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

Epigenetic Studies on the DNA and Protein Level

Mohammad Edris Parsa
aus
München

2017

Erklärung

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

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München den 19.02.2018

(Mohammad Edris Parsa)

Dissertation eingereicht am: 16.11.2017
Erstgutachter: Prof. Dr. Thomas Carell
Zweitgutachter: PD Dr. Stylianos Michalakis
Tag der mündlichen Prüfung: 15.12.2017

Für meine Familie

We think too much and feel too little.

Charlie Chaplin

Danksagung

Nicht nur diese Arbeit, sondern mein gesamter Werdegang wäre ohne die Unterstützung von den verschiedensten Menschen, von Kollegen, Freunden und meiner Familie nicht vorstellbar. Mein Beitrag hierzu ist nur als einer von vielen zu verstehen und ich möchte in den nächsten Zeilen einigen Menschen danken, die zum Gelingen dieses Prozesses beigetragen haben.

Meinem Doktorvater *Professor Dr. Thomas Carell* danke ich für die spannende und herausfordernde Themenstellung. Ich danke ihm für sein Vertrauen in mich und meine Arbeit, für die Übertragung von Verantwortung und Freiheit. Insbesondere möchte ich ihm dafür danken, dass er mich in schwierigen Zeiten unterstützt hat.

PD Dr. Stylianos Michalakis danke ich für die Übernahme des Zweitgutachtens. Ich danke ihm ebenfalls für seine Anregungen und sein Feedback in diversen Meetings. Bei den weiteren Mitgliedern der Prüfungskommission, *Prof. Dr. Anja Hoffmann-Röder*, *PD Dr. Dietmar E. Martin*, *Prof. Dr. Konstantin Karaghiosoff* und *Prof. Dr. Lena Daumann*, bedanke ich mich für die Mitwirkung an der mündlichen Prüfung.

Dr. Markus Müller danke ich für viele anregende Diskussionen und die stetige Hilfe in vielen Belangen. Frau *Slava Gärtner* möchte ich für die Abwicklung aller bürokratischen Aufgaben danken.

Bei *Kerstin Kurz* und Frau *Sabine Voss* bedanke ich mich für ihren Einsatz und für die massenspektrometrische Unterstützung. Insbesondere bei *Kristof Hufnagel* möchte ich mich in aller Form bedanken. Seine Motivation und die damit einhergende Arbeit haben viele meiner und unserer Studien erst möglich gemacht. Sein Hilfsbereitschaft und sein Fleiß waren mir eine große Hilfe. Danke! Ich wünsche dir alles Gute und eine spannende, erfolgreiche und erfüllende Zukunft.

Ich möchte mich auch bei allen Mitarbeitern der Fakultät bedanken, deren oft nicht offensichtliches Mitwirken, die Forschung erst ermöglicht. *Heidi Buchholz*, *Claudia Brackelmann* und *Michael Gayer* seien hier stellvertretend genannt.

Meinen ehemaligen Praktikanten, *Cornelia Unger*, *Felix Hagelskamp*, *Felix Metzner*, *Si-*

mon Veth, *Till Reinhardt*,*Lisa Haddick*, *Matthias Heiß*, *Ilya Jourjine*, *Christoph Habiger*, *Fabio Raith*, *Roksana Chorazewicz* danke ich für ihre große Einsatzbereitschaft während ihres Praktikums. Ich war sicherlich nicht der einfachste Betreuer und ich hoffe ihr konntet mir meine Ungeduld nachsehen. Insbesondere *Felix Hagelskamp*, *Felix Metzner*, *Lisa Haddick*, *Matthias Heiß* und *Simon Veth* haben mich in besonderen Maße unterstütz. Es war mir eine Freude Teil eurer Ausbildung gewesen zu sein und ich bin davon überzeugt, dass ihr einen spannenden Weg bestreiten werdet.

Meinem Labor F4.001b mit seinen aktuellen und ehemaligen Mitgliedern *Korbinian Brunner*, *Dr. Felix Gnerlich*, *Dr. Johannes Harder*, *Dr. Arne Schröder*, *René Rahimoff* und *Martin Rossa* danke ich für die freundschaftliche Atmosphäre und für eine unvergessliche Zeit. Die vielen Abende im Labor, die Musik, unsere gemeinsamen Unternehmungen und Gespräche werden mir fehlen. Ich habe in euch nicht nur Kollegen, sondern gute Freunde gefunden und hoffe auch in Zukunft in Kontakt zu bleiben. Bei einigen anderen ehemaligen Mitgliedern des Arbeitskreises möchte ich mich ebenfalls namentlich bedanken. Insbesondere bei *Dr. Iacovos N. Michaelides*, *Dr. Andrea Künzel*, *Dr. Simon Geiger*, *Dr. Jessica Steinbacher*, *Dr. Benjamin Hackner* und *Dr. David Eisen*.

Ich bedanke mich bei allen Mitgliedern des Arbeitskreises Carell von denen ich mit den meisten auf die eine oder andere Art kooperiert habe. Für das Korrekturlesen dieser Arbeit bedanke ich mich bei *Dr. Markus Müller*, *Dr. Fabio Spada*, *Franziska Traube* und *Angie Kirchner*.

Ich möchte mich bei meinen Mitbewohnern bedanken, insbesondere bei *Christoph Körner* und *Philipp Grauke*. Seit nunmehr über 20 Jahren sind wir Freunde und ich freue mich auf die nächsten 20. Auch *Marlen Cieplik*, *Lorenz Michl*, *Johannes Binnsack*, *Martina Schernhammer*, *Florian Seitzl*, *Christopher Schramm* und *Martina Sorgenfrei* möchte ich für die gemeinsame Zeit danken. Neben meinen Vermietern *Lothar und Elisabeth Hentschel* möchte ich auch allen weiteren Bewohner der Alramstr. 18 danken. Sie waren eine große Bereicherung und ich habe meine Zeit in "der WG" sehr genossen.

Ich habe das Glück im Studium viele bereichernde Freundschaften geschlossen zu haben. Sie alle aufzuzählen würde den Rahmen dieser Danksagung bei weitem sprengen. Besonderer Dank gewährt *Andreas Fetzer*, *Dr. Mario Ellwart*, *David Konrad* und *Dr. Shu-An Liu*. Ich freue mich darauf euch oft wieder zu sehen.

Claudia war in unterschiedlichster Form, zu unterschiedlichsten Zeiten, mal mehr, mal weniger Teil meines Studiums, meiner Promotion und meines Lebens. Unsere Freundschaft war oft kompliziert aber nie beendet. Insbesondere in den letzten Monaten dieser Promotionsphase hast du mich in besonderer Weise unterstützt.

Wibke möchte ich von ganzen Herzen danken. Ihre Unterstützung und ihr Rat haben mich die letzten zwei Jahre begleitet. Sie war sowohl in Zeiten von Erfolgen als auch Rückschlägen für mich da und dafür kann ich ihr nicht genug danken. Deine Selbstlosigkeit, deine Empathie und dein großes Herz werden mir immer ein Vorbild sein. Durch dich habe ich viel neues kennengelernt, bin geduldiger geworden und habe viel über mich selbst gelernt. Und nochmal: "Du hast recht!" Ich möchte hier ebenfalls *Eva* und *Roberto* danken, die mir immer zur Seite standen und eine große Hilfe waren in letzten zwei Jahren.

Meiner Familie kann ich nicht genug danken. Ohne deren seelischen und auch finanziellen Beistand wäre weder mein Studium noch meine Promotion möglich gewesen. Die Opferbereitschaft meiner Eltern, die Großherzigkeit meiner Schwestern und unser Zusammenhalt sind die Grundlage all meiner Errungenschaften, der vergangenen sowie der zukünftigen. Ich habe großes Glück Teil dieser Familie sein zu dürfen und werde immer für euch da sein.

List of Publications

Parts of this PhD thesis were already published, submitted or presented on scientific conferences.

Publications:

- **E. Parsa**, A. S. Schröder, T. Carell, Modifizierte DNA-Basen erweitern das Verständnis der Genregulation - Einsatz moderner Massenspektrometrie als Werkzeug in der Epigenetik, *Naturwss. Rundschau.* **2015**, *68*, 500-505.

- A. S. Schröder, O. Kotljarova, **E. Parsa**, K. Iwan, N. Raddaoui, T. Carell, Synthesis of (R)-Configured 2′-Fluorinated mC, hmC, fC, and caC Phosphoramidites and Oligonucleotides, *Org. Lett.* **2016**, *18*, 4368-4371.

- 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, *Chem. Commun.* **2016**, *52*, 14361-14364.

- 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. Bio.* **2017**, *accepted*.

- N. Kitsera, J. Allgayer, **Edris Parsa**, N. Geier, M. Rossa, T. Carell, A. Khobta, Functional impacts of 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxycytosine at a single hemi-modified CpG dinucleotide in a gene promoter. *Nucleic Acid Res.* **2017**, *accepted*.

Conferences:

- **E. Parsa**, M. Stadlmeier, F. Traube, A. Künzel, T. Carell, Elucidating the interactome and post-translational modification of TET enzymes, *CNRS - Jacques Monod Conference "DNA methylation and demethylation"* **2015**, Roscoff, France.

- **E. Parsa**, A. F. Künzel, F. R. Traube, T. Carell, "Investigation of functional TET enzymes – *in vitro*, *in vivo* and proteomic studies, *EMBO Conference Chemical Biology* **2016**, Heidelberg, Germany.

Contents

List of Figures

Zusammenfassung

Multizelluläre Organismen, wie beispielsweise der Mensch, besitzen eine Vielzahl an unterschiedlichen Zelltypen, mit zum Teil höchst unterschiedlichen Funktionen und Formen. Dennoch haben alle Zellen eines gemeinsam, ihre DNS Sequenz. Offensichtlich ist die reine Sequenzinformation also nicht ausreichend, um Unterschiede in Morphologie und Funktion hinreichend zu erklären. Das Forschungsgebiet der Epigenetik beschäftigt sich mit der Fragestellung, wie dieses Phänomen zustande kommt.

Im Rahmen dieser Doktorarbeit konnte ich unter anderem neuartige Test-Moleküle etablieren, die für zukünftige epigenetische Studien verwendet werden können. In enger Zusammenarbeit mit *Dr. Arne Schröder* konnte ich zeigen, dass 2'-(R)-Fluor modifizierte DNA-Bausteine von der epigenetischen Maschinerie der Zelle erkannt und prozessiert werden. Dies galt sowohl für DNA Methyltransferasen, als auch für die erst kürzlich charakterisierten TET-Enzyme. Die Fluor-modifizierten DNA Bausteine sind darüber hinaus nicht toxisch und konnten so, für weitere Studien, biologischen Systemen zugeführt werden. Hierdurch konnte ich zusammen mit *Katharina Iwan* und anderen, einen von Carell *et al.* postulierten Stoffwechselweg, nämlich die aktive Demethylierung von mC über einen direkten C-C Bindungsbruch, erstmals *in vivo* nachweisen. Hier zeigte sich, dass sowohl 2'-(R)-F-fC, als auch das natürliche Derivat fC direkt umgesetzt werden zu 2'-(R)-F-C bzw. C, ohne dass dabei die klassische DNA Reparaturmaschinerie zum Einsatz kommt.

Darüber hinaus gelang es mir im Rahmen dieser Promotion Einblicke in die Regulation von TET Enzymen zu erhalten. Ich etablierte in Zusammenarbeit mit *Dr. Benjamin Hackner* ein neuartiges Proteinexpressions- und Aufreinigungs-Protokoll, welches verschiedene Studien innerhalb der Forschungsgruppe Carell ermöglichte. Zusammen mit *Franziska Traube*, *Dr. Andrea Künzel* und anderen zeigte ich, dass ein Netzwerk von metabolischen Enzymen die Aktivität von TET3 in Hirngewebe steuert. Darüber hinaus konnte ich zusammen mit *Dr. Andrea Künzel* und *Franziska Traube* Unterschiede im Protein-Interaktions-Kontext von TET1 und TET3 identifizieren, was Rückschlüsse auf

deren unterschiedliche Funktionen innerhalb der Zelle erlaubt. Die neuartige Expressions-Methode erlaubte mir zusammen mit *Michael Stadlmeier* und *Dr. Benjamin Hackner* die post-translationalen Modifikationen von TET Enzymen zu charakterisieren. Ich konnte über Mutations-Studien zeigen, dass einzelne Phosphorylierungen im katalytischen Zentrum die Aktivität von TET Enzymen beeinflussen können.

Im Rahmen einer Kooperation mit der Gruppe von *Dr. Andriy Khobta* untersuchte ich darüber hinaus die Auswirkung epigenetischer Basen auf die Transkription und den Zusammenhang mit der DNA-Reparatur Maschinerie.

Summary

Although all cells present in a multicellular organism share the same DNA sequence, cells can differ in both function and morphology from each other. Obviously, the sequence information alone is not sufficient to explain these differences. Epigenetics tries to understand, how this phenomenon can be explained.

In the course of this thesis, novel molecular tools were established in order to answer some of these tempting questions. In close collaboration with *Dr. Arne Schröder*, I demonstrated that 2'-(R)-Fluorine modified DNA building blocks are compatible with the epigenetic machinery inside the cell. Both DNA methyltransferases as well as the recently characterized TET enzymes are able to recognize and process these non-canonical DNA derivatives. These modified DNA building blocks are non-toxic and could therefore be fed to various biological systems. This enabled further studies that could show that active demethylation of mC can occur through a direct C-C bond cleavage reaction. This pathway was often postulated, but never clearly demonstrated. Together with *Katharina Iwan* and others, I could show that this demethylation, occurs *in vivo* to a significant extent and that both 2'-(R)-F-fC and the canonical derivative fC are converted to 2'-(R)-F-C and C, respectively.

Another study focused on investigating the regulation of TET enzymes, which are responsible for the oxidation of mC. Together with *Dr. Benjamin Hackner*, I established a novel expression and purification protocol enabling various studies in the Carell group. Together with *Franziska Traube*, *Dr. Andrea Künzel* and others, I could demonstrate that a network of metabolic enzymes control the activity of TET3 in brain tissue. Furthermore, different interactomes of TET1 and TET3 were identified. This may give insight into the distinct functional contexts of TET1 and TET3 within an organism. The new expression protocol also enabled the characterization of post-translational modifications of TET enzymes. Together with *Michael Stadelmeier* and *Dr. Benjamin Hackner*, I mapped the phosphorylation sites of these enzymes. I subsequently performed mutation studies and was able to demonstrate that single phosphorylations in the active center

can affect the activity of TET enzymes.

In cooperation with the group of *Dr. Andriy Khobta* I also investigated the functional impact of epigenetic bases upon transcription and the interplay with DNA repair.

Introduction

1 "Cracking the Code" - a historical perspective

"Today, we are learning the language in which God created life."

This quite ardent expression by former US president Bill Clinton was only one of many stated during a press conference held in June 2000 in the East Room of the White House. Together with Tony Blair, Craig Venter, Francis Collins and others, the completion of the first draft of the human genome was announced and celebrated. A wide range of people, including journalists, politicians and scientists believed that this remarkable milestone would mark the end of several diseases and pave the way for a more healthful future and longer lifespans. Unfortunately, these promises could not be kept in their entirety. As will be discussed in the following chapters, the human genome is far more complex than its sole sequence, and as outstanding as this scientific breakthrough was, it neither marked an end of genetic research nor was it the beginning. The history of genetic research is rich in key discoveries and characters, and the latter are often not credited enough. It is worth mentioning, for example that the history of genetics or heritage did not start with Charles Darwin's book "On the Origin of Species"[1], nor with Gregor Mendel's "Versuche über Pflanzen-Hybriden"[2] as is widely believed. The general idea of heritage can be traced back to Hippocrates (ca. 460-370 BC).[3] About 410 BC, Hippocrates outlined a theory later called *pangenesis* by Charles Darwin. Hippocrates' theory of the *inheritance of acquired characteristics* (IAC) assumes that heredity relies on the production of specific "seeds" by the parents. Both Hippocrates and Darwin believed that the whole organism participates in the production these seeds, named "*gemmules*" by Darwin. These "seeds" are then transferred to the sperm and the egg and are subsequently transmitted from the parent to the offspring. For a long period of

time this theory was generally accepted, with Jean-Baptiste de Lamarck only being one of its most prominent supporters.[3,4] Although Darwin and Lamarck are generally recognized as opponents, there are many intersections in their studies, and Darwin's work on *pangenesis* complements rather than contradicts his work on natural selection.[5,6]

Although the theory of IAC was widely accepted for about 2000 years, it was almost completely rejected in the 20th century. Darwin noted on this matter:

> "You will think me very self-sufficient, when I declare that I feel *sure* if Pangenesis is now still born it will thank God at some future time reappear, begotten by some other Father, & christened by some other name."[7]

He should be right, as recent research supports the theory of IAC in parts and the "Lamarckian" ideas are revisited.[8–13] There were two major reasons why *pangenesis* was a "stillborn" and rejected. One being due to Darwin's cousin Francis Galton. He actually intended to prove his cousin's theory and aimed to demonstrate that *gemmules* do indeed exist in the body. He was convinced that these *gemmules* circulate in the blood, hence he performed blood transfusion experiments on rabbits, trying to transfer characteristics from one breed to another. Yet his attempts failed and he claimed: "I have (...) arrived at definite results, negativing, in my opinion, beyond all doubt the truth of the doctrine of Pangenesis".[14,15] It should anyway be noted that Darwin himself never argued that *gemmules* must be located in the blood, Darwin actually took into account that *gemmules* can be located even outside of the body.[16] The second major reason why Darwin's theory of *pangenesis* was not well accepted in the community was due to the work of the influential German scientist August Weismann. His work on mice ended the era of *pangenesis* for a long period of time. In his experiments, Weismann cut the tails of the animals for many generations and observed that their offspring continued to develop long tails.[17] Based on his studies Weisman proposed the *germ-plasm theory*, suggesting that the hereditary information is only located in sperm and egg cells and is not altered by environmental influences or affected by other cells. In other words, it is not possible to inherit "acquired characteristics" from somatic cells, which only carry out ordinary functions. Weismann therefore rejected the idea of the *gemmules* and concluded that no information can pass from the soma to the germ plasm (*Weismann barrier*). Other scientist like Hugo de Vries modified Darwin's theory of *pangenesis*. De Vries proposed that *gemmules* do exist, but cannot be transferred between cells; they only travel intracellu-

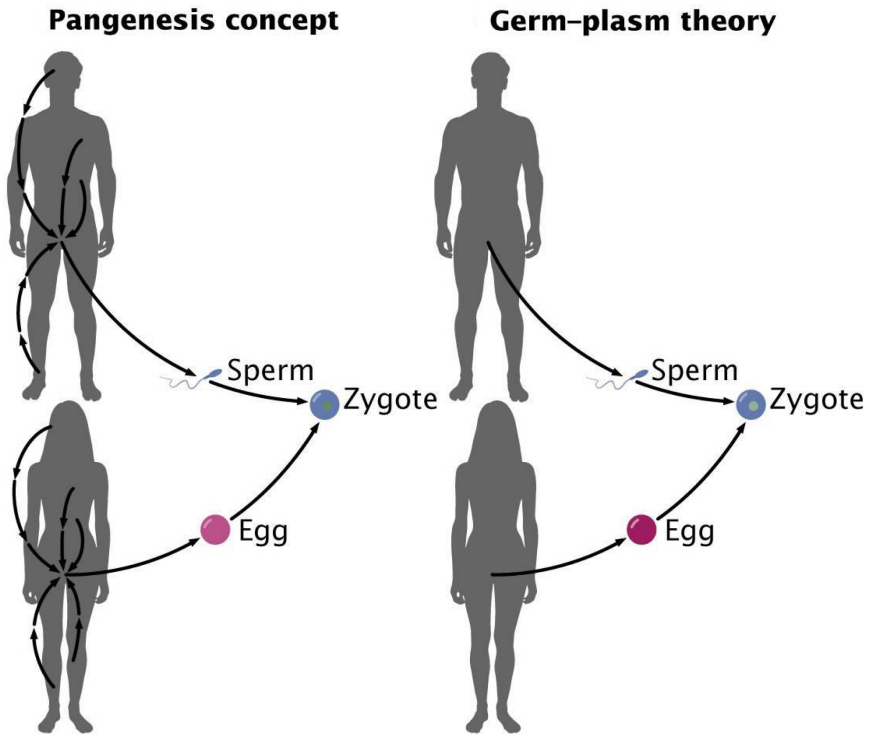


Figure 1: Hippocrates proposed the theory of IAC, later named *pangenesis* by Charles Darwin. According to this theory, *gemmules* are produced by somatic cells throughout the body and transferred to the sperm/egg cells. Weisman and others believed that only germ cells contribute to heredity, and that acquired characteristics are therefore not passed on to the offspring. The figure is adapted from Pierce.[18]

larly. Hugo de Vries was also the first scientist to postulate that different phenotypes have different hereditary carriers/particles. He termed these particles *pangenes*, a term that was later altered to *genes* by Johannsen.[19,20]

These mostly theoretical discussions were accompanied by scientific discoveries that facilitated further research and enabled the scientific community to develop more and more sophisticated theories. Some of the most outstanding scientific breakthroughs were Gregor Mendel's experiments on plant hybridization in 1865, at a time when DNA was not yet discovered.[2] Mendel was the first to develop statistical methods in heredity research. Due to his work being published in a rather obscure scientific journal, it was unfortunately neglected for many years. In the meantime (1869), Friedrich Miescher discovered a completely new chemical entity in cells, which he termed *nuclein*, some-

thing now known as DNA.[21–23]. The function and nature of this novel ingredient of cells, however remained elusive. Later, in 1882, chromosomes were identified and observed during cell division by Walther Flemming.[24] Flemming was unfortunately not familiar with Mendel's work and thus could not connect his findings with the Mendelian laws of heredity. It actually took 30 years until Mendel's work was rediscovered. Three researchers, Carl Correns, Hugo de Vries and Erich von Tschermak, went on to independently reproduce Mendel's results.[23] At the same time, two scientist made discoveries that could physically explain the Mendelian laws of heritage, namely Theodor Boveri and Walter Sutton (1902). Their *Boveri-Sutton chromosome theory* could for the first time be used to interpret Mendel's observations. They could show, among other things, that sperm and egg both contribute the same number of chromosomes, and that individual chromosomes impact development in different ways. They proposed that each "individual chromosome possesses different qualities".[25–27]. This may now sound self-evident, but the theory remained discussed highly controversial, until Thomas Hunt Morgan and his student Alfred Sturtevant proved unequivocally that genes (or "qualities") do reside on chromosomes. They were able to draw the first chromosome map in 1913.[28]. Another milestone in genetic research was *Griffith's Experiment* in 1928, when Frederick Griffith was able to demonstrate that the heredity material can be transferred between two organisms.[29] About 15 years later in 1944, it was clearly proven by Avery, MacLeod and McCarty that this "transforming particle" is not a protein, as was widely believed, but actually DNA.[30] Consequently it took about a century from the discovery of DNA in the year 1869 to ascertain that this molecule did indeed contain the hereditary information (1944). George Beadle and Edward Tatum tried to connect the concepts of "genes" and "proteins" more precisely and proposed that one gene is necessary for one enzyme. This "one gene, one enzyme" theory was later expanded to the *central dogma of molecular biology.*[31–33] This concept is briefly summed up by "DNA makes RNA and RNA makes protein". Yet the discovery of splicing[34] and reverse transcription[35,36] unveiled a more complex flow of genetic information.

Having identified DNA as the heritage material, researchers now aimed to gain a more comprehensive mechanistic understanding of this biopolymer. Erwin Chargaff identified distinct proportions of the four DNA bases that compose DNA. The amount of adenine equals the amount of thymine and guanine equals cytosine, something now known as the *Chargaff rules.*[37] Chargaff's findings could be further explained by yet another scientific breakthrough, solving the molecular structure of DNA. This structure revealed a double

helical architecture of DNA and C:G/A:T base pairing. The structure was elucidated through joint efforts of Maurice Wilkins, Rosalind Franklin, James Watson and Francis Crick and published in several papers 1953.[38–42] In one of these papers Watson and Crick realized that "the specific pairing (...) immediately suggests a possible copying mechanism for the genetic material." This "copying mechanism", the semi-conservative replication mechanism, was demonstrated 1958 by Matthew Meselson and Franklin Stahl and could explain clearly how heredity is achieved on a molecular level.[43] It remained elusive, however, how the nucleotide sequence of each gene corresponded to the amino acid sequence of a protein. In 1961 this genetic code was decrypted by the work of Francis Crick and others.[44,45] The next logical step for scientists was, to sequence the DNA and in 1972 Walter Fiers and co-workers were the first to sequence a protein-coding gene.[46] Later in 1977 Frederick Sanger, Allan Maxam and Walter Gilbert developed further sequencing methods for DNA.[47–50] In the meantime researchers also became interested in manipulating DNA. The discovery and utilization of restriction enzymes or the development of methods like PCR by Mullis further drove research and enabled novel studies and breakthroughs.[51,52] The generation of designed plasmids and their subsequent transformation resulted in the first drug developed and produced based on DNA technology.[53–55]

In the 1990s, sequencing became a more and more dominant tool of molecular genetics and in 1995 the first complete genome of a free living organism was sequenced, *Haemophilus influenzae*.[56] Only a year later, the DNA sequence of *Saccharomyces cerevisiae* was published, representing the first eukaryotic genome.[57] The first draft of the human genome was published 2001 by the Human Genome Project and the private company Celera.[58,59] Only a few years later, the complete (99%) sequence of the human genome was made public.[60] This sequence revealed that less than 5% of the genome encodes for proteins (about 20.000 genes). Consequently, this would mean that the vast majority of the genome is "junk DNA", a term that is actually not suited and leads to misinterpretations.[61] Follow-up programs of the Human Genome Project aimed to understand the molecular function of these non-protein-coding sequences. Namely the ENCODE project (Encyclopedia of DNA Elements) is to be mentioned here, which aimed to identify all functional elements of the human genome.[62] The extensive collaboration of several research groups could show that in fact the vast majority of the human genome have at least one biochemical function.[63–65] Researchers are just beginning to understand and connect the results of the ENCODE project and many further studies

will shed more light on the complex organization of the genome. The next chapter will introduce yet another layer of complexity that further explains why understanding our genome is far more complicated than anticipated in June 2000.

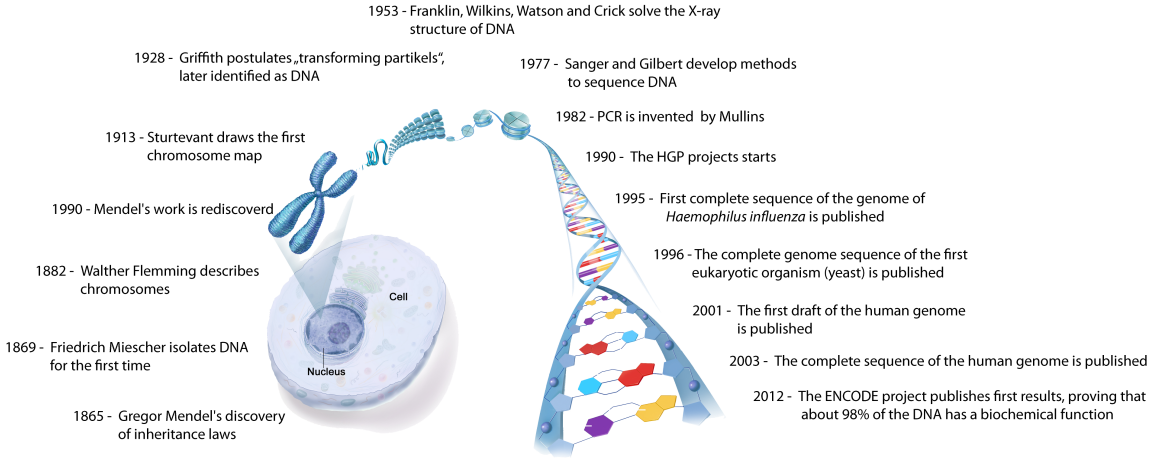


Figure 2: The history of genetics is rich in key discoveries, starting with Mendel's mathematical description of heritage, through the structural elucidation of the DNA, resulting in the genomics era. These are selected discoveries.

2 The Second Layer of Information

Despite the fact that every cell in a multicellular organism shares the same DNA sequence, these organisms in general do not appear as a homogenous mass, but have parts of distinct morphology and function. Various structures are formed by specialized cells and perform different tasks. The human body for example consists of more than 200 different cell types.[66] It is evident that the sole DNA sequence is not sufficient to describe these phenomena and another level of information besides the sequence must be present that controls cell specific patterns. This level is called *epigenetics*. The idea of epigenetics (although not the term) was first introduced by C.H. Waddington.[67] He defined epigenetics as "all those events which lead to the unfolding of the genetic program for development". This definition was later specified by Robin Holliday who described epigenetics as "nuclear inheritance which is not based on differences in DNA sequence", a definition coming close to what nowadays is termed as epigenetics.[68] There is another rather intriguing picture to describe epigenetics. One can image the DNA sequence as a piece of music. The pattern of notes resembles the patterns of A, G, T and C. Just like a piece of music, however, this pattern may be interpreted by various artists in different ways. In the interpretation of the genetic sequence, the modality seems to be based on non-canonical DNA bases and other epigenetic systems.
Three different epigenetic mechanisms are known to date: chromatin remodeling[69–71], the chemical modification of histones[72–77] and the chemical modification of DNA. The latter will be further discussed.

2.1 Chemical Modifications of DNA

For about half a century, 5-methylcytosine (mC) was the only known chemical modification in mammalian DNA and the "fifth" base of the genome, besides the four canonical DNA bases adenine (A), guanine (G), thymine (T) and cytosine (C). In recent years however more and more chemical derivatives of canonical DNA bases were identified in the genome of various organisms. 5-hydroxymethylcytosine (hmC)[78,79], 5-formylcytosine (fC)[80,81] and 5-carboxycytosine (caC)[81,82] were characterized in the DNA from stem cells and brain tissue. Also a derivate of T, namely 5-hydroxymethyluracil (hmU), could be identified and quantified in the DNA of murine embryonic stem cells (mESCs).[83]

This study by Pfaffeneder *et al.* demonstrated that hmU is generated in a regulated fashion and is not necessarily a DNA lesion. The most recent addition in the chemical diversity of eukaryotic DNA is N6-methyladenosine (m^6A), although its existence in mammalian cells is controversially discussed.[84–87]

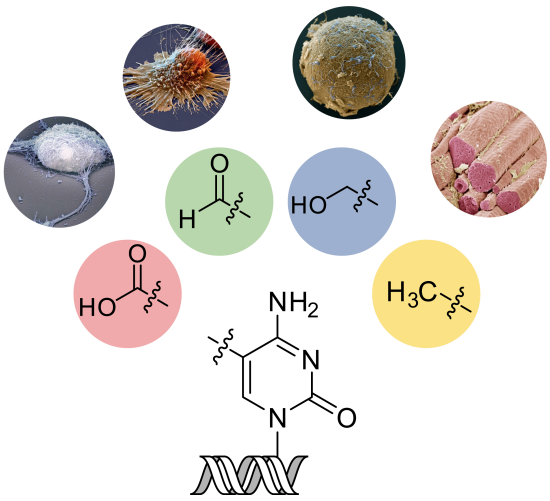


Figure 3: The vast complexity of life and the ability to form diverse cell types is in part based on rather simple chemical modifications of DNA (besides the chemical modification of proteins). Four derivatives of cytosine (mC, hmC, fC and caC) are depicted, which are studied intensely.

However, despite of massive research efforts, the function of these chemically modified DNA bases is not yet fully understood. The chemical derivatives of cytosine (see Figure 3) are most likely involved in embryonic development, memory and diseases like cancer.[88–94] It seems impressive how rather simple chemical modifications are part of a machinery that enable diverse expression patterns, resulting in a diverse cell type population, but can also be responsible for cognitive function and diseases.
If not otherwise stated the notation of canonical and non-canonical bases in the following refers to a DNA context.

2.1.1 5-Methylcytosine

The DNA base 5-methylcytosine (mC) was first described in the year 1925 and positively identified 1950.[95–97] The methyl group is transferred from S-adenosyl methionine (SAM) to the C5 position of cytosine by DNA Methyltransferases (DNMTs).[98,99] In mammals three DNMTs are known, DNMT1, DNMT3a and DNMT3b. Whereas DNMT1 is a "maintenance" methyltransferase, DNMT3a and DNMT3b are needed for the *de novo* methylation of cytosine.[100,101] The mechanism of this process is depicted in Figure 4B. SAM acts as a co-substrate and methyl donor. DNMTs are using a "flipping" mechanism and a thiol as a nucleophile.[102,103] This thiol attacks the C6 position of cytosine. A protonation of N3 via a glutamate residue facilitates this reaction. Next the C4-C5 double bond of cytosine attacks the methyl group of SAM. Subsequent *syn*-elimination re-establishes the aromatic character and results in mC.

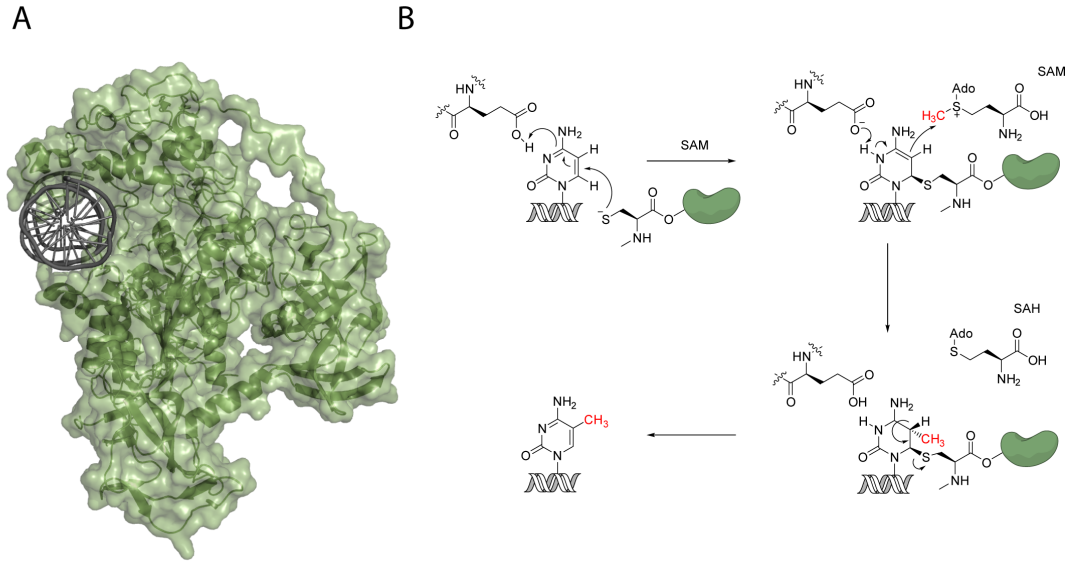


Figure 4: A) Crystal structure of mouse DNMT1 in complex with DNA (PDB code 3PT6). **B** Mechanism of a DNA methyltransferase. First, nucleophilic addition of a thio group from a cysteine residue occurs at position 6 of the flipped out cytosine. A glutamate supplies the proton for N3, facilitating this reaction. The C4-C5 double bond of the base then attacks the methyl group of SAM. A *syn* elimination yields the product mC.

In mammals about 4%-5% of all cytosines are methylated.[83,104] Most of the methylation occurs in a CpG context.[105] This motif occurs frequently in promoter regions, forming so

called CpG islands, while it is generally underrepresented in the genome.[106,107] About 72% of human promoters are rich in CpGs (GpG islands), which are in general hypomethylated.[105,108,109] In the year 1975, Holliday and Riggs demonstrated that mC has an epigenetic function and is responsible for gene silencing.[110,111] In general, mC is necessary for processes like genomic imprinting, or the repression of repetitive elements.[112,113] Depending on its position, however, mC can have distinct effects, e.g. if being present inside the gene body functions related to splicing are discussed.[114–116] Whereas a cytosine methylation at repetitive sequences e.g. centromeres are important for the stability of chromosomes.[117]

The methyl group points towards the major groove of the DNA double helix, facilitating an interaction with proteins and therefore allowing biological effects. These effects are, as mentioned earlier, mainly transcriptional repression, e.g. through the recruitment of repressing transcription factors like the methyl CpG binding protein 2 (MeCP2).[118,119] 5-methylcytosine is also able to recruit histone modifying enzymes[120,121] and *vice versa*, hence controlling the density of DNA and its accessibility for transcription.[122–126] As tissue specific promoters also contain CpG dinucleotides, the relevance of mC for cell differentiation is supported.[127–129] The cells are hereby enabled to influence tissue specific protein expression and control cell fate.[130] The methylation of cytosine is essential for normal embryonic development and stem cell differentiation. DNMT1 knockout mice die only a few days after fertilization.[131] DNMT3a and DNMT3b knockout mice do develop, but die shortly after birth.[132] It should however be mentioned that the DNA of some organisms like *Caenorhabditis elegans* do not contain mC and the presence in *Drosophila* is still a matter of disagreement in the community.[133]

Methylation patterns are dynamic and can change due to various stimuli, hence nature has established mechanisms to remove the methyl group (for details see chapter 3). The paternal genome for example is heavily demethylated a few hours after fertilization and before the first cell division.[134,135] The factor needed for this process is contributed by the oocyte.[136] This demethylation is significantly faster than the demethylation of the maternal genome. After erasure of the methylation patterns, the DNA is re-methylated to establish and "program" cell specific patterns.[137,138] It is possible to revert this program and to re-program somatic cells into stem cells, a discovery awarded with the Nobel prize for physiology 2012. John B. Gurdon and Shinya Yamanaka were honored for their research in re-programming somatic cells into *induced pluripotent stem*

cells (iPSC).[139,140] This process involves several methylation and demethylation events, hence changing the epigenome.[141–143] A comprehensive understanding of this process could accelerate the process of stem cell therapies without the ethical dilemmas of using embryonic stem cells.[144–147] Other modified DNA bases seem to be part of the demethylation processes[88,148–150] and the next chapters will discuss possible functions and the metabolism of these novel DNA modifications.

2.2 Distribution and Functional Context of Oxidized mC Derivatives

2.2.1 5-Hydroxymethylcytosine

It was only recently in the year 2009 that two groups identified 5-hydroxymethylcytosine (hmC) as the direct oxidation product of 5-methylcytosine (mC) and thus the "sixth" base of the genome of higher organisms.[78,79] The DNA base hmC was identified before, e.g. in phages[97,151], however its presence in higher organisms was in general considered a oxidative lesions.[152–154] Also, earlier quantification data of hmC in higher organisms could not be reproduced by others and the quantified values of hmC differed comparing to novel studies.[104,155–158] These novel quantification studies by Carell and co-workers were performed in several mouse tissues and revealed that the level of hmC varies in-between organs.[104,159,160] While in most organs hmC levels make up about 0.05%-0.15% per C, in brain the levels raise to more than 1% per C. The Carell research group also demonstrated that the levels of hmC rise, correlating with the age of the organism and then remain constant. In adult human brain tissue these levels can reach up to 1.2% of all cytosines.[104,159,161–165] In contrast to mC which is also known to occur non CpG context[166,167], more than 99% of all hmC is present in this sequence context.[168] In addition, hmC can, again in contrast to mC[105], be distributed asymmetrically in the DNA double helix.[168–174] Most of the hmC is positioned in euchromatin regions of the genome[175,176], hence regions with high gene density and activity.[177,178] In this euchromatin regions, especially distal regulatory enhancers, gene promoters, gene bodies, and the proximity of transcription start sites are regions of high hmC content.[176,179] About half of the hmC is associated with cis-regulatory elements where transcription factors bind.[168,180–182] hmC is mostly present in promoters with low to mid CpG levels and as

the CpG density correlates with gene expressions, hmC is correlated with weak to mid-expressed genes.[168,183–185] Highly expressed gene promoters do not contain hmC.[168,181]

The positions of hmC are conserved between human and murine tissue and hmC is most likely connected to the epigenetic control of transcription.[176,179,186] If hmC is present in the gene body, this seems to correlate with active transcription.[169,187–190] However hmC is not in general connected to active gene expression. Robertson *et al.* demonstrated a repressive effect of hmC in HeLa cells if present in a promoter.[191] Other studies including a recently published paper by the Khopta group also revealed a negative effect of promoter-located hmC on gene expression.[192,193] As mentioned earlier, hmC accumulates during aging. In addition, it seems to be a stable mark in the genome, hence it is likely that hmC specific reader proteins are expressed.[194–197] The impact of hmC by a DNA-protein interaction may be due to the perturbation of the DNA binding of proteins or by attraction/rejection of specific binders.[158,175,198] For example, hmC prevents the binding of several mC-binding proteins, therefore affecting gene expression.[154,190] Several enzymes have differential binding affinity to hmC and mC, e.g. CXXC containing enzymes or SRA domain containing enzymes.[199–201] UHRF1, which is essential for the maintenance of mC, binds both mC and hmC, but conflicting studies were published regarding the preferred affinity.[194,202–205] Several studies also discovered hmC binding proteins with to date unknown functions like THY28, which may be involved in apoptosis.[194,206,207] Besides that, it is worth mentioning and striking that apparently hmC is involved in memory formation, learning, and neurological diseases like anxiety, Alzheimer's or Huntington disease.[208–213] Other studies suggest that hmC may also be involved in DNA repair.[214,215] In summary, it seems evident that hmC has diverse biological roles, although a detailed molecular understanding of this novel DNA base is still missing and its precise function remains elusive.

2.2.2 5-Formylcytosine and 5-Carboxylcytosine

Only two years after the discovery of hmC, Pfaffeneder *et. al* and Ito *et al.* identified the subsequent oxidation product of hmC in stem cell DNA, 5-formylcytosine (fC).[80,81] He *et al.* identified 5-carboxylcytosine (caC) the same year.[82] As depicted in Figure 5 the levels of fC and caC are significantly lower compared to hmC. Only about 0.02%-0.001% of all cytosines are fC or caC respectively.[83] Again these levels depend on the age of

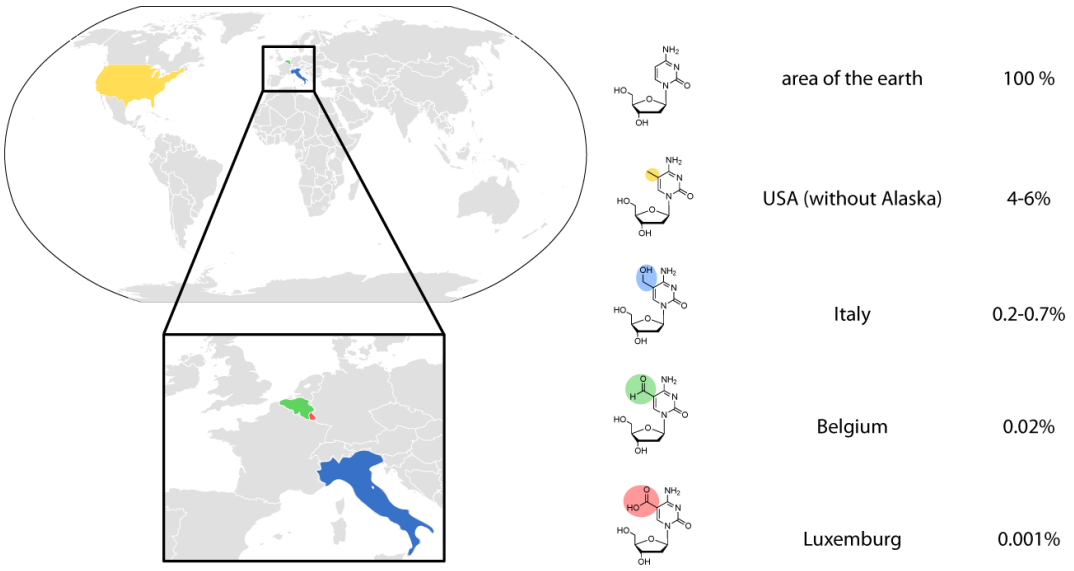


Figure 5: If the area of the landmass of earth would represent the amount of C in the genome, mC would make up about 4%-6% of this area, or roughly the area of the United States (without Alaska). Italy would represent hmC and the levels of fC and caC would be represented by Belgium and Luxemburg with 0.02% and 0.001%.[216]

the organism, but in contrast to hmC, the levels do not rise with age but decline shortly after birth.[159] The highest levels of fC are found in embryonic stem cells, while hardly present in somatic cells. The oxidation product caC is not detectable in somatic cells.[83] Sequencing studies in murine embryonic stem cells and neural progenitor cells were performed in wild type and Thymine-DNA glycosylase (TDG) knockout cell lines, as the glycosylase TDG is known to process fC and caC (see chapter 3.3.2), therefore removing it from the genome.[82,217–222] The levels of fC and caC in the TDG knockout cells were about eight times higher than in the wild type, which may hint towards a dynamic regulation inside the cell.[217–219] Like hmC, fC and caC are also mostly present in regulatory regions of the genome or inside the gene body. fC is also asymmetrically positioned like hmC.[170,220] The highest levels of fC and caC are detected in distal regulatory elements, in the regions of active and bivalent enhancers.[217,219] Bivalent enhancers are inactive regions of the genome, that can be activated very quickly.[223] It would be intriguing to interpret this localization at bivalent regions as a cellular mechanism to quickly switch on a gene after methylation via oxidation reactions and subsequent removal with TDG (or another mechanism).[222] Indeed Zhu *et al.* demonstrated very recently that the production of fC in promoters precedes the upregulation of gene expression.[224] The function of

fC and caC, present in gene bodies is not yet fully understood. They may interact with RNA Polymerase II (RNAPII) and therefore fine-tune gene transcription.[225,226] Splicing may also be influenced by fC and caC, like for mC.[179] It should be mentioned that some sequencing studies contradict each other, which may be due to the low abundance of fC and caC. While Raiber *et al.* and the group of Neri *et al.* found high levels of fC and caC in promoters of highly expressed genes[218,221], Shen *et al.* found increases fC levels in promoters with low to medium activity.[217] Improved sequencing methods may clarify these discrepancies in the future.

As mentioned levels of fC and caC rise in TDG knockouts, supporting the idea that maybe oxidative mC derivatives are necessary to prevent hypermethylation and therefore protect from transcriptional repression. For example, caC is accumulates at OCT4, NANOG, SOX2 and ESRRB binding sites in TDG-knockout mESCs.[217] These sites are perhaps regions of high oxidation and subsequent removal.[217] fC and caC are most likely connected with gene transcription and/or cell differentiation, which is supported by studies on neural progenitors and their differentiation to neurons or glia.[227] While cell specific promoters in progenitors cells are heavily methylated (causing these genes to be inactive) during differentiation neuron specific promoters get demethylated and therefore activated. This demethylation is TET and fC and caC dependent and a TDG knockdown causes an accumulation of fC and caC in these regions.[227]
Like hmC the oxidation products fC and caC may also recruit specific binders, as at least fC seems to be a partially stable mark in the genome.[194,197,228] Also, fC levels can reach high levels at specific local regions.[220] fC interacting partners, identified by the Carell/Vermeulen group[194] and the Reik group[197] include proteins like TDG or members of the NuRD complex, a histone deacetylase complex that represses gene expression.[229,230] Forkhead-box-proteins (FOX) were also identified as fC binders. FOX proteins are eukaryotic transcription factors, that influence the local chromatin structure and are connected to embryonic development and metabolic processes.[231,232]. In addition, fC positions in the genome correlate with the binding sites of the transcription factor p300, which supports the idea of fC being a distinct epigenetic mark.[219] It is unclear whether fC is perturbing the DNA structure and hence influence binding of proteins to the DNA, as contradicting studies are published.[233,234] In general, the formyl group of fC would be an attractive position for protein-DNA interactions. *In vitro*, the formation of a Schiff base could be shown for fC[235], hence a reaction with lysine rich histones would be intriguing, especially since the lysines in histone tails are subject to

post-translational modifications, which have the potential to fine-tune this reaction (e.g. by mono- or dimethylation). Indeed two recent studies identified DNA-Protein cross-links between fC and the nucleosome core particles, both *in vitro* and *in vivo*.[236,237] The biological relevance of these cross-links has yet to be studied in more detail. Not much has been published so far for caC, probably due to its very low abundance. An interaction with DNMT1 was identified by Carell and co-workers, although the relevance of this finding has yet to be assessed.[194] In addition the CXXC domain of TET3 has been identified as a specific reader for caC.[238]

2.3 TET Enzymes

The enzymatic oxidation of mC to hmC, fC and caC is catalyzed by ten-eleven-translocation enzymes (TET enzymes).[78,80–82] In vertebrates, three isoforms of TET enzymes are known[239] and all of them are capable of oxidizing mC up to caC. Initially, the human TET1 enzyme was identified as a fusion partner of KMT2A (MLL) in cancer.[240,241] In 2009 it was then characterized as an orthologue of *Trypanosoma brucei* base J-binding protein (JBP1, JBP2).[78,239,242–244] TET enzymes are key players in the pluripotency network during embryonic development and their expression is crucial.[245] For example TET1 and TET2 are regulated by OCT4 and while the levels of these enzymes decrease during differentiation, the levels of TET3 rise.[246,247]

2.3.1 Mechanism and Substrates

The exact mechanism of TET enzymes is not yet fully understood, however similarities to other oxygenase enzymes are to be expected.[186,248–253] Mechanistically the process can be described as two single electron reductions of iron (see Figure 6). The reaction starts with the coordination of α-ketoglutarate to the iron (II) species. The substrate mC facilitates the coordination of molecular oxygen in the active center of the enzyme, resulting in a decarboxylation of α-ketoglutarate and an active iron (IV)-oxo species. A proton is then abstracted from mC via a radical process. The substrate is subsequently hydroxylated, again via a radical mechanism. Whether or not the oxidation of mC up to caC occurs in an iterative or processive fashion is currently debated as contradicting

studies have been published.[254,255] This *chemical processivity*, however, has to be distinguished from *genetic processivity*. The latter describes the genetic outcome *in vivo*, hence whether a genomic regions is oxidized to hmC or further to fC and caC. This genetic processivity is depending on the chromatin state and accessibility, as well as transcription factors and the supply with α-ketoglutarate.[170,171,217,256–258]

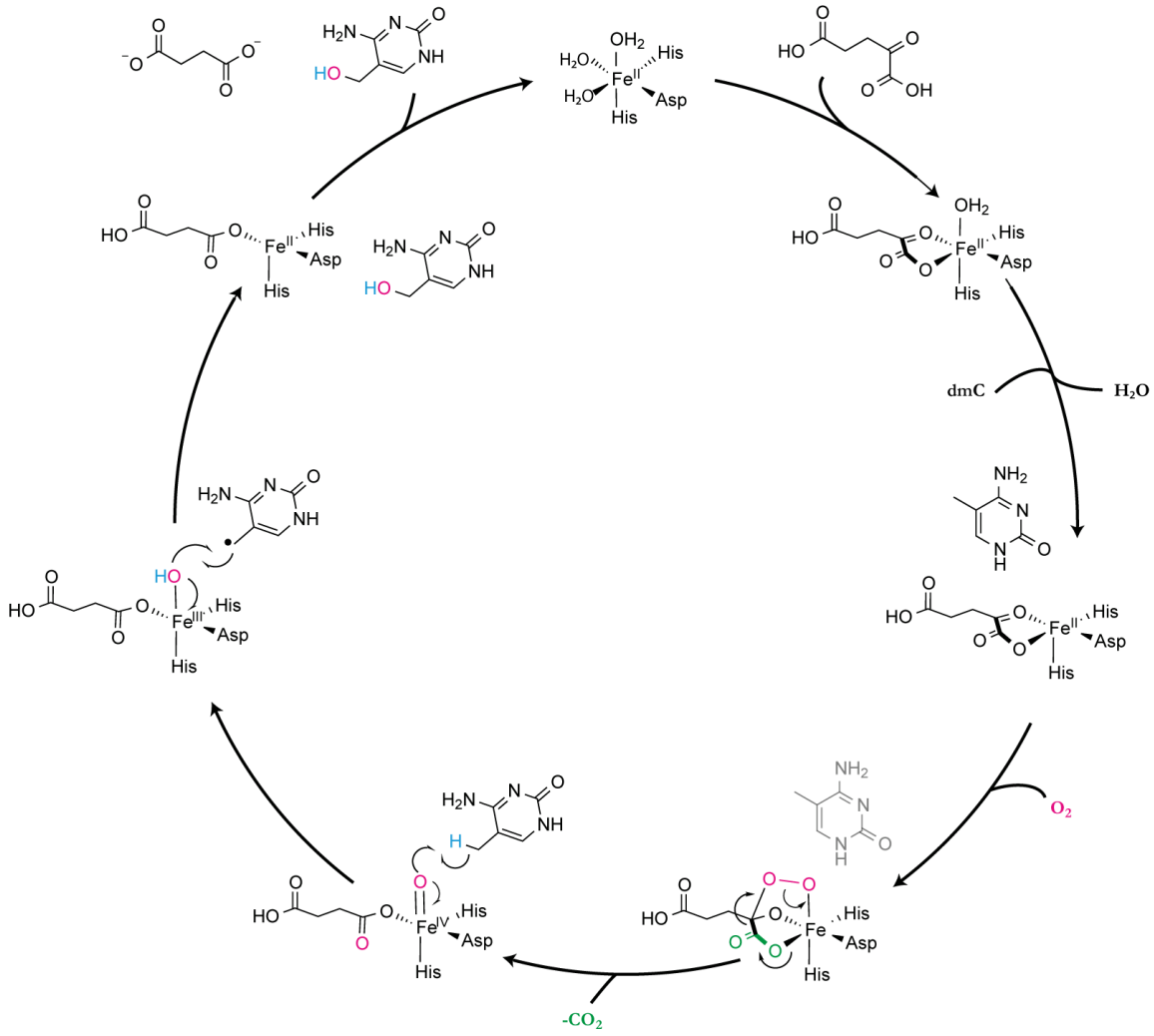


Figure 6: Mechanism of iron dependent dioxygenases. α-ketoglutarate coordinates the iron in the active center of the enzyme and replaces a prior coordinated water molecules. One water molecule completes the octahedral structure of the complex. The substrate (mC) weakens the binding of the water molecule to iron, enabling molecular oxygen to coordinate. The subsequent decarboxylation is not yet fully understood, but results in an active ferryl-oxo species. This reacts in a radical manner with the substrate. In the last step of the reaction cycle the product (hmC) and succinate are released.

The substrate specificity of TET enzymes was studied by Hu and colleagues.[259] They demonstrated in their study that TET enzymes prefer mC compared to hmC or fC as a substrate. It should also be mentioned that TET enzymes prefer mCpG dinucleotides rather than a mCpA, mCpT or mCpC context.[249,251] TET proteins are able to oxidize several combinations of substrates e.g. mC, hmC or fC paired with various modified bases.[257,260,261] Carell and co-workers showed in 2014 that TET enzymes also recognize thymine as a substrate and oxidize it to hmU in a controlled and regulated manner.[83] The oxidation of mC in a RNA context is also possible.[262]

2.3.2 Structural Properties of TET Enzymes

The C-terminal domains of TET enzymes harbor the catalytic active center. It is formed from a double stranded beta-helix (DSBH) and a cysteine-rich domain, wrapping around the DSBH and stabilizing the structure.[249–252]. The DSBH motif is important for DNA and substrate binding and brings together the three components, iron (II), α-ketoglutarate and mC. A conserved arginine enables the binding of α-ketoglutarate, while the iron itself is coordinated by two histidines and one aspartate.[249,251] TET enzymes have a variable insert in their DSBH region, pointing away from the catalytic center.[249] These motifs differ in their length and resemble the C-terminal domain (CTD) of RNA Polymerase II (RNAPII) and are dispensable for catalytic activity.[263] The catalytic domain of TET enzymes alone, is sufficient for oxidation and nuclear localization.[78,245,264] The full length TET1 and TET3 enzymes also have a CXXC domain. It consists of about 60 amino acids and is located at the N terminus. Splicing versions without the CXXC domain are however also expressed.[265,266] Notably, these isoforms are expressed in different tissue and show different activities.[238,265,266] A chromosomal inversion in TET2 led to a loss of its CXXC domain. This motif is now a distinct gene, called IDAX.[267]

2.3.3 Non-Enzymatic Effects of TET Enzymes and their Regulation

Besides their enzymatic function as oxidizing enzymes, TET enzymes can influence cellular processes by other mechanisms.[88,186,268] For example, it is interesting that TET proteins prefer regions of the genome that do not correspond to regions of high mC and hmC

levels.[105,109,188,269] TET enzymes also influence transcription differently. While TET1 and TET3 repress gene expression in mESCs[188] TET2 is activating transcription.[270] These differences can be explained by the recruitment of diverse binding partners, e.g. TET1 recruits the repressive SIN3 or NuRD complex[188,271–274], while TET2 can activate gene expression by the interaction with O-GlcNAc transferase (OGT), which can activate gene expression by influencing the chromatin.[270,275–277]. TET enzymes are also involved in telomere elongation and chromatin stability.[278,279]

TET enzymes themselves can be regulated in various ways, e.g. by controlling the supply of their co-substrate. α-ketoglutarate is produced from isocitrate by IDH1, IDH2 and IDH3.[280] Overexpression of IDH1 and IDH2 result in elevated hmC levels and downregulation in decreased levels.[281,282] Mutated IDH versions produce 2-hydroxglutarate which inhibits TET enzyme by competing with α-ketoglutarate.[281,283] Other small molecules like fumarate, succinate, glutamate or glutamine also can decrease or increase hmC levels.[256,284,285] Vitamin C is known as a stimulating agent for TET activity. This effect may be due to vitamin C acting as a "co-substrate" (as stated by minor *et al.*[286]) and directly interacting with the catalytic domain of TET enzymes. In addition it may promote the recycling of iron(II).[257,286–288] Other pathways of regulation include the supply with molecular oxygen[285,289–291] and iron[292], both affecting the oxidation of mC. TET enzymes may furthermore be regulated post-translationally (will also be discussed in chapter 10) and post-transcriptionally. Several microRNAs regulate TET enzymes[293–297] and the translation of TET1 mRNA is facilitated by the protein DAZL.[298] After translation TET enzymes are modified by post-translational modifications, e.g. monoubiquitination, acetylation, GlcNacylation, PARylation and phosphorylation.[299–302] Last but not least, the amount of TET enzymes, and hence indirectly the activity, may be regulated by protein-protein interaction and proteolysis. If IDAX is overexpressed, TET2 degradation is increased, while depletion results in increased TET2 levels.[267] On the other hand TET activity may be enhanced by UHRF2.[194] The degradation of TET enzymes seems to be calpain dependent proteolysis in murine ESC, while in cancer cells, the ubiquitin-proteasome pathways could regulate TET levels.[247,303]

3 Demethylation of 5-Methylcytosine

In mammals demethylation of mC is required at several developmental stages and time points.[179] In consequence, formerly repressed genes are able to be expressed again.[304–306] This demethylation process can either be global[305,307–309] or occur at specific loci.[198,310–313] Global demethylation mainly occurs during embryogenesis whereas locus-specific demethylation of mC is a process that is crucial e.g. for neural plasticity.[198,310–312]

Demethylation of mC is in principle possible via two pathways; active and passive demethylation. Passive in this context describes the dilution of mC during cell division and DNA replication respectively. In this case the methyl information is not passed on to the daughter DNA strand by DNMT1.[314] In theory a passive demethylation is also possible via the dilution of hmC, fC or caC.[315] Several studies showed that DNMT1 is less efficient in a hmC/C, fC/C and caC/C context compared to the native mC/C.[202,316–319] If this replication dependent dilution is repeated several times, mC (and possibly hmC, fC and caC) will get diluted over time, resulting in a demethylated region of the genome.[316,320] The active demethylation pathway is not coupled to DNA replication and will be described in more detail in chapter 3.3.[304]

3.1 Global Demethylation of 5-Methylcytosine

As mentioned, global demethylation events take place during embryogenesis. Only very few DNA loci are excluded from this event.[307] This process takes places during the zygote phase and establishes the totipotent character of the cell.[137] Later during development a second global demethylation wave is necessary for the development of primordial germ cells (PGCs).[307,321–323]

The global demethylation event during the zygote phase is replication independent as it occurs before the first cell division. Here the paternal pronucleus is rapidly demethylated.[134,135] In fact studies by Iqbal *et al.* and others showed that this demethylation event is coupled to an oxidation of mC to hmC and further to fC and caC.[320,324,325] This oxidation is carried out by TET3, which is in contrast to TET1 and TET2 highly

expressed in the zygote.[326] Therefore it is not surprising that a TET3 knockdown or knockout perturb correct embryonic development, as essential genes for totipotency like OCT4 or NANOG are not demethylated in the paternal pronucleus.[326,327] It should be noted, that in contrast to mC, no mechanism is known to date that passes on hmC, fC or caC to the daughter DNA strand. Therefore these epigenetic marks are diluted hand in hand with cell division and hence the genome gets demethylated.[316] It is surprising that although TET3 is present in the zygote and the nuclei are no longer confined to seperate cell, it oxidizes only the paternal mC. The maternal pronucleus seems to be protected of this reaction. Several protection mechanism, e.g. proteins like Stella and histone modifications (H3K9me2) work hand in hand to prevent the oxidation of maternal mC.[327–329] The maternal mC is therefore removed in a replication dependent manner.[330] It is still not clear why the paternal and maternal DNA are demethylated differently. It should be noted, however that the methylation levels of the paternal and maternal DNA differ. While the paternal DNA is hypermethylated with about 90% of the CpGs being methylated, the maternal mC level in CpG context accounts for only about 40%.[331]

As mentioned above, embryogenesis is not the only phase of global demethylation. The formation of primordial germ cells (PGCs) also involves global demethylation of mC.[321–323,332] This process can be characterized in two phases. The first one being a passive demethylation[333,334] and a subsequent active demethylation.[321,332,335] This second, active demethylation seems to involve TET1.[304]

3.2 Locus Specific Demethylation of 5-Methylcytosine

Demethylation is not limited to embryonic or germ cell development. Also somatic cells undergo demethylation at specific DNA loci.[336–341] This process and detailed mechanism is still under investigation and possible pathways will be discussed in chapter 3.3.[306] One main somatic system is prone for loci dependent demethylation, neurons.[310,342–344] It is very exciting that these events occur at promoter regions of genes like BDNF (Brain-derived neurotrophic factor), involved in neural plasticity and learning.[310,338,345] Demethylation can occur after e.g. electric stimulation[346,347] and is at least in some cases a TET1 dependent oxidation to hmC and maybe fC/caC.[346] Also more complex stimuli can trigger demethylation, e.g. contextual fear conditioning results in locus spe-

cific demethylation in the promoter region of BDNF.[310,348] It becomes evident, that both, synaptic plasticity and the establishment of memory is a complex clockwork of demethylation, *de novo* methylation and also histone modifications in neurons.[310,349,350]

3.3 Active Demethylation Pathways

In plants, the demethylation of mC is carried out by the base excision repair (BER) machinery. A family of glycosylases, namely Demeter (Dme) recognize mC and remove it from the genome.[351] This process results in the replacement of mC by C. Dme enzymes are capable of demethylating mC not only in the CpG context, but also in a CpNpG and CpNpN (N = A, T or C) sequence context *in vitro*.[352] In mammals, no Dme orthologues have been identified to date. Nevertheless, a similar glycosylase activity has been proposed for the Thymine-DNA Glycosylase (TDG) and the Methyl-CpG-Binding Protein 4 (MBD4). However, their activity was significantly lower compared to their native substrates (T-G mismatches).[353–356] Also MBD4 deficient mice are vital and do not exhibit strong effects on demethylation of the paternal DNA in the zygote.[357,358] These data suggest a minor role for DNA glycosylase activity for the demethylation of mC. Other possibilities were therefore proposed, but it remained an unsolved question how demethylation in mammals occurs in detail. The discovery of hmC and the TET enzymes moved the research field in a new direction. It seems plausible that hmC may be an intermediate in an active oxidative demethylation pathway (described in chapter 3.3.2 and 3.3.3). Three pathways for active demethylation will be discussed in the following (see Figure 7). Two involve the base excision repair system, the other one is based on a direct C-C bond cleavage reaction.[90,179,304,315]

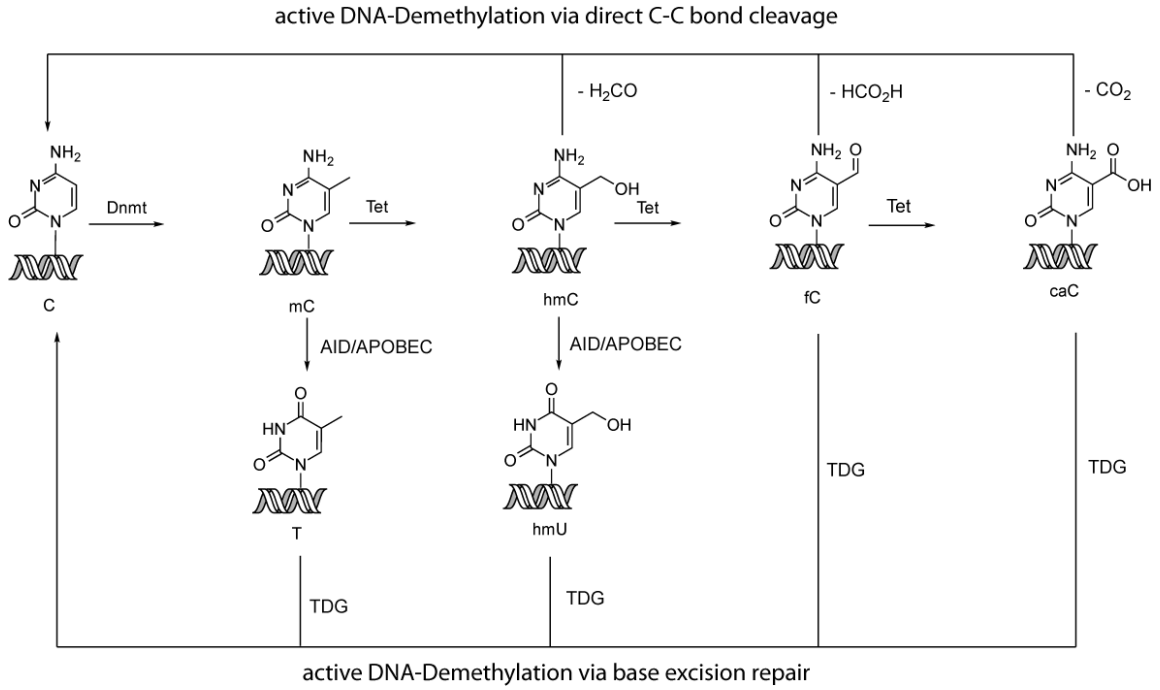


Figure 7: Besides passive demethylation three active pathways for the removal of modified cytosines were proposed. One being the direct removal of the chemical modification from the base (upper pathway). This pathway avoids the BER-machinery and results in unmodified cytosine. The two other possible pathways (down) would utilize the BER machinery, in case of fC and caC directly or following a deamination of mC or hmC, respectively.

