Si(CH₃)₂(^tBu)), 0.13 (s, 3H, Si(CH₃)₂(^tBu)), 0.12 (s, 3H, Si(CH₃)₂(^tBu)), 0.10 (s, 3H, Si(CH₃)₂(^tBu)), 0.08 (s, 3H, Si(CH₃)₂(^tBu)). **¹³C-NMR (201 MHz, CDCl₃, ppm):** δ = 165.0 (dd, ¹J_{C-N} = 21.8 Hz, ¹J_{C-N} = 8.54 Hz 4-C), 154.9 (sd, ¹J_{C-N} = 9.2 Hz, 2-C), 141.5 (C-6), 94.5 (C-5), 93.4 (d, ¹J_{C-F} = 192.3 Hz, C-2'), 88.9 (d, ²J_{C-F} = 33.4 Hz, C-1'), 83.1 (C-4'), 68.0 (d, ²J_{C-F} = 17.3 Hz, C-3'), 60.4 (C-5'), 26.1 (Si(C(CH₃)₃)), 25.8 (Si(C(CH₃)₃)), 18.6 (Si(C(CH₃)₃)), 18.2 (Si(C(CH₃)₃)), -4.5 (Si(CH₃)₂(^tBu)), -4.9 (Si(CH₃)₂(^tBu)), -5.3 (Si(CH₃)₂(^tBu)), -5.4 (Si(CH₃)₂(^tBu)). **¹⁹F-NMR (376 MHz, CDCl₃, ppm):** δ = -201.3 (ddd, ²J_{H-F} = 51.8 Hz, ³J_{H-F} = 20.8, ³J_{H-F} = 15.6). **¹⁵N-NMR (41 MHz, CDCl₃, ppm):** δ = -288.1, -288.3. **HRMS (ESI⁺):** calc. for C₂₁H₄₁FN¹⁵N₂O₄Si₂⁺ [M]⁺: 476.2555, found: 476.2558. **HRMS (ESI⁻):** calc. for C₂₁H₃₉FN¹⁵N₂O₄Si₂⁻ [M-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).

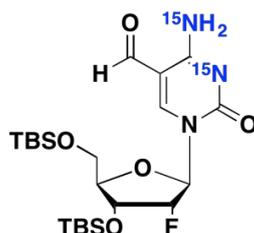
3',5'-bis-*O*-(*tert*-butyl(dimethyl)silyl)-5-iodo-2'-deoxy-2'-(*R*)-fluoro-(N^3,N^4 - $^{15}\text{N}_2$)-cytidine (35)



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 chromatography (5% MeOH/DCM) of the crude mixture yielded **35** (37 mg, 60 μmol , quant.) as an orange oil.

$^1\text{H-NMR}$ (400 MHz, CDCl_3 , ppm): δ = 7.99 (s, 1H, 6-H), 5.98 (dd, $^3J_{\text{H-F}} = 16.0$ Hz, $^3J = 1.8$ Hz, 1H, 1'-H), 4.81 (ddd, $^2J_{\text{H-F}} = 52.3$ Hz, $^3J = 4.4$ Hz, $^3J = 1.8$ Hz, 1H, 2'-H), 4.19 (ddd, $^3J_{\text{H-F}} = 18.9$ 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(CH₃)₂(^tBu)), 0.89 (s, 9 H, Si(CH₃)₂(^tBu)), 0.16 (s, 3H, Si(CH₃)₂(^tBu)), 0.15 (s, 3H, Si(CH₃)₂(^tBu)), 0.10 (s, 3H, Si(CH₃)₂(^tBu)), 0.08 (s, 3H, Si(CH₃)₂(^tBu)). **$^{13}\text{C-NMR}$ (101 MHz, CDCl_3 , ppm):** δ = 164.0 (dd, $^1J_{\text{C-N}} = 23.0$ Hz, $^1J_{\text{C-N}} = 6.0$ Hz, 4-C), 154.5 (sd, $^1J_{\text{C-N}} = 8.9$ Hz, 2-C), 146.4 (C-6), 57.0 (C-5), 93.1 (d, $^1J_{\text{C-F}} = 193.3$ Hz, C-2'), 89.2 (d, $^2J_{\text{C-F}} = 33.0$ Hz, C-1'), 83.9 (C-4'), 68.8 (d, $^2J_{\text{C-F}} = 16.5$ Hz, C-3'), 61.1 (C-5'), 57.4 (C-5), 26.5 (Si(C(CH₃)₃)), 25.8 (Si(C(CH₃)₃)), 18.9 (Si(C(CH₃)₃)), 18.2 (Si(C(CH₃)₃)), -4.4 (Si(CH₃)₂(^tBu)), -4.7 (Si(CH₃)₂(^tBu)), -4.8 (Si(CH₃)₂(^tBu)), -4.9 (Si(CH₃)₂(^tBu)). **$^{19}\text{F-NMR}$ (376 MHz, CDCl_3 , ppm):** δ = -201.7 (ddd, $^2J_{\text{H-F}} = 52.3$ Hz, $^3J_{\text{H-F}} = 19.0$, $^3J_{\text{H-F}} = 16.8$). **$^{15}\text{N-NMR}$ (41 MHz, CDCl_3 , ppm):** δ = -280.4, -280.5. **HRMS (ESI⁺):** calc. for C₂₁H₄₀FIN¹⁵N₂O₄Si₂⁺ [M]⁺: 602.1521, found: 602.1530. **HRMS (ESI⁻):** calc. for C₂₁H₃₈FIN¹⁵N₂O₄Si₂⁻ [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).

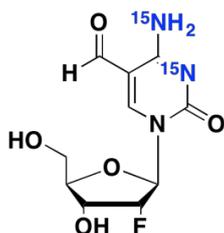
3',5'-bis-*O*-(*tert*-butyl(dimethyl)silyl)5-formyl-2'-deoxy-2'-(*R*)-fluoro-(*N*³,*N*⁴-¹⁵N₂)-cytidine (36)



The reaction vessel of a high pressure CO-autoclave was charged with **35** (220 mg, 0.37 mmol, 1.0 eq), PPh₃ (58 mg, 0.2 mmol, 0.6 eq), Pd₂(dba)₃CHCl₃ (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.

¹H-NMR (599 MHz, CDCl₃, ppm): δ = 9.50 (s, 1H, CHO), 8.53 (s, 1H, 6-H), 8.26 (dd, ¹J_{H-N} = 92.9 Hz, ¹J = 6.0 Hz, 1H, ¹⁵NH₂), 7.29 (d, ¹J_{H-N} = 92.9 Hz, 1H, ¹⁵NH₂), 6.02 (dd, ³J_{H-F} = 16.3 Hz, ³J = 2.0 Hz, 1H, 1'-H), 4.92 (dd, ²J_{H-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(CH₃)₂(^tBu)), 0.89 (s, 9 H, Si(CH₃)₂(^tBu)), 0.14 (s, 3H, Si(CH₃)₂(^tBu)), 0.11 (s, 3H, Si(CH₃)₂(^tBu)), 0.10 (s, 3H, Si(CH₃)₂(^tBu)), 0.08 (s, 3H, Si(CH₃)₂(^tBu)). **¹³C-NMR (151 MHz, CDCl₃, ppm):** δ = 187.3 (CHO), 162.9 (dd, ¹J_{C-N} = 20.9 Hz, ¹J_{C-N} = 5.1 Hz, C-4), 153.2 (C-6), 153.0 (C-2), 105.7 (C-5), 93.0 (d, ¹J_{C-F} = 192.9 Hz, C-2'), 89.8 (d, ²J_{C-F} = 34.4 Hz, C-1'), 83.8 (C-4'), 68.2 (d, ²J_{C-F} = 16.5 Hz, C-3'), 60.7 (C-5'), 26.2 (Si(C(CH₃)₃)), 25.7 (Si(C(CH₃)₃)), 18.8 (Si(C(CH₃)₃)), 18.2 (Si(C(CH₃)₃)), -4.3 (Si(CH₃)₂^tBu), -4.8 (Si(CH₃)₂(C(CH₃)₃)), -4.8 (Si(CH₃)₂^tBu), -5.2 (Si(CH₃)₂^tBu). **¹⁹F-NMR (377 MHz, CDCl₃, ppm):** δ = -200.9 (ddd, ²J_{H-F} = 51.7 Hz, ³J_{H-F} = 19.9, ³J_{H-F} = 15.7). **¹⁵N-NMR (41 MHz, CDCl₃, ppm):** δ = -175.2, -286.0. **HRMS (ESI⁺):** calc. for C₂₂H₄₁FN¹⁵N₂O₅Si₂⁺ [M]⁺: 504.2504, found: 504.2513. **HRMS (ESI⁻):** calc. for C₂₂H₃₉FN¹⁵N₂O₅Si₂⁻ [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).

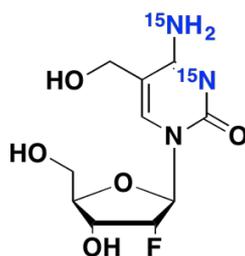
5-formyl-2'-deoxy-2'-(R)-fluoro-(N^3,N^4 - $^{15}\text{N}_2$)-cytidine (20)



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/H₂O --> 13% MeCN/H₂O in 30 minutes, 5.0 mL/min) for cell-feeding purposes.

$^1\text{H-NMR}$ (800 MHz, D₂O, ppm): δ = 9.54 (s, 1H, CHO), 8.90 (s, 1H, 6-H), 6.10 (d, $^3J_{\text{H-F}} = 17.7$ Hz, 1H, 1'-H), 5.22 (dd, $^2J_{\text{H-F}} = 52.3$ Hz, $^3J = 4.3$ Hz, 1H, 2'-H), 4.39 (ddd, $^3J_{\text{H-F}} = 24.2$ 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). **$^{13}\text{C-NMR}$ (201 MHz, D₂O, ppm):** δ = 190.3 (CHO), 162.8 (dd, $^1J_{\text{C-N}} = 21.8$ Hz, $^1J_{\text{C-N}} = 6.5$ Hz, C-4), 155.0 (C-6), 154.9 (C-2), 105.9 (C-5), 93.7 (d, $^1J_{\text{C-F}} = 186.0$ Hz, C-2'), 89.9 (d, $^2J_{\text{C-F}} = 34.9$ Hz, C-1'), 82.5 (C-4'), 67.2 (d, $^2J_{\text{C-F}} = 16.7$ Hz, C-3'), 59.1 (C-5'). **$^{19}\text{F-NMR}$ (376 MHz, D₂O, ppm):** δ = -201.6 (ddd, $^2J_{\text{H-F}} = 52.3$ Hz, $^3J_{\text{H-F}} = 24.0$, $^3J_{\text{H-F}} = 17.7$). **$^{15}\text{N-NMR}$ (41 MHz, D₂O, ppm):** δ = -199.9, -200.4. **HRMS (ESI⁺):** calc. for C₁₀H₁₃FN¹⁵N₂O₅⁺ [M+H]⁺: 276.0774, found: 276.0778. **HRMS (ESI⁻):** calc. for C₁₀H₁₁FN¹⁵N₂O₅⁻ [M-H]⁻: 274.0629, found: 274.0629.

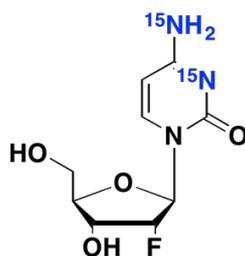
5-hydroxymethyl-2'-deoxy-2'-(R)-fluoro-(N^3, N^4 - $^{15}\text{N}_2$)-cytidine (37**)**



First, **20** (1.00 mg, 3.60 μmol , 1.0 eq) was dissolved in MeOH (1.0 mL), $\text{CeCl}_3 \cdot 7 \text{H}_2\text{O}$ (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 lyophilization. Purification of **37** was achieved through semipreparative HPLC (Waters Nucleodur C18 ec, 0% MeCN/ H_2O \rightarrow 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.

$^1\text{H-NMR}$ (800 MHz, D_2O , ppm): δ = 7.85 (s, 1H, 6-H), 5.92 (dd, $^3J_{\text{H-F}} = 19.4 \text{ Hz}$, $^3J = 1.1 \text{ Hz}$, 1H, 1'-H), 5.07 (dd, $^2J_{\text{H-F}} = 52.9 \text{ Hz}$, $^3J = 4.5 \text{ Hz}$, 1H, 2'-H), 4.35 (s, 2H, CH_2OH), 4.28 (ddd, $^3J_{\text{H-F}} = 22.0 \text{ Hz}$, $^3J = 8.9 \text{ Hz}$, $^3J = 4.5 \text{ Hz}$, 1H, 3'-H) 4.07 (dt, $^3J = 8.9 \text{ Hz}$, $^3J = 3.3 \text{ Hz}$, 1H, 4'-H), 3.97 (dd, $^2J = 13.0 \text{ Hz}$, $^3J = 2.4 \text{ Hz}$, 1H, 5'-H), 3.78 (dd, $^2J = 13.0 \text{ Hz}$, $^3J = 4.0 \text{ Hz}$, 1H, 5'-H). **$^{13}\text{C-NMR}$ (201 MHz, D_2O , ppm):** δ = 165.2 (dd, $^1J_{\text{C-N}} = 21.2 \text{ Hz}$, $^1J_{\text{C-N}} = 6.7 \text{ Hz}$, C-4), 156.9 (C-2), 140.9 (C-6), 106.5 (C-5), 93.7 (d, $^1J_{\text{C-F}} = 184.4 \text{ Hz}$, C-2'), 89.8 (d, $^2J_{\text{C-F}} = 35.0 \text{ Hz}$, C-1'), 82.1 (C-4'), 67.7 (d, $^2J_{\text{C-F}} = 16.1 \text{ Hz}$, C-3'), 59.6 (C-5'), 57.7 (CH_2OH). **$^{19}\text{F-NMR}$ (376 MHz, D_2O , ppm):** δ = -200.5 (ddd, $^2J_{\text{H-F}} = 52.9 \text{ Hz}$, $^3J_{\text{H-F}} = 22.0$, $^3J_{\text{H-F}} = 19.4$). **$^{15}\text{N-NMR}$ (41 MHz, D_2O , ppm):** δ = -201.9, -202.0. **HRMS (ESI $^+$):** calc. for $\text{C}_{10}\text{H}_{15}\text{FN}^{15}\text{N}_2\text{O}_5^+$ [$\text{M}+\text{H}$] $^+$: 278.0931, found: 278.0933. **HRMS (ESI $^-$):** calc. for $\text{C}_{10}\text{H}_{13}\text{FN}^{15}\text{N}_2\text{O}_5^-$ [$\text{M}-\text{H}$] $^-$: 276.0785, found: 276.0787.

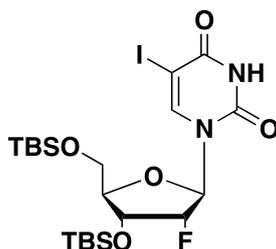
2'-deoxy-2'-(R)-fluoro-(N³,N⁴-¹⁵N₂)-cytidine (19)



34 (117 mg, 0.25 mmol, 1.0 eq) was dissolved in EtOAc (3.5 mL) and pyridine (98 μ L). Subsequently HF•pyridine (70% solution, 123 μ L, 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 H₂O and then purified by HPLC which yielded 44 mg (0.18 mmol, 72%) of **19** as a colourless solid.

¹H-NMR (800 MHz, D₂O, 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-F} = 19.8 Hz, 1H, 1'-H), 5.15 (ddd, ²J_{H-F} = 52.9 Hz, ³J = 4.9 Hz, ³J = 1.1 Hz, 1H, 2'-H), 4.34 (ddd, ³J_{H-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). **¹³C-NMR (201 MHz, D₂O, ppm):** δ = 162.6 (C4), 149.9 (C2), 140.6 (C6), 96.0 (C5), 93.2 (d, ¹J_{C-F} = 193.8 Hz, C2'), 89.9 (d, ²J_{C-F} = 35.7 Hz, C1'), 82.2 (C4'), 67.8 (d, ²J_{C-F} = 15.5 Hz, C3'), 59.9 (C5'). **¹⁹F-NMR (376 MHz, D₂O, ppm):** δ = -200.3 (dt, ²J_{F-H} = 53.1 Hz, ³J_{F-H} = 19.0 Hz). **HRMS (ESI⁻):** calc. for C₁₁H₁₅FN¹⁵N₂O₆⁻ [M+OAc]⁻: 306.0891, found: 306.0894.

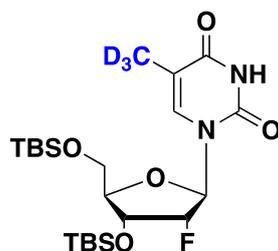
3',5'-bis-O-(tert-butyl(dimethyl)silyl)- 5-iodo-2'-deoxy-2'-(R)-fluoro -uridine (38)



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 I₂ (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 NaHCO₃ and a saturated solution of Na₂S₂O₃ was added until the solution became colourless. The aqueous phases were extracted with EtOAc (2x50 mL) and the combined organic fractions dried over Na₂SO₄. Column chromatography (20% EtOAc/Hex) yielded the iodinated compound as a brownish foam (469 mg, 0.78 mmol, 74%).