3.3.1 Active Demethylation via Deamination

One possible pathway for the active removal of mC from genomic DNA starts with a deamination of mC to T. The resulting T-G mismatch is then repaired by the glycosylase TDG or MBD4. Deaminases like the Activation-Induced Deaminase (AID) and the Apolipoprotein B pre-mRNA Editing enzyme catalytic polypeptide(APOBEC) have been proposed as mC deaminases.[359–363] Both enzymes are capable of deaminating mC to T *in vitro* and in an *E. coli* assay.[361,364] Some studies suggest, that deamination is initiating demethylation e.g. in zebrafish or Primordial Germ Cells (PCGs).[365] However, there are also studies showing only minor contributions of deamination to demethylation of mC, e.g. bisulfite sequencing studies revealed a small contribution of deamination to the active demethylation of mC in PGCs.[366] Also several studies demonstrated a strong

correlation in the activity of the deaminases, depending on the size of the substituent at C5, making mC worse a substrate than C.[367,368] In addition the deamination of mC to T could not be reproduced for dsDNA, making it at least unlikely that this mechanism is relevant *in vivo*.[364] Another class of enzymes that are reported to be capable of deaminating mC are DNA methyltransferases.[340] In general it seems attractive to have an enzyme that is able to catalyze both, the methylation and demethylation (via deamination) of cytosine. This would facilitate the cell to rapidly and dynamically regulate its mC levels.[340] One proposed mechanism for the deamination reaction is depicted in Figure 8 (right). Chemically the saturation of the C5-C6 bond in the covalent protein-DNA intermediate would facilitate the nucleophilic attack of water at the C4 position. The stabilized amino group would act as a leaving group and its substitution by water would result in T.[369] The observed deamination, however, was only proven indirectly by

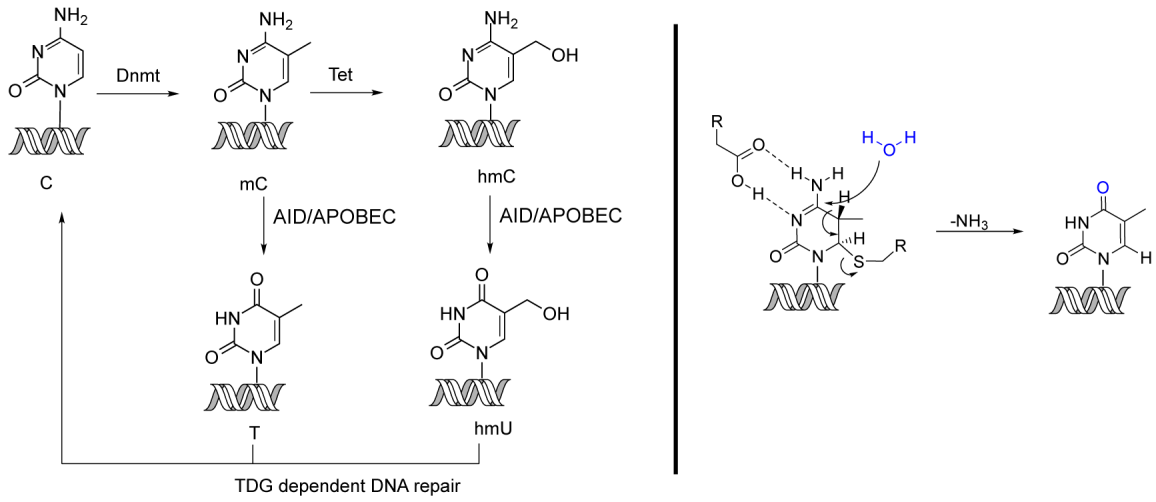


Figure 8: AID and APOBEC might deaminate mC or hmC respectively. The resulting T or hmU mismatch (T:G, hmU:G) would subsequently be repaired by the DNA glycosylase TDG followed by the BER machinery. A nucleophilic attack at C6 would saturate the C5-C6 bond. The subsequent attack of water at C4 would result in the deaminated species.

a Ligation-mediated-PCR assay after potential deamination and AP-site cleavage. Also, the reaction took place at very low SAM concentrations.[340] Therefore the biological relevance of this setup remains unclear, as SAM is ubiquitous in the cell and necessary for various biological processes, while this process would require rapid spatial changes in the SAM-level.[370,371]

Besides mC also hmC could potentially be deaminated and then further processed as hmU. Studies showed that overexpression of AID or other members of the AID/APOBEC family did decrease hmC levels in HEK cells.[346] But also contradicting studies are published showing only minor effects of AID/APOBEC on oxidative mC derivatives.[372] Also the Carell research group could demonstrate that the steady-state level of hmU is not dependent on hmC deamination. Instead hmU is enzymatically produced via the oxidation of T by TET enzymes in a regulated manner.[83] Last but not least, as mentioned the deamination activity is very weak in case of a large C5 substituent, hence the weak activity towards mC is even lower for hmC.[367,368,372] In summary, these results at least challenge the idea of active demethylation through deamination of mC to T or hmC to hmU.

3.3.2 Active Demethylation via TDG Excision

Although mC, as a substrate would ease the complexity of active demethylation, it is unlikely that TDG is a mC glycosylase.[373] TDG is known to function as a Thymine-DNA-glycosylase and repairs T-G mismatches. However the repair of mutagenic mC to T deaminations seems to be only one part of the story and TDG's relevance for active demethylation is more prominent.[179,373,374] In contrast to other glycosylases, TDG interacts with transcription factors, *de novo* methyltransferases, histone modifying enzymes and also TET enzymes.[260,375–378] The association of DNMTs with TDG influences their activity.[377] Methylation activity of DNMT3a is for example reduced upon TDG interaction, whereas the activity of TDG is enhanced. These interactions suggest a more prominent function of TDG in epigenetics. The importance of TDG for normal development is supported by the fact that TDG knock-out mice die after several days of development and cell lines derived from these organisms have a dysfunctional gene expression profile. It is worth mentioning that the the lethality occurs after the time of active demethylation.[218,379,380]

TDG as a "demethylating" enzymes gained much more attention after the discovery of fC and caC, as it could be demonstrated that TDG is capable of processing these bases.[82,222,374,381–383] It should be noted that TDG removes fC more efficiently from DNA than caC, although its affinity towards caC seems to be higher.[222,384] TDG is capable of recognizing various combinations of substrates at a CpG site, e.g. caC/mC,

caC/hmC or caC/caC.[260] These studies are the foundation of a putative demethylation pathway depicted in Figure 9. mC is oxidized by TET enzymes to hmC and further to fC and caC. These two non-canonical DNA bases are then removed from the genome by TDG.[82,222] The resulting abasic site is repaired by the BER machinery.[385] However, as described in chapter 3.1, embryonic development goes hand in hand with the rapid removal of mC, a process that involves TET dependent oxidation.[134,135,320,324] The described pathway through TDG involves the BER machinery and therefore introduces strand breaks, which would be potentially mutagenic. Elevated levels of DNA strand breaks are observed during differentiation[386], but the removal of fC and caC does not result in an accumulation of intermediates of the BER pathway.[387] This would argue either for a TDG independent pathway, which also has been proposed[388] or that TDG acts in a complex with other enzymes, therefore rapidly processing abasic sites and avoiding harmful intermediates.[389] The potential mutagenic behavior of strand breaks could be bypassed by an active demethylation pathway as described in the following section.

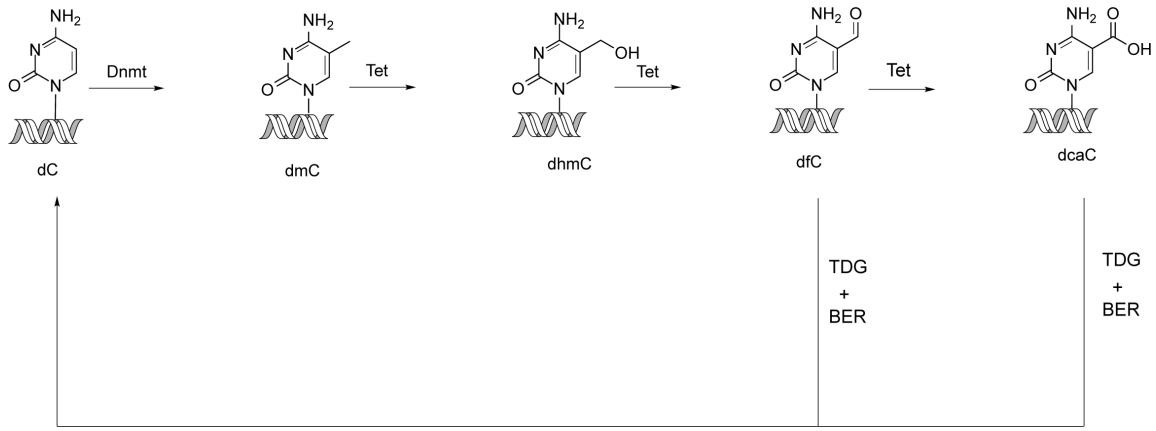


Figure 9: TET enzymes oxidize mC to hmC and further to fC and caC. The last two oxidation products of mC are recognized and processed by the mismatch specific glycosylase TDG. The resulting abasic site is repaired by the BER machinery and cytosine is reincorporated. This pathway includes potentially mutagenic strand breaks.

3.3.3 Active Demethylation via direct C-C Bond Cleavage

From a chemical perspective, fC and caC are prone for a direct C-C bond cleavage reaction at the C5 position. In theory, also an hmC-based bond cleavage is possible.

These reactions would generate an unmodified cytosine residue without the need of the BER machinery and the generation of abasic sites, hence avoiding potentially mutagenic strand breaks. The putative enzyme would therefore be a dehydroxylase, deformylase or decarboxylase. In this scenario, the decarboxylation of caC is the most favorable reaction, as Carell and co-workers could demonstrate.[390] The proportion of reactivity between caC, fC and hmC is roughly estimated as 30:3:1. A nucleophilic attack at the C6 position, e.g. by a cysteine could facilitate the decarboxylation/deformylation reaction depicted in Figure 10. Several studies could show that these dehydroxylation, deformy-

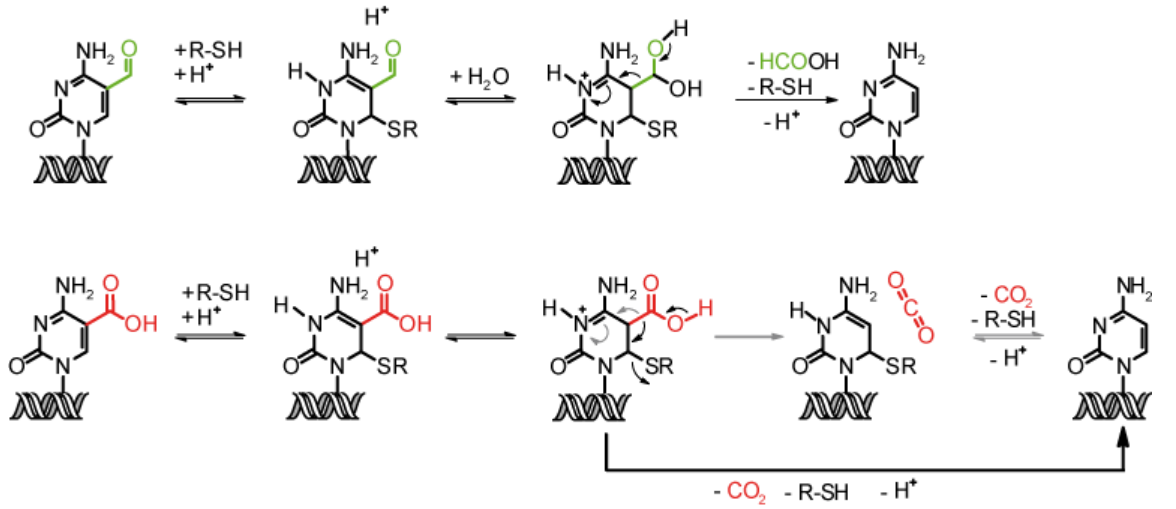


Figure 10: A nucleophilic attack, e.g. by a cysteine at C6 would facilitate the demethylation reaction. For the deformylation an additional attack at the aldehyde is necessary, e.g. by water. The figure is adapted from Schiesser *et al.*[390]

lation and decarboxylation reactions are in principle possible. A mutated version of the DNA methyltransferase HhaII could add or remove aldehydes, also on hmC.[391] This reactivity was also shown for DNMT3a and DNTM3b. Catalyzing this reaction would enable DNMTs to add and remove the methyl group without the need of another enzyme.[392] Last but not least it was demonstrated that DNMTs can directly demethylate mC to C.[393]. This direct removal of the chemical modification could also be detected for caC.[394] However these studies were mostly artificial and the contribution of these reactions to a native systems is not clear.

In general oxidative demethylation reactions are known to occur in a native context, however with other substrates. Lysine residues for example, can be demethylated in an oxidative fashion by Jumonji and amine oxidases, however in this case the reac-

tion takes place at a C-N bond.[395–397] ALKBH5/FTO are enzymes that are capable of demethylating RNA bases by oxidation reactions.[398,399] Also for DNA, an oxidative demethylation pathway is known. Thymine can be oxidized by the enzyme Thymin-7-hydroxlase to hmU, fU and caU which is then processed by the Isoorotat decarboxylase (IDCase, *Cordyceps militaris*), resulting in uracil.[400,401] This enzyme is *in vitro* also capable of directly forming cytosine from carboxylcytosine, however only on the free DNA base.[402] A similar decarboxylation activity could be shown by Carell and coworkers in stem cell lysate. The study utilized isotopically labeled caC and incorporated the base in a synthetic oligonucleotide. After incubation with stem cell lysate heavy labeled cytosine could be detected via mass spectrometry.[403] It is not clear, if this was a specifically, biochemically catalyzed reaction and no enzyme could so far be linked to this process. A recent study by the Carell group further supports the idea of a direct C-C bond cleavage *in vivo*. This new data suggest rather a deformylation reaction than a decarboxylation reaction. Iwan *et al.* demonstrated that fC is directly deformylated to cytosine in various cell types.[404] Like decarboxylations also deformylations are known to occur in nature, yet again for other substrates. Lanosterin for example, an intermediate of the cholesterol biosynthesis, is deformylated. Also during the degradation of fatty acids deformylation reactions are observed.[405–408]

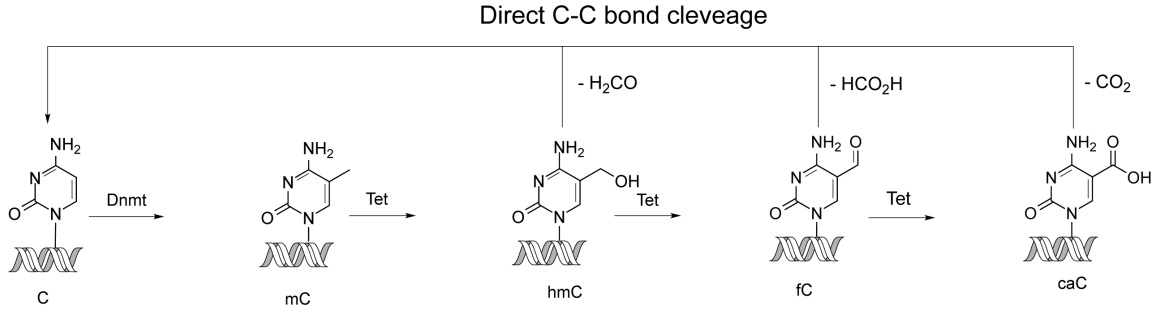


Figure 11: Possible conversion of hmC, fC and caC to C by a direct C-C bond cleavage reaction, releasing formaldehyde, formic acid or carbon dioxide. This pathway would avoid potentially mutagenic strand breaks.

Results and Publications

4 Synthesis of (R)-Configured 2′-Fluorinated mC, hmC, fC, and caC Phosphoramidites and Oligonucleotides

Investigating the metabolism and function of novel epigenetic bases can be facilitated by the use of stable analogues that are easy to detect and bioisostere. The present study reports the synthesis of 2′-fluorinated phosphoramidite analogous of 5-methylcytosine (mC), 5-hydroxymethylcytosine (hmC), 5-formylcytosine (fC), and 5-carboxycylcytosine (caC). They are not cleaved by thymine-DNA glycosylase (TDG) and M.SsI, a DNA Methyltransferase is capable of methylating 2′-fluorinated cytosine, demonstrating that the 2′-fluorination does not alter epigenetic recognition.
For this study, I performed the *in vitro* methylation assay and subsequent sample preparation for the LC-MS/MS analysis of the reaction. I further participated in the data evaluation of the analysis.

Organic LETTERS

Letter

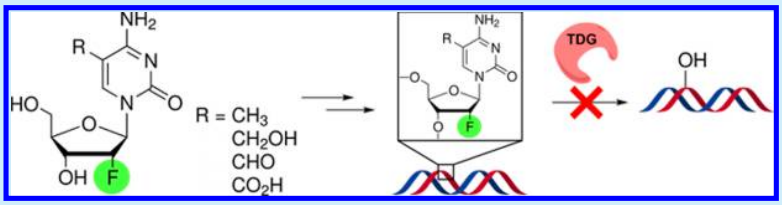
pubs.acs.org/OrgLett

Synthesis of (*R*)-Configured 2′-Fluorinated mC, hmC, fC, and caC Phosphoramidites and Oligonucleotides

Arne S. Schröder, Olga Kotljarova, Edris Parsa, Katharina Iwan, Nada Raddaoui, and Thomas Carell*

Center for Integrated Protein Science, Department of Chemistry, Ludwig-Maximilians-Universität München, Butenandtstraße 5-13, 81377 Munich, Germany

Supporting Information



ABSTRACT: Investigation of the function of the new epigenetic bases requires the development of stabilized analogues that are stable during base excision repair (BER). Here we report the synthesis of 2′-(*R*)-fluorinated versions of the phosphoramidites of 5-methylcytosine (mC), 5-hydroxymethylcytosine (hmC), 5-formylcytosine (fC), and 5-carboxycytosine (caC). For oligonucleotides containing 2′-(*R*)-F-fdC, we show that these compounds cannot be cleaved by the main BER enzyme thymine-DNA glycosylase (TDG).

Fluorine is an element that is used in medicinal chemistry to replace H atoms in pharmaceutically active molecules with astonishing effects. Fluorine substitution stabilizes molecules to extend their lifetimes in the bloodstream, and often it increases the affinities of molecules for their biological targets by increasing their lipophilicities.[1] In nucleoside chemistry, for example, fluorination of dC at the 2′ position creates molecules like gemcitabine (**1**), which are used as antimetabolites in modern cancer therapy.[2] The 2′-F substitution has several effects. Most importantly, a 2′-(*R*)-configuration as in 2′-(*R*)-F-dC (**2**) stabilizes the C3′-endo conformation of the ribose sugar so that the base becomes RNA-like.[3] A fluorine at C2′ also blocks the activity of glycosylases, thereby stabilizing the base during base excision repair (BER).[4] We are currently investigating the chemistry that occurs at the nucleoside 2′-deoxycytidine (dC, **3**) that leads to the formation and removal of the methylated and subsequently oxidized epigenetic dC derivatives 5-methyl- (mdC, **4**), 5-hydroxymethyl- (hmdC, **5**), 5-formyl- (fdC, **6**), and 5-carboxy-2′-deoxycytidine (cadC, **7**) (Figure 1).[5] Nucleosides **5**−**7** are products of consecutive enzymatic oxidation of **4** by the action of ten-eleven-translocation enzymes (Tet enzymes), which use molecular oxygen and α-ketoglutarate to perform the oxidation chemistry.[6] Current data suggest that fdC and cadC are removed from the genome by BER via the enzyme thymine-DNA glycosylase (TDG).[5d,7] Other data predict that the bases may undergo some kind of deformylation/decarboxylation reaction, which would convert fdC and cadC directly back into the canonical base dC.[8] In order to distinguish these processes, it is important to have tool molecules that cannot be repaired by BER. This would allow one to decipher chemical processes at fdC and cadC beyond BER. Here we report the synthesis of

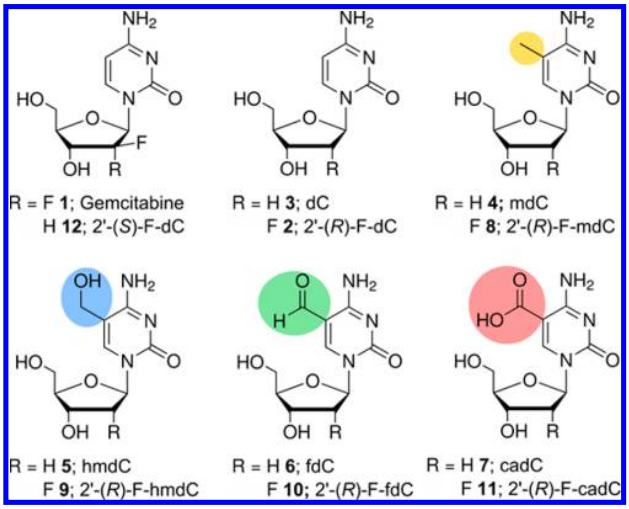


Figure 1. Overview of epigenetically relevant nucleosides and 2′-fluoro nucleosides that are important in this context.

the 2′-(*R*)-fluorinated versions of mdC (**8**), hmdC (**9**), fdC (**10**), and cadC (**11**). We have developed phosphoramidite building blocks for the incorporation of these bases into DNA strands, and we show that these nucleosides are indeed stable during BER. With the plan in mind to investigate epigenetic processes directly in the genome of stem cells, we realized that the 2′-*arabino*-configured compound 2′-(*S*)-F-dC (**12**) might be too toxic. Indeed, when we evaluated the toxicity of the *ribo*-

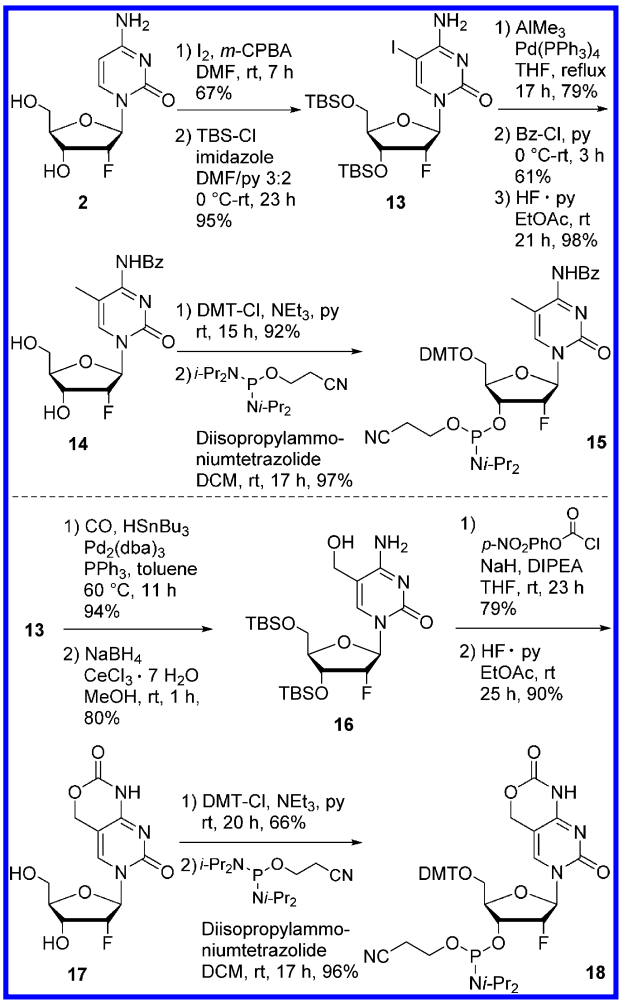
Received: July 20, 2016
Published: August 19, 2016

DOI: 10.1021/acs.orglett.6b02110
Org. Lett. 2016, 18, 4368−4371

configured compound 2′-(*R*)-F-dC against **12** in stem cells (see the Supporting Information), we noted a strongly reduced toxicity for 2′-(*R*)-F-dC. This is already interesting because it is believed that the 2′-(*S*)-F configuration has a much smaller impact on the overall DNA structure.[3,9] Our stem cell data are, however, in full agreement with toxicity studies in rats and woodchucks showing that feeding of 2′-(*R*)-F-dC at up to 500 mg kg^{-1} day^{-1} is possible without considerable toxicity effects.[10]

For the synthesis of the 2′-(*R*)-F-xdC nucleosides and phosphoramidites **15**, **18**, **21**, and **24** (see Scheme 1), we started with 2′-(*R*)-F-dC (**2**), which was iodinated at C5 with elemental iodine and *m*-CPBA.[11] Subsequent silylation yielded TBS-protected 5-iodo-2′-(*R*)-F-dC **13**. The needed methylation was best carried out under Kumada conditions with trimethylaluminum.[12] This furnished the 2′-(*R*)-F-mdC compound in 79% yield. Notably, the use of other methyl-transferring agents such as MeMgCl resulted in a 1:1 mixture of methylated and dehalogenated products. We believe that the exocyclic amine requires complete deprotonation to avoid a 1,3-proton shift from the exocyclic amine to the Pd-activated C5-position. Further protection with BzCl and silyl deprotection with Olah's reagent furnished 2′-(*R*)-F-mdC derivative **14**, which was converted into the 2′-(*R*)-F-mdC phosphoramidite building block **15** using standard procedures.[13]

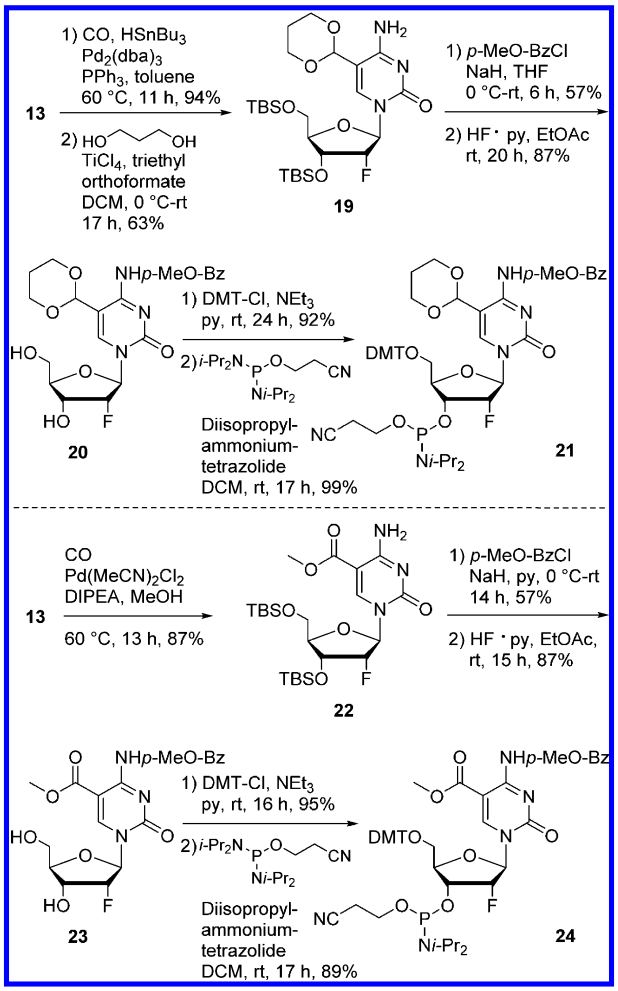
Scheme 1. Synthesis of 2′-(*R*)-F-mdC and 2′-(*R*)-F-hmdC Phosphoramidite Building Blocks 15 and 18



For the synthesis of the 2′-(*R*)-F-hmdC phosphoramidite **18**, we started from intermediate **13**. Carbonylative Stille coupling with tributyltin hydride and reduction of the formyl group under Luche conditions yielded 2′-(*R*)-F-hmdC derivative **16**.[14] The exocyclic amine together with the hydroxyl group was protected as a carbamate using *p*-nitrophenyl chloroformate.[15] Efficient conversion required full deprotonation of both functional groups with NaH prior to addition of the protecting reagent. Final silyl deprotection, DMT protection, and synthesis of the hmdC phosphoramidite building block **18** with Bannwarth's reagent furnished the 2′-(*R*)-F-hmC phosphoramidite in high yield (34% over six steps from **13**).

Regarding 2′-(*R*)-F-fdC phosphoramidite building block **21**, we performed a carbonylative Stille coupling reaction of **13** with tributyltin hydride (see Scheme 2). Subsequent masking of the formyl group as a 1,3-dioxane unit with 1,3-propanediol and $TiCl_4$ as the activating Lewis acid provided compound **19**. For the protection of the exocyclic amine, we chose *p*-$MeOC_6H_4COCl$ as recently reported.[16] The electron-pushing methoxy unit strongly enhances the stability of the amine protecting group during solid-phase DNA synthesis, and this is

Scheme 2. Synthesis of 2′-(*R*)-F-fdC and 2′-(*R*)-F-cadC Phosphoramidite Building Blocks 21 and 24



DOI: 10.1021/acs.orglett.6b02110
Org. Lett. 2016, 18, 4368–4371

strictly required in order to obtain oligonucleotides in high yields. Again, satisfactory yields were obtained only when the exocyclic amine was deprotonated with NaH prior to addition of *p*-MeOC$_6$H$_4$COCl. Final silyl deprotection yielded **20**, which was converted into 2′-(*R*)-F-fdC phosphoramidite building block **21** using standard procedures.

Starting from intermediate **13**, we next developed the synthesis of the 2′-(*R*)-F-cadC phosphoramidite building block. The synthesis of the methyl ester was achieved using Pd^0-mediated CO insertion in methanol.[17] Because of the electron-withdrawing nature of the ester moiety, we decided to use *p*-MeOC$_6$H$_4$COCl for stable protection of the exocyclic amine. Conversion of **23** using standard procedures delivered 2′-(*R*)-F-cadC phosphoramidite building block **24** in just five steps in an overall yield of 36% starting from **13**.

To examine the ability to prepare oligonucleotides containing 2′-(*R*)-F-xdC, we prepared the corresponding **ODN1a–d** (see Figure 2). The modified nucleotides were

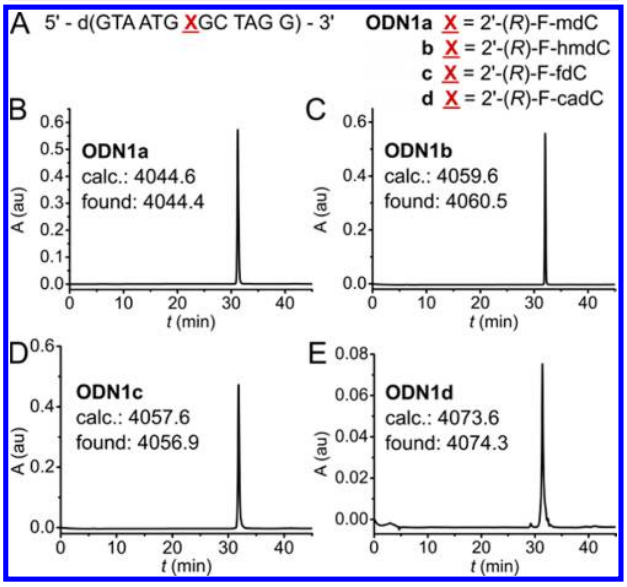


Figure 2. (A) Sequence of the synthesized **ODN1a–d** with incorporation of the corresponding 2′-(*R*)-F-xdC phosphoramidite building blocks. (B–D) Reversed-phase HPL chromatograms and MALDI-TOF data for the corresponding purified **ODN1a–d** after basic and, in the case of 2′-(*R*)-F-fdC, acidic cleavage from the resin and deprotection.

placed in a CpG context. The solid-phase syntheses were performed using standard phosphoramidite conditions.[18] For the 2′-(*R*)-F nucleosides, the coupling times were increased from 30 to 180 s to ensure good coupling yields. For deprotection of the oligonucleotides containing 2′-(*R*)-F-mdC and 2′-(*R*)-F-fdC, including cleavage from the solid support, we first treated the solid-phase material with saturated aqueous ammonia solution (18 h, 25–28 °C). Subsequently, the oligonucleotide containing 2′-(*R*)-F-fdC was exposed to aqueous acetic acid (80%) at 20 °C until MALDI-TOF/MS analysis indicated complete hydrolysis of the 1,3-dioxane unit (~6 h). Because of the carbamate and ester units, the oligonucleotides containing 2′-(*R*)-F-hmdC and 2′-(*R*)-F-cadC were deprotected with NaOH (0.4 M in 4:1 methanol/water) for 18 h. This procedure avoided the formation of aminomethyl and amide moieties.[11,17a] Analytical reversed-phase HPLC directly after deprotection showed in all cases just one major product. After purification, the corresponding oligonucleotides were obtained in 20–52% yield and high purity (>95%). MALDI-TOF/MS spectra showed the expected masses, confirming the presence of the 2′-(*R*)-F-xdC bases in the ODNs. In summary, the synthesized 2′-(*R*)-F-xdC phosphoramidite building blocks enabled the synthesis of oligonucleotides containing the corresponding fluorinated nucleosides.

We next started to evaluate the extent to which the 2′-(*R*)-F substitution would affect typical epigenetic processes. First, we wanted to know whether the H-to-F chemical mutation influences the activity of methyltransferases (see Figure 3).

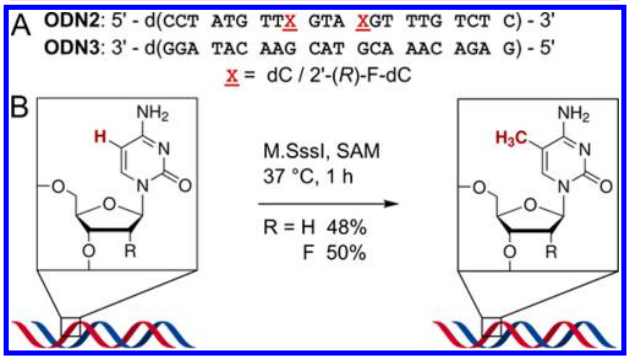


Figure 3. (A) Sequences of the synthesized **ODN2** and **ODN3** with incorporation of dC or 2′-(*R*)-F-dC nucleoside. (B) The methylation assay of **ODN2** and **ODN3** with methyltransferase M.SssI showed that the fluoro label in 2′-(*R*)-F-dC has no influence on the level of methylation.

To study this, we synthesized **ODN2** having either dC or 2′-(*R*)-F-dC in a CpG context. After hybridization of **ODN2** with **ODN3**, they were incubated with methyltransferase M.SssI. To determine the level of mdC or 2′-(*R*)-F-mdC, we digested the DNA strands to the nucleoside level and performed UHPLC-MS/MS (QQQ) analysis. As the verification of our hypothesis, we observed methylation of dC (48%) and 2′-(*R*)-F-dC (50%). This demonstrates that the 2′-(*R*)-F substitution does not affect the native behavior of the DNA and that 2′-(*R*)-F-xdC nucleosides are suitable tools for the investigation of the active demethylation beyond base excision repair.

In 2011 and 2012, the groups of Drohat[7a] and Cheng[19] showed that fdC and cadC are excised by human TDG (hTDG). Previously, glycosylase activity was blocked with fluorinated DNA bases (2′-F-(*S*)-cadC, 2′-F-(S/*R*)-dU).[4a,c] In order to determine whether the 2′-(*R*)-F-fdC compounds would block hTDG activity, we synthesized oligonucleotides **ODN4** with either the fdC or F-fdC nucleoside at a central position and hybridized the strands to the complementary oligonucleotide **ODN5**. After hybridization and incubation with hTDG, the DNA strand was treated with piperidine.[7a,19,20] Subsequently, we analyzed the products by HPLC (see Figure 4). As expected, we detected complete strand cleavage for the fdC-containing **ODN4**. However, in the case of the **ODN4** containing 2′-(*R*)-F-fdC, we did not observe any strand cleavage products. Thus, we proved that the 2′-(*R*)-F label indeed inhibits the hTDG activity, blocking BER of fdC.

In summary, we have synthesized 2′-(*R*)-F phosphoramidite building blocks of the epigenetically relevant nucleosides. These building blocks enabled the synthesis of oligonucleotides

DOI: 10.1021/acs.orglett.6b02110
Org. Lett. 2016, 18, 4368–4371

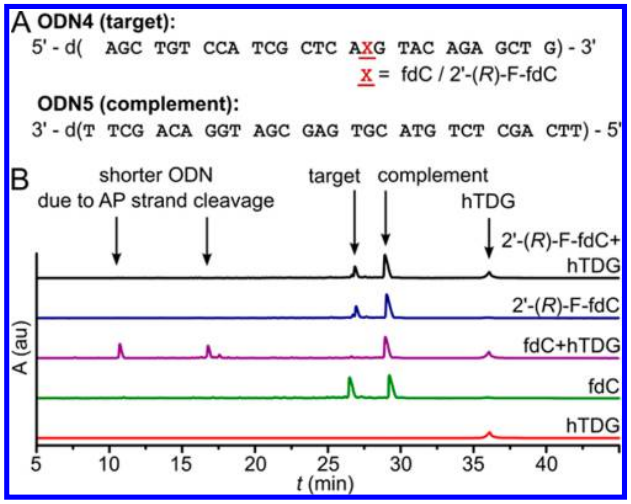


Figure 4. (A) Sequence of synthesized **ODN4** and **ODN5** for the hTDG glycosylation assay. (B) Reversed-phase HPL chromatogram of the hTDG glycosylation assay.

containing 2′-(*R*)-F-mdC, 2′-(*R*)-F-hmdC, 2′-(*R*)-F-fdC, and 2′-(*R*)-F-cadC in high yields and quality. Furthermore, we showed that a 2′-(*R*)-F label on fdC blocks the activity of the critical TDG enzyme, thus inhibiting base excision repair of this base. The 2′-(*R*)-F label is consequently the ideal tool for analysis of the epigenetic metabolism beyond base excision.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.orglett.6b02110.

Synthesis of all compounds and oligonucleotides and details of the assays (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: Thomas.Carell@lmu.de. Website: http://www.carellgroup.de.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Nadine Hinkel, Korbinian Krieger, Stefan Marchner, and Robert Rampmaier (all Ludwig-Maximilians-Universität München) for practical assistance. Furthermore, we thank the Deutsche Forschungsgemeinschaft (SFB749), the Fonds der Chemischen Industrie (predoctoral fellowship for A.S.S.), and the Excellence Cluster (CiPS^{M}, EXC114 and GRK 2062) for support.

■ REFERENCES

(1) (a) Gillis, E. P.; Eastman, K. J.; Hill, M. D.; Donnelly, D. J.; Meanwell, N. A. *J. Med. Chem.* **2015**, *58*, 8315−8359. (b) Purser, S.; Moore, P. R.; Swallow, S.; Gouverneur, V. *Chem. Soc. Rev.* **2008**, *37*, 320−330.

(2) Parker, W. B. *Chem. Rev.* **2009**, *109*, 2880−2893.

(3) (a) Sinnott, M. L. *Chem. Rev.* **1990**, *90*, 1171−1202. (b) Williams, A. A.; Darwanto, A.; Theruvathu, J. A.; Burdzy, A.; Neidigh, J. W.; Sowers, L. C. *Biochemistry* **2009**, *48*, 11994−12004. (c) Berger, I.; Tereshko, V.; Ikeda, H.; Marquez, V. E.; Egli, M. *Nucleic Acids Res.* **1998**, *26*, 2473−2480.

(4) (a) Dai, Q.; Lu, X.; Zhang, L.; He, C. *Tetrahedron* **2012**, *68*, 5145−5151. (b) Schaerer, O. D.; Verdine, G. L. *J. Am. Chem. Soc.* **1995**, *117*, 10781−10782. (c) Schärer, O. D.; Kawate, T.; Gallinari, P.; Jiricny, J.; Verdine, G. L. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 4878−4883.

(5) (a) Kriaucionis, S.; Heintz, N. *Science* **2009**, *324*, 929−930. (b) Tahiliani, M.; Koh, K. P.; Shen, Y.; Pastor, W. A.; Bandukwala, H.; Brudno, Y.; Agarwal, S.; Iyer, L. M.; Liu, D. R.; Aravind, L.; Rao, A. *Science* **2009**, *324*, 930−935. (c) Pfaffeneder, T.; Hackner, B.; Truß, M.; Münzel, M.; Müller, M.; Deiml, C. A.; Hagemeier, C.; Carell, T. *Angew. Chem., Int. Ed.* **2011**, *50*, 7008−7012. (d) 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.

(6) (a) Loenarz, C.; Schofield, C. J. *Chem. Biol.* **2009**, *16*, 580−583. (b) Ito, S.; Shen, L.; Dai, Q.; Wu, S. C.; Collins, L. B.; Swenberg, J. A.; He, C.; Zhang, Y. *Science* **2011**, *333*, 1300−1303.

(7) (a) Maiti, A.; Drohat, A. C. *J. Biol. Chem.* **2011**, *286*, 35334−35338. (b) Raiber, E.-A.; Beraldi, D.; Ficz, G.; Burgess, H. E.; Branco, M. R.; Murat, P.; Oxley, D.; Booth, M. J.; Reik, W.; Balasubramanian, S. *Genome Biol.* **2012**, *13*, R69. (c) Zhang, L.; Lu, X.; Lu, J.; Liang, H.; Dai, Q.; Xu, G.-L.; Luo, C.; Jiang, H.; He, C. *Nat. Chem. Biol.* **2012**, *8*, 328−330.

(8) (a) Schiesser, S.; Hackner, B.; Pfaffeneder, T.; Müller, M.; Hagemeier, C.; Truss, M.; Carell, T. *Angew. Chem., Int. Ed.* **2012**, *51*, 6516−6520. (b) Liutkevičiūtė, Z.; Kriukienė, E.; Ličytė, J.; Rudytė, M.; Urbanavičiūtė, G.; Klimašauskas, S. *J. Am. Chem. Soc.* **2014**, *136*, 5884−5887. (c) 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) Lee, S.; Bowman, B. R.; Ueno, Y.; Wang, S.; Verdine, G. L. *J. Am. Chem. Soc.* **2008**, *130*, 11570−11571.

(10) (a) Richardson, F. C.; Tennant, B. C.; Meyer, D. J.; Richardson, K. A.; Mann, P. C.; Mcginty, G. R.; Wolf, J. L.; Zack, P. M.; Bendele, R. A. *Toxicol. Pathol.* **1999**, *27*, 607−617. (b) Kierdaszuk, B.; Krawiec, K.; Kazimierczuk, Z.; Jacobsson, U.; Johansson, N. G.; Munch-petersen, B.; Eriksson, S.; Shugar, D. *Nucleosides Nucleotides* **1999**, *18*, 1883−1903.

(11) Münzel, M.; Globisch, D.; Trindler, C.; Carell, T. *Org. Lett.* **2010**, *12*, 5671−5673.

(12) Hirota, K.; Kitade, Y.; Kanbe, Y.; Isobe, Y.; Maki, Y. *Synthesis* **1993**, *1993*, 213−215.

(13) Bannwarth, W.; Trzeciak, A. *Helv. Chim. Acta* **1987**, *70*, 175−186.

(14) Luche, J. L. *J. Am. Chem. Soc.* **1978**, *100*, 2226−2227.

(15) Sammet, B. *Synlett* **2009**, *2009*, 3050−3051.

(16) Schröder, A. S.; Steinbacher, J.; Steigenberger, B.; Gnerlich, F. A.; Schiesser, S.; Pfaffeneder, T.; Carell, T. *Angew. Chem., Int. Ed.* **2014**, *53*, 315−318.

(17) (a) Münzel, M.; Lischke, U.; Stathis, D.; Pfaffeneder, T.; Gnerlich, F. A.; Deiml, C. A.; Koch, S. C.; Karaghiosoff, K.; Carell, T. *Chem. - Eur. J.* **2011**, *17*, 13782−13788. (b) Nomura, Y.; Haginoya, N.; Ueno, Y.; Matsuda, A. *Bioorg. Med. Chem. Lett.* **1996**, *6*, 2811−2816.

(18) (a) Beaucage, S. L.; Caruthers, M. H. *Tetrahedron Lett.* **1981**, *22*, 1859−1862. (b) Sinha, N. D.; Biernat, J.; McManus, J.; Köster, H. *Nucleic Acids Res.* **1984**, *12*, 4539−4557.

(19) Hashimoto, H.; Hong, S.; Bhagwat, A. S.; Zhang, X.; Cheng, X. *Nucleic Acids Res.* **2012**, *40*, 10203−10214.

(20) Bennett, M. T.; Rodgers, M. T.; Hebert, A. S.; Ruslander, L. E.; Eisele, L.; Drohat, A. C. *J. Am. Chem. Soc.* **2006**, *128*, 12510−12519.

DOI: 10.1021/acs.orglett.6b02110
Org. Lett. 2016, 18, 4368−4371

5 2′-(R)-Fluorinated mC, hmC, fC and caC triphosphates are substrates for DNA polymerases and TET-enzymes

In order to gain a more comprehensive insight in the metabolism and function of novel epigenetic DNA bases, test molecules are necessary to dissect various pathways. These molecules should be non-toxic and bioisostere. In the present study 2′-fluorinated triphosphates were synthesized and incorporated via polymerase chain reaction into long double stranded oligonucleotides. It could be demonstrated that these 2′-fluorinated derivatives are good substrates for DNA polymerases and TET enzymes.
For this study I supported the synthesis and purification of the PCR products. Further, I developed and optimized the expression and purification protocol for TET enzymes out of HEK cells. In addition, I established the *in vitro* TET assay and performed the sample preparation for the subsequent LC-MS/MS analysis.

ChemComm

COMMUNICATION

View Article Online
View Journal | View Issue

Cite this: *Chem. Commun.*, 2016, 52, 14361

Received 14th September 2016,
Accepted 25th November 2016

DOI: 10.1039/c6cc07517g

www.rsc.org/chemcomm

2′-(*R*)-Fluorinated mC, hmC, fC and caC triphosphates are substrates for DNA polymerases and TET-enzymes†

A. S. Schröder,‡ E. Parsa,‡ K. Iwan, F. R. Traube, M. Wallner, S. Serdjukow and T. Carell*

A deeper investigation of the chemistry that occurs on the newly discovered epigenetic DNA bases 5-hydroxymethyl-(hmdC), 5-formyl-(fdC), and 5-carboxy-deoxycytidine (cadC) requires chemical tool compounds, which are able to dissect the different potential reaction pathways in cells. Here we report that the 2′-(*R*)-fluorinated derivatives F-hmdC, F-fdC, and F-cadC, which are resistant to removal by base excision repair, are good substrates for DNA polymerases and TET enzymes. This result shows that the fluorinated compounds are ideal tool substances to investigate potential C–C-bond cleaving reactions in the context of active demethylation.

While all cells of a multicellular organism have an identical DNA sequence, their morphology and function differ to a great extent (*i.e.* neurons *vs.* adenocytes). This is possible because these cells have different sets of genes in active and passive states.[1] This programming of the individual genomes occurs on the DNA level by the chemical modification of deoxycytidine (dC, **1**), which is potentially methylated to give 5-methyldeoxycytidine (mdC, **2**). In the years 2009[2,3] and 2011[4,5] it was shown that mdC is additionally oxidized by the action of TET-enzymes[6] to give 5-hydroxymethyl-(hmdC, **3**), 5-formyl-(fdC, **4**), and 5-carboxy-deoxycytidine (cadC, **5**). Although the exact function of these oxidized deoxycytidine derivatives is not clear, it is believed that they play a fundamental role during the epigenetic programming of the genome that leads to activated and passivated genes.[7] Although it is unclear how this is mechanistically achieved we have learned that fdC and cadC are substrates of the base excision repair (BER) process.[8–10] The glycosylase TDG is able to recognize fdC and cadC (but not hmdC) upon which cleavage of the glycosidic bond occurs that transforms fdC and cadC into abasic sites (**6**), which are repaired *via* the insertion of "fresh" dC to give finally a demethylated position as shown in Fig. 1.

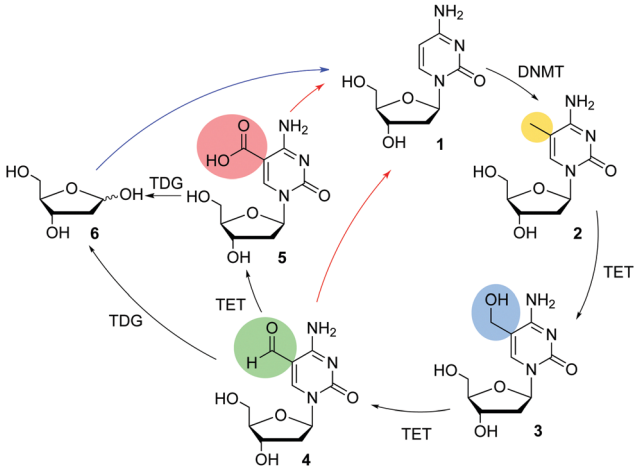


Fig. 1 Putative active demethylation pathways. DNMT: DNA methyltransferase, TET: ten-eleven-translocation methylcytosine dioxygenase, TDG: thymine DNA glycosylase. Red arrows: deformylation/decarboxylation, blue arrow: BER based active demethylation.

From a chemical point of view a direct deformylation of fdC or decarboxylation of cadC are attractive alternative pathways (C–C bond cleavage pathway, red arrows in Fig. 1) that would allow the direct conversion of fdC and cadC into dC, without the need to create potentially harmful abasic sites.[11–13] These sites are especially dangerous when they are generated on both strands in the duplex because then harmful double strand breaks are generated. In order to study the potential C–C bond cleavage pathway we need tool substances that are TET substrates but resist repair by BER. Recently we reported that 2′-fluorinated versions of fdC and cadC are BER resistant.[14] Here, we report that the triphosphates of F-mdC (**7**), F-hmdC (**8**), F-fdC (**9**) and F-cadC (**10**), in which the 2′-center is (*R*)-configured, are good substrates for various DNA polymerases and that this property can be used to generate long DNA strands containing multiple 2′-(*R*)-fluorinated mdCs, hmdCs, fdC, and cadCs using only slightly adapted PCR protocols. We finally show that the 2′-fluorinated compounds are also good substrates for the

Center for Integrated Protein Science, Department of Chemistry, Ludwig-Maximilians-Universität München, Butenandtstraße 5-13, 81377 Munich, Germany. E-mail: Thomas.carell@lmu.de; Web: www.carellgroup.de

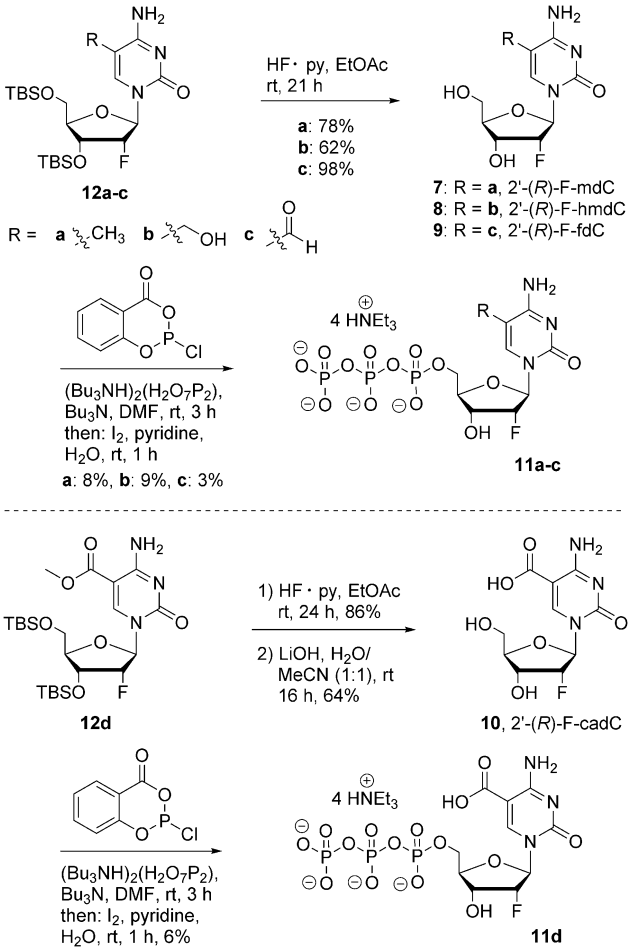
† Electronic supplementary information (ESI) available. See DOI: 10.1039/c6cc07517g

‡ These authors contributed equally.

TET enzymes and this together makes them ideal tool substances to study active demethylation *via* the putative C–C bond cleavage.

For the synthesis of the 2′-(*R*)-F-xdC triphosphates **11a–d** (see Scheme 1), we started with the corresponding, silyl protected 2′-(*R*)-F-xdC nucleosides **12a–d**.[14] Silyl deprotection of **11a–c** with Olah's reagent furnished the 2′-(*R*)-fluorinated nucleosides of mdC (**7**), hmdC (**8**) and fdC (**9**) in good yields between 62% and 98%. Regarding the nucleoside 2′-(*R*)-F-cadC (**10**), we first deprotected the silyl groups prior to saponification with LiOH in H_2O/MeCN (1 : 1). Using Eckstein conditions, it was possible to convert the nucleosides **7–10** into the corresponding 2′-(*R*)-fluorinated triphosphate mdCTP (**11a**), hmdCTP (**11b**), fdCTP (**11c**) and cadCTP (**11d**) in one-pot-reactions.[15–17] After extensive purification by preparative HPL chromatography, we obtained yields in the range of 3–9%. These yields are remarkable, particularly in light of the fact that no protecting groups were used, *e.g.* on the benzylic hydroxyl group of 2′-(*R*)-F-hmdC (**8**).

We next examined if the triphosphates are able to function as substrates for DNA polymerases. For this purpose, we started with primer extension studies to screen for suitable polymerases (see ESI†).[18,19] Remarkably, all triphosphates with the exception of 2′-(*R*)-F-cadCTP (**11d**) were accepted by the DNA-polymerase *Phusion*. This is surprising, because *Phusion* exhibits robust proof-reading activity and as such may hinder incorporation of unnatural triphosphates. The here examined triphosphates have not only a chemical group on the base but feature in addition a fluorine atom at C2′. The result suggests that the 2′-(*R*)-F derivatives mimic the natural situation quite closely.[17] The fluorine atom is bioisosteric to a H-atom[20] and it seems that its (*R*)-configuration, which most likely leads to a C3′-*endo* conformation of the sugar pucker does not hinder DNA polymerase based incorporation.[21–24] The incorporation of 2′-(*R*)-F-hmdCTP (**11b**) and 2′-(*R*)-F-fdCTP (**11c**) was furthermore possible with the polymerase *KOD-XL*, and here the yield were slightly higher. Incorporation of 2′-(*R*)-F-cadCTP (**11d**) was finally achieved with the polymerase *Therminator*. We believe that the problems associated with this base are caused by the carboxylic acid at the 5-position of the base, which carries a negative charge.[17]



Scheme 1 Synthesis of 2′-(*R*)-F-substituted triphosphates of mdC (**11a**; 2′-(*R*)-F-mdCTP), hmdC (**11b**; 2′-(*R*)-F-hmdCTP), fdC (**11c**; 2′-(*R*)-F-fdCTP) and cadC (**11d**; 2′-(*R*)-F-cadCTP).

With the 2′-(*R*)-F-xdCTP (**11a–d**) and the knowledge of which polymerases to use at hand, we next searched for appropriate conditions for the PCR (see Fig. 2). As a template, we selected a fragment of the OCT4 promoter sequence (see Fig. 2A). This sequence is known for its high density of epigenetically relevant nucleosides and, due to the high amount of CpG units it is usually considered to be a difficult template for PCR. The primers for the PCR were designed to yield an 81 base pair product containing 14–15 2′-(*R*)-F-xdC bases depending on the primer. 4–5 dCs are present in the primer and these are of course not exchanged during PCR. For the reaction, we fully replaced the dC triphosphate by the corresponding 2′-(*R*)-F-xdCTP. Hence, full length PCR products can only be formed if the appropriate 2′-(*R*)-F-xdCTP is accepted and incorporated by the polymerase. Forward and reverse primer were annealed to the template at 55 °C. The elongation of the primers was best performed at 72 °C for only 25 seconds. For complete extension of the primer, we extended the final elongation time to 5 min. The experimental results of the PCR are shown in Fig. 2B. For gel electrophoresis analysis, we had to use a tris-borate buffer system instead of tris-acetate to avoid "smearing" of the 2′-(*R*)-F-cadC product caused by the additional carboxyl groups present at this base.

Remarkably, we obtained full length PCR products for all 2′-(*R*)-F-xdCTPs showing again that a 2′-(*R*)-F-substitution hardly affects the procession of the polymerase. Further proof of the correct incorporation of 2′-(*R*)-F-xdCTP was gained by LC-MS/MS analysis. To this end, the PCR products were fully digested to the nucleoside level. The sugar phosphate backbone was first cleaved with nuclease S1 and *Antarctic phosphatase*, resulting in the oligo- and 5′-monophosphates as well as nucleosides. Further hydrolysis down to the nucleoside level was realized with snake venom phosphodiesterase I.[25–28] The resulting nucleoside mixture was subsequently analyzed by UHPLC-MS/MS (QQQ) (see Fig. 2C and D). As an example, the UV-trace of the digested 2′-(*R*)-F-fdC PCR product is shown. The clean chromatogram shows only the expected nucleosides dA, dT, dG, dC (from the primer) and the 2′-(*R*)-F-fdC compound. The quantification of the 2′-(*R*)-F-nucleosides using exact calibration curves of the synthesized nucleosides confirms the incorporation. Most importantly,

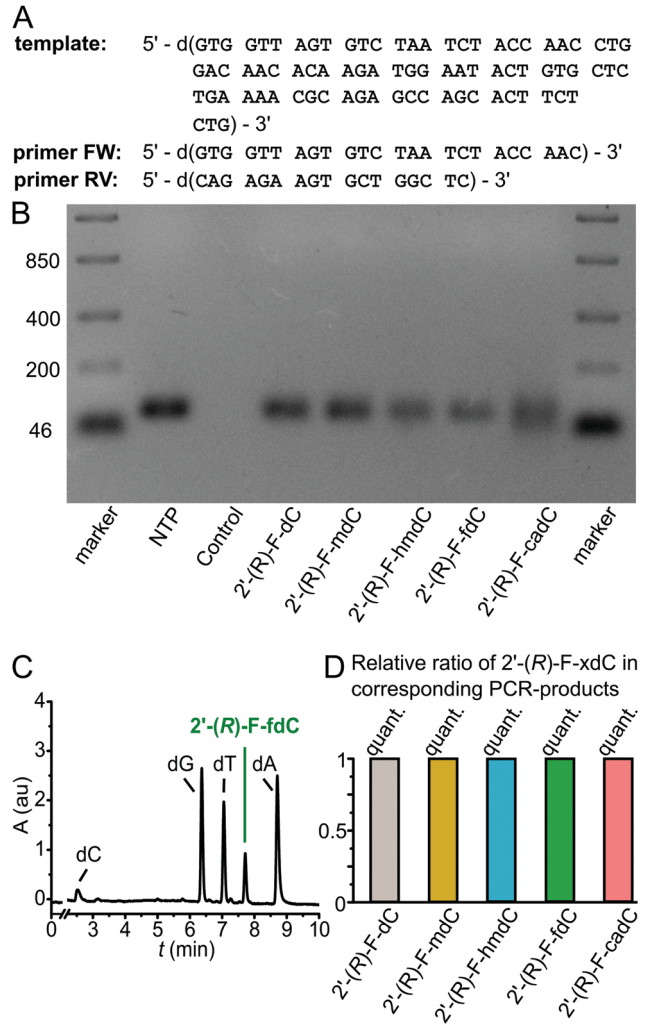


Fig. 2 Synthesis of 2′-(*R*)-F-xdC containing OCT4 promoter fragments by PCR. (A) Sequence of the template and primers used for the PCR (FW: forward, RV: reverse). (B) Result of the PCR analyzed by gel electrophoresis. The retention of the band corresponds to the expected PCR products. (Control: PCR with all NTP, but without a polymerase. 2′-(*R*)-F-xdC: PCR with all NTP except dCTP but with corresponding 2′-(*R*)-F-xdC.) (C) UV trace of UHPL-chromatogram after enzymatic digestion of the 2′-(*R*)-F-fdC PCR-product. (D) Quantification data of 2′-(*R*)-F-xdC PCR-products after enzymatic digestion and UHPLC-MS/MS analysis.

we did not detect any side products during the PCR reaction. Overall, the data show, that the 2′-(*R*)-F-xdCTPs can be incorporated despite the 2′-(*R*)-fluoro label *via* PCR into long oligonucleotides.

We next turned our attention to the question, if 2′-(*R*)-F-xdCs can be oxidized by TET enzymes despite the presence of the 2′-(*R*)-fluoro substituent. To this end, we overexpressed fused constructs of GFP and the catalytic domain of TET1 (TET1cd) in HEK293T cells and isolated functional TET1cd with the help of agarose beads coated with anti-GFP-antibodies. To obtain sufficient amounts of the enzyme we treated the cells with sodium butyrate. This compound acts as an HDAC inhibitor which leads to increased protein expression.[29] High concentrations of benzonase and rigorous washing yielded pure and functional TET1cd. This was confirmed by incubation with an ODN containing mdC in which mdC was efficiently oxidized (see ESI†).

We then added the 81 basepair long OCT4 promoter fragment containing 14–15 2′-(*R*)-F-mdCs, depending on forward or reverse strand, respectively. After incubating the PCR-product with TET1cd in a reducing buffer (see ESI†) for 3 h at 37 °C, we isolated the fragment, digested it down to the nucleoside level as described above and analyzed the obtained nucleoside mixture using UHPLC-MS/MS (QQQ). As shown in Fig. 3B the MS-trace showed clearly that the 2′-(*R*)-F-mdC starting material was not only oxidized to 2′-(*R*)-F-hmdC but also to the higher oxidized species 2′-(*R*)-F-fdC and 2′-(*R*)-F-cadC. Exact quantification data show that TET1cd oxidation produced 7.9% 2′-(*R*)-F-hmdC, 3.3% 2′-(*R*)-F-fdC and 0.2% 2′-(*R*)-F-cadC. Again, no side products like deamination to 2′-(*R*)-F-dT or background C–C bond cleavage, which would provide 2′-(*R*)-F-dC, were detected. In order to study if the TET protein oxidizes the 2′-F-mdC with an efficiency comparable to the non-fluorinated mdC, we added the same 81 basepair long OCT4 promoter fragment containing mdC instead of 2′-(*R*)-F-mdC to pure and functional TET1cd. After digestion and LC-MS/MS analysis, we detected now only fdC and cadC but no hmdC showing that the non-fluorinated mdC is the slightly better substrate (see ESI,† Table S2). However, we see for mdC and 2′-F-mdC oxidation up to the fdC and cadC level. This shows that the 2′-(*R*)-fluoro analogs can report TET activity, although the F-atom does reduce the TET activity to some extent.