¹H-NMR (400 MHz, CDCl₃, ppm): δ = 9.34 (s, 1H, NH), 7.98 (s, 1H, 6-H), 6.08 (dd, ³J_{H-F} = 14.3 Hz, ³J = 3.8 Hz, 1H, 1'-H), 4.82 (dt, ²J_{H-F} = 52.5 Hz, ³J = 4.2 Hz, 1H, 2'-H), 4.27 (dt, ³J_{H-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(CH₃)₂(^tBu)), 0.90 (s, 9 H, Si(CH₃)₂(^tBu)), 0.16 (s, 3H, Si(CH₃)₂(^tBu)), 0.15 (s, 3H, Si(CH₃)₂(^tBu)), 0.10 (s, 3H, Si(CH₃)₂(^tBu)), 0.08 (s, 3H, Si(CH₃)₂(^tBu)). **¹³C-NMR (101 MHz, CDCl₃, ppm):** δ = 160.0 (4-C), 149.9 (2-C), 144.0 (C-6), 92.4 (d, ¹J_{C-F} = 195.3 Hz, C-2'), 87.4 (d, ²J_{C-F} = 33.7 Hz, C-1'), 85.1 (C-4'), 69.6 (d, ²J_{C-F} = 15.6 Hz, C-3'), 69.2 (C-5'), 61.7 (C-5), 26.3 (Si(C(CH₃)₃)), 25.7 (Si(C(CH₃)₃)), 18.7 (Si(C(CH₃)₃)), 18.2 (Si(C(CH₃)₃)), -4.6 (Si(CH₃)₂^tBu), -5.0 (Si(CH₃)₂^tBu), -5.1 (Si(CH₃)₂^tBu), -5.1 (Si(CH₃)₂^tBu). **HRMS (ESI⁺):** calc. for C₂₁H₃₉FIN₂O₅Si₂⁺ [M+H]⁺: 601.1421, found:601.1432.

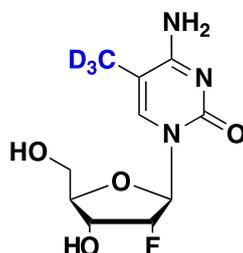
3',5'-bis-O-[tert-butyl(dimethyl)silyl]- 5-(D₃)-methyl 2'-deoxy-2'-(R)-fluoro-uridine (39)



38 (100 mg, 0.167 mmol, 1.0 eq) was dissolved in THF (4 mL) and Ni(dppp)Cl₂ (45 mg, 0.083 mmol, 0.5 eq) was added. The mixture was cooled to 0 °C and CD₃MgI (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. NH₄Cl (10 mL). The aqueous phase was extracted twice with EtOAc (15 mL), combined organic fractions dried over Na₂SO₄ and volatiles removed *in vacuo*. The crude mixture was finally purified through column chromatography (20% EtOAc/Hex) to yield 30 mg (0.061 mmol, 37%) of **39**.

¹H-NMR (400 MHz, CDCl₃, ppm): δ = 8.82 (s, 1H, NH), 7.38 (s, 1H, 6-H), 6.07 (dd, ³J_{H-F} = 15.5 Hz, ³J = 3.6 Hz, 1H, 1'-H), 4.83 (dt, ²J_{H-F} = 52.9 Hz, ³J = 4.2 Hz, 1H, 2'-H), 4.29 (dt, ³J_{H-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(CH₃)₂(^tBu)), 0.91 (s, 9 H, Si(CH₃)₂(^tBu)), 0.12 - 0.18 (m, 12H, Si(CH₃)₂(^tBu)). **¹³C-NMR (101 MHz, CDCl₃, ppm):** δ = 163.7 (4-C), 150.2 (2-C), 135.5 (C-6), 111.6 (5-C), 92.3 (d, ¹J_{C-F} = 194.2 Hz, C-2'), 87.4 (d, ²J_{C-F} = 33.5 Hz, C-1'), 84.5 (C-4'), 69.5 (d, ²J_{C-F} = 15.8 Hz, C-3'), 61.7 (C-5'), 26.0 (Si(C(CH₃)₃)), 25.7 (Si(C(CH₃)₃)), 18.5 (Si(C(CH₃)₃)), 18.2 (Si(C(CH₃)₃)), -4.7 (Si(CH₃)₂^tBu), -5.1 (Si(CH₃)₂^tBu), -5.3 (Si(CH₃)₂^tBu), -5.3 (Si(CH₃)₂^tBu). **HRMS (ESI⁺):** calc. for C₂₂H₃₉D₃FN₂O₅Si₂⁺ [M+H]⁺: 492.2799, found: 492.2805.

5-(D₃)-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 POCl₃ (12 μL, 0.12 mmol, 1.2 eq) were added at rt. The mixture was stirred for two hours and then poured into sat. NH₄Cl (10 mL). The resulting aqueous phase was quickly extracted with EtOAc (10 mL), dried over Na₂SO₄ and the solvents removed *in vacuo*. The residue was redissolved in dioxane (3.0 mL) and NH₄OH (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/H₂O in 45 mins) to give **18** as a white powder.

¹H-NMR (400 MHz, D₂O, ppm): δ = 7.54 (s, 1H, 6-H), 5.87 (dd, ³J_{H-F} = 19.8 Hz, ³J = 1.2 Hz, 1H, 1'-H), 5.00 (dd, ²J_{H-F} = 53.0 Hz, ³J = 4.6 Hz, 1H, 2'-H), 4.23 (ddd, ³J_{H-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). **¹³C-NMR (101 MHz, D₂O, ppm):** δ = 166.2 (C4), 157.0 (C2), 139.1 (C6), 104.4 (C5), 93.7 (d, ¹J_{C-F} = 184.5 Hz, C2'), 89.7 (d, ²J_{C-F} = 35.1 Hz, C1'), 82.1 (C4'), 67.8 (d, ²J_{C-F} = 16.6 Hz, C3'), 59.8 (C5').

HRMS (ESI⁺): calc. for C₁₀H₁₂D₃FN₂O₄⁺ [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 3.2.1

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

* Beitrag der Autoren zu gleichen Teilen

Supporting Information

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

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

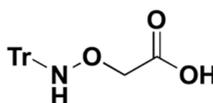
Center for Integrated Protein Science at the Department of Chemistry, LMU Munich,
Butenandtstrasse 5-13, Munich 81377, Germany

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*, *TCl*, *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 F₂₅₄*), 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 H₂O, pH = 4.3 (referred to as buffer A) and 0.25 mM ammonium formate in 80% MeCN/H₂O (referred to as buffer B) was used.

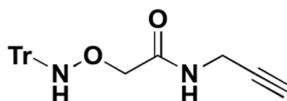
Synthesis of Hydroxylamine 1 and internal standards 9a/b and 10a/b

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



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

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

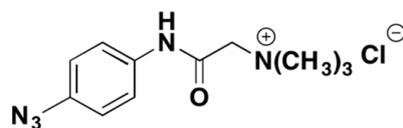


N-trityl protected aminoxyacetic acid **6** (2.50 g, 7.50 mmol, 1.0 eq) was suspended in DCM (40 mL) and was subsequently charged with TBTU (2.89 g, 9.00 mmol, 1.2 eq), DIPEA (1.60 mL, 9.00 mmol, 1.2 eq) and propargylamine (1.40 mL, 22.6 mmol, 3.0 eq). The suspension was stirred at rt, whereupon

after 15 hours a clear yellowish solution was formed. The mixture was diluted with EtOAc (300 mL), the organic phase was washed with NH₄Cl (300 mL) and NaHCO₃ (300 mL) and then dried over Na₂SO₄. Volatiles were finally removed *in vacuo* and the crude mixture was purified via column chromatography (10% EtOAc --> 40% EtOAc/Hex). **7** (2.57g, 6.93 mmol, 92%) was yielded as a colorless solid.

¹H-NMR (300 MHz, CDCl₃): δ/ppm = 7.37–7.22 (m, 15H, (C₆H₅)₃C), 6.59 (s, 1H, (C₆H₅)₃C-NH-O), 5.81 (bs, 1H, O=C-NH), 4.25 (s, 2H, O-CH₂C=O), 3.85 (dd, ³J = 5.5 Hz, ⁴J = 2.6 Hz, 2H, HN-CH₂), 2.15 (t, ⁴J = 2.6 Hz, 1H, C≡C-H). **¹³C-NMR (75 MHz, CDCl₃):** δ/ppm = 169.1 (C=O), 143.9 (3C, 3 × O-NH-C-C), 129.0 (6C, C_{Ar}-H), 128.2 (6C, C_{Ar}-H), 127.4 (3C, C_{tert}-H), 79.3 (C≡C-H), 74.6 (C(C₆H₅)₃), 73.4 (O-CH₂), 71.7 (C≡C-H), 28.8 (NH-CH₂). **HRMS (ESI⁺):** calc. for C₂₄H₂₂N₂NaO₂ [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)

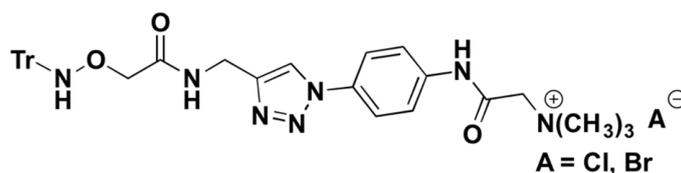


Betaine **3a** (0.30 g, 2.56 mmol, 1.0 eq) was first dried on high vac at 180 °C for 20 minutes. After cooling to rt, the colorless solid was suspended in DMF (25 mL). 4-Azidoanilin hydrochloride **2** (0.54g, 3.17 mmol, 1.2 eq), TBTU (0.99 g, 3.07 mmol, 1.2 eq) and DIPEA (1.10 mL, 6.32 mmol, 2.4 eq) 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/H₂O/7N NH₃ in methanol = 90:10:0.6:0.6) and **4a** was yielded as the corresponding phenyl triazolate salt. The salt was then redissolved in H₂O (50 mL) and was acidified to pH = 1. The aqueous phase was extracted with Et₂O until TLC analysis of the organic phase fractions showed no UV absorption. The aqueous layer was subsequently neutralized with conc. NH₃ and the chloride salt of **4a** (0.62 g, 2.30 mmol, 90%) was yielded as a brownish powder.

¹H-NMR (300 MHz, dmsO d⁶): δ/ppm = 11.22 (s, 1H, NH), 7.69 (d, ³J = 8.9 Hz, 2H, CH=C-NH), 7.12 (d, ³J = 8.4 Hz, 2H, CH=C-N₃), 4.42 (s, 2H, CH₂), 3.30 (s, 9H, N(CH₃)₃). **¹³C-NMR (101 MHz, dmsO d⁶):** δ/ppm = 162.0 (C=O), 135.0 (NH-C=CH), 134.9 (N₃-C=CH), 121.2 (2C, NH-C=CH), 119.5 (2C, N₃-C=CH), 64.3 (CH₂),

53.4 (3C, N(CH₃)₃). **HRMS (ESI⁺)**: calc. for C₁₁H₁₆N₅O⁺ [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.

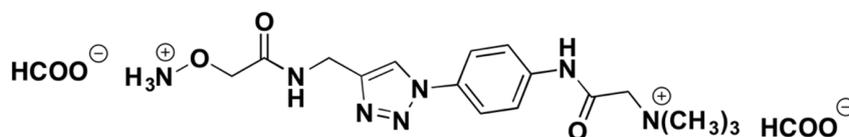
***N,N,N*-trimethyl-2-oxo-2-((4-(4-((2-((tritylamino)oxy)acetamido)methyl)-1*H*-1,2,3-triazole-1-yl)phenyl)amino)ethanaminium chloride/bromide (8a)**



First, a mixture of DCM and H₂O (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•SMe₂ (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/H₂O/7N NH₃ in methanol = 80:20:0.6:0.6). **8a** was yielded as a slightly yellow brownish solid (0.32 g, 0.50 mmol, 77%).

¹H-NMR (300 MHz, dmsO d⁶): δ /ppm = 11.67 (s, 1H, NH-C₆H₄), 8.55 (s, 1H, CH₂-C=CH-N), 8.34 (s, 1H, Ph₃C-NH), 8.32 (t, ³J = 5.8 Hz, 1H, O=C-NH-CH₂), 7.91–7.84 (m, 4H, C₆H₄), 7.34–7.19 (m, 15H, C(C₆H₅)₃), 4.53 (s, 2H, (CH₂-N(CH₃)₃), 4.45 (d, ³J = 5.8, 2H, NH-CH₂), 3.85 (s, 2H, N-O-CH₂), 3.33 (s, 9H, N(CH₃)₃). **¹³C-NMR (101 MHz, dmsO d⁶)**: δ /ppm = 169.7 (O=C-NH-CH₂), 162.4 (O=C-CH₂-N), 146.0 (CH₂-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 (CH₂-C=CH-N), 120.5 (4C, N-C=CH-CH=C-N), 73.7 (C(C₆H₅)₃), 73.2 (O-CH₂), 64.4 (CH₂-N(CH₃)₃), 53.4 (N(CH₃)₃), 33.8 (NH-CH₂). **HRMS (ESI⁺)**: calc. for C₃₅H₃₈N₇O₃⁺ [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.

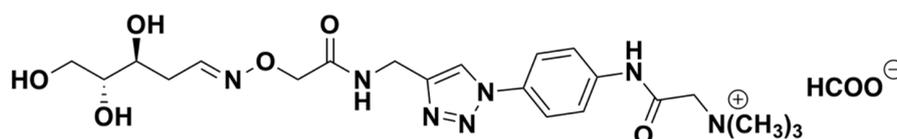
2-((4-(4-((2-(Aminoxy)acetoamido)methyl)-1H-1,2,3-triazole-1-yl)phenyl-amino)-N,N,N-trimethyl-2-oxoethanaminium formate (1a)



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 NH₃ 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.

¹H-NMR (400 MHz, D₂O): δ/ppm = 8.40 (s, 1H, CH₂-C=CH-N), 8.21 (s, 1H, HCOO), 7.58 (d, ³J = 9.2 Hz, 2H, CH-CH=C-N₃), 7.52 (d, ³J = 9.2 Hz, 2H, CH-CH=C-NH), 4.53 (s, 2H, N-O-CH₂), 4.27 (s, 2H, NH-CH₂), 4.22 (s, 2H, CH₂-N(CH₃)₃), 3.36 (s, 9H, N(CH₃)₃). **¹³C-NMR (101 MHz, D₂O):** δ/ppm = 169.7 (O=C-NH-CH₂), 162.4 (O=C-CH₂-N), 146.0 (CH₂-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 (CH₂-C=CH-N), 120.5 (4C, N-C=CH-CH=C-N), 73.7 (C(C₆H₆)₃), 73.2 (O-CH₂), 64.4 (CH₂-N(CH₃)₃), 53.4 (N(CH₃)₃), 33.8 (NH-CH₂). **HRMS (ESI⁺):** calc. for C₁₆H₂₄N₇O₃⁺ [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).

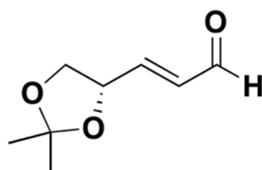
N,N,N-Trimethyl-2-oxo-2-((4-(4-((2-(((3S,4R)-3,4,5-trihydroxypentylidene)amino)oxy)acet-amido)-methyl)-1H-1,2,3-triazol-1-yl)phenyl)amino)ethanaminium formate (9a)



1a (50.0 mg, 0.12 mmol, 1.0 eq) and 2'-desoxyribose (182 mg, 1.36 mmol, 11.8 eq) were dissolved in H₂O (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 syringe filter 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.

¹H-NMR (600 MHz, D₂O): δ/ppm = 8.46 (s, 1H, HCOO), 8.34 (s, 1H, CH₂-C=CH-N), 7.79 (d, J=9.0 Hz, 2H, CH-CH=C-N₃), 7.74 - 7.71 (m, 8H, CH-CH=C-NH, C1'-H^A), 7.08 (t, ³J = 5.4 Hz, 1H, C1'-H^B), 4.67 (s, 2H, N-O-CH₂^B), 4.63 (s, 2H, NH-CH₂), 4.62 (s, 2H, N-O-CH₂^A), 4.35 (s, 2H, CH₂-N(CH₃)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(CH₃)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). **¹³C-NMR (150 MHz, D₂O):** δ/ppm = 172.4 (O=C-NH-CH₂), 170.9 (HCOO), 162.7 (O=C-CH₂-N), 153.5 (C1'^A), 153.1 (C1'^B), 145.1 (CH₂-C=C), 136.8 (N-C=CH-CH), 133.5 (N-C=CH-CH), 122.5 (2C, CH=C-NH), 122.3 (CH₂-C=CH-N), 121.9 (2C, CH=C-N₃), 74.2 (C4'), 74.0 (C4'), 71.7 (N-O-CH₂^B), 71.5 (N-O-CH₂^A), 69.0 (C3'^A), 68.8 (C3'^B), 65.1 (CH₂-N(CH₃)₃), 62.3 (C5'), 54.3 (N(CH₃)₃), 34.1 (NH-CH₂), 32.4 (C2'^A), 29.2 (C2'^B). **HRMS (ESI⁺):** calc. for C₂₁H₃₂N₇O₆⁺ [M]⁺: 478.2409; found: 478.2404.