In summary, here we show that the 2′-(*R*)-F triphosphates of the epigenetically relevant nucleosides mdC, hmdC, fdC and cadC can be efficiently incorporated into long oligonucleotides

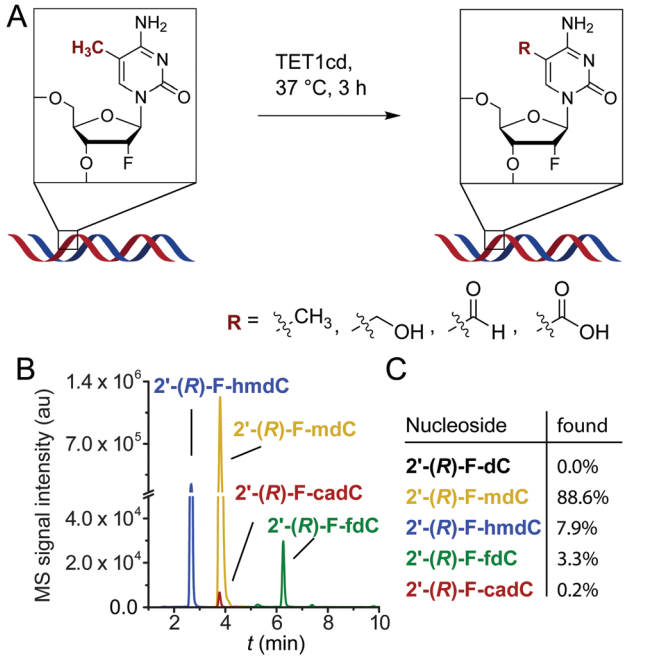


Fig. 3 TET1cd oxidation assay. (A) Scheme of the TET1cd dependent oxidation reaction. (B) UHPL-MS/MS-trace of 2′-(*R*)-F-xdC nucleosides after enzymatic digestion of the DNA fragment. (C) Quantification data of the product after complete enzymatic digestion and UHPLC-MS/MS (QQQ) analysis.

using PCR. This provides oligonucleotides of sufficient length for detailed mechanistic studies. Importantly, while the 2′-(*R*)-F substituent blocks the BER-based removal of fdC and cadC it is a good substrate for the TET enzymes.

We thank Jessica Furtmeier for practical assistance (Ludwig-Maximilians-Universität München). Furthermore, we thank the Deutsche Forschungsgemeinschaft (SFB749, SFB1032, CA275-8/5) for financial support. Additional support was obtained from the Fonds der Chemischen Industrie (predoctoral fellowship for A. S. S.), the Boehringer Ingelheim Fonds (predoctoral fellowship for F. R. T.), the Excellence Cluster (CiPS^{M}, EXC114), the DFG GRK2062 and the Marie Curie Training and Mobility network (Clickgene) is acknowledged.

Notes and references

1 A. Avgustinova and S. A. Benitah, *Nat. Rev. Mol. Cell Biol.*, 2016, **17**, 643–658.
2 M. Tahiliani, K. P. Koh, Y. Shen, W. A. Pastor, H. Bandukwala, Y. Brudno, S. Agarwal, L. M. Iyer, D. R. Liu, L. Aravind and A. Rao, *Science*, 2009, **324**, 930–935.
3 S. Kriaucionis and N. Heintz, *Science*, 2009, **324**, 929–930.
4 T. Pfaffeneder, B. Hackner, M. Truß, M. Münzel, M. Müller, C. A. Deiml, C. Hagemeier and T. Carell, *Angew. Chem., Int. Ed.*, 2011, **50**, 7008–7012.
5 S. Ito, L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He and Y. Zhang, *Science*, 2011, **333**, 1300–1303.
6 C. Loenarz and C. J. Schofield, *Chem. Biol.*, 2009, **16**, 580–583.
7 A. Perera, D. Eisen, M. Wagner, S. K. Laube, A. F. Künzel, S. Koch, J. Steinbacher, E. Schulze, V. Splith, N. Mittermeier, M. Müller, M. Biel, T. Carell and S. Michalakis, *Cell Rep.*, 2015, **11**, 283–294.
8 A. Maiti and A. C. Drohat, *J. Biol. Chem.*, 2011, **286**, 35334–35338.
9 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 and G.-L. Xu, *Science*, 2011, **333**, 1303–1307.
10 E.-A. Raiber, D. Beraldi, G. Ficz, H. E. Burgess, M. R. Branco, P. Murat, D. Oxley, M. J. Booth, W. Reik and S. Balasubramanian, *Genome Biol.*, 2012, **13**, 1–11.
11 S. Schiesser, T. Pfaffeneder, K. Sadeghian, B. Hackner, B. Steigenberger, A. S. Schröder, J. Steinbacher, G. Kashiwazaki, G. Höfner, K. T. Wanner, C. Ochsenfeld and T. Carell, *J. Am. Chem. Soc.*, 2013, **135**, 14593–14599.
12 Z. Liutkevičiūtė, E. Kriukienė, J. Ličytė, M. Rudytė, G. Urbanavičiūtė and S. Klimašauskas, *J. Am. Chem. Soc.*, 2014, **136**, 5884–5887.
13 S. Schiesser, B. Hackner, T. Pfaffeneder, M. Müller, C. Hagemeier, M. Truss and T. Carell, *Angew. Chem., Int. Ed.*, 2012, **51**, 6516–6520.
14 A. S. Schröder, O. Kotljarova, E. Parsa, K. Iwan, N. Raddaoui and T. Carell, *Org. Lett.*, 2016, **18**, 4368–4371.
15 J. Ludwig and F. Eckstein, *J. Org. Chem.*, 1989, **54**, 631–635.
16 J. Caton-Williams, M. Smith, N. Carrasco and Z. Huang, *Org. Lett.*, 2011, **13**, 4156–4159.
17 B. Steigenberger, S. Schiesser, B. Hackner, C. Brandmayr, S. K. Laube, J. Steinbacher, T. Pfaffeneder and T. Carell, *Org. Lett.*, 2013, **15**, 366–369.
18 T. Ono, M. Scalf and L. M. Smith, *Nucleic Acids Res.*, 1997, **25**, 4581–4588.
19 C. G. Peng and M. J. Damha, *Can. J. Chem.*, 2008, **86**, 881–891.
20 R. S. Rowland and R. Taylor, *J. Phys. Chem.*, 1996, **100**, 7384–7391.
21 M. L. Sinnott, *Chem. Rev.*, 1990, **90**, 1171–1202.
22 A. A. Williams, A. Darwanto, J. A. Theruvathu, A. Burdzy, J. W. Neidigh and L. C. Sowers, *Biochemistry*, 2009, **48**, 11994–12004.
23 I. Berger, V. Tereshko, H. Ikeda, V. E. Marquez and M. Egli, *Nucleic Acids Res.*, 1998, **26**, 2473–2480.
24 S. Lee, B. R. Bowman, Y. Ueno, S. Wang and G. L. Verdine, *J. Am. Chem. Soc.*, 2008, **130**, 11570–11571.
25 T. Ando, *Biochim. Biophys. Acta, Nucleic Acids Protein Synth.*, 1966, **114**, 158–168.
26 M. Rina, C. Pozidis, K. Mavromatis, M. Tzanodaskalaki, M. Kokkinidis and V. Bouriotis, *Eur. J. Biochem.*, 2000, **267**, 1230–1238.
27 T. Nihei, G. L. Cantoni and With the technical assistance of Rachelle Rothenberg, *J. Biol. Chem.*, 1963, **238**, 3991–3998.
28 A. S. Schröder, J. Steinbacher, B. Steigenberger, F. A. Gnerlich, S. Schiesser, T. Pfaffeneder and T. Carell, *Angew. Chem., Int. Ed.*, 2014, **53**, 315–318.
29 M. G. Riggs, R. G. Whittaker, J. R. Neumann and V. M. Ingram, *Nature*, 1977, **268**, 462–464.

6 5-Formylcytosine to Cytosine Conversion by C-C Bond Cleavage *in vivo*

The methylation of cytosine is crucial for various cellular processes. It is an essential part of the epigenetic network. However, a comprehensive understanding of the demethylation of 5-methylcytosine is still missing. To date the best studied pathway for the demethylation is based on the oxidation of 5-methylcytosine (mC) to 5-hydroxymethylcytosine (hmC), 5-formylcytosine (fC) and 5-carboxylcytosine (caC). Through a base excision repair it is possible to replace fC and caC with unmodified cytosine. The present study investigates an alternative pathway that does not involve potential mutagenic base excision repair. By using 2′-fluorinated DNA building blocks it was possible to demonstrate that 5-formylcytosine is converted to cytosine by a direct C-C bond cleavage reaction.
For this study, I participated in processing the DNA for subsequent LC-MS/MS analysis.

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

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

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

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

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

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

RESULTS

Isotopically labeled fdC is directly converted into dC

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

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

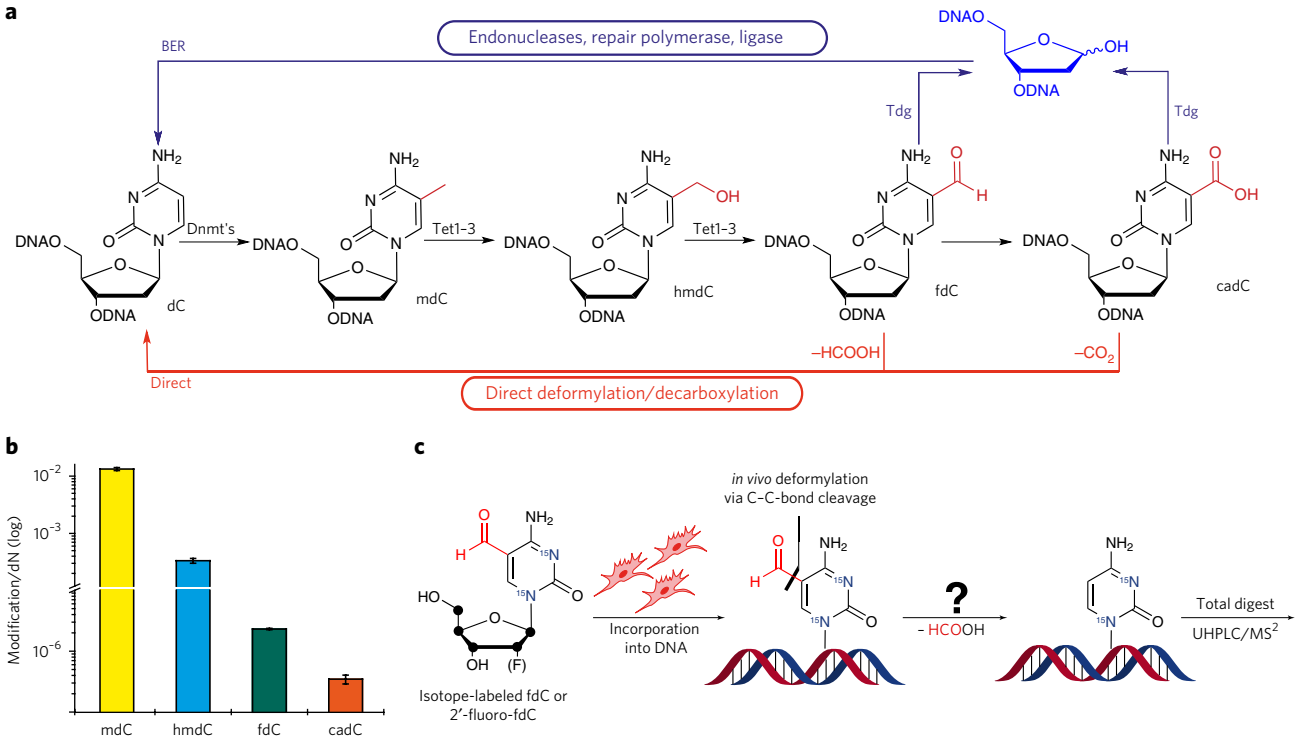


Figure 1 | Isotope tracing studies. (**a**) Suggested pathways of active demethylation. Thymine-DNA glycosylase (Tdg)-based cleavage of the glycosidic bond of 5-formyl-dC (fdC) and 5-carboxy-dC (cadC) results in an abasic site, which initiates a BER process that leads to the replacement of fdC and cadC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

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

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

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

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

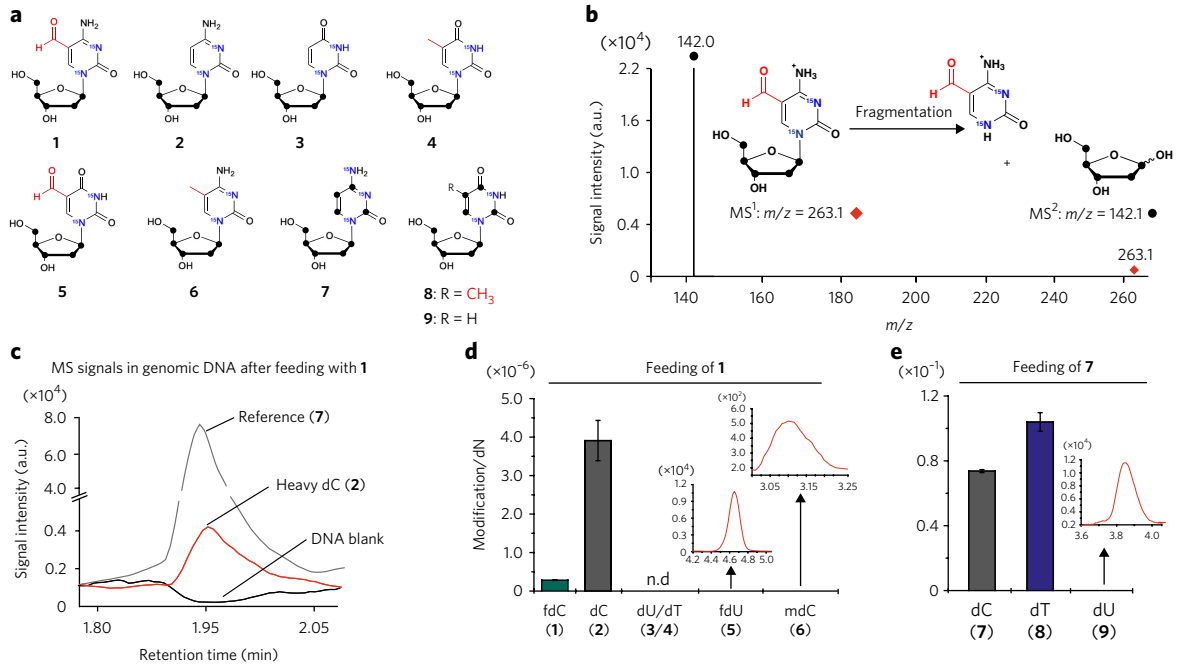


Figure 2 | Conversion of isotopically labeled fdC into dC in mESCs. (**a**) Overview of the compounds that may be detected after feeding of **1** or **7** to mESCs. (**b**) Feeding of **1** to mESCs results in incorporation of the isotopologue into the genomic DNA, as proven by its fingerprint MS transition.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

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

2′-fluorinated cytosines detect biochemical conversions

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

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

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

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

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

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

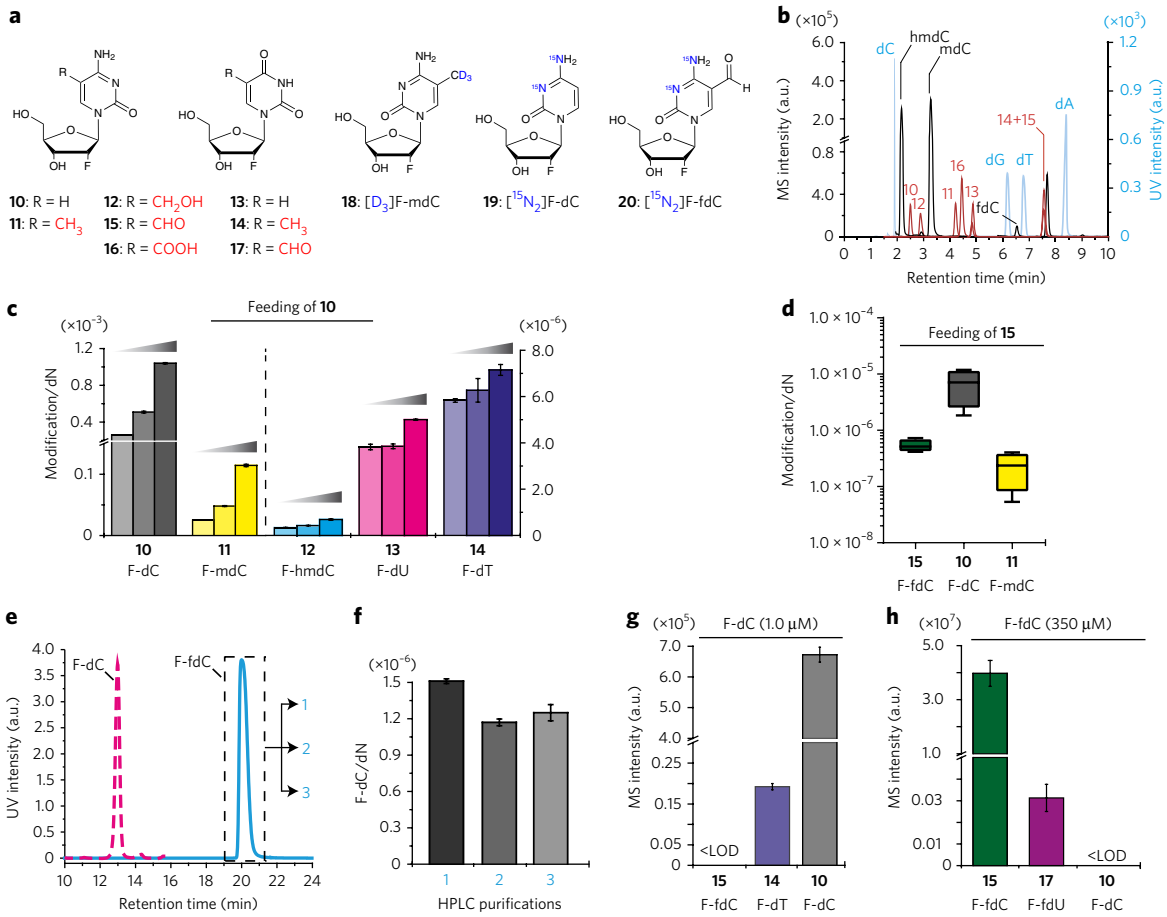


Figure 3 | F-fdC is converted into F-dC within the genome. (**a**) Chemical structures of 2′-fluorinated dC and dU derivatives that were investigated and internal standards used. (**b**) Resulting UV and MS traces of the nucleosides under investigation. Light blue, canonical bases; black, natural dC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

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

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

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

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

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

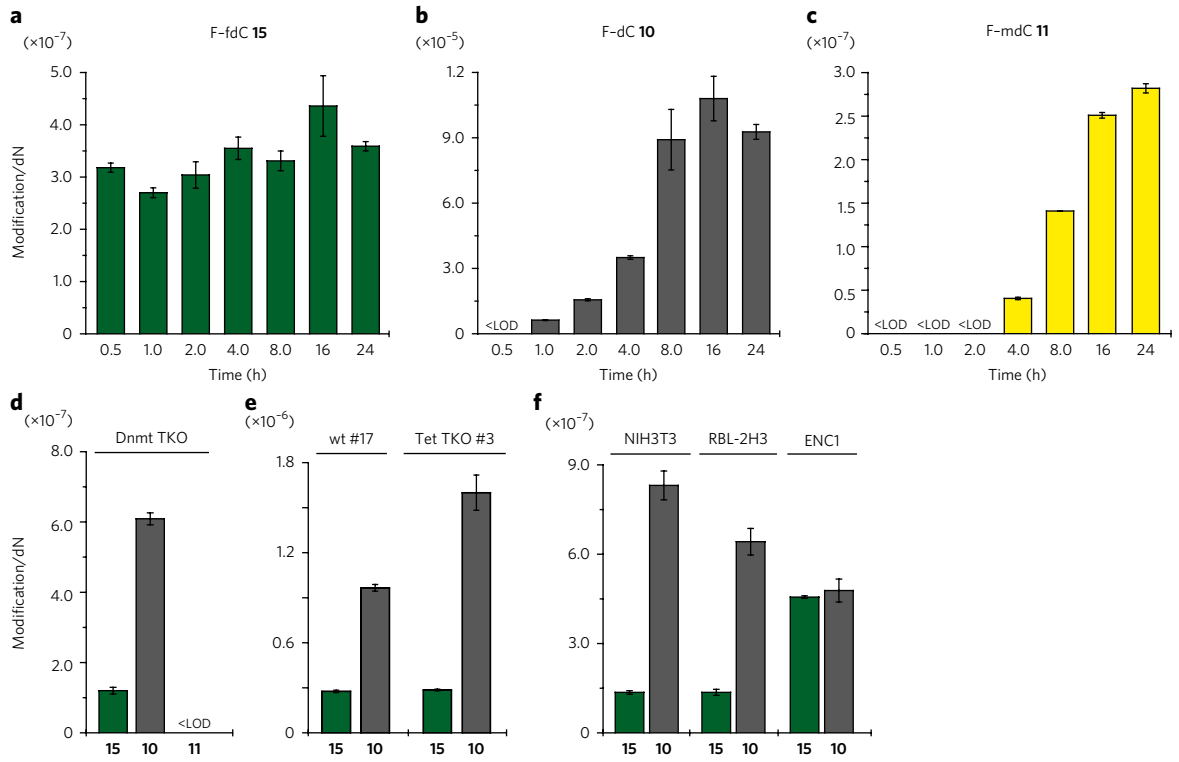


Figure 4 | Demodification of 2′-fluorinated fdC is a rapid process, does not require Dnmt or Tet enzymes and occurs also in somatic cell types. (**a**–**c**) Time-course study showing the genomic build-up of F-fdC **15** (**a**), F-dC **10** (**b**) and F-mdC **11** (**c**) upon metabolic labeling with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

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

Dnmt or Tet enzymes are not required for demodification

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

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

DISCUSSION

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

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

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

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

METHODS

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

References

1. Smith, Z.D. & Meissner, A. DNA methylation: roles in mammalian development.. *Nat. Rev. Genet* **14**, 204–220 (2013).
2. Schübeler, D. Function and information content of DNA methylation. *Nature* **517**, 321–326 (2015).
3. Jeltsch, A. & Jurkowska, R.Z. Allosteric control of mammalian DNA methyltransferases – a new regulatory paradigm *Nucleic Acids Res.* **44**, 8556–8575 (2016).
4. Tahiliani, M. *et al.* Conversion of 5-methylcytosine to -hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930–935 (2009).
5. Kriaucionis, S. & Heintz, N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **324**, 929–930 (2009).
6. Pfaffeneder, T. *et al.* The discovery of 5-formylcytosine in embryonic stem cell DNA. *Angew. Chem. Int. Ed. Engl.* **50**, 7008–7012 (2011).
7. Ito, S. *et al.* Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333**, 1300–1303 (2011).
8. He, Y.F. *et al.* Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333**, 1303–1307 (2011).
9. Globisch, D. *et al.* Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS One* **5**, e15367 (2010).
10. Münzel, M., Globisch, D. & Carell, T. 5-Hydroxymethylcytosine, the sixth base of the genome. *Angew. Chem. Int. Ed. Engl.* **50**, 6460–6468 (2011).
11. Pfaffeneder, T. *et al.* Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA. *Nat. Chem. Biol.* **10**, 574–581 (2014).
12. Wagner, M. *et al.* Age-dependent levels of 5-methyl-, 5-hydroxymethyl-, and 5-formylcytosine in human and mouse brain tissues. *Angew. Chem. Int. Edn Engl.* **54**, 12511–12514 (2015).
13. Branco, M.R., Ficz, G. & Reik, W. Uncovering the role of 5-hydroxymethylcytosine in the epigenome. *Nat. Rev. Genet.* **13**, 7–13 (2011).
14. Wu, H. & Zhang, Y. Mechanisms and functions of Tet protein-mediated 5-methylcytosine oxidation. *Genes Dev.* **25**, 2436–2452 (2011).
15. Bachman, M. *et al.* 5-Formylcytosine can be a stable DNA modification in mammals. *Nat. Chem. Biol.* **11**, 555–557 (2015).
16. Su, M. *et al.* 5-Formylcytosine could be a semipermanent base in specific genome sites. *Angew. Chem. Int. Ed. Engl.* **55**, 11797–11800 (2016).
17. Raiber, E.A. *et al.* 5-Formylcytosine alters the structure of the DNA double helix. *Nat. Struct. Mol. Biol.* **22**, 44–49 (2015).
18. Song, C.X. *et al.* Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. *Cell* **153**, 678–691 (2013).
19. Kellinger, M.W. *et al.* 5-formylcytosine and 5-carboxylcytosine reduce the rate and substrate specificity of RNA polymerase II transcription. *Nat. Struct. Mol. Biol.* **19**, 831–833 (2012).
20. Zhu, C. *et al.* Single-cell 5-formylcytosine landscapes of mammalian early embryos and ESCs at single-base resolution. *Cell Stem Cell* **20**, 720–731.e5 (2017).
21. Hill, P.W., Amouroux, R. & Hajkova, P. DNA demethylation, Tet proteins and 5-hydroxymethylcytosine in epigenetic reprogramming: an emerging complex story. *Genomics* **104**, 324–333 (2014).
22. Wu, X., Inoue, A., Suzuki, T. & Zhang, Y. Simultaneous mapping of active DNA demethylation and sister chromatid exchange in single cells. *Genes Dev.* **31**, 511–523 (2017).
23. Maiti, A. & Drohat, A.C. Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. *J. Biol. Chem.* **286**, 35334–35338 (2011).
24. Guo, F. *et al.* Active and passive demethylation of male and female pronuclear DNA in the mammalian zygote. *Cell Stem Cell* **15**, 447–459 (2014).
25. Wu, S.C. & Zhang, Y. Active DNA demethylation: many roads lead to Rome. *Nat. Rev. Mol. Cell Biol.* **11**, 607–620 (2010).
26. Schiesser, S. *et al.* Deamination, oxidation, and C-C bond cleavage reactivity of 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxycytosine. *J. Am. Chem. Soc.* **135**, 14593–14599 (2013).
27. Liutkevičiūtė, Z. *et al.* Direct decarboxylation of 5-carboxylcytosine by DNA C5-methyltransferases. *J. Am. Chem. Soc.* **136**, 5884–5887 (2014).
28. Schiesser, S. *et al.* Mechanism and stem-cell activity of 5-carboxycytosine decarboxylation determined by isotope tracing. *Angew. Chem. Int. Ed. Engl.* **51**, 6516–6520 (2012).
29. Jekunen, A. & Vilpo, J.A. 5-Methyl-2′-deoxycytidine. Metabolism and effects on cell lethality studied with human leukemic cells in vitro. *Mol. Pharmacol.* **25**, 431–435 (1984).
30. Vilpo, J.A. & Vilpo, L.M. Biochemical mechanisms by which reutilization of DNA 5-methylcytosine is prevented in human cells. *Mutat. Res.* **256**, 29–35 (1991).
31. Schröder, A.S. *et al.* Synthesis of (*R*)-configured 2′-fluorinated mC, hmC, fC, and caC phosphoramidites and oligonucleotides. *Org. Lett.* **18**, 4368–4371 (2016).
32. Schröder, A.S. *et al.* 2′-(*R*)-Fluorinated mC, hmC, fC and caC triphosphates are substrates for DNA polymerases and TET-enzymes. *Chem. Commun. (Camb.)* **52**, 14361–14364 (2016).
33. Blaschke, K. *et al.* Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells. *Nature* **500**, 222–226 (2013).

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

Acknowledgments

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

Author contributions

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

Competing financial interests

The authors declare no competing financial interests.

Additional information

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

ONLINE METHODS

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

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

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

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

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

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

DNA digestion. 0.5–10 µg of genomic DNA in 35 µL H_2O were digested as follows: An aqueous solution (7.5 µL) of 480 µM $ZnSO_4$, containing 42 U nuclease S1 (*Aspergillus oryzae*, Sigma-Aldrich), 5 U Antarctic phosphatase (New England BioLabs) and specific amounts of labeled internal standards were added, and the mixture was incubated at 37 °C for 3 h. After addition of 7.5 µL of a 520 µM $[Na]_2$-EDTA solution, containing 0.2 U snake venom phosphodiesterase I (*Crotalus adamanteus*, USB corporation), the sample was incubated for 3 h at 37 °C or overnight and then stored at −20 °C. Prior to UHPLC–MS/MS analysis, samples were filtered by using an AcroPrep Advance 96 filter plate 0.2 µm Supor (Pall Life Sciences).

LC/MS-MS analysis of DNA samples. Quantitative UHPLC–MS/MS analysis of digested DNA samples was performed using an Agilent 1290 UHPLC system equipped with a UV detector and an Agilent 6490 triple quadrupole mass spectrometer. Prior to every measurement series, external calibration curves were measured to quantify the levels of the F-nucleosides (**Supplementary Fig. 12**). Additionally, (*R*)-2′-[$^{15}N_2$]F-dC (**19**), (*R*)-2′-[$^{15}N_2$]F-hmdC (**37**) and (*R*)-2′-[$^{15}N_2$]F-fdC (**20**) were used to validate the resulting peaks by co-injection. For exact quantification of fluorinated nucleosides also internal quantification with stable isotope dilution techniques for F-fdC, F-dC and F-mdC were developed (**Supplementary Fig. 13**). Natural nucleosides were quantified with the stable isotope dilution technique. An improved method, based on earlier published work[26,28,52–54] was developed, which allowed the concurrent analysis of all nucleosides in one single analytical run[11]. The source-dependent parameters were as follows: gas temperature 80 °C, gas flow 15 L/min (N_2), nebulizer 30 psi, sheath gas heater 275 °C, sheath gas flow 11 L/min (N_2), capillary voltage 2,500 V in the positive ion mode, capillary voltage −2,250 V in the negative ion mode and nozzle voltage 500 V. The fragmentor voltage was 380 V/ 250 V. Delta EMV was set to 500 (positive mode) and 800 (negative mode). Compound-dependent parameters are summarized in **Supplementary Tables 1–4**. Chromatography was performed by a Poroshell 120 SB-C8 column (Agilent, 2.7 µm, 2.1 mm × 150 mm) at 35 °C using a gradient of water and MeCN, each containing 0.0085% (v/v) formic acid, at a flow rate of 0.35 mL/min: 0–4 min; 0– 3.5% (v/v) MeCN; 4–7.9 min; 3.5–5% MeCN; 7.9–8.2 min; 5–80% MeCN; 8.2–11.5 min; 80% MeCN; 11.5–12 min; 80–0% MeCN; 12–14 min; 0% MeCN. The effluent up to 1.5 min and after 12 min was diverted to waste by a Valco valve. The autosampler was cooled to 4 °C. The injection volume was 39 µL.

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

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

doi:10.1038/nchembio.2531

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

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

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

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

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

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

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

7 Functional impacts of 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxycytosine at a single hemi-modified CpG dinucleotide in a gene promoter

Enzymatic oxidation of mC to hmC, fC and caC is connected to active DNA demethylation and epigenetic reprogramming in mammals. However, it is yet not clear whether the mC oxidation products have autonomous epigenetic or regulatory functions in the genome. This study used an artificial upstream promoter constituted of one cAMP response element (CRE) to measure the impact of mC in a hemi-methylated CpG on the promoter activity and further explored the consequences of hmC, fC, and caC in the same system. All modifications induced mild impairment of the CREB transcription factor binding. The decrease of the gene expression by mC or hmC was proportional to the impairment of CREB binding and had a steady character over at least 48 hours. In contrast, promoters containing single fC or caC underwent further progressive loss of activity, up to almost complete repression. This decline was strongly (caC) or partly (fC) dependent on TDG. The results thus indicate that fC and caC can provide a signal for perpetuation and enhancement of the repressed transcriptional state by a mechanism that requires base excision repair.

For this study, I synthesized and purified the highly modified oligonucleotides containing mC, hmC, fC and caC. These were then used for the generation of the CRE element. In addition, I participated the synthesis and purification of the 2′-fluorinated oligonucleotides.

Nucleic Acids Research, 2017 **1**
doi: 10.1093/nar/gkx718

Functional impacts of 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxycytosine at a single hemi-modified CpG dinucleotide in a gene promoter

Nataliya Kitsera[1], Julia Allgayer[2], Edris Parsa[3], Nadine Geier[1], Martin Rossa[3], Thomas Carell[3] and Andriy Khobta[1,2,*]

[1]Institute of Toxicology, University Medical Center of the Johannes Gutenberg University Mainz, Mainz 55131, Germany, [2]Institute of Pharmacy and Biochemistry, Johannes Gutenberg University of Mainz, Mainz 55128, Germany and [3]Center for Integrated Protein Science at the Department of Chemistry, Ludwig-Maximilians-Universität München, Munich 81377, Germany

Received February 28, 2017; Revised July 21, 2017; Editorial Decision August 05, 2017; Accepted August 08, 2017

ABSTRACT

Enzymatic oxidation of 5-methylcytosine (5-mC) in the CpG dinucleotides to 5-hydroxymethylcytosine (5-hmC), 5-formylcytosine (5-fC) and 5-carboxycytosine (5-caC) has central role in the process of active DNA demethylation and epigenetic reprogramming in mammals. However, it is not known whether the 5-mC oxidation products have autonomous epigenetic or regulatory functions in the genome. We used an artificial upstream promoter constituted of one cAMP response element (CRE) to measure the impact of 5-mC in a hemi-methylated CpG on the promoter activity and further explored the consequences of 5-hmC, 5-fC, and 5-caC in the same system. All modifications induced mild impairment of the CREB transcription factor binding to the consensus 5′-TGACGTCA-3′ CRE sequence. The decrease of the gene expression by 5-mC or 5-hmC was proportional to the impairment of CREB binding and had a steady character over at least 48 h. In contrast, promoters containing single 5-fC or 5-caC underwent further progressive loss of activity, up to an almost complete repression. This decline was dependent on the thymine-DNA glycosylase (TDG). The results thus indicate that 5-fC and 5-caC can provide a signal for perpetuation and enhancement of the repressed transcriptional state by a mechanism that requires base excision repair.

INTRODUCTION

DNA methylation at cytosine residues plays essential roles in transcriptional regulation and genome maintenance in vertebrates (1–3). Most abundant and best studied context in which 5-methylcytosine is present in the genomes is symmetric methylation of CpG dinucleotides. CpG methylation can be copied to daughter DNA strands during replication and thus constitutes an epigenetic mark propagated through cell division (4,5). CpG methylation in the promoter regions regulates transcriptional activity of genes by at least two mechanisms. One is direct modulation of binding of the methylation-sensitive transcription factors to the regulatory motifs containing methylated CGs (6–8). Another is reinforcement and long-term propagation of chromatin-mediated gene silencing by recruitment of the methyl-CpG-binding domain (MBD) proteins and, in turn, histone deacetylases (1,2). Both these modes of action appear to be important for silencing and transcriptional repression of genes whose promoters contain long-range arrays rich in CpG dinucleotides, known as CpG islands (CGIs). In the non-CGI promoters, functions of CpG methylation are even more complex and the outcomes for transcriptional activity may be very variable for genes controlled by different transcription factors and in different cell lineages (9–11). DNA demethylation mechanisms add another complexity level to the regulation of gene transcription by erasing 5-methylcytosine or converting it to other types of base modifications whose regulatory functions still have to be explored (12–15). Thus, currently the best characterized DNA demethylation pathway in vertebrates mediated by the ten eleven translocation (TET) family oxygenases (16–21) generates 5-hmC, 5-fC and 5-caC, which may have independent functions in the regulation of gene expression (13,22). However, understanding of functional consequences of cytosine demethylation products at specific regulatory elements is hindered by the extremely low levels of these modifications in the genome (17,19,23) and by the dynamic character of 5-fC and 5-caC which are efficiently removed by DNA repair mechanisms (15,18,24).

*To whom correspondence should be addressed. Tel: +49 6131 17 9271; Fax: +49 6131 17 8499; Email: khobta@uni-mainz.de

Transitions between the methylated and unmethylated states of CpG dinucleotides require hemi-modified states in which 5-mC or its oxidation products are present in only one DNA strand (25). Functional impact of such modifications on the regulation of gene expression is unknown because deliberate hemi-modification of chosen cytosine residues in DNA cannot be generated enzymatically. To overcome this hurdle, we have previously proposed an efficient procedure for a site-specific incorporation of synthetic oligonucleotides containing various base modifications into plasmid-borne reporter genes with the help of sequence-specific nicking endonucleases (26). With these tools, we have already obtained important insights into the mechanisms of excision repair of several types of structurally defined DNA lesions in functional reporter genes (27–29). Here, we applied an analogous approach to measure the functional impacts of 5-mC, 5-hmC, 5-fC and 5-caC on the regulation of gene transcription. Our strategy was to generate expression constructs containing each of these cytosine modifications in a CpG dinucleotide within an artificial promoter composed of a cyclic adenosine monophosphate (cAMP) response element (CRE) upstream from a TATA box-containing RNA polymerase II basal promoter. We chose CRE because it contains a CpG dinucleotide in the consensus sequence and because it was known that binding of the specific transcription factors (e.g., CREB and C/EBP alpha) and the induced transcription can be either negatively or positively regulated by the symmetric CpG methylation (8,30). Moreover the presence of 5-mC and 5-hmC in just one DNA strand significantly reduce, whereas the C>U substitution in either one or both DNA strands increase the affinity of CREB binding (31,32). The aim of present work was to compare the functional significance of CpG hemi-methylation and hemi-hydroxymethylation for the CRE-driven gene expression and to investigate the impacts of the DNA demethylation intermediates 5-fC and 5-caC on the promoter activity.

MATERIALS AND METHODS

Cell lines and reporter vectors

Isogenic HeLa-derived cell lines in which TDG, SMUG1 or UNG1/2 DNA glycosylases were knocked down by stable expression of the specific shRNA and the corresponding control HeLa/pEpS cell line were generated and characterised by western blotting and the specific DNA glycosylase activity testing previously (29). The sequence cloned into the pENTR/pSuper+ vector (Addgene, Cambridge, MA, USA) for expression of shRNA which efficiently targets human TDG was 5′ GATCCGGGAACGAAATATGGACGTTCAACTCG AGTTGAACGTCCATATTTCGTTCTTTTTGGAAA. The parental vector pENTR/pSuper+ was used as a scrambled transcript control. Vectors pCRE-uno with a unique full-length 5′-TGACGTCA-3′ consensus CRE sequence upstream from the RNA polymerase II transcription initiation region of the reporter EGFP gene and pCMV1111 with four CRE sequences were described previously (33). The vectors contain tandem sites for the Nb.BsrDI nicking endonuclease spaced with the 18-nucleotide interval around the 5′-TGACGTCA-3′ consensus CRE sequence (Figure 1A). The derived vectors pCRE-uno-C (with inverse orientation of the Nb.BsrDI sites but otherwise identical to pCRE-uno) and pCRE-zero (in which CRE was mutated to produce an inert 5′-ACTGACTG-3′ sequence) were generated by PCR with the overlapping primers using PfuTurbo® polymerase (Agilent Technologies, Waldbronn, Germany). The template DNA was digested with DpnI (Thermo Fischer Scientific, Darmstadt, Germany)

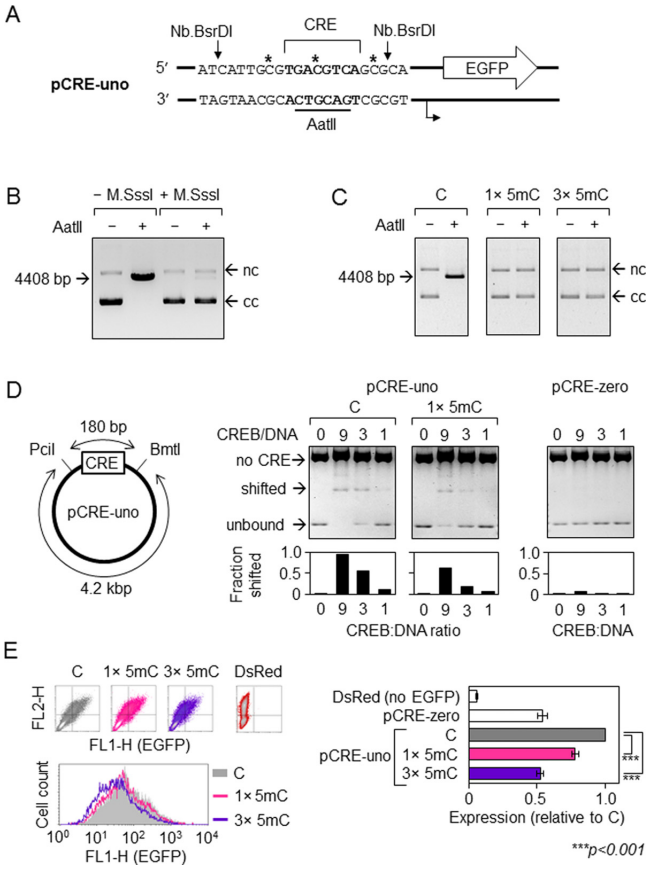


Figure 1. Impact of CpG hemi-methylation in the minimal CRE promoter on the CREB binding and the reporter EGFP gene expression. (**A**) Artificial CRE-uno promoter: transcription start (broken arrow), CRE sequence (bold), AatII site (underlined), Nb.BsrDI nicking positions (vertical arrows), and positions of 5-mC/5-hmC/5-fC/5-caC in the incorporated synthetic oligonucleotides (asterisks). (**B**) Inhibition of the AatII restriction endonuclease by the M.SssI methylation of the central CG-dinucleotide in CRE. Agarose gel electrophoresis of pCRE-uno incubated with AatII. Arrows indicate migration positions of the linearized vector (4408 bp) and of the covalently closed (cc) and nicked (nc) forms of circular vector DNA. (**C**) Verification of the incorporation of synthetic oligonucleotides containing one or three 5-mC (1 × 5mC, 3 × 5mC) by inhibition of the AatII cleavage. (**D**) Scheme of the PciI/BmtI digest of the pCRE-uno vector and electrophoretic mobility shift assay (EMSA) of the PciI/BmtI fragments under the indicated CREB:DNA molar ratios. The shorter (180 bp) fragment contains CRE with no or one 5-mC in the central CG-dinucleotide. The right panel shows control CRE-less vector (pCRE-zero) incubated with CREB in parallel. (**E**) EGFP expression in HeLa cells transfected with pCRE-uno containing none, one (in the central CG-dinucleotide of CRE) or three 5mC. Representative FACS data and relative EGFP expression values (mean ± SD, $n = 7$). Expression in the absence of an EGFP-coding vector (DsRed) denotes the lower detection limit. CRE-less vector (pCRE-zero) was included in two experiments as a reference for the basal expression level. *P*-values calculated by the Student's *t*-test.

followed by direct repair in the ultracompetent Escherichia coli scs-8 (Agilent Technologies).

Synthetic oligonucleotides

DNA CE-phosphoramidites Bz-dA, Bz-dC, iBu-dG, T and Bz-mC were obtained from Glen Research (Sterling, VA) or Link Technologies (Bellshill, Scotland). 5-HmC, 5-fC and 5-caC phosphoramidites were synthesized and solid-phase synthesis of the 18-mer 5′-CATTGCGTGACGTCAGCG deoxyribo-oligonucleotides was performed using the standard protocols described previously (22). Underlined cytosines show positions in which 5-mC/5-hmC/5-fC/5-caC were incorporated, as specified in the text. Synthesized DNA chains were HPLC-purified with VP 250/10 Nucleosil 100-7 C 18 columns (Macherey-Nagel, Düren, Germany) and the quality determined by MALDI-MS (22). 2′-(*R*)-fluorinated derivatives of 5-fC and 5-caC were synthetised and handled as described previously (34). The sequence for incorporation of the 2′-fluorinated nucleotides was 5′-CATTGCGTGACGTCAGCG. Oligonucleotides 5′-CATTGCGTGA[THF/S-THF]GTCAGCG containing the abasic site analog tetrahydrofuran with either phosphodiester (THF) or phosphorothioate (S-THF) 5′-linkage as well as 5′-CATTGCGTGAC[8-oxoG]TCAGCG containing 8-oxo-7,8-dihydroguanine were purchased from BioSpring GmbH (Frankfurt am Main, Germany).

Generation of the reporter constructs with 5-mC/5-hmC/5-fC/5-caC specifically positioned in the CRE sequence

Site-specific double nicks on both sides of the CRE sequence in the vectors pCRE-uno ('top' strand), pCRE-uno-C ('bottom' strand) or pCMV1111 were generated by the Nb.BsrDI endonuclease (NEB GmbH, Frankfurt am Main, Germany). Synthetic 18-mer oligonucleotides containing 5-mC/5-hmC/5-fC/5-caC, the 2′-(*R*)-fluorinated derivatives of 5-fC/5-caC or THF/S-THF in the specified positions were used to displace the excised native DNA strand fragment and seamlessly ligated into vector DNA by the strand exchange protocol described previously (26). The respective control oligonucleotide without modifications was always ligated in parallel for every independent vector preparation. The presence of 5-mC/5-hmC/5-fC/5-caC in the covalently closed vector DNA was verified by the inhibition of cleavage by the AatII restriction endonuclease (NEB).

Electrophoretic mobility shift assays

Vectors pCRE-uno and pCRE-zero were digested with PciI and BmtI (both NEB) to produce the CRE-less 4228 bp fragment and the 180 bp fragment with one (pCRE-uno) or no CRE (pCRE-zero) and cleaned up. Binding reactions contained 10 nM DNA (400 ng in 15 μl reaction volume) and proportional variable amounts of the purified CREB protein (BioCat, Heidelberg, Germany) in the binding buffer composed of 20 mM Tris-HCl (pH 8.0), 50 mM KCl, 5 mM $MgCl_2$, 1 mM dithiothreitol and 5% glycerol (all reagents from Sigma-Aldrich, Seelze, Germany) supplemented with 1 mg/ml bovine serum albumin (NEB). After 30-min incubation at 37°C reactions were chilled on ice and fragments separated by agarose gel electrophoresis in the 0.5 × TBE buffer (Bio-Rad Laboratories, Dreieich, Germany) under cooling conditions, as described previously (35). Band intensities were quantified following the ethidium bromide staining with the help of GelDocTM XR+ molecular imager and the Image LabTM software (Bio-Rad).

Transfections and gene expression analyses

Exponentially growing in six-well plates cells were transfected with the help of Effectene (QIAGEN, Hilden, Germany) with the EGFP reporter constructs (with or without the specified modifications) in combination with the tracer pDsRed-Monomer-N1 vector (Clontech, Saint-Germain-en-Laye, France), 400 ng each vector. The method for quantitative determination of EGFP expression in transiently transfected cells by flow cytometry was validated by titration of the EGFP-encoding vector and described in detail previously (36). Briefly, formaldehyde-fixed cells were equilibrated in phosphate buffered saline (PBS) and analysed using FACSCalibur™ and the CellQuest™ Pro software (Beckton Dickinson GmbH, Heidelberg, Germany). FSC/SSC gating was routinely used to exclude fragmented and aggregated cells and DsRed signal (FL2-H) was applied as additional gating marker to select for the effectively transfected cells. After the exclusion of untransfected cells, EGFP fluorescence (FL1-H) distribution plots were generated and average EGFP expression per cell determined as the median of the distribution. To assess variability between independent experiments, expression levels of constructs with cytosine modifications were calculated in each experiment relative to the expression of the control construct (without modifications) as a ratio of the respective FL1-H values.

RESULTS

Generation of expression constructs containing the C5 atom modifications of defined cytosines

We have previously generated the EGFP-encoding pCRE-uno reporter vector in which the basal RNA polymerase II promoter is coupled with a single CRE as a minimal upstream regulatory element. The CRE-motif is flanked by strand-specific nicking sites for the endonuclease Nb.BsrDI (Figure 1A), which are neutral for the regulation of gene expression, whereas CRE itself measurably activates transcription of the reporter EGFP gene (33). We sought to use these properties for the investigation of consequences of defined cytosine modifications at a single CpG dinucleotide for the regulation of the CRE-driven gene expression. For this purpose, we have introduced 5-mC (or 5-hmC/5-fC/5-caC) into the vector DNA sequence by substitution of the Nb.BsrDI-excised native DNA strand fragment for the synthetic strands containing the respective modifications at the central CG-dinucleotide of the CRE sequence.

As a proof of principle, we first generated expression constructs containing one or three 5-mC residues in the substituted DNA strand (Figure 1A). The pCRE-uno CRE sequence accommodates a unique recognition sequence for the AatII restriction endonuclease, whose activity is inhibited by the CpG methylation (Figure 1B), which en-

ables easy verification of the incorporation of 5-mC with the synthetic oligonucleotide (Figure 1C). Subsequently, we used the same approach for the incorporation of oligonucleotides containing 5-hmC, 5-fC or 5-caC and demonstrated that all the modifications were efficiently integrated into vector DNA (Supplementary Figure S1).

CREB binding and gene expression in the presence of the CpG hemi-methylation

We have analysed by the electrophoretic mobility shift assay the influence of single 5-mC in the central CpG dinucleotide of the CRE consensus sequence on binding of the transcription factor CREB (Figure 1D). Comparison with the nearly identical vector pCRE-zero in which the CRE site has been deleted showed that CREB binds specifically to the restriction fragment containing the CRE-motif and that single 5-mC in the central CpG dinucleotide significantly reduces the band shifting. The equivalent levels of band shifting in the presence of 5-mC, were achieved by increasing the CREB:DNA ratio by the factor of ∼2.5 (Figure 1D and not shown data), suggesting that single methyl group causes a 2- to 3-fold reduction of CREB affinity to its binding site.

To check whether the impaired CREB binding has a functional significance for the gene transcription in cells, we next measured the impact of CpG hemi-methylation on the EGFP expression levels in transfected human cells. Analyses in HeLa cells showed that single 5-mC incorporated into the central CpG nucleotide decreased the expression by approximately one quarter. Hemi-methylation of all three CpG dinucleotides within the exchanged DNA fragment resulted in the halved expression level with respect to the expression of unmethylated control templates (Figure 1E). This quantitatively corresponded to the expression level of the CRE-less vector, suggesting that hemi-methylation at these three positions reduces gene transcription to the basal levels and, hence, nullifies the activatory effect of CREB. Accordingly, we derive that hemi-methylation of the central CpG alone caused an ∼2-fold reduction of the CREB-dependent fraction of transcription. For comparison, total methylation of the pCRE-uno vector by M.SssI almost completely obliterated the EGFP expression, thus indicating a very efficient repression of both induced and basal transcription (Supplementary Figure S2). Together, the results indicate that even a single methyl group—if present at the critical CpG dinucleotide—can measurably reduce the level of gene transcription in cells by direct reduction of the affinity of the transcription factor binding.

The impact of 5-hmC on the gene expression: comparison with 5-mC

We have next generated expression constructs containing synthetic 5-hmC at the same CpG dinucleotides as described above for 5-mC and determined the expression levels in Hela cells under the same conditions. As 5-mC, 5-hmC caused a significant reduction in the EGFP expression levels if present in the central CpG dinucleotide of the CRE sequence and even stronger reduction when all three CpG dinucleotides of the incorporated synthetic DNA were modified by the hydroxymethyl group (Figure 2A). Quantitatively, these effects were similar to the magnitude of the inhibitory effects of 5-mC at the same positions. CREB binding to the modified CRE was also impaired by single 5-hmC to at least the same extent as by 5-mC (Figure 2B), which suggests that the observed decrease in the gene transcription levels in cells was caused primarily by the impaired binding of the transcriptional activator. It has to be noted, however, that the presence of additional methyl or hydroxymethyl groups at the CpG sites outside from the minimal consensus CRE sequence did not cause further prominent decrease of CREB binding in the band-shifting assays (Supplementary Figures S3 and S4), thus indicating that the binding mode and the impact of 5-hmC on the gene expression in cells may be additionally influenced by other factors.

Because 5-hmC has been proposed to play a functional role in DNA demethylation (by either TDG-dependent (16,17,21) or -independent (37,38) pathways, we wondered whether the magnitudes of the repressory effects of both CpG modifications (5-mC and 5-hmC) would vary in the

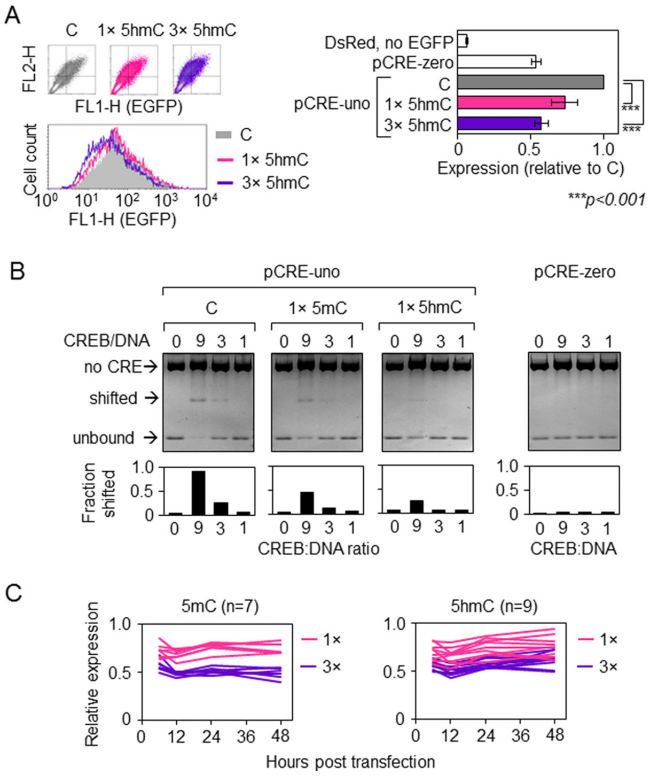


Figure 2. Impacts of one or three 5-hmC in the minimal CRE promoter on the EGFP gene expression and comparison with the respective effects of 5-mC. (**A**) EGFP expression in HeLa cells transfected with pCRE-uno containing none, one or three 5-hmC. Representative fluorescence scatter plots and the correspondent overlaid fluorescence distribution plots of HeLa cells 24 h after transfection (on the left) as well as mean EGFP expression values determined relative to the construct containing unmodified oligonucleotide ('C'). Mean of nine independent experiments ± SD; *P*-values calculated by the Student's *t*-test. (**B**) CREB binding to CRE containing one 5-mC or 5-hmC in the central CG-dinucleotide detected by EMSA of the pCRE-uno PciI/BmtI fragment. (**C**) EGFP expression in HeLa cells transfected with pCRE-uno containing either one (1×) or three (3×) of the indicated CpG modifications in the promoter fragment (relative to pCRE-uno without CpG modifications). Lines represent independent transfections.

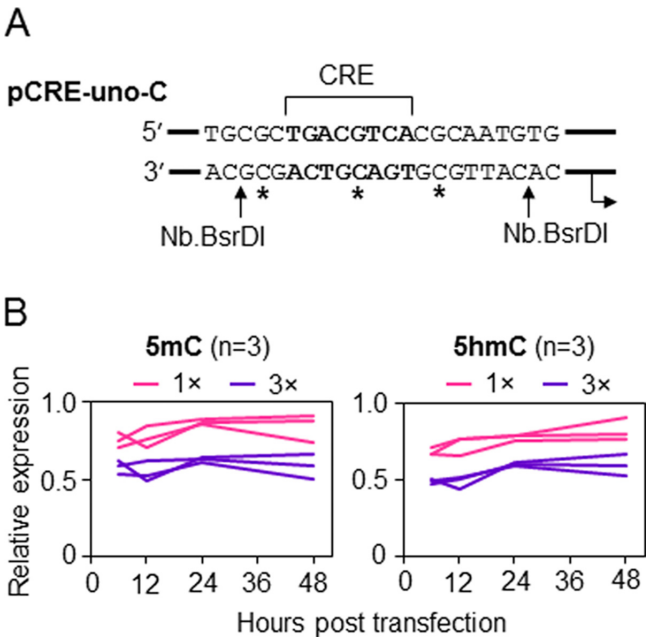


Figure 3. Impacts of 5-mC and 5-hmC in the 'bottom' DNA strand of the minimal CRE promoter on the EGFP gene expression. (**A**) Scheme of the promoter fragment of the pCRE-uno-C vector used for modification of CG-dinucleotides in the bottom strand (asterisks). (**B**) EGFP expression in HeLa cells transfected with pCRE-uno containing either one (1×) or three (3×) of the indicated modifications in the promoter fragment (relative to pCRE-uno without modifications). Lines represent independent transfections.

course of time. However, the results showed that the effects of both modifications persisted for at least 48 hours post transfection. During the whole time course, constructs with 5-mC and 5-hmC showed on average equal extents of transcriptional inhibition (Figure 2C), suggesting that 5-hmC has not been removed from CRE. The values for 5-hmC had a somewhat higher interexperimental noise; however, there was no clear tendency to recovery of the gene expression levels over the time course. The results thus indicate that, at least in the cell model used, hemi-methylated and hemi-hydroxymethylated CpG dinucleotides have overall very similar negative impacts on CREB binding and the CRE-activated transcription.

The effects of 5-mC and 5-hmC are manifested in both DNA strands

It is not known, whether biological consequences of hemi-methylation (or hemi-hydroxymethylation) depend on the DNA strand which is methylated. Because only one strand of the pCRE-uno plasmid (Figure 1A, top strand) can be substituted for a synthetic strand, we have generated an analogous pCRE-uno-C plasmid with flipped BsrDI sites, in which the complementary DNA strand can be replaced (Figure 3A, bottom strand). With CRE sequence as a reference point, we introduced 5-mC and 5-hmC at exactly the same CpG dinucleotides as previously in the top strand and performed time course expression analyses in HeLa cells (Figure 3B). Analogously to the previous experiments (Figure 2C), the results showed that either 5-mC or 5-hmC in the central CpG of the CRE sequence resulted in losses of approximately one fourth of the total gene expression level. Additional modifications at the neighbouring CpG dinucleotides caused ulterior decrease of the gene expression resulting in the loss of approximately half of the total transcriptional activity, which corresponds to residual activity equivalent to the basal transcription level of the CRE-less vector described above. Once again, the effects of both modifications persisted for at least 48 hours. Overall, we conclude that the effects of 5-mC and 5-hmC at the same hemi-modified CpG dyads on the transcriptional activation/repression of the CRE-regulated promoter are qualitatively and quantitatively very similar, regardless of the DNA strand affected by the modification.

Potent and largely indirect repression by 5-fC and 5-caC

The TET-induced mechanism of enzymatic DNA demethylation described previously (16–21) foresees transient generation of 5-fC and perhaps 5-caC. Although active DNA demethylation is regarded as a potential mechanism for re-expression of genes originally silenced by DNA methylation, functional impact of these oxidation products on gene expression has not been sufficiently characterized to date. We have used the experimental setup described above to characterize the impacts of these cytosine modifications on the CRE-driven gene transcription. Because of the variable band-shifting activity between different batches of the recombinant CREB protein, we had to adjust the protein amounts in these experiments with respect to the conditions used in Figures 1D and 2B in order to achieve measurable band shifting with unmodified ('C') DNA (Figures 4A and 5A). We found that both 5-fC and 5-caC to some extent interfere with CREB-binding *in vitro*, however, do not prevent the binding completely, i.e. CREB binding properties to its target sequence were altered by these modifications to similar extents as earlier observed for hemi-methylated and hemi-hydroxymethylated CpG.

Expression analyses of the constructs containing one or three 5-fC in and around the CRE sequence showed that this modification leads to decreased reporter gene expression in the transfected HeLa cells (Figure 4B and C). The impairment of the EGFP expression was initially very mild (as judged from the fluorescence measured at the earliest time point of 6 h); however, the magnitude of the negative impact of 5-fC on the gene expression greatly increased by the 24 h time point after transfection (Figure 4B). At 24 h post-transfections, the degrees of the inhibition of gene expression by 5-fC and 5-caC clearly exceeded the respective effects of 5-mC and 5-hmC reported above (Figures 1–3). In-depth monitoring of the expression of constructs containing one or three hemi-modified CpG sites over a time span of 6–48 h post-transfection revealed a very strong and steady decline of the EGFP expression with no signs of recovery over the whole observation period, suggesting that 5-fC induced persistent repression of the gene transcription (Figure 4C). Thereby, the gene repression could not be accounted solely to the inhibition of CREB binding, as concluded from both strength and kinetics of the observed effects.

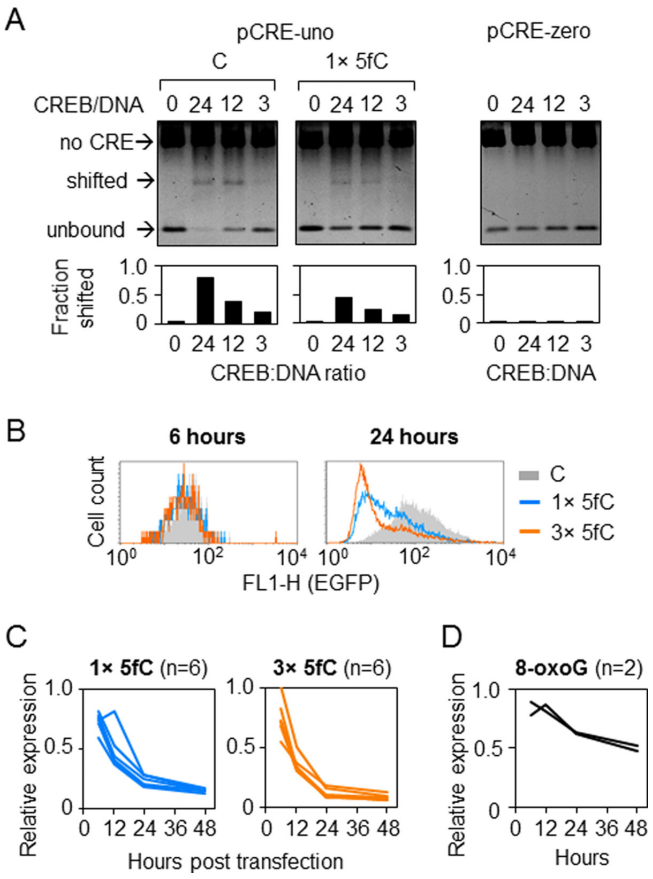


Figure 4. Effects of 5-fC in the minimal CRE promoter. (**A**) CREB binding to CRE containing one 5-fC in the central CG-dinucleotide detected by EMSA. (**B**) Representative fluorescence distribution plots of HeLa cells 24 h after transfection with constructs containing 5-fC in one (1×) or three (3×) CG-dinucleotides (overlaid with the reference 'C' construct obtained with unmodified synthetic oligonucleotide). (**C**) Time-course of the EGFP expression (relative to 'C') in HeLa cells transfected with constructs containing 1× or 3 × 5-fC. Lines show independent transfections. Mean values for each time point are highly significantly different from 1 ($P < 0.001$, heteroscedastic Student's *t*-test). (**D**) Time-course of the EGFP expression in HeLa cells transfected with constructs containing single 8-oxoG in the CRE (two independent experiments).

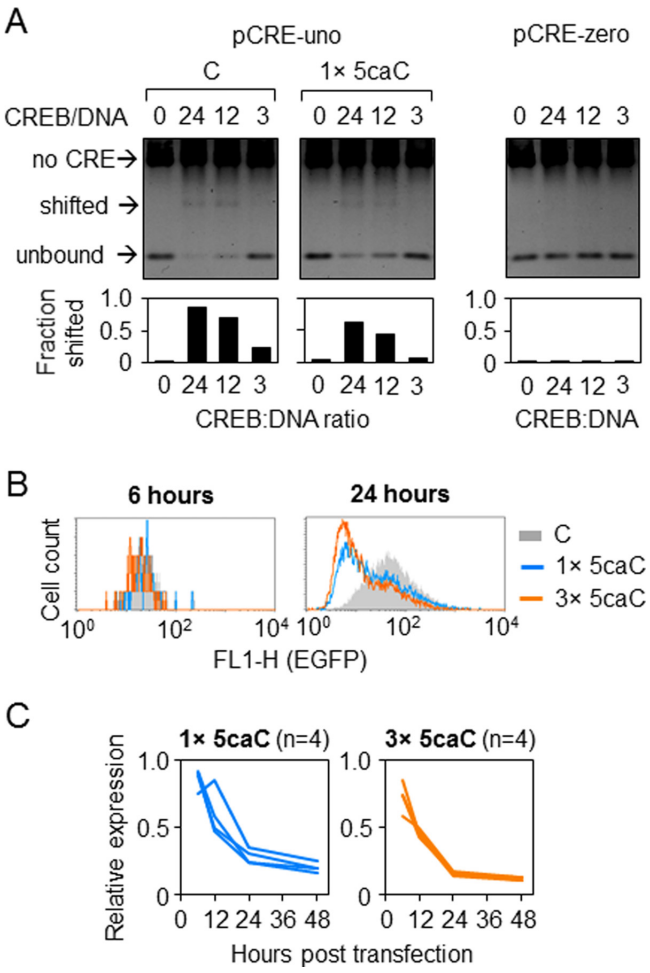


Figure 5. Effects of 5-caC in the minimal CRE promoter. (**A**) CREB binding to CRE containing one 5-caC in the central CG-dinucleotide detected by EMSA. (**B**) Representative fluorescence distribution plots of HeLa cells 24 h after transfection with constructs containing 5-caC in one (1×) or three (3×) CG-dinucleotides overlaid with the reference 'C' construct. (**C**) Time-course of the EGFP expression (relative to 'C') in HeLa cells transfected with constructs containing 1× or 3 × 5-caC. Lines show independent transfections. Mean values for each time point are highly significantly different from 1 ($P < 0.001$, heteroscedastic Student's *t*-test).

It is interesting to note that a similar gradual transcriptional repression was previously reported for reporter constructs containing various nucleobase modifications and has been identified as a hallmark of DNA lesions undergoing base excision repair (BER) (27,29,33). The role of BER in the gene repression has been inferred based on the requirements of the lesion-specific DNA glycosylases and of the strand-cleaved reaction intermediate generated by the apurinic/apyrimidinic site endonuclease APE1 (33). To test whether BER of a modification situated in the CRE sequence would affect the gene expression in a comparable way, we have placed single 8-oxo-7,8-dihydroguanine (8-oxoG) as a representative BER substrate into the CRE sequence. The results indeed showed a time-dependent decline of the gene expression (Figure 4D). Thus, 8-oxoG and 5-fC in the CRE sequence induced qualitatively similar transcriptional responses; however, the negative impact of 5-fC on the gene expression was much more powerful.