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



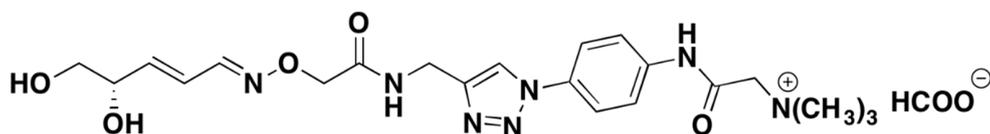
Methyl (*2E*)-3-[(*4S*)-2,2-dimethyl-1,3-dioxolan-4-yl]prop-2-enoate **11** (0.20 g, 1.08 mmol, 1.0 eq) was dissolved in DCM (2.0 mL) and cooled to -78 °C. DIBAL-H (2.20 mL, 2M in toluene, 2.1 eq) was added and the yellowish mixture was slowly warmed to rt. After 90 minutes, DCM (5.0 mL) and H₂O (4.0 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 Na₂SO₄. Volatiles were removed under reduced pressure and the allylic alcohol was isolated in quantitative yield and used without further purification.

¹H-NMR (400 MHz, CDCl₃): δ/ppm = 5.88 (dt, ³J = 15.4 Hz, ⁴J = 5.0 Hz, 1H, 5'-H) 5.65 (dd, ³J = 15.6 Hz, ⁴J = 7.6 Hz, 1H, 5'-H), 4.47 (q, ³J = 7.3 Hz, 1H, 4'-H), 4.08 (d, ³J = 5.1 Hz, 2H, 1'-H), 4.30 (dd, ³J = 8.2 Hz, ⁴J = 6.1 Hz, 1H, 3'-H), 3.53 (t, ³J = 7.9 Hz, 1H, 2'-H), 2.34 (br s, 1H, CH₂-OH), 1.36 (s, 3H, O-C(CH₃)(CH₃)-O), 1.32 (s, 3H, O-C(CH₃)(CH₃)-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 overnight. After the addition of saturated Na₂SO₄ (10 mL) and a solution of Na₂S₂O₃ (171 mg, dissolved in 10 mL H₂O), the mixture was extracted with DCM (3 x 15 mL) and dried over Na₂SO₄. 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.

¹H-NMR (400 MHz, CD₂Cl₂): δ/ppm = 9.50 (d, ³J = 7.6 Hz, 1H, 1'-CHO), 6.70 (dd, ³J = 15.6 Hz, ⁴J = 5.3 Hz, 1H, 3'-H), 6.23 (dt, ³J = 15.6 Hz, ⁴J = 5.8 Hz, 1H, 2'-H), 4.73 (q, ³J = 6.8 Hz, 1H, 4'-H), 4.18 (dd, ³J = 8.4 Hz, ⁴J = 6.8 Hz, 1H, 5'-H), 3.67 (dd, ³J = 8.4 Hz, ⁴J = 6.8 Hz, 1H, 5'-H), 1.39 (s, 3H, O-C(CH₃)(CH₃)-O), 1.35 O-C(CH₃)(CH₃)-O). **¹³C-NMR (101 MHz, CD₂Cl₂):** δ/ppm = 193.0 (-CHO), 153.4 (3'-C), 132.1 (2'-C), 110.3 (C_{quart}), 74.9 (4'-C), 68.7 (5'-C), 26.2 (O-C(CH₃)(CH₃)-O), 25.4 (O-C(CH₃)(CH₃)-O). **HRMS (EI):** calc. for C₈H₁₁O₃⁺ [M - H]⁺: 155.0708; found: 155.0707.

2-((4-(4-((2-(((1E,2E)-3-((S)-2,2-Dimethyl-1,3-dioxolane-4-yl)allylidene)-amino)oxy)acetamido)methyl)-1H-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 H₂O and CHCl₃ (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.

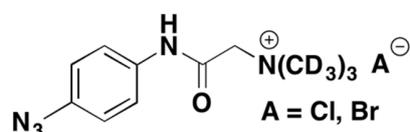
¹H-NMR (400 MHz, D₂O): δ/ppm = 8.31 (s, 1H, HCOO), 8.12 (s, 1H, CH₂-C=CH-N), 7.84 (d, 1H, 1'-H), 7.59 – 7.53 (m, 4H, CH-CH=C-N₃, CH-CH=C-NH), 6.20 – 6.03 (m, 2H, 2' + 3'-H's), 4.54 – 4.52 (m, 1H, 4'-H), 4.50 (s, 2H, N-O-CH₂), 4.47 (s, 2H, NH-CH₂), 4.20 (s, 2H, CH₂-N(CH₃)₃), 4.02 – 3.98 (m, 1H, 5'-H), 3.45 – 3.50 (m, 1H, 5'-H), 3.27 (s, 9H, CH₂-N(CH₃)₃), 1.26 (s, 3H, O-C(CH₃)(CH₃)-O), 1.24 (s, 3H, O-C(CH₃)(CH₃)-O). **HRMS (ESI⁺):** calc. for C₂₄H₃₄N₇O₅⁺ [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 • H₂O (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 lyophilization. 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.

¹H-NMR (400 MHz, D₂O): δ/ppm = 8.53 (s, 1H, HCOO), 8.30 (s, 1H, CH₂-C=CH-N), 8.01 (s, d, ³J = 8.9 Hz, 1H, 1'-H), 7.78 – 7.70 (m, 4H, CH-CH=C-N₃, CH-CH=C-NH), 6.31 – 6.32 (m, 2H, 2' + 3'-H's), 4.63 (s, 2H, N-O-CH₂), 4.62 (s, 2H, NH-CH₂), 4.35 – 4.32 (m, 3H, CH₂-N(CH₃)₃ + 4'-H), 3.61 (dd, ¹J = 11.7 Hz, ³J = 4.4, 1H,

5'-H), 3.51 (dd, $^1J = 11.7$ Hz, $^3J = 6.5$, 1H, 5'-H), 3.22 (s, 9H, $\text{CH}_2\text{-N}(\text{CH}_3)_3$). $^{13}\text{C-NMR}$ (101 MHz, D_2O): $\delta/\text{ppm} = 172.3$ ($\text{O}=\text{C-NH-CH}_2$), 170.9 (HCOO), 162.7 ($\text{O}=\text{C-CH}_2\text{-N}$), 153.7 (1'-C), 144.5 ($\text{CH}_2\text{-C}=\text{C}$), 142.8 (3'-C), 136.9 ($\text{N-C}=\text{CH-CH}$), 133.6 ($\text{N-C}=\text{CH-CH}$), 123.0 (2'-C), 122.5 (2C, $\text{CH}=\text{C-NH}$), 122.3 ($\text{CH}_2\text{-C}=\text{CH-N}$), 122.0 (2C, $\text{CH}=\text{C-N}_3$), 72.0 (N-O-CH_2) 71.6 (4'-C), 64.4 (5'-C), 65.1 ($\text{CH}_2\text{-N}(\text{CH}_3)_3$), 54.3 ($\text{N}(\text{CH}_3)_3$), 34.1 (NH-CH_2). **HRMS (ESI⁺)**: calc. for $\text{C}_{21}\text{H}_{30}\text{N}_7\text{O}_5^+$ [M]⁺: 460.2303; found: 460.2305.

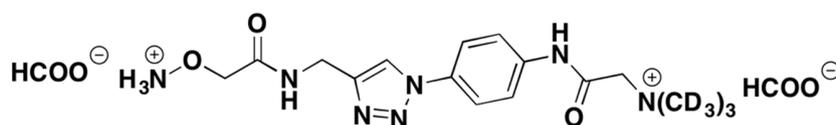
2-((4-Azidophenyl)amino-*N,N,N*-tri(methyl-*d*₃)-2-oxoethanaminium chloride (4b)



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.

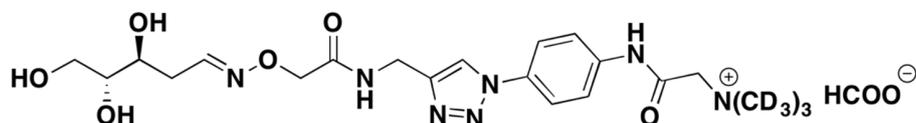
$^1\text{H-NMR}$ (600 MHz, D_2O): $\delta/\text{ppm} = 7.39$ (d, $^3J=8.6$, 2H, $\text{CH-CH}=\text{C-NH}$), 7.04 (d, $^3J=8.5$, 2H, $\text{CH-CH}=\text{C-N}_3$), 4.18 (s, 2H, CH_2). $^{13}\text{C-NMR}$ (150 MHz, D_2O , ppm): $\delta/\text{ppm} = 162.7$ (C=O), 137.5 ($\text{NH-C}=\text{CH}$), 132.5 ($\text{N}_3\text{-C}=\text{CH}$), 123.5 (2C, $\text{NH-C}=\text{CH}$), 119.6 (2C, $\text{N}_3\text{-C}=\text{CH}$), 65.0 (CH_2). **HRMS (ESI⁺)**: calc. for $\text{C}_{11}\text{H}_7\text{D}_9\text{N}_5\text{O}^+$ [M]⁺: 243.1914; found: 243.1916.

2-((4-(4-((2-(Aminoxy)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.

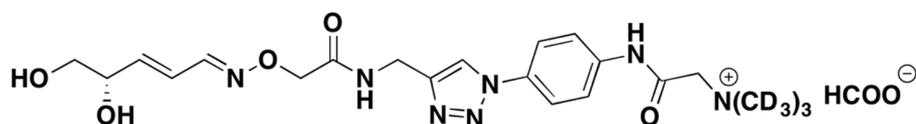
***N,N,N*-Tri(methyl-*d*₃)-2-oxo-2-((4-(4-((2-(((3*S*,4*R*)-3,4,5-trihydroxypentylidene)amino)oxy)acetamido)-methyl)-1*H*-1,2,3-triazol-1-yl)phenyl)amino)ethanaminium formate (9b)**



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

¹H-NMR (600 MHz, D₂O): δ/ppm = 8.46 (s, 2H, HCOO), 8.35 (s, 1H, CH₂-C=CH-N), 7.81 (d, ³J = 7.5 Hz, 2H, CH-CH=C-N₃), 7.75 - 7.73 (m, 8H, CH-CH=C-NH, C1'-H^A), 7.08 (t, ³J = 5.4 Hz, 0.1H, C1'-H^B), 4.67 (s, 2H, N-O-CH₂^B), 4.64 (s, 2H, NH-CH₂), 4.62 (s, 2H, N-O-CH₂^A), 4.34 (s, 2H, CH₂-N(CH₃)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'-H₂), 3.67–3.53 (m, 2H, 1x C5'-H₂, C4'-H), 2.80–2.68 (m, 2H, C2'-H₂^B), 2.59 - 2.54 (m, 1H, C2'-H^A), 2.43–2.34 (m, 1H, C2'-H^A). **¹³C-NMR (150 MHz, D₂O, ppm):** δ/ppm = 172.4 (O=C-NH-CH₂), 170.9 (HCOO), 162.8 (O=C-CH₂-N), 153.5 (C1'^A), 153.1 (C1'^B), 145.1 (CH₂-C=C), 136.8 (N-C=CH-CH), 133.6 (N-C=CH-CH), 122.6 (2C, CH=C-NH), 122.4 (CH₂-C=CH-N), 122.0 (2C, CH=C-N₃), 74.2 (C4'), 74.0 (C4'), 71.7 (N-O-CH₂^B), 71.5 (N-O-CH₂^A), 69.0 (C3'^A), 68.8 (C3'^B), 64.9 (CH₂-N(CD₃)₃), 62.3 (C5'), 53.3 (N(CD₃)₃), 34.1 (NH-CH₂), 32.4 (C2'^A), 29.2 (C2'^B). **HRMS (ESI⁺):** calc. for C₂₁H₂₃D₉N₇O₆⁺ [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**.

¹H-NMR (800 MHz, D₂O): δ/ppm = 8.47 (s, 1H, HCOO), 8.32 (s, 2H, CH₂-C=CH-N), 8.02 (d, ³J = 9.4 Hz, 1H, 1'-H), 7.80 - 7.71 (m, 4H, CH-CH=C-N₃, CH-CH=C-NH), 6.34 - 6.26 (m, 2H, 2' + 3'-H's), 4.65 (s, 2H, N-O-CH₂), 4.63 (s, 2H, NH-CH₂), 4.35 - 4.32 (m, 3H, CH₂-N(CH₃)₃ + 4'-H), 3.63 (dd, ¹J = 11.7 Hz, ³J = 4.4, 1H, 5'-H), 3.53 (dd, ¹J = 11.7 Hz, ³J = 6.5, 1H, 5'-H). **¹³C-NMR (150 MHz, D₂O):** δ/ppm = 172.3 (O=C-NH-CH₂), 170.9 (HCOO), 162.8 (O=C-CH₂-N), 153.7 (1'-C), 145.2 (CH₂-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, CH=C-NH), 122.3 (CH₂-C=CH-N), 122.0 (2C, CH=C-N₃), 72.1 (N-O-

CH₂) 71.6 (4'-C), 64.4 (5'-C), 65.1 (CH₂-N(CH₃)₃), 53.3 (N(CD₃)₃), 34.1 (NH-CH₂). **HRMS (ESI⁺)**: calc. for C₂₁H₂₁D₉N₇O₅⁺ [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* [¹³C₁₀¹⁵N₅] at a concentration of 100 µM for three days, followed by treatment with 200 µM labeled dG* for two days; dC* [¹³C₉¹⁵N₃] and dT* [¹³C₁₀¹⁵N₂] 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.

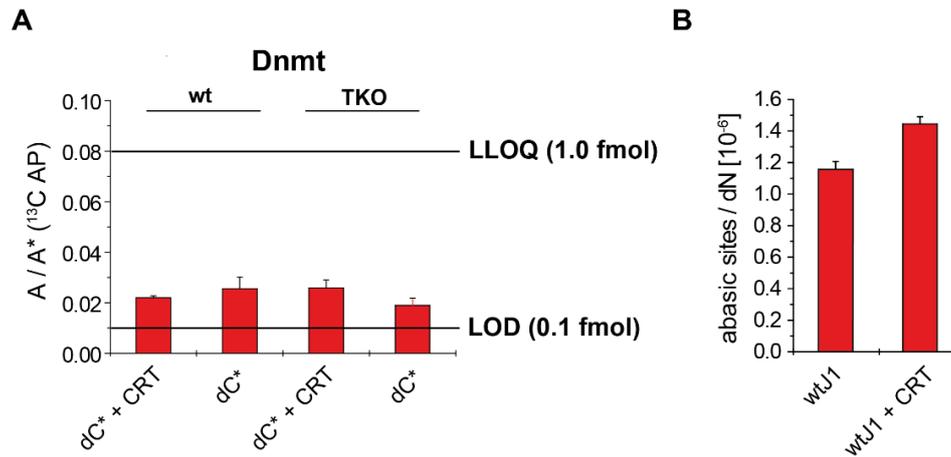


Figure S1: AP-sites after inhibition of Ape1. **(A)** A/A* ratios of labeled AP-sites in wild type and Dnmt TKO mESCs were over the LOD but below the LLOQ. **(B)** Quantification of unlabeled AP-sites of the dC* feeding experiment show an increase by treatment with the inhibitor CRT0044876.

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 μ M 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 μ L for *genomic tips 20G* or 100 μ L for *genomic tips 100G*) and RNase A (2.0 μ L/20G, 10 μ L/100G) at 50 °C for one hour. After 30 minutes, additional RNase A (2.0 μ L or 10 μ L, 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 μ M 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 ddH₂O (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 µg, *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.

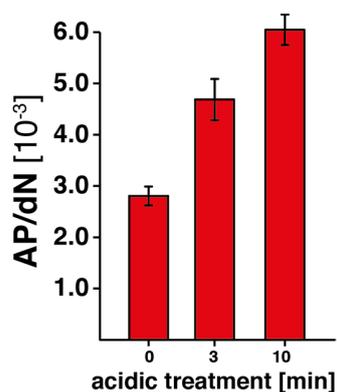


Figure S2: Artificial generation of AP-sites under harsh conditions.

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.

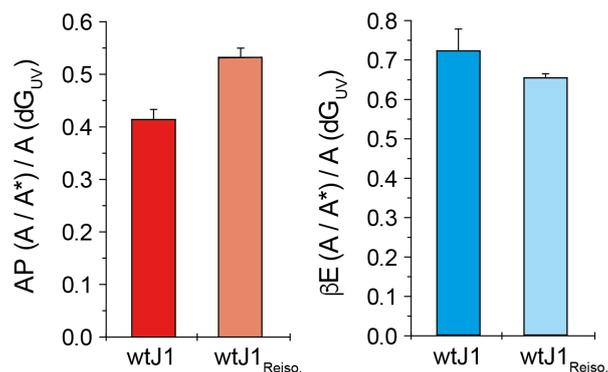


Figure S3: A/A* ratios of AP- and βE-sites normalized by dG. The DNA of wild type mESCs shows similar values of the BER intermediates before and after reisolation.