Further, we have analysed the CREB binding (Figure 5A) as well as the EGFP expression (Figure 5B and C) with 5-caC incorporated into the pCRE-uno reporter vector. The results showed that responses to 5-caC essentially recapitulated the response to 5-fC, i.e. mild inhibition of CREB binding and time-dependent progressive repression of transcription in transfected HeLa cells. Together, the results indicate that transcriptional responses to 5-fC and 5-caC largely resemble the response to DNA lesions, as exemplified by 8-oxoG. The observed transcriptional repression of the reporter constructs containing 5-fC or 5-caC further suggests that also oxidation of 5-mC/5-hmC to 5-fC/5-caC in the process of TET-induced DNA demethylation may not necessarily lead to transcriptional re-activation.

TDG mediates BER-dependent transcriptional repression by 5-fC and 5-caC

Because BER intermediate has been previously implicated in transcriptional repression or silencing induced by various DNA lesions (27,29,33,36), we questioned whether the progressive inhibition of expression of the reporter vectors containing 5-fC or 5-caC is similarly mediated by BER. Both 5-fC and 5-caC are efficiently removed from DNA *in vivo* and *in vitro* by TDG (16,18); therefore we performed gene expression analyses of vectors containing these modifications in cells with knocked down TDG expression (Supplementary Figure S5). The negative effects of 5-caC in pCRE-uno on the reporter gene expression were greatly reduced in the cell line stably expressing the TDG-specific short hairpin (sh) RNA but not in the isogenic cell lines with knocked down UNG1/2 or SMUG1 DNA glycosylases and not in the cell line stably transfected with non-specific shRNA (Figure 6A, left panel). The results thus indicate that excision of 5-caC by TDG is a prerequisite of ulterior transcriptional silencing of the CRE-controlled expression constructs. The magnitude of the inhibitory effects of 5-fC on the gene expression was also significantly moderated by the TDG knockdown, however to a lesser extent than it was observed for 5-caC.

To verify whether the effects of TDG knockdown were specific to 5-fC/5-caC, we further generated expression constructs carrying AP site (THF) in the same position in CRE (Supplementary Figure S6). As expected, the expression analyses showed a strong decrease of EGFP expression induced by THF but not by the APE1-resistant AP site (S-THF). Importantly, the magnitude of the effect of THF was not at all alleviated by the TDG knockdown. We thus conclude that BER reactions downstream from the base excisions step were not affected.

Because TDG knockdown had quantitatively only modest impact on the expression of constructs containing 5-caC and especially 5-fC, we sought to increase the dynamic range for measurements of the EGFP expression. Therefore, we performed analogous experiments in the pCMV1111 vector which contains a strong promoter with multiple regulatory elements including four consensus CRE sequences, one of which was chosen for targeted incorporation of synthetic C/5-fC/5-caC (Supplementary Figure S1B). Remarkably, also in the context of the strong promoter, both 5-caC and 5-fC induced dynamic decrease of the gene expression and in both cases the effects were significantly moderated by the TDG knockdown (Figure 6A, right panel). Also in this promoter the effect of TDG knockdown was somewhat smaller for 5-fC than 5-caC, but the difference was far less pronounced than in pCRE-uno.

The results obtained in the TDG knockdown cells underpinned the role of BER of 5-fC and 5-caC in the transcriptional repression by these cytosine modifications (Figure 6A); however, since the recovery of gene expression was not complete, they did not exclude a potential contribution of other mechanisms. To eliminate the base excision activity completely, we have applied a chemical approach. It was previously reported that sugar 2′-fluorination of the 5-fC and 5-caC nucleotides prevents the TDG excision activity completely while having only minimal impact on the physical and functional properties of the DNA helix (34). Therefore, we next generated the pCMV1111 expression constructs carrying single 5-fC/5-caC deoxyribonucleotides or their 2′-fluorinated analogs in the specified position in CRE (Supplementary Figure S7). The expression analyses in HeLa cells explicitly showed that 2′-fluorination of deoxyribose fully reversed the gene repression by 5-fC and 5-caC (Figure 6B and C). This effect was exactly reproduced in the pCRE-uno vector (Supplementary Figure S8). The results thus show that regulatory promoter elements retain their activatory functions and that 5-fC and 5-caC do not lead to transcriptional repression when the base excision is inhibited, e.g. by the deoxyribose 2′-fluorination.

DISCUSSION

There are at least two mechanisms for regulation of gene transcription by 5-mC in DNA. It was previously suggested that selective methylation of CpG dinucleotides present in upstream promoter elements can interfere with recruitment of the specific transcription factors (6–8). On the other hand, CpG methylation can recruit general co-repressors via the MBD binding (1,2). Using a minimal CRE-driven promoter with an artificially introduced hemi-methylated CpGs in defined positions, we show here that a single methyl group in the central CpG of the CRE sequence causes a 2- to 3-fold reduction of the specific CREB binding affinity and an equivalent decrease in the CRE-dependent gene expression in human cells (Figure 1). This proportionality suggests that impaired CREB binding has direct functional relationship with the gene expression. Nevertheless, hemi-methylation of the neighboured CpGs clearly added to the decreased gene expression in cells (Figures 2 and 3), even though CREB binding was not further impaired by the modifications outside of the CRE consensus sequence (Figures 2 and 3, Supplementary Figures S3 and S4), which indicates that impaired CREB binding is not the only mechanism for the declined gene expression if more than one 5-mC is present. Indeed, massive methylation of the reporter vector by M.SssI resulted in a complete transcriptional silencing of the reporter constructs (Supplementary Figure S2). This mechanism appears to dominate over the CREB-induced transcriptional activation, since the enzymatic methylation eliminated both induced and basal expression. This notion is supported by further observation that selective partial demethylation of the CRE region did not rescue the expression of the M.SssI methylated reporter constructs (data not shown). In summary, the results demonstrate that methylation and even hemi-methylation of single CpG dinucleotides in critical gene regulatory elements can have substantial negative impacts on the gene expression by directly impeding the transcription factor binding. It is necessary to stress, however, that this effect can only be functionally significant when the surrounding DNA sequence is unmethylated.

The outcome of CRE hydroxymethylation on the gene expression was essentially the same as of methylation (Figures 1–3), thus suggesting that the functional modes of inhibition of the activator-induced gene transcription by these two modifications are similar or common. As for 5-mC, the magnitude of negative effect of single 5-hmC on the gene ex-

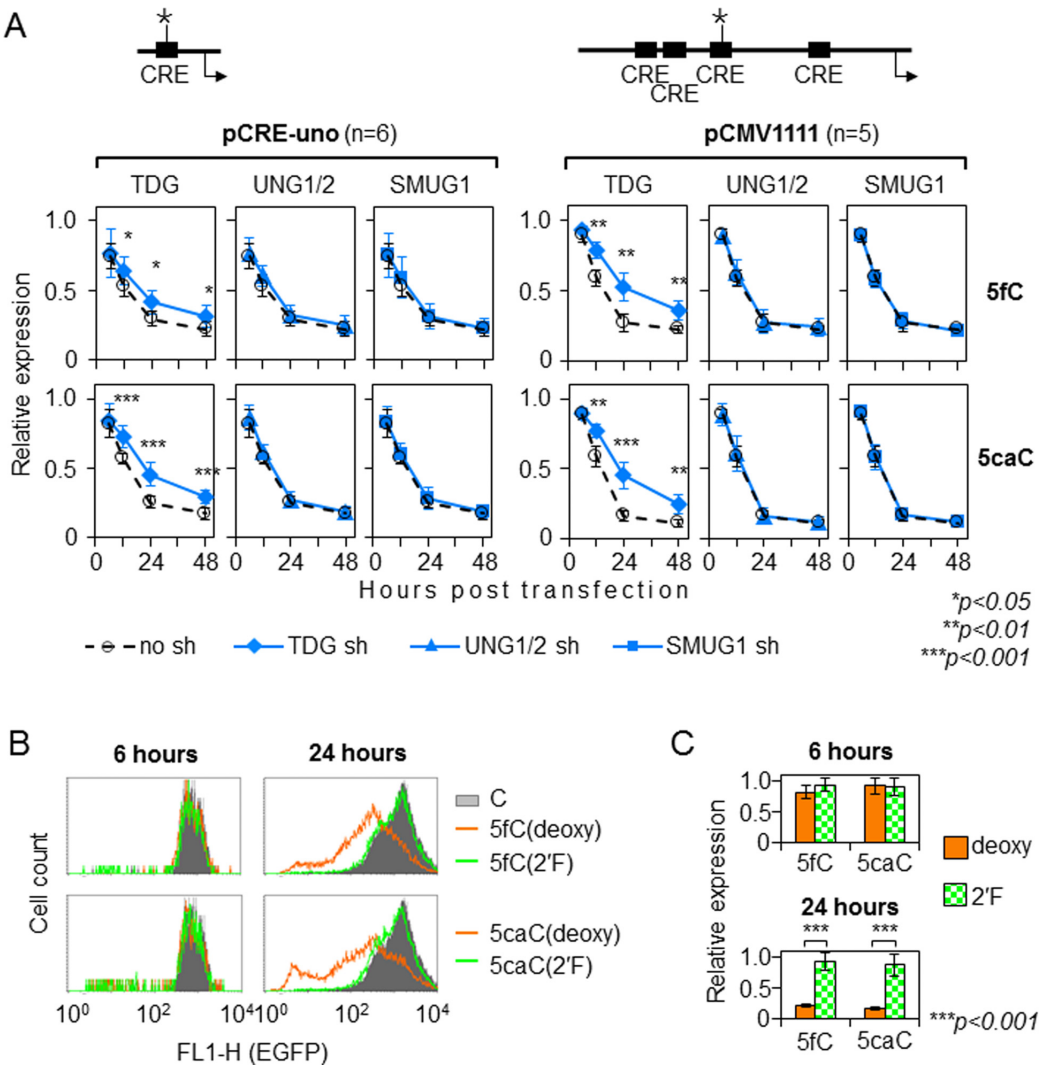


Figure 6. Impact of BER on the expression of reporter constructs containing single 5-fC/5-caC in the CRE sequence. (**A**) Effects of 5-fC and 5-caC on EGFP expression in HeLa-derived cell lines with stable shRNA knockdown of the UNG1/2, TDG or SMUG1 DNA glycosylases as compared to the control isogenic cell line ('no sh'). Expression of the constructs containing 5-fC or 5-caC in the pCRE-uno (left panel) and pCMV1111 (right panel) vectors at each time point is presented relative to the respective reference constructs containing cytosine (mean ± SD; *P*-values calculated by the Student's *t*-test). (**B, C**) EGFP expression in HeLa cells at the specified times after transfection with pCMV1111 containing single 5-fC or 5-caC as deoxyribonuceotides (deoxy) or the respective 2′-(R)-fluorinated derivatives (2′F). Representative FACS data (B) and relative EGFP expression values (mean ±SD, n = 5, *P*-values calculated by the Student's t-test) (C).

pression is consistent with the impediment of CREB binding to CRE sequence by the modification. This result suggests that the effects of both 5-mC and 5-hmC are likely mediated by prevention of binding of activatory proteins specific for unmodified CpG rather than by recruitment of co-repressors via a specific interaction with 5-mC or 5-hmC. Such a mechanism is in agreement with proteomics data which identified little overlap between the spectra of specific interactors of 5-mC and 5-hmC, whereas binding of a much larger set of proteins, including several families of transcriptional activators, was disrupted by either modification (22).

The degrees of inhibition of gene expression by single 5-mC and 5-hmC stayed steady over the whole time-course of the gene expression analyses (Figures 2 and 3), suggesting that epigenetic state of the promoter region remained stable over at least 48 hours. In a marked contrast, the expression of constructs containing 5-fC or 5-caC showed strong negative dynamics, indicating a progressive silencing of the affected promoter (Figures 4 and 5). Intriguingly, the spectrum of putative 5-fC and 5-caC readers previously identified by the proteomics approaches includes candidates with potential gene silencing functions such as DNMT1 and the components of the Mi-2/NuRD nucleosome remodeling deacetylase complex (22,39). On the other hand, the enrichment of DNA repair proteins including TDG and MPG in the sets of putative 5-fC and 5-caC readers raises the possibility that proteins bound to products or the DNA repair reactions have been also classified as interactors in these experimental setups.

We have previously reported that excision of various base modifications by the specific DNA glycosylases in concert with the apurinic/apyrimidinic site endonuclease commonly initiates transcriptional silencing (27,29,33). Now we show that TDG clearly contributes to progressive transcriptional silencing induced by 5-caC and 5-fC (Figure 6A), which strongly suggests the key role of BER in this process. Incomplete recovery of the gene expression levels under the conditions of TDG knockdown should likely be attributed to low residual TDG activity in cells, since complete abolition of excision of 5-fC/5-caC by the deoxyribose 2′-fluorination fully restored the expression (Figure 6B and C). Nevertheless, based on these results it still cannot be ruled out that some other DNA glycosylase (or endonuclease) could also contribute to processing of 5-fC and/or 5-caC. Thus, NEIL1, NEIL2 and NEIL3 interact with TET1 and have been implicated in transcriptional activation of TET-oxidised M.SssI-methylated vector DNA in the absence of TDG (40). Even though the evidence is missing that NEIL proteins autonomously excise 5-fC, another study showed that NEIL1 and NEIL2 accelerate the TDG turnover and thus may stimulate excision when TDG is limited (15). Since our findings implicate a BER intermediate in transcriptional silencing induced by 5-fC and 5-caC, it is intriguing to suggest that balance between the activity of individual enzymes within BER or coordination between the BER steps in a given cell type or during a particular differentiation phase may have profound consequences for the regulation of gene expression by these DNA modifications. In our system, both generation of AP site and subsequent strand cleavage by APE1 are required for the gene silencing (Figure 6, Supplementary Figure S6). Taking into account intrinsically low processivity of TDG, it would be important to investigate, whether the switches exist which would halt one of these steps thereby possibly preventing the gene silencing. Such mechanisms could be critical to safeguard the proper gene function following the restoration of unmethylated CpG in the genome.

In summary, our findings have implications for understanding of transcription regulatory functions of hemimethylated (or hemi-modified by 5-hmC, 5-fC or 5-caC) CpG dinucleotides. Such CpG modifications can arise in cells during DNA demethylation or de novo DNA methylation reactions. The results indicate that 5-mC and 5-hmC act as stable regulatory marks predominantly by decreasing binding of transcriptional activator CREB to the modified target motif. The results further suggest that 5-fC and 5-caC trigger dynamic changes in the promoter activity. Strikingly, in the minimal CRE-regulated promoter employed here, both 5-fC and 5-caC functioned primarily as repressory marks. The impacts of 5-fC and 5-caC on the promoter activity are only partly attributable to the direct impairment of CREB binding and the negative effects on the gene expression strongly advance with time in a TDG-dependent fashion. Extrapolating present results to the situation during active DNA demethylation in cells, we suggest that generation of 5-fC and 5-caC in the regulatory promoter elements and further processing of these modifications by BER by default leads to perpetuation and enhancement of the repressed transcriptional state and that transcriptional reactivation would require additional signals.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors thank Bork Lühnsdorf, Bernd Epe, Alexander Ishchenko, Lars Schomacher and Michael Musheev for discussions of the experimental data. The authors are grateful to Bernd Epe and his group members for free access to the FACS instrument and other laboratory facilities.

FUNDING

Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) [KH 263/1, KH263/2, Heisenberg Fellowship KH 263/3 to A.K; GRK 2062 Fellowship to M.R.]. Funding for open access charge: DFG, German Research Foundation.
Conflict of interest statement. None declared.

REFERENCES

1. Schubeler,D. (2015) Function and information content of DNA methylation. *Nature*, **517**, 321–326.
2. Jones,P.A. (2012) Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.*, **13**, 484–492.
3. Deaton,A.M. and Bird,A. (2011) CpG islands and the regulation of transcription. *Genes Dev.*, **25**, 1010–1022.
4. Wigler,M., Levy,D. and Perucho,M. (1981) The somatic replication of DNA methylation. *Cell*, **24**, 33–40.
5. Gruenbaum,Y., Cedar,H. and Razin,A. (1982) Substrate and sequence specificity of a eukaryotic DNA methylase. *Nature*, **295**, 620–622.
6. Ben-Hattar,J. and Jiricny,J. (1988) Methylation of single CpG dinucleotides within a promoter element of the Herpes simplex virus tk gene reduces its transcription in vivo. *Gene*, **65**, 219–227.
7. Watt,F. and Molloy,P.L. (1988) Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. *Genes Dev.*, **2**, 1136–1143.
8. Iguchi-Ariga,S.M. and Schaffner,W. (1989) CpG methylation of the cAMP-responsive enhancer/promoter sequence TGACGTCA abolishes specific factor binding as well as transcriptional activation. *Genes Dev.*, **3**, 612–619.
9. Hodges,E., Molaro,A., Dos Santos,C.O., Thekkat,P., Song,Q., Uren,P.J., Park,J., Butler,J., Rafii,S., McCombie,W.R. *et al.* (2011) Directional DNA methylation changes and complex intermediate states accompany lineage specificity in the adult hematopoietic compartment. *Mol. Cell*, **44**, 17–28.
10. Stadler,M.B., Murr,R., Burger,L., Ivanek,R., Lienert,F., Scholer,A., van Nimwegen,E., Wirbelauer,C., Oakeley,E.J., Gaidatzis,D. *et al.* (2011) DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature*, **480**, 490–495.
11. Ziller,M.J., Gu,H., Muller,F., Donaghey,J., Tsai,L.T., Kohlbacher,O., De Jager,P.L., Rosen,E.D., Bennett,D.A., Bernstein,B.E. *et al.* (2013) Charting a dynamic DNA methylation landscape of the human genome. *Nature*, **500**, 477–481.
12. Cadet,J. and Wagner,J.R. (2014) TET enzymatic oxidation of 5-methylcytosine, 5-hydroxymethylcytosine and 5-formylcytosine. *Mut. Res.*, **764–765**, 18–35.
13. Hahn,M.A., Szabo,P.E. and Pfeifer,G.P. (2014) 5-Hydroxymethylcytosine: a stable or transient DNA modification? *Genomics*, **104**, 314–323.
14. Klungland,A. and Robertson,A.B. (2016) Oxidized C5-methyl cytosine bases in DNA: 5-hydroxymethylcytosine; 5-formylcytosine; and 5-carboxycytosine. *Free Radic. Biol. Med.*, **107**, 62–68.
15. Schomacher,L., Han,D., Musheev,M.U., Arab,K., Kienhofer,S., von Seggern,A. and Niehrs,C. (2016) Neil DNA glycosylases promote substrate turnover by Tdg during DNA demethylation. *Nat. Struct. Mol. Biol.*, **23**, 116–124.

16. He,Y.F., Li,B.Z., Li,Z., Liu,P., Wang,Y., Tang,Q., Ding,J., Jia,Y., Chen,Z., Li,L. *et al.* (2011) Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science*, **333**, 1303–1307.
17. Ito,S., Shen,L., Dai,Q., Wu,S.C., Collins,L.B., Swenberg,J.A., He,C. and Zhang,Y. (2011) Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science*, **333**, 1300–1303.
18. Maiti,A. and Drohat,A.C. (2011) 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.
19. Pfaffeneder,T., Hackner,B., Truss,M., Munzel,M., Muller,M., Deiml,C.A., Hagemeier,C. and Carell,T. (2011) The discovery of 5-formylcytosine in embryonic stem cell DNA. *Angew. Chem.*, **50**, 7008–7012.
20. Lister,R., Mukamel,E.A., Nery,J.R., Urich,M., Puddifoot,C.A., Johnson,N.D., Lucero,J., Huang,Y., Dwork,A.J., Schultz,M.D. *et al.* (2013) Global epigenomic reconfiguration during mammalian brain development. *Science*, **341**, 1237905.
21. Weber,A.R., Krawczyk,C., Robertson,A.B., Kusnierczyk,A., Vagbo,C.B., Schuermann,D., Klungland,A. and Schar,P. (2016) Biochemical reconstitution of TET1-TDG-BER-dependent active DNA demethylation reveals a highly coordinated mechanism. *Nat. Commun.*, **7**, 10806.
22. Spruijt,C.G., Gnerlich,F., Smits,A.H., Pfaffeneder,T., Jansen,P.W., Bauer,C., Munzel,M., Wagner,M., Muller,M., Khan,F. *et al.* (2013) Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell*, **152**, 1146–1159.
23. Kriaucionis,S. and Heintz,N. (2009) The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science*, **324**, 929–930.
24. Zhang,L., Lu,X., Lu,J., Liang,H., Dai,Q., Xu,G.L., Luo,C., Jiang,H. and He,C. (2012) Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA. *Nat. Chem. Biol.*, **8**, 328–330.
25. Kress,C., Thomassin,H. and Grange,T. (2006) Active cytosine demethylation triggered by a nuclear receptor involves DNA strand breaks. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 11112–11117.
26. Luhnsdorf,B., Kitsera,N., Warken,D., Lingg,T., Epe,B. and Khobta,A. (2012) Generation of reporter plasmids containing defined base modifications in the DNA strand of choice. *Anal. Biochem.*, **425**, 47–53.
27. Kitsera,N., Stathis,D., Luhnsdorf,B., Muller,H., Carell,T., Epe,B. and Khobta,A. (2011) 8-Oxo-7,8-dihydroguanine in DNA does not constitute a barrier to transcription, but is converted into transcription-blocking damage by OGG1. *Nucleic Acids Res.*, **39**, 5926–5934.
28. Kitsera,N., Gasteiger,K., Luhnsdorf,B., Allgayer,J., Epe,B., Carell,T. and Khobta,A. (2014) Cockayne syndrome: varied requirement of transcription-coupled nucleotide excision repair for the removal of three structurally different adducts from transcribed DNA. *PLoS One*, **9**, e94405.
29. Luhnsdorf,B., Epe,B. and Khobta,A. (2014) Excision of uracil from transcribed DNA negatively affects gene expression. *J. Biol. Chem.*, **289**, 22008–22018.
30. Rishi,V., Bhattacharya,P., Chatterjee,R., Rozenberg,J., Zhao,J., Glass,K., Fitzgerald,P. and Vinson,C. (2010) CpG methylation of half-CRE sequences creates C/EBPalpha binding sites that activate some tissue-specific genes. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 20311–20316.
31. Moore,S.P., Toomire,K.J. and Strauss,P.R. (2013) DNA modifications repaired by base excision repair are epigenetic. *DNA Repair*, **12**, 1152–1158.
32. Verri,A., Mazzarello,P., Biamonti,G., Spadari,S. and Focher,F. (1990) The specific binding of nuclear protein(s) to the cAMP responsive element (CRE) sequence (TGACGTCA) is reduced by the misincorporation of U and increased by the deamination of C. *Nucleic Acids Res.*, **18**, 5775–5780.
33. Allgayer,J., Kitsera,N., Bartelt,S., Epe,B. and Khobta,A. (2016) Widespread transcriptional gene inactivation initiated by a repair intermediate of 8-oxoguanine. *Nucleic Acids Res.*, **44**, 7267–7280.
34. Schroder,A.S., Parsa,E., Iwan,K., Traube,F.R., Wallner,M., Serdjukow,S. and Carell,T. (2016) 2'-(R)-Fluorinated mC, hmC, fC and caC triphosphates are substrates for DNA polymerases and TET-enzymes. *Chem. Commun.*, **52**, 14361–14364.
35. Alessandri,M., Beretta,G.L., Ferretti,E., Mancia,A., Khobta,A. and Capranico,G. (2004) Enhanced CPT sensitivity of yeast cells and selective relaxation of Ga14 motif-containing DNA by novel Gal4-topoisomerase I fusion proteins. *J. Mol. Biol.*, **337**, 295–305.
36. Khobta,A., Anderhub,S., Kitsera,N. and Epe,B. (2010) Gene silencing induced by oxidative DNA base damage: association with local decrease of histone H4 acetylation in the promoter region. *Nucleic Acids Res.*, **38**, 4285–4295.
37. Grin,I. and Ishchenko,A.A. (2016) An interplay of the base excision repair and mismatch repair pathways in active DNA demethylation. *Nucleic Acids Res.*, **44**, 3713–3727.
38. Xue,J.H., Xu,G.F., Gu,T.P., Chen,G.D., Han,B.B., Xu,Z.M., Bjoras,M., Krokan,H.E., Xu,G.L. and Du,Y.R. (2016) Uracil-DNA glycosylase UNG promotes Tet-mediated DNA demethylation. *J. Biol. Chem.*, **291**, 731–738.
39. Iurlaro,M., Ficz,G., Oxley,D., Raiber,E.A., Bachman,M., Booth,M.J., Andrews,S., Balasubramanian,S. and Reik,W. (2013) A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation. *Genome Biol.*, **14**, R119.
40. Muller,U., Bauer,C., Siegl,M., Rottach,A. and Leonhardt,H. (2014) TET-mediated oxidation of methylcytosine causes TDG or NEIL glycosylase dependent gene reactivation. *Nucleic Acids Res.*, **42**, 8592–8604.

8 A Network of Metabolic Enzymes Controls Tet3 Activity in the Brain

8.1 Introduction

Tet enzymes are α-ketoglutarate-dependent dioxygenases, which oxidize the DNA base 5-methylcytosine (mC) to 5-hydroxymethylcytosine (hmC)[78,79], 5-formylcytosine (fC)[80] and 5-carboxycytosine (caC).[81,82] The currently available information suggests that these bases have epigenetic functions[194,196,228,409] and are directly involved in active demethylation processes.[217] In this respect it was shown that fC and caC are excised by the thymine DNA glycosylase (TDG), resulting in abasic sites that are repaired by the insertion of unmodified dC-bases.[82,222] High levels of hmC, fC and caC are observed in mouse embryonic stem cells (mESC)[83], but the highest hmC levels were measured in adult brain.[104,158,196] Here, hmC was found to affect transcription.[168,183,184,188,190] The hmC levels are in addition dynamically regulated by neural activity.[346,410–412] In mammals, three Tet enzymes are known (Tet1-3) of which Tet3 and a Tet3 isoform lacking the CXXC domain (Tet3^{-CXXC}) are the most abundant forms in differentiated tissues, including brain.[265,413–416] The Tet-performed oxidation of mC requires α-ketoglutarate as a co-substrate, which is converted to succinate and CO_2.[251,268] This strict dependence on α-ketoglutarate[256] creates an unsolved supply problem. α-ketoglutarate is typically manufactured in the mitochondria and it is unclear how it reaches the Tet enzymes in the cell nucleus. Since α-ketoglutarate availability was shown to be rate determining in various tissues and mESC, the controlling of its supply could be a means for regulating the Tet activity.[256,417]

8.2 Generation of Tet3 interactome fishing baits

In order to obtain information about Tet3 interaction partners in brain tissue that may help supplying Tet3 with α-ketoglutarate, we performed an affinity proteomics study and concentrated in our analysis on enzymes that are involved in the biosynthesis of α-ketoglutarate. For the experiment, we developed the workflow depicted in Fig. 12A. First, we overexpressed full length Tet3 and Tet3^{-CXXC} as N-terminal GFP fusions (GFP-

Tet3, GFP-Tet3 CXXC) in HEK293T cells. We next immobilized the GFP-Tet3 fusion proteins on anti-GFP nanobody-coated magnetic beads. Contaminating HEK293T proteins were removed by extensive washing with up to 1 M NaCl. This did not lead to Tet3 loss from the beads (Extended Data Fig. 1A). In order to verify that Tet3 was still bound to the beads, we treated the beads subsequently with 200 mM glycine (pH = 2.5), which led to the elution of a protein with a molecular weight of about 220 kDa, as expected for the GFP-Tet3 fusion protein (Extended Data Fig. 1A). Next, we investigated if the immobilized Tet3 proteins are functionally active. To this end, the stringently washed Tet3-loaded beads were added to a solution of an oligonucleotide with a single mC in a central position plus Fe^{2+} and α-ketoglutarate. Sucessful conversion of the mC base to the oxidized forms hmC, fC and caC was determined by analysis on a triple quadrupole mass spectrometer as described earlier.[83] For comparison, we performed the same experiment with the immobilized catalytic domain of Tet1 (Tet1cd) as a positive control. A sample without a Tet enzyme served as a negative control. As depicted in Extended Data Fig. 1B, oxidation of mC within DNA is clearly observed for Tet1cd and also for the beads covered with full length Tet3, but not for the negative control. As expected, Tet1cd is more active. Importantly, the Tet3-covered beads oxidize the base mC to all expected products hmC, fC and caC in the presence of α-ketoglutarate and Fe^{2+}.

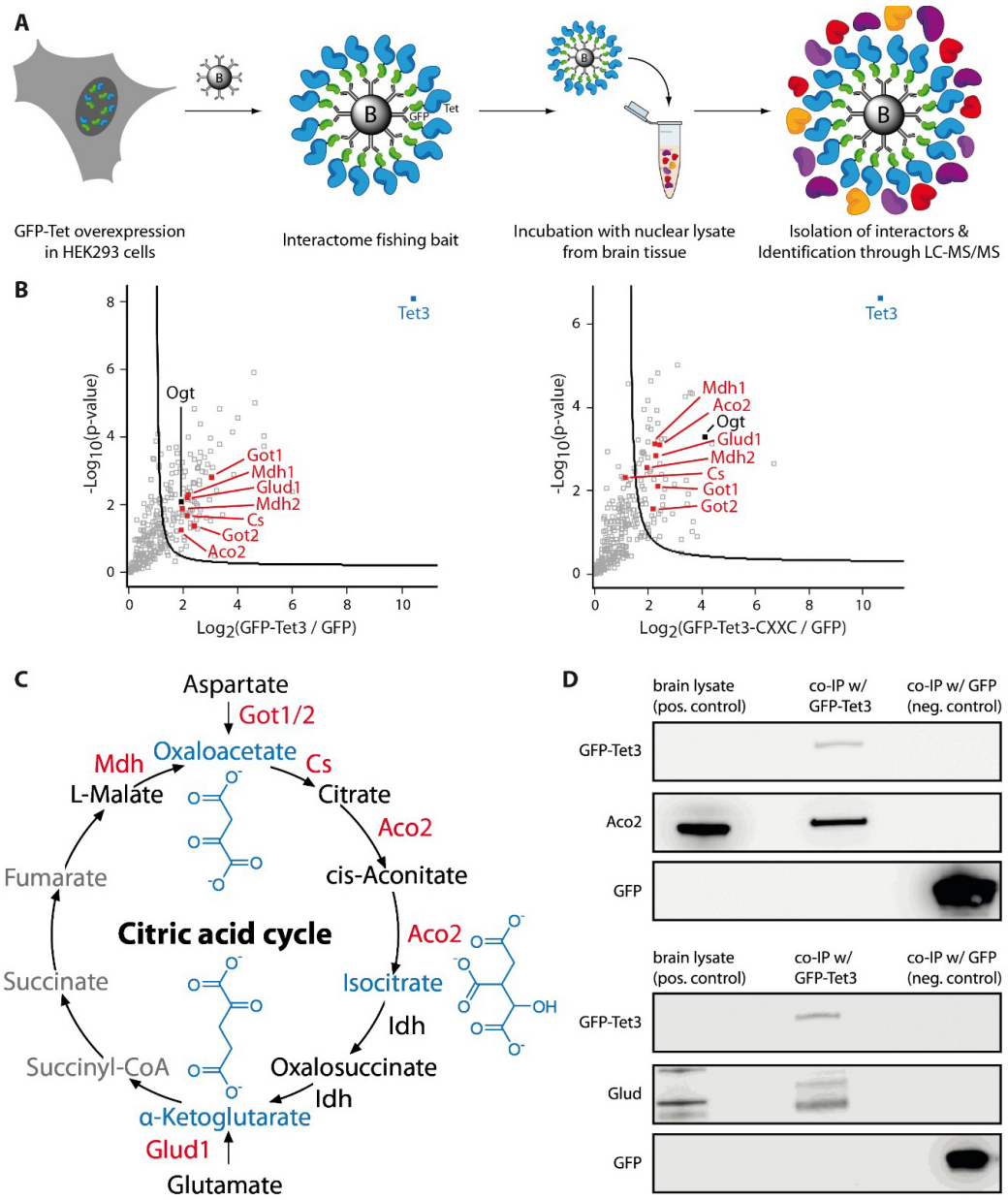


Figure 12: Tet3 interacts with Glud1, Got 1/2 and enzymes from the citric acid cycle in mouse brain. (A) Schematic representation of the modified co-immunoprecipitation (co-IP) workflow. B: anti-GFP magnetic beads. (B) Volcano plots of the interaction partners found in the modified Tet3-co-IPs in adult mouse brain for Tet3 and Tet3^{-CXXC} (FDR = 0.05; s0 = 2). (C) The citric acid cycle with the Tet3 interacting proteins marked in red. Aco2 = aconitase 2; Cs = citrate synthase; Glud1 = glutamate dehydrogenase1; Got1/2 = glutamic-oxaloacetic transaminase; Idh = isocitrate dehydrogenase; Mdh = malate dehydrogenase. (D) Western blots of the modified Tet3-co-IP using antibodies against Aco2 and Glud.

8.3 Tet3 interacts with metabolic enzymes

The beads, coated with active Tet3 or Tet3^{-CXXC}, were next used as baits to fish for interacting proteins in nuclear lysates of whole adult mouse brains. The samples were analysed using quantitative label-free mass spectrometry.[418] Four biological replicates were measured in order to obtain statistically meaningful quantification data. In the experiments, beads coated with unfused GFP were used as the negative control. The statistical analysis of the data shows, that this method provides highly reproducible data sets. The obtained data show a high correlation within the replicates as determined by the Pearson correlation (Extended Data Fig. 2A). The intensity profiles show the expected data distribution across the samples (Extended Data Fig. 2B). We detected a number of interesting interactors such as Pura, Uba1, Uchl1 and Mark1-3 (Extended Data Fig. 3). Particularly informative, however, is the detection of Ogt, which is a known Tet interactor (Fig. 12B).[419] This observation provides a first validation of the obtained data. In agreement with an earlier study[194], we also observed structural and extracellular matrix proteins due to their sheer abundance in the histologically heterogeneous brain samples. By focussing the analysis on interaction partners able to supply α-ketoglutarate, we detected proteins of the citric acid cycle in the data set (TCA cycle, Fig. 12B). We found specifically malate dehydrogenase (Mdh1/2), citrate synthase (Cs) and aconitase2 (Aco2), all involved in the biosynthesis of α-ketoglutarate from malate (Fig. 12C). The only other enzyme, which is needed to convert isocitrate into α-ketoglutarate, isocitrate dehydrogenase (Idh), is suprisingly missing in the dataset. The proteomics data show furthermore a potential association of Tet3 with the two glutamate-oxaloacetate transaminases (Got1 and Got2), which convert aspartic acid into oxaloacetate. If these found enzymes are indeed catalytically active in close association with Tet3, this would consequently lead to an increased local concentration of isocitrate in close vicinity of Tet3 (Fig. 12C). In the data set, we detected to our suprise in addition the enzyme glutamate dehydrogenase (Glud1) as a potential Tet3 interactor (Fig. 12B). Because Glud1 deaminates glutamate to α-ketoglutarate, this interaction would establish a direct TCA-independent α-ketoglutarate supply (Fig. 12C). Glutamate is an abundant amino acid but it is also a mediator of excitatory signals in the mammalian central nervous system.[420] The interaction of Tet3 with Glud1 could therefore be a potential link between neural activity and mC to hmC oxidation (Fig. 12C).[410,412]

8.4 Verification of the interactions and localization of the interactors

In order to gain support for the observed interactions, we used antibodies against Aco2 and Glud1 to verify their presence on the Tet3 covered beads. Fig. 12D (Extended Data Fig. 4A) shows that both proteins are indeed clearly detectable. We next performed co-IP of endogenous Tet3 using nuclear lysate from adult mouse brain. First, we confirmed the specificity of the anti-Tet3 antibody (Fig. 13A and Extended Data Fig. 4B).28 We then confirmed the interaction of endogenous Tet3 with Glud1 and Aco2 by western blotting analysis (Fig. 13B, Extended Data Fig. 4B). Additional investigation of the proteins that co-precipitated with endogenous Tet3 using mass spectrometry and LFQ confirmed also the presence of the previously identified α-ketoglutarate biosynthesis enzymes Glud1, Aco2, Cs, Got2, Mdh2, which were significantly enriched in the endogenous Tet3-co-IP compared to the negative control (Fig. 13C). The proteins Got1 and Mdh1 were also found in the data-set, but they were not significantly enriched (p-value 0.0978 for Got1 and 0.1418 for Mdh1). We next investigated the interaction of Glud1, Aco2 and Got2 with Tet3 directly in the nuclei of brain cells. To this end we performed immunohistochemistry (IHC) experiments and proximity ligation assays (PLA).[421] We used for the studies mouse hippocampus because of the high hmC content. We first tested the Tet3 antibody regarding unspecific interactions in Tet-deficient mESCs and found no unspecific signals at the applied concentrations (Extended Data Fig. 5A). In the hippocampal slices, we detected for Tet3 and also for Glud1 clear localization in the cell nucleus (Fig. 13D, Extended Data Fig. 5B). This is suprising for Glud1 given that it is a mitochondrial protein. Some nuclear localization can also be detected for Aco2 (Fig. 13D, Extended Data Fig. 5B). Most important, however, are the PLA data, which support for all three proteins Glud1, Aco2 and Got2 a direct interaction with Tet3. We detected numerous PLA signals in the case of Tet3+Glud1 (Fig. 13E). Spots were also detected for Tet3+Aco2 (Fig. 13E) and Tet3+Got2 (Extended Data Fig. 5D). No spots were obtained in the corresponding negative control experiment (Extended Data Fig. 5D). A positive control experiment with the known interactor Ogt provided spot numbers comparable to Glud1 (Fig. 13E).

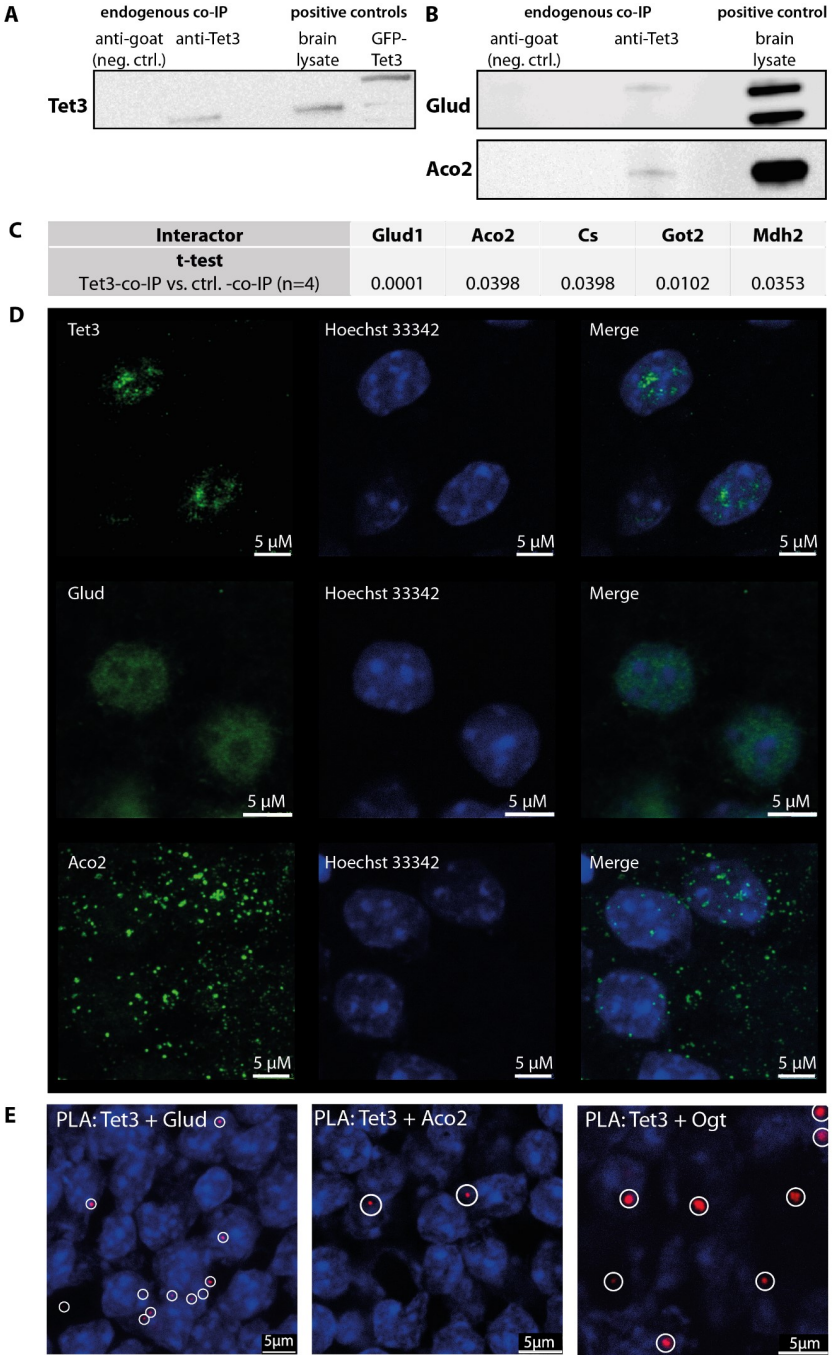


Figure 13: Validation of Tet3 interactors. (A) Western blot shows the specificity of the anti-Tet3 antibody. (B) Western blot results of the endogenous Tet3-co-immunoprecipitation (co-IP) confirm Glud and Aco2 as Tet3 interactors in mouse brain. (C) Mass spectrometry analysis of endogenous Tet3-co-IP and ctrl.-co-IP (with anti-goat antibody). (D) Immunohistochemistry of Tet3, Glud1 and Aco2 in the mouse hippocampus shows that Tet3 and Glud1 are present in the cell nucleus. Aco2 is mostly located in the mitochondria. (E) In situ proximity ligation assay (PLA) shows the direct interaction (red signals; encircled for better visibility) between Glud1/Tet3 and Aco2/Tet3 in mouse hippocampal nuclei. Tet3/Ogt-PLA served as positive control.

Taken together, co-IP and PLA show that in hippocampus, Tet3 associates with the enzymes Aco, Cs, Mdh2 and Got2. Particular strong signals were detected with Glud1 confirming the possibility that in brain cells mC to hmC oxidation is performed with α-ketoglutarate that is manufactured from the neurotransmitter amino acid glutamate.

8.5 Functional study of the Glud1-Tet3 interaction

We next studied the functional connection between Glud1 and Tet3. For the experiment, we co-expressed GFP-Tet3FL together with FLAG-tagged Glud1 (FLAG-Glud1) in HEK293T cells (Extended Data Fig. 6A). PLA shows that GFP-Tet3FL interacts with FLAG-GLud1, whereas we could not observe an interaction in the control experiment with GFP and FLAG-Glud1 (Fig. 14A). To study whether Glud1 influences Tet3 activity, we compared the levels of mC, hmC, fC and caC of HEK293T cell DNA of cells that were either coexpressing GFP-Tet3FL and FLAG-Glud1 or only GFP-Tet3FL after 24h and after 48h (Fig. 14B). Since the GFP-Tet3FL/FLAG-Glud1 coexpressing cells did not express as much of the GFP-Tet3FL as the cells that were only expressing GFP-Tet3FL, we normalized the hmC, fC and caC levels to the GFP-signal (relative fluorescent unit (RFU)) that we determined using fluorescence-based flow cytometry. The overall mC levels were lower for only GFP-Tet3FL-expressing HEKcells, but the difference was not significant (24 h: unpaired t-test, n=3 GFP-Tet3FL, n=3 GFP-Tet3FL+FLAG-Glud1, p-value 0.2507; 48h: unpaired t-test, n=4 GFP-Tet3FL, n=4 GFP-Tet3FL+FLAG-Glud1, p-value 0.2158). However, for hmC and fC, we measured significant differences (24h: unpaired t-test with Welch's correction, n=3 GFP-Tet3FL, n=3 GFP-Tet3FL+FLAG-Glud1, p-values 0.0148 (hmC) and 0.0053 (fC); 48h: unpaired t-test with Welch's correction, n=4 GFP-Tet3FL, n=4 GFP-Tet3FL+FLAG-Glud1, p-value 0.0073 (hmC) and 0.0494 (fC)). The data show that FLAG-Glud1 functionally interacts with GFP-Tet3FL and that it strongly stimulates GFP-Tet3 activity. A similar stimulation was also observed in experiments with Tet1FL and Tet2FL showing that the glutamate supply path provides α-ketoglutatrate not only for Tet3 (Extended Data Fig. 6B/C).

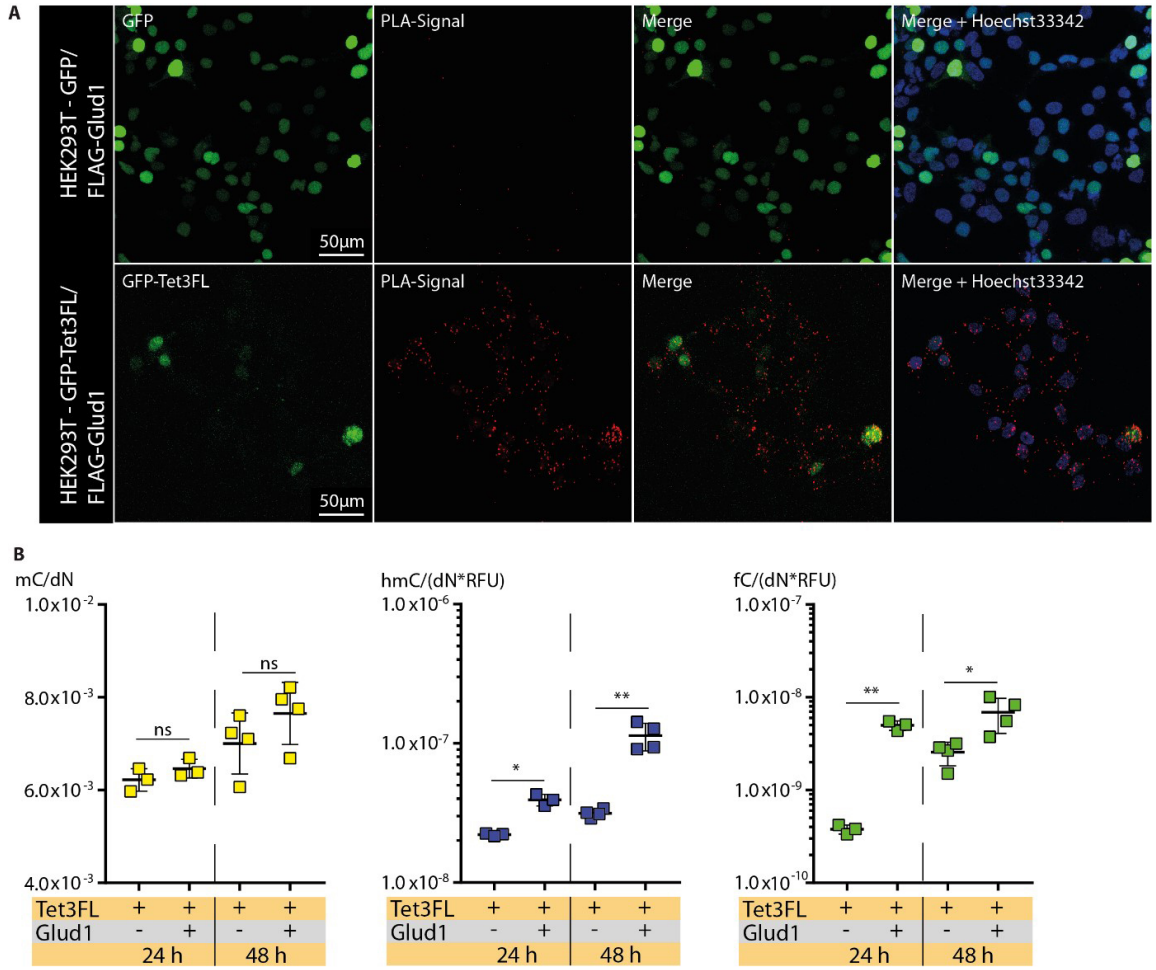


Figure 14: Functional interaction of Glud1 and Tet3. . (A) Proximity ligation assay in HEK293 T cells either coexpressing GFP-Tet3FL and FLAG-Glud1 or GFP and FLAG-Glud1. (B) Cytosine modification levels in HEK293T cells, co-transfected with either GFP-Tet3FL and FLAG-Glud1 or GFP-Tet3FL and an empty vector (pESG iba45) after 24 h and 48 h; +: transfected, -: not transfected; ns = not significant; p-values *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$; plot presents mean with standard deviation (SD). Each data point represents one biological replicate for each modification.

In order to obtain physiologically more relevant evidence of the Glud1-based Tet3 stimulation, we next studied the hmC levels in mouse hippocampal slices (n=5) with and without KCl-induced neuronal depolarization (25 mM versus 3 mM KCl, for 6 h, Fig.15A) in presence and absence of the Glud1 inhibitor R16237. To test the inhibitor, we first co-transfected HEK293T cells with GFP-Tet3 and FLAG-Glud1 and could see that R162-treated cells showed a 60% lower hmC and a 40% lower fC level, compared to the untreated cells (n=2, Fig. 15B, Extended Data Fig. 6B). The data show that the in-

hibitor can effectively reduce Glud1 activity and that the stimulatory effect of Glud1 on Tet3 activity depends on the catalytic activity of Glud1 and not only on its mere presence. Next, we tested the effect of R162 in the hippocampal slice experiment. While the mC levels remain unchanged (Fig. 15C), we observed that the hmC levels increase significantly (Fig. 15C) when the neurons are depolarized with 25 mM KCl (RM one-way ANOVA Tukey's multiple comparisons test, p-value 0.0007). This hmC-raising effect could be again blocked with R162 (RM one-way ANOVA Tukey's multiple comparisons test, p-value 0.0094) showing that the catalytic activity of Glud1 is needed to raise the hmC levels in response to neuronal depolarization. This is in perfect agreement with the idea that the α-ketoglutarate is rate limiting.[256,417]

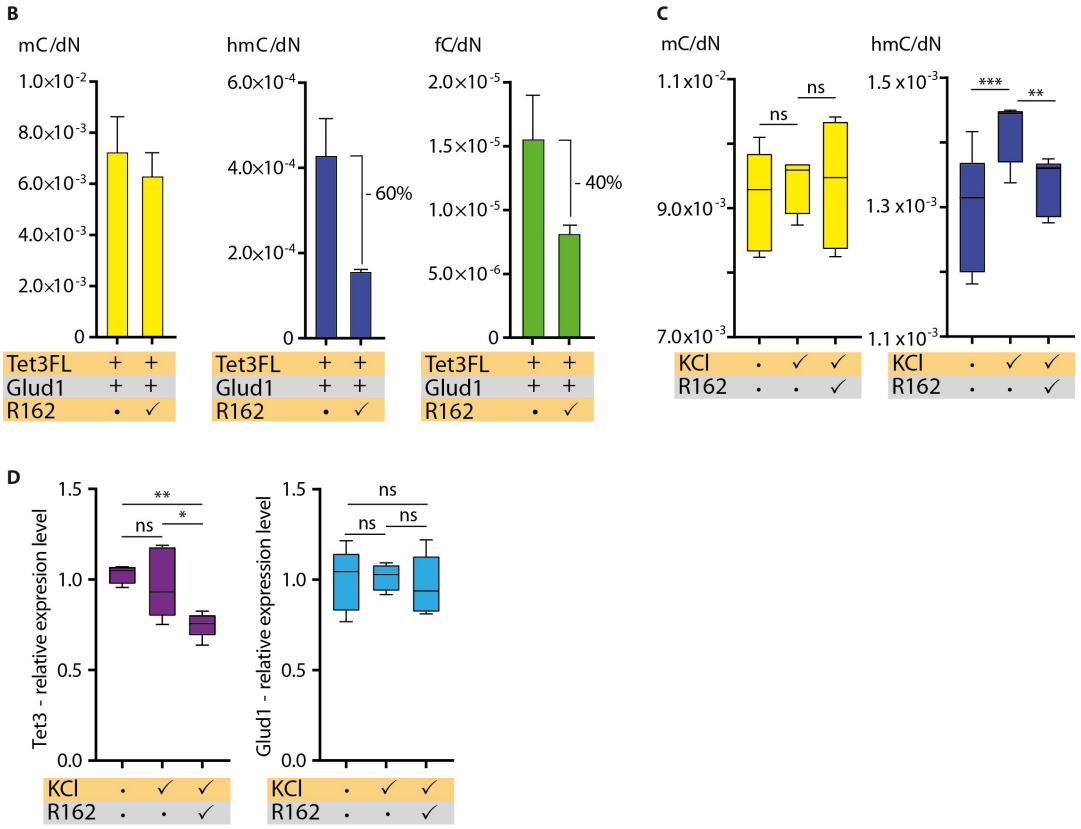


Figure 15: Interaction of Glud1 and Tet3 in the hippocampus. (B) HEK293T (n=2) cells were co-transfected with GFP-Tet3FL and FLAG-Glud1 and fed with 20 µM Glud1 inhibitor R162. Bar graph presents mean and SD. (C) – (D) Tukey boxplot; hippocampal neurons (n=5) were depolarized in 25 mM KCl buffer with and without 20 µM R162. Neurons in physiological buffer served as control. (C) Levels of mC and hmC and (D) relative expression levels of Tet3 and Glud1 (RT-qPCR data). •no addition, ✓addition ; ns = not significant; p-values *: $p < 0.05$, **: $p< 0.01$, ***: $p < 0.001$.

8.6 Effect of TCA cycle intermediates on Tet3 activity

The observed association of Tet3 with Mdh, Got1/2, CS and Aco in absence of Idh suggests the accumulation of isocitrate and to some extent oxaloacetate in close proximity to Tet3. In order to investigate how these metabolites may affect Tet3 activity, we incubated mC-containing DNA with GFP-Tet3cd loaded agarose beads in a buffer system containing ascorbic acid and DTT, for keeping the iron in the active Fe^{2+}-state (Fig. 16A). While the DNA oligonucleotide (Fig. 16A 1-1) is stable in water under these conditions (Fig. 16B), we noted to our surprise upon addition of Tet3cd (no α-ketoglutarate) severe degradation of the oligonucleotide (MALDI-TOF analysis, Fig. 15C). This degradation was observed with other DNA sequences and other Tet enzymes as well (Extended Data Fig. 8). Deeper analysis of the degradation reaction with an oligonucleotide, in which we replaced the T bases with U in order to enable detection of also mC-deamination reactions, showed predominant hydrolysis of the phosphodiesters and cleavage of glycosidic bonds. These degradation pathways are explainable with the known chemistry of high-spin Fe^{2+}-centers, that are the reactive centers in active Tet-enzymes. These centers are Lewis acidic, which causes phosphordiester hydrolysis[422], and redox active, which can explain the observed glycosidic bond cleavage (Fig. 16C).[423,424] Once we added α-ketoglutarate to the model study, DNA degradation was fully stopped. Instead, the expected oxidation of mC to hmC and fC is detected (Fig. 16D). The observation of DNA degradation in the absence of α-ketoglutarate confirms that the reactive Fe^{2+} center needs full coordination to prevent potential DNA damage.[424–426] It is not clear which molecules coordinate in vivo to the Fe^{2+}-center in the absence of α-ketoglutarate, when the enzyme is in an inactive mode, but it is tempting to speculate that the TCA metabolites are to some extent involved. Indeed, when we replaced in these studies the α-ketoglutarate by either oxaloacetate (Fig. 16E) or isocitrate (Fig. 16F), we detected in both cases neither oxidation of mC nor DNA degradation. Deeper analysis of the degradation reaction with an oligonucleotide, in which we replaced the T bases with U in order to enable detection of also mC-deamination reactions, showed predominant hydrolysis of the phosphodiesters.

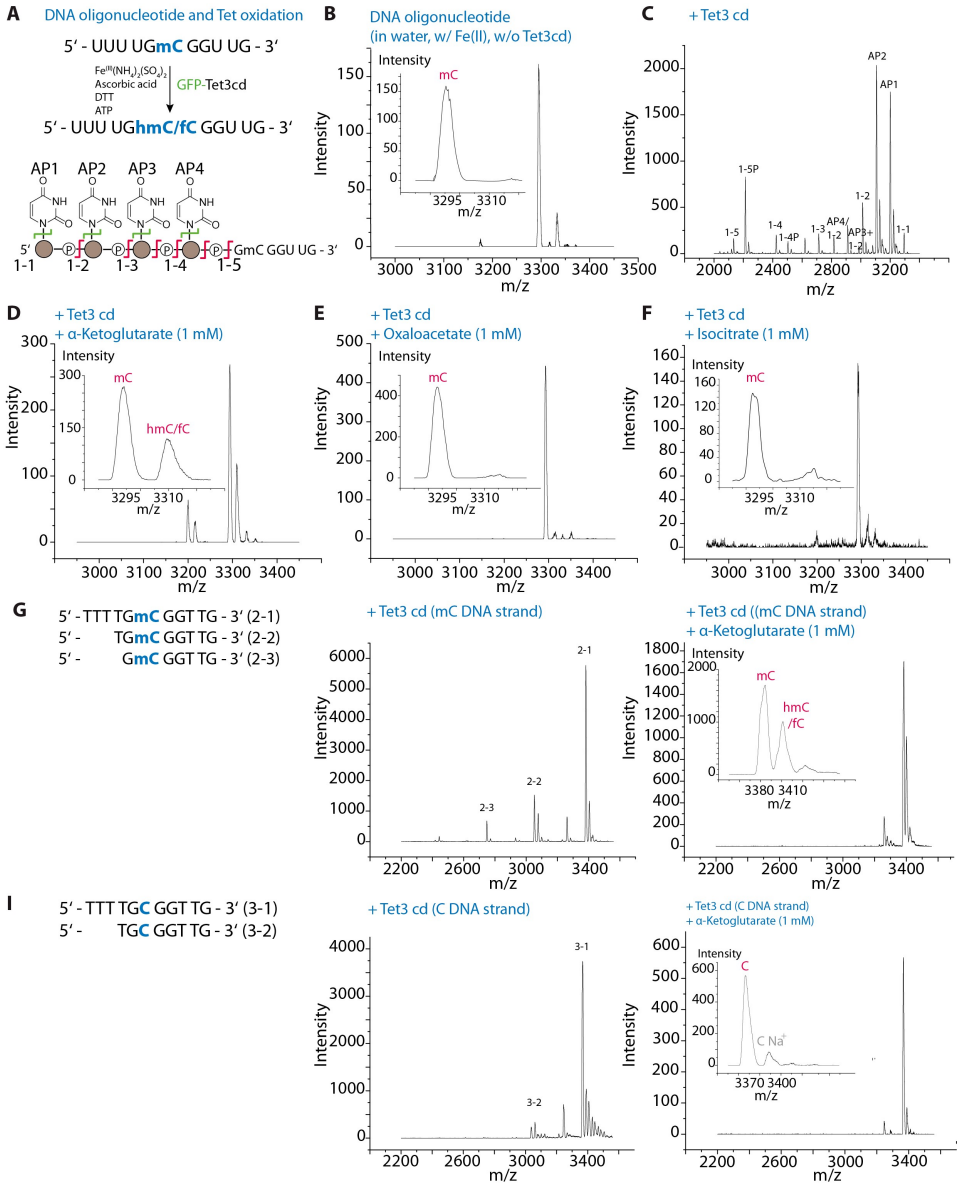


Figure 16: Tet activity influenced by metabolites (I). (A) DNA-oligonucleotide, including degradation pathways (AP = abasic site), and buffer ingredients for in vitro activity tests. (B) – (F) Metabolites prevent Tet3-induced decomposition of mC-containing DNA. MALDI-TOF analysis shows peaks of mC- (3294.5 m/z) and hmC/fC-containing DNA (3310.8 m/z). (G) – (I) In vitro assay with mC and only C-containing DNA oligo. DNA oligo 5'-TTT TGmC GGT TG-3' MALDI-TOF m/z mC: 3381.6; hmC/fC: 3397.6; 2-1 3381.6 Da; 2-2 3051 Da; 2-3 2746.3. DNA oligo 5'-TTT TGC GGT TG-3' MALDI-TOF m/z 3367.5.

Ligand docking simulation using Maestro 10.7 (Extended Data Fig. 9A) with the crystal

structure of the catalytic domain of Tet2 (PDB ID: 4NM6) confirm that both metabolites are able to bind in the active site (Extended Data Fig. 9B). If binding in the active site occurs, this has to induce a competitive inhibitory effect and indeed, when we added increasing amounts of both metabolites to the oxidation assay containing the Tet3-beads, the mC-containing oligonucleotide, and 1 mM α-ketoglutarate, we detected for both metabolites at millimolar concentrations an inhibitory effect (Fig. 16A and Extended Data Fig. 10) quite similar to what has been observed for 2-hydroxyglutarate[281] and recently also for succinate and fumarate.[285] We next investigated, if isocitrate influences the mC to hmC oxidation in cellulo (Fig. 15H, Extended Data Fig. 11). To this end, HEK293T cells were transfected with Tet3FL and 10 mM of isocitrate was added to the medium 4 h after transfection. Two days after transfection, we analysed the levels of mC, hmC, fC and caC and detected a significant reduction of the hmC and caC values (unpaired t-test, n=3, p-value = 0.0480 and 0.0493, resp.). The mC levels in contrast did not change.

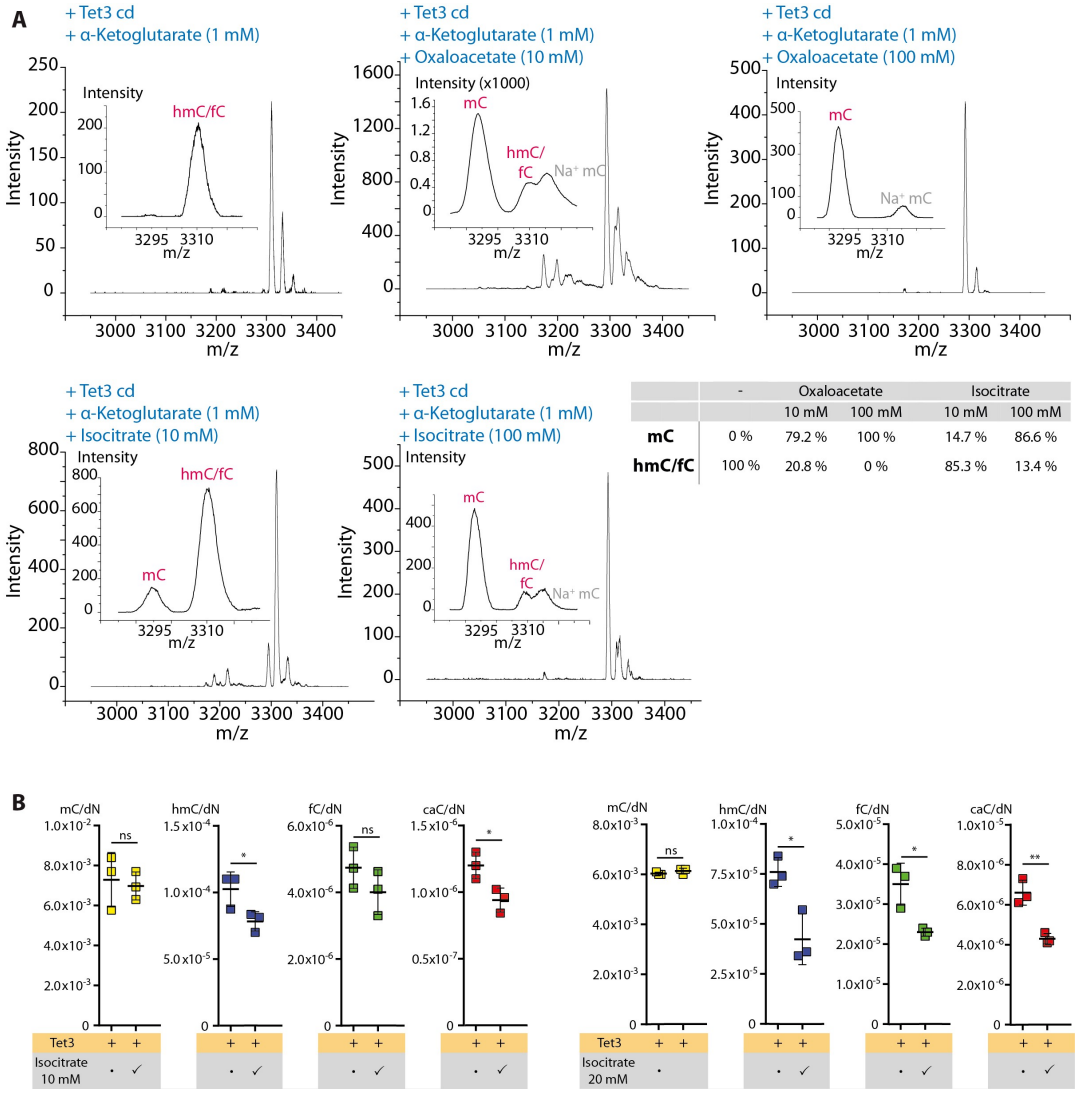


Figure 17: Tet activity influenced by metabolites.). (A) Oxaloacetate and isocitrate decrease Tet3 activity in vitro. (B) Isocitrate lowers Tet3 activity in HEK293T cells, which were transfected with GFP-Tet3 and fed with 10 mM or 20 mM isocitrate (+ ✓). Cells that were transfected, but not fed with isocitrate (+ •) served as controls. ns = not significant; Unpaired t-test (n=3) *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$; plot presents mean with SD. Each data point represents one biological replicate for each modification.