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

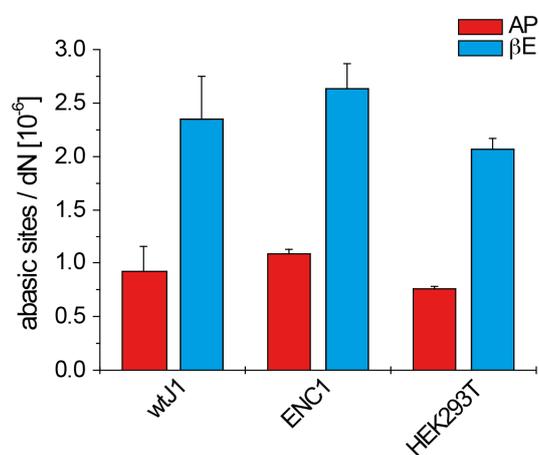


Figure S4: Quantification of AP- and βE-intermediates in wild type mESCs, ENC1 neural stem cells and somatic HEK293T cells shows similar values for the BER intermediates in all three cell lines. The quantification results of one biological replicate is depicted.

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 ddH₂O 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'

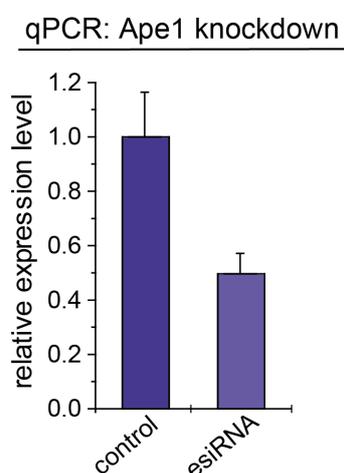


Figure S5: Relative mRNA expression levels show an estimated knockdown efficiency of about 50% using Ape1 (Apex1) esiRNA. As housekeeping reference gene alpha-tubulin was used.

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 $^\circ\text{C}$ /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 $^\circ\text{C}$ /1400 rpm. Derivatized DNA was then precipitated via addition of NaOAc (3.3 μL , 3M), vortexing and incubation at 37 $^\circ\text{C}$ /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 $^\circ\text{C}$, 15000 rpm). The supernatant was removed and washing steps were carried out (1x75% *i*-PrOH, 10 $^\circ\text{C}$, 15000 rpm, 30 minutes; 2x75% cold EtOH, 4 $^\circ\text{C}$, 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 ddH₂O 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 H₂O) was incubated in an aqueous solution containing 480 µM ZnSO₄, 5 U Antarctic phosphatase (*New England BioLabs*) and 42 U nuclease S1 (*Aspergillus 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 µg 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 work⁶⁻¹⁰, 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 H₂O 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 (N₂), nebulizer 30 psi, sheath gas heater 275 °C, sheath gas flow 11 l/min (N₂), 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	<i>wide</i>	158.03	<i>wide</i>	170	5	5	Positive
5cadC	272.09	<i>wide</i>	156.04	<i>wide</i>	170	5	5	Positive
$[^{15}N_2, D_2]5hmdC$	262.12	<i>enhanced</i>	146.07	<i>enhanced</i>	40	27	1	Positive
5hmdC	258.11	<i>enhanced</i>	142.06	<i>enhanced</i>	40	27	1	Positive
$[D_3]5mdC$	245.13	<i>enhanced</i>	129.09	<i>enhanced</i>	30	60	1	Positive
5mdC	242.11	<i>enhanced</i>	126.07	<i>enhanced</i>	30	60	1	Positive
dC	228.12	<i>enhanced</i>	112.05	<i>enhanced</i>	25	5	5	Positive
$[^{13}C_9, ^{15}N_3]dC$	240.12	<i>enhanced</i>	119.06	<i>enhanced</i>	25	5	5	Positive
Time segment 3.8–5.8 minutes								
$[D_2]5hmdU$	259.09	<i>wide</i>	216.08	<i>wide</i>	48	7	5	Negative
5hmdU	257.08	<i>wide</i>	214.07	<i>wide</i>	48	7	5	Negative
$[^{15}N_2]5fdU$	257.06	<i>wide</i>	213.05	<i>wide</i>	48	6	5	Negative
5fdU	255.06	<i>wide</i>	212.06	<i>wide</i>	48	6	5	Negative
Time segment 5.8–8.1 minutes								
$[^{15}N_5]8oxodG$	289.08	<i>wide</i>	173.04	<i>wide</i>	90	9	7	Positive
8oxodG	284.10	<i>wide</i>	168.05	<i>wide</i>	90	9	7	Positive
dG	268.10	<i>wide</i>	152.06	<i>wide</i>	75	45	3	Positive
$[^{13}C_{10}, ^{15}N_5]dG$	283.12	<i>wide</i>	162.06	<i>wide</i>	75	45	3	Positive
$[^{15}N_2]5fdC$	258.09	<i>wide</i>	142.04	<i>wide</i>	50	5	5	Positive
5fdC	256.09	<i>wide</i>	140.05	<i>wide</i>	50	5	5	Positive
dT	243.1	<i>enhanced</i>	127.05	<i>enhanced</i>	35	40	3	Positive
$[^{13}C_{10}, ^{15}N_2]dT$	255.12	<i>wide</i>	130.07	<i>wide</i>	50	8	5	Positive
Time segment 8.1–12.2 minutes								
1-Naphthyl-Oxim	500.24	<i>wide</i>	472.23	<i>wide</i>	5	19	5	Positive
9b_1	487.30	<i>wide</i>	459.29	<i>wide</i>	38	19	5	Positive
9b_2	487.30	<i>wide</i>	201.18	<i>wide</i>	38	40	5	Positive
$[^{13}C_5]9a_1$	483.26	<i>wide</i>	455.25	<i>wide</i>	38	19	5	Positive
$[^{13}C_5]9a_2$	483.26	<i>wide</i>	192.13	<i>wide</i>	38	40	5	Positive
9a_1	478.24	<i>wide</i>	450.23	<i>wide</i>	38	19	5	Positive
9a_2	478.24	<i>wide</i>	192.13	<i>wide</i>	38	40	5	Positive
10b_1	469.29	<i>wide</i>	441.28	<i>wide</i>	38	19	3	Positive
10b_2	469.29	<i>wide</i>	201.18	<i>wide</i>	38	33	3	Positive
$[^{13}C_5]10a_1$	465.23	<i>wide</i>	437.22	<i>wide</i>	38	19	3	Positive
$[^{13}C_5]10a_2$	465.23	<i>wide</i>	192.13	<i>wide</i>	38	34	3	Positive
10a_1	460.23	<i>wide</i>	432.22	<i>wide</i>	38	19	3	Positive
10a_2	460.23	<i>wide</i>	192.13	<i>wide</i>	38	34	3	Positive
1b	371.25	<i>wide</i>	343.24	<i>wide</i>	5	19	5	Positive
1a	362.19	<i>wide</i>	334.19	<i>wide</i>	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; 1.5 - 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	<i>wide</i>	459.29	<i>wide</i>	38	13	2	Positive
9b_2	487.30	<i>wide</i>	201.18	<i>wide</i>	38	25	2	Positive
9a_1	478.24	<i>wide</i>	450.23	<i>wide</i>	38	13	2	Positive
9a_2	478.24	<i>wide</i>	192.13	<i>wide</i>	38	25	2	Positive
1b	371.25	<i>wide</i>	343.24	<i>wide</i>	5	10	1	Positive
1a	362.19	<i>wide</i>	334.19	<i>wide</i>	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.

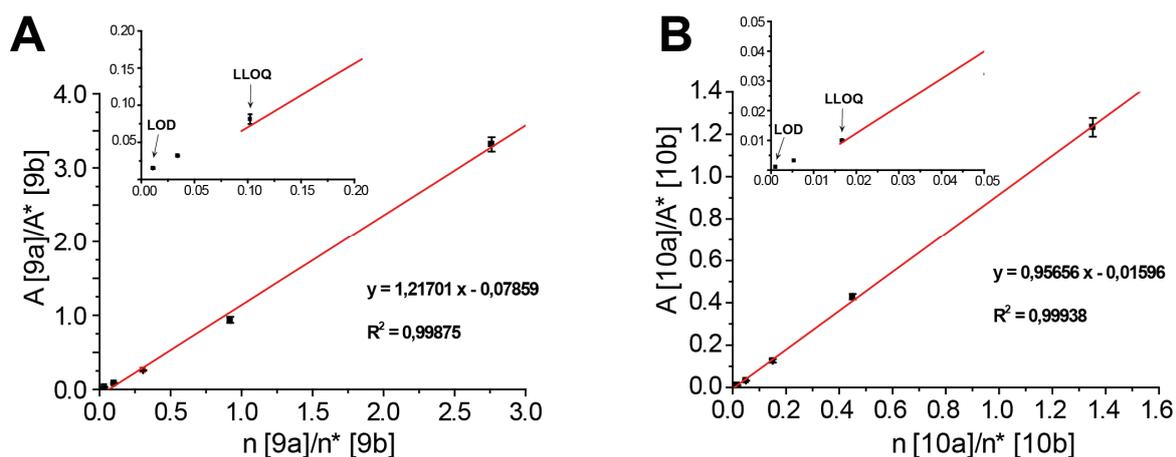


Figure S6: Calibration curves for **(A)** AP- (**9**) and **(B)** β E-site (**10**).

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 [μ g]. The total amounts of nucleosides were obtained by using the average molar mass of $308.91 \text{ g mol}^{-1}$ 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 CHCl₃/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 CHCl₃/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 ddH₂O (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), Na₂EDTA (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-C₁₈* 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 H₂O and buffer B = 10 mM TEAB, pH = 7.5 in 80% MeCN/H₂O). 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.

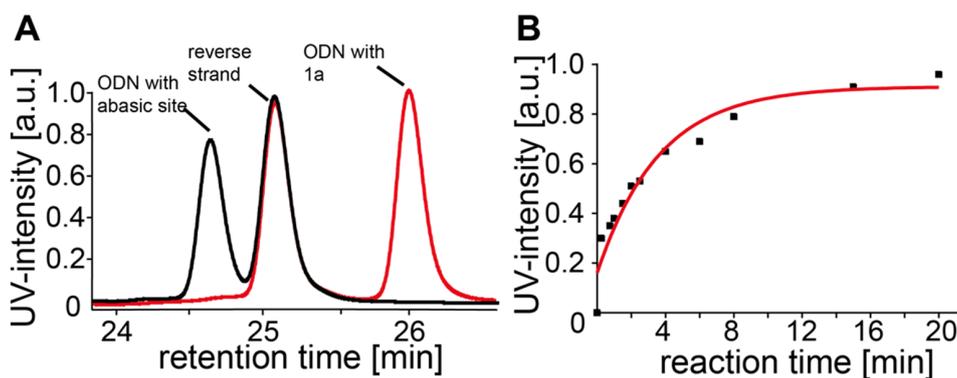


Figure S7: Reaction kinetics on oligo with defined abasic site. (A) Obtained UV-signals of ODN with abasic site and reverse strand before (black lines) and after derivatization with **1a**. **(B)** Normalized UV signals of ODN + **1** after specific time points.

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

An 8oxoG containing synthetic 13mer 5'-GTAATG8oxoGGCTAGG-3' and its counter strand 5'-CCTAGCCATTAC-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.

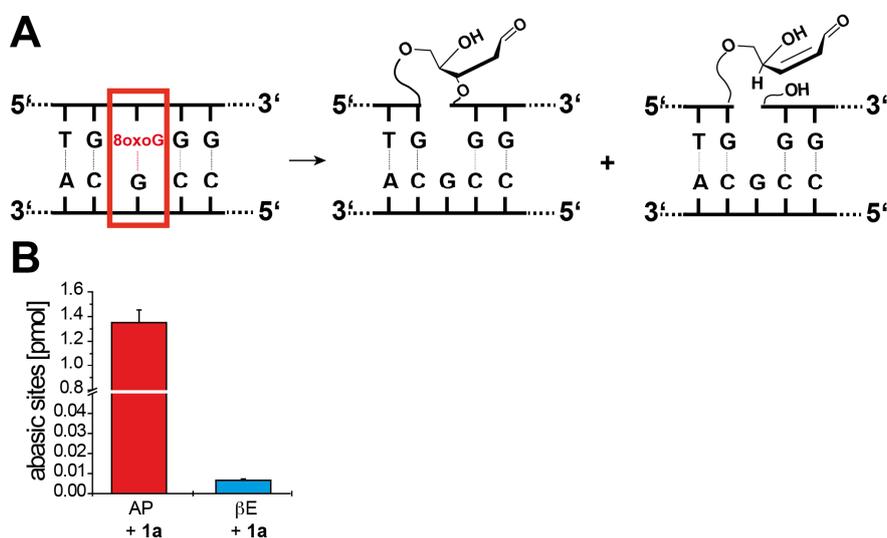


Figure S8: Reaction of an 8oxodG containing oligonucleotide with hOGG1. (A) Reaction of hOGG1 creating a defined AP- and β E-site inside a DNA duplex. **(B)** Quantification of the generated AP- and β E-sites after treatment of the oligonucleotide with **1a**.

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).

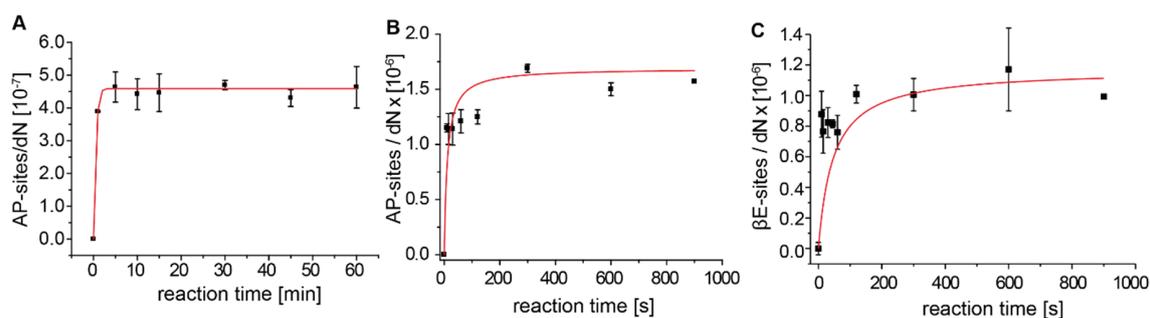


Figure S9: Derivatization of gDNA with **1** shows a fast derivatization reaction and does not artificially generate abasic sites.

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 3.2.2

M. Su, A. Kirchner, S. Stazzoni, M. Müller, M. Wagner, A. Schröder, T. Carell, 5-Formylcytosine Could Be a Semipermanent Base in Specific Genome Sites, *Angew. Chem., Int. Ed. Engl.* **2016**, *55*, 11797-11800.

Supporting Information

5-Formylcytosine Could Be a Semipermanent Base in Specific Genome Sites

*Meng Su, Angie Kirchner, Samuele Stazzoni, Markus Müller, Mirko Wagner, Arne Schröder, and Thomas Carell**

anie_201605994_sm_miscellaneous_information.pdf

Table of contents

1. General methods of organic synthesis
2. Synthesis of the hydroxylamine linker
3. ^1H -NMR spectra of the linker
4. General methods for oligonucleotide synthesis
5. Crosslinking studies with the synthesized strands
6. Experimental details of the genomic fdC profiling study
7. Quantification modeling
8. Droplet digital PCR data

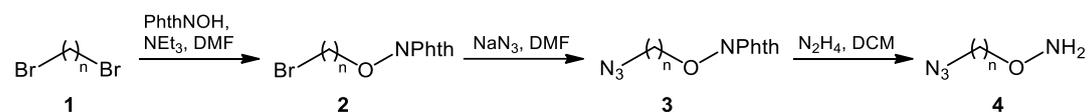
1. General methods of organic synthesis

Chemicals were purchased from *Sigma-Aldrich* and used without further purification. The solvents for organic synthesis were of reagent grade and purified by distillation. Solutions were concentrated *in vacuo* on a *Heidolph* rotary evaporator with a *Vario PC2001* diaphragm pump by *Vacuubrand*. All mixed solvent systems are reported as v/v solutions. All reactions were monitored by thin-layer chromatography (TLC), performed on *Merck* 60 (silica gel F₂₅₄) plates. Chromatographic purification of products was accomplished using flash column chromatography on silica gel (230-400 mesh) purchased from *Merck*.

¹H- and ¹³C-NMR spectra were recorded in deuterated solvents on *Bruker ARX* 400 spectrometers and calibrated to the residual solvent peak. Chemical shifts (δ , ppm) are quoted relative to the residual solvent peak as internal standard and coupling constants (*J*) are corrected and quoted to the nearest 0.1 Hz. Multiplicities are abbreviated as follows: s = singlet, d = doublet, t = triplet, m = multiplet.

2. Synthesis of the hydroxylamine linker

O-(4-Azidobutyl)hydroxylamine **4**



1,4-Dibromobutane **1** (5.9 mL, 49.4 mmol, 2.0 eq.) was added to a solution of *N*-hydroxyphthalimide (PhthNOH, 4.0 g, 24.5 mmol, 1.0 eq.) and triethylamine (7.5 mL, 53.6 mmol, 2.2 eq.) in anhydrous dimethylformamide. The mixture was stirred at room temperature for 24 h. The reaction was diluted with water, and the aqueous phase was extracted three times with ethyl acetate. The combined organic phases were dried over MgSO₄, filtered and concentrated to give the crude product **2** (5.02 g, 16.9 mmol, 0.69 eq.) as a white solid. The residue was dissolved in anhydrous dimethylformamide and sodium azide (1.32 g, 20.6 mmol, 0.85 eq.) was added. The mixture was stirred at room temperature for 2 h, diluted with water and extracted with ethyl acetate three times. The combined organic phases were dried over MgSO₄, filtered and concentrated. The crude was purified by flash chromatography on silica gel (*iso*-hexane/ethyl acetate 10:1 → 2:1) to give **3** (3.97 g, 15.2 mmol, 0.62 eq.) as a yellow oil. The oil was redissolved in hydrazine monohydrate (1.1 mL, 22.8 mmol, 0.93 eq.) and dichloromethane (10 mL). The mixture was stirred at room temperature for 24 h and then filtered. The solution was diluted with dichloromethane and washed with NaCl aq. three times. The combined organic phases were dried over MgSO₄,

filtered and concentrated to give **4** (1.81 g, 13.9 mmol, 57% yield over three steps) as a colorless oil.

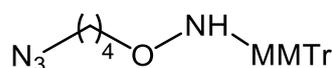
$R_f = 0.42$ (DCM/ MeOH 10:1).

$^1\text{H-NMR}$ (400 MHz, CDCl_3): $\delta = 3.70$ (t, $^3J_{\text{H,H}} = 5.6$ Hz, 2H, O-CH₂-CH₂), 3.38 (t, $^3J_{\text{H,H}} = 6.4$ Hz, 2H, CH₂-N₃), 1.68–1.64 (m, 4H, CH₂-CH₂-CH₂-CH₂).

$^{13}\text{C-NMR}$ (100 MHz, CDCl_3): $\delta = 75.1$ (-O-CH₂-CH₂), 51.2 (CH₂-N₃), 25.6 (CH₂), 25.5 (CH₂).

HRMS (ESI+): calculated for $\text{C}_4\text{H}_{11}\text{ON}_4^+$ $[\text{M}+\text{H}]^+$: 131.0927, found: 131.0928.

***O*-(4-Azidobutyl)-*N*-[(4-methoxyphenyl)diphenylmethyl]hydroxylamine (**5**)**



5

O-(4-Azidobutyl)hydroxylamine **4** (1.88 g, 14.4 mmol, 1.0 eq.) was dissolved in anhydrous dichloromethane (40 mL). 4-Monomethoxytritylchloride (MMTr-Cl, 4.91 g, 15.9 mmol, 1.1 eq.) and diisopropylethylamine (5.0 mL, 28.9 mmol, 2.0 eq.) was added to the mixture at 0°C. The reaction was stirred at room temperature for 2 h, diluted with dichloromethane, washed with saturated NaHCO_3 , dried over MgSO_4 , filtered and concentrated. The crude was purified by flash chromatography on silica gel (*iso*-hexanes/ethyl acetate 15:1 + 3% triethylamine) to give **5** (4.89 g, 12.2 mmol, 84%) as a yellowish oil.

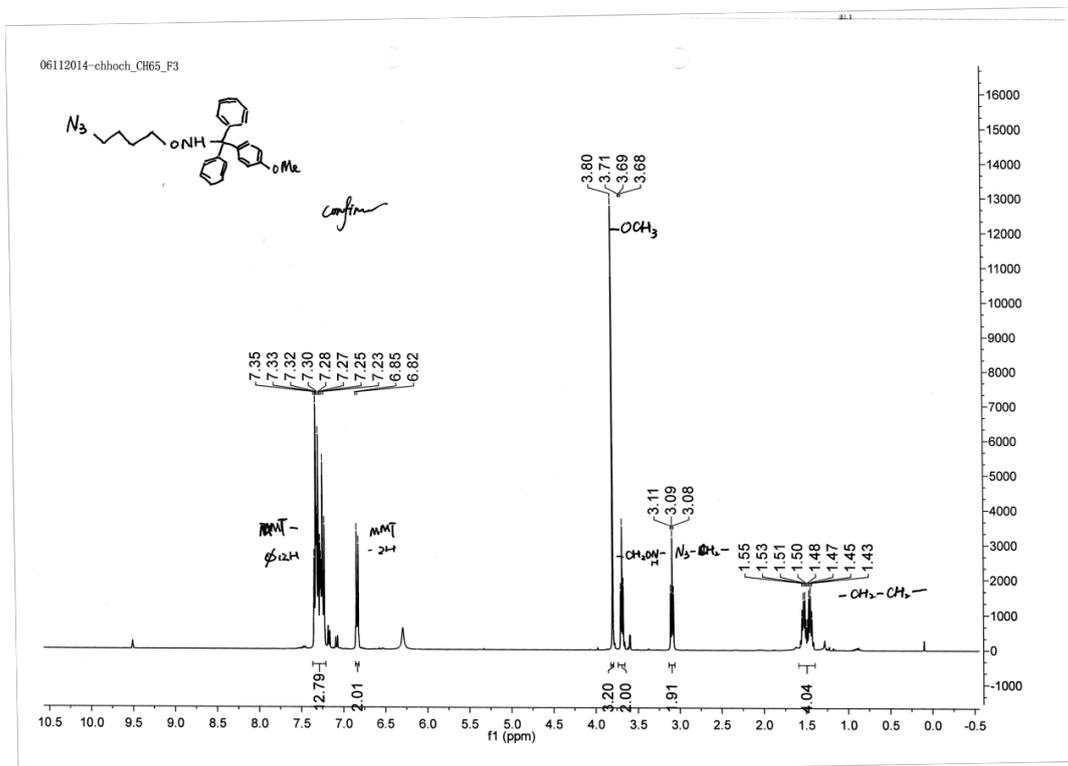
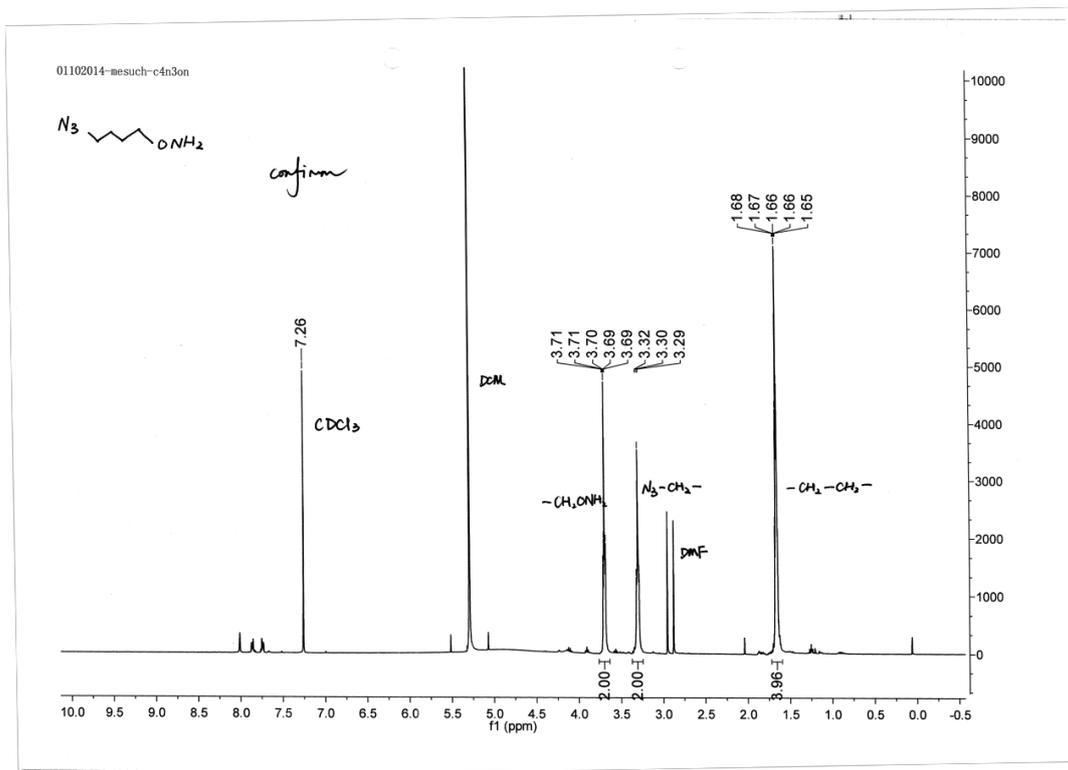
$R_f = 0.64$ (*iso*-hexane/ ethyl acetate 10:1 + 3% triethylamine).

$^1\text{H-NMR}$ (400 MHz, CDCl_3): $\delta = 7.31$ –7.18 (m, 12H, $12 \times \text{C}_{\text{Ar}}\text{H}$), 6.79 (d, $^3J_{\text{H,H}} = 8.8$ Hz, 2H, $2 \times \text{CH}_3\text{-O-C-CH}$), 3.76 (s, 3H, O-CH₃), 3.65 (t, $^3J_{\text{H,H}} = 6.0$ Hz, 2H, O-CH₂-CH₂), 3.05 (t, $^3J_{\text{H,H}} = 6.8$ Hz, 2H, N₃-CH₂), 1.52–1.37 (m, 4H, N₃-CH₂-CH₂-CH₂-CH₂).

$^{13}\text{C-NMR}$ (100 MHz, CDCl_3): $\delta = 158.3$ (CH₃-O-C), 144.6 ($2 \times \text{C}_{\text{Ar}}$), 136.6 (C_{Ar}), 130.2 ($2 \times \text{C}_{\text{Ar}}$), 129.0 ($4 \times \text{C}_{\text{Ar}}$), 127.6 ($4 \times \text{C}_{\text{Ar}}$), 126.7 ($2 \times \text{C}_{\text{Ar}}$), 112.9 ($2 \times \text{C}_{\text{Ar}}$), 77.2 (O-NH-C), 73.2 (O-CH₂), 55.2 (O-CH₃), 51.1 (N₃-CH₂), 25.8 (N₃-CH₂-CH₂), 25.5 (O-CH₂-CH₂).

HRMS (ESI-): calculated for $\text{C}_{25}\text{H}_{27}\text{O}_4\text{N}_4^-$ $[\text{M}+\text{HCO}_2]^-$: 447.2038, found: 447.2040.

3. ¹H-NMR spectra of the linker



4. General methods for oligonucleotide synthesis

DNA Oligonucleotide synthesis was performed on an Applied Biosystems Incorporated 394 automated synthesizer. Phosphoramidites and solid supports columns were purchased from *Glen Research*, *Link Technology*, and *ChemGene Corporation*. Oligodeoxynucleotides were synthesized in a 1 μmol scale with standard DNA synthesis cycles (trityl off mode). Coupling time for modified nucleosides was extended to 10 min. The oligonucleotide was cleaved using conc. ammonium hydroxide aq. at 25 $^{\circ}\text{C}$ for 17 h. The aqueous solution was then collected and evaporated in a *SpeedVac* concentrator, and the pellet was redissolved in ddH₂O.

Analytical RP-HPLC was performed using a *Macherey-Nagel* Nucleodur 100-3 C18ec column on 2695 Separation Module equipped with a Waters Alliance 2996 Photodiode Array Detector using a flow of 0.5 mL/min. Semi-preparative RP-HPLC was performed using a *Macherey-Nagel* C18 column (5 mm, 9.4 \times 250 mm) on *Waters* Breeze 2487 Dual λ Array Detector, 1525 Binary HPLC Pump. Conditions: Buffer A, 0.1 M TEAA (triethylammonium acetate) in water; buffer B, 0.1 M TEAA in 80% acetonitrile.

The purified fractions were concentrated and characterized by Matrix Assisted Laser Desorption Ionization Time of Flight (MALDI-TOF) on a Bruker Daltonics Autoflex II instrument. The concentration of the oligonucleotide solutions was calculated from the UV absorbance at 260 nm on a Nanodrop ND-1000 spectrophotometer. Extinction coefficients of the oligonucleotides at 260 nm were calculated by addition of the extinction coefficients of the individual nucleobases: dA 15.0 L/(mmol \cdot cm), dC 7.1 L/(mmol \cdot cm), dG 12.0 L/(mmol \cdot cm), dT 8.4 L/(mmol \cdot cm), mdC 7.8 L/(mmol \cdot cm), hmdC 8.7 L/(mmol \cdot cm), fdC 11.3 L/(mmol \cdot cm), cadC 7.1 L/(mmol \cdot cm). The 1,2,3-triazole and abasic monomer are transparent at 260 nm.

Click reaction on the solid support with azide linkers and deprotection

After solid phase synthesis (0.2 μmol scale, approx. 50% yield for 13 mer, calculated as 0.1 μmol), the solid support was suspended in dimethyl sulfoxide (80 μL) and acetonitrile (25 μL). To the mixture, CuSO₄ aq. (100 mM, 50 μL , 5.0 μmol , 50 eq.), sodium ascorbate aq. (500 mM, 20 μL , 10 μmol , 100 eq.), *N,N*-diisopropylethylamine solution in acetonitrile (200 mM, 75 μL , 15 μmol , 150 eq.), solution of **5** in dimethyl sulfoxide (100 mM, 50 μL , 5.0 μmol , 50 eq.) were added. The reaction was conducted at 25 $^{\circ}\text{C}$ for 24 h. Afterwards, the solid support was washed with dimethyl sulfoxide, dilute NaHCO₃ aq., acetonitrile, ether and air-dried to a powder. The solid phase was then cleaved with conc. aqueous NH₃ at 25 $^{\circ}\text{C}$ for 17 h, purified by HPLC and concentrated. The removal of the MMTr protection group on the hydroxylamine was carried out by dissolving the obtained oligonucleotide in acetic acid aq. (20%, 200 μL) at 25 $^{\circ}\text{C}$ for 30 min, precipitated by addition of sodium acetate solution (3 M, 60 μL) and EtOH (1040 μL), and then purified again with HPLC.

Schiff base formation between fdC-oligonucleotides and probe strands

A mixture of fdC containing oligonucleotides (**T1,2**), or T/C/mC/hmC/caC/Abasic containing oligonucleotides (**T3-8**) in control experiments (15 μ M, 20 μ L, 0.3 nmol, 1 eq.), probe strands (15 μ M, 20 μ L, 0.3 nmol, 1 eq.), NaCl aq. (1 M, 15 μ L), NaOAc aq. (pH = 6.0, 100 mM, 15 μ L) and ddH₂O (80 μ L) was prepared to make a final volume of 150 μ L (oligonucleotide working concentration 2 μ M). The mixture was heated to 85 °C for 5 min then slowly cooled down to 25 °C in 3 h. A first aliquot (15 μ L) was taken and quenched before 5.4 μ L of 4-methoxyaniline solution (250 mM, ddH₂O/DMSO, v/v 9/1, acidified to pH = 5.5 with acetate acid) was added to give a catalyst working concentration of 10 mM. The reaction was conducted at 25 °C, 500 rpm for 24 h. Aliquots (15 μ L) were taken at 1, 2, 4, 6, 8, 12, 24 h and quenched by addition of loading buffer. All the samples were heated at 85 °C for 3 min followed by PAGE assay as mentioned above.

When using 1,4-benzenediamine as the catalyst, a 10 mM stock solution of 10 mM in 0.5% acetic acid aq. was prepared.

Melting point experiments

Melting profiles were measured on a JASCO V-650 spectrometer using quartz glass cuvettes with 10 mm path length. The samples contained 100 mM NaCl, 10 mM NaOAc buffer pH 6.0 and 1 μ M of each strand in a final volume of 200 μ L. The measurement was repeated three times with independent sample. Before the measurement, the oligonucleotides were hybridized by slowly cooling down the samples from 85 °C to room temperature. The solutions were covered with silicon oil and tightly plugged. Absorbance was recorded in the forward and reverse direction at temperatures from 25 °C (or 15 °C) to 85 °C with a slope of 1 °C/min. T_M values were calculated as the zero-crossing of 2nd derivate of the 339 nm background corrected change in hyperchromicity at 260 nm.

Table S1. Synthesized oligonucleotides in this study. (Letters in bold and italic stand for 2'-*O*-propargyl nucleosides or epigenetic modifications.)