8.7 Conclusion

Tet enzymes oxidize mC to hmC and fC in the cell nucleus and for this oxidation they need oxygen and α-ketoglutarate as essential co-substrates.[78–81] α-ketoglutarate is an intermediate of the TCA cycle and hence manufactured predominantly in the mitochondrium. This generates a supply problem. Several recent studies show that the α-ketoglutarate supply is rate limiting for Tet oxidation.[256,417] In this study, data from proteomics experiments, endogenous co-IPs and PLAs show that in neurons Tet3 is directly associated in the cell nucleus with some of the key TCA proteins involved in the biosynthesis of α-ketoglutarate. In addition, we find Tet3 associated with Got1/2 and Glud1. Glud1 uses NAD+ to oxidize and hydrolyse glutamate directly to α-ketoglutarate. Glutamate is a major neurotransmitter and glutamatergic neurons are known to be the dominant neurons that show synaptic plasicity, which in turn is a known trigger of changing hmC levels.[412] Although glutamate is an abundant amino acid, it is tempting to speculate that the direct Glud1-Tet3 interaction establishes a functional link between brain function and epigenetics. We show that Glud1 and Tet3 are directly associated with each other and that this connection stimulates Tet3, leading to a substantial increase of the hmC-levels. Studies with the Glud1 inhibitor R162 show that this is caused by the direct biosynthesis of α-ketoglutarate close to the active site of Tet3. This will lead to a strongly increased effective molarity of the co-substrate. When α-ketoglutarate is rate limiting[256,417], we have to ask the question what is bound to the reactive Fe^{2+}-center in its absence. We show that in the absence of a potential ligand, the Fe^{2+}-center has the tendency to damage DNA.[427] This degrading effect is totally stopped in the presence of oxaloacetate and isocitrate, which are the biosynthesis intermediates that are formed by Got2 and the other TCA proteins found associated with Tet3. Although we are unable to study in vivo, which ligand is bound in the absence of α-ketoglutarate to the Fe^{2+}-center, it is conceivable that these TCA intermediates with their chemical structures so close to α-ketoglutarate provide the needed protective function.

8.8 Methods

8.8.1 Cell culture and transfection

The HEK293T cells (ATCC) were cultivated at 37 °C in water saturated, CO_2-enriched (5%) atmosphere. DMEM (10% FBS) or rpmI (10% FBS) was used as growing medium. When reaching a confluence of 70% to 80% the cells were passaged. The transfection was performed in p150-petri dishes. Five million cells were used in 20 mL of medium. After seeding, the cells were incubated at previously described cultivation conditions for 24 h to reach a confluence of 40% to 80%. 10 μg of DNA and 30 µL of the transfection reagent jetPRIME® purchased from Polyplus Transfection were used as described by the manufacturer. All expression plasmids for GFP-Tet and CXXC-GFP constructs were described previously.[265] To increase the transcription rate in transfected cells, 4 h after transfection the medium was removed and sodium butyrate (final conc. 4 mM) treated medium was added. After 48 h, the cells were harvested and immediately used for nuclear extract preparation.

8.8.2 Nuclear extract preparation

Nuclear lysate of HEK293T cells was prepared as described previously.[194] The complete adult mouse brains (Mus musculus, BL6J wt, female, 110 days old) were lysed according to a protocol that was published before.[428] The resulting nuclear extract was then treated with 25 U benzonase (Merck Millipore) for 30 min on ice and subsequently centrifuged for 15 min at 15000 rpm. The supernatant containing the nuclear lysate was transferred to a fresh tube. A Bradford protein assay (Bio-Rad) was performed according to the manufacturer´s instructions to measure the protein concentration.

8.8.3 GFP-Tet saturated co-immunoprecipitation

20 µl of anti-GFP beads (Chromotek) were washed three times with wash buffer (10 mM Tris/Cl pH 7.5; 150 mM NaCl; 0.5 mM EDTA) and then incubated for 15 min on ice with nuclear extract of GFP-Tet overexpressing HEK293T cells. To ensure the saturation of

the beads with the GFP fusion construct, different amounts of lysate were tested and monitored using a Tecan Reader. The GFP-Tet loaded beads were then washed twice with 150 mM NaCl solution containing 10 mM HEPES, pH 7.5. Another two wash steps with 1 M NaCl solution containing 10 mM HEPES, pH 7.5 and two wash steps with Lysis Buffer C (20 mM Hepes pH = 7.5, 420 mM NaCl, 2 mM $MgCl_2$, 0.2 mM EDTA, 20% (v/v) glycerol) followed. The saturated GFP-Tet beads were subsequently incubated with 200 µg nuclear brain extract for 15 min on ice. Following, they were washed twice with wash buffer. To elute the bound proteins, 50 µl of 200 mM glycine pH 2.5 were added and the solution was vortexed for 30 seconds. In order to gain more yield the elution step was repeated.

8.8.4 Co-IP of endogenous Tet3 protein

Co-IP of endogenous Tet3 was performed, using nuclear brain extract. When the interactors were subsequently analysed by mass spectrometry, 250 µg of nuclear brain extract, 1 µg of antibody (anti-Tet3 antibody (Abiocode Tet3 (N1) antibody R1092-1, rabbit polyclonal) or anti-goat antibody as negative control (SigmaAldrich G4018, rabbit polyclonal) and 10 µL of Dynabeads Protein G (ThermoFisher) were used per replicate; for the analysis by Western Blot four times the amount of nuclear brain extract, antibodies and Dynabeads Protein G were used. The nuclear brain extract was incubated with the antibody for 1 h at 4 °C on a tube rotator and the Dynabeads Protein G were washed three times with wash buffer (10 mM Tris pH 7.5, 150 mM NaCl, 0.5 mM EDTA). Afterwards, the Dynabeads were added to the lysate and incubated for 30 min at 4 °C. After incubation, the beads were washed three times with wash buffer. Last, proteins were eluted with 30 µL of 1% (v/v) formic acid for 15 min at room temperature (mass spectrometry analysis) or with 50 µL of SDS loading buffer (50 mM Tris pH 6.8, 100 mM DTT, 2% (w/v) SDS, 10% (v/v) glycerol, 0.1% (w/v) bromphenolblue) for 10 min at 70 °C.

8.8.5 LC-MS/MS analysis

Samples for the mass spectrometer were reduced by the addition of 100 mM TCEP and subsequent incubation for 1 h at 60 °C on a shaker at 650 rpm. They were then alkylated

by adding 200 mM iodoacetamide and incubating for 30 min at room temperature in the dark. Following, the samples were digested with 0.5 µg trypsin (Promega) at 37 °C for 16 h. Afterwards, they were incubated for 5 min at 100 °C and subsequently 1 mM phenylmethylsulphonylfluoride was added. StageTips were utilized to purify the samples for mass spectrometry.[429] The samples were analyzed with an UltiMate 3000 nano liquid chromatography system (Dionex, Fisher Scientific) attached to an LTQ-Orbitrap XL (Fisher Scientific). They were desalted and concentrated on a µ-precolumn cartridge (PepMap100, C18, 5 µM, 100 Å, size 300 µm i.d. x 5 mM) and further processed on a custom made analytical column (ReproSil-Pur, C18, 3 µM, 120 Å, packed into a 75 µm i.d. x 150 mM and 8 µm picotip emitter). The samples were processed via a 127 min multi-step analytical separation at a flow rate of 300 nl/min. The gradient with percentages of solvent B was programmed as follows: 1% for 1 minute; 1% - 4% over 1 minute; 4% - 24% over 100 minutes; 24% - 60% over 8 minutes; 60% - 85% over 2 minutes; 85Mass spectrometric analysis was done with a full mass scan in the mass range between m/z 300 and 1700 at a resolution of 60000. Following this survey scan five scans were performed using the ion trap mass analyzer at a normal resolution setting and wideband CID fragmentation with a normalized collision energy of 35. Signals with an unrecognized charge state or a charge state of 1 weren't picked for fragmentation. To avoid supersampling of the peptides, an exclusion list was implemented with the following settings: after 2 measurements in 30 seconds, the peptide was excluded for 90 seconds.

8.8.6 LFQ data processing

The MaxQuant software (version 1.5.0.25) was used for LFQ. Quantification was performed with four biological replicates for Tet3-saturated co-IP. GFP alone served here as control. Three biological replicates were used for the endogenous co-IP and the co-IP with anti-goat antibody served as control. The Andromeda search engine was used in combination with uniprot databases (Mus musculus). A maximum of two missed cleavage sites was allowed. The main search peptide tolerance was set to 4.5 ppm. Carbamidomethyl (C) was set as static modification. Variable modifications were Acetyl (Protein N-term) and Oxidation (M). The LFQ algorithm was applied with default settings. The option "match between runs" was also used. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE48

partner repository with the dataset identifier PXD004518. LFQ data was analyzed with the Perseus software (version 1.5.0.9). The LFQ intensities were log transformed and only proteins identified in at least three samples were retained. As one of the GFP control quadruplicates contained only 64 proteins instead of >400, this replicate was removed from the data set. Gene ontology analyses were performed with the Database for Annotation, Visualization and Integrated Discovery (DAVID Bioinformatics Resources 6.7).

8.8.7 Western Blotting

When the GFP-Tet saturated co-IP was analyzed by western blotting (tank (wet) electrotransfer) instead of mass spectrometry, the proteins were eluted at 70 °C for 15 min with SDS loading buffer (100 mM Tris pH = 6.8, 100 mM DTT, 2% (w/v) SDS, 10% (v/v) glycerol, 0.05% (w/v) bromphenol blue). The samples were loaded on a 4-15% precast polyacrylamide gel (Bio-Rad) and MagicMark XP Standard (ThermoFisher) was used as a protein standard. The gel was run at constant 150 V for 70 min. For blotting, we used a PVDF blotting membrane (GE Healthcare) and pre-cooled Towbin blotting buffer (25 mM Tris, 192 mM glycine, 20% (v/v) methanol, 0.038% (w/v) SDS). The membrane was activated for 5 min in abs. methanol; the filter paper was equilibrated for 15 min in Towbin buffer and the precast gel was equilibrated for 5 min after running in Towbin buffer. Western blotting was performed at 4 °C for 11 h at constant 35 V. After blotting, the PVDF membrane was blocked for 1 h at room temperature using 5% (w/v) milk powder in TBS-T (20 mM Tris/Cl pH = 7.5, 150 mM NaCl, 0.1% (v/v) Tween-20). The primary antibodies were diluted 1:1000 in 15 mL of 5% (w/v) milk powder in TBS-T. After blocking, the membrane was cut in three parts (part1: proteins < 40 kDa; part2: proteins > 40 kDa, but < 120 kDa; part3: proteins > 120 kDa). The blocking suspension was discarded and the diluted primary antibodies were added (anti-GFP antibody (CellSignaling 2555S, from mouse) to part1 and part3; anti-Aco2 antibody (Abcam ab110321, from mouse) or anti-Glud antibody (Thermo Scientific PA5-19267, from goat) to part2) for 11 h at 4 °C. After incubation, the primary antibodies were discarded and the membrane was washed three times with TBS-T. HRP-conjugated secondary antibodies (all from Sigma-Aldrich) were diluted 1:10000 in 5% (w/v) milk powder in TBS-T and added for 1 h at room temperature. Afterwards, the membrane was washed two times with TBS T and one time with TBS

before SuperSignal West Pico Chemiluminescent Substrate (Thermo Scientific) was used for imaging. The western blots were performed twice (two biological replicates), yielding the same result. When the endogenous co-IP was analyzed by western blotting, the same conditions were used, but the membrane was not cut after blotting. To test the specificity of the anti-Tet3 antibody used in the endogenous IP (Abiocode R1092), the PVDF membrane was incubated with same anti-Tet3 antibody (Abiocode R1092). To avoid detection of the antibody fragments, TidyBlot from Biorad (HRP conjugated Western blot detection reagent) was used. To analyse the Tet3-Glud interaction, the PVDF membrane was incubated with anti-Glud antibody; to analyse the Tet3-Aco2 interaction, the PVDF membrane was incubated with anti-Aco2 antibody. The western blots were performed twice (two biological replicates), yielding the same result.

8.8.8 Immunohistochemistry

Immunohistochemistry (IHC) experiments were performed as previously described13 with minor modifications. In brief, 12 μm thick coronar cryo-sections of snap-frozen adult mouse brain were fixed on slides using 4% paraformaldehyde (PFA) in 0.1 M phosphate buffered solution, pH 7.4 (0.1 M PB). After three times washing with 40 μL 0.1 M PB, the slices were incubated over night with primary antibody solution at 4 °C in a humidity chamber. Anti-Tet3 antibody (Biomol AC-R1092-1, from rabbit or SigmaAldrich HPA050845, from rabbit), anti-Glutamate-Dehydrogenase antibody (Thermo Scientific PA5-19267, from goat) and anti-Aconitase2 antibody (Abcam ab110321, from mouse) were used as primary antibodies. The primary antibodies were diluted (anti-Tet3 1:200, anti-Glud 1:100, anti-Aco2 1:200) in 0.1 M PB, containing 5% (v/v) blocking reagent (Chemiblocker, CB, Millipore) and 0.3% (v/v) Triton-X solution. For the negative controls, no primary antibodies were added. For secondary detection we used Alexa488-anti-rabbit (1:800, Cell Signaling Technologies), Cy2-anti-goat (1:200, Jackson ImmunoResearch) and Cy3-anti-mouse (1:400, Jackson ImmunoResearch) diluted in 0.1 M PB, containing 3% (v/v) CB. Cell nuclei were stained with Hoechst 33342 (5 μg/mL), which was applied for 5 min in the dark at room temperature. After mounting (Mountant Permafluor, ThermoScientific), the slices were analyzed using a Leica SP8 confocal laser scanning microscope (Leica, Wetzlar). The IHC experiments were performed in two biological replicates, yielding the same result.

8.8.9 Proximity Ligation Assay

The proximity ligation assay (PLA) was carried out as described[416] using Duolink InSitu Orange Starter Kit (Sigma) with slight modifications. The slices were rehydrated and fixed as described above. Before the anti-Glud and anti-Tet3 primary antibodies or the anti-Aco2 and anti-Tet3 antibodies or the anti-Got2 (Thermo Scientific MA5-15595, from mouse) and the anti-Tet3 primary antibodies were added, the slices were blocked with 10% CB. Anti-Tet3, anti-Glud and anti-Aco2 antibodies were diluted as decribed in the IHC; the anti-Got2 antibody was diluted 1:200 in 0.1 M PB containing 5% CB and 0.3% Triton-x 100. The primary antibodies were added over night at 4 °C in a humidity chamber; for the negative control no primary antibody was added. The next day, the slices were washed two times with Washbuffer A (Sigma). For the PLA Probe Solution, 10 µL of PLA Probe (+) Anti-Rabbit (Sigma), 10 µL of PLA Probe ( ) Anti-Mouse (Sigma) (for anti-Aco2 and anti-Got2) or 10 µL of PLA Probe (-) Anti-Goat (Sigma) (for anti-Glud) and 30 µL of Antibody Diluent were mixed and added. The slices were incubated in a pre-warmed humidity chamber for 1 h at 37 °C. In the next step, the slices were washed two times with Washbuffer A and the Ligation solution, which had been prepared previously by mixing 1.0 µL of 1 U/µL Ligase, 10 µL of Ligation Stock (all from Sigma) and 29 µL of bidest. H_2O, were added. After 30 min of incubation in a pre-warmed humidity chamber at 37 °C, the Ligation solution was tapped off and the slices were washed two times with 100 µL of Washbuffer A. For the amplification reaction, 0.5 µL of Polymerase 10 U/µL, 10 µL of Amplification Stock (all from Sigma) and 39.5 µL of bidest. H_2O were mixed and added. The slices were incubated for 100 min at 37 °C in a pre-warmed humidity chamber. The slices were washed two times with Washbuffer B (Sigma), one time with 0.01% Washbuffer B and one time with 0.1 M PB. Cell nuclei were stained with Hoechst 33342 solution. After mounting (Mountant Permafluor, ThermoScientific), the slices were analyzed using a Leica SP8 confocal laser scanning microscope (Leica, Wetzlar). The PLA was performed once for Glud as well as Aco2 and twice for Got, yielding the same results. The negative controls did not show any signal.

8.8.10 Isocitrate in vivo experiment

The experiment was done in three biological replicates and technical triplicates each. 3×105 HEK293T cells were seeded per 6-well. 24 h after seeding, the cells were transfected with 600 ng of vector DNA coding for GFP-Tet3 (full-length construct). 4 h after transfection, the medium was changed and 10 mM of isocitrate (DL-isocitric acid trisodium salt, Sigma-Aldrich) were added from a 1 M stock solution to the medium. 24 h after transfection, the medium was changed again and 10 mM of isocitrate were added. 48 h after transfection the genomic DNA was isolated and the subsequent analysis of the mC-, hmC-, fC- and caC-levels were performed as described earlier. 3 5 µg of genomic DNA were analysed per technical replicate. Levels of 8-oxoG were analysed to show that random oxidation processes do not impair the result. Statistical analysis was done using GraphPad Prism 7.0. The samples were tested for normal distribution (D'Agostino Pearson test) and a unpaired t-test was performed (n=3 for each condition; only biological replicates (means of the technical replicates) were taken into account). Variance was determined by standard deviation, showing similarity between the groups.

8.8.11 In vitro activity test

For the activity tests, approximately 10 million HEK293T cells were used. The cells were transfected with Tet3 catalytic domain (Tet3cd), according to the protocol above. The cells were harvested, centrifuged and lysed with 5 mL of RIPA buffer (Chromotek), supplemented with 250 U benzonase (Merck Millipore) and protease inhibitor (Roche). After 30 min on ice the suspension was centrifuged at 10000 x g and 4 °C for 15 min. The supernatant consisted of nuclear and cytoplasmic proteins. Per activity test, 100 µL of anti-GFP beads (Chromotek) were washed three times with wash buffer (10 mM Tris/Cl pH 7.5; 150 mM NaCl; 0.5 mM EDTA) and then incubated for 1 h at 4 °C with the supernatant. The GFP-Tet3cd loaded beads were washed with GFP wash buffer (Chromotek). For the second wash step GFP wash was supplemented with 2 mM $ZnSO_4$ and 10 µL Nuclease S1. The beads were incubated with this buffer for 30 min on ice. Another two wash steps with 1 M NaCl solution containing 10 mM HEPES, pH 7.5 and two wash steps with GFP wash buffer followed. The beads were centrifuged at 2500 x g and 4 °C for 5 min and the supernatant discarded. Next, the beads were

split into four test tubes and water and 0.1 mM Fe(II)$(NH_4)_2(SO_4)_2$ 6 x H_2O (final conc.; total volume 50 µL) was added. 1 mM α-ketoglutarate (Na^+ salt) were added to sample1; 1 mM oxaloacetate were added to sample2 and 1 mM isocitrate was added to sample3. To sample4, no cosubstrate was added. As an additional control, one sample (sample 6) was prepared without Tet and without co-substrate. 1000 pmol of an Oligo (5'–UUUUGmCGGUUG–3') was added and the reaction mixture incubated at 37 °C for 1.5 h. After centrifugation (15000 rpm), the supernatant was desalted using a desalting membrane and analyzed via MALDI-TOF. The *in vitro* activity test was performed five times, yielding the same result.

8.8.12 In vitro activity test with inhibitors

The samples were prepared as described above, but reaction buffer (50 mM HEPES pH=7.5, 100 mM NaCl, 2 mM Vitamin C, 1.2 mM ATP, 2.5 mM DTT) and 2000 pmol of the oligo were used instead of 1000 pmol and the samples were incubated at 3 h at 37 °C. 1 mM of α-ketoglutarate were added to each sample. To the positive control, neither oxaloacetate nor isocitrate was added. 10 mM and 100 mM of oxaloacetate or isocitrate were added in order to test the inhibitory potential. After incubation, the samples were centrifuged (rt, 5000 x g) and the supernatant was used for MALDI-TOF analysis. In order to calculate the relation of mC-containing oligo (substrate) to hmC/fC-containing oligo (product), OriginPro 2016G 64-bit software was used. The relation of mC to hmC/fC was calculated from MALDI-TOF analysis via peak integration. Peaks of the mC-containing oligo (3294.5 Da) were integrated from m/z 3290 – 3300 and peaks of the hmC/fC-containing oligo (3310 Da) were integrated from m/z 3305 – 3312. The in vitro activity test with inhibitors was performed twice, yielding the same result.

8.8.13 Co-expression of Tet3 and Glud1 in HEK293T cells

The experiment was done in three biological replicates. 3 x 105 HEK293T cells were seeded per 6-well. 24 h after seeding, the cells were transfected with 2.0 µg of vector DNA coding for GFP-Tet3 (full-length construct or catalytic domain) and for the co-transfection additionally with 2.0 µg of vector DNA coding for Glud1. 4 h after transfection, the medium was changed. 36 h after transfection, cells were harvested and

GFP-Tet3 levels were determined using fluorescence-activated cell sorting (BD FACS Cantor II, BD Biosciences, USA; FACS parameters FSC 132 V, SSC 407 V, GFP 308 V). Next, the genomic DNA was isolated and the subsequent analysis of the mC-, hmC- and fC- were performed in technical triplicates as described earlier.3 Levels of 8-oxoG were analysed to show that random oxidation processes do not impair the result. 5 µg of genomic DNA was analysed per technical replicate. Levels of hmC and fC of Tet3-Glud co-transfected cells were normalized to Tet3 levels of only Tet3-transfected cells. Statistical analysis was done using GraphPad Prism 7.0. The samples (including biological and technical replicates) were tested for normal distribution (D'Agostino Pearson test) and an unpaired t-test was performed (only biological replicates (means of the technical replicates), n=3 for each condition). Variance was determined by standard deviation, showing similarity between the groups. If R162 was applied (two biological replicates), it was added at a concentration of 20 µM 4 h after transfection.

8.8.14 Synthesis of the Glud1-inhibitor R162

Synthesis of the inhibitor R162 was done according to the literature.[430,431] Prior to use for the cell culture, the compound was purified via preparative HPLC (Nucleosil VP 250/10 C18 column from Macherey Nagel, 100% MeCN for 10 minutes).

8.8.15 Depolarization of hippocampal neurons

Animals:
4 – 5 weeks old C57BL6/J (Charles River, Sulzfeld, Germany) wildtype mice (4 x male, 2 x female) were used. All procedures concerning animals conform to the German animal protection laws and were approved by the local authority (Regierung von Oberbayern).

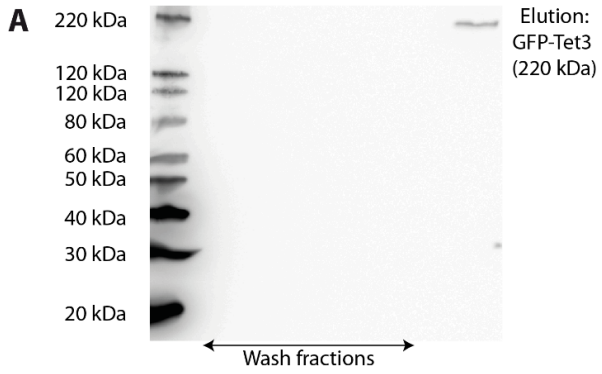
Hippocampal slices:
The mice (n=6) were euthanized by cervical dislocation. Acute transverse hippocampal slices (400 µm thick) were prepared as described previously.[432,433] In brief, the brain was removed, the hippocampi of each hemisphere were dissected and cut using a MX-TS tissue slicer (Siskiyou Cooperation, OR). The slices were collected in an oxygenated (95% O_2, 5% CO_2) physiological solution (118 mM NaCl, 3 mM KCl, 1 mM NaH_2PO_4,

25 mM $NaHCO_3$, 10 mM Glucose, 1.5 mM $CaCl_2$, 1 mM $MgCl_2$, 0.1% DMSO) at 37 °C until the hippocampi of all replicates were cut. Then the slices were distributed to three different conditions: oxygenated physiological solution, oxygenated 25 mM KCl solution (118 mM NaCl, 25 mM KCl, 1 mM NaH_2PO_4, 25 mM $NaHCO_3$, 10 mM Glucose, 1.5 mM $CaCl_2$, 1 mM $MgCl_2$, 0.1 % DMSO) and oxygenated 25 mM KCl solution supplemented with inhibitor R-162 (118 mM NaCl, 25 mM KCl, 1 mM NaH_2PO_4, 25 mM $NaHCO_3$, 10 mM Glucose, 1.5 mM $CaCl_2$, 1 mM $MgCl_2$, 20 μM R-162, 0.1 % DMSO). 6 – 10 slices were pooled for each replicate. After a 6 h incubation time, the slices were transferred into reaction tubes, snap frozen in liquid nitrogen and stored at -80 °C until use. DNA isolation and subsequent analysis of mC and hmC levels were performed as described earlier.12, 13 Levels of 8-oxoG were analysed to show that random oxidation processes do not impair the result. Measurements were done in technical duplicates or triplicates, depending on the amount of isolated genomic DNA. Per replicate, 1 μg of genomic DNA was analysed. Statistical analysis after Triple-Quad analysis of mC and hmC levels: Statistical analysis was done using GraphPad Prism 7.0. For n=6, p-values of 0.0091 (physiological conditions vs. 25 mM KCl) and 0.0193 (25 mM KCl vs. 25 mM KCl + 20 μM R162) were calculated using repeated measures (RM) one-way ANOVA Tukey's multiple comparisons test. However, we excluded one mice (mouse 6) from further analysis, since no change of the hmC levels upon KCl treatment could be observed and therefore it was not possible to study the effect of R162 on hmC dynamics. Three outliers within the technical replicates (mouse 3, physiological conditions, 0.00147 hmC/dN; mouse 5, 25 mM KCl 0.00113 hmC/dN and 0.00242 hmC/dN) were also excluded from further analysis. Assuming normal distribution of the mC and hmC levels for the mouse hippocampal region, RM one-way ANOVA was statistically significant for hmC: total number of values = 5 (p-value 0.0009) and there was significant matching (p-value 0.0031). There was no significant difference between physiological conditions and depolarization conditions with R162 (p-value 0.1583 physiological solution/25 mM KCl + 20 μM R162). For the mC-levels, the RM one-way ANOVA was not statistically significant (p-value 0.5156), but the matching was also effective (p-value 0.0023). Tukey's multiple comparisons test did not show significant differences between the different conditions (p-value 0.6054 physiological solution/25 mM KCl; p-value 0.5483 25 mM KCl/25 mM KCl + 20 μM R162; p-value 0.9945 physiological solution/25 mM KCl + 20 μM R162). For the statistical analysis only biological replicates (means of the technical replicates) were considered.

8.9 Contributions

Franziska Traube, Dr. Andrea Künzel, PD Dr. Stylianos Michalakis, Dr. Markus Müller and Prof. Dr. Thomas Carell designed the experiments. Dr. Andrea Künzel developed the saturated co-IP workflow and did the saturated co-IP, including the proteomics measurement. Franziska Traube and Dr. Andrea Künzel did the endogenous co-IPs, including the proteomics measurements, the Western blots, the IHC, the PLAs and analysed the data. Edris Parsa developed the expression protocol in HEK cells. Franziska Traube did the co-transfection experiments and the isocitrate feeding experiment. Franziska Traube, Edris Parsa, Thomas Wildenhof performed the *in vitro* Tet activity assays. Anna Geserich and PD Dr. Stylianos Michalakis performed and analysed the IHC and PLA experiments. René Rahimoff synthesized R162. Constanze Scheel, Victoria Splith and PD Dr. Stylianos Michalakis performed the depolarization of hippocampal neurons. Franziska Traube, Angie Kirchner and Katharina Iwan isolated DNA, performed Triple-Q experiments and analysed Triple-Q data. Michael Stadlmeier helped with the mass spectrometry studies. PD Dr. Stylianos Michalakis provided mouse samples. Dr. Markus Müller and Dr. Fabio Spada discussed and interpreted results. Prof. Jürgen Cox, helped with the statistical analysis of the proteomics data. Prof. Dr. Thomas Carell designed and supervised the whole study and interpreted data.

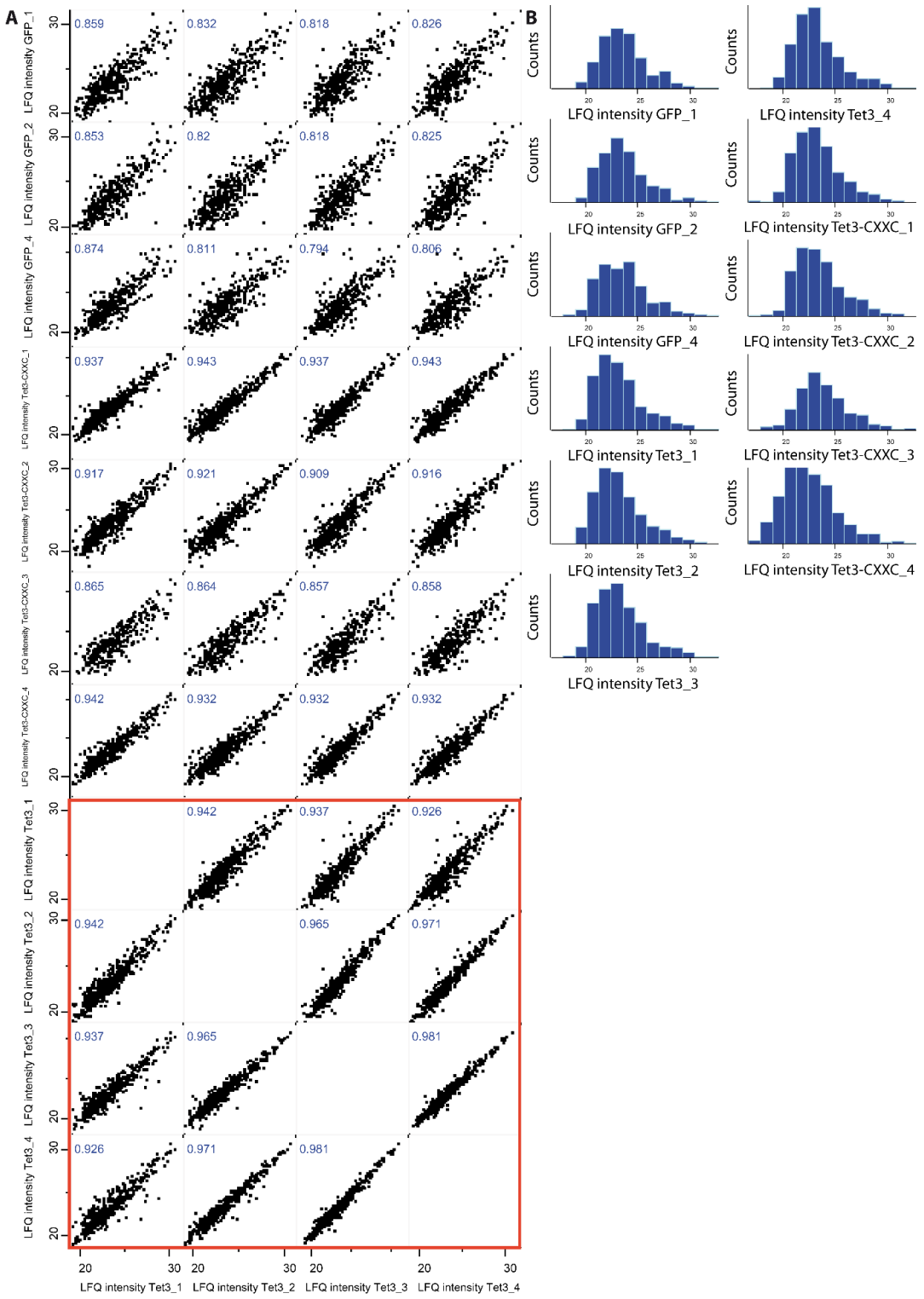
8.10 Supllementary Data



B

	Tet3FL	w/o Tet (neg. ctrl.)	Tet1cd (pos. ctrl.)
mC	85 %	100 %	20 %
hmC	13.5 %	-	61 %
fC	1 %	-	10 %
caC	0.5 %	-	9 %

Extended Data Fig. 1: Purification of GFP-Tet3. . (A) Complete Western Blot showing the steps of Tet3FL purification using an anti-GFP antibody to detect GFP-Tet3. In the wash fractions, no GFP-Tet3 could be detected. In the elution fraction, only the GFP-Tet3 full-length was detected. (B) Tet3 is active on the GFP nanobeads. The table shows the result of an in vitro activity assay in technical duplicates of Tet1 catalytic domain (Tet1cd) as a positive control and of Tet3 full length protein (Tet3FL) on the beads.



Extended Data Fig. 2: Data Quality. (A) Multi scatter plot comparing the LFQ intensities from the four biological replicates of the Tet3 experiment with itself (red rectangle) and the pull downs using beads loaded with GFP and Tet3 lacking the CXXC domain. Blue numbers indicate the Pearson correlation. (B) Histograms showing counts of label-free quantification (LFQ) intensities of the data set.

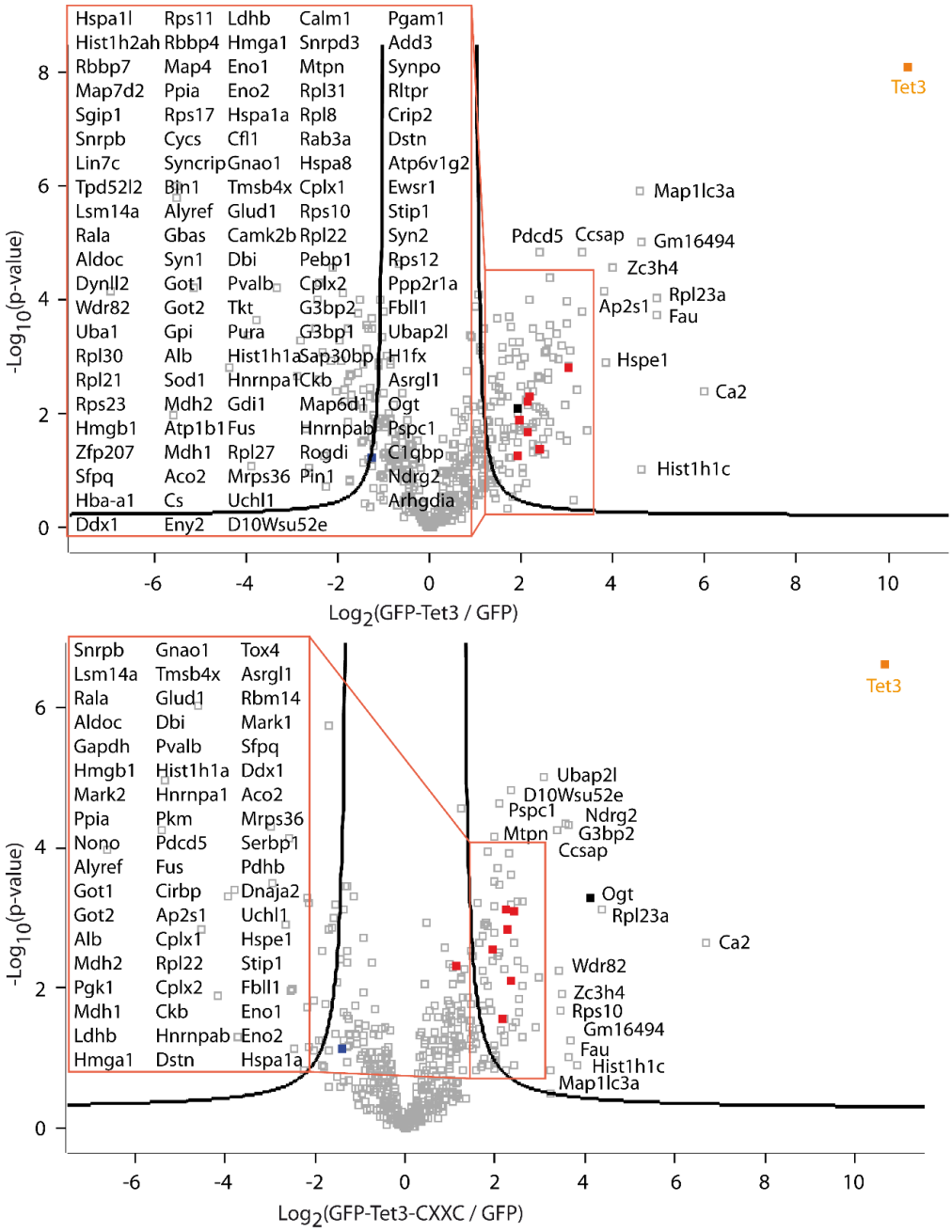
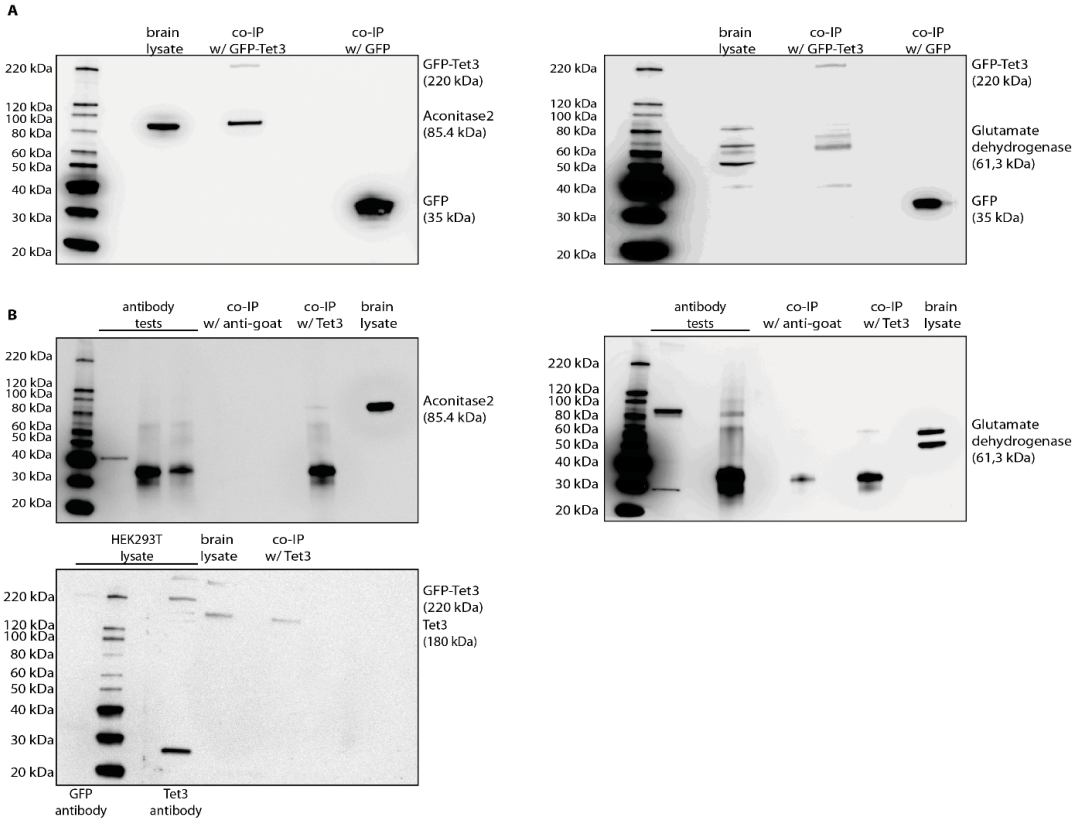
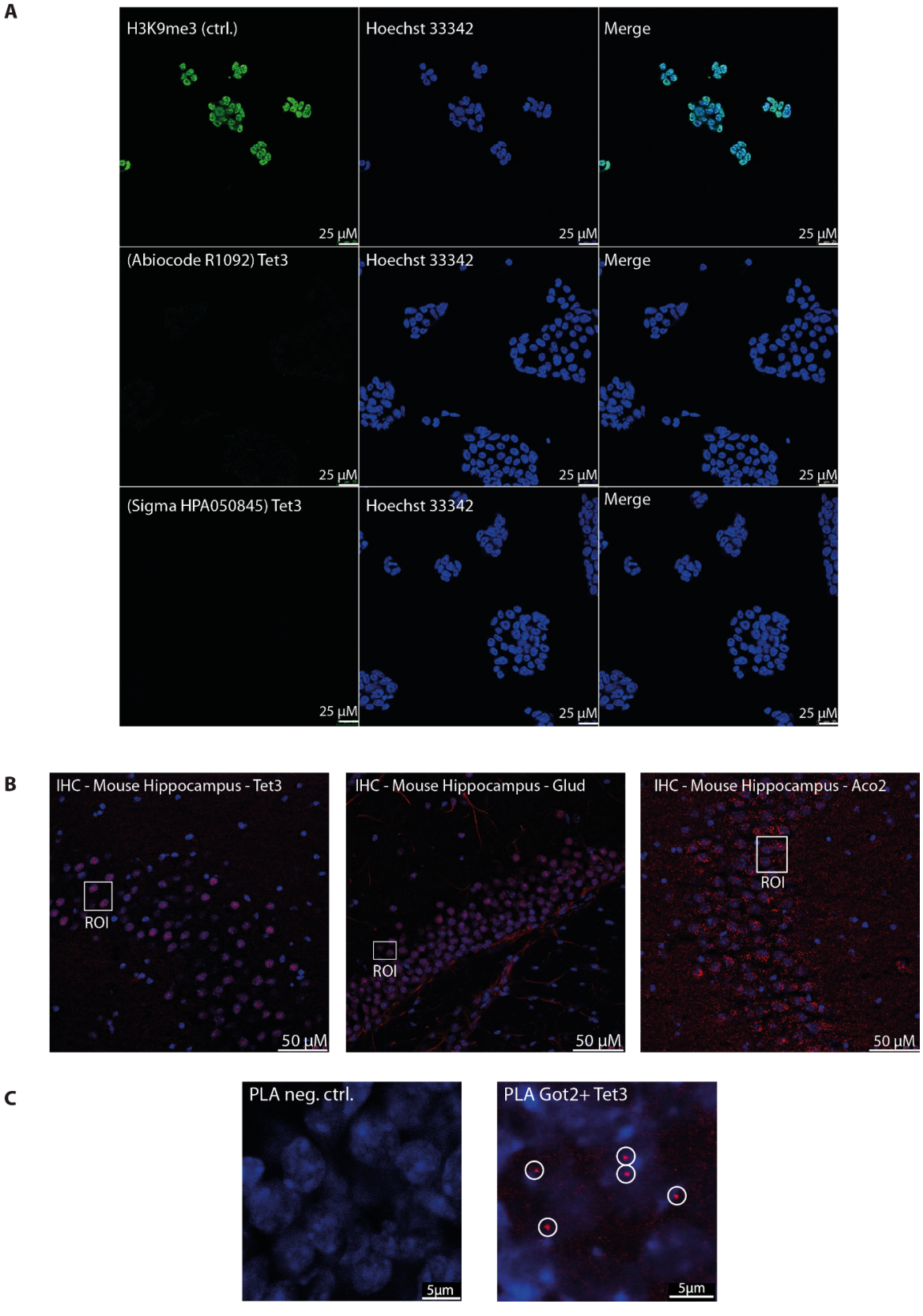


Fig. S2: Volcano plots of the interaction partners found in the modified Tet3-Co-Immunoprecipitation proteomics study in adult mouse brain for Tet3 and Tet3-CXXC (FDR = 0.05; s0 = 2).

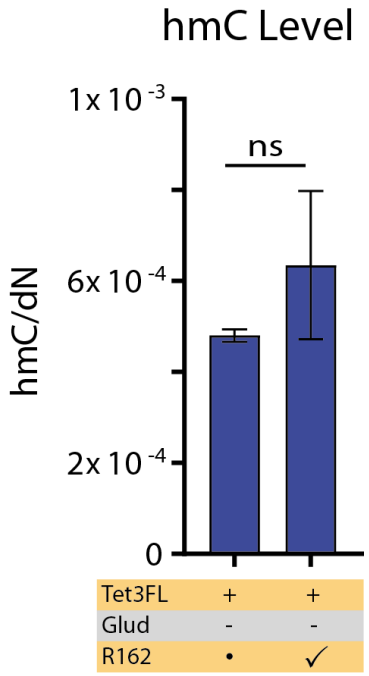
Extended Data Fig. 3: Volcano plots of co-immunoprecipitation (co-IP) experiments. Volcano plots of the interaction partners found in the modified Tet3-co-IP proteomics study in adult mouse brain for Tet3 and Tet3-CXXC (FDR = 0.05; s0 = 2) and interactors found in Tet3 endogenous co-IP study.



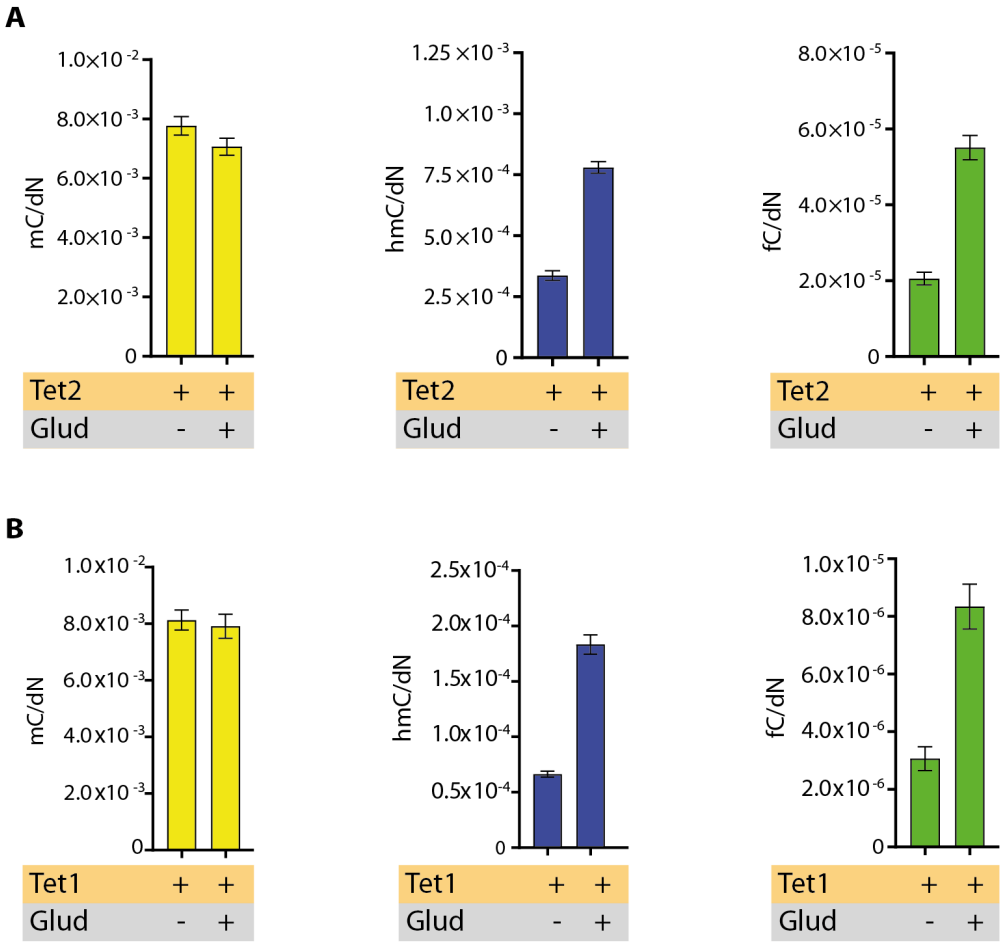
Extended Data Fig. 4: Complete Western Blots after Tet3 saturated and endogenous co-immunoprecipitation (co-IP). (A) Western Blots after Tet3 saturated co-IP using anti-aconitase2 antibody and anti-glutamate-dehydrogenase antibody. (B) Western Blots after Tet3 endogenous co-IP using anti-aconitase2 antibody and anti-glutamate-dehydrogenase antibody and anti-Tet3 antibody. Anti-GFP-antibody was also used to detect GFP-Tet3FL in HEK293T cell lysate.


A
H3K9me3 (ctrl.)
Hoechst 33342
Merge
(Abiocode R1092) Tet3
(Sigma HPA050845) Tet3
25 µM
B
IHC - Mouse Hippocampus - Tet3
IHC - Mouse Hippocampus - Glud
IHC - Mouse Hippocampus - Aco2
ROI
50 µM
C
PLA neg. ctrl.
PLA Got2+ Tet3
5µm

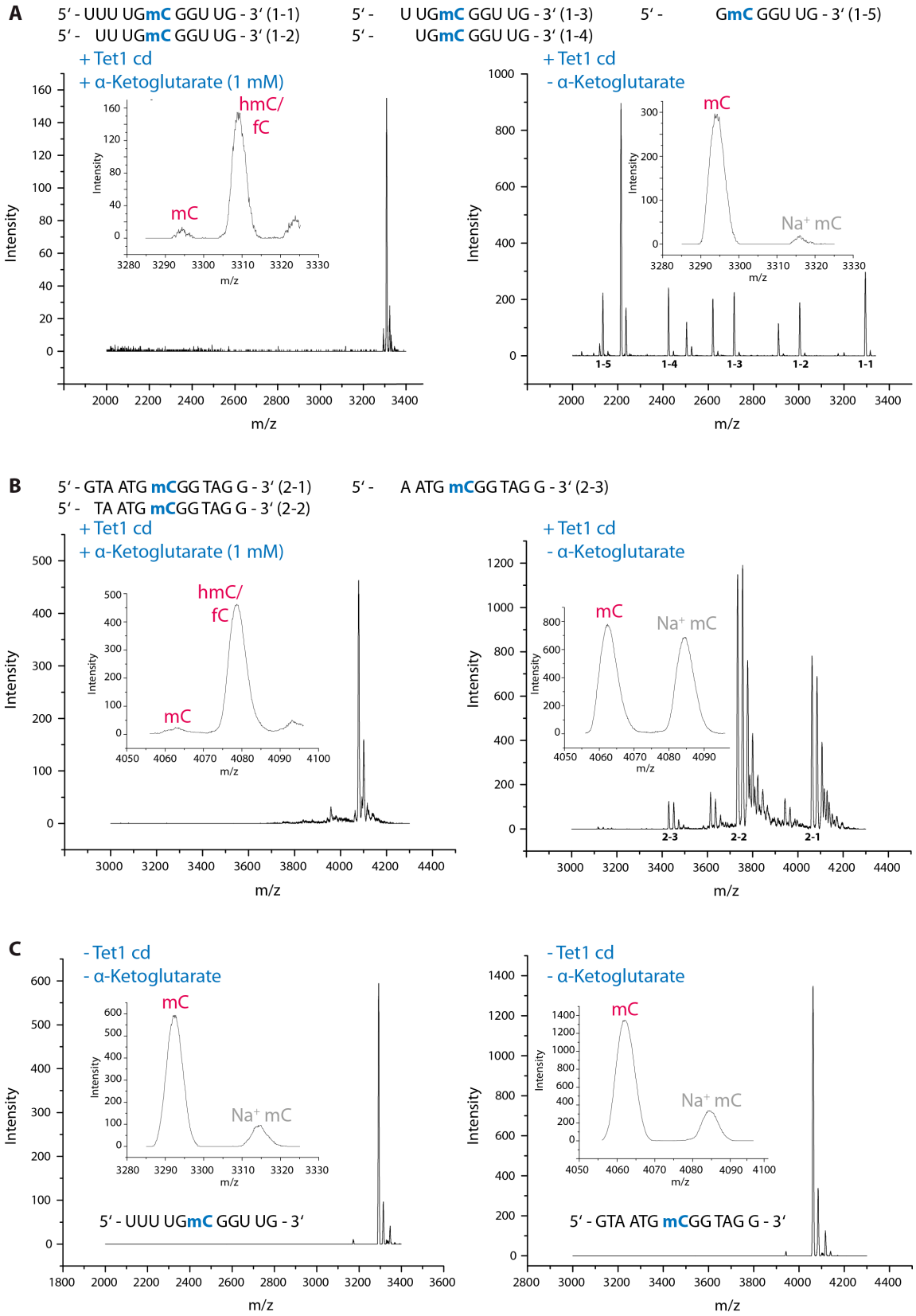
Extended Data Fig. 5: Microscopy data.(A) Test of Tet3 antibody from Abiocode and Sigma in Tet triple knock-out mouse embryonic stem cells. Chromatinmarker H3K9me3 served as a positive control that the immunocytochemistry (ICC) worked in principle. For all three ICCs the same secondary antibody was used. (B) Overview of the hippocampal regions for immunohistochemistry (IHC) using either an anti-Tet3 or an anti-Glud antibody or an anti-Aco2 antibody. ROI marks the region that was chosen for close-up images in Fig. 2C. (C) PLAs in mouse hippocampi with different antibody combinations. For the negative control, no primary antibody was used. The encirceled red dots indicate a positive signal.



Extended Data Fig. 6: Effect of R162 on Tet3 activity. Cells that only expressed Tet3 (+ - • /+ -✓) served as a control that shows that R162 did not impair Tet3 activity. All measurements were done in biological and technical duplicates (n = 2 for unpaired t-test).

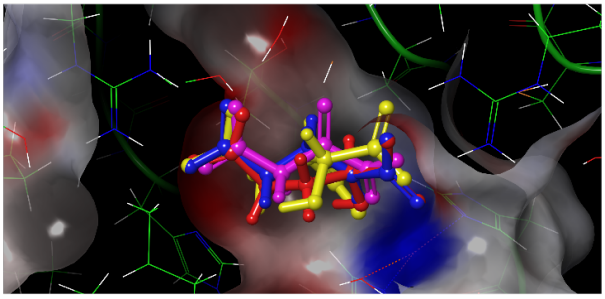


Extended Data Fig. 7: Glud1 and Tet1/Tet2 co-transfections in HEK cells and depolarization of hippocampal neurons. (A), (B) HEK293T cells were co-transfected with GFP-Tet2 and Glud1 or GFP-Tet1 and Glud1. Global levels of mC, hmC and fC were compared to each other. The measurement was done in biological and technical triplicates.


A
5′ - UUU UGmC GGU UG - 3′ (1-1)
5′ - UU UGmC GGU UG - 3′ (1-2)
5′ - U UGmC GGU UG - 3′ (1-3)
5′ - UGmC GGU UG - 3′ (1-4)
5′ - GmC GGU UG - 3′ (1-5)
+ Tet1 cd
+ α-Ketoglutarate (1 mM)
hmC/ fC
mC
+ Tet1 cd
- α-Ketoglutarate
mC
Na+ mC
1-5
1-4
1-3
1-2
1-1
B
5′ - GTA ATG mCGG TAG G - 3′ (2-1)
5′ - TA ATG mCGG TAG G - 3′ (2-2)
5′ - A ATG mCGG TAG G - 3′ (2-3)
+ Tet1 cd
+ α-Ketoglutarate (1 mM)
hmC/ fC
mC
+ Tet1 cd
- α-Ketoglutarate
mC
Na+ mC
2-3
2-2
2-1
C
- Tet1 cd
- α-Ketoglutarate
mC
Na+ mC
5′ - UUU UGmC GGU UG - 3′
- Tet1 cd
- α-Ketoglutarate
mC
Na+ mC
5′ - GTA ATG mCGG TAG G - 3′
Intensity
m/z

Extended Data Fig. 8. Tet1cd activity in presence and absence of the co-substrate. (A) Tet1cd was fully active in reaction buffer in the presence of α-ketoglutarate (DNA oligo: 5'-UUU UGmC GGU UG-3'; MALDI-TOF m/z mC: 3294.5; hmC/fC: 3310.8). When only Fe(II) and Tet1cd are present, the DNA oligo is hydrolysed: 1-1 3294.5 Da; 1-2 3004.8 Da; 1-3 2714.6 Da; 1-4: 2424.4 Da; 1-5: 2134.3 Da. (B) Tet1cd was also active on DNA-oligo 5'-GTA ATG mCGG TAG G-3'. MALDI-TOF analysis shows a small peak of mC- (4066 m/z) and a large peak of hmC/fC-containing DNA (4082 m/z). When no co-substrate is present, the oligomer was also hydrolysed: 3-1 4066.6 Da; 3-2 3737.3 Da; 3-3: 3431.1 Da. However, the degradation was to a lesser extent than for the other oligomers. (C) If no Tet enzyme is present, both DNA oligomers are stable in water in the presence of Fe(II).

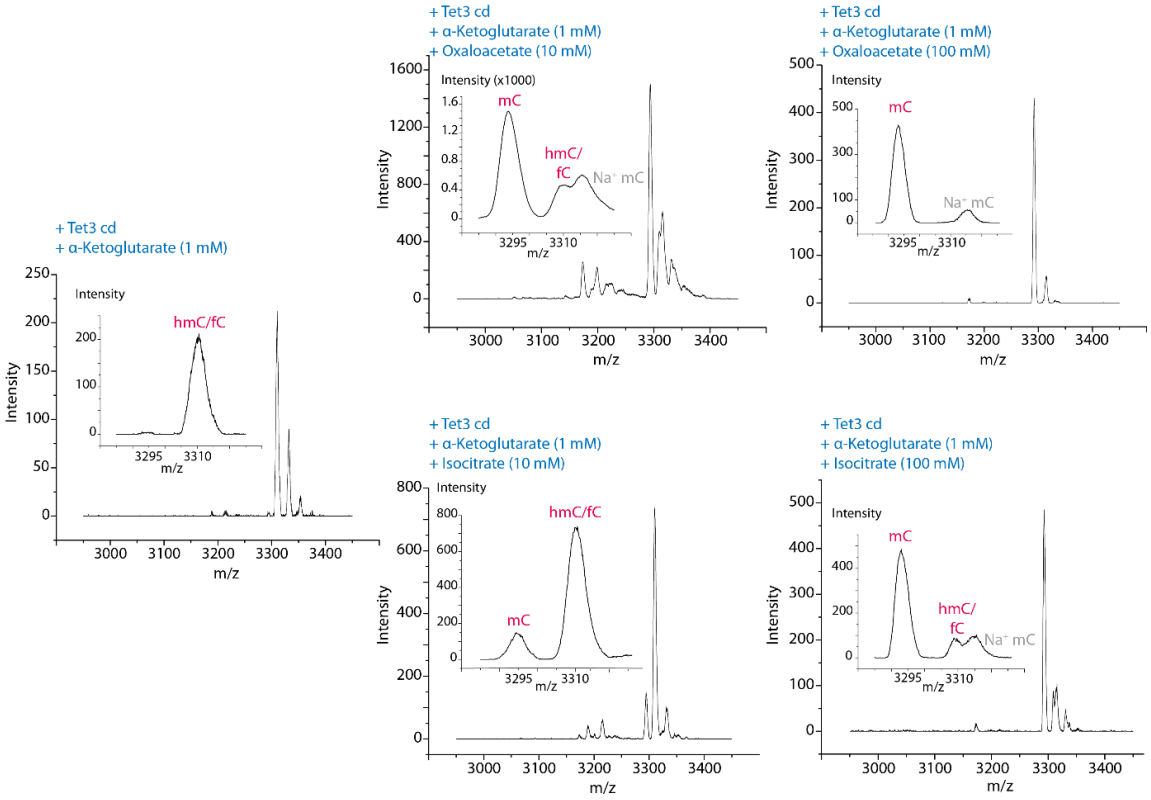
A



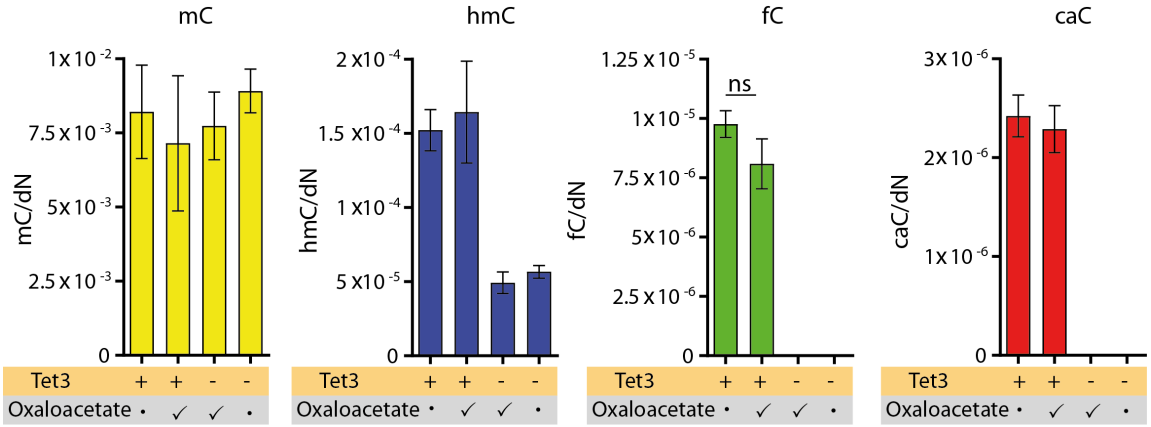
B

Metabolite	Docking score
α-Ketoglutarate	-12.701
2-Hydroxyglutarate	-12.329
Isocitrate	-11.875
Oxaloacetate	-11.381

Extended Data Fig. 9: Calculation of docking scores for different metabolites in the binding pocket of Tet2. Different metabolites were fitted inside the catalytic domain of Tet2 using Maestro 10.7 with standard settings (Glide ligand docking). (A) *upalpha*-Ketoglutarate (blue), 2-hydroxyglutarate (red), isocitrate (yellow) and oxaloacetate (purple) fit inside the binding pocket. (B) The docking scores for the different metabolites. The lower the score, the better the ligand fits in the binding pocket.



Extended Data Fig. 10: MALDI-TOF spectra of in vitro inhibitor test. The substrate peak (mC) was expected at 3294.5 Da, the product peak (hmC/fC) was expected at 3310 Da. Peaks at m/z 3315 represent Na^+ adduct peak of the original substrate.



Extended Data Fig. 11: Oxaloacetate feeding of Tet3-transfected HEK293T cells.HEK293T cells were transfected with Tet3 and fed with 10 mM oxaloacetate (+ ✓). Global levels of mC, hmC, fC and caC were compared to cells that were transfected, but not fed with oxaloacetate (+ •). Untransfected cells that were either fed with oxaloacetate (-✓) or not (- •) served as additional controls.

9 The Functional Context of TET1 and TET3 in Eu- and -Heterochromatin

9.1 Introduction

Tet enzymes are α-ketoglutarate dependent oxidases, which oxidize the DNA base 5-methylcytosine (mC) in a stepwise fashion to 5-hydroxymethylcytosine (hmC), 5-formylcytosine (fC) and 5-carboxycytosine (caC).[78–81,188] Of these oxidation products, hmC is the most abundant modification. It is strongly enriched in embryonic stem cells and in brain.[104,160,196,414] fC was recently found in human brain as well, but in all analyzed samples the fC and particularly the caC levels are small in comparison to hmC.[159] Three different Tet enzymes are known (Tet1-3) of which Tet1 and Tet3 show a dynamic adjustment of their expression during differentiation. The highest expression levels of Tet1 are found in mouse embryonic stem cells (mESCs), whereas its expression decreases during differentiation. Tet3 expression in contrast is weak in mESCs (Fig. S1), but it is quickly upregulated during differentiation.[80,246,265,414,434] Two Tet3 isoforms, one containing a CpG-binding CXXC domain (Tet3) and one lacking it (Tet3^{-CXXC}), are dominantly expressed in somatic cell types and tissues, including neural progenitor cells (NPCs) and neurons.[187,265,413,415] The question of why vertebrates need different Tet enzymes to perform the oxidation of mC to hmC and fC is a major unsolved problem.

9.2 Materials and Methods

9.2.1 Cell culture and transfection

The HEK293T cells were cultivated at 37 °C in water saturated, CO_2-enriched (5%) atmosphere. DMEM (10% FBS) was used as growing medium. When reaching a confluence of 70% to 80% the cells were passaged. The transfection was performed in p150-petri dishes. Five million cells were used in 20 mL of medium. After seeding, the cells were incubated at previously described cultivation conditions for 24 hours to reach a confluence of 60% to 80%. 10µg of DNA and 30µL of the transfection reagent

jetPRIME® purchased from Polyplus Transfection were used as described by the manufacturer. All expression plasmids for GFP-Tet were described previously.[265] To increase the transcription rate in transfected cells, 4 hours after transfection the medium was removed and sodium butyrate (4 mM) treated medium was added. After 48 hours the cells were harvested. mESCs were routinely passaged under FBS/2i conditions and were weaned off 2i (partial priming) for four days as described[80] before preparation of nuclear extracts. Except for experiments shown in Fig. S4, wild type J1 mESCs were used throughout this study. Tet TKO mESCs (Fig. S4) were described previously.[435] The FGF-2/EGF-dependent NPC line ENC1 was cultured as previously described.[265]

9.2.2 Nuclear extract preparation

Nuclear lysate of cell lines (HEK293T, mESCs, NPCs) was prepared as described in.[194] The complete adult mouse brains were lysed according to the protocol by.[428] The resulting nuclear extract was then treated with 25 U Benzonase for 30 minutes on ice and subsequently centrifuged for 15 minutes at 15000 rpm. The supernatant containing the nuclear lysate was transferred to a fresh tube. A Bradford protein assay (Bio-Rad) was performed according to the manufacturer´s instructions to measure the protein concentration.