Entry	Description	5'-----3'	Calc.	Exptl.
T1	ODN-fC	GTAATG fC GCTAGG	4040.9	4036.2
T2	fC-shift	GTAAT fC GCTAGG	4000.9	3999.6
T3	fC-T	GTAATG T GCTAGG	4027.7	4024.7
T4	fC-C	GTAATG C GCTAGG	4012.7	4008.6
T5	fC-mC	GTAATG mC GCTAGG	4026.7	4021.6
T6	fC-hmC	GTAATG hmC GCTAGG	4042.7	4039.6
T7	fC-caC	GTAATG caC GCTAGG	4056.7	4054.9
T8	fC-Abasic	GTAATG AP GCTAGG	3919.6	3916.4
	Probe-1-alkyne	CC U AGCGCATTAC	3932.7	3930.5
P1	Probe-1-MMTr	CC U AGCGCATTAC	4335.2	4335.1
	Probe-1-ONH ₂	CC U* AGCGCATTAC	4084.8 ^[a]	4084.2
	Probe-2-MMTr	CC U ATCGCATTAC	4310.2	4310.3
P2	Probe-2-ONH ₂	CC U* ATCGCATTAC	4059.8 ^[a]	4064.3

[a] contains one sodium ion

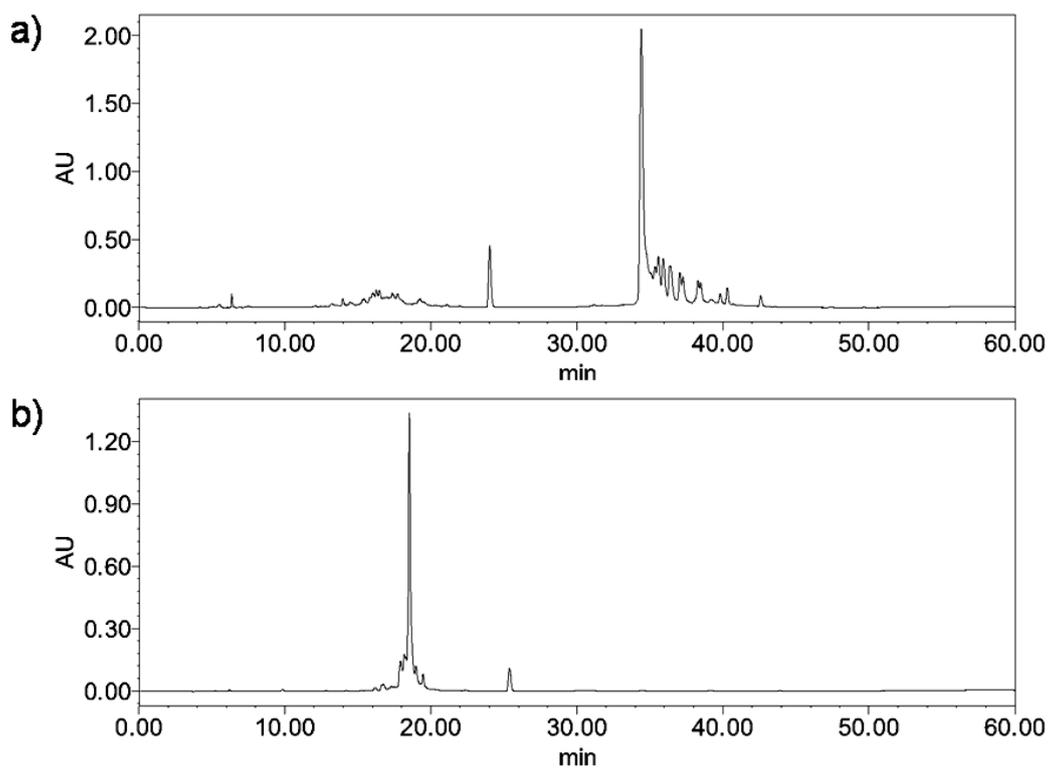


Figure S1. Typical HPLC trace of crude product a) **P1-MMTr** and b) **P1-ONH₂**. Conditions: buffer A, 0.1 M TEAA; buffer B, 0.1 M TEAA in 80% acetonitrile, linear gradient from 0% to 60% B over 45 min. Retention time: (a) 34.4 min, (b) 18.5 min. AU = arbitrary unit.

5. Crosslinking studies with the synthesized strands

T1: 5'-GTAATGfCGCTAGG-3' T2: 5'-GTAATfCCGCTAGG-3'
P1: 3'-CATTACGCGAU*CC-5' P1: 3'-CATTACGCGAU*CC-5'

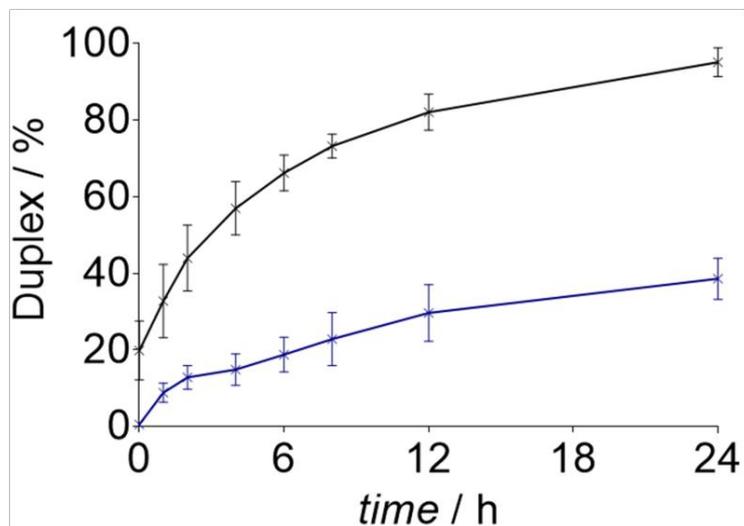


Figure S2. Quantification of the DNA duplex formation during the reaction using the catalyst 4-methoxyaniline. Black line: duplex formation between T1 and P1, blue line: duplex formation between T2 and P1. Error bars represent the standard error of the mean calculated from three replicates.

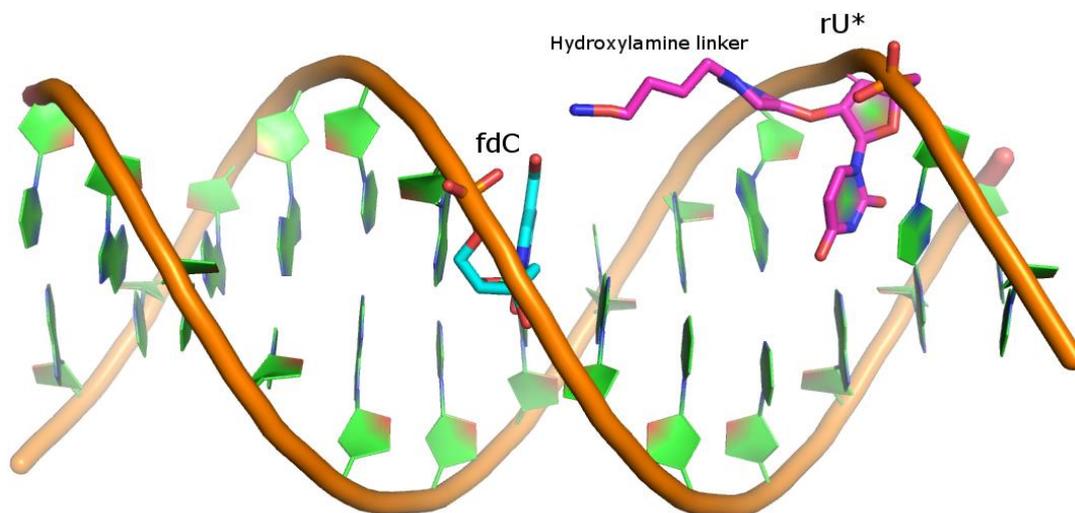
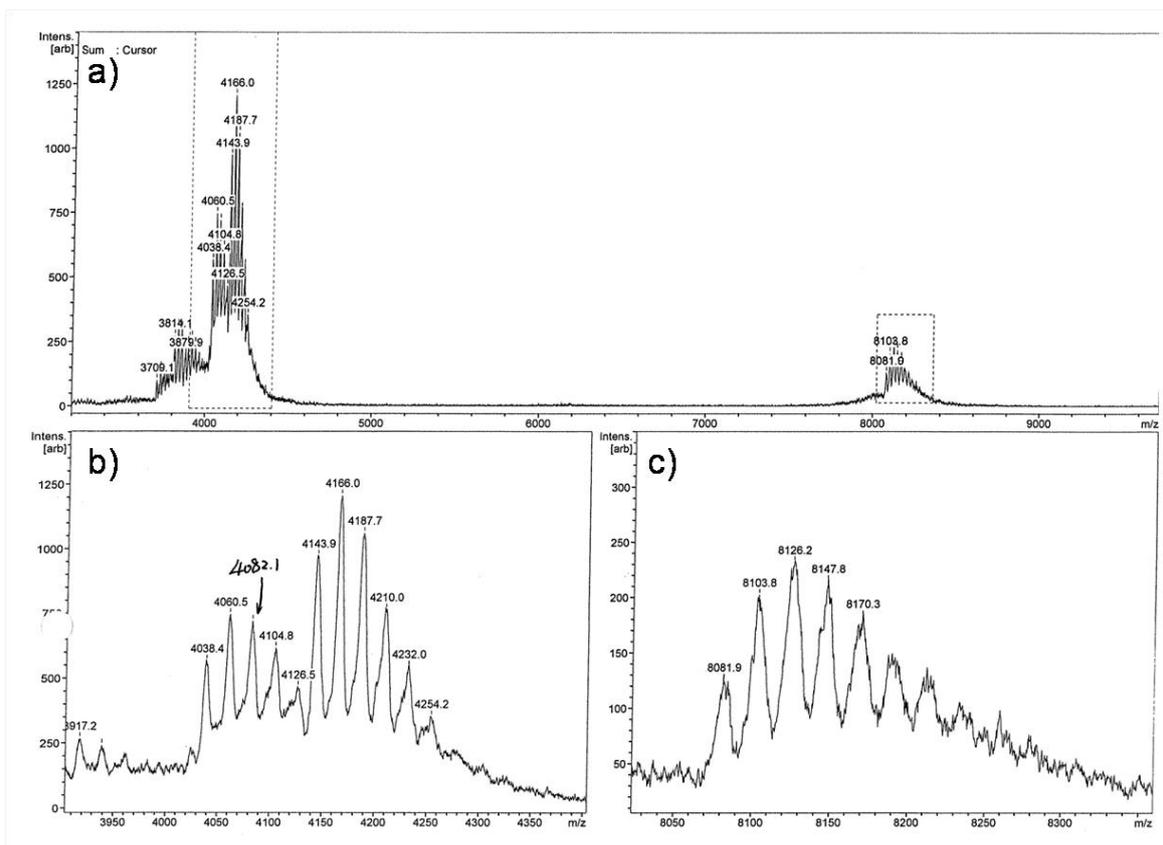


Figure S3. Model representation of a duplex showing the position of the hydroxylamine linker relative to the fdC on the complementary strand.



No.	5'-----3'	Calc.	Exptl.
P1	CCU*AGCGCATTAC	4061.8	4060.5
T1	GTAATGfCGCTAGG	4040.9	4038.4
T1	GTAATGfCGCTAGG	4146.0 ^[a]	4143.9
T1:P1		8084.7	8081.9

[a] conjugate with 4-methoxyaniline

Figure S4. MALDI-TOF mass spectrum of the crosslinked duplex **T1:P1** and single strands. a) Overall MALDI-TOF spectrum; b) peaks corresponding to single strands **P1** and **P1**- 4-methoxyaniline conjugate; c) peaks corresponding to linked duplex **T1:P1**. Conditions: 10 μ M oligonucleotides, 100 mM NaCl, 10 mM NaOAc buffer pH 6.0.

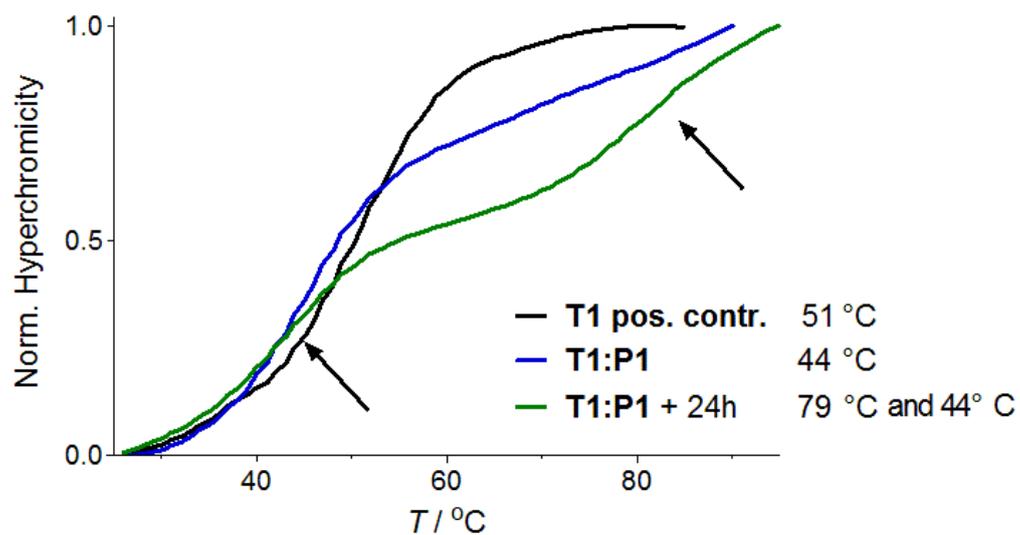


Figure S5. Melting curves of duplex **T1:P1** after reannealing or after 24 h incubation without catalyst compared with duplex **T1** and its counter strand (positive control). Conditions: 1 μ M oligonucleotides, 100 mM NaCl, 10 mM NaOAc buffer pH 6.0, the final volume of 200 μ L.

6. Experimental details of the genomic fdC profiling study

Cell culture and genomic DNA isolation

J1 wild type stem cells (strain 129/SvJae),^[1] Dnmt TKO (J1, strain 129/SvJae),^[2] Tdg^{+/-} (E14, strain 129/Ola) and the Tdg^{-/-} cell line (E14, strain 129/Ola),^[3] were routinely maintained on gelatinized plates in DMEM (Sigma-Aldrich) supplemented with 10% FBS (PAN Biotech), 1x MEM-essential amino acids (NEAA), 0.2 mM L-alanyl-L-glutamine, 1x penicillin-streptomycin, 0.1 mM β -mercaptoethanol (all from Sigma-Aldrich), 1000 U/ml mouse recombinant LIF (ORF Genetics), 1 μ M PD 0325901 and 3 μ M CHIR 99021 (2i; both from Axon Medchem). In these conditions, global genomic mC levels are very low and its oxidized derivatives are even lower, as we described previously.^[4] For the experiments, the cultures were passaged twice (over five days), in DMEM supplemented with FBS and LIF, but lacking 2i. With this strategy, primed mESC cultures were obtained and oxidized cytosine derivatives reached reproducibly higher and stable levels.^[4] In case of the experiment using J1 wild type and Dnmt TKO cells, the cultures were passaged every second day over a period of six days.

Mouse embryonic stem cells were lysed directly in the plates with RLT-buffer (Qiagen). The lysates were homogenized with a TissueLyser MM400 (Retsch) for 1 min at 30 Hz and centrifuged for 5 min at 21000 xg. Then genomic DNA was isolated using the Zymo Quick gDNA Midi Kit according to the manufacturer's instruction. The concentration was measured using a Nanodrop ND-1000 (Peqlab).

Probe crosslinking

The gDNA solution obtained above (1.2 μ g), the fdC probe (**P**) (1 μ M, 2 μ L), NaH₂PO₄-Na₂HPO₄ buffer (200 mM, pH = 6.0, 2 μ L), NaCl aq. (1.5 M, 2 μ L), and ddH₂O were mixed to a final volume of 18 μ L. The mixture was heated to 95 °C for 3 min, and then cooled down rapidly to 25 °C. 1,4-Benzenediamine aq. (10 mM, 2 μ L) was added and the reaction vial shaken (300 rpm) for 6 h at 25 °C. First, the mixture was neutralized with Na₂HPO₄ aq. (200 mM, 40 μ L), and then purification with *NEB Monarch* PCR DNA Cleanup Kit using the binding buffer (120 μ L) and eluting with the elution buffer (Tris-EDTA) (30 μ L). The eluted solution was quantified with the Nanodrop and 22-32 ng/ μ L was obtained. UV spectra confirmed the main peak centered at 260 nm.

Ligation

The crosslinked gDNA solution (300 ng), reporter strand (**R**) (20 nM, 1 μ L), Ampligase reaction buffer (10 \times , 2 μ L), Ampligase from *Epicenter* (5 U/ μ L, 2 μ L, 10 U) and ddH₂O were mixed to a final volume of 20 μ L. The mixture was heated to 95 °C for 3 min, and then to 94 °C for 1 min, 60 °C for 1 h and back to 94 °C for 10 cycles. Then, the reaction mixture was diluted with Tris-HCl buffer (200 mM, pH = 7.6, 50 μ L) before purification with *NEB Monarch* PCR DNA Cleanup Kit using the binding buffer (140 μ L) and eluting with the elution buffer (10 μ L). The eluted solution was quantified with the Nanodrop, obtaining 20-32 ng/ μ L. UV spectra confirmed the main peak centered at 260 nm.