9.2.3 GFP-Tet saturated Co-Immunoprecipitation

20 µl anti-GFP beads (Chromotek) were washed three times with wash buffer (10 mM Tris/Cl pH 7.5; 150 mM NaCl; 0.5 mM EDTA) and then incubated for 15 minutes on ice with nuclear extract of GFP-Tet overexpressing HEK293T cells. To ensure the saturation of the beads with the GFP fusion construct, different amounts of lysate and different incubation times were tested and monitored using a Tecan Reader (Fig. S2A, nonlinear fit using GraphPad Prism 7.02). The GFP-Tet loaded beads were then washed twice with 150 mM NaCl solution containing 10 mM HEPES. Another two wash steps with 1 M NaCl solution containing 10 mM HEPES and two wash steps with Lysis Buffer C (20 mM Hepes pH = 7.5, 420 mM NaCl, 2 mM $MgCl_2$, 0.2 mM EDTA, 20% (v/v) glycerol) followed. The saturated GFP-Tet beads were subsequently incubated with 200 µg nuclear extract of choice (mESCs, NPCs) for 15 minutes on ice. Following, they

were washed twice with wash buffer. To elute the bound proteins, 50 µl of 200 mM glycine pH 2.5 was added and the solution was vortexed for 30 seconds. To gain more yield the elution step was repeated.

9.2.4 GFP-Tet saturated Co-Immunoprecipitation

The reversed Co-IP was carried out with FACT/ Supt16H antibody (PA5-18443, Thermo Scientific), PP1 gamma antibody (PA5-21671, Thermo Scientific) and Anti-Goat IgG antibody (G4018, Sigma-Aldrich), which served as a negative control. For each reversed Co-IP, 10 µL of Dynabeads Protein G (Life Technologies) were washed three times with 500 µL of Washbuffer 3 (PBS, 0.5% (w/v) octyl β-D-glucopyranosid) prior to addition of 1 µg of antibody. The antibody was dissolved in 50 µL of Washbuffer 3, added to the Dynabeads and mixed for 1 hour at room temperature on a tube rotator. Then, the unbound antibody was removed and the beads with the bound antibody were washed two times with 500 µL of Washbuffer 3. In the next step, 250 µg of nuclear lysate from mESCs, when PP1 gamma antibody or anti-goat IgG antibody were used, or 250 µg of nuclear lysate from NCPs, when FACT/Supt16H antibody was used, were added in a total volume of 400 µL of Lysis Buffer C (20 mM Hepes pH = 7.5, 420 mM NaCl, 2 mM $MgCl_2$, 0.2 mM EDTA, 20% (v/v) glycerol) to the beads and gently mixed by pipetting up and down. The lysate was incubated with the beads for 1 hour at room temperature on a tube rotator. In order to remove unbound and non-specifically bound protein after incubation, the supernatant was discarded and the beads were washed two times with 500 µL of Washbuffer 1 (10 mM Hepes pH = 7.5, 150 mM NaCl, 0.5 mM EDTA), two times with 500 µL of Washbuffer 2 (10 mM Hepes pH = 7.5, 1 M NaCl, 0.5 mM EDTA) and two times with 500 µL of Washbuffer 3. Afterwards, 250 µg of nuclear lysate from GFP-Tet3 transfected HEK293T cells, which had not been treated previously with butyric acid, were added in a total volume of 400 µL of Lysis Buffer C and gently mixed by pipetting up and down. The lysate and the beads were incubated for 2 hours at 4 °C on a tube rotator. Finally, the supernatant was discarded, the beads were washed three times with 500 µL of Washbuffer 3 and the bound protein complexes were eluted by addition of 30 µL of 1% (v/v) formic acid and incubation for 15 minutes at room temperature in a thermomixer (700 rpm). The eluate was collected and 10 µL of 1 M Tris were added in order to neutralize the pH; then the samples were prepared for LC-MS/MS analysis.

9.2.5 LC-MS/MS analysis

Samples for the mass spectrometer were reduced by the addition of 100 mM TCEP and subsequent incubation for 1 hour at 60 °C on a shaker at 650 rpm. They were then alkylated by adding 200 mM iodoacetamide and incubating for 30 minutes at room temperature in the dark. Following, the samples were digested with 0.5 µg trypsin (Promega) at 37 °C for 16 hours. Afterwards, they were incubated for 5 minutes at 100 °C and subsequently 1 mM phenylmethylsulphonylfluoride was added. StageTips were utilized to purify the samples for mass spectrometry.[429] The samples were analyzed with an UltiMate 3000 nano liquid chromatography system (Dionex, Fisher Scientific) attached to an LTQ-Orbitrap XL (Fisher Scientific). They were desalted and concentrated on a µ-precolumn cartridge (PepMap100, C18, 5 µM, 100 Å, size 300 µm i.d. x 5 mM) and further processed on a custom made analytical column (ReproSil-Pur, C18, 3 µM, 120 Å, packed into a 75 µm i.d. x 150 mM and 8 µm picotip emitter). The samples were processed via a 127 minutes multi-step analytical separation described below at a flow rate of 300 nl/min. Only LC-MS grade solvents were used (solvent A: water + 0.1% formic acid; solvent B: acetonitrile + 0.1% formic acid). The gradient with percentages of solvent B was programmed as follows: 1% for 1 minute; 1% - 4% over 1 minute; 4% - 24% over 100 minutes; 24% - 60% over 8 minutes; 60% - 85% over 2 minutes; 85% for 5 minutes; 85% - 1% over 2 minutes; 1% for 8 minutes. Mass spectrometric analysis was done with a full mass scan in the mass range between m/z 300 and 1700 at a resolution of 60000. Following this survey scan, five scans were performed using the ion trap mass analyzer at a normal resolution setting and wideband CID fragmentation with a normalized collision energy of 35. Signals with an unrecognized charge state or a charge state of 1 were not picked for fragmentation. A dynamic exclusion list was implemented with the following settings: after two measurements in 30 seconds, the peptide was excluded from selection and fragmentation for 90 seconds.

9.2.6 Mass spectrometric data processing

The MaxQuant software (version 1.5.0.25) was used for the mass spectrometric data processing. The Andromeda search engine was used in combination with Uniprot databases (Mus musculus). A maximum of two missed cleavage sites was allowed. The main search

peptide tolerance was set to 4.5 ppm. Carbamidomethyl (C) was set as static modification. Variable modifications were Acetyl (Protein N-term) and Oxidation (M). minimal peptide length was set to 7.

9.2.7 LFQ data processing

For LFQ, mass spectrometric data processing was used in combination with the LFQ option in the MaxQuant software (version 1.5.0.25). Quantification was performed with four biological replicates. The Co-IP with GFP alone served as control. The LFQ algorithm was applied with default settings (LFQ min. ratio count: 2, Fast LFQ, LFQ min. number of neighbours: 3, LFQ average number of neighbours: 6). The option "match between runs" was also used. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE[436] partner repository with the dataset identifier PXD005711. LFQ data was analyzed with the Perseus software (version 1.5.0.9). The LFQ intensities were log transformed and only proteins identified in at least three samples were retained. Gene ontology analyses were performed with the Database for Annotation, Visualization and Integrated Discovery (DAVID Bioinformatics Resources 6.7).

9.2.8 Proximity Ligation Assay

The proximity ligation assay (PLA) was carried out using Duolink InSitu Orange Starter Kit (Sigma). 30000 2d primed mESCs were seeded in each µ-plate well (ibidi) and transfected with GFP-Tet3 and GFP as control using lipofectamine 2000 (Life Technologies). The cells were then incubated for 2 days at 37 °C. Afterwards, the medium was removed and 200 µL of Washbuffer A (PBS, 0.9 mM $CaCl_2$, 0.5 mM $MgCl_2$, 0.02% (v/v) Tween-20) were added immediately in order to prevent the cells from drying out. The cells were washed two times with 200 µL of Washbuffer A and then fixed on the surface. For the fixation, 200 µL of 3.7% (v/v) paraformaldehyde in Washbuffer B (PBS, 0.9 mM $CaCl_2$, 0.5 mM $MgCl_2$) were added and the cells were incubated for 15 minutes at room temperature. Afterwards, the fixation solution was removed and the cells were washed two times with 200 µL of Washbuffer A. In the next step, the cells were incubated for 1 hour at room temperature in 200 µL of Blocking Solution (3% (w/v) BSA in Washbuffer

B). In the meantime, the antibody solutions were prepared. 1 µL of PP1 gamma antibody (PA5-21671, Thermo Scientific) and 1 µL of GFP (4B10) Mouse antibody (2955S, Cell Signaling) were dissolved in 100 µL of Antibody Diluent (Sigma), respectively. The antibodies were added after having removed the Blocking Solution and the cells were incubated overnight at 4 °C in a humid atmosphere. The next day, the Antibody Diluent was discarded and the cells were washed two times with 200 µL of Washbuffer A. For the PLA Probe Solution, 10 µL of PLA Probe (+) Anti-Rabbit (Sigma), 10 µL of PLA Probe (-) Anti-Mouse (Sigma) and 30 µL of Antibody Diluent were mixed and added. The cells were incubated in a humidity chamber for 1 hour at 37 °C. In the next step, the PLA Probe Solution was removed, the cells were washed two times with 200 µL of Washbuffer A and 50 µL of the Ligation Solution, which had been prepared previously by mixing 1.0 µL of 1U/µL Ligase, 10 µL of Ligation Stock (all from Sigma) and 29 µL of bidest. H_2O, were added. After 30 minutes of incubation in a humidity chamber at 37 °C, the Ligation Solution was discarded and the cells were washed two times with 200 µL of Washbuffer A. For the amplification reaction, 0.5 µL of Polymerase 10 U/µL, 10 µL of Amplification Stock (all from Sigma) and 39.5 µL of bidest. H_2O were mixed and added to the cells. The cells were incubated for 100 minutes at 37 °C in a humidity chamber. The Amplification Solution was discarded; the cells were washed two times with 200 µL of Washbuffer B and one time with 200 µL of 0.01% Washbuffer B. Finally, 100 µL of Washbuffer B were added and the cells were analyzed by confocal microscopy in the fluorescence modus (Leica DMI4000B, filter I3 for GFP, filter N2.1 for PLA amplification product).

9.2.9 RNA isolation

RNA from mESC and ENC1 was isolated using peqGold Total RNA Kit (peqLab) according to the manufacturer's manual.

9.2.10 Real-time quantitative polymerase chain reaction (qPCR)

Prior to the qPCR, 1 µg of RNA was transcribed to 1µg of cDNA using the iScript cDNA Synthesis Kit (BioRad) according to the manufacturer's manual. The final cDNA concentration was 50 ng/µL. For the qPCR, the cDNA was diluted to 25 ng/µL with

nuclease-free bidest. H_2O and 10 µL of iTaq Universal SYBR Green Supermix (Bio-Rad) were added per 2 µL of cDNA dilution. For each primer pair, the forward and the reverse primer were mixed together in equal amounts and diluted with nuclease-free bidest. H_2O to a final concentration of 1.25 µM. 12µL of cDNA/iTaq mastermix and 8 µL of primer mix were added per well. 96-well PCR plates (Axygen Scientific) were used in combination with Flat 8 Cap Strips and the samples were spun down at 2000 x g for 5 minutes using an Eppendorf centrifuge 5810 R. Each sample was measured in triplicates using an Eppendorf Mastercycler Realplex4 with the following PCR conditions: Step 1 95 °C (2:00 min), Step 2 95 °C (0:15 min), Step 3 55 °C (0:15 min), Step 4 72 °C (0:20 min); Step 2 – 4 were repeated 40 times.

Tet1 forward primer: 5' CCA GGA AGA GGC GAC TAC GTT 3'
Tet1 reverse primer: 5' TTA GTG TTG TGT GAA CCT GAT TTA TTG T 3'
Tet2 forward primer: 5' ACT TCT CTG CTC ATT CCC ACA GA 3'
Tet2 reverse primer: 5' TTA GCT CCG ACT TCT CGA TTG TC 3'
Tet3 (all) forward primer: 5' GAG CAC GCC AGA GAA GAT CAA 3'
Tet3 (all) reverse primer: 5' CAG GCT TTG CTG GGA CAA TC 3'

9.2.11 In vitro activity test

For the activity tests, approximately 10 million HEK293T cells were used. The cells were transfected with Tet1 catalytic domain (Tet1cd) or Tet3 full length protein (Tet3FL), according to the protocol above. The cells were harvested, centrifuged and lysed with 5 mL of RIPA buffer (Chromotek), supplemented with 250 U benzonase (Merck Millipore) and protease inhibitor (Roche). After 30 minutes on ice the suspension was centrifuged at 10000 x g and 4 °C for 15 minutes. The supernatant consisted of nuclear and cytoplasmic proteins. Per activity test, 100µL of anti-GFP beads (Chromotek) were washed three times with wash buffer (10 mM Tris/Cl pH 7.5; 150 mM NaCl; 0.5 mM EDTA) and then incubated for 1 hour at 4 °C with the supernatant. The GFP-Tet loaded beads were washed with GFP wash buffer (Chromotek). For the second wash step GFP wash was supplemented with 2 mM $ZnSO_4$ and 10µL Nuclease S1. The beads were incubated with this buffer for 30 minutes on ice. Another two wash steps with 1 M NaCl solution containing 10 mM HEPES, pH 7.5 and two wash steps with GFP wash buffer followed. The beads were centrifuged at 2500 x g and 4 °C for 5 minutes and the supernatant

discarded. Next, reaction buffer (50 mM HEPES pH=7.5, 100 mM NaCl, 2 mM Vitamin C, 1.2 mM ATP, 2.5 mM DTT, 0.1 mM $Fe(II)(NH_4)_2(SO_4)_2$ 6 x H_2O) and 1 mM α-ketoglutarate (Na^+ salt) was added. As an additional control, one sample was prepared without Tet. 1000 pmol of an Oligo (5'-UUUUGmCGGUUG–3') was added and the reaction mixture incubated at 37 °C for 1.5 hours. After centrifugation (15000 rpm), the supernatant containing the oligonucleotide was transferred into a new tube. The DNA was precipitated from 3 M NaOAc (pH = 5.2) and dissolved in 30µL of water.

9.2.12 Immunocytochemistry

Immunocytochemistry (ICC) experiments were performed as previously described with minor modifications. In brief, 30000 3 d primed mESCs or ENC1 cells were seeded in eachµ-plate well (ibidi). The next day (4 d priming for mESC), cells were fixed (10 min) on slides using 4% (v/v) paraformaldehyde (PFA) in 0.1 M phosphate buffered saline with 0.5 mM $MgCl_2$ and 0.9 mM $CaCl_2$, pH 7.4 (0.1 M PBS). After three times washing with 0.1 M PBS, the cells were permeabilized using 0.3% (v/v) TritonX in 0.1M PBS. After washing for three times, cells were incubated with 0.1 M PBS with 5% blocking reagent (Chemiblocker, CB, Millipore) for 30 min. Primary antibodies were added and incubated for two hours at room temperature in a humidity chamber. Anti-Tet3 antibody (Abiocode M1092-3 (dilution 1:500), from mouse), anti-Tet1 antibody (Active Motif 61741 (dilution 1:300), from rat) and anti-H3K9me3 antibody (Active Motif H3K9me3 (dilution 1:500), from rabbit) were used as primary antibodies for ICC with chromatin markers. Anti-Tet3 antibody (Abiocode M1092-4a (dilution 1:2000), from mouse), anti-Ppp1cc antibody (Thermo Scientific PA5-21671 (dilution 1:100), from rabbit), anti-Tet1 antibody (Active Motif Active Motif 61741 (dilution 1:50), from rat) and anti-Ywhag antibody (Thermo Scientific PA5-29690 (dilution 1:100), from rabbit) were used as primary antibodies for validation of previously discovered Tet interactors in mESC. Anti-Tet3 antibody (Sigma HPA050845 (dilution 1:100), from rabbit) and anti-Supt16 antibody (Thermo Scientific PA5-18443 (dilution 1:200), from goat) were used as primary antibodies for validation of previously discovered Tet interactors in ENC1. The primary antibodies were diluted in 0.1 M PBS, containing 5% (v/v) blocking reagent and 0.3% (v/v) Triton-X solution. For the negative controls, no primary antibodies were added. After incubation, cells were washed three times with 0.1 M PBS, containing 2% (v/v) blocking reagent. For detection, we used Alexa488-anti-rat (1:600,

Cell Signaling Technologies), Alexa488-anti-mouse (1:600, Cell Signaling Technologies), Alexa555-anti-rat (1:600, Cell Signaling Technologies), Alexa555-anti-rabbit (1:600, Cell Signaling Technologies), Alexa555-anti-mouse (1:600, Cell Signaling Technologies) and Cy3-anti-goat (1:400, Jackson ImmunoResearch) diluted in 0.1 M PBS, containing 3% (v/v) CB as secondary antibodies and incubated for 1 h at room temperature. After incubation, cells were washed three times with 0.1 M PBS. Cell nuclei were stained with Hoechst 33342 (5µg/mL), which was applied for 5 min in the dark at room temperature. After mounting (Mountant Permafluor, ThermoScientific), the slides were analyzed using a Leica SP8 confocal laser scanning microscope (Leica, Wetzlar). Co-localization was analyzed using ImageJ 2.0.0, Colocalization Threshold. The ICC experiments were performed in two biological replicates, yielding the same result.

9.3 Results

In order to clarify the functional roles of Tet1 and both Tet3 isoforms, we performed an affinity proteomics study to decipher the interacting proteins and protein complexes. To allow a direct comparison between different cell types, we used a Tet saturated co-immunoprecipitation (Co-IP) workflow as recently reported and as depicted in Fig. 16A. In short, GFP-Tet fusion protein coated magnetic beads were generated. The beads were intensively washed with up to 1 M NaCl in order remove unspecifically bound proteins. This intensive washing removed contaminating HEK proteins efficiently but did not impair the binding of GFP-Tet to the GFP-nanobody beads (Fig. S2B). We showed that the different Tet proteins bound to the beads possess the expected catalytic activity proving that the Tet enzymes are functional intact. Importantly, we see oxidation by the different Tet-covered beads of mC to all expected products hmC, fC and caC in the presence of -ketoglutarate and Fe^{2+}. These functional Tet-loaded beads are used as baits for Co-IP of Tet interactors in lysates of different cell types. This target protein saturated Co-IP workflow enables not only a robust comparison between samples but also the study of low abundant bait proteins. We analyzed the interactomes of Tet1, Tet3 and Tet3^{-CXXC} (Fig. 16B) in mESCs and NPCs. For quantitative analysis, label-free quantification technology[418] was applied with GFP-covered beads as controls. In each case, we analyzed four biological replicates to enable statistically reliable quantification. The data show a close correlation within the replicates as determined by the Pearson correlation (Fig. S3A), proving high data quality. The LFQ intensities for each sample

are distributed as expected (Fig. S3B). The principle component analysis of the data shows the clustering of the different Tet Co-IPs and GFP control in mESCs as expected (Fig. S3C). In total, 1497 proteins were identified, of which 61 are enriched with a -log(two-way ANOVA p-value interaction) parameter of > 2 (Fig. 16C). The correlation analysis of the obtained data uncovers distinct interactomes of the individual Tet enzymes in the two investigated cell types. Interestingly, we found the cell type specific differences of the interactomes most pronounced, but identified also Tet paralog-specific clusters. A strong cell-type specific cluster contains proteins involved in gene expression and nucleic acid binding, which are strongly enriched in NPCs, but depleted in mESCs (green box). The cluster contains proteins involved in gene expression and nucleic acid and protein binding (Table S1). Analysis of individual proteins shows that the kinase Camk2b is enriched. Another identified protein is Suds3, which is part of the Sin3a complex. SudS3 was already discovered as an interactor of Tet1[188], which supports our data. Furthermore, Hmgb1 is strongly enriched in the NPC data set, which is involved in the organization of DNA and regulation of transcription. Hmga1 is another high mobility group protein regulating transcription and is highly enriched in NPCs in contrast to mESCs. The cell-type specific cluster for mESCs (cyan box) shows nucleic acid binding in the nucleus as significant gene ontology (GO) terms for molecular function and cellular compartment (Table S2).

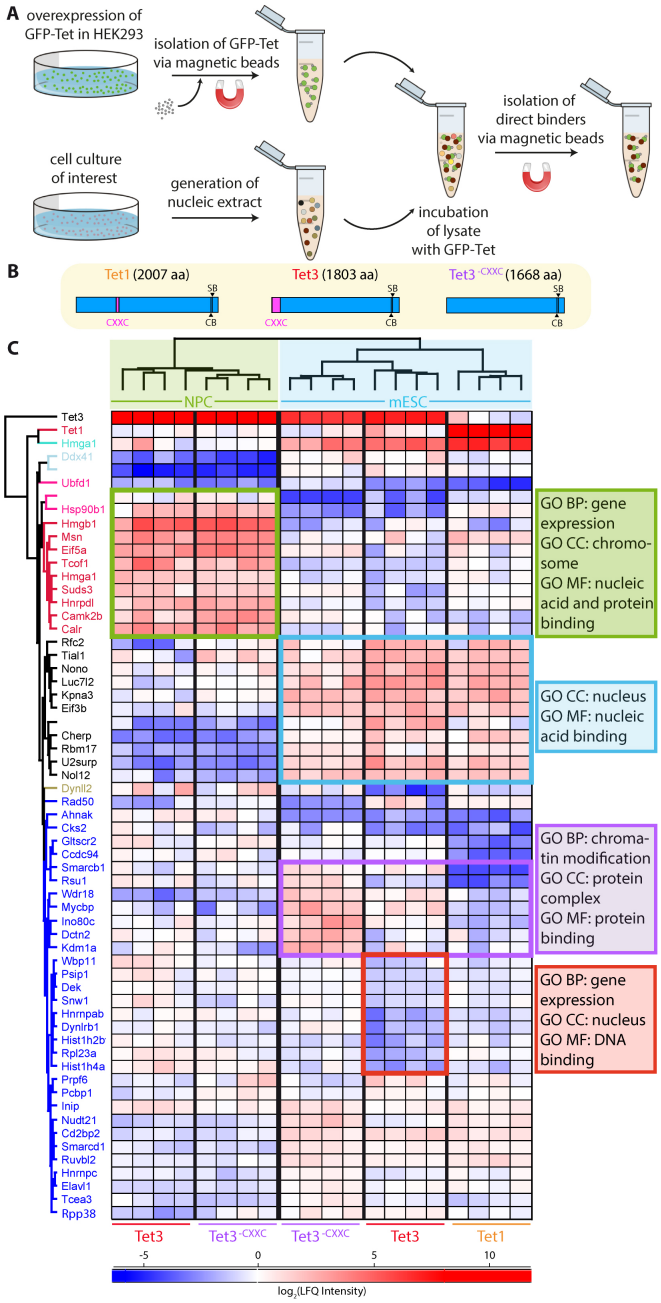


Figure 18: Experimental design and data overview. (A) Tet-saturated co-immunoprecipitation workflow. (B) Schematic representation of the Tet constructs used for this study. aa = amino acids; SB = substrate binding; CB = co-substrate binding (C) Correlation-based clustering of the log2 transformed label-free quantification (LFQ) intensities normalized by row mean subtraction of the GFP control. A selection of Gene Ontology (GO) terms are given (BP = biological process, CC = cellular compartment, MF = molecular function). The gradient from blue to red indicates the increase of enrichment ascertained by a two-way ANOVA test (-log(two-way ANOVA p-value interaction) > 2); mESC = mouse embryonic stem cells; NPC = neural progenitor cells.

For example, Nono is a strong Tet interactor in contrast to NPCs. This protein is involved in transcriptional regulation. Strikingly, Kpna3 is especially enriched in mESCs as well. This protein functions in nuclear protein import by binding to NLS motifs. Moreover, Rfc2, which is involved in replication, is also enriched.

The paralog specific clusters are less pronounced and show significant separation only in mESCs. Most interesting is the observation that in mESCs, proteins involved in chromatin remodeling are enriched in experiments performed with Tet3^{-CXXC} and they are repelled in the case of Tet1 (purple box, Table S3)). Another observation is that Tet3 seems to be less involved in gene expression than Tet1 and Tet3^{-CXXC} (red box, Table S4). A detailed analysis of individual reader proteins is shown in Fig. 18.

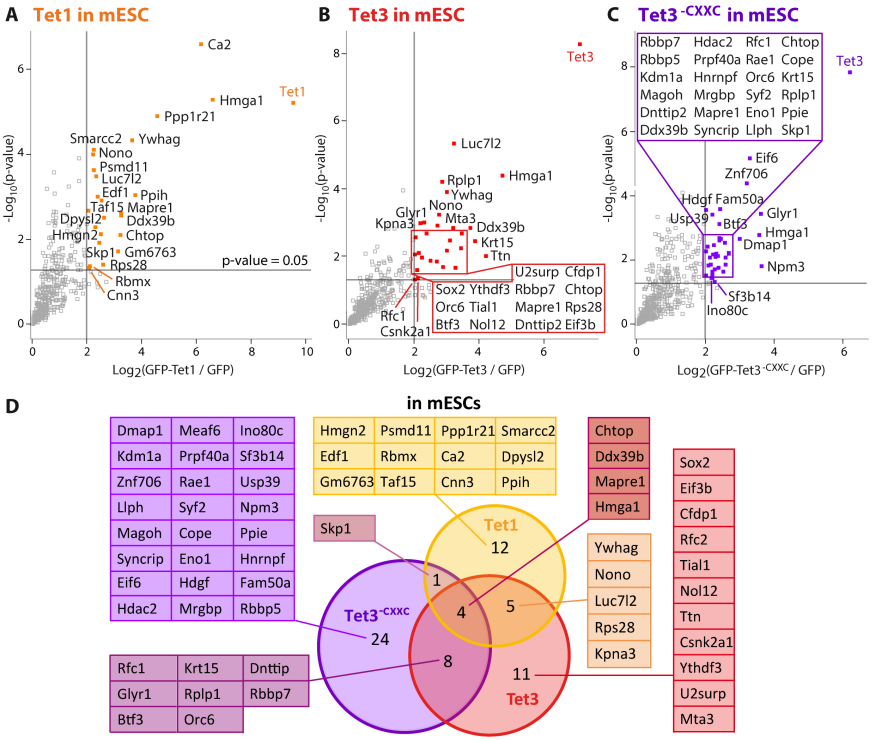


Figure 19: Interaction partners of Tet family members in mouse embryonic stem cells (mESCs). (A-C) Volcano plots of the Co-IP experiments with Tet1, Tet3 and Tet3^{-CXXC} in mESCs. (D) Venn diagram showing the interaction partners of Tet1, Tet3 and Tet3^{-CXXC} in mESCs.

The Tet1 interactome in mESCs reveals 22 attracted proteins (Fig. 17A), twelve of which bind to Tet1 only (Fig. 17D yellow box). Two of those proteins are involved in

transcription regulation in combination with chromatin remodeling (Hmgn2, Smarcc2). Three of the enriched proteins have functions in transcription initiation and regulation of cell differentiation (Hmga1, Edf1, Taf15). Regarding the specific Tet1 interactors, enrichment of Hmgn2 is notable because this protein stabilizes open chromatin structures. Additionally, the protein Alyref (Gm6763), which is part of the TREX complex, is found to interact specifically with Tet1. TREX is coupling mRNA transcription to mRNA processing and nuclear export. Other components of this complex (Chtop and Ddx39b) are found to interact with all Tet proteins in mESCs. Skp1, which is found both in the Tet1 and Tet3$^{\text{-CXXC}}$ interactomes, is part of the SCF ubiquitin ligase complex. This complex mediates the ubiquitination of proteins involved in transcription as well as signal transduction and cell cycle progression. The combined data therefore suggest that Tet1 acts in transcriptional regulation.[437] Another Tet1 interactor, which was also found for Tet3 and Tet3$^{\text{-CXXC}}$, is Ywhag. This protein is a phosphopeptide binding regulatory adapter protein and we could confirm the interaction in mESC using immunocytochemistry (ICC) (Fig. S4). We next studied the interaction partners of Tet3$^{\text{-CXXC}}$ (Fig. 17C). With this experiment, we also wanted to exclude the possibility that the CXXC domain, which is known to mediate binding to CpG containing DNA, pulls down general DNA binding proteins that are not Tet3 specific. Upon comparing the interaction partners of Tet3 and Tet3$^{\text{-CXXC}}$, it is evident that more proteins and especially a larger set of DNA-binding proteins are enriched in the Co-IPs with Tet3$^{\text{-CXXC}}$. This excludes the possibility that the DNA-binding capability of the CXXC domain distorts our experimental results (Fig. 17D). In total, we detected 12 proteins that bind to both Tet3 and Tet3$^{\text{-CXXC}}$, while 24 proteins are specifically enriched for Tet3$^{\text{-CXXC}}$. Some of these exclusive Tet3$^{\text{-CXXC}}$ interactors are transcriptional repressors (Znf706, Eno1, Hdgf). Besides this, many identified proteins are involved in chromatin remodeling (Kdm1a, Hdac2, Dmap1, Meaf6, Mrgbp, Ino80c, Rbbp5, Rbbp7). The data show that among all the Tet proteins, Tet3$^{\text{-CXXC}}$ specifically recruits proteins that interact with histones, perform DNA modification and are histone modification reader. This idea is supported by a GO analysis that reveals gene expression (p-value = 5.4 x 10^{-0}) and chromatin modification (p-value = 5.1 x 10^{-4}) as significantly enriched biological processes. Analysis of the interactome of Tet3 in mESCs reveals a slightly richer interactome with nine proteins previously identified in the Tet1 pulldown and an additional 19 interaction partners (Fig. 17B, D). Strikingly, we detected a specific interaction of Tet3 with the pluripotency and transcription factor Sox2.[438] Regarding the other pluripotency factors already known to interact with Tet proteins, (Nanog and Oct4), we were

able to confirm Oct4 as a weak Tet3 interaction partner (log2(GFP-Tet3/GFP) = 2.5; –log(p-value) = 1.1).[439,440] Other Tet3 interactors are Ywhag and Csnk2a1, which is the catalytic subunit of a serine/threonine-protein kinase complex, indicating a possible regulation via phosphorylation. In the Tet3 interactome in mESCs, proteins involved in regulation of transcription (Mta3, Btf3, Sox2, Nono) and replication (Rfc1, Rfc2, Orc6) dominate. Moreover, the histone-binding protein Rbbp7 could be identified as a Tet3 interactor, together with Glyr1, which binds chromatin modifications. These data indicate that Tet3 has a higher tendency to interact with proteins associated with chromatinized DNA than Tet1. Interestingly, we detected Chtop as an interactor of all investigated Tet proteins, which supports the recent discovery that hmC recruits the Chtop-methylosome complex to initiate transcription.[441] Weak interactions may not be detectable when the exogenous bead-bound protein has to compete with endogenously expressed Tet protein. We therefore repeated the Tet1 experiment in Tet triple knockout cells to see if proteins would appear that were otherwise not identified by our method because the exogenous Tet-bait may be unable to disrupt existing complexes. This experiment, however, provided no significant differences, arguing that our Tet-baits are correctly presenting the interaction profiles (Fig. S5). It is to be expected that interaction profiles change in the course of differentiation. To investigate this process, we analyzed NPCs, which are more differentiated compared to mESCs. For this study, we used both Tet3 and Tet3^{-CXXC} as they are the dominant isoforms in these cells.[265] Again, Tet3^{-CXXC} has more interaction partners than Tet3 (Fig. 18A and 18B). Among the proteins, which bind to both Tet3 isoforms, the Hmg proteins (Hmgn1, Hmgn2, Hmgb1) are abundant (Fig. 18C). These proteins alter the accessibility of the DNA duplex in the chromatin environment. One of the interactors of Tet3 is Lgals1, which is involved in the regulation of apoptosis, proliferation and differentiation. Another enriched protein is Sub1, which is a general co-activator cooperating with TAFs as a mediator between activators and the transcriptional machinery.

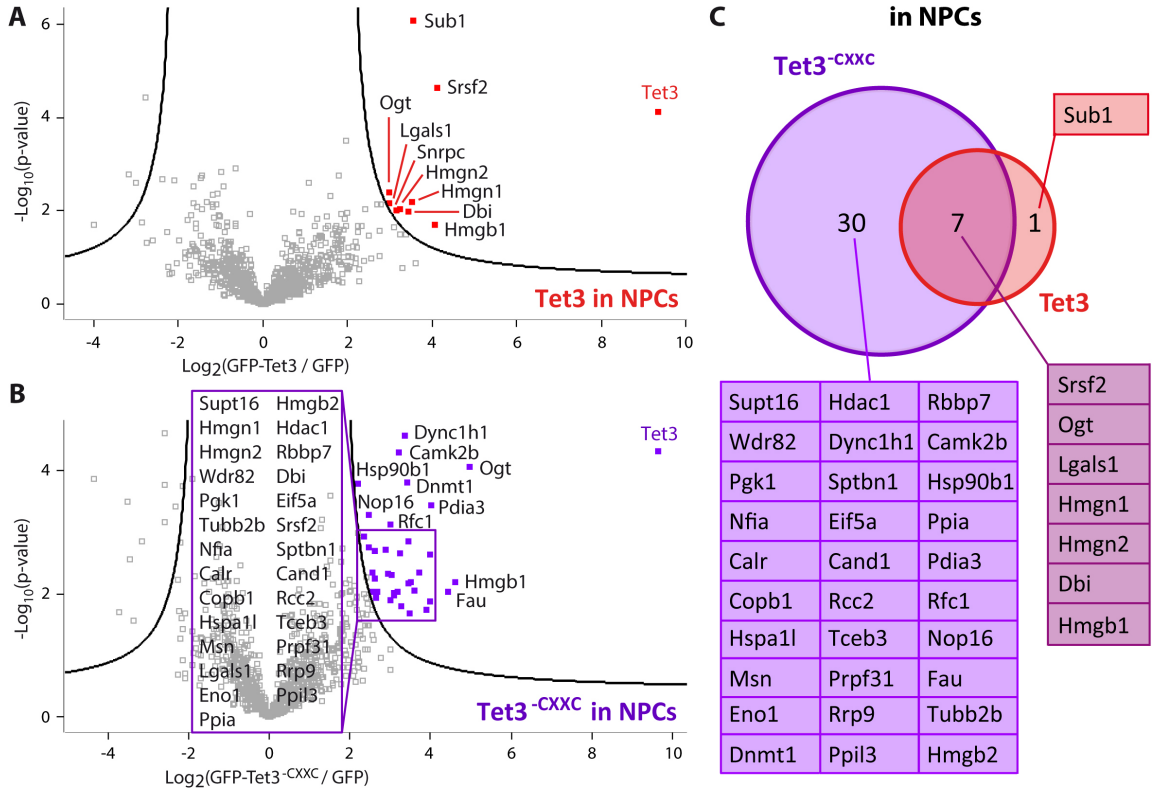


Figure 20: Interaction partners of Tet family members in neural progenitor cells (NPCs). (A-B) Volcano plots of the results from the modified co-immunoprecipitations with Tet3 and Tet3^{-CXXC} in NPCs (s0 = 2; false discovery rate (FDR) = 0.05). (C) Venn diagram showing the interaction partners of Tet3 and Tet3^{-CXXC} in NPCs.

Seven of the nine proteins enriched in the Tet3 dataset are also interactors of Tet3^{-CXXC}. The observation that Tet3^{-CXXC} has more interaction partners, might find its rationale in the fact that Tet3^{-CXXC} is more abundant in neuronal cells.[265] A protein that specifically interacts with Tet3^{-CXXC} is Dnmt1, responsible for maintenance of DNA methylation and itself a CXXC domain containing protein. Another protein that binds to Tet3^{-CXXC} is Rcc2, which is a regulator of chromosome condensation. Other Tet3^{-CXXC} interacting proteins are Ogt, Hdac1, Rbbp7, Hmgb2, Wdr82 and Supt16, which are all involved in chromatin organization and nucleosome modification. Notably, Ogt is present in our data, which is a well-known binder of the Tet enzymes. This again supports our data.[302,419] Tet3^{-CXXC} interactors also include proteins involved in DNA transcription regulation and replication such as Rfc1, Tceb3 and Nfia. Furthermore, Eno1 is found,

which acts as transcriptional repressor binding to the myc promoter. Another interaction partner is Eif5a that regulates apoptosis. This protein is thought to play a major role in brain development, possibly connecting Tet3$^{\text{-CXXC}}$ with neuronal development.[413,415] In order to uncover general trends behind the data, we analyzed the statistically enriched Tet1, Tet3 and Tet3$^{\text{-CXXC}}$ interactors in the respective volcano plots. As most of the interacting proteins are associated with chromatin modification, we focused particularly on protein complexes involved in chromatin remodeling such as NuA4, SWI/SNF, NuRD, FACT and PTW/PP1 (Fig. 19A, Table S5). Proteins, which show a strong interaction (more than twofold enriched with a p-value < 0.05) are color-coded (Table S6). This analysis confirms that Tet1 has only weak interactions with chromatin remodeling complexes. Only few contacts are observed with the SWI/SNF chromatin remodeling complex (Smarcc1, Smarcc2, Smarcd1), the NuA4 histone acetylase complex (Ruvbl2), the PTW/PP1 complex (Ppp1ca) and the NuRD deacetylase complex (Rbbp4, Mbd3). In contrast to Tet1, the data reveal a statistically highly significant enrichment of chromatin remodelers for Tet3 (student´s t-test p-value = 0.0096 for Tet3 and student´s t-test p-value = 0.0004 for Tet3$^{\text{-CXXC}}$) (Fig. 19C). This finding supports the idea that a function of Tet3 is to operate on more chromatinized DNA. Particularly strong Tet3 interactions are Arid1a of the SWI/SNF complex, Dmap1 and Ruvbl2 of the NuA4 complex, Wdr82, Ppp1ca and Ppp1cc of the PTW/PP1 complex and Mta3 and Rbbp7 of the NuRD complex. An even stronger correlation with chromatin remodeling is detected for Tet3$^{\text{-CXXC}}$ that extends also to the FACT complex, arguing that Tet3 and particularly Tet3$^{\text{-CXXC}}$ operate in stronger chromatinized DNA.

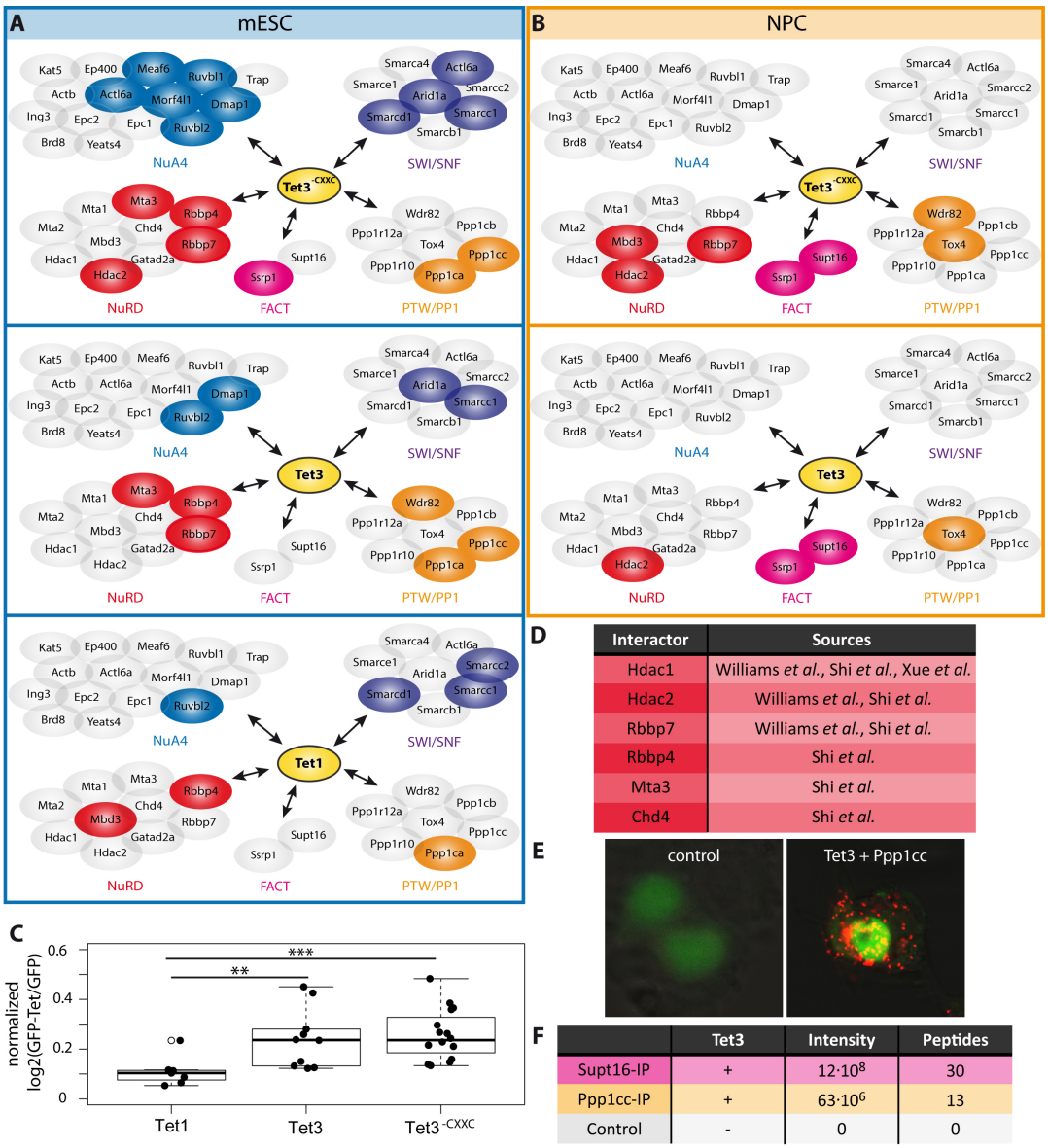


Figure 21: Chromatin modifying complexes interact with Tets. (A) Complexes associated with Tet1, Tet3 and Tet3^{-CXXC} in mouse embryonic stem cells (mESC) and neural progenitor cells (NPC) (B). Colored proteins have a p-value < 0.05 and are more than twofold enriched. (C) Boxplot of the log2(GFP-Tet/GFP) values normalized for Tet1, Tet3 and Tet3^{-CXXC} of the significantly enriched complex members from (A). (D) List of the Tet interactors that have already been identified.[188,271,442] (E) Tet3 was identified (+) in the reversed Co-IP experiments with Supt16 and Ppp1cc. The intensity and the number of unique peptides of Tet3 are indicated. (F) Immunocytochemistry experiments show co-localization of Ppp1cc and Tet3 in mESC and Supt16 and Tet3 in NPC. (G) Proximity Ligation Assay of GFP-Tet3 and Ppp1cc in mESCs. Green signals represent GFP-Tet3, whereas the red spots mark the interaction of Tet3 with Ppp1cc.

Most of the NuRD complex members have already been discovered as Tet interactors, validating our data (Fig. 19D). Therefore, we chose Ppp1cc of the PTW/PP1 complex to further validate the proteomics data. This complex is of particular interest as it consists of many phosphatases of the Ppp1 family, which could be involved in the regulation of the Tet proteins. We performed a reversed pull down experiment with an antibody against Ppp1cc. As shown in Fig. 19E, Tet3 was indeed identified in the Ppp1cc Co-IP dataset. It was identified based on 13 unique peptides. The interaction of Tet3 with Ppp1cc in mESCs was further confirmed by ICC and fluorescence microscopy (Fig. 19F). The colocalized pixel map shows overlapping signals of two different antibodies from grey to white. A white dot signifies that the signal of both antibodies is very strong at this spot, which indicates a clear co-localization. Moreover, we conducted a proximity ligation assay (PLA) in primed to substantiate further the physical interaction of Tet3 with Ppp1cc. The confocal analysis indeed provided a strong co-localization signal (Fig. 19G). Interesting are also the data that we obtained for Tet3 and Tet3$^{\text{-CXXC}}$ in NPCs. Here, a strong interaction with Supt16 and Ssrp1, which are the components of the FACT complex (Fig. 19B, Table S5, S6), is detected. The FACT complex is essential for productive RNA Pol-II transcription on chromatinized DNA in non-terminally differentiated cells.[443] This result supports the idea that the function of Tet3 and Tet3$^{\text{-CXXC}}$ is to operate on chromatinized DNA. In order to gain independent support for the Tet3-FACT interaction, we performed again a reversed Co-IP experiment (Fig. 19E) with an anti-Supt16 antibody and ICC in NPC (Fig. 19F) confirming Tet3 as an interaction partner of Supt16. To gain further support for the association of Tet3 with heterochromatinized DNA, we used ICC to examine the nuclear localization of Tet1 and Tet3 in heterochromatin. We used an antibody against the heterochromatin marker H3K9me3 to localize heterochromatin areas inside the cell nucleus (Fig. 20). A second antibody against Tet1 and Tet3 was used to localize the different Tet enzymes. As depicted in Fig. 20, the obtained data confirm that Tet3 in contrast to Tet1 indeed co-localizes with the heterochromatin marker H3K9me3. Which fully supports the proteomics data.

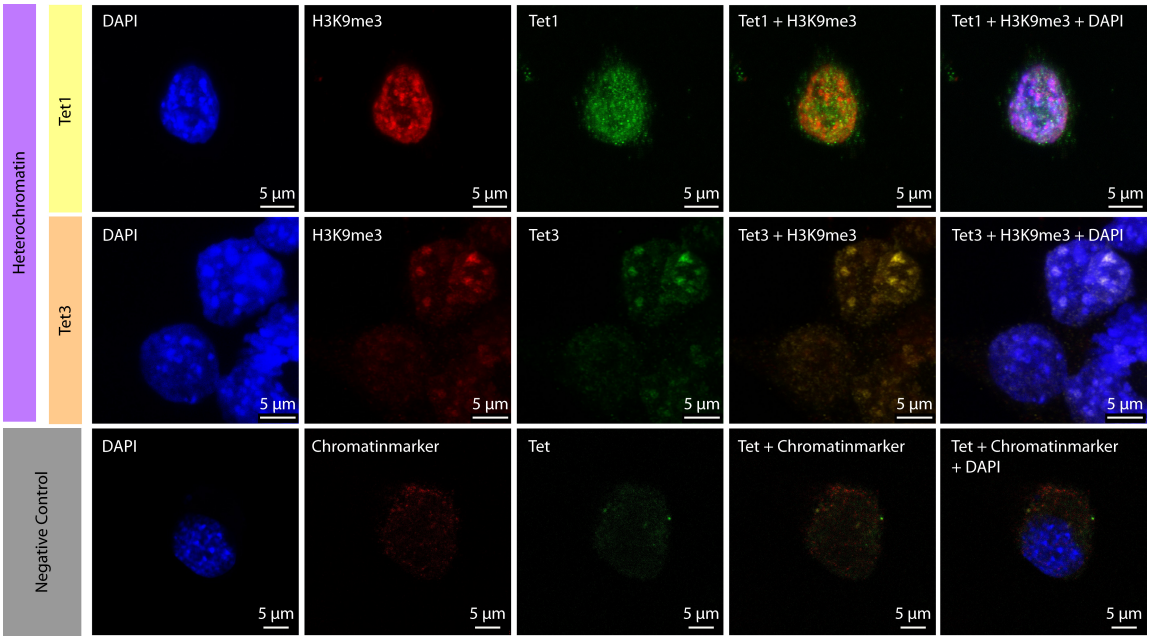


Figure 22: Tet3 co-localizes with the heterochromatin marker H3K9me3 in the nucleus of mouse embryonic stem cells (mESCs). The cells were treated with anti-H3K9me3 antibodies (red) and anti-Tet1 (green) or anti-Tet3 (green) antibodies respectively. Hoechst 33342 was used to stain DNA. The negative control was performed by using only secondary antibodies and were recorded with the same settings.

9.4 Discussion

The large-scale proteomics data show distinct cell type specific and Tet paralog-specific interactomes for Tet1, Tet3 and Tet3^{-CXXC}. Notably, we found many interactors involved in the transfer and recognition of phosphate groups. By adding (kinases Camk2b and Csnk2a1), removing (phosphatases Ppp1cc and Ppp1ca) or protecting (Ywhag) phosphate groups, the activity of the Tet enzymes could be modulated. Another interesting interactor is Dnmt1, which maintains the methylation status after replication. Further experiments are needed to prove that this is a functional interaction in vivo, which could be a way to inherit not only the methylation, but also the hydroxymethylation status of the DNA. Moreover, our data uncover that the Tet3 interactome is dominated by proteins regulating chromatin structure and DNA accessibility in a chromatinized environment. Furthermore, the data suggest that Tet1 operates mainly on DNA that is present in an open, active chromatin structure to maintain the hmC levels.[437] This

is in accord with its dominant expression in mESCs, where the genome has been proposed to be only loosely chromatinized.[444–446] During the differentiation process, which involves both gene activation and silencing, Tet3 is upregulated.[80,246,265,414,434] In agreement with this expression profile, we observe that Tet3 is able to interact intensively with chromatin remodeling factors (SWI/SNF), histone-modifying complexes (NuRD, NuA4) and with proteins of the PTW/PP1 complex. All of these regulate chromatin structure and DNA accessibility. Although a few of these interactions were also observed with Tet1, they are significantly stronger for Tet3 (student´s t-test p-value $<$ 0.01). Upregulation of Tet3 during stem cell development finds a rationale in this way as it allows accessing more strongly chromatinized genome areas. The idea that Tet3 operates on chromatinized DNA was finally confirmed by immunocytochemistry and fluorescence microscopy showing that Tet3 colocalizes with the heterochromatin maker H3K9me3. In support of this conclusion, the NPC data show a robust interaction of Tet3 and Tet3$^{\text{-CXXC}}$ with Ssrp1 and Supt16, which form the FACT complex that is required for transcription on chromatinized DNA. Our data thus support the idea that Tet3 and Tet3$^{\text{-CXXC}}$ in NPCs are the Tet family members responsible for the production of hmC in transcribed gene bodies.[434,447]

9.5 Contributions

Dr. Andrea Künzel and *Franziska Traube* performed the experiments and the proteomics measurements and analyzed data. *Victoria Split*h and *Dr. Stylianos Michalakis* did the fluorescence microscopy. *Edris Parsa* performed the Tet activity assays. *Dr. David Eisen*, *Michael Stadlmeier* and *Dr. Benjamin Hackner* helped with the mass spectrometry studies. *Angie Kirchner* performed cell culture experiments. *Dr. Markus Müller* and *Dr. Stylianos Michalakis* discussed and interpreted results. *Dr. Fabio Spada* performed and supervised the cell culture studies and discussed and interpreted data. *Dr. Jürgen Cox* helped with the statistical analysis of the data. *Prof. Dr. Thomas Carell* designed and supervised the study and interpreted data.

9.6 Supplementary Data

Table S1:Selection of gene ontology (GO) terms for the proteins in the green box of the cluster analysis in Fig. 16C

GO term	protein count (GO term)	p% (count GO termin/count total list)	p-value
BP: gene expression	8	80	3.4×10^{-3}
CC: chromosome	3	30	7.1×10^{-2}
MF: nucleic acid binding	8	80	9.0×10^{-4}
MF: protein binding	8	80	5.4×10^{-2}

The number of proteins enriched for a specific GO term (protein count), the percentage of the protein count enriched for the specific GO term compared to the total number of proteins in the box and the p-value for the enrichment are given. BP = biological process, CC = cellular compartment, MF = molecular function

Table 1: Table S2: Selection of gene ontology (GO) terms for the proteins in the cyan box of the cluster analysis in Fig. 16C.

GO term	protein count (GO term)	p% (count GO termin/count total list)	p-value
CC: nucleus	9	90	1.2×10^{-3}
MF: nucleic acid binding	9	90	6.6×10^{-5}

The number of proteins enriched for a specific GO term (protein count), the percentage of the protein count enriched for the specific GO term compared to the total number of proteins in the box and the p-value for the enrichment are given. CC = cellular compartment, MF = molecular function

Table S3: Selection of gene ontology (GO) terms for the proteins in the purple box of the cluster analysis in Fig. 16C.

GO term	protein count (GO term)	p% (count GO termin/count total list)	p-value
BP: chromatin remodeling	2	29	4.2×10^{-2}
CC: chromosome	4	57	2.2×10^{-3}
MF: transcription coactivator activity	3	43	5.5×10^{-4}

The number of proteins enriched for a specific GO term (protein count), the percentage of the protein count enriched for the specific GO term compared to the total number of proteins in the box and the p-value for the enrichment are given. BP = biological process, CC = cellular compartment, MF = molecular function

Table S4: Selection of gene ontology (GO) terms for the proteins in the green box of the cluster analysis in Fig. 16C.

GO term	protein count (GO term)	p% (count GO termin/count total list)	p-value
BP: gene expression	7	88	3.1×10^{-3}
CC: nucleus	2	88	7.8×10^{-3}
MF: DNA binding	4	50	4.5×10^{-2}

The number of proteins enriched for a specific GO term (protein count), the percentage of the protein count enriched for the specific GO term compared to the total number of proteins in the box and the p-value for the enrichment are given. BP = biological process, CC = cellular compartment, MF = molecular function

Table S5: Log2(GFP-Tet/GFP) values of the chromatin remodelling complex members which are more than twofold enriched and have a p-value < 0.05 in at least one of the Tet co-IPs compared to the control.

Complex Members	Tet1 Co-IP mESC	Tet3$^{\text{-CXXC}}$ Co-IP mESC	Tet3 Co-IP mESC	Tet1 Co-IP NPC	Tet3$^{\text{-CXXC}}$ Co-IP NPC
Actl6a	pos	1.34	pos	neg	neg
Arid1a	pos	0.92	1.67	neg	neg
Dmap1	pos	3.00	1.84	pos	pos
Hdac2	pos	2.23	neg	1.98	1.98
Mbd3	1.11	pos	neg	pos	1.92
Mta3	pos	1.66	3.20	neg	neg
Ppp1cc	neg	1.84	1.99	neg	neg
Rbbp7	pos	2.39	3.02	pos	3.25
Ruvbl2	0.99	0.86	1.07	neg	neg
Smarcc1	0.62	0.99	0.89	pos	pos
Tox4	pos	neg	neg	1.09	pos
Wdr82	pos	pos	1.69	pos	3.09
Ruvbl1	pos	1.34	pos	neg	neg
Morf4l1	pos	1.63	neg	neg	pos
Ppp1ca	0.83	1.31	0.87	neg	pos
Rbbp4	1.08	1.51	0.94	pos	pos
Mta2	pos	pos	pos	neg	pos
Ppp1r12a	neg	neg	neg	pos	pos
Smarcd1	0.51	0.83	pos	neg	neg

Complex Members	Tet1 Co-IP mESC	Tet3^{-CXXC} Co-IP mESC	Tet3 Co-IP mESC	Tet1 Co-IP NPC	Tet3^{-CXXC} Co-IP NPC
Smarcc2	2.24	neg	pos	neg	pos
Smarca4	pos	pos	pos	neg	pos
Ppp1r10	neg	pos	pos	neg	neg
Smarcb1	neg	pos	pos	neg	neg
Gatad2a	neg	neg	neg	pos	pos
Chd4	pos	pos	pos	pos	pos
Mta1	pos	pos	pos	pos	pos
Hdac1	pos	pos	neg	pos	2.57
Meaf6	pos	2.27	neg	pos	pos
Smarce1	neg	pos	neg	pos	neg
Supt16	pos	pos	neg	1.75	2.36
Ssrp1	pos	1.42	pos	1.84	1.84

log2(GFP-Tet/GFP) values for the Tet1, Tet3 and Tet3^{-CXXC} Co-IPs in mouse embryonal stem cells (mESCs) and neural progenitor cells (NPCs) are given.

Table S6: Tet normalized values of the chromatin remodelling complex members which are more than twofold enriched and have a p-value < 0.05 in at least one of the Tet Co-IPs compared to the control.

Complex Members	Tet1 mESC normalized	Tet3$^{\text{-CXXC}}$ mESC normalized	Tet3 mESC normalized	Tet3 NPC normalized	Tet3$^{\text{-CXXC}}$ NPC normalized
Ppp1cc		0.30	0.28		
Actl6a		0.22			
Arid1a		0.15	0.24		
Dmap1		0.48	0.26		
Hdac2		0.36		0.21	0.21
Mbd3	0.12				0.20
Mta3		0.27	0.45		
Rbbp7		0.38	0.43		0.34
Ruvbl2	0.10	0.14	0.15		
Smarcc1	0.06	0.16	0.13		
Tox4				0.12	
Wdr82			0.24		0.32
Ruvbl1		0.22			
Morf4l1		0.26			
Ppp1ca	0.09	0.21	0.12		
Rbbp4	0.11	0.24	0.13		
Smarcc2	0.23				
Smarcd1	0.05	0.13			
Meaf6		0.37			

Complex Members	Tet1 mESC normalized	Tet3^{-CXXC} mESC normalized	Tet3 mESC normalized	Tet3 NPC normalized	Tet3^{-CXXC} NPC normalized
Ssrp1		0.23		0.20	0.19
Supt16				0.19	0.24

The normalized values for the Tet1, Tet3 and Tet3^{-CXXC} Co-IPs in mouse embryonal stem cells (mESCs) and neural progenitor cells (NPCs) were calculated by dividing the log2 value of the complex member by the log2 value of Tet1, Tet3 or Tet3^{-CXXC}.

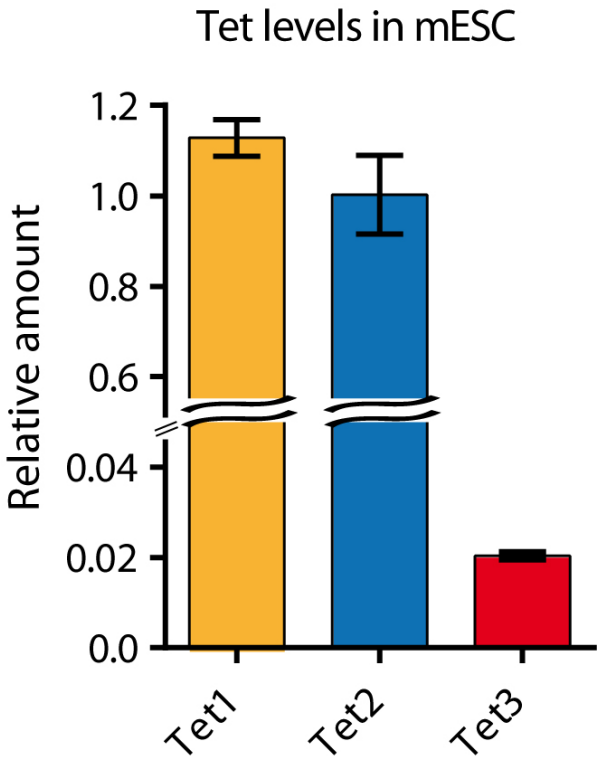


Figure S1: Tet levels in mouse embryonic stem cells (mESC). RT-qPCR analysis of Tet1, Tet2 and Tet3 mRNA levels in mESCs shows that Tet1 and Tet2 are highly expressed whereas the amount of Tet3 mRNA is reduced but not zero.

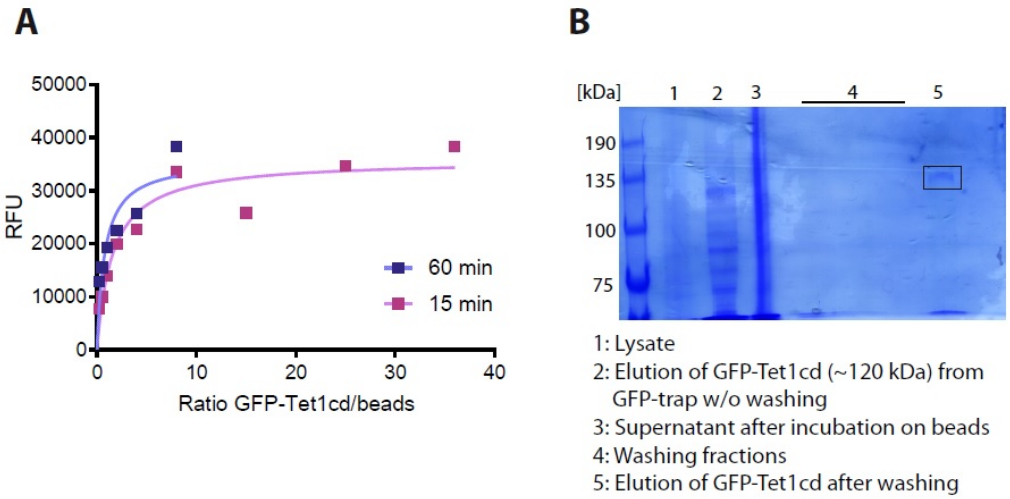


Fig. S2: Validation of the Co-Immunoprecipitation workflow using Tet-loaded beads.
(A) Saturation of GFP-Tet1cd on GFP nanobody magnetic beads. (B) Successful purification of GFP-Tet1cd shown by Coomassie stained 5 % SDS PAGE gel.

Figure S2: Validation of the Co-Immunoprecipitation workflow using Tet-loaded beads. (A) Saturation of GFP-Tet1cd on GFP nanobody magnetic beads. (B) Successful purification of GFP-Tet1cd shown by Coomassie stained 5% SDS PAGE gel.

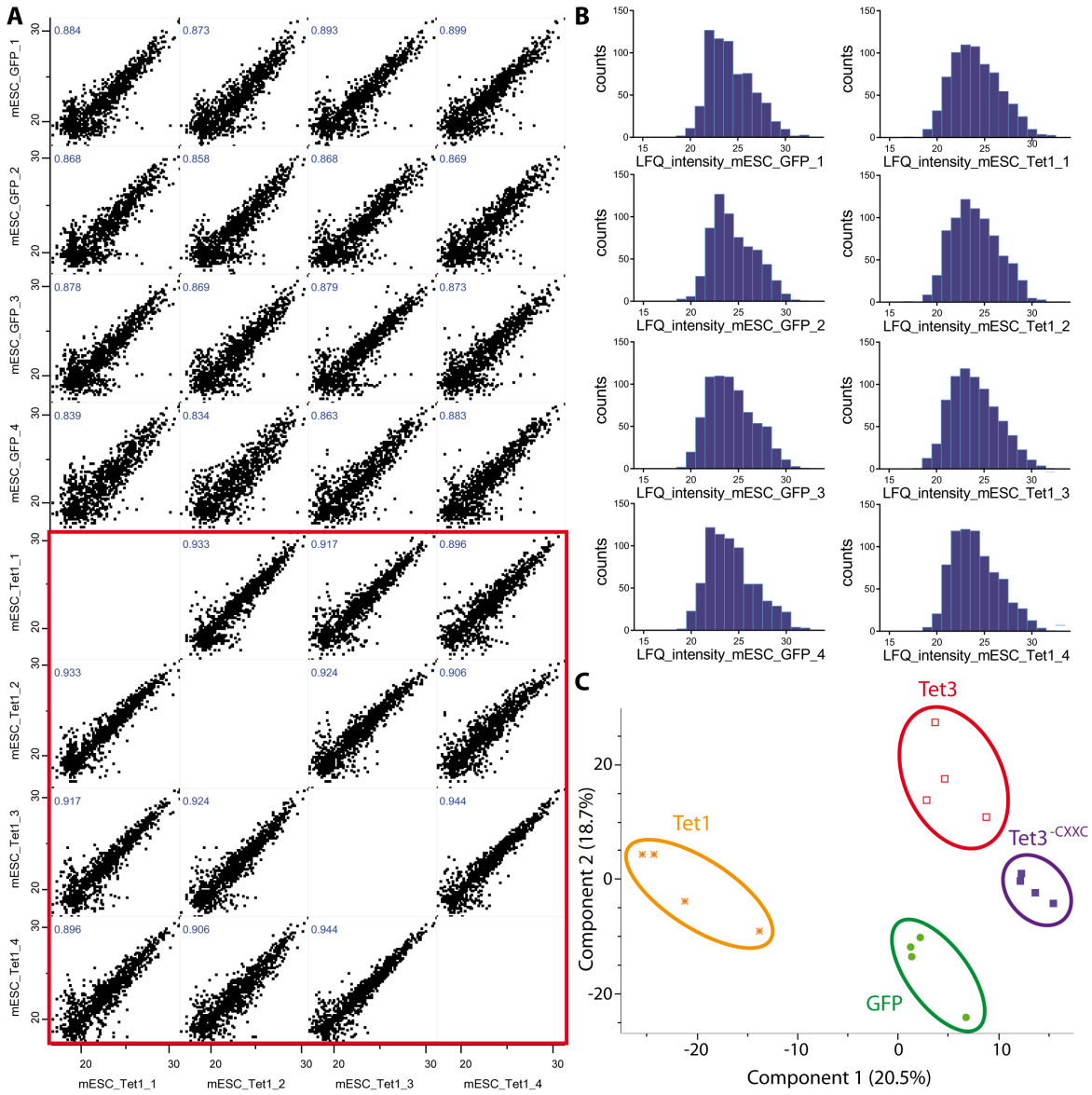


Figure S3: Data quality.(A) Multi scatter plot comparing the LFQ intensities from the four biological replicates of the Tet1 experiment with itself (red rectangle) and the control GFP Co-Immunoprecipitation (Co-IP) in mouse embryonic stem cells (mESCs). Blue numbers indicate the Pearson correlation. (B) Histograms showing counts of LFQ intensities in the four different biological replicates of the control (GFP) and the Tet1 Co-IPs in mESCs. (C) Principal component analysis of the data from the different Tet Co-IPs and the GFP control experiment in mESCs.

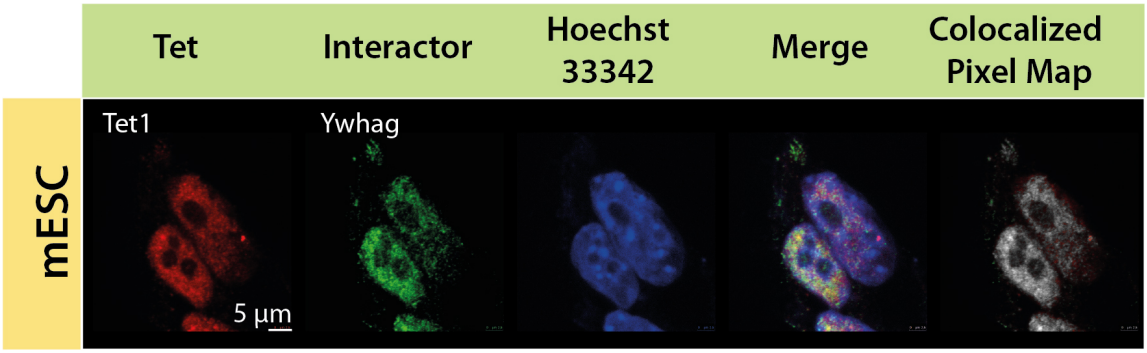


Figure S4: Co-localization of Tet1 and Ywhag in the mouse embryonic stem cells (mESCs). Immunocytochemistry in combination with fluorescence microscopy confirms that Tet1 and Ywhag interact in mESCs.

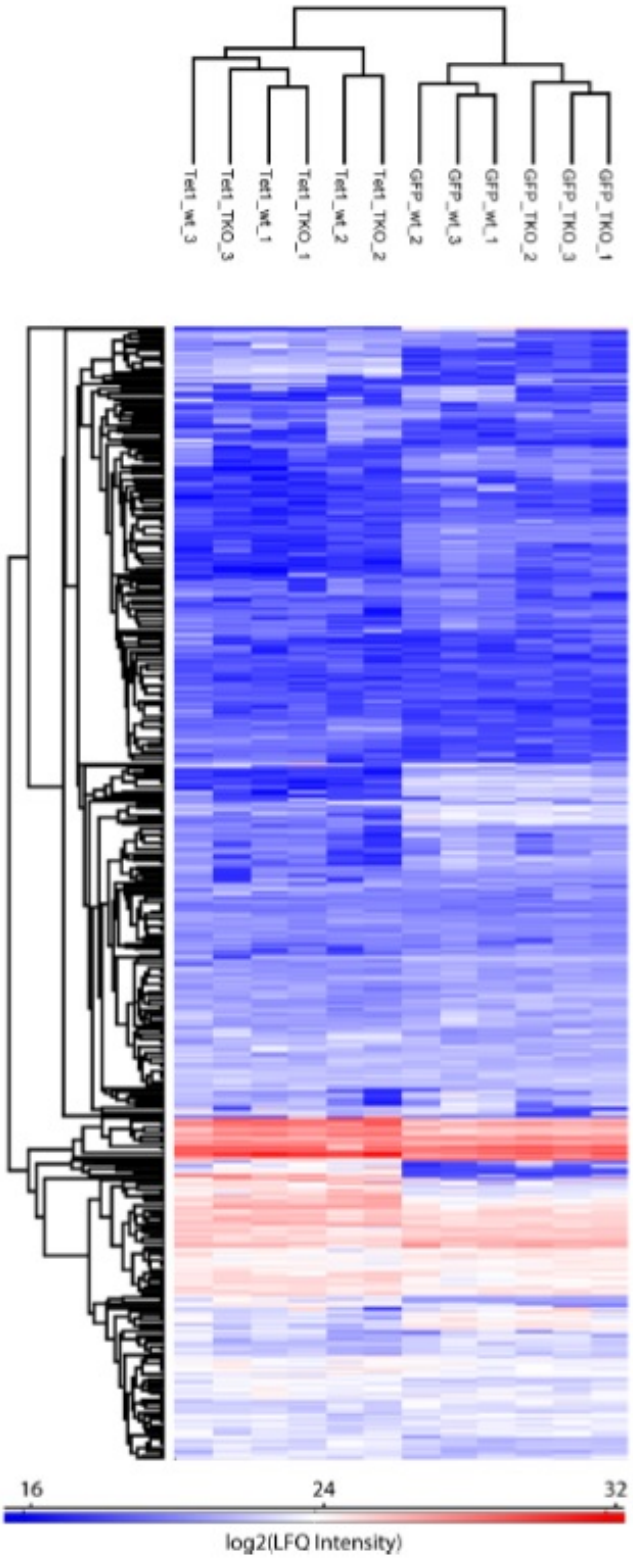

Tet1_wt_3
Tet1_TKO_3
Tet1_wt_1
Tet1_TKO_1
Tet1_wt_2
Tet1_TKO_2
GFP_wt_2
GFP_wt_3
GFP_wt_1
GFP_TKO_2
GFP_TKO_3
GFP_TKO_1
16
24
32
log2(LFQ Intensity)

Figure S5: The Co-Immunoprecipitation workflow using Tet-covered beads is not biased by competing endogenous Tet.(A) Multi scatter plot comparing the LFQ intensities from the four biological replicates of the Tet1 experiment with itself (reCluster heat map comparing three replicates of the Tet1 Co-IP and the GFP control in wildtype (wt) and triple knockout (TKO) mouse embryonic stem cells.

10 Conserved Phosphorylations of TET Enzymes in the Active Center control their activity

10.1 Introduction

Research in DNA epigenetics gained new focus when Tahiliani *et al.* and Kriaucionis *et al.* identified 5-hydroxymethylcytosine (hmC) as the direct, enzymatic oxidation product of 5-methylcytosine (mC).[78,79] The methylation of cytosine at position 5 by DNA methyltransferases (DNMTs) contributes significantly to the epigenetic machinery in higher organisms. The subsequent oxidation to hmC, fC and caC is carried out by a family of α-ketoglutarate dependent oxygenases, namely TET enzymes.[80–82,239] In vertebrates, three TET enzymes are known with distinct expression patterns and functions. While TET1 and TET2 are expressed in murine embryonic stem cells (mESCs) and are considered to oxidize transcription start sites (TSS) and gene bodies, TET3 is expressed in the brain and oocyte where it oxidizes the paternal pronuclear DNA.[326,327,448] Controlling the activity of these oxidation events is critical both in a spatial and timely manner. Several small molecules like 2-hydroxyglutarate or ascorbate can influence TET activity.[281,286,287,449–451] Unpublished work by the Carell group further investigated the regulation of TET enzymes by metabolites and interacting proteins (see chapter 8 and 9). Another general possibility for the control of an enzymes activity are post-translational modifications (PTMs). They are dynamic and easy to write and erase. PTMs have diverse roles and are ubiquitous in nature. A wide range of modifications are known, e.g. methylation, acylation, glycosylation, ubiquitination and phosphorylation.[452–457] One prominent example for the impact of post-translational modifications upon cellular processes, is the phosphorylation of the C-terminal domain (CTD) of RNA polymerase II (RNAPII). The CTD of RNAPII is modified at various sites at different time points of transcription, resulting in diverse effects.[458,459] As the CTD of RNAPII exhibits a homology to parts of the C-terminal catalytic domain of TET1 it is intriguing to investigate this motif for phosphorylation patterns.[263]

10.2 Results

We first characterized the phosphorylation patterns of TET1, TET2 and TET3 in full length and also the catalytic domain of TET1 (TET1cd). These proteins were either expressed, as described in chapter 8.8, via overexpression in HEK293T cells or isolated in a more native state from mESCs without overexpression (in collaboration with the Leonhardt group, LMU Munich).[460] Subsequently the enzymes were analyzed by LC-MS/MS together with *Michael Stadlmeier* for their PTMs. As we were interested in general mechanistic insights, we focused our interest on the catalytic domain of the TET enzymes and concentrated on highly conserved motifs, between mouse, human and all three TET enzymes. A small excerpt of the region of interest is depicted in Figure 21. One specific amino acid, S1874, located in a highly conserved peptide (W**S**DSFE) was found to be phosphorylated in TET1, in all peptides. In case of TET2 and TET3 this motif was identified in both, the phosphorylated state and unmodified.

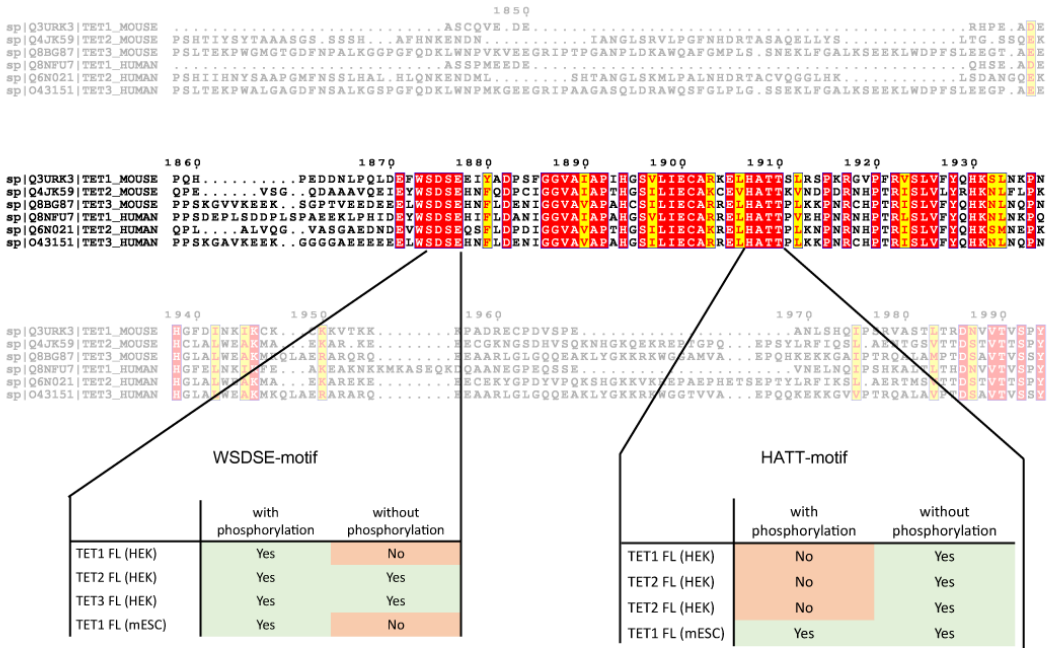


Figure 23: TET enzymes were isolated from different cell lines and analyzed for phosphorylation sites. Fully conserved amino acids are marked in red. Two phosphorylations in highly conserved regions were identified, namely Serin 1874 and either threonine 1909 or 1910. In part phosphorylated species as well as unphosphorylated peptides were identified.

Also the endogenous full-length TET1, isolated from mESCs was phosphorylated at S1874. In this sample however, another conserved peptide, HA**TT** could be identified in both, the phosphorylated, as well as the unmodified state. It was, however, not clear, whether the first or second threonine was phosphorylated. Samples isolated out of HEK cells did not show this phosphorylation. In order to gain insight into the function of these phosphorylations we conducted mutation studies. We started our study by mutating S1874, which, in case of TET1, was always observed in a phosphorylated state. We generated both, the alanine and glutamate mutant, mimicking the non-phosphorylated and phosphorylated state of the amino acid, respectively. After transient expression of the mutated protein constructs, we measured the levels of the oxidation products of mC, as described earlier by Carell and co-workers.[83] The expression levels of the wild type and mutants were normalized to the GFP-fluorescence, as described in chapter 8.8

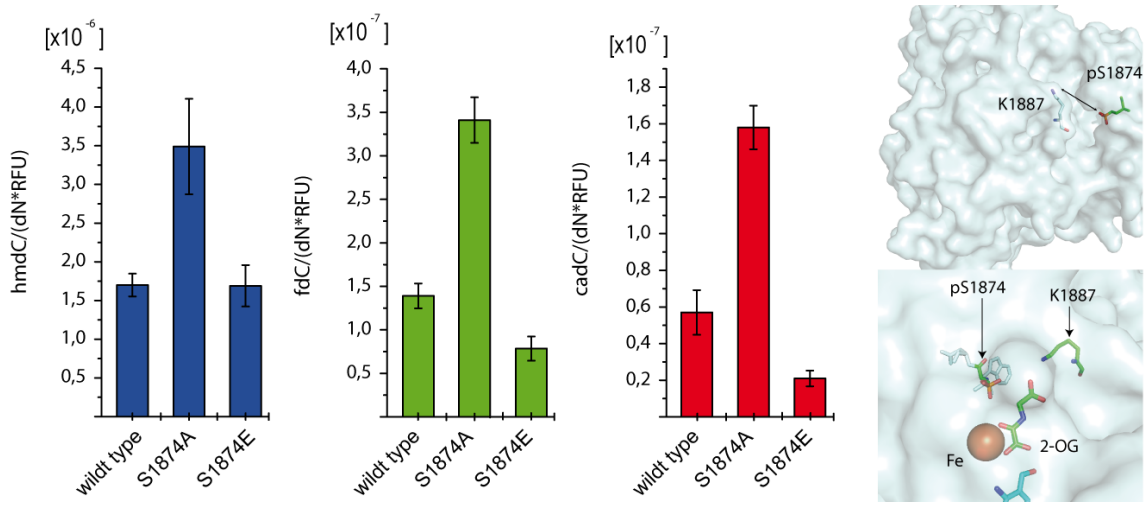


Figure 24: Mutating S1874 to either alanine or glutamate impacts TET activity. S1874A is more active, while the phosphorylation-mimic S1874E behaves almost like the wild type. The enhanced activity of S1874A may be explained by its position. S1874 is located opposite of K1887 in a groove. A phosphorylation (S1874E) could "close" this groove, which is located near the active center. This groove may be involved in co-substrate supply, hence by the attraction of the negatively charged phosphate and positively charged lysine residue, this supply may be limited. This is not possible for S1874A, hence the groove would permanently be "open", therefore enhancing TET activity.

The alanine mutant (S1874A) increases TET activity, while the activity of the glutamate mutant (S1874E) is comparable to the wild type. The latter was expected, as this site was (in case for TET1) always characterized in a phosphorylated state. The posi-

tion of S1874 in the TET2 crystal structure may explain these results. S1874 is located opposite of a lysine (K1887) in a groove. This structural element is located near the active center, where the iron is coordinated by the co-substrate α-ketoglutarate (2-OG) (see Figure 22). It is possible that this groove is involved in co-substrate supply and therefore it may be involved in TET regulation. If S1874 is phosphorylated (pS1874) it may attract, with its negatively charged phosphate group, the positively charged, flexible lysine side-chain. This could close the cavity, hence perturbating the supply of α-ketoglutarate to the active center. As no phosphorylation is possible for S1874A, the groove would always be in an "open" state, which may explain the enhanced activity of TET.

Besides the EW**S**DS-motif, we identified a phosphorylated threonine in the HA**TT** motif from stem-cell TET1 full length. This site was found in both, a phosphorylated and unphosphorylated state. As the exact position of the phosphorylation was not clear, both sites were subjected to mutation studies.

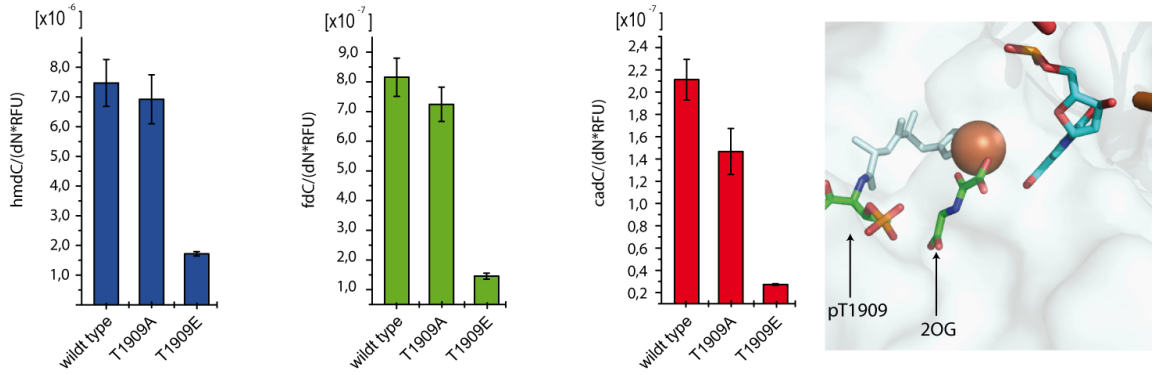


Figure 25: Mutating T1909 to either alanine or glutamate impacts TET activity. While T1909A behaves like the wild type, T1909E is catalytically almost inactive. The crystal structure reveals the close proximity of T1909 to the iron center, which is coordinated with α-ketoglutarate (2-OG). If T1909 is phosphorylated (pT1909) the binding of α-ketoglutarate may be disturbed, due to the negative charges and steric demand of the phosphate group. This could explain the inactive enzyme, as no oxidation is possible without the co-substrate.

The phospho-mimic T1909E has a negative effect upon oxidation of mC. Localizing T1909 in the crystal structure reveals a very close proximity to the iron center, where α-ketoglutarate is coordinated. If phosphorylated (pT1909), there might be repelling

effects of the negatively charged phosphate and the negativity charged α-ketoglutarate. Also the large phosphate group may sterically hinder the coordination of α-ketoglutarate to the iron. This would explain the almost catalytically dead T1909E mutant. As the position of the phosphorylation was not certain, we also investigated the effects of the mutations on the amino acid T1910. Based on its position in the crystal structure, the same trend as for T1909 is expected for T1910. It is also in close proximity to the iron center (see Figure 24).

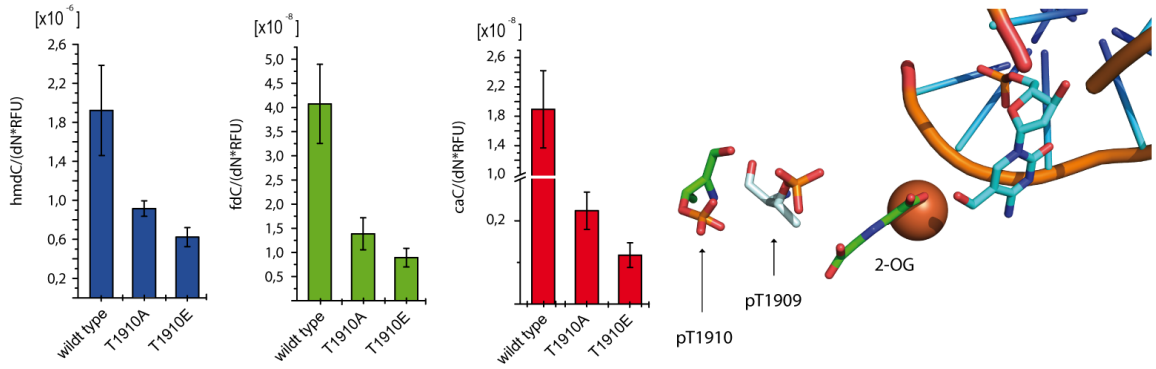


Figure 26: Mutating T1910 to either alanine or glutamate does show the expected trend. As for T1909 the alanine mutant of T1910 is more active than the glutamate mutant. However the general activity of both mutants was lower compared to the wild type, which may be due to structural perturbations of the mutants.

As depicted in Figure 24, the same effect as for T1909 (T1880 in TET2) is visible, however the general activity of the mutated TET enzymes was lower. This may be due to several reason, e.g. the mutation may interfere with protein folding.

10.3 Summary and Outlook

It is to date an unsolved question how TET enzymes are fine-tuned in their activity. Besides the regulation on the transcription/translation level, it seems plausible to regulate the activity of TET enzymes in a reversible, dynamic way, by post-translational modifications. A comprehensive analysis of the phosphorylation patterns of TET1, TET2 and TET3 identified highly conserved motifs that are (in part) dynamically phosphorylated. This study could, for the first time, demonstrate how single phosphorylations in the

active center of TET enzymes can alter their activity significantly. Highly conserved motifs around the active site (WSDSE, HATT) may therefore be involved in regulating the activity of TET enzymes. In all three investigated amino acid positions (S1874, T1909 and T1910), the mutation to glutamate (phosphorylation-mimic) had a negative effect on the oxidation of mC. It seems plausible that the cell may "activate" TET enzymes by removing a phosphate in the active center and subsequently "switching it off" by a simple phosphorylation, e.g. at T1909. A dephosphorylation of S1874 may boost TET activity, if massive oxidation is necessary.
Further research should concentrate on elucidating how these phosphorylation patterns are established and controlled, both in regard of the responsible kinases and phosphatases, but also regarding the time-points. Initial proteomic studies performed by *Dr. Andrea Künzel* and *Franziska Traube* identified several kinases and phosphatases, which interact with TET1cd and indeed preliminary data suggest that CDK11 may be responsible for TET1cd phosphorylation (see Figure 25). *Franziska Traube* performed

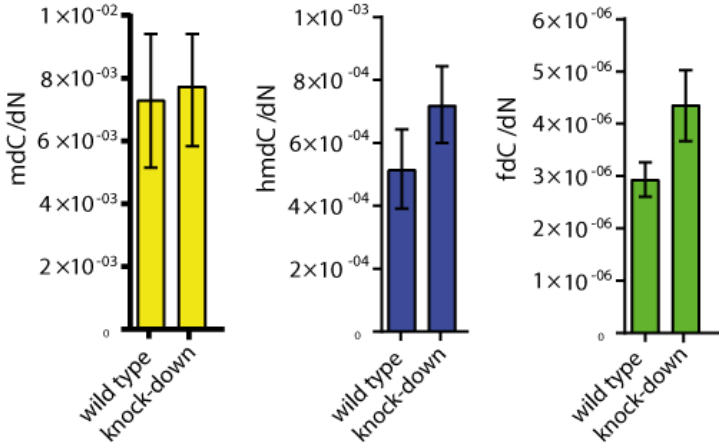


Figure 27: esi-RNA mediated knock-down of CDK11 results in elevated levels of hmC and fC compared to the control, supporting the data obtained by mutation studies.

esi-RNA mediated knock-down experiments for CDK11, resulting in elevated levels of hmC and fC. This fits well to the mutation studies, as a knock-down of CDK11 would result in less phosphorylations. The mutations of TET1cd (Figure 22, Figure 23 and Figure 24) all show that a phosphorylation has a negative impact upon oxidation of mC. However, more detailed studies are necessary in order to validate the interaction between TET1cd and CDK11 and to further elucidate the biochemical mechanism of TET1cd phosphorylation.

10.4 Experimental Procedures

Michael Stadlmeier performed LC-MS sample preparations, measurements and data evaluation. *Franziska Traube* assisted by cell culture- and protein purification-support. The initial development of the transfection and purification protocol, as described in chapter 8.8 and Schröder *et al.* was supported by *Dr. Benjamin Hackner.*[461]

10.4.1 Transfection and flow cytometric measurements

The GFP-TET1cd construct[83] was transfected according and subsequently the GFP levels were determined by flow cytometry as described in chapter 8.8. The following isolation of genomic DNA and quantification was conducted as described in Pfaffeneder *et al.* and chapter 8.8.[83]

10.4.2 Generation of mutated versions of TET1cd

The GFP-TET1cd construct[83] was too large in order to directly generate mutated versions of the protein by means of site directed mutagenesis. Therefore a small fragment of the DNA sequence was synthesized *de novo* (GeneArt) and used for site directed mutagenesis (Agilent Quick Change II) according to the manufacturers protocol. This fragment was cloned into the TET1cd-backbone by the use of NotI and SpeI.

TET1cd fragment (base 4008-4588 of the GFP-Tet1cd construct[83])

DNA sequence
The restriction sites are depicted bold.

ACTAGTCCATCTGAGCAGCTAACTTCTAACCAGTCAAACCAGCAGCTCCCTCT
CCTCAGCAATTCTCAGAAACTGGCTTCCTGTCAGGTGGAAGATGAGCGGCACC
CTGAAGCGGATGAGCCTCAGCACCCCGAGGACGATAACTTGCCTCAACTTGAT
GAATTCTGGTCAGACAGTGAGGAGATCTACGCCGATCCTTCCTTTGGTGGCGT
GGCGATAGCACCCATTCACGGCTCGGTGCTCATTGAGTGCGCTCGGAAGGAGC

TTCATGCTACCACCTCTTTGCGCTCCCCCAAACGAGGGGTCCCTTTTCGTGTG
TCCCTTGTATTCTACCAGCACAAAAGCCTAAACAAGCCTAATCATGGTTTTGA
TATCAACAAAATTAAGTGTAAATGCAAAAAAGTAACGAAAAAAAAGCCCGCAG
ACCGGGAGTGTCCTGATGTATCCCCCGAAGCCAATTTATCACACCAAATTCCT
TCTCGAGTTGCATCAACCTTAACCCGAGACAATGTTGTTACCGTGTCCCCATA
CTCTCTCACTCATGTTGCGGGACCCTACAATCGTTGGGTCTAA**GCGGCCGC**

Protein sequence
UniProtKB-Q3URK3 (TET1 - MOUSE)

T S P S E Q L T S N Q S N Q Q L P L L S N S Q K L A S C Q V E D E R
H P E A D E P Q H P E D D N L P Q L D E F W **S** D S E E I Y A D P S F G G V
A I A P I H G S V L I E C A R K E L H A **T** **T** S L R S P K R G V P F R V S L V
F Y Q H K S L N K P N H G F D I N K I K C K C K K V T K K K P A D R E C P
D V S P E A N L S H Q I P S R V A S T L T R D N V V T V S P Y S L T H V A G
P Y N R W V

DNA Primers used for the site directed mutagenesis

Primer	Sequence (5'-3')
S1874A-FWD	AGATCTCCTCACTGTCTGCCCAGAATTCATCAAGTTG
S1874A-REV	CAACTTGATGAATTCTGGGCAGACAGTGAGGAGATCT
S1874E-FWD	GGCGTAGATCTCCTCACTGTCCTCCCAGAATTCATCAAGTTGAGG
S1874E-REV	CCTCAACTTGATGAATTCTGGGAGGACAGTGAGGAGATCTACGCC
T1909A-FWD	CGCAAAGAGGTGGCAGCATGAAGCTCCTTCCG
T1909A-REV	CGGAAGGAGCTTCATGCTGCCACCTCTTTGCG
T1909E-FWD	-GCGCAAAGAGGTCTCAGCATGAAGCTCCTTCCGAGCGC
T1909E-REV	GCGCTCGGAAGGAGCTTCATGCTGAGACCTCTTTGCGC
T1910A-FWD	GGGAGCGCAAAGAGGCGGTAGCATGAAGCTC
T1910A-REV	GAGCTTCATGCTACCGCCTCTTTGCGCTCCC
T1910E-FWD	GTTTGGGGGAGCGCAAAGACTCGGTAGCATGAAGCTCCTTC
T1910E-REV	GAAGGAGCTTCATGCTACCGAGTCTTTGCGCTCCCCCAAAC

Bibliography

[1] C. Darwin, *On the Origin of Species by Means of Natural Selection*, Murray, London, **1859**.

[2] G. Mendel, "Versuche über Pflanzen-Hybriden", *Verhandlungen des naturforschenden Vereines in Brünn* **1866**, *42*, 3–47.

[3] C. Zirkle, "The Inheritance of Acquired Characters and the Provisional Hypothesis of Pangenesis", *Am. Nat.* **1935**, *69*, 417–445.

[4] J. de Monet de Lamarck, *Philosophie zoologique: ou Exposition des considérations relative à l'histoire naturelle des animaux*, Dentu et L'Auteur, **1809**.

[5] C. Darwin, *The Variation of Animals and Plants Under Domestication*, J. Murray, **1868**.

[6] Y. Liu, "A new perspective on Darwin's Pangenesis", *Biol. Rev.* **2008**, *83*, 141–149.

[7] Darwin Correspondence Project, "Letter no. 5918", `http://www.darwinproject.ac.uk/DCP-LETT-5918` (visited on 09/25/2017).

[8] E. J. Richards, "Inherited epigenetic variation - revisiting soft inheritance", *Nat. Rev. Genet.* **2006**, *7*, 395–401.

[9] Y. Liu, "Like father like son", *EMBO Rep.* **2007**, *8*, 798–803.

[10] H. Sano, "Inheritance of acquired traits in plants: Reinstatement of Lamarck", *Plant Signaling & Behavior* **2010**, *5*, 346–348.

[11] M. J. West-Eberhard, "Toward a Modern Revival of Darwin's Theory of Evolutionary Novelty", *Philos. Sci.* **2008**, *75*, 899–908.

[12] Y. .-.-S. Liu, X. M. Zhou, M. X. Zhi, X. J. Li, Q. L. Wang, "Darwin's contributions to genetics", *J. Appl. Genet.* **2009**, *50*, 177–184.

[13] G. L. Geison, "Darwin and Heredity: the Evolution of His Hypothesis of Pangenesis", *J. Hist. Med. Allied Sci.* **1969**, *XXIV*, 375–411.

[14] F. Galton, "Experiments in Pangenesis, by Breeding from Rabbits of a Pure Variety, into Whose Circulation Blood Taken from other Varieties Had Previously Been Largely Transfused", *Proc. Roy. Soc. London* **1870**, *19*, 393–410.

[15] M. Bulmer, "The Development of Francis Galton's Ideas on the Mechanism of Heredity", *J. Hist. Biol.* **1999**, *32*, 263–292.

[16] C. Darwin, "Pangenesis", *Nature* **1871**, 502–503.

[17] A. Weismann, B. C. J. G. Knight, S. Poulton, Edward Bagnall, S. Schönland, S. Shipley, A. E., "Essays upon heredity and kindred biological problems", **1889**.

[18] B. A. Pierce, *Genetic: A Conceptual Approach*, 2nd ed., W.H.Freeman & Co Ltd, **2005**.

[19] H. D. Vries, *Intracellular Pangenesis*, **1889**.

[20] W. Johannsen, *Elemente der exakten Erblichkeitslehre*, Gustav Fischer, Jena, **1909**.

[21] F. Miescher, "Über die chemische Zusammensetzung der Eiterzellen", *Med.-Chem. Untersuchungen* **1871**, *4*, 441–460.

[22] F. Miescher, *Die histochemischen und physiologischen Arbeiten von Friedrich Miescher*, Vogel, Leipzig, **1897**.

[23] R. Dahm, "Friedrich Miescher and the discovery of DNA", *Dev. Biol.* **2005**, *278*, 274–288.

[24] W. Flemming, *Zellsubstanz, Kern und Zelltheilung*, Vogel, **1882**.

[25] E. W. Crow, J. F. Crow, "100 years ago: Walter Sutton and the chromosome theory of heredity", *Genetics* **2002**, *160*, 1–4.

[26] T. Boveri, *Ergebnisse über die Konstitution der chromatischen Substanz des Zellkerns*, G. Fischer, Jena, **1904**.

[27] W. S. Sutton, "On the morphology of the chromosome group in Brachystola magna", *Biol. Bull.* **1902**, *4*, 24–39.

[28] A. Sturtevant, "The linear arrangement of six sex-linked factors in Drosophila, as shown by their mode of association", *J. Exp. Zool.* **1913**, *14*, 43–59.

[29] F. Griffith, "The Significance of Pneumococcal Types", *J Hyg (Lond)* **1928**, *27*, 113–159.

[30] O. T. Avery, C. M. MacLeod, M. McCarty, "Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types: Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from Pneumococcus Type III", *J. Exp. Med.* **1944**, *79*, 137–158.

[31] G. W. Beadle, E. L. Tatum, "Genetic Control of Biochemical Reactions in Neurospora", *Proc. Natl. Acad. Sci. U.S.A.* **1941**, *27*, 499–506.

[32] F. H. C. Crick, "On Protein Synthesis", *Symp. Soc. Exp. Biol.* **1958**, *XII*.

[33] F. H. Crick, "Central dogma of molecular biology.", *Nature* **1970**, *227*, 561–563.

[34] L. T. Chow, R. E. Gelinas, T. R. Broker, R. J. Roberts, "An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA", *Cell* **1977**, *12*, 1–8.

[35] D. Baltimore, "Viral RNA-dependent DNA Polymerase: RNA-dependent DNA Polymerase in Virions of RNA Tumour Viruses", *Nature* **1970**, *226*, 1209–1211.

[36] H. M. Temin, S. Mizutani, "Viral RNA-dependent DNA Polymerase: RNA-dependent DNA Polymerase in Virions of Rous Sarcoma Virus", *Nature* **1970**, *226*, 1211–1213.

[37] C. Tamm, H. S. Shapiro, R. Lipshitz, E. Chargaff, "Distribution Density of Nucleotides within a Desoxyribonucleic Acid Chain", *J. Biol. Chem.* **1953**, *203*, 673–688.

[38] J. Watson, F. Crick, "A Structure for Deoxyribose Nucleic Acid", *Nature* **1953**, 737–738.

[39] M. H. F. Wilkins, A. R. Stokes, H. R. Wilson, "Molecular Structure of Deoxypentose Nucleic Acids", *Nature* **1953**, 738–740.

[40] R. E. Franklin, R. Gosling, "Molecular Configuration in Sodium Thymonucleate", *Nature* **1953**, 740–741.

[41] J. Watson, F. Crick, "Genetical Implications of the structure of Deoxyribonucleic Acid", *Nature* **1953**, 964–967.

[42] R. E. Franklin, R. Gosling, "Evidence for 2-Chain Helix in Crystalline Structure of Sodium Deoxyribonucleate", *Nature* **1953**, 156–157.

[43] M. Meselson, F. W. Stahl, "The Replication of DNA in Escherichia Coli", *Proc. Natl. Acad. Sci. U.S.A.* **1958**, *44*, 671–682.

[44] F. H. Crick, L. Barnett, S. Brenner, R. J. Watts-Tobin, "General nature of the genetic code for proteins", *Nature* **1961**, *192*, 1227–32.

[45] C. Yanofsky, "Establishing the Triplet Nature of the Genetic Code", *Cell* **2007**, *128*, 815–818.

[46] W. M. Jou, G. Haegeman, M. Ysebaert, W. Fiers, "Nucleotide Sequence of the Gene Coding for the Bacteriophage MS2 Coat Protein", *Nature* **1972**, *237*, 82–88.

[47] F. Sanger, S. Nicklen, A. R. Coulson, "DNA sequencing with chain-terminating inhibitors", *Proc. Natl. Acad. Sci. U.S.A.* **1977**, *74*, 5463–5467.

[48] F. Sanger, G. M. Air, B. G. Barrell, N. L. Brown, A. R. Coulson, J. C. Fiddes, C. A. Hutchison, P. M. Slocombe, M. Smith, "Nucleotide sequence of bacteriophage phiX174 DNA", *Nature* **1977**, *265*, 687–695.

[49] J. M. Heather, B. Chain, "The sequence of sequencers: The history of sequencing DNA", *Genomics* **2016**, *107*, 1–8.

[50] A. M. Maxam, W. Gilbert, "A new method for sequencing DNA", *Proc. Natl. Acad. Sci. U.S.A.* **1977**, *74*, 560–564.

[51] K. Mullis, H. Erlich, N. Arnheim, G. Horn, R. Saiki, S. Scharf, 4683202, US Patent 4,683,195, **1987**.

[52] W. A. M. Loenen, D. T. F. Dryden, E. A. Raleigh, G. G. Wilson, N. E. Murray, "Highlights of the DNA cutters: a short history of the restriction enzymes", *Nucleic Acids Res.* **2014**, *42*, 3–19.

[53] S. N. Cohen, A. C. Y. Chang, H. W. Boyer, R. B. Helling, "Construction of Biologically Functional Bacterial Plasmids In Vitro", *Proc. Natl. Acad. Sci. U.S.A.* **1973**, *70*, 3240–3244.

[54] D. V. Goeddel, D. G. Kleid, F. Bolivar, H. L. Heyneker, D. G. Yansura, R. Crea, T. Hirose, A. Kraszewski, K. Itakura, A. D. Riggs, "Expression in Escherichia coli of chemically synthesized genes for human insulin", *Proc. Natl. Acad. Sci. U.S.A.* **1979**, *76*, 106–110.

[55] S. N. Cohen, "DNA cloning: A personal view after 40 years", *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 15521–15529.

[56] R. Fleischmann, M. Adams, O. White, R. Clayton, E. Kirkness, A. Kerlavage, C. Bult, J. Tomb, B. Dougherty, J. Merrick, e. al. et, "Whole-genome random sequencing and assembly of Haemophilus influenzae Rd", *Science* **1995**, *269*, 496–512.

[57] A. Goffeau, B. G. Barrell, H. Bussey, R. W. Davis, B. Dujon, H. Feldmann, F. Galibert, J. D. Hoheisel, C. Jacq, M. Johnston, E. J. Louis, H. W. Mewes, Y. Murakami, P. Philippsen, H. Tettelin, S. G. Oliver, "Life with 6000 Genes", *Science* **1996**, *274*, 546–567.

[58] T. H. G. Project, "Initial sequencing and analysis of the human genome", *Nature* **2001**, *409*, 860–921.

[59] J. C. Venter, M. D. Adams, "The Sequence of the Human Genome", *Science* **2001**, *291*, 1304–1351.

[60] C. Human Genome Sequencing, "Finishing the euchromatic sequence of the human genome", *Nature* **2004**, *431*, 931–945.

[61] A. F. Palazzo, T. R. Gregory, "The Case for Junk DNA", *PLos Genet.* **2014**, *10*, 1–8.

[62] E. P. Consortium, "The ENCODE (ENCyclopedia Of DNA Elements) Project", *Science* **2004**, *306*, 636–640.

[63] E. P. C. The, "Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project", *Nature* **2007**, *447*, 799–816.

[64] J. R. Ecker, W. A. Bickmore, I. Barroso, J. K. Pritchard, Y. Gilad, E. Segal, "Genomics: ENCODE explained", *Nature* **2012**, *489*, 52–55.

[65] E. P. C. The, "An Integrated Encyclopedia of DNA Elements in the Human Genome", *Nature* **2012**, *489*, 57–74.

[66] U. Welsch, T. Deller, *Sobotta Lehrbuch Histologie*, Elsevier,Urban&FischerVerlag, **2011**.

[67] C. H. Waddington, *An introduction to modern genetics*, New York,The Macmillan company, **1939**, p. 456.

[68] R. Holliday, "Epigenetics: An overview", *Dev. Genet.* **1994**, *15*, 453–457.

[69] L. Mohrmann, C. P. Verrijzer, "Composition and functional specificity of SWI2/SNF2 class chromatin remodeling complexes", *Biochim. Biophys. Acta* **2005**.

[70] E. J. Yoo, Y. K. Jang, M. A. Lee, P. Bjerling, "rp3, a chromodomain helicase/ATPase DNA binding protein, is required for heterochromatin silencing in fission yeast.", *Biochem. Biophys. Res. Commun.* **2002**, *295*, 970–974.

[71] J. Walfridsson, O. Khorosjutina, P. Matikainen, C. M. Gustafsson, K. Ekwall, "A genome-wide role for CHD remodelling factors and Nap1 in nucleosome disassembly.", *EMBO J.* **2007**, *26*, 2868–2879.

[72] D. Y. Lee, J. J. Hayes, D. Pruss, A. P. Wolffe, "A positive role for histone acetylation in transcription factor access to nucleosomal DNA", *Cell* **1993**.

[73] A. Kuzmichev, T. Jenuwein, P. Tempst, D. Reinberg, "Different EZH2-containing complexes target methylation of histone H1 or nucleosomal histone H3", *Mol. Cell* **2004**.

[74] Y. Shiio, R. N. Eisenman, "Histone sumoylation is associated with transcriptional repression.", *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13225–13230.

[75] T. Jenuwein, C. D. Allis, "Translating the histone code", *Science* **2001**, *293*, 1074–1080.

[76] S. B. Rothbart, B. D. Strahl, "Interpreting the language of histone and DNA modifications", *Biochim. Biophys. Acta Gene Regul. Mech.* **2014**, *1839*, 627–643.

[77] B. D. Strahl, C. D. Allis, "The language of covalent histone modifications.", *Nature* **2000**, *403*, 41–45.

[78] 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.

[79] S. Kriaucionis, N. Heintz, "The Nuclear DNA Base 5-Hydroxymethylcytosine Is Present in Purkinje Neurons and the Brain", *Science* **2009**, *324*, 929–930.

[80] T. Pfaffeneder, B. Hackner, M. Truß, M. Münzel, M. Müller, C. A. Deiml, C. Hagemeier, T. Carell, "The discovery of 5-formylcytosine in embryonic stem cell DNA.", *Angew. Chem. Int. Ed.* **2011**, *50*, 7008–7012.

[81] 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.

[82] 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.

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

[84] Y. Fu, G.-Z. Luo, K. Chen, X. Deng, M. Yu, D. Han, Z. Hao, J. Liu, X. Lu, L. C. Dore, X. Weng, Q. Ji, L. Mets, C. He, "N(6)-Methyldeoxyadenosine Marks Active Transcription Start Sites in Chlamydomonas", *Cell* **2015**, *161*, 879–892.

[85] G. Zhang, H. Huang, D. Liu, Y. Cheng, X. Liu, W. Zhang, R. Yin, D. Zhang, P. Zhang, J. Liu, C. Li, B. Liu, Y. Luo, Y. Zhu, N. Zhang, S. He, C. He, H. Wang, D. Chen, "N6-Methyladenine DNA Modification in Drosophila", *Cell* **2015**, *161*, 893–906.

[86] E. L. Greer, M. A. Blanco, L. Gu, E. Sendinc, J. Liu, D. Aristizábal-Corrales, C.-H. Hsu, L. Aravind, C. He, Y. Shi, "DNA Methylation on N6-Adenine in C. elegans", *Cell* **2015**, *161*, 868–878.

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

[88] 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.

[89] Y. Alaghband, T. W. Bredy, M. A. Wood, "The role of active DNA demethylation and Tet enzyme function in memory formation and cocaine action", *Neurosci. Lett.* **2016**, *625*, 40–46.

[90] 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.

[91] L. I. Kroeze, B. A. van der Reijden, J. H. Jansen, "5-Hydroxymethylcytosine: An epigenetic mark frequently deregulated in cancer", *Biochim. Biophys. Acta Rev. Cancer* **2015**, *1855*, 144–154.

[92] V. López, A. Fernández, M. Fraga, "The role of 5-hydroxymethylcytosine in development, aging and age-related diseases", *Ageing Res Rev.* **2017**, *37*, 28–38.

[93] A. Klungland, A. B. Robertson, "Oxidized C5-methyl cytosine bases in DNA: 5-Hydroxymethylcytosine; 5-formylcytosine; and 5-carboxycytosine", *Free Radic. Biol. Med.* **2017**, *107*, 62–68.

[94] A. A. Rawłuszko-Wieczorek, A. Siera, P. P. Jagodziński, "TET proteins in cancer: Current 'state of the art'", *Crit. Rev. Oncol. Hematol.* **2015**, *96*, 425–436.

[95] T. B. Johnson, R. D. Coghill, "Researches on Pyrimidines. C111. The Discovery of 5-methyl-cytosine in Tuberculinic acid, the nucleic acid of the Tubercle Bacillus 1", *J. Am. Chem. Soc.* **1925**, *47*, 2838–2844.

[96] G. R. Wyatt, "Occurrence of 5-Methyl-Cytosine in Nucleic Acids", *Nature* **1950**, *166*, 237–238.

[97] G. R. Wyatt, S. S. Cohen, "A New Pyrimidine Base from Bacteriophage Nucleic Acids", *Nature* **1952**, *170*, 1072–1073.

[98] E. Li, Y. Zhang, "DNA Methylation in Mammals", *Cold Spring Harbor Perspect. Biol.* **2014**, *6*, DOI `10.1101/cshperspect.a019133`.

[99] R. J. Klose, A. P. Bird, "Genomic DNA methylation: the mark and its mediators", *Trends Biochem. Sci.* **2006**, *31*, 89–97.

[100] M. G. Goll, T. H. Bestor, "Eukaryotic cytosine methyltransferases", *Annu. Rev. Biochem.* **2005**, *74*, 481–514.

[101] R. Z. Jurkowska, T. P. Jurkowski, A. Jeltsch, "Structure and function of mammalian DNA methyltransferases.", *Chembiochem* **2011**, *12*, 206–222.

[102] G. Vilkaitis, E. Merkiene, S. Serva, E. Weinhold, S. Klimasauskas, "The Mechanism of DNA Cytosine-5 Methylation", *J. Biol. Chem.* **2001**, *276*, 20924–20934.

[103] Q. Du, Z. Wang, V. L. Schramm, "Human DNMT1 transition state structure", *Proc. Natl. Acad. Sci. U.S.A.* **2016**, *113*, 2916–2921.

[104] 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.

[105] P. A. Jones, "Functions of DNA methylation: islands, start sites, gene bodies and beyond", *Nat. Rev. Genet.* **2012**, *13*, 484–492.

[106] C. Coulondre, J. H. Miller, P. J. Farabaugh, W. Gilbert, "Molecula Basis of Base Substitution Hotspots in Escherichia Coli", *Nature* **1978**, *274*, 775–780.

[107] A. P. Bird, "DNA Methylation and the Frequency of CpG in Animal DNA", *Nucleic Acids Res.* **1980**, *8*, 1499–1504.

[108] S. Saxonov, P. Berg, D. L. Brutlag, "A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters.", *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 1412–1417.

[109] A. M. Deaton, A. Bird, "CpG islands and the regulation of transcription", *Genes Dev.* **2011**, *25*, 1010–1022.

[110] A. D. Riggs, "X inactivation differentiation and DNA methylation.", *Cytogenet. Cell Genet.* **1975**, *14*, 9–25.

[111] R. Holliday, J. E. Pugh, "Dna Modification Mechanisms and Gene Activity During Development", *Science* **1975**, *187*, 226–232.

[112] Z. D. Smith, A. Meissner, "DNA methylation: roles inmammalian development", *Nat. Rev. Genet.* **2013**, *14*, 204–220.

[113] T. H. Bestor, J. R. Edwards, M. Boulard, "Notes on the role of dynamic DNA methylation in mammalian development", *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112*, 6796–6799.

[114] 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 baseresolution show widespread epigenomicdifferences", *Nature* **2010**, *462*, 315–322.

[115] K. Flores, F. Wolschin, J. J. Corneveaux, A. N. Allen, M. J. Huentelman, G. V. Amdam, "Genome-wide association between DNA methylation and alternative splicing in an invertebrate.", *BMC Genomics* **2012**, *13*, 480.

[116] A. K. Maunakea, I. Chepelev, K. Cui, K. Zhao, "Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition", *Cell Res.* **2013**, *23*, 1256–1269.

[117] A. H. Moarefi, F. ChEdin, "ICF Syndrome Mutations Cause a Broad Spectrum of Biochemical Defects in DNMT3B-Mediated De Novo DNA Methylation", *J. Mol. Biol.* **2011**, *409*, 758–772.

[118] X. S. Nan, H. H. Ng, C. A. Johnson, C. D. Laherty, B. M. Turner, R. N. Eisenman, A. Bird, "Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex", *Nature* **1998**, *393*, 386–389.

[119] H. H. Ng, Y. Zhang, B. Hendrich, C. A. Johnson, B. M. Turner, H. Erdjument-Bromage, P. Tempst, D. Reinberg, A. Bird, "MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex", *Nat. Genet.* **1999**, *23*, 58–61.

[120] J. P. Thomson, P. J. Skene, J. Selfridge, T. Clouaire, J. Guy, S. Webb, A. R. W. Kerr, A. Deaton, R. Andrews, K. D. James, D. J. Turner, R. Illingworth, A. Bird, "CpG islands influence chromatin structure via the CpG-binding protein Cfp1.", *Nature* **2010**, *464*, 1082–1086.

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

[122] H. Hui Ng, A. Bird, "Histone deacetylases: silencers for hire", *Trends Biochem. Sci.* **2000**, *25*, 121–126.

[123] J. C. Rice, C. D. Allis, "Histone methylation versus histone acetylation: new insights into epigenetic regulation", *Curr. Opin. Cell Biol.* **2001**, *13*, 263–273.

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

[125] J. Wang, S. Hevi, J. K. Kurash, H. Lei, F. Gay, J. Bajko, H. Su, W. Sun, H. Chang, G. Xu, F. Gaudet, E. Li, T. Chen, "The lysine demethylase LSD1 (KDM1) is required for maintenance of global DNA methylation.", *Nat. Genet.* **2009**, *41*, 125–129.

[126] Q. Zhao, G. Rank, Y. T. Tan, H. Li, R. L. Moritz, R. J. Simpson, L. Cerruti, D. J. Curtis, D. J. Patel, C. D. Allis, J. M. Cunningham, S. M. Jane, "PRMT5-mediated methylation of histone H4R3 recruits DNMT3A, coupling histone and DNA methylation in gene silencing", *Nat. Struct. Mol. Biol.* **2009**, *16*, 304–311.

[127] J. Schug, W.-P. Schuller, C. Kappen, J. M. Salbaum, M. Bucan, C. J. Stoeckert, "Promoter features related to tissue specificity as measured by Shannon entropy", *Genome Biol.* **2005**, *6*, R33.

[128] K. Shiota, "DNA methylation profiles of CpG islands for cellular differentiation and development in mammals.", *Cytogenet. Genome Res.* **2004**, *105*, 325–334.

[129] W. Reik, "Epigenetic Reprogramming in Mammalian Development", *Science* **2001**, *293*, 1089–1093.

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

[131] E. LI, T. H. Bestor, R. Jaenisch, "Targeted Mutation of the Dna Methyltransferase Gene Results in Embryonic Lethality", *Cell* **1992**, *69*, 915–926.

[132] 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.

[133] T. L. Dunwell, G. P. Pfeifer, "Drosophila genomic methylation: new evidence and new questions", *Epigenomics* **2014**, *6*, 459–461.

[134] 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.

[135] W. Mayer, A. Niveleau, J. Walter, R. Fundele, T. Haaf, "Embryogenesis: Demethylation of the zygotic paternal genome", *Nature* **2000**, *403*, 501–502.

[136] N. Beaujean, J. E. Taylor, M. McGarry, J. O. Gardner, I. Wilmut, P. Loi, G. Ptak, C. Galli, G. Lazzari, A. Bird, L. E. Young, R. R. Meehan, "The effect of interspecific oocytes on demethylation of sperm DNA", *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 7636–7640.

[137] Z. D. Smith, M. M. Chan, T. S. Mikkelsen, H. Gu, A. Gnirke, A. Regev, A. Meissner, "A unique regulatory phase of DNA methylation in the early mammalian embryo", *Nature* **2012**, *484*, 339–344.

[138] X. S. Ma, X. G. Wang, L. Qin, C. L. Song, F. Lin, J. M. Song, C. C. Zhu, H. L. Liu, "De novo DNA methylation of the paternal genome in 2-cell mouse embryos", *Genet. Mol. Res.* **2014**, *13*, 8632–8639.

[139] J. B. Gurdon, "The developmental capacity of nuclei taken from intestinal epithelium cells of feeding tadpoles.", *J. Embryol. Exp. Morphol.* **1962**, *10*, 622–640.

[140] K. Takahashi, S. Yamanaka, "Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors.", *Cell* **2006**, *126*, 663–676.

[141] K. Kim, A. Doi, B. Wen, K. Ng, R. Zhao, P. Cahan, J. Kim, M. J. Aryee, H. Ji, L. I. R. Ehrlich, A. Yabuuchi, A. Takeuchi, K. C. Cunniff, H. Hongguang, S. Mckinney-Freeman, O. Naveiras, T. J. Yoon, R. A. Irizarry, N. Jung, J. Seita, J. Hanna, P. Murakami, R. Jaenisch, R. Weissleder, S. H. Orkin, I. L. Weissman, A. P. Feinberg, G. Q. Daley, "Epigenetic memory in induced pluripotent stem cells", *Nature* **2010**, *467*, 285–290.

[142] R. K. Ng, J. B. Gurdon, "Epigenetic memory of active gene transcription is inherited through somatic cell nuclear transfer.", *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 1957–1962.

[143] J. B. Gurdon, D. A. Melton, "Nuclear reprogramming in cells.", *Science* **2008**, *322*, 1811–1815.

[144] B. Walia, N. Satija, R. P. Tripathi, G. U. Gangenahalli, "Induced Pluripotent Stem Cells: Fundamentals and Applications of the Reprogramming Process and its Ramifications on Regenerative Medicine", *Stem Cell Rev.* **2012**, *8*, 100–115.

[145] A. Seifinejad, M. Tabebordbar, H. Baharvand, L. A. Boyer, G. H. Salekdeh, "Progress and Promise Towards Safe Induced Pluripotent Stem Cells for Therapy", *Stem Cell Rev.* **2010**, *6*, 297–306.

[146] M. K. Carpenter, J. Frey-Vasconcells, M. S. Rao, "Developing safe therapies from human pluripotent stem cells.", *Nat. Biotechnol.* **2009**, *27*, 606–613.

[147] Y. Jung, G. Bauer, J. A. Nolta, "Concise Review: Induced Pluripotent Stem Cell-Derived Mesenchymal Stem Cells: Progress Toward Safe Clinical Products", *Stem Cells* **2012**, *30*, 42–47.

[148] J. P. Scott-Browne, C.-W. J. Lio, A. Rao, "TET proteins in natural and induced differentiation", *Curr. Opin. Genet. Dev.* **2017**, *46*, 202–208.

[149] K. D. Rasmussen, K. Helin, "Role of TET enzymes in DNA methylation, development, and cancer", *Genes Dev.* **2016**, *30*, 733–750.

[150] C. A. Doege, K. Inoue, T. Yamashita, D. B. Rhee, S. Travis, R. Fujita, P. Guarnieri, G. Bhagat, W. B. Vanti, A. Shih, R. L. Levine, S. Nik, E. I. Chen, A. Abeliovich, "Early-stage epigenetic modification during somatic cell reprogramming by Parp1 and Tet2", *Nature* **2012**, *488*, 652–655.

[151] G. R. Wyatt, S. Cohen, "The bases of the nucleic acids of some bacterial and animal viruses: the occurrence of 5-hydroxymethylcytosine", *Biochem. J.* **1953**, *55*, 774–782.

[152] J. R. Wagner, 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.

[153] S. Zuo, R. J. Boorstein, G. W. Teebor, "Oxidative damage to 5-methylcytosine in DNA", *Nucleic Acids Res.* **1995**, *23*, 3239–3243.

[154] V. Valinluck, H. H. Tsai, D. K. Rogstad, A. Burdzy, A. Bird, L. C. Sowers, "Oxidative damage to methyl-CpG sequences inhibits the binding of the methyl-CpG binding domain (MBD) of methyl-CpG binding protein 2 (MeCP2)", *Nucleic Acids Res.* **2004**, *32*, 4100–4108.

[155] R. M. Kothari, V. Shankar, "5-Methylcytosine content in the vertebrate deoxyribonucleic acids: species specificity.", *J. Mol. Evol.* **1976**, *7*, 325–329.

[156] E. Privat, L. C. Sowers, "Photochemical Deamination and Demethylation of 5-Methylcytosine", *Chem. Res. Toxicol.* **1996**, *9*, 745–750.

[157] N. W. Penn, R. Suwalski, C. O'Riley, K. Bojanowski, R. Yura, "The presence of 5-hydroxymethylcytosine in animal deoxyribonucleic acid", *Biochem. J.* **1972**, *126*, 781–790.

[158] M. Münzel, D. Globisch, T. Carell, "5-Hydroxymethylcytosine, the Sixth Base of the Genome", *Angew. Chem. Int. Ed.* **2011**, *50*, 6460–6468.

[159] 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, "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.

[160] M. Münzel, D. Globisch, T. Brückl, M. Wagner, V. Welzmiller, S. Michalakis, M. Müller, M. Biel, T. Carell, "Quantification of the Sixth DNA Base Hydroxymethylcytosine in the Brain", *Angew. Chem. Int. Ed.* **2010**, *49*, 5375–5377.

[161] S.-G. Jin, X. Wu, A. X. Li, G. P. Pfeifer, "Genomic mapping of 5-hydroxymethylcytosine in the human brain", *Nucleic Acids Res.* **2011**, *39*, 5015–5024.

[162] S. Dzitoyeva, H. Chen, H. Manev, "Effect of aging on 5-hydroxymethylcytosine in brain mitochondria", *Neurobiol. Aging* **2012**, *33*, 2881–2891.

[163] H. Chen, S. Dzitoyeva, H. Manev, "Effect of aging on 5-hydroxymethylcytosine in the mouse hippocampus", *Restor. Neurol. Neurosci.* **2012**, *30*, 237–245.

[164] L. Chouliaras, D. L.A. van den Hove, G. Kenis, S. Keitel, P. R. Hof, J. van Os, H. W.M. Steinbusch, C. Schmitz, B. P.F. Rutten, "Age-Related Increase in Levels of 5-Hydroxymethylcytosine in Mouse Hippocampus is Prevented by Caloric Restriction", *Curr. Alzheimer Res.* **2012**, *9*, 536–544.

[165] T. F. Kraus, V. Guibourt, H. A. Kretzschmar, "5-Hydroxymethylcytosine, the "Sixth Base", during brain development and ageing", *J. Neural Transm.* **2015**, *122*, 1035–1043.

[166] R. Lister, E. A. Mukamel, J. R. Nery, M. Urich, C. A. Puddifoot, N. D. Johnson, J. Lucero, Y. Huang, A. J. Dwork, M. D. Schultz, M. Yu, J. Tonti-Filippini, H. Heyn, S. Hu, J. C. Wu, A. Rao, M. Esteller, C. He, F. G. Haghighi, T. J. Sejnowski, M. M. Behrens, J. R. Ecker, "Global Epigenomic Reconfiguration During Mammalian Brain Development", *Science* **2013**, *341*, 1237905.

[167] 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.

[168] 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.

[169] G. Ficz, M. R. Branco, S. Seisenberger, F. Santos, F. Krueger, T. A. Hore, C. J. Marques, S. Andrews, W. Reik, "Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation", *Nature* **2011**, *473*, 398–402.

[170] 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*, 1231–40.

[171] Z. Sun, N. Dai, J. G. Borgaro, A. Quimby, D. Sun, I. R. C. Jr., Y. Zheng, Z. Zhu, S. Guan, "A Sensitive Approach to Map Genome-wide 5-Hydroxymethylcytosine and 5-Formylcytosine at Single-Base Resolution", *Mol. Cell* **2015**, *57*, 750–761.

[172] M. J. Booth, M. R. Branco, G. Ficz, D. Oxley, F. Krueger, W. Reik, S. Balasubramanian, "Quantitative Sequencing of 5-Methylcytosine and 5-Hydroxymethylcytosine at Single-Base Resolution", *Science* **2012**, *336*, 934–937.

[173] D.-Q. Shi, I. Ali, J. Tang, W.-C. Yang, "New Insights into 5hmC DNA Modification: Generation, Distribution and Function", *Front. Genet.* **2017**, *8*, 100.

[174] D. Mooijman, S. S. Dey, J.-C. Boisset, N. Crosetto, A. van Oudenaarden, "Single-cell 5hmC sequencing reveals chromosome-wide cell-to-cell variability and enables lineage reconstruction", *Nat Biotech* **2016**, *34*, 852–856.

[175] S. Kriaucionis, M. Tahiliani, "Expanding the Epigenetic Landscape: Novel Modifications of Cytosine in Genomic DNA", *Cold Spring Harbor Perspect. Biol.* **2014**, *6*, a018630.

[176] L. Shen, Y. Zhang, "5-Hydroxymethylcytosine: generation, fate, and genomic distribution", *Curr. Opin. Cell Biol.* **2013**, *25*, 289–296.

[177] B. Li, M. Carey, J. L. Workman, "The role of chromatin during transcription", *Cell* **2007**, *128*, 707–719.

[178] A. I. Lamond, W. C. Earnshaw, "Structure and function in the nucleus", *Science* **1998**, *280*, 547–553.

[179] H. Wu, Y. Zhang, "Reversing DNA Methylation: Mechanisms, Genomics, and Biological Functions", *Cell* **2014**, *156*, 45–68.

[180] H. Stroud, S. Feng, S. M. Kinney, S. Pradhan, S. E. Jacobsen, "5-Hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells", *Genome Biol.* **2011**, *12*, R54.

[181] 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.

[182] G. A. Maston, S. K. Evans, M. R. Green, "Transcriptional regulatory elements in the human genome", *Annu. Rev. Genomics Hum. Genet.* **2006**, *7*, 29–59.

[183] H. Wu, A. C. D'Alessio, S. Ito, Z. Wang, K. Cui, K. Zhao, Y. E. Sun, Y. Zhang, "Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells", *Genes Dev.* **2011**, *25*, 679–684.

[184] Y. Xu, F. Wu, L. Tan, L. Kong, L. Xiong, J. Deng, A. Barbera, L. Zheng, H. Zhang, S. Huang, J. Min, T. Nicholson, T. Chen, G. Xu, Y. Shi, K. Zhang, Y. G. Shi, "Genome-wide Regulation of 5hmC, 5mC and Gene Expression by Tet1 Hydroxylase in Mouse Embryonic Stem Cells", *Mol. Cell* **2011**, *42*, 451–464.

[185] 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.

[186] B. Delatte, R. Deplus, F. Fuks, "Playing TETris with DNA modifications", *EMBO J.* **2014**, *33*, 1198–1211.

[187] M. A. Hahn, R. Qiu, X. Wu, A. X. Li, H. Zhang, J. Wang, J. Jui, S.-G. Jin, Y. Jiang, G. P. Pfeifer, Q. Lu, "Dynamics of 5-hydroxymethylcytosine and chromatin marks in Mammalian neurogenesis.", *Cell Rep* **2013**, *3*, 291–300.

[188] K. Williams, J. Christensen, M. T. Pedersen, J. V. Johansen, P. A. Cloos, J. Rappsilber, K. Helin, "TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity", *Nature* **2011**, *473*, 343–348.

[189] H. Wu, A. C. D'Alessio, S. Ito, K. Xia, Z. Wang, K. Cui, K. Zhao, Y. Eve Sun, Y. Zhang, "Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells", *Nature* **2011**, *473*, 389–393.

[190] M. Mellen, P. Ayata, S. Dewell, S. Kriaucionis, N. Heintz, "MeCP2 Binds to 5hmC Enriched within Active Genes and Accessible Chromatin in the Nervous System", *Cell* **2012**, *151*, 1417–1430.

[191] J. Robertson, A. B. Robertson, A. Klungland, "The presence of 5-hydroxymethylcytosine at the gene promoter and not in the gene body negatively regulates gene expression", *Biochem. Biophys. Res. Commun.* **2011**, *411*, 40–43.

[192] N. Kitsera, J. Allgayer, E. Parsa, N. Geier, M. Rossa, T. Carell, A. Khobta, "Functional impacts of 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxycytosine at a single hemi-modified CpG dinucleotide in a gene promoter", *Nucleic Acids Res.* **2017**, DOI 10.1093/nar/gkx718.

[193] C. You, D. Ji, X. Dai, Y. Wang, "Effects of Tet-mediated Oxidation Products of 5-Methylcytosine on DNA Transcription in vitro and in Mammalian Cells", *Sci. Rep.* **2014**, *4*, 7052.

[194] C. G. Spruijt, F. Gnerlich, A. H. Smits, T. Pfaffeneder, P. W. T. C. Jansen, C. Bauer, M. Münzel, 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, M. Vermeulen, "Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives.", *Cell* **2013**, *152*, 1146–1159.

[195] M. A. Hahn, P. E. Szabo, G. P. Pfeifer, "5-Hydroxymethylcytosine: A stable or transient DNA modification?", *Genomics* **2014**, *104*, 314–323.

[196] 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.

[197] 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*, 1–11.

[198] M. Santiago, C. Antunes, M. Guedes, N. Sousa, C. J. Marques, "TET enzymes and DNA hydroxymethylation in neural development and function - how critical are they?", *Genomics* **2014**, *104*, 334–340.

[199] N. P. Blackledge, J. P. Thomson, P. J. Skene, "CpG Island Chromatin Is Shaped by Recruitment of ZF-CxxC Proteins", *Cold Spring Harbor Perspect. Biol.* **2013**, *5*, a018648.

[200] C. Xu, C. Bian, R. Lam, A. Dong, J. Min, "The structural basis for selective binding of non-methylated CpG islands by the CFP1 CXXC domain", *Nat. Commun.* **2011**, *2*, 227.

[201] L. M. Johnson, M. Bostick, X. Zhang, E. Kraft, I. Henderson, J. Callis, S. E. Jacobsen, "The SRA methyl-cytosine-binding domain links DNA and histone methylation", *Curr. Biol.* **2007**, *17*, 379–384.

[202] 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.

[203] C. Frauer, T. Hoffmann, S. Bultmann, V. Casa, M. C. Cardoso, I. Antes, H. Leonhardt, "Recognition of 5-Hydroxymethylcytosine by the Uhrf1 SRA Domain", *PLoS One* **2011**, *6*, e21306.

[204] 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.

[205] T. Zhou, J. Xiong, M. Wang, N. Yang, J. Wong, B. Zhu, R.-M. Xu, "Structural Basis for Hydroxymethylcytosine Recognition by the SRA Domain of UHRF2", *Mol. Cell* **2014**, *54*, 879–886.