Droplet digital PCR

ddPCR experiments were performed on a *Bio-Rad* QX100 ddPCR System. For one reaction, gDNA (6 ng), four primers (18 μ M each, 1 μ L), two TaqMan probes (5 μ M each, 1 μ L), digital PCR Supermix for Probes (no dUTP, 2 \times , 10 μ L), and ddH₂O were mixed to a final volume of 20 μ L with primer working concentration of 900 nM and TaqMan probe working concentration of 250 nM.

PCR cycles were conducted on a *Bio-Rad* T100 Thermal cycler. PCR cycle: 95°C for 10 min, 94°C for 30 sec and specific annealing temperature (64°C) for 1 min for 35 or 40 cycles, then 98°C for 10 min and cooled down to 12°C. A temperature ramp of 2°C/s was used. Droplet generation and counting were conducted according to the manufacturer's instructions, i.e. reaction mixture prepared as above (20 μ L) and ddPCR droplet generate oil (70 μ L) were used per reaction. The accounted droplet number was retained in 10000-18000. FAM for detection amplicon was set to channel 1; HEX for reference amplicon was set to channel 2.

Each percentage value represents the averages and standard deviations from the mean of at least two technical replicates and two biological replicates. LC-MS quantification were conducted according to the previous report.^[5]

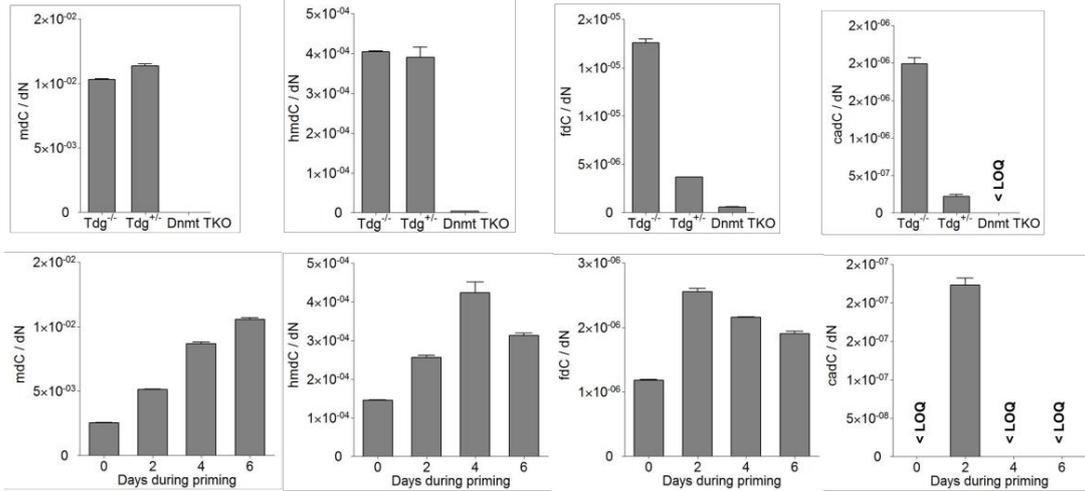
References

- [1] E. Li, T. H. Bestor, R. Jaenisch, *Cell* **1992**, *69*, 915-926.
- [2] A. Tsumura, T. Hayakawa, Y. Kumaki, S. Takebayashi, M. Sakaue, C. Matsuoka, K. Shimotohno, F. Ishikawa, E. Li, H. R. Ueda, J. Nakayama, M. Okano, *Genes Cells* **2006**, *11*, 805-814.
- [3] 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.

- [4] T. Pfaffeneder, F. Spada, M. Wagner, C. Brandmayr, S. K. Laube, D. Eisen, M. Truss, J. Steinbacher, B. Hackner, O. Kotljarova, D. Schuermann, S. Michalakis, O. Kosmatchev, S. Schiesser, B. Steigenberger, N. Raddaoui, G. Kashiwazaki, U. Muller, C. G. Spruijt, M. Vermeulen, H. Leonhardt, P. Schar, M. Muller, T. Carell, *Nat Chem Biol* **2014**, *10*, 574-581.
- [5] M. Wagner, J. Steinbacher, T. F. J. Kraus, S. Michalakis, B. Hackner, T. Pfaffeneder, A. Perera, M. Müller, A. Giese, H. A. Kretzschmar, T. Carell, *Angew. Chem., Int. Ed.* **2015**, *54*, 12511-12514.



Figure S6. Sequence detail of the detection strategy: a) Illustration of the detection strategy with sequence details for locus 1, i.e. 30,020,539th position on chromosome 16 (MM9); b) synthesized and purchased oligonucleotides for locus 1 and 2. Bold and red letters represent nucleoside modifications or functional group: p, phosphate group at 5' terminus; F, FAM; H, HEX; Q, BHQ-1. MALDI-TOF: **P2**: calc.7951.5, found 7948.5, **P3** calc. 7672.5, found 7671.0, contain one sodium ion.



	mdC		hmdC		fdC		cadC	
	pro dN	STABW%						
Tdg ^{-/-}	1.0E-02	1.1	4.0E-04	1.1	1.8E-05	3.6	2.0E-06	7.3
Tdg ^{+/-}	1.1E-02	2.0	3.9E-04	11.3	3.7E-06	0.8	2.2E-07	22.3
Dnmt TKO	3.0E-06	28.0	3.8E-06	18.8	6.0E-07	6.8	*	*
WT0	2.5E-03	1.7	1.5E-04	0.6	1.2E-06	1.9	*	*
WT2	5.1E-03	1.6	2.6E-04	3.6	2.6E-06	3.3	2.2E-07	6.8
WT4	8.7E-03	2.4	4.2E-04	11.2	2.2E-06	0.8	*	*
WT6	1.1E-02	1.7	3.1E-04	3.1	1.9E-06	3.4	*	*

Figure S7. Global 5mC, 5hmC, fdC, and 5caC quantification using LC-MS: Tdg^{-/-}, Tdg^{+/-}, and Dnmt TKO mES cells and 0, 2, 4, 6 days during priming of wild-type mES cells. *: <LOQ, below the limit of quantification.

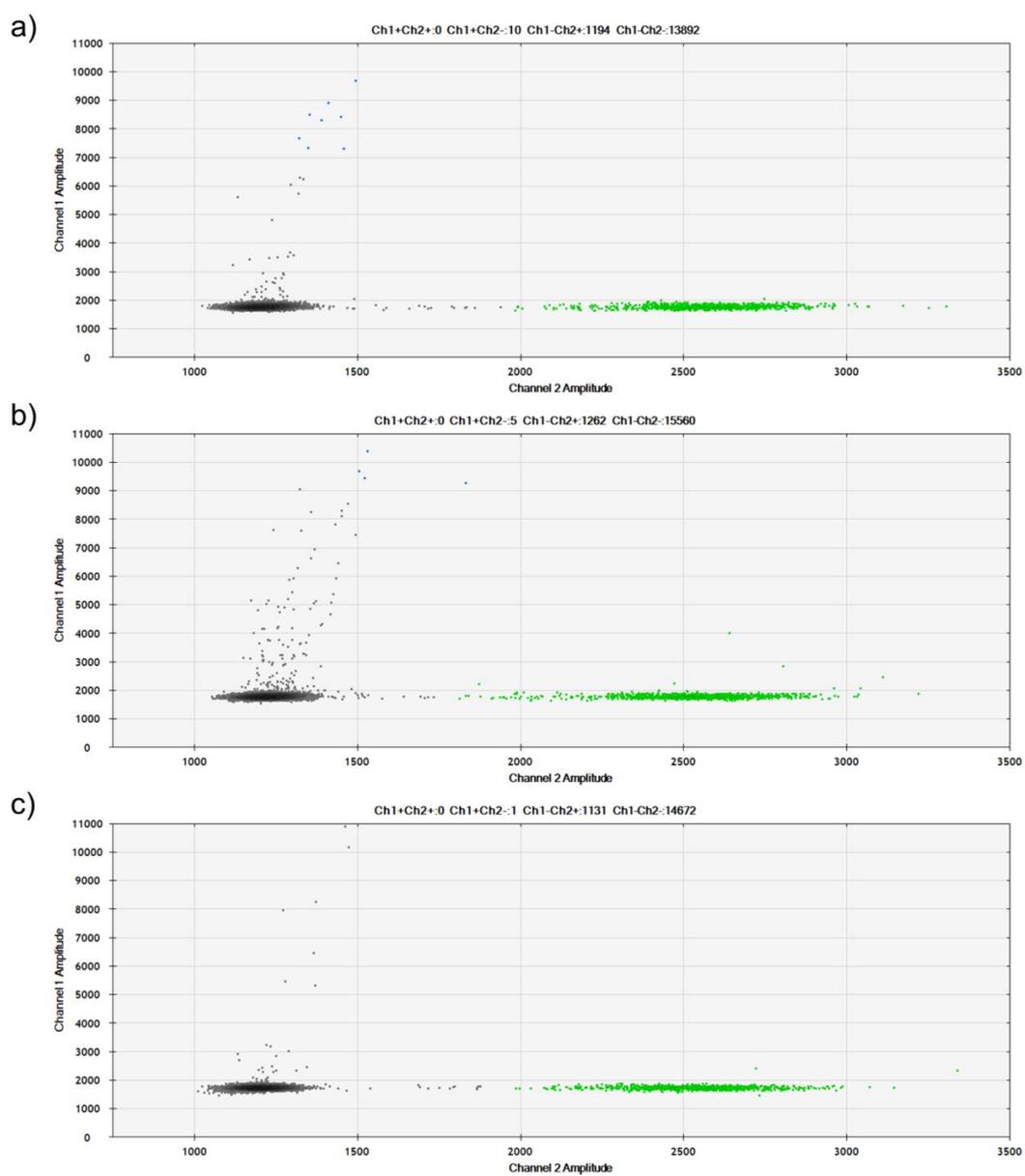


Figure S8. 2-D plot of droplet fluorescence for negative control of locus 1: a) **P3**, instead of **P2**, was used; b) no reporter stand **R1**; c) no Ampligase.

7. Quantification modeling

The encapsulation maximum of one target amplicon in one droplet to generate a positive or negative signal is the ideal scenario for our situation. If a droplet contains more than one detection amplicon, for example, one contains one fdC and one cytosine, it will show a positive signal, and the negative cytosine signal vanishes.

The probability for two or more detection amplicons to get into one droplet can be calculated according to the Poisson distribution, i.e. a discrete random variable X complies the Poisson distribution with parameter $\lambda > 0$, if, for $k = 0, 1, 2, \dots$, the probability mass function of X is given by:

$$f(k; \lambda) = \Pr(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

where e is Euler's number and $k!$ is the factorial of k .

Table S2 Poisson distribution probabilities of genome copies in the droplet.

input ng	λ	k				
		1	2	3	4	5
3	0.05	4.8%	0.1%	0.0%	0.0%	0.0%
6	0.10	9.0%	0.5%	0.0%	0.0%	0.0%
9	0.15	12.9%	1.0%	0.0%	0.0%	0.0%
10	0.17	14.1%	1.2%	0.1%	0.0%	0.0%
15	0.25	19.5%	2.4%	0.2%	0.0%	0.0%
20	0.33	23.9%	4.0%	0.4%	0.0%	0.0%
30	0.50	30.3%	7.6%	1.3%	0.2%	0.0%
40	0.67	34.2%	11.4%	2.5%	0.4%	0.1%

For example, the mass of a mouse genome is approximately 3.0 pg (3.0×10^{-12} g). If 30 ng for a 20 μ L reaction is used, 10,000 genomes will be distributed into 20,000 droplets. So, λ equals to $10,000 / 20,000 = 0.50$. Let $X = 1$, then $f(1; 0.50) = 0.303$; let $X = 2$, then $f(2; 0.50) = 0.076$. This means 30.3% of the droplets, instead 50% of the droplet, contain a single copy while 7.6% of the droplets contain two copies. Extensive distribution probabilities are listed in Table S2. If less than 9 ng is settled in a 20 μ L reaction, the probability to have two copies inside one droplet will be lower than 1%. For the ease of calculation, 6 ng gDNA is used for each reaction of 20 μ L, corresponding to ca. 90 copy/ μ L.

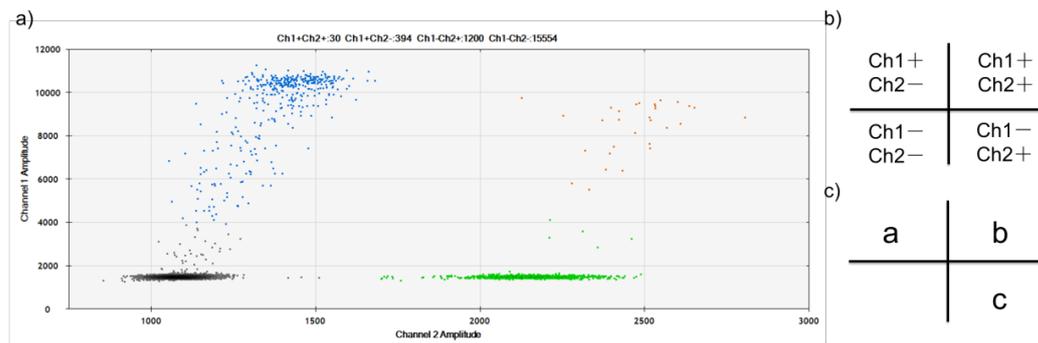


Figure S9. ddPCR output and modeling: a) 2-D plot of droplet fluorescence for illustration; b) clusters separation in four quadrants; c) algebraic simplification of counting numbers of the clusters.

As shown in Figure S8, Ch1+Ch2+ (yellow) refers to droplets with both positive signals; Ch1+Ch2- (blue) refers to droplets with only detection (report strand) signal; Ch1-Ch2+ (green) refers to droplets with only reference (gDNA) signal; Ch1-Ch2- (black) refers to droplets without target locus and ligated product; AD refers to all the droplets accepted; resolution refers to the separation of the clusters.

In principle, Ch1+Ch2+ shows the droplets that contain fdC sites in the target locus; Ch1+Ch2- indicates false-positive signals due to unspecific amplification and the dissociated ligated products; Ch1-Ch2+ shows the droplets containing only the target gDNA.

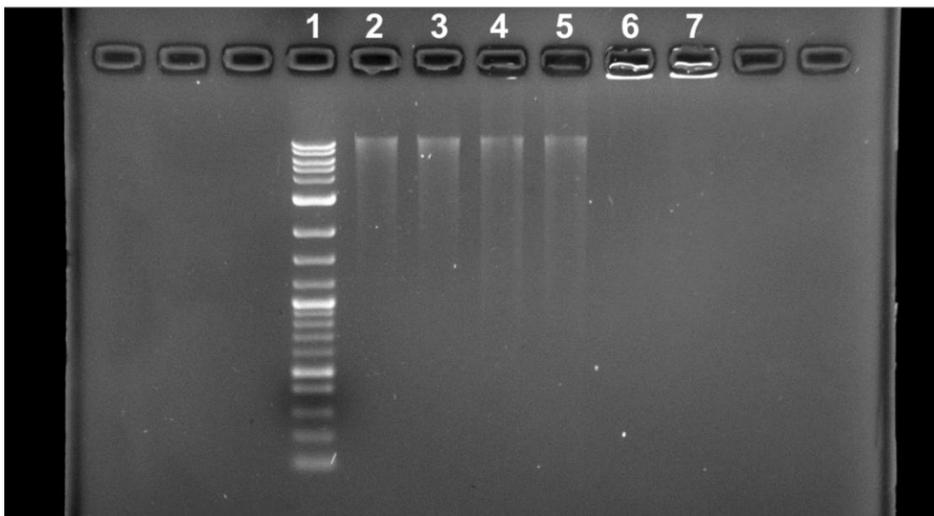


Figure S10. Agarose gel showing gDNA degradation: Line 1, log 2 marker; line 2,3, gDNA (150 ng) after crosslinking; line 4,5, gDNA (150 ng) after ligation cycle, 95°C for 3min, then 10 cycles of 94°C for 1min and 60°C for 1h; line 6,7, gDNA (150 ng).

Without considering the dissociation of the ligated products, the unreacted probe which remained in the system will cause unspecific amplification, i.e. Ch1+Ch2- signals. Catalyst, acid buffer, and ligation cycles will cause gDNA degradation (Figure S9, giving more Ch1+Ch2- false-negative signals. However, Ch1+Ch2- and Ch1+Ch2- / Ch1-Ch2- resolution do not play a role in the mathematical modelling that we used.

Assuming that all fdC at the target site is converted to the reporter strand via crosslinking and ligation, the yield is 100%. Assume that there are less than 150 copies in 1 μ L so that the Poisson distribution is exclusive in our model.