[206] M. M. Compton, J. M. Thomson, A. H. Icard, "The analysis of cThy28 expression in avian lymphocytes", *Apoptosis* **2001**, *6*, 299–314.

[207] X. Z. Jiang, H. Toyota, T. Yoshimoto, E. Takada, H. Asakura, J. Mizuguchi, "Anti-IgM-induced down-regulation of nuclear Thy28 protein expression in Ramos B lymphoma cells", *Apoptosis* **2003**, *8*, 509–519.

[208] L. A. Papale, A. Madrid, S. Li, R. S. Alisch, "Early-life stress links 5-hydroxymethylcytosine to anxiety-related behaviors", *Epigenetics* **2017**, *12*, 264–276.

[209] E. Dong, S. G. Dzitoyeva, F. Matrisciano, P. Tueting, D. R. Grayson, A. Guidotti, "Brain-Derived Neurotrophic Factor Epigenetic Modifications Associated with Schizophrenia-like Phenotype Induced by Prenatal Stress in Mice", *Biol. Psychiatry* **2015**, *77*, 589–596.

[210] M. Nishioka, M. Bundo, K. Kasai, K. Iwamoto, "DNA methylation in schizophrenia: progress and challenges of epigenetic studies", *Genome Med.* **2012**, *4*, 96.

[211] S. Al-Mahdawi, S. A. Virmouni, M. A. Pook, "The emerging role of 5-hydroxymethylcytosine in neurodegenerative diseases", *Front. Neurosci.* **2014**, *8*, 397.

[212] L. Chouliaras, D. Mastroeni, E. Delvaux, A. Grover, G. Kenis, P. R. Hof, H. W. Steinbusch, P. D. Coleman, B. P. Rutten, D. L. van den Hove, "Consistent decrease in global DNA methylation and hydroxymethylation in the hippocampus of Alzheimer's disease patients", *Neurobiol. Aging* **2013**, *34*, 2091–2099.

[213] F. Wang, Y. Yang, X. Lin, J.-Q. Wang, Y.-S. Wu, W. Xie, D. Wang, S. Zhu, Y.-Q. Liao, Q. Sun, Y.-G. Yang, H.-R. Luo, C. Guo, C. Han, T.-S. Tang, "Genome-wide loss of 5-hmC is a novel epigenetic feature of Huntington's disease", *Hum. Mol. Genet.* **2013**, *22*, 3641.

[214] A. B. Robertson, J. Robertson, M. Fusser, A. Klungland, "Endonuclease G preferentially cleaves 5-hydroxymethylcytosine-modified DNA creating a substrate for recombination", *Nucleic Acids Res.* **2014**, *42*, 13280–13293.

[215] G. Kafer, X. Li, T. Horii, I. Suetake, S. Tajima, I. Hatada, P. Carlton, "5-Hydroxymethylcytosine Marks Sites of DNA Damage and Promotes Genome Stability", *Cell Rep* **2016**, *14*, 1283–1292.

[216] E. Parsa, A. Schröder, T. Carell, "Modifizierte DNA-Basen erweitern das Verständnis der Genregulation", *Naturwiss. Rundschau.* **2015**, *68*, 500–505.

[217] 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.

[218] E.-A. Raiber, D. Beraldi, G. Ficz, H. E. Burgess, M. R. Branco, P. Murat, D. Oxley, M. J. Booth, W. Reik, S. Balasubramanian, "Genome-wide distribution of 5-formylcytosine in embryonic stem cells is associated with transcription and depends on thymine DNA glycosylase", *Genome Biol.* **2012**, *13*, 1–11.

[219] 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, "Genome-wide Profiling of 5-Formylcytosine Reveals Its Roles in Epigenetic Priming", *Cell* **2013**, *153*, 678–691.

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

[221] F. Neri, D. Incarnato, A. Krepelova, S. Rapelli, F. Anselmi, C. Parlato, C. Medana, F. Dal Bello, S. Oliviero, "Single-Base Resolution Analysis of 5-Formyl and 5-Carboxyl Cytosine Reveals Promoter DNA Methylation Dynamics", *Cell Rep* **2015**, *10*, 674–683.

[222] A. Maiti, A. C. Drohat, "Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine potential implications for active demethylation of CpG sites", *J. Biol. Chem.* **2011**, *286*, 35334–35338.

[223] M. Bulger, M. Groudine, "Functional and Mechanistic Diversity of Distal Transcription Enhancers", *Cell* **2011**, *144*, 327–339.

[224] C. Zhu, Y. Gao, H. Guo, B. Xia, J. Song, X. Wu, H. Zeng, K. Kee, F. Tang, C. Yi, "Single-Cell 5-Formylcytosine Landscapes of Mammalian Early Embryos and ESCs at Single-Base Resolution", *Cell Stem Cell* **2017**, *20*, 720–731.

[225] 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.

[226] L. Wang, Y. Zhou, L. Xu, R. Xiao, X. Lu, L. Chen, J. Chong, "Molecular basis for 5-carboxycytosine recognition by RNA polymerase II elongation complex", *Nature* **2015**.

[227] L. Wheldon, A. Abakir, Z. Ferjentsik, T. Dudnakova, S. Strohbuecker, D. Christie, N. Dai, S. Guan, J. Foster, I. Corrêa Jr, M. Loose, J. Dixon, V. Sottile, A. Johnson, A. Ruzov, "Transient Accumulation of 5-Carboxylcytosine Indicates Involvement of Active Demethylation in Lineage Specification of Neural Stem Cells", *Cell Rep* **2014**, *7*, 1353–1361.

[228] 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.

[229] Y. Xue, J. Wong, G. T. Moreno, M. K. Young, J. Côté, W. Wang, "NURD, a Novel Complex with Both ATP-Dependent Chromatin-Remodeling and Histone Deacetylase Activities", *Mol. Cell* **1998**, *2*, 851–861.

[230] Y. Zhang, H. H. Ng, H. Erdjument-Bromage, P. Tempst, A. Bird, D. Reinberg, "Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation", *Genes Dev.* **1999**, *13*, 1924–1935.

[231] A. S. Lalmansingh, S. Karmakar, Y. Jin, A. K. Nagaich, "Multiple modes of chromatin remodeling by Forkhead box proteins", *Biochim. Biophys. Acta Gene Regul. Mech.* **2012**, *1819*, 707–715.

[232] P. Carlsson, M. Mahlapuu, "Forkhead Transcription Factors: Key Players in Development and Metabolism", *Dev. Biol.* **2002**, *250*, 1–23.

[233] 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.* **2015**, *22*, 44–49.

[234] 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.

[235] J. Hu, X. Xing, X. Xu, F. Wu, P. Guo, S. Yan, Z. Xu, J. Xu, X. Weng, X. Zhou, "Selective chemical labelling of 5-formylcytosine in DNA by fluorescent dyes.", *Chemistry* **2013**, *19*, 5836–5840.

[236] 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.

[237] S. Ji, H. Shao, Q. Han, C. L. Seiler, N. Tretyakova, "Reversible DNA-protein cross-linking at epigenetic DNA marks", *Angew. Chem. Int. Ed.* **2017**, DOI 10.1002/anie.201708286.

[238] S.-G. Jin, Z.-M. Zhang, T. L. Dunwell, M. R. Harter, X. Wu, J. Johnson, Z. Li, J. Liu, P. E. Szabó, Q. Lu, G.-l. Xu, J. Song, G. P. Pfeifer, "Tet3 reads 5-carboxylcytosine through its CXXC domain and is a potential guardian against neurodegeneration", *Cell Rep* **2016**, *14*, 493–505.

[239] 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.

[240] R. B. Lorsbach, J. Moore, S. Mathew, S. C. Raimondi, S. T. Mukatira, J. R. Downing, "TET1, a member of a novel protein family, is fused to MLL in acute myeloid leukemia containing the t(10;11)(q22;q23)", *Leukemia* **2003**, *17*, 637–41.

[241] 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.

[242] Z. Yu, P.-A. Genest, B. ter Riet, K. Sweeney, C. DiPaolo, R. Kieft, E. Christodoulou, A. Perrakis, J. M. Simmons, R. P. Hausinger, H. G. van Luenen, D. J. Rigden, R. Sabatini, P. Borst, "The protein that binds to DNA base J in trypanosomatids has features of a thymidine hydroxylase", *Nucleic Acids Res.* **2007**, *35*, 2107.

[243] P. Borst, R. Sabatini, "Base J: Discovery, Biosynthesis, and Possible Functions", *Annu. Rev. Microbiol.* **2008**, *62*, 235–251.

[244] L. J. Cliffe, R. Kieft, T. Southern, S. R. Birkeland, M. Marshall, K. Sweeney, R. Sabatini, "JBP1 and JBP2 are two distinct thymidine hydroxylases involved in J biosynthesis in genomic DNA of African trypanosomes", *Nucleic Acids Res.* **2009**, *37*, 1452–1462.

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

[246] K. P. Koh, A. Yabuuchi, S. Rao, Y. Huang, K. Cunniff, J. Nardone, A. Laiho, M. Tahiliani, C. A. Sommer, G. Mostoslavsky, R. Lahesmaa, S. H. Orkin, S. J. Rodig, G. Q. Daley, A. Rao, "Tet1 and Tet2 Regulate 5-Hydroxymethylcytosine Production and Cell Lineage Specification in Mouse Embryonic Stem Cells", *Cell Stem Cell* **2011**, *8*, 200–213.

[247] Y. Wang, Y. Zhang, "Regulation of TET protein stability by calpains", *Cell Rep* **2014**, *6*, 278–284.

[248] F. P. Guengerich, "Introduction: Metals in Biology: α-Ketoglutarate/Iron-Dependent Dioxygenases", *J. Biol. Chem.* **2015**, *290*, 20700–20701.

[249] H. Hashimoto, J. E. Pais, X. Zhang, L. Saleh, Z.-Q. Fu, N. Dai, I. R. Corrêa, Y. Zheng, X. Cheng, "Structure of a Naegleria Tet-like dioxygenase in complex with 5-Methylcytosine DNA", *Nature* **2014**, *506*, 391–395.

[250] H. Hashimoto, J. E. Pais, N. Dai, J. Correa, I. R., X. Zhang, Y. Zheng, X. Cheng, "Structure of Naegleria Tet-like dioxygenase (NgTet1) in complexes with a reaction intermediate 5-hydroxymethylcytosine DNA", *Nucleic Acids Res.* **2015**, *43*, 10713–10721.

[251] 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*, 1545–55.

[252] H. Hashimoto, X. Zhang, P. M. Vertino, X. Cheng, "The Mechanisms of Generation, Recognition, and Erasure of DNA 5-Methylcytosine and Thymine Oxidations", *J. Biol. Chem.* **2015**, *290*, 20723–33.

[253] V. K. Ponnaluri, J. P. Maciejewski, M. Mukherji, "A mechanistic overview of TET-mediated 5-Methylcytosine oxidation", *Biochem. Biophys. Res. Commun.* **2013**, *436*, 115–20.

[254] E. Tamanaha, S. Guan, K. Marks, L. Saleh, "Distributive Processing by the Iron (II)/alpha-Ketoglutarate-Dependent Catalytic Domains of the TET Enzymes Is Consistent with Epigenetic Roles for Oxidized 5-Methylcytosine Bases", *J. Am. Chem. Soc.* **2016**, *138*, 9345–9348.

[255] D. J. Crawford, M. Y. Liu, C. S. Nabel, X.-J. Cao, B. A. Garcia, R. M. Kohli, "Tet2 Catalyzes Stepwise 5-Methylcytosine Oxidation by an Iterative and de novo Mechanism", *J. Am. Chem. Soc.* **2016**, *138*, 730–733.

[256] H. Yang, H. Lin, H. Xu, L. Zhang, L. Cheng, B. Wen, J. Shou, K. Guan, Y. Xiong, D. Ye, "TET-catalyzed 5-methylcytosine hydroxylation is dynamically regulated by metabolites", *Cell Res.* **2014**, *24*, 1017–1020.

[257] X. Wu, Y. Zhang, "TET-mediated active DNA demethylation: mechanism, function and beyond", *Nat. Rev. Genet.* **2017**, *18*, 517–534.

[258] B. Xia, D. Han, X. Lu, Z. Sun, A. Zhou, Q. Yin, H. Zeng, M. Liu, X. Jiang, W. Xie, C. He, C. Yi, "Bisulfite-free, base-resolution analysis of 5-formylcytosine at the genome scale", *Nat. Meth.* **2015**, *12*, 1047–1050.

[259] L. Hu, J. Lu, J. Cheng, Q. Rao, Z. Li, H. Hou, Z. Lou, L. Zhang, W. Li, W. Gong, M. Liu, C. Sun, X. Yin, J. Li, X. Tan, P. Wang, Y. Wang, D. Fang, Q. Cui, P. Yang, C. He, H. Jiang, C. Luo, Y. Xu, "Structural insight into substrate preference for TET-mediated oxidation", *Nature* **2015**, *527*, 118–122.

[260] A. R. Weber, C. Krawczyk, A. B. Robertson, A. Kuśnierczyk, C. B. Vågbø, 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.

[261] M. Y. Liu, H. Torabifard, D. J. Crawford, J. E. DeNizio, X.-J. Cao, B. A. Garcia, G. A. Cisneros, R. M. Kohli, "Mutations along a TET2 active site scaffold stall oxidation at 5-hydroxymethylcytosine", *Nat. Chem. Biol.* **2017**, *13*, 181–187.

[262] L. Fu, C. R. Guerrero, N. Zhong, N. J. Amato, Y. Liu, S. Liu, Q. Cai, D. Ji, S.-G. Jin, L. J. Niedernhofer, G. P. Pfeifer, G.-L. Xu, Y. Wang, "Tet-mediated formation of 5-hydroxymethylcytosine in RNA.", *J. Am. Chem. Soc.* **2014**, *136*, 11582–11585.

[263] A. K. Upadhyay, J. R. Horton, X. Zhang, X. Cheng, "Coordinated methyl-lysine erasure: structural and functional linkage of a Jumonji demethylase domain and a reader domain.", *Curr. Opin. Struct. Biol.* **2011**, *21*, 750–760.

[264] H. Zhang, X. Zhang, E. Clark, M. Mulcahey, S. Huang, Y. G. Shi, "TET1 is a DNA-binding protein that modulates DNA methylation and gene transcription via hydroxylation of 5-methylcytosine", *Cell Res.* **2010**, *20*, 1390–1393.

[265] 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.

[266] W. Zhang, W. Xia, Q. Wang, A. J. Towers, J. Chen, R. Gao, Y. Zhang, C.-a. Yen, A. Y. Lee, Y. Li, C. Zhou, K. Liu, J. Zhang, T.-P. Gu, X. Chen, Z. Chang, D. Leung, S. Gao, Y.-h. Jiang, W. Xie, "Isoform Switch of TET1 Regulates DNA Demethylation and Mouse Development", *Mol. Cell* **2016**, *64*, 1062–1073.

[267] 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.

[268] X. Lu, B. S. Zhao, C. He, "TET family proteins: oxidation activity, interacting molecules, and functions in diseases", *Chem. Rev.* **2015**, *115*, 2225–2239.

[269] H. Wu, Y. Zhang, "Tet1 and 5-hydroxymethylation: a genome-wide view in mouse embryonic stem cells", *Cell Cycle* **2011**, *10*, 2428–2436.

[270] Q. Chen, Y. Chen, C. Bian, R. Fujiki, X. Yu, "TET2 promotes histone O-GlcNAcylation during gene transcription.", *Nature* **2013**, *493*, 561–564.

[271] F. T. Shi, H. Kim, W. Lu, Q. He, D. Liu, M. A. Goodell, M. Wan, Z. Songyang, "Ten-eleven translocation 1 (Tet1) is regulated by O-linked N-acetylglucosamine transferase (Ogt) for target gene repression in mouse embryonic stem cells", *J. Biol. Chem.* **2013**, *288*, 20776–20784.

[272] O. Yildirim, R. Li, J.-H. Hung, P. Chen, X. Dong, L.-S. Ee, Z. Weng, O. Rando, T. Fazzio, "Mbd3/NURD Complex Regulates Expression of 5-Hydroxymethylcytosine Marked Genes in Embryonic Stem Cells", *Cell* **2011**, *147*, 1498–1510.

[273] J. Ramirez, J. Hagman, "The Mi-2/NuRD complex: a critical epigenetic regulator of hematopoietic development, differentiation and cancer", *Epigenetics* **2009**, *4*, 532–536.

[274] M. P. Torchy, A. Hamiche, B. P. Klaholz, "Structure and function insights into the NuRD chromatin remodeling complex", *Cell. Mol. Life Sci.* **2015**, *72*, 2491–2507.

[275] B. A. Lewis, J. A. Hanover, "O-GlcNAc and the epigenetic regulation of gene expression", *J. Biol. Chem.* **2014**, *289*, 34440–34448.

[276] G. W. Hart, C. Slawson, G. Ramirez-Correa, O. Lagerlof, "Cross talk between O-GlcNAcylation and phosphorylation: roles in signaling, transcription, and chronic disease", *Annu. Rev. Biochem.* **2011**, *80*, 825–858.

[277] R. Fujiki, W. Hashiba, H. Sekine, A. Yokoyama, T. Chikanishi, S. Ito, Y. Imai, J. Kim, H. H. He, K. Igarashi, J. Kanno, F. Ohtake, H. Kitagawa, R. G. Roeder, M. Brown, S. Kato, "GlcNAcylation of histone H2B facilitates its monoubiquitination", *Nature* **2011**, *480*, 557–560.

[278] F. Lu, Y. Liu, L. Jiang, S. Yamaguchi, Y. Zhang, "Role of Tet proteins in enhancer activity and telomere elongation", *Genes Dev.* **2014**, DOI 10.1101/gad.248005.114.

[279] J. Yang, R. Guo, H. Wang, X. Ye, Z. Zhou, J. Dan, H. Wang, P. Gong, W. Deng, Y. Yin, S. Mao, L. Wang, J. Ding, J. Li, D. L. Keefe, M. M. Dawlaty, J. Wang, G. Xu, L. Liu, "Tet Enzymes Regulate Telomere Maintenance and Chromosomal Stability of Mouse ESCs", *Cell Rep* **2016**, *15*, 1809–1821.

[280] J.-A. Losman, W. G. Kaelin, "What a difference a hydroxyl makes: mutant IDH, (R)-2-hydroxyglutarate, and cancer", *Genes Dev.* **2013**, *27*, 836–852.

[281] W. Xu, H. Yang, Y. Liu, Y. Yang, P. Wang, S.-H. Kim, S. Ito, C. Yang, P. Wang, M.-T. Xiao, L.-x. Liu, W.-q. Jiang, J. Liu, J.-y. Zhang, B. Wang, S. Frye, Y. Zhang, Y.-h. Xu, Q.-y. Lei, K.-L. Guan, S.-m. Zhao, Y. Xiong, "Oncometabolite 2-Hydroxyglutarate Is a Competitive Inhibitor of α-Ketoglutarate-Dependent Dioxygenases", *Cancer Cell* **2011**, *19*, 17–30.

[282] C. G. Lian, Y. Xu, C. Ceol, F. Wu, A. Larson, K. Dresser, W. Xu, L. Tan, Y. Hu, Q. Zhan, C. W. Lee, D. Hu, B. Q. Lian, S. Kleffel, Y. Yang, J. Neiswender, A. J. Khorasani, R. Fang, C. Lezcano, L. M. Duncan, R. A. Scolyer, J. F. Thompson, H. Kakavand, Y. Houvras, L. I. Zon, J. Mihm, M. C., U. B. Kaiser, T. Schatton, B. A. Woda, G. F. Murphy, Y. G. Shi, "Loss of 5-hydroxymethylcytosine is an epigenetic hallmark of melanoma", *Cell* **2012**, *150*, 1135–1146.

[283] M. E. Figueroa, O. Abdel-Wahab, C. Lu, P. S. Ward, J. Patel, A. Shih, Y. Li, N. Bhagwat, A. Vasanthakumar, H. F. Fernandez, M. S. Tallman, Z. Sun, K. Wolniak, J. K. Peeters, W. Liu, S. E. Choe, V. R. Fantin, E. Paietta, B. Lowenberg, J. D. Licht, L. A. Godley, R. Delwel, P. J. Valk, C. B. Thompson, R. L. Levine, A. Melnick, "Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation", *Cancer Cell* **2010**, *18*, 553–567.

[284] M. Xiao, H. Yang, W. Xu, S. Ma, H. Lin, H. Zhu, L. Liu, Y. Liu, C. Yang, Y. Xu, S. Zhao, D. Ye, Y. Xiong, K.-L. Guan, "Inhibition of α-KG-dependent histone and DNA demethylases by fumarate and succinate that are accumulated in mutations of FH and SDH tumor suppressors", *Genes Dev.* **2012**, *26*, 1326–1338.

[285] T. Laukka, C. J. Mariani, T. Ihantola, J. Z. Cao, J. Hokkanen, J. William G Kaelin, L. A. Godley, P. Koivunen, "Fumarate and Succinate Regulate Expression of Hypoxia-Inducible Genes via TET Enzymes", *J. Biol. Chem.* **2015**, *291*, 4256–4265.

[286] E. A. Minor, B. L. Court, J. I. Young, G. Wang, "Ascorbate induces ten-eleven translocation (Tet) methylcytosine dioxygenase-mediated generation of 5-hydroxymethylcytosine.", *J. Biol. Chem.* **2013**, *288*, 13669–13674.

[287] K. Blaschke, K. T. Ebata, M. M. Karimi, J. A. Zepeda-Martínez, P. Goyal, S. Mahapatra, A. Tam, D. J. Laird, M. Hirst, A. Rao, M. C. Lorincz, M. Ramalho-Santos, "Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells.", *Nature* **2013**, *500*, 222–226.

[288] R. Yin, S.-Q. Mao, B. Zhao, Z. Chong, Y. Yang, C. Zhao, D. Zhang, H. Huang, J. Gao, Z. Li, Y. Jiao, C. Li, S. Liu, D. Wu, W. Gu, Y.-G. Yang, G.-L. Xu, H. Wang, "Ascorbic Acid Enhances Tet-Mediated 5-Methylcytosine Oxidation and Promotes DNA Demethylation in Mammals", *J. Am. Chem. Soc.* **2013**, *135*, 10396–10403.

[289] Y.-P. Tsai, H.-F. Chen, S.-Y. Chen, W.-C. Cheng, H.-W. Wang, Z.-J. Shen, C. Song, S.-C. Teng, C. He, K.-J. Wu, "TET1 regulates hypoxia-induced epithelial-mesenchymal transition by acting as a co-activator", *Genome Biol.* **2014**, *15*, 513.

[290] C. J. Mariani, A. Vasanthakumar, J. Madzo, A. Yesilkanal, T. Bhagat, Y. Yu, S. Bhattacharyya, R. H. Wenger, S. L. Cohn, J. Nanduri, A. Verma, N. R. Prabhakar, L. A. Godley, "TET1-Mediated Hydroxymethylation Facilitates Hypoxic Gene Induction in Neuroblastoma", *Cell Rep* **2014**, *7*, 1343–1352.

[291] 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.

[292] B. Zhao, Y. Yang, X. Wang, Z. Chong, R. Yin, S.-H. Song, C. Zhao, C. Li, H. Huang, B.-F. Sun, D. Wu, K.-X. Jin, M. Song, B.-Z. Zhu, G. Jiang, J. M. Rendtlew Danielsen, G.-L. Xu, Y.-G. Yang, H. Wang, "Redox-active quinones induces genome-wide DNA methylation changes by an iron-mediated and Tet-dependent mechanism", *Nucleic Acids Res.* **2014**, *42*, 1593–1605.

[293] X. Lv, H. Jiang, Y. Liu, X. Lei, J. Jiao, "MicroRNA-15b promotes neurogenesis and inhibits neural progenitor proliferation by directly repressing TET3 during early neocortical development", *EMBO Rep.* **2014**, *15*, 1305–1314.

[294] S. Song, K. Ito, U. Ala, L. Kats, K. Webster, S. Sun, M. Jongen-Lavrencic, K. Manova-Todorova, J. Teruya-Feldstein, D. Avigan, R. Delwel, P. Pandolfi, "The Oncogenic MicroRNA miR-22 Targets the TET2 Tumor Suppressor to Promote Hematopoietic Stem Cell Self-Renewal and Transformation", *Cell Stem Cell* **2013**, *13*, 87–101.

[295] S. Song, L. Poliseno, M. Song, U. Ala, K. Webster, C. Ng, G. Beringer, N. Brikbak, X. Yuan, L. Cantley, A. Richardson, P. Pandolfi, "MicroRNA-Antagonism Regulates Breast Cancer Stemness and Metastasis via TET-Family-Dependent Chromatin Remodeling", *Cell* **2013**, *154*, 311–324.

[296] X. Fu, L. Jin, X. Wang, A. Luo, J. Hu, X. Zheng, W. M. Tsark, A. D. Riggs, H. T. Ku, W. Huang, "MicroRNA-26a targets ten eleven translocation enzymes and is regulated during pancreatic cell differentiation", *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 17892–17897.

[297] J. Cheng, S. Guo, S. Chen, S. Mastriano, C. Liu, A. D'Alessio, E. Hysolli, Y. Guo, H. Yao, C. Megyola, D. Li, J. Liu, W. Pan, C. Roden, X.-L. Zhou, K. Heydari, J. Chen, I.-H. Park, Y. Ding, Y. Zhang, J. Lu, "An Extensive Network of TET2-Targeting MicroRNAs Regulates Malignant Hematopoiesis", *Cell Rep* **2013**, *5*, 471–481.

[298] M. Welling, H.-H. Chen, J. Muñoz, M. U. Musheev, L. Kester, J. P. Junker, N. Mischerikow, M. Arbab, E. Kuijk, L. Silberstein, P. V. Kharchenko, M. Geens, C. Niehrs, H. van de Velde, A. van Oudenaarden, A. J. Heck, N. Geijsen, "DAZL regulates Tet1 translation in murine embryonic stem cells", *EMBO Rep.* **2015**, *16*, 791–802.

[299] T. Nakagawa, L. Lv, M. Nakagawa, Y. Yu, C. Yu, A. C. D'Alessio, K. Nakayama, H. Y. Fan, X. Chen, Y. Xiong, "CRL4(VprBP) E3 ligase promotes monoubiquitylation and chromatin binding of TET dioxygenases", *Mol. Cell* **2015**, *57*, 247–260.

[300] F. Ciccarone, E. Valentini, M. Zampieri, P. Caiafa, "5mC-hydroxylase activity is influenced by the PARylation of TET1 enzyme", *Oncotarget* **2015**, *6*, 24333–24347.

[301] Q. Zhang, X. Liu, W. Gao, P. Li, J. Hou, J. Li, J. Wong, "Differential Regulation of Ten-Eleven Translocation Family of Dioxygenases by O-Linked β-N-Acetylglucosamine Transferase OGT.", *J. Biol. Chem.* **2014**, *289*, 5986–5996.

[302] C. Bauer, K. Gobel, N. Nagaraj, C. Colantuoni, M. Wang, U. Müller, E. Kremmer, A. Rottach, H. Leonhardt, "Phosphorylation of TET proteins is regulated via O-GlcNAcylation by the glycosyltransferase OGT", *J. Biol. Chem.* **2015**, *290*, 4801–4812.

[303] Y. W. Zhang, Z. Wang, W. Xie, Y. Cai, L. Xia, H. Easwaran, J. Luo, R. C. Yen, Y. Li, S. B. Baylin, "Acetylation Enhances TET2 Function in Protecting against Abnormal DNA Methylation during Oxidative Stress", *Mol. Cell* **2017**, *65*, 323–335.

[304] S. C. Wu, Y. Zhang, "Active DNA demethylation: many roads lead to Rome", *Nat. Rev. Mol. Cell Biol.* **2010**, *11*, 607–620.

[305] H. Lee, T. Hore, W. Reik, "Reprogramming the Methylome: Erasing Memory and Creating Diversity", *Cell Stem Cell* **2014**, *14*, 710–719.

[306] F. M. Piccolo, A. G. Fisher, "Getting rid of DNA methylation", *Trends Cell Biol.* **2014**, *24*, 136–143.

[307] I. Cantone, A. G. Fisher, "Epigenetic programming and reprogramming during development", *Nat. Struct. Mol. Biol* **2013**, *20*, 282–289.

[308] W. Sun, M. Guan, X. Li, "5-hydroxymethylcytosine-mediated DNA demethylation in stem cells and development.", *Stem Cells Dev.* **2014**, *23*, 923–930.

[309] S. Yamaguchi, K. Hong, R. Liu, A. Inoue, L. Shen, K. Zhang, Y. Zhang, "Dynamics of 5-methylcytosine and 5-hydroxymethylcytosine during germ cell reprogramming", *Cell Res.* **2013**, *23*, 329–339.

[310] I. B. Zovkic, M. C. Guzman-Karlsson, J. D. Sweatt, "Epigenetic regulation of memory formation and maintenance", *Learn. Mem.* **2013**, *20*, 61–74.

[311] L. Wen, F. Tang, "Genomic distribution and possible functions of DNA hydroxymethylation in the brain", *Genomics* **2014**, *104*, 341–346.

[312] W. Sun, L. Zang, Q. Shu, X. Li, "From development to diseases: the role of 5hmC in brain", *Genomics* **2014**, *104*, 347–51.

[313] T. Kato, K. Iwamoto, "Comprehensive DNA methylation and hydroxymethylation analysis in the human brain and its implication in mental disorders", *Neuropharmacology* **2014**, *80*, 133–139.

[314] S. Kagiwada, K. Kurimoto, T. Hirota, M. Yamaji, M. Saitou, "Replication-coupled passive DNA demethylation for the erasure of genome imprints in mice", *EMBO J.* **2013**, *32*, 340–53.

[315] R. M. Kohli, Y. Zhang, "TET enzymes, TDG and the dynamics of DNA demethylation", *Nature* **2013**, *502*, 472–479.

[316] A. Inoue, Y. Zhang, "Replication-Dependent Loss of 5-Hydroxymethylcytosine in Mouse Preimplantation Embryos", *Science* **2011**, *334*, 194–194.

[317] V. Valinluck, L. C. Sowers, "Endogenous Cytosine Damage Products Alter the Site Selectivity of Human DNA Maintenance Methyltransferase DNMT1", *Cancer Res.* **2007**, *67*, 946–950.

[318] 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.

[319] P. Giehr, C. Kyriakopoulos, G. Ficz, V. Wolf, J. Walter, "The Influence of Hydroxylation on Maintaining CpG Methylation Patterns: A Hidden Markov Model Approach", *PLoS Comput. Biol.* **2016**, *12*, 1–16.

[320] 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.

[321] S. Seisenberger, S. Andrews, F. Krueger, J. Arand, J. Walter, F. Santos, C. Popp, B. Thienpont, W. Dean, W. Reik, "The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells", *Mol. Cell* **2012**, *48*, 849–862.

[322] S. Guibert, T. Forne, M. Weber, "Global profiling of DNA methylation erasure in mouse primordial germ cells", *Genome Res.* **2012**, *22*, 633–641.

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

[324] K. Iqbal, S.-G. Jin, G. P. Pfeifer, P. E. Szabó, "Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine", *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 3642–3647.

[325] 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 Methylomes in Mammals", *Cell* **2014**, *157*, 979–991.

[326] 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.

[327] 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.

[328] T. Nakamura, Y. Arai, H. Umehara, M. Masuhara, T. Kimura, H. Taniguchi, T. Sekimoto, M. Ikawa, Y. Yoneda, M. Okabe, S. Tanaka, K. Shiota, T. Nakano, "PGC7/Stella protects against DNA demethylation in early embryogenesis", *Nat. Cell Biol.* **2007**, *9*, 64–71.

[329] T. Nakamura, Y. J. Liu, H. Nakashima, H. Umehara, K. Inoue, S. Matoba, M. Tachibana, A. Ogura, Y. Shinkai, T. Nakano, "PGC7 binds histone H3K9me2 to protect against conversion of 5mC to 5hmC in early embryos", *Nature* **2012**, *486*, 415–419.

[330] F. Santos, B. Hendrich, W. Reik, W. Dean, "Dynamic Reprogramming of DNA Methylation in the Early Mouse Embryo", *Dev. Biol.* **2002**, *241*, 172–182.

[331] H. Kobayashi, T. Sakurai, M. Imai, N. Takahashi, A. Fukuda, O. Yayoi, S. Sato, K. Nakabayashi, K. Hata, Y. Sotomaru, Y. Suzuki, T. Kono, "Contribution of Intragenic DNA Methylation in Mouse Gametic DNA Methylomes to Establish Oocyte-Specific Heritable Marks", *PLos Genet.* **2012**, *8*, e1002440.

[332] S. Gkountela, K. X. Zhang, T. A. Shafiq, W. W. Liao, J. Hargan-Calvopina, P. Y. Chen, A. T. Clark, "DNA Demethylation Dynamics in the Human Prenatal Germline", *Cell* **2015**, *161*, 1425–1436.

[333] R. Ohno, M. Nakayama, C. Naruse, N. Okashita, O. Takano, M. Tachibana, M. Asano, M. Saitou, Y. Seki, "A replication-dependent passive mechanism modulates DNA demethylation in mouse primordial germ cells", *Development* **2013**, *140*, 2892–2903.

[334] Z. D. Smith, A. Meissner, "The simplest explanation: passive DNA demethylation in PGCs", *EMBO J.* **2013**, *32*, 318–321.

[335] Y. Kawasaki, J. Lee, A. Matsuzawa, T. Kohda, T. Kaneko-Ishino, F. Ishino, "Active DNA demethylation is required for complete imprint erasure in primordial germ cells", *Sci. Rep.* **2014**, *4*, 3658.

[336] A. Bird, "Il2 transcription unleashed by active DNA demethylation", *Nat. Immunol.* **2003**, *4*, 208–209.

[337] D. Bruniquel, R. H. Schwartz, "Selective, stable demethylation of the interleukin-2 gene enhances transcription by an active process", *Nat. Immunol.* **2003**, *4*, 235–240.

[338] K. Martinowich, D. Hattori, H. Wu, S. Fouse, F. He, Y. Hu, G. Fan, Y. E. Sun, "DNA Methylation-Related Chromatin Remodeling in Activity-Dependent Bdnf Gene Regulation", *Science* **2003**, *302*, 890–893.

[339] S. Kangaspeska, B. Stride, R. Metivier, M. Polycarpou-Schwarz, D. Ibberson, R. P. Carmouche, V. Benes, F. Gannon, G. Reid, "Transient cyclical methylation of promoter DNA", *Nature* **2008**, *452*, 112–115.

[340] R. Metivier, R. Gallais, C. Tiffoche, C. Le Peron, R. Z. Jurkowska, R. P. Carmouche, D. Ibberson, P. Barath, F. Demay, G. Reid, V. Benes, A. Jeltsch, F. Gannon, G. Salbert, "Cyclical DNA methylation of a transcriptionally active promoter", *Nature* **2008**, *452*, 45–50.

[341] S. K. T. Ooi, T. H. Bestor, "The colorful history of active DNA demethylation", *Cell* **2008**, *133*, 1145–1148.

[342] D. P. Gavin, K. A. Chase, R. P. Sharma, "Active DNA demethylation in post-mitotic neurons: A reason for optimism", *Neuropharmacology* **2013**, *75*, 233–245.

[343] N. N. Karpova, "Role of BDNF epigenetics in activity-dependent neuronal plasticity", *Neuropharmacology* **2014**, *76*, 709–718.

[344] I. A. Qureshi, M. F. Mehler, "An evolving view of epigenetic complexity in the brain", *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **2014**, 369.

[345] P. Bekinschtein, M. Cammarota, L. M. Igaz, L. R. Bevilaqua, I. Izquierdo, J. H. Medina, "Persistence of Long-Term Memory Storage Requires a Late Protein Synthesis- and BDNF- Dependent Phase in the Hippocampus", *Neuron* **2007**, *53*, 261–277.

[346] J. Guo, Y. Su, C. Zhong, G.-l. Ming, H. Song, "Hydroxylation of 5-Methylcytosine by TET1 Promotes Active DNA Demethylation in the Adult Brain", *Cell* **2011**, *145*, 423–434.

[347] D. K. Ma, M.-H. Jang, J. U. Guo, Y. Kitabatake, M.-l. Chang, N. Pow-anpongkul, R. A. Flavell, B. Lu, G.-l. Ming, H. Song, "Neuronal Activity–Induced Gadd45b Promotes Epigenetic DNA Demethylation and Adult Neurogenesis", *Science* **2009**, *323*, 1074–1077.

[348] F. D. Lubin, T. L. Roth, J. D. Sweatt, "Epigenetic Regulation of bdnf Gene Transcription in the Consolidation of Fear Memory", *J. Neurosci.* **2008**, *28*, 10576–10586.

[349] C. A. Miller, J. D. Sweatt, "Covalent Modification of DNA Regulates Memory Formation", *Neuron* **2007**, *53*, 857–869.

[350] T. L. Roth, J. D. Sweatt, "Regulation of chromatin structure in memory formation", *Curr. Opin. Neurobiol.* **2009**, *19*, 336–342.

[351] J.-K. Zhu, "Active DNA demethylation mediated by DNA glycosylases.", *Annu. Rev. Genet.* **2009**, *43*, 143–166.

[352] J. Penterman, D. Zilberman, J. H. Huh, T. Ballinger, S. Henikoff, R. L. Fischer, "DNA demethylation in the Arabidopsis genome", *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 6752–6757.

[353] M. Frémont, M. Siegmann, S. Gaulis, R. Matthies, D. Hess, J. P. Jost, "Demethylation of DNA by purified chick embryo 5-methylcytosine-DNA glycosylase requires both protein and RNA", *Nucleic Acids Res.* **1997**, *25*, 2375–2380.

[354] J. P. Jost, S. Schwarz, D. Hess, H. Angliker, F. V. Fuller-Pace, H. Stahl, S. Thiry, M. Siegmann, "A chicken embryo protein related to the mammalian DEAD box protein p68 is tightly associated with the highly purified protein-RNA complex of 5-MeC-DNA glycosylase", *Nucleic Acids Res.* **1999**, *27*, 3245–3252.

[355] B. Zhu, Y. Zheng, D. Hess, H. Angliker, S. Schwarz, M. Siegmann, S. Thiry, J. P. Jost, "5-methylcytosine-DNA glycosylase activity is present in a cloned G/T mismatch DNA glycosylase associated with the chicken embryo DNA demethylation complex", *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 5135–5139.

[356] B. Zhu, Y. Zheng, H. Angliker, S. Schwarz, S. Thiry, M. Siegmann, J.-P. Jost, "5-Methylcytosine DNA glycosylase activity is also present in the human MBD4 (G/T mismatch glycosylase) and in a related avian sequence", *Nucleic Acids Res.* **2000**, *28*, 4157–4165.

[357] F. Santos, W. Dean, "Epigenetic reprogramming during early development in mammals", *Reproduction* **2004**, *127*, 643–651.

[358] 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.

[359] N. Navaratnam, R. Sarwar, "An overview of cytidine deaminases", *Int. J. Hematol.* **2006**, *83*, 195–200.

[360] 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.

[361] S. G. Conticello, "The AID/APOBEC family of nucleic acid mutators", *Genome Biol.* **2008**, *9*, 229.

[362] C. Prochnow, R. Bransteitter, X. S. Chen, "APOBEC deaminases-mutases with defensive roles for immunity", *Sci. China C Life Sci.* **2009**, *52*, 893–902.

[363] D. M. Franchini, S. K. Petersen-Mahrt, "AID and APOBEC deaminases: balancing DNA damage in epigenetics and immunity", *Epigenomics* **2014**, *6*, 427–443.

[364] H. D. Morgan, W. Dean, H. A. Coker, W. Reik, S. K. Petersen-Mahrt, "Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: implications for epigenetic reprogramming.", *J. Biol. Chem.* **2004**, *279*, 52353–52360.

[365] 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.

[366] C. Popp, W. Dean, S. Feng, S. J. Cokus, S. Andrews, M. Pellegrini, S. E. Jacobsen, W. Reik, "Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency", *Nature* **2010**, *463*, 1101–1105.

[367] 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.

[368] G. Rangam, K.-M. Schmitz, A. J. A. Cobb, S. K. Petersen-Mahrt, "AID Enzymatic Activity Is Inversely Proportional to the Size of Cytosine C5 Orbital Cloud", *PLoS One* **2012**, *7*, e43279.

[369] Z. M. Svedruzic, "Mammalian Cytosine DNA Methyltransferase Dnmt1: Enzymatic Mechanism, Novel Mechanism-Based Inhibitors, and RNA-directed DNA Methylation", *Curr. Med. Chem.* **2008**, *15*, 92–106.

[370] M. A. Grillo, S. Colombatto, "S-adenosylmethionine and protein methylation", *Amino Acids* **2005**, *28*, 357–62.

[371] M. R. Bauerle, E. L. Schwalm, S. J. Booker, "Mechanistic Diversity of Radical S-Adenosylmethionine (SAM)-dependent Methylation", *J. Biol. Chem.* **2015**, *290*, 3995–4002.

[372] E. K. Schutsky, C. S. Nabel, A. 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.

[373] D. Cortazar, C. Kunz, Y. Saito, R. Steinacher, P. Schär, "The enigmatic thymine DNA glycosylase", *DNA Repair* **2007**, *6*, 489–504.

[374] S. R. Dalton, A. Bellacosa, "DNA demethylation by TDG", *Epigenomics* **2012**, *4*, 459–467.

[375] D. S. Chen, M. J. Lucey, F. Phoenix, J. Lopez-Garcia, S. M. Hart, R. Losson, L. Buluwela, R. C. Coombes, P. Chambon, P. Schär, S. Ali, "T : G mismatch-specific thymine-DNA glycosylase potentiates transcription of estrogen-regulated genes through direct interaction with estrogen receptor alpha", *J. Biol. Chem.* **2003**, *278*, 38586–38592.

[376] S. Um, M. Harbers, A. Benecke, B. 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.

[377] 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.* **2007**, *35*, 390–400.

[378] M. Tini, A. Benecke, 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.

[379] S. Cortellino, J. Xu, M. Sannai, R. Moore, E. Caretti, A. Cigliano, M. Le Coz, K. Devarajan, A. Wessels, D. Soprano, L. K. Abramowitz, M. S. Bartolomei, F. Rambow, M. R. Bassi, T. Bruno, M. Fanciulli, C. Renner, A. J. Klein-Szanto, Y. Matsumoto, D. Kobi, I. Davidson, C. Alberti, L. Larue, A. Bellacosa, "Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair", *Cell* **2011**, *146*, 67–79.

[380] 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. Schär, "Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability", *Nature* **2011**, *470*, 419–423.

[381] L. Zhang, X. Lu, J. Lu, H. Liang, Q. Dai, G.-L. Xu, C. Luo, H. Jiang, C. He, "Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA.", *Nat. Chem. Biol.* **2012**, *8*, 328–330.

[382] X. Lu, D. Han, B. S. Zhao, C. X. Song, L. S. Zhang, L. C. Dore, C. He, "Base-resolution maps of 5-formylcytosine and 5-carboxylcytosine reveal genome-wide DNA demethylation dynamics", *Cell Res.* **2015**, *25*, 386–389.

[383] A. B. Sjolund, A. G. Senejani, J. B. Sweasy, "MBD4 and TDG: Multifaceted DNA glycosylases with ever expanding biological roles", *Mutat. Res. Fundam. Mol. Mech. Mutagen.* **2013**, *743*, 12–25.

[384] L. Zhang, X. Lu, J. Lu, H. Liang, Q. Dai, G.-L. Xu, C. Luo, H. Jiang, C. He, "Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA", *Nat. Chem. Biol.* **2012**, *8*, 328–330.

[385] H. E. Krokan, M. Bjoras, "Base excision repair", *Cold Spring Harbor Perspect. Biol.* **2013**, *5*, a012583.

[386] M. Wossidlo, J. Arand, V. Sebastiano, K. Lepikhov, M. Boiani, R. Reinhardt, H. Schöler, J. Walter, "Dynamic link of DNA demethylation, DNA strand breaks and repair in mouse zygotes", *EMBO J.* **2010**, *29*, 1877–1888.

[387] 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.

[388] 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*, 447–458.

[389] 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.

[390] S. Schiesser, T. Pfaffeneder, K. Sadeghian, B. Hackner, B. Steigenberger, A. S. Schröder, J. Steinbacher, G. Kashiwazaki, G. Höfner, K. T. Wanner, C. Ochsenfeld, T. Carell, "Deamination, Oxidation, and C–C Bond Cleavage Reactivity of 5-Hydroxymethylcytosine, 5-Formylcytosine, and 5-Carboxycytosine", *J. Am. Chem. Soc.* **2013**, *135*, 14593–14599.

[391] Z. Liutkeviciute, G. Lukinavicius, V. Masevicius, D. Daujotyte, S. Klimasauskas, "Cytosine-5-methyltransferases add aldehydes to DNA", *Nat. Chem. Biol.* **2009**, *5*, 400–402.

[392] C.-C. Chen, K.-Y. Wang, C.-K. J. Shen, "The Mammalian de Novo DNA Methyltransferases DNMT3A and DNMT3B Are Also DNA 5-Hydroxymethylcytosine Dehydroxymethylases", *J. Biol. Chem.* **2012**, *287*, 33116–33121.

[393] C.-C. Chen, K.-Y. Wang, C.-K. J. Shen, "DNA 5-Methylcytosine Demethylation Activities of the Mammalian DNA Methyltransferases", *J. Biol. Chem.* **2013**, *288*, 9084–9091.

[394] Z. Liutkevičiutė, E. Kriukienė, J. Ličytė, M. Rudytė, G. Urbanavičiutė, S. Klimašauskas, "Direct Decarboxylation of 5-Carboxylcytosine by DNA C5- Methyltransferases", *J. Am. Chem. Soc.* **2014**, *136*, 5884–5887.

[395] R. J. Klose, K. Yamane, Y. Bae, D. Zhang, H. Erdjument-Bromage, P. Tempst, J. Wong, Y. Zhang, "The transcriptional repressor JHDM3A demethylates trimethyl histone H3 lysine[thinsp]9 and lysine[thinsp]36", *Nature* **2006**, *442*, 312–316.

[396] S. M. Kooistra, K. Helin, "Molecular mechanisms and potential functions of histone demethylases", *Nat. Rev. Mol. Cell Biol.* **2012**, *13*, 297–311.

[397] S. Markolovic, T. M. Leissing, R. Chowdhury, S. E. Wilkins, X. Lu, C. J. Schofield, "Structure–function relationships of human JmjC oxygenases—demethylases versus hydroxylases", *Curr. Opin. Struct. Biol.* **2016**, *41*, 62–72.

[398] J. Nieminuszczy, E. Grzesiuk, "Bacterial DNA repair genes and their eukaryotic homologues: 3. AlkB dioxygenase and Ada methyltransferase in the direct repair of alkylated DNA", *Acta Biochim. Pol.* **2007**, *54*, 459–468.

[399] 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.

[400] C. K. Liu, C. A. Hsu, M. T. Abbott, "Catalysis of three sequential dioxygenase reactions by thymine 7-hydroxylase", *Arch. Biochem. Biophys.* **1973**, *159*, 180–187.

[401] J. A. Smiley, M. Kundracik, D. A. Landfried, S. Barnes, V. R., 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*, 256–64.

[402] S. Xu, W. Li, J. Zhu, R. Wang, Z. Li, G.-L. Xu, J. Ding, "Crystal structures of isoorotate decarboxylases reveal a novel catalytic mechanism of 5-carboxyl-uracil decarboxylation and shed light on the search for DNA decarboxylase", *Cell Res.* **2013**, *23*, 1296–1309.

[403] S. Schiesser, B. Hackner, T. Pfaffeneder, M. Müller, M.ller, C. Hagemeier, 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.

[404] 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.* **2017**, accepted.

[405] 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*, 1117–25.

[406] 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 14alpha-demethylase (CYP51) from Leishmania infantum", *J. Biol. Chem.* **2011**, *286*, 26838–48.

[407] 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.

[408] K. G. Aukema, T. M. Makris, S. A. Stoian, J. E. Richman, E. Munck, 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.

[409] H. Hashimoto, Y. O. Olanrewaju, Y. Zheng, G. G. Wilson, X. Zhang, X. Cheng, "Wilms tumor protein recognizes 5-carboxylcytosine within a specific DNA sequence", *Genes Dev.* **2014**, *28*, 2304–2313.

[410] X. Li, W. Wei, Q.-Y. Zhao, J. Widagdo, D. Baker-Andresen, C. R. Flavell, A. D'Alessio, Y. Zhang, T. W. Bredy, "Neocortical Tet3-mediated accumulation of 5-hydroxymethylcytosine promotes rapid behavioral adaptation", *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 7120–7125.

[411] G. A. Kaas, C. Zhong, D. E. Eason, D. L. Ross, R. V. Vachhani, G.-l. Ming, J. R. King, H. Song, J. D. Sweatt, "TET1 Controls CNS 5-Methylcytosine Hydroxylation, Active DNA Demethylation, Gene Transcription, and Memory Formation", *Neuron* **2013**, *79*, 1086–1093.

[412] H. Yu, Y. Su, J. Shin, C. Zhong, J. U. Guo, Y.-L. Weng, F. Gao, D. H. Geschwind, G. Coppola, G.-l. Ming, H. Song, "Tet3 regulates synaptic transmission and homeostatic plasticity via DNA oxidation and repair", *Nat Neurosci* **2015**, *18*, 836–843.

[413] T. Li, D. Yang, J. Li, Y. Tang, J. Yang, W. Le, "Critical Role of Tet3 in Neural Progenitor Cell Maintenance and Terminal Differentiation", *Molecular Neurobiology* **2015**, *51*, 142–154.

[414] A. Szwagierczak, S. Bultmann, C. S. Schmidt, F. Spada, H. Leonhardt, "Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA", *Nucleic Acids Res.* **2010**, *38*, e181.

[415] Y. Xu, C. Xu, A. Kato, W. Tempel, J. Abreu, C. Bian, Y. Hu, D. Hu, B. Zhao, T. Cerovina, J. Diao, F. Wu, H. He, Q. Cui, E. Clark, C. Ma, A. Barbara, G. Veenstra, G. Xu, U. Kaiser, X. Liu, S. Sugrue, X. He, J. Min, Y. Kato, Y. Shi, "Tet3 CXXC Domain and Dioxygenase Activity Cooperatively Regulate Key Genes for Xenopus Eye and Neural Development", *Cell* **2012**, *151*, 1200–1213.

[416] A. Perera, D. Eisen, M. Wagner, S. Laube, A. Künzel, S. Koch, J. Steinbacher, E. Schulze, V. Splith, N. Mittermeier, M. Müller, M. Biel, T. Carell, S. Michalakis, "TET3 Is Recruited by REST for Context-Specific Hydroxymethylation and Induction of Gene Expression", *Cell Rep* **2015**, *11*, 283–294.

[417] B. W. Carey, L. W. S. Finley, J. R. Cross, C. D. Allis, C. B. Thompson, "Intracellular alpha-ketoglutarate maintains the pluripotency of embryonic stem cells", *Nature* **2014**, *518*, 413.

[418] J. Cox, M. Y. Hein, C. A. Luber, I. Paron, N. Nagaraj, M. Mann, "Accurate Proteome-wide Label-free Quantification by Delayed Normalization and Maximal Peptide Ratio Extraction, Termed MaxLFQ", *Mol. Cell Proteomics* **2014**, *13*, 2513–2526.

[419] P. Vella, A. Scelfo, S. Jammula, F. Chiacchiera, K. Williams, A. Cuomo, A. Roberto, J. Christensen, T. Bonaldi, K. Helin, D. Pasini, "Tet Proteins Connect the O-Linked N-acetylglucosamine Transferase Ogt to Chromatin in Embryonic Stem Cells", *Molecular Cell* **2013**, *49*, 645–656.

[420] B. S. Meldrum, "Glutamate as a Neurotransmitter in the Brain: Review of Physiology and Pathology", *J. Nutr.* **2000**, *130*, 1007.

[421] S. Fredriksson, M. Gullberg, J. Jarvius, C. Olsson, K. Pietras, S. M. Gustafsdottir, A. Ostman, U. Landegren, "Protein detection using proximity-dependent DNA ligation assays", *Nat Biotech* **2002**, *20*, 473–477.

[422] T. Nishino, K. Morikawa, "Structure and function of nucleases in DNA repair: shape, grip and blade of the DNA scissors", *Oncogene* **2002**, *21*, 9022–9032.

[423] E. S. Henle, S. Linn, "Formation, Prevention, and Repair of DNA Damage by Iron/Hydrogen Peroxide", *J. Biol. Chem.* **1997**, *272*, 19095–19098.

[424] W. K. Pogozelski, T. D. Tullius, "Oxidative Strand Scission of Nucleic Acids:-Routes Initiated by Hydrogen Abstraction from the Sugar Moiety", *Chem. Rev.* **1998**, *98*, PMID: 11848926, 1089–1108.

[425] A. C. Mello-Filho, R. Meneghini, "Iron is the intracellular metal involved in the production of DNA damage by oxygen radicals", *Mutat. Res. Fundam. Mol. Mech. Mutagen.* **1991**, *251*, 109–113.

[426] D. R. Lloyd, D. H. Phillips, "Oxidative DNA damage mediated by copper(II), iron(II) and nickel(II) Fenton reactions: evidence for site-specific mechanisms in the formation of double-strand breaks, 8-hydroxydeoxyguanosine and putative intrastrand cross-links", *Mutat. Res. Fundam. Mol. Mech. Mutagen.* **1999**, *424*, 23–36.

[427] M. Pitié, G. Pratviel, "Activation of DNA Carbon-Hydrogen Bonds by Metal Complexes", *Chemical Reviews* **2010**, *110*, PMID: 20099805, 1018–1059.

[428] I. Guillemin, M. Becker, K. Ociepka, E. Friauf, H. G. Nothwang, "A subcellular prefractionation protocol for minute amounts of mammalian cell cultures and tissue", *PROTEOMICS* **2005**, *5*, 35–45.

[429] N. A. Kulak, G. Pichler, I. Paron, N. Nagaraj, M. Mann, "Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells", *Nat Meth* **2014**, *11*, 319–324.

[430] K. V. S. N. Murty, R. Pal, K. Datta, D. Mal, "Glucose Promoted Claisen Rearrangement of 1-Allyloxy Anthraquinones", *Synth. Commun.* **1994**, *24*, 1287–1292.

[431] H. Sharghi, G. Aghapour, "Claisen Rearrangement of Allyloxyanthraquinones with Silver/Potassium Iodide in Acetic Acid as a New and Efficient Reagent", *J. Org. Chem.* **2000**, *65*, 2813–2815.

[432] T. Kleppisch, A. Pfeifer, P. Klatt, P. Ruth, A. Montkowski, R. Fässler, F. Hofmann, "Long-Term Potentiation in the Hippocampal CA1 Region of Mice Lacking cGMP-Dependent Kinases Is Normal and Susceptible to Inhibition of Nitric Oxide Synthase", *J. Neurosci.* **1999**, *19*, 48–55.

[433] S. Michalakis, T. Kleppisch, S. A. Polta, C. T. Wotjak, S. Koch, G. Rammes, L. Matt, E. Becirovic, M. Biel, "Altered synaptic plasticity and behavioral abnormalities in CNGA3-deficient mice", *Genes Brain Behav.* **2011**, *10*, 137–148.

[434] M. Kim, Y.-K. Park, T.-W. Kang, S.-H. Lee, Y.-H. Rhee, J.-L. Park, H.-J. Kim, D. Lee, D. Lee, S.-Y. Kim, Y. S. Kim, "Dynamic changes in DNA methylation and hydroxymethylation when hES cells undergo differentiation toward a neuronal lineage", *Hum. Mol. Genet.* **2014**, *23*, 657–667.

[435] X. Hu, L. Zhang, S.-Q. Mao, Z. Li, J. Chen, R.-R. Zhang, H.-P. Wu, J. Gao, F. Guo, W. Liu, G.-F. Xu, H.-Q. Dai, Y. Shi, X. Li, B. Hu, F. Tang, D. Pei, G.-L. Xu, "Tet and TDG Mediate DNA Demethylation Essential for Mesenchymal-to-Epithelial Transition in Somatic Cell Reprogramming", *Cell Stem Cell* **2014**, *14*, 512–522.

[436] J. A. Vizcaíno, A. Csordas, N. del-Toro, J. A. Dianes, J. Griss, I. Lavidas, G. Mayer, Y. Perez-Riverol, F. Reisinger, T. Ternent, Q.-W. Xu, R. Wang, H. Hermjakob, "2016 update of the PRIDE database and its related tools", *Nucleic Acids Research* **2016**, *44*, D447–D456.

[437] C. Jin, Y. Lu, J. Jelinek, S. Liang, M. R. Estecio, M. C. Barton, J.-P. J. Issa, "TET1 is a maintenance DNA demethylase that prevents methylation spreading in differentiated cells", *Nucleic Acids Research* **2014**, *42*, 6956–6971.

[438] J. Ding, X. Huang, N. Shao, H. Zhou, D.-F. Lee, F. Faiola, M. Fidalgo, D. Guallar, A. Saunders, P. V. Shliaha, H. Wang, A. Waghray, D. Papatsenko, C. Sánchez-Priego, D. Li, Y. Yuan, I. R. Lemischka, L. Shen, K. Kelley, H. Deng, X. Shen, J. Wang, "Tex10 Coordinates Epigenetic Control of Super-Enhancer Activity in Pluripotency and Reprogramming", *Cell Stem Cell* **2015**, *16*, 653–668.

[439] Y. Costa, J. Ding, T. W. Theunissen, F. Faiola, T. A. Hore, P. V. Shliaha, M. Fidalgo, A. Saunders, M. Lawrence, S. Dietmann, S. Das, D. N. Levasseur, Z. Li, M. Xu, W. Reik, J. C. R. Silva, J. Wang, "Nanog-dependent function of Tet1 and Tet2 in establishment of pluripotency", *Nature* **2013**, *495*, 370–374.

[440] J. Ding, H. Xu, F. Faiola, A. Ma'ayan, J. Wang, "Oct4 links multiple epigenetic pathways to the pluripotency network", *Cell Res* **2012**, *22*, 155–167.

[441] H. Takai, K. Masuda, T. Sato, Y. Sakaguchi, T. Suzuki, T. Suzuki, R. Koyama-Nasu, Y. Nasu-Nishimura, Y. Katou, H. Ogawa, Y. Morishita, H. Kozuka-Hata, M. Oyama, T. Todo, Y. Ino, A. Mukasa, N. Saito, C. Toyoshima, K. Shirahige, T. Akiyama, "5-Hydroxymethylcytosine Plays a Critical Role in Glioblastomagenesis by Recruiting the CHTOP-Methylosome Complex", *Cell Rep* **2014**, *9*, 48–60.

[442] S. Xue, C. Liu, X. Sun, W. Li, C. Zhang, X. Zhou, Y. Lu, J. Xiao, C. Li, X. Xu, B. Sun, G. Xu, H. Wang, "TET3 Inhibits Type I IFN Production Independent of DNA Demethylation", *Cell Rep* **2016**, *16*, 1096–1105.

[443] H. Garcia, D. Fleyshman, K. Kolesnikova, A. Safina, M. Commane, G. Paszkiewicz, A. Omelian, C. Morrison, K. Gurova, "Expression of FACT in mammalian tissues suggests its role in maintaining of undifferentiated state of cells", *Oncotarget* **2011**, *2*, 783–796.

[444] K. Ahmed, H. Dehghani, P. Rugg-Gunn, E. Fussner, J. Rossant, D. P. Bazett-Jones, "Global Chromatin Architecture Reflects Pluripotency and Lineage Commitment in the Early Mouse Embryo", *PLOS ONE* **2010**, *5*, 1–13.

[445] S. Efroni, R. Duttagupta, J. Cheng, H. Dehghani, D. J. Hoeppner, C. Dash, D. P. Bazett-Jones, S. L. Grice, R. D. McKay, K. H. Buetow, T. R. Gingeras, T. Misteli, E. Meshorer, "Global Transcription in Pluripotent Embryonic Stem Cells", *Cell Stem Cell* **2008**, *2*, 437–447.

[446] E. Meshorer, D. Yellajoshula, E. George, P. J. Scambler, D. T. Brown, T. Misteli, "Hyperdynamic Plasticity of Chromatin Proteins in Pluripotent Embryonic Stem Cells", *Dev. Cell* **2006**, *10*, 105–116.

[447] A. A. Sérandour, S. Avner, F. Oger, M. Bizot, F. Percevault, C. Lucchetti-Miganeh, G. Palierne, C. Gheeraert, F. Barloy-Hubler, C. L. Péron, T. Madigou, E. Durand, P. Froguel, B. Staels, P. Lefebvre, R. Métivier, J. Eeckhoute, G. Salbert, "Dynamic hydroxymethylation of deoxyribonucleic acid marks differentiation-associated enhancers", *Nucleic Acids Res.* **2012**, *40*, 8255–8265.

[448] Y. Huang, L. Chavez, X. Chang, X. Wang, W. A. Pastor, J. Kang, J. A. Zepeda-Martínez, U. J. Pape, S. E. Jacobsen, B. Peters, A. Rao, "Distinct roles of the methylcytosine oxidases Tet1 and Tet2 in mouse embryonic stem cells", *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 1361–1366.

[449] J. I. Young, S. Züchner, G. Wang, "Regulation of the Epigenome by Vitamin C", *Annu. Rev. Nutr.* **2015**, *35*, 545–564.

[450] L. Dang, D. W. White, S. Gross, B. D. Bennett, M. A. Bittinger, E. M. Driggers, V. R. Fantin, H. G. Jang, S. Jin, M. C. Keenan, K. M. Marks, R. M. Prins, P. S. Ward, K. E. Yen, L. M. Liau, J. D. Rabinowitz, L. C. Cantley, C. B. Thompson, M. G. Vander Heiden, S. M. Su, "Cancer-associated IDH1 mutations produce 2-hydroxyglutarate", *Nature* **2009**, *462*, 739–744.

[451] J. Chen, L. Guo, L. Zhang, H. Wu, J. Yang, H. Liu, X. Wang, X. Hu, T. Gu, Z. Zhou, J. Liu, J. Liu, H. Wu, S.-Q. Mao, K. Mo, Y. Li, K. Lai, J. Qi, H. Yao,

G. Pan, G.-L. Xu, D. Pei, "Vitamin C modulates TET1 function during somatic cell reprogramming", *Nat. Genet.* **2013**, *45*, 1504–1509.

[452] X. Yang, K. Qian, "Protein O-GlcNAcylation: emerging mechanisms and functions", *Nat. Rev. Mol. Cell Biol.* **2017**, *18*, 452–465.

[453] K. M. Harlen, L. S. Churchman, "The code and beyond: transcription regulation by the RNA polymerase II carboxy-terminal domain", *Nat. Rev. Mol. Cell Biol.* **2017**, *18*, 263–273.

[454] J.-S. Seeler, A. Dejean, "SUMO and the robustness of cancer", *Nat. Rev. Cancer* **2017**, *17*, 184–197.

[455] B. R. Sabari, D. Zhang, C. D. Allis, Y. Zhao, "Metabolic regulation of gene expression through histone acylations", *Nat. Rev. Mol. Cell Biol.* **2017**, *18*, 90–101.

[456] A. Hirano, Y.-H. Fu, L. J. Ptacek, "The intricate dance of post-translational modifications in the rhythm of life", *Nat. Struct. Mol. Biol.* **2016**, *23*, 1053–1060.

[457] Y.-C. Wang, S. E. Peterson, J. F. Loring, "Protein post-translational modifications and regulation of pluripotency in human stem cells", *Cell Res.* **2014**, *24*, 143–160.

[458] J.-P. Hsin, J. L. Manley, "The RNA polymerase II CTD coordinates transcription and RNA processing", *Genes Dev.* **2012**, *26*, 2119–2137.

[459] H. P. Phatnani, A. L. Greenleaf, "Phosphorylation and functions of the RNA polymerase II CTD", *Genes Dev.* **2006**, *20*, 2922–2936.

[460] U. Rothbauer, K. Zolghadr, S. Muyldermans, A. Schepers, M. C. Cardoso, H. Leonhardt, "A Versatile Nanotrap for Biochemical and Functional Studies with Fluorescent Fusion Proteins", *Mol. Cell. Proteomics* **2008**, *7*, 282–289.

[461] 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", *Chem. Commun.* **2016**, *52*, 14361–14364.

Ja, wir werden alles, alles noch einmal in Frage stellen. Und wir werden nicht mit Siebenmeilenstiefeln vorwärtsgehen, sondern im Schneckentempo. Und was wir heute finden, werden wir morgen von der Tafel streichen und erst wieder anschreiben, wenn wir es noch einmal gefunden haben. Und was wir zu finden wünschen, das werden wir, gefunden, mit besonderem Misstrauen ansehen. Also werden wir an die Beobachtung der Sonne herangehen mit dem unerbittlichen Entschluss, den Stillstand der Erde nachzuweisen! Und erst wenn wir gescheitert sind, vollständig und hoffnungslos geschlagen und unsere Wunden leckend, in traurigster Verfassung, werden wir zu fragen anfangen, ob wir nicht doch recht gehabt haben und die Erde sich dreht! Sollte uns aber dann jede andere Annahme als diese unter den Händen zerronnen sein, dann keine Gnade mehr mit denen, die nicht geforscht haben und doch reden.

Bertolt Brecht, Leben des Galilei