Let $a = \text{Ch1+Ch2-}$, $b = \text{Ch1+Ch2+}$, $c = \text{Ch1-Ch2+}$, (Figure S8) $A = \text{Accepted droplet for the experiment entry}$, respectively, a' , b' , c' , and A' for the control, i.e. TET knockout cell line.

Let $\eta = \text{fdC content of the target site}$.

Then,

$$\eta = \frac{a + b - \frac{A}{A'}(a' + b')}{b + c + (a - \frac{A}{A'}a')/\eta}$$

where $a - \frac{A}{A'}a'$ refers to the degraded gDNA copy containing fdC at the target site

that does not show in Ch2, $(a - \frac{A}{A'}a')/\eta$ refers to all the degraded gDNA copies.

So,

$$\eta = \frac{(b - \frac{A}{A'}b')}{b + c}$$

Herein, in this ideal model, without considering the dissociation of the ligated products, η is independent of a , a' , and c' , i.e. genome degradation and unspecific ligation do not affect fdC percentage. Ch1+Ch2- only be resulted from the dissociation of the ligated products. Also, as shown in Figure S6 a-c, b' can be omitted. Simplified η'

$$\eta' = \frac{a + b}{b + c}$$

is calculated to indicate relative abundance of fdC at the target site.

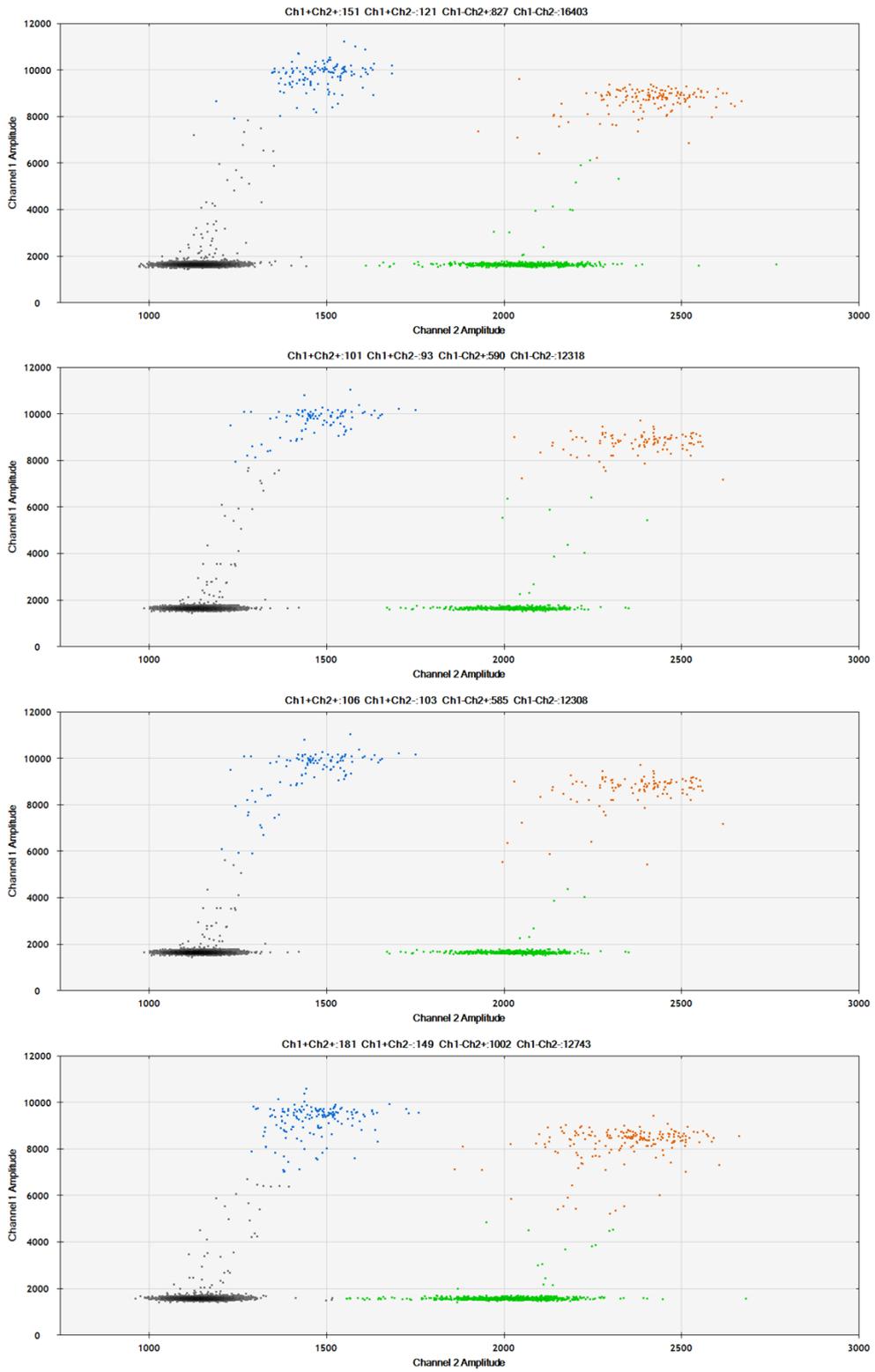
In reality, neither the fdC probe covers all the target sites nor the reporter strand ligates to all the target-linked probe. Therefore, only relative quantification is possible in our model.

8. Droplet digital PCR data

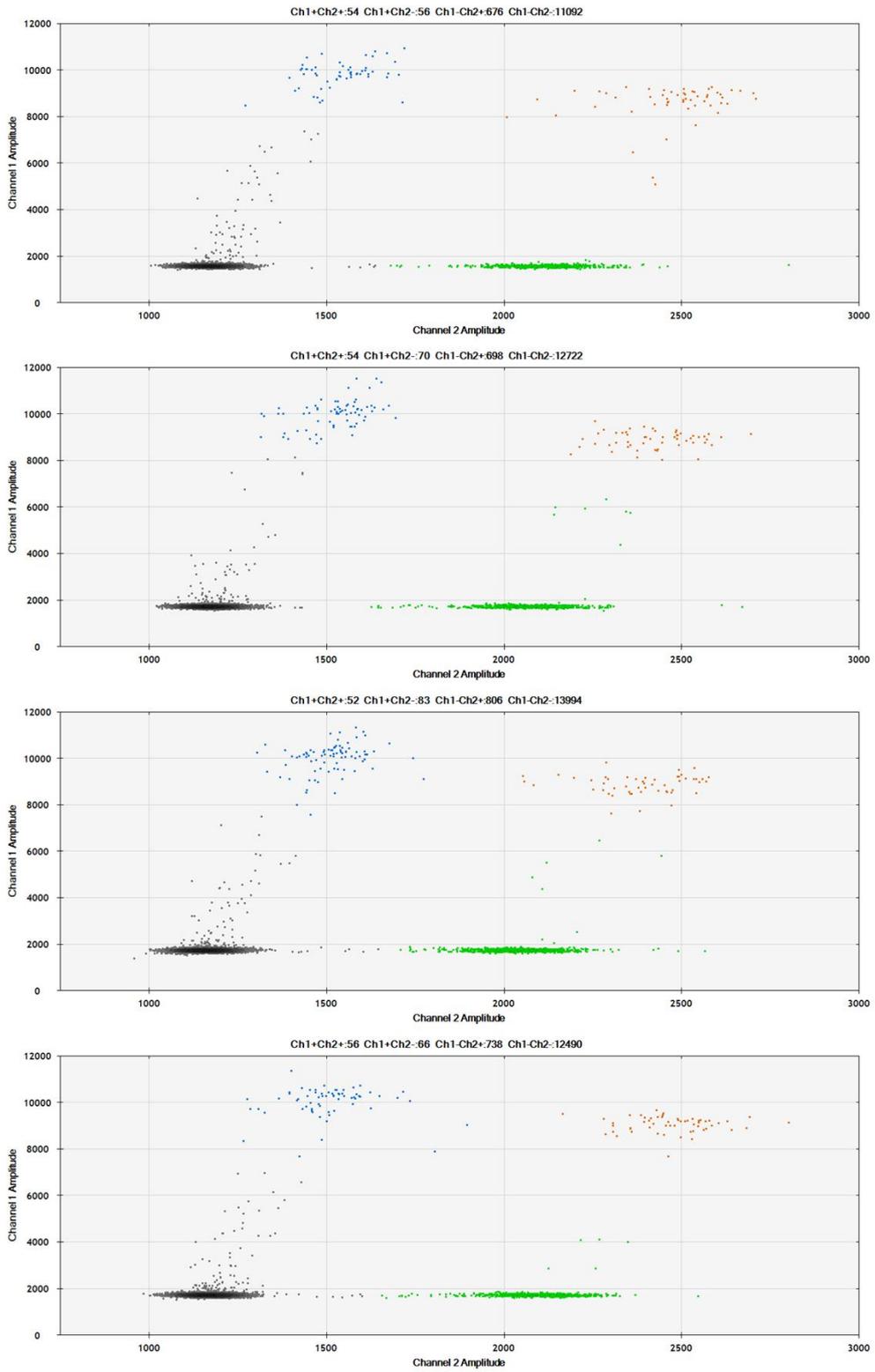
Raw data of fdC detection in Tdg^{-/-}, Tdg^{+/-}, Dnmt TKO cells for locus 1. (AD: accepted droplets)

	Ch1	Ch2	1+2+	1+2-	1-2+	1-2-	AD	η	Average
Tdg ^{-/-}	17.6	67.6	151	121	827	16399	17502	27.8%	28.5%
Tdg ^{-/-}	17.6	63.7	101	93	590	12318	14896	28.1%	
Tdg ^{-/-}	18.9	63.7	106	103	585	12308	13102	30.2%	
Tdg ^{-/-}	28.8	103	181	149	1002	12743	14075	27.9%	
Tdg ^{+/-}	11.6	74.0	54	56	676	11085	11878	15.1%	15.7%
Tdg ^{+/-}	11.3	67.2	54	70	698	12722	13544	16.5%	
Tdg ^{+/-}	10.7	69.6	52	83	806	13994	14935	15.7%	
Tdg ^{+/-}	10.8	72.0	56	66	738	12490	13350	15.4%	
Dnmt TKO	4.0	115.0	10	29	1049	10327	11415	3.7%	5.2%
Dnmt TKO	3.1	61.1	11	17	625	11913	12566	4.4%	
Dnmt TKO	4.1	58.0	14	30	620	12508	13174	6.9%	
Dnmt TKO	2.8	60.6	8	21	600	11484	12119	4.8%	
Dnmt TKO	5.7	96.0	35	28	986	11998	13047	6.2%	

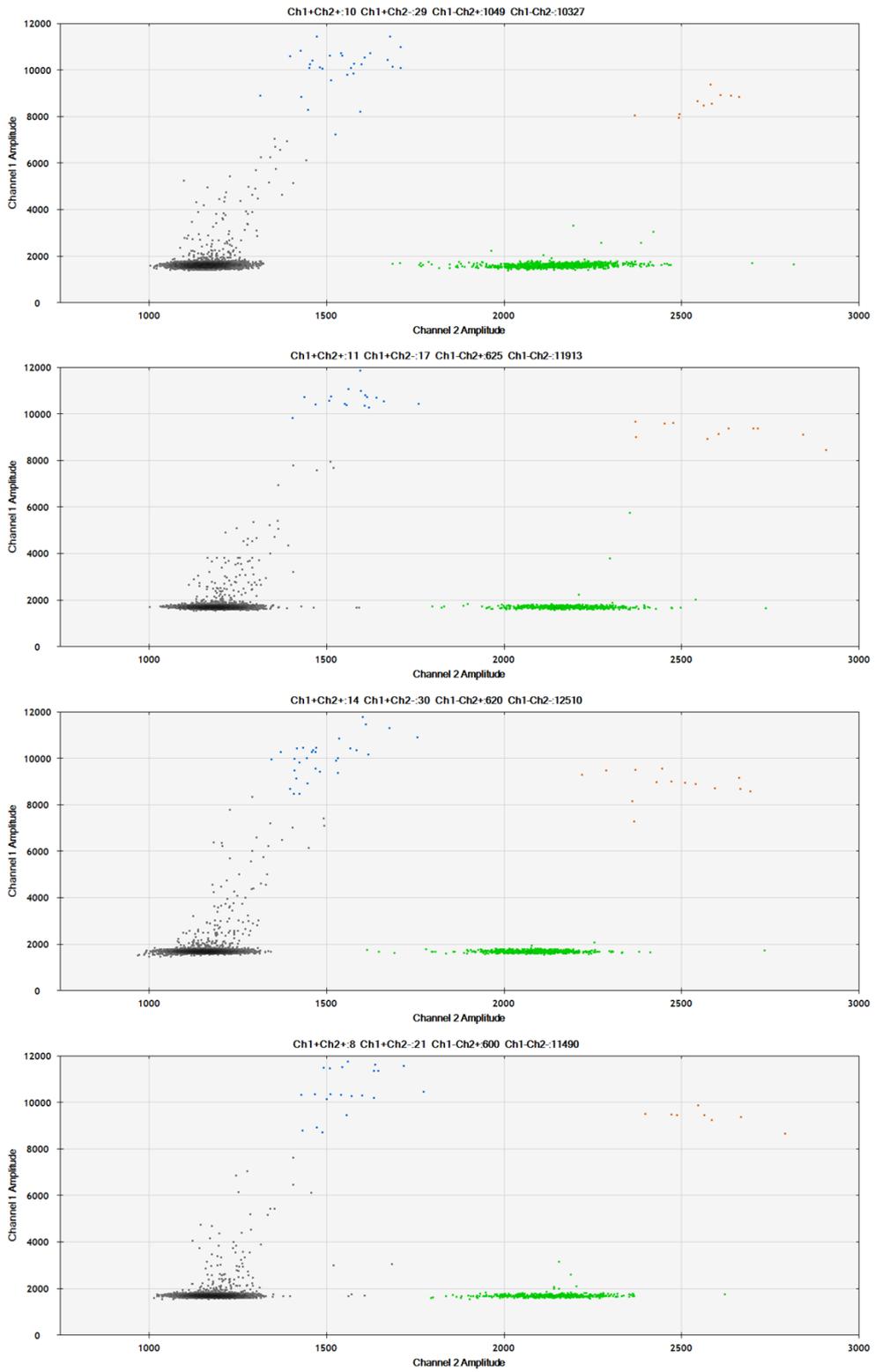
Locus 1 Tdg^{-/-} mES cell sample



Locus 1 Tdg^{+/-} mES cell sample.



Locus 1 Dnmt TKO mES cell sample.



Locus 1 Raw data of fdC detection in wild-type cells during priming.

	Ch1	Ch2	1+2+	1+2-	1-2+	1-2-	AD	η	Average
WT0	5.4	108	20	34	1017	10787	11858	5.21%	6.06%
WT0	6.8	110	30	58	1336	13841	15265	6.44%	
WT0	6.6	112	26	53	1255	12738	14072	6.17%	
WT0	7.8	108	23	58	1241	13060	14382	6.41%	
WT2	9.7	89.2	40	98	1193	15552	16883	11.19%	10.90%
WT2	10.2	88.7	35	105	1249	15948	17337	10.90%	
WT2	10.8	93.0	35	96	1168	14476	15771	10.89%	
WT2	10.2	96.0	54	94	1279	15637	17064	11.10%	
WT2	6.3	56.2	19	65	744	15582	16410	11.01%	
WT2	6.0	57.3	27	54	759	15780	16620	10.31%	
WT4	6.2	68.0	19	30	501	8702	9252	9.42%	8.76%
WT4	5.4	68.4	26	46	811	13942	14821	8.60%	
WT4	4.5	59.0	14	24	477	9449	9964	7.74%	
WT4	5.8	64.5	23	43	688	12579	13333	9.28%	
WT6	11.1	117	42	67	1081	10691	11881	9.71%	8.67%
WT6	8.2	114	28	41	888	9000	9957	7.53%	
WT6	9.8	120	32	41	820	7909	8802	8.57%	
WT6	8.5	114	43	57	1086	11010	12196	8.86%	

Locus 2 Raw data of fdC detection in Tdg^{-/-}, Tdg^{+/-} cells

Well		Ch1	Ch2	1+2+	1+2-	1-2+	1-2-	AD	η	Average
H04	Tdg ^{-/-}	10.5	52	55	92	650	15671	16468	20.9%	19.8%
E06	Tdg ^{-/-}	20.8	115	88	133	1084	11285	12590	18.9%	
F06	Tdg ^{-/-}	22.5	114	98	173	1226	12805	14302	20.5%	
G06	Tdg ^{-/-}	20.1	109	94	149	1178	12927	14348	19.1%	
H06	Tdg ^{-/-}	21.8	116	90	159	1179	12141	13569	19.6%	
B09	Tdg ^{+/-}	5.5	66	28	49	884	15696	16657	8.4%	9.2%
A10	Tdg ^{+/-}	5.9	59	13	29	400	8002	8444	10.2%	
E01	Tdg ^{+/-}	8.3	96	43	78	1299	15749	17169	9.0%	
G01	Tdg ^{+/-}	8.1	93	39	73	1205	15070	16387	9.0%